Nirmal Kumar Betchoo
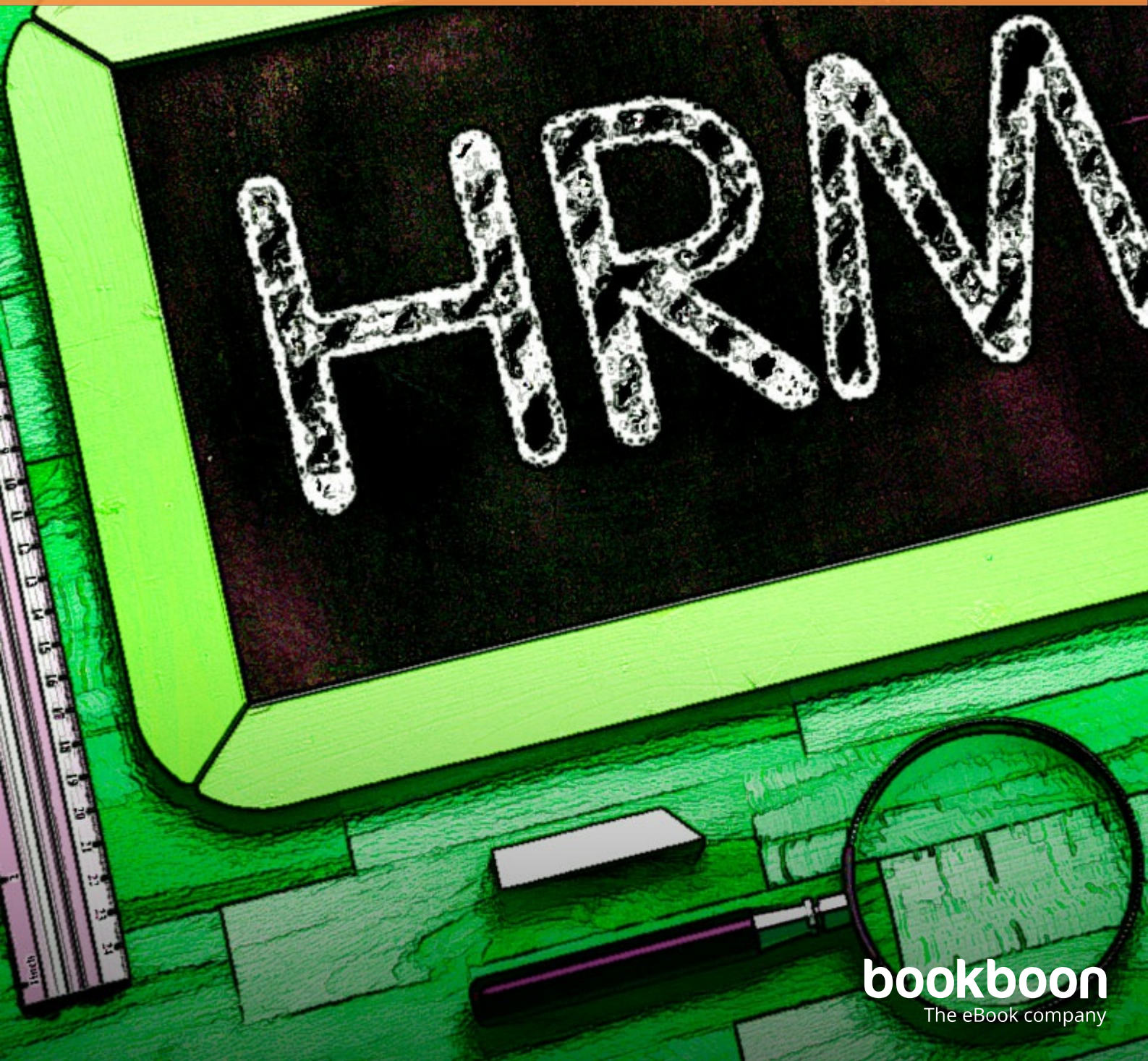
Numerical Analysis in HRM

The Research Toolkit

NIRMAL KUMAR BETCHOO

NUMERICAL ANALYSIS IN HRM
THE RESEARCH TOOLKIT

CONTENTS

PREFACE

This book titled 'Numerical Analysis in HRM, The Research Toolkit' is presented as a sequel to the book 'Applied Research in HRM' where the former publication was focused on working out qualitative data in human resource management. This book can be either independently or concurrently used with the other publication as its focus is on quantitative or numerical analysis. It is common to say that as students mature and move to more advanced learning concepts, they must also master abstract notions that are usually quantitative in nature. Basically, the first idea that comes to the mind is that information might be referenced from any statistical publication or from Internet sources.

To better address the need of advanced learners and practitioners of human resource management, this publication comes as a toolkit for researchers. Quantitative analysis is firstly complex to manipulate and interpret compared to qualitative data but, in essence, better provides scientific insight into research. Technically, a numerical synthesis of information might offer evidence of well undertaken and debated research. The fact that numerical values are used to interpret information or support hypotheses proves that the mastering of quantitative concepts is very important to HRM research.

It is an assumption that social sciences students and, in this perspective, human resource students seem to be more versed in qualitative than quantitative data. This perception is generalised and could appear like biased when it is known that the human resource function englobes the learning, assimilation and mastering of mathematical concepts and numerical data that help in better displaying and understanding information.

This publication aims at providing the needed guidance to HRM students by taking into consideration that their level of understanding of mathematical concepts might be weaker than those in the accounting or scientific area. To counter this view, the book exposes each quantitative concept in a simple and logical manner allowing the student firstly grasp basic statistical information and learn how to tackle problems that are linked with the discipline.

The book then moves to hypothesis formulation and testing by depicting the choice between the proposed hypothesis and the null hypothesis. It progresses on developing concepts like the normal distribution curve, the t- and the z-tests along with advanced concepts like the Chi-Square, ANOVA, etc. There are additional chapters covering concepts like probability and estimated values, regression and correlation. To assist in effective research planning, current concepts like the GANTT Chart and the Fishbone diagram are also addressed.

The quantitative aspect combined with qualitative research could be the platform to develop the mixed-methods research. Since HRM effectively combines both research types, this book will help in blending concepts learnt in the former book 'Applied Research in HRM' with this easy-to-read textbook. This publication addresses the needs and expectations of students, in an international context, willing to master HRM through the scientific and numerical approach. It is hoped that this painstaking work of data manipulation and interpretation addresses the needs for both undergraduate and advanced learners in HRM today.

NOTES FROM THE AUTHOR

This publication 'Numerical Analysis in HRM, The Research Toolkit' is the fourth collaboration of the author with Ventus Publishing after the successful spell of 'Managing Workplace Diversity' and 'Public Sector Management: A Millennial Insight'. All these publications have been carefully and painstakingly written for tertiary level students with the intention of introducing the concepts in a simple and straightforward manner without much frills. They look more like an instructor's resource material offered in a university compared with a recommended textbook.

This book titled 'Numerical Analysis in HRM, The Research Toolkit' is offered free online and aligns with the author's aim to share knowledge in a free world. It is heart fulfilling to learn that the past two books have been archived in some official US websites, used in various places like India, Myanmar, Eastern Europe and the West Indies with lecturers and professors from Croatia, Latvia, Kazakhstan or Poland reading the materials through the portal Reserachgate. It is hoped that this endeavour will also accompany this book given that it is targeted to international social sciences students who are all too easily upset by research analysis. This is what the author, with long experience in teaching research techniques, has sought to do and, with the collaboration of the publishing director, Mrs Karin Jakobsen, hopes that a learning facility is offered freely worldwide in a bold and constantly evolving educational environment.

Nirmal Kumar Betchoo
August 2017

NOTES FROM THE REVIEWER

Reviewing '*Numerical Analysis in HRM*' is a privilege. George Orwell, William Strunk, and Elwyn Brooks White, among others, inspired grammatical and stylistic changes. Such changes make sentences active, positive, and shorter. Words are plainer by letter or syllable to increase clarity and boldness. Simple verbs replace phrases, vivid verbs adverbs.

Hope readers find trustworthy views for understanding human resource management with depth and breadth. May my colleague and associate's book help think, speak, and act about this topic with reliability, proficiency, honesty, and integrity.

Cyrus Jones, Graduate, Johns Hopkins University
July 2017

1 QUANTITATIVE TECHNIQUES IN HRM

If the major part of this book has been devoted to applied research in the qualitative or descriptive form, another useful component in HR research is quantitative. In HR, researchers consider the mixed method technique's proficiency comprising a combination of theory and figures. By the way, the HR function deals with the recording, assessment, and evaluation of data presented to the human resource information system, be it either paper documents or digital data.

Quantitative techniques involve using figures, statistical techniques, mathematical principles, and determining hypotheses. These demand a certain level of mathematical proficiency from the researcher applies her knowledge to a social science like human resource management. A variety of techniques should come to use.

Quantitative techniques contribute to the scientific approach in research by partly overcoming displaying too much non-mathematical information. It is a known fact diagrams, tables, and charts display information in a more meaningful way by synthesising information and using few words. In this way, information is displayed fast, with proficiency. Quantitative data has also relevant as such information must be presented in the most precise way. Information can then be tabulated, forecasted, and interpreted from such scientific information.

A useful argument favouring quantitative data comes from HR students might be stereotyped as innumerate. This is not an idiosyncrasy of such students but rather a perception behavioural sciences often bear the task of overly using qualitative information. This idea can be contested as the same sciences need numerical data to support arguments.

This section briefly addresses quantitative data without being exhaustive. It broadly explains the use and application of qualitative data in research bur purports that such information, if correctly used, provides better insights both to the presentation and discussion of research. In research findings, data analysis, and presentation, quantitative information is of paramount importance. Please never underestimate its use.

This section also covers the statistical techniques used in quantitative research namely hypothesis formulation, the application of t and z-tests, the use of ANOVA, among others, to help HR students become familiar with such terms but also use them in the most ingenuous manner. This cannot be accepted a software simply does the calculation for the student to the expense of himself understanding why and how the formula has worked and under which condition it applies. Without being too long, the basic techniques are introduced in the following pages.

Mean, Mode and Median

In statistics, mean is the average of all the data, this is the 'central' value of a set of numbers. Median is the middle value of the set of data and mode is the one that occurs most often in a set of data. The difference between the largest and smallest data is the range. This online mean median mode calculator helps you to calculate average and range, ascending order of the numbers, mean, median and mode of the given data.

Example:

The average number of hours of work for 6 employees in the production department are as follows:

$$42,38,39,34,41,45$$

The mean is the sum of all numbers divided by the total number of employees;

42+38+39=34+41=45 divided by 6 = 39.83333

To 2 decimal places, the value will be 39.83
To 1 decimal place this will be 39.8

The range is 11 meaning difference between 34 and 45.

The mean value implies the average number of hours worked by the employees. Such a mean value infers despite lower and higher working hours, the trend is 39.8 for the employees in general. If a bell-shaped curve were drawn, the values would be represented as:
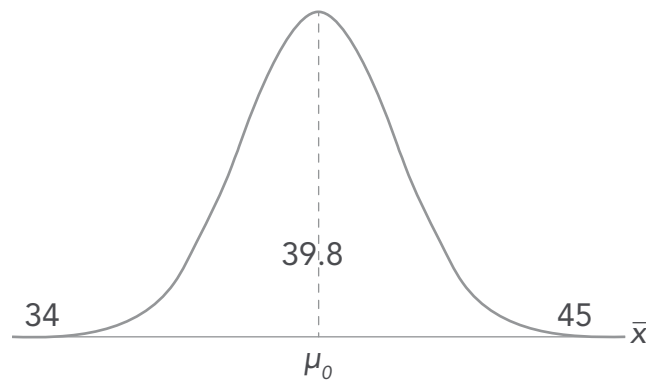
The **mean** of a sample or a population is computed by adding all of the observations and dividing by the number of observations.

In the general case, the mean can be calculated, using one of the following equations:

Population mean = μ = $\sum X$ / N OR Sample mean = x = $\sum x$ / n

where $\sum X$ is the sum of all the population observations, N is the number of population observations, $\sum X$ is the sum of all the sample observations, and n is the number of sample observations.

Mode

The mode is the most frequently appearing value in a population or sample. Suppose a sample of five workers is drawn and their wage is assessed as follows in Euros:

1000, 1200, 1000,1200, 1000, 1350, 1000, 1225.

The most common occurrence is 1,000 € because it occurs four times in this model. Remember that modes are more realistic when their occurrences exceed others. If there is just one more than another occurrence, the modal value is weak.

Application of the mode:

- Modal wages of employees.
- Modal number of hours worked.
- Modal number of documents processed by an employee.

Median

The median is a simple measure of central tendency. To find the median, one arranges the observations in order from smallest to largest value. If there is an odd number of observations, the median is the middle value. If there is an even number of observations, the median is the average of the two middle values.

In a sample of four employees, one might want to compute the median annual income. Suppose the incomes are €2,500 for the first employee; €3,500 for the second; €2,800 for the third; and €2,900 for the fourth.

Arranging in order is as follows:

2500, 2800, 2900 and 3500

There is no exact middle value from even numbers, here there are 4 counts only.

The median is the 2nd and the 3rd figure combined: 2800 + 2900 /2 = €2850

In case if there are odd numbers, the middle number is easily obtained. For instance, we add 4000.
Rearranging in order is as follows:

2500, 2800, 2900, 3500, 4000

The median or middle number is better spotted as: 2900
The median is €2,900.

It infers that the middle salary level is €2,900 in this group of employees.

Quartile concept

Quartiles divide a rank-ordered data set into four equal parts. The values dividing each part are called the first, second, and third quartiles; and are denoted by Q_1, Q_2, and Q_3, respectively. The concept of quartile helps a human resource manager effectively divide a range of data into equal units and highlighting which one is the lowest and which one is the highest.

In pay structures, the quartile concept works. For instance, top managers will be in the upper quartile, operatives might be in the lower quartile and middle level employees are in the second quartile.

Wage determination takes place in many societies in this way and generally, the quartiles help in sorting out whether a minimal wage can be provided or not and at which level.

Example

Selective list of wages in an organisation in euros.

2000, 3000, 2800, 2600,3200,4000.

First principle, the middle value or median should be located.
Rearranging the values:

20000, 2600, 2800, 3000, 3200, 3400

In this case, there are even values (6), the median is
2800+3000/2 = 2900

The set is sliced:

20000, 26|00, 2800, | 3000, 32|00, 3400

The lower quartile is €2600.
The upper quartile is €3200.

The inter-quartile range is 3200-2600=€600.

The red marks represent where the slices have been exactly cut.

Another example

Now, let us add €2400 in the same line which now becomes an odd count (7)

The trend is displayed as follows:

20000, 2400, 2600, 2800, 3000, 3200, 3400

The median value is clear cut: 2800

The set is now sliced as follows:

20000, 24|00, 2600, 28|00, 3000, 32|00, 3400

The lower quartile is €2400.
The upper quartile is €2800.
The inter-quartile range is 2800-2400=€400.

Averaging values

There might be a condition where values could be averaged and not just spotted through slicing. For instance, there are just four values provided.

2000, 3000, 2800, 2600

Each are re-arranged:

2000, 2600, 2800,3000

The set is sliced:

2000, **|** 2600, **|** 2800, **|** 3000

The median is 2600+ 2800/2 = €2700.
The lower quartile is 2000+2600/2 = €2300.
The upper quartile is 2800 + 3000/2 = €2900.

Standard deviation

Standard deviation is a numerical value to indicate how widely individuals in a group vary. If individual observations vary greatly from the group mean, the standard deviation is large and vice versa.

Please distinguish the standard deviation of a population and the standard deviation of a sample. Each have different notations and are computed differently. The standard deviation of a population is denoted by σ and the standard deviation of a sample, by *s*.

The standard deviation of a population is defined by the following formula

$$\sigma = \sqrt{[\sum (X_i - X)^2 / N]}$$

where σ is the population standard deviation, X is the population mean, X_i is the *i*th element from the population, and N is the number of elements in the population.

The standard deviation of a sample is defined by slightly different formula

$$s = \sqrt{[\sum (x_i - x)^2 / (n - 1)]}$$

where s is the sample standard deviation, x is the sample mean, x_i is the ith element from the sample, and n is the number of elements in the sample.

The standard deviation is equal to the square root of the variance.

Example

The wages of the employees are:

2000, 2400, 2600, 2800, 3000, 3200, 3400

Calculation of standard deviation:

Mean value is 2771.

Methodology: $(2000-2771)^2 + (2400-2771)^2 + (2600-2771)^2 + (2800-2771)^2 + (3000-2771)^2 + (3400-2771)^2$ divided by 7 for the population SD and divided by 7-1 for the sample SD and finally square rooted for both.

The values stand as:

Total Numbers, N=7
Sum = 19400
Mean = 2771
Sample standard deviation: s = 482
Population Standard Deviation, σ = 446.

Variance

The variance is a numerical value used to indicate how widely individuals in a group vary. It is important to distinguish between the variance of a population and the variance of a sample. Each have different notations, and are computed differently.

The variance of a population is denoted by $σ^2$; and the variance of a sample, by s^2.
The variance of a population is defined by the following formula:

$$σ^2 = \sum (Xi - X)^2 / N$$

where $σ^2$ is the population variance, X is the population mean, X_i is the ith element from the population, and N is the number of elements in the population.

The variance of a sample is defined by slightly different formula:

$$s^2 = \sum (xi - x)^2 / (n - 1)$$

where s^2 is the sample variance, x is the sample mean, x_i is the ith element from the sample, and n is the number of elements in the sample. Using this formula, the variance of the sample is an unbiased estimate of the variance of the population.

The variance equals the square of the standard deviation. Researchers use the variance to get raw estimates prior to studying the standard deviation.

Example

A set of data is provided for employees working for one week in HR company

40, 41, 42, 43, 47, 48, 50, 52

The population size is eight.
The mean is 45.35 to 2 d.p.
The Variance (σ^2) is 17.48 to 2 d.p.

The figure varies if the set is a population, i.e. all the employees or a sample; a selection of employees in the firm.

The sample variance is 19.98 to 2 d.p.

References

Alcula.com (2017) http://www.alcula.com/calculators

Easycalculation (2017) https://www.easycalculation.com

Miniwebtool.com (2017) http://www.miniwebtool.com

Multiple-Choice Questions

1. An essential averaged value used in quantitative research is the
   A. mode.
   B. median.
   C. quartile.
   D. mean.

2. The median value in the following set 1000,2000,3000,4000 is:
   A. 2100.
   B. 2000.
   C. 2500.
   D. 3000.

3. The modal value in the following set 4,4,5,5,5,6,6,7,8 is
   A. 5.
   B. 4.
   C. 6.
   D. 7.

4. The lower quartile in the following set 1200, 1300, 1400, 1500, 1600 is
   A. 1200.
   B. 1250.
   C. 1550.
   D. 1600.

5. The upper quartile in the following set 1200, 1300, 1400, 1500, 1600, 1700 is
   A. 1600.
   B. 1300.
   C. 1450.
   D. 1500.

6. For a population of eight, the N value in standard deviation is
   A. not stated.
   B. unknown.
   C. 7.
   D. 8.

7. For a sample of 15, the number value in standard deviation is
   A. not stated.
   B. unknown.
   C. 14.
   D. 15.

8. A squared value of the standard deviation is
   A. variance.
   B. median.
   C. sum of squares.
   D. square root.

9. A mean value of 4.187777 to 2 decimal places is
   A. 4.1.
   B. 4.19.
   C. 4.2.
   D. 4.18.

10. A small standard deviation means that the set of data is
   A. close.
   B. distant.
   C. dispersed.
   D. deviated.

Answers: 1D 2C 3A 4B 5A 6D 7C 8A 9B 10A

2 THE NORMAL DISTRIBUTION CURVE CONCEPT

A major component of research and hypothesis testing is based on the use of the normal distribution or the Poisson distribution curve. This concept is used in most parametric and non-parametric tests with the perception all elements of a population could be within the distribution. Researchers tend to say respondents or items under observation are normally distributed. The normal distribution curve is a wonderful concept used in research simply because of its general application.

For instance, students participating in an examination are expected to be under normal distribution. A very low percentage might fail the examination while the performance progresses and a majority of the students attain an acceptable pass rate. There will also be a few students scoring distinction and they will progress to the other end of the curve known as a tail.

Apart from students, many variables are easily included in the normal distribution curve although there might be criticisms to it. For example, the performance of businesses in any sector could be assessed in the same way. A few companies might be high flyers in business, some others could be underperforming and the majority could be operating just at an average level.

The normal distribution curve does not rightly mean that at one tail – the left side, everything is counted as very weak or near zero. This is a misconception because is only the lower level that is counted. Say, for a test, a large number of students score 70 or above. The middle value or mean is estimated at 72.5 and the lowest mark is 65. This also fits in the normal distribution curve with the left tail denoted with 65, the right tail at, say, 85 and the middle value at 72.5. Still, the concept applies probably with a fairly low variance and standard deviation.

There is also the concept of skewness of the distribution curve especially when abnormal situations occur. During the 2008 financial crisis 2008 that stormed markets, company performance declined drastically. In this context, the performances were positively skewed with mean performances moving to the left leaving a long tail to the right.

In the class example, if students score much higher than the average value, performance will also be skewed provided there is greater tendency to move to higher marks. In this case, there will be a negative skew with mean marks exceeding the median value. These are exceptions however.

This section addresses the normal distribution curve, the confidence intervals, skew concepts and how the curve later helps in the development of hypotheses for research.

Normal distribution

Normal distribution is the most important and most widely used distribution in statistics. This is sometimes called the bell curve although the tonal qualities of such a bell would be less than pleasing. It is also called the 'Gaussian curve' after the mathematician Karl Friedrich Gauss.

**Normal Curve
Standard Deviation**
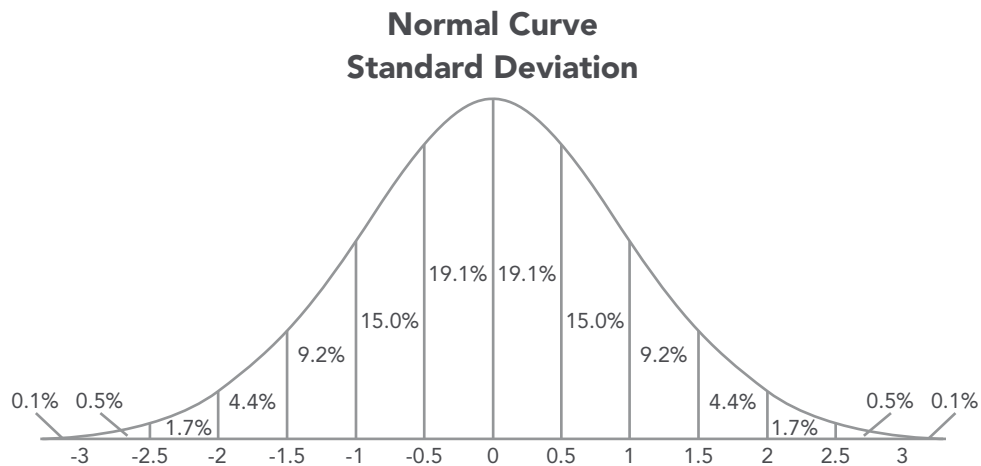
The diagram also denotes standard deviations as they move from the mean. If the curve is equally distributed according to the deviations of, say 0.5, the maximum deviations are + and -3. Generally, all elements or subjects fit in the curve but there might be significance if the values of the standard deviation reach + 0r − 2.5.

The concept of the tail

An important tool for analysis remains the tail. The normal distribution curve has two tails, one to the right and one to the left. Positive values or influences generally meet the right tail; then weaker values will be on the left tail. Note when the significant deviation is zero, there is the mean value only under normal distribution. Researchers will use the values of standard deviation and check them in a table to see how significant they might be the moment when the values sought for research enter the critical zone of + or − 2.5.

Tests could be carried out by any researcher in both directions. Suppose the outcome of a test is unknown, Researchers can choose a two-tailed test to see in which direction the trend moves. This could be the effect of implementing shift time at work and the relation with employees' productivity.

If employees are productive with a shift system, the test should be done on the right tail; a positive value.

If employees get discouraged and become counterproductive, the test should be done to the left tail with lower or negative value.

If the effect is unknown or non-estimated, the two-tailed test is applied just to see in which direction will the values and the standard deviation move.
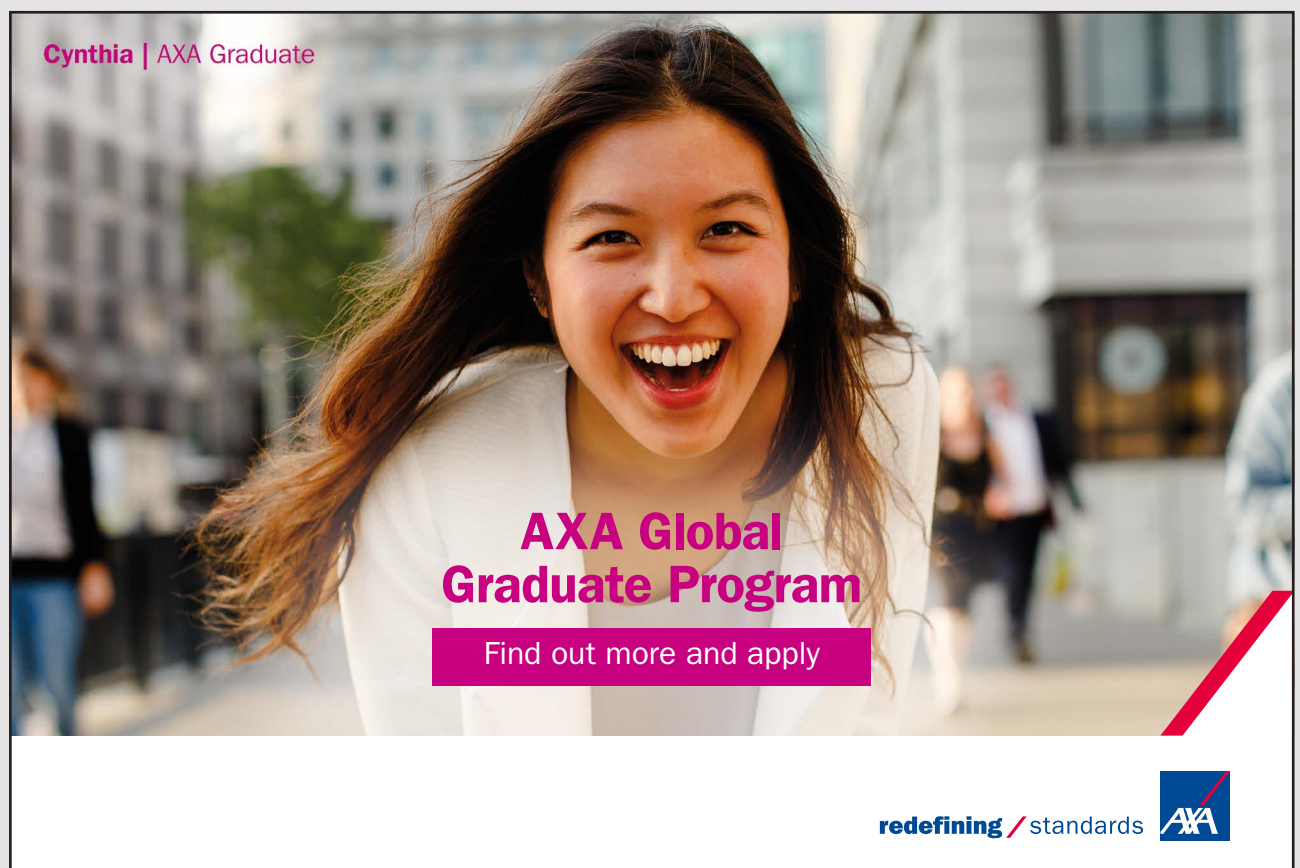
Further concepts

Normal distributions are symmetric around their mean. The mean, median, and mode of a normal distribution are equal. The area under the normal curve is equal to 1.0.

Normal distributions are denser in the centre and less dense in the tails.

Normal distributions are defined by two parameters, the mean (μ) and the standard deviation (σ).

68% of the area of a normal distribution is within one standard deviation of the mean (1 SD).

About 95% of the area of a normal distribution is within two standard deviations of the mean (2 SD)
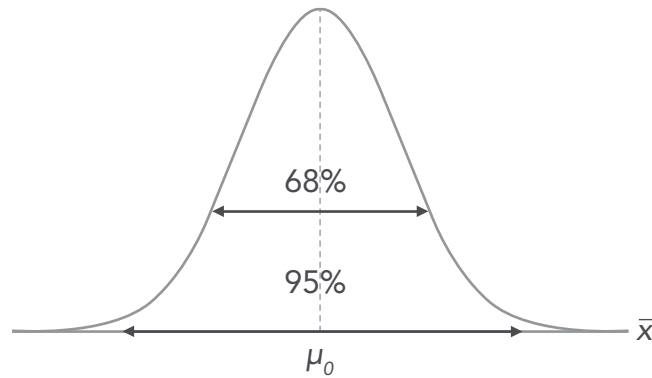
The normal distribution is important because this describes the statistical behaviour of many real-world events. For example, performance of businesses, general working hours of the workforce in a particular country, management practices in organisations ranging from laisser-faire to autocratic.

The shape of a normal distribution is determined by the mean and standard deviation. The standard normal distribution is the normal distribution with a mean of zero and a standard deviation of one. When the mean value moves either right or to the left, then the standard deviation increases or decreases.

The Normal Distribution has no skew. A Normal Distribution is not skewed.
It is perfectly symmetrical and the Mean is exactly at the peak.

Mean=Mode=Median
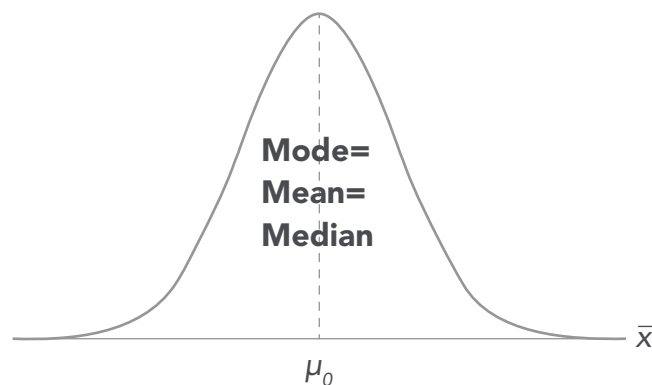
The normal random variable

The normal random variable of a standard normal distribution is called a **standard score** or a **z-score**. The normal random variable *X* from any normal distribution can be transformed into a *z* score from a standard normal distribution via the following equation:

$$z = (X - \mu) \; / \; \sigma$$

where *X* is a normal random variable, μ is the mean, and σ is the standard deviation. As any normal random variable can be changed into a *z* score, the standard normal distribution provides a useful frame of reference. This is the normal distribution that generally appears in the appendix of statistics textbooks.

Sample question on the normal random variable

The HRM Company has found that an average job lasts 1000 hours with a standard deviation of 100 hours. Assume working hours are normally distributed. What is the probability a randomly selected work lasts 1200 hours or less?

Solution

The mean score is 1000.
The standard deviation is 100.
The normal random variable, for which we want to find a cumulative probability, is 1200.

Normal variable X = 1200, Mean 1000, SD 100

Z value = 1200-1000/100 = 2.

Curve skewness

Data can be skewed, meaning this tends to have a long tail on one side or the other. Measures of central location and spread are useful for summarising a distribution of data. They also facilitate the comparison of two or more sets of data. Yet not every measure of central location and spread is well suited to each data set. For example, because the normal distribution – or bell-shaped curve – is symmetrical, the mean, median, and mode have the same value. Still in practice observed data rarely approach this ideal shape. As a result, the mean, median, and mode usually differ.

Negative Skew

A curve – see below – could be negatively skewed. Researchers sometimes say it is 'skewed to the left'. This can be better recalled as the left peak as the mean is also on the left of the peak. In the normal curve, as stated earlier, the mean=mode-median.

In the negatively skewed curve, the mode will represent the peak value, the mean value moves to the left and the median value is slightly greater but still less than the mode. Note as a principle, the mode is the peak value.

Example

Due to an adverse condition, workers' productivity has declined considerably. The modal or most common value is say 30 compared to 50. Then the mean value could be 25, the median value 27.
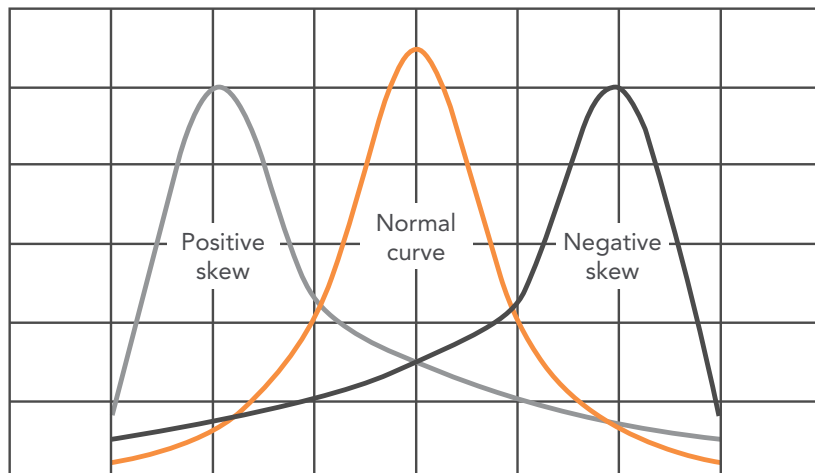
Positive Skew

Another curve – see above – could be positively skewed. Researchers sometimes say it is 'skewed to the right'. This can be better recalled as the right peak as the mean is also on the left of the peak. In the normal curve, as stated earlier, the mean=mode-median.

In the positively skewed curve, the mode represents peak value, the mean value moves to the right, and the median value is slightly greater but still less than the mode. Note as a principle, the mode is the peak value in this case as well.

Example

Due to a positive condition, workers' productivity improved considerably. The modal or most common value is say 70 compared to 50. The mean value could be 75, the median value 72. The skewness normally represents a positive condition.

Skewness calculation

The following outputs are recorded by employees:

$$20, 60, 10, 10, 20, 30$$

Using a statistical calculator, the following findings are obtained.

Mean: 25, Variance: 291.7, Standard deviation: 17.07
Range: 50, Median: 20
Mean deviation: 13,33, Skewness: 1.204

The curve is slightly skewed to the right but is a negative skew.

The Confidence Interval concept

A confidence interval gives an estimated range of values which is likely to include an unknown population parameter, the estimated range being calculated from a given set of sample data. Confidence intervals are expected to show limits of accuracy in a project and therefore explain how much a prediction is valuable or not. Within the confidence limits, a lower limit might suggest a weaker performance and the same goes for a higher limit. This could come the prediction of a poll or any survey undertaken in an HR department.

A confidence interval pushes the comfort threshold of user researchers and managers. To compute a 95% confidence interval, a researcher needs three pieces of data:-

- The mean (for continuous data) or proportion.
- The standard deviation which describes how dispersed the data is around the average.
- The sample size.

Confidence limits

Confidence limits are the lower and upper boundaries/values of a confidence interval, that is, the values which define the range of a confidence interval.

The upper and lower bounds of a 95% confidence interval are the 95% confidence limits. These limits may be taken for other confidence levels, for example, 90%, 99%, 99.9%.

The confidence level is the probability value associated with a confidence interval. It is often expressed as a percentage. For example, 0.05 confidence value =5%. Then the confidence level is equal to (1-0.05) = 0.95, i.e. a 95% confidence level.

Suppose an opinion poll predicted that, if the election were held today, the Worker's Party would win 60% of the vote.

The researcher might attach a 95% confidence level to the interval 60% plus or minus 3%.

That is, she thinks the Worker's Party may likely get between 57% and 63% of total vote.

Answer

$$60 + 3 = 63\%$$
$$60-3 = 57\%$$
$$95\% \text{ CI: } 0.57 \leq \text{CI} \geq 0.63$$

Continuous data example

Imagine a researcher asked 50 respondents how satisfied each were with recent experience of your research theme: Motivation on a 4 Likert-point scale, with 1 = not at all satisfied and 4 = extremely satisfied.

Find the mean by adding the scores for each of the 50 respondents and divide by the total responses – 50.

For this example, there is an average response of six. Compute the standard deviation. There is a sample standard deviation of 1.2.

- Compute the standard error (S_e) by dividing the standard deviation by the square root of the sample size: $1.2/ \sqrt{(50)} = 0.17$.
- Compute the margin of error by multiplying the standard error by 2. [17 × 2 = 0.34.]
- Compute the confidence interval by adding the margin of error to the mean from Step 1 and then subtracting the margin of error from the mean:

6+.34=6.34
6-.34=5.66

95% CI: 5.66≤ CI ≥6.34

Another example

Scores of respondents of a survey regarding their satisfaction of their manager were as follows:

2, 6, 4, 1, 7, 3, 6, 1, 7, 1, 6, 5, 1, 1

Compute the 95% confidence interval.

Answer

Mean: 2+6+4+1+ 7+ 3+ 6+ 1+ 7+1+ 6+ 5+ 1+ 1/ 14 = 3.64

Standard deviation: 2.47 – Using calculator

Standard error: 2.47 divided by $\sqrt{14}$ (14 respondents) =0.66

Margin of error: 0.66 × 2 = 1.34

CI: +3.64 0r -1.34

95% CI: -1.34≤ CI ≥3.64

Degrees of freedom

A data set contains a number of observations, say, *n*. They constitute *n* individual pieces of information. These pieces of information can be used to estimate either parameters or variability.

In general, each item estimated costs one degree of freedom. The remaining degrees of freedom are used to estimate variability. Researchers count them properly to get a fair assessment.

Example

There are *n* observations.

There is one parameter – the mean – that needs to be estimated.

That leaves *n-1* degrees of freedom for estimating variability.

Eight operatives in a sample: *df*=8-1=7

Two samples: There are *n1+n2* observations.

There are two means to be estimated.

That leaves *n1+n2-2* degrees of freedom for estimating variability.

Nine operatives in sample 1, 7 in sample 2: *df*=9+7 -2 = 14.

Multiple-Choice Questions

1. The normal distribution curve assumes…could be included within the distribution.
   A. some elements of a population.
   B. all elements of a population.
   C. selected elements of a population.
   D. sufficient elements of a population.

2. The maximum deviations in a normal distribution curve are
    A. -1 and +1.
    B. -0.5 and +0.5.
    C. -3 and +3.
    D. -2.5 and +2.5.

3. Normal distributions are symmetric around their
    A. mode.
    B. median.
    C. mean.
    D. standard deviation.

4. In a positively skewed curve, the tail is longer to the
    A. inside.
    B. middle.
    C. right.
    D. left.

5. In a negatively skewed curve, the mode is…than the might.
    A. lower.
    B. higher.
    C. greater.
    D. wider.

6. …of the area of a normal distribution is within one standard deviation of the mean.
    A. 86%.
    B. 8%.
    C. 95%.
    D. 68%.

7. A normal variable X is 1350, the mean is 1000 and the SD 100. The z value will be
    A. +3.5.
    B. -3.5.
    C. +2.5.
    D. -1.0.

8. Confidence intervals are expected to show…in a project.
    A. limits of accuracy.
    B. limits of errors.
    C. statistical errors.
    D. degree of confidence.

9. The manager predicts employee absences on Monday in the range of 45 + or − 6, the confidence limit is written as
   A. 95% CI: 39≤ CI ≥41.
   B. 95% CI: 39≤ CI ≥51.
   C. 95% CI: 51≤ CI ≥39.
   D. 95% CI: 45≤ CI ≥51.

10. There are eight operatives in a sample and six operatives in another sample. The degrees of freedom will be
   A. 6.
   B. 14.
   C. 7.
   D. 8.

Answers: 1B 2C 3C 4D 5A 6D 7A 8A 9B 10C

3 HYPOTHESIS TESTING IN HRM

An important part of HR based research comes from the formulating hypotheses. For quantitative research to be valid, results must be proved. The basic way to prove a result is to find if a hypothesis worked positively or negatively or not at all. The tendency of any researcher is to see the hypothesis formulated might work.

A hypothesis can be defined as any argument formulated by a human resource researcher. For instance, she might believe the application of suitable HR strategy might positively impact on stress at work. By quoting such an evidence, she develops a hypothesis. The solution cannot be generalised because it has to be proven. On formulating a hypothesis, the researcher takes the precaution of stating hat the hypothesis will not work. This becomes a null hypothesis since this invalidates the hypothesis under question.

The hypothesis test must happen in the most proficient manner. The sample under question should an accurate to either prove or disapprove a hypothesis. If such sample is incorrect, failing to align with the population under question, then errors are likely. These errors could be categorised as Type I or II. Such errors might also state the limitation of the research but do not necessarily question the competence of the researcher. Errors coming from the sample might simply limit the hypothesis to a single random sample.

In hypothesis testing, t and z-tests are generally performed by those embarking on initial human resource projects. Advanced parametric and non-parametric tests will be applied at advanced levels of study. An important concept in the tests is the p value. This is the critical value that researchers will look at in research. There are conventions namely a p-value which is less than 5% remains critical and suggests that the hypothesis has to be accepted. Any value greater than 5% means the null hypothesis must be accepted.

Hypothesis tests depend on values which are positive and negative. These tell the direction of the value as seen from the previous chapter's two-tailed concept. These do not state an increased or decreased value but essentially decide upon the direction of the outcome. Researchers who are confident of an outcome like effective HR strategies will increase employee satisfaction could generally say a one-tailed test might be performed since these are knowledgeable of the outcome. If researchers are uncertain, each might prefer a two-tailed test.

Hypothesis testing cannot ascertain the result of the tests undertaken. This is a scientific statistical test in relation to a sample under study while other techniques like observation, panel of expert opinion might also disapprove this.

The logic of hypothesis testing

The logic of hypothesis testing is deep-rooted in an understanding of the sampling distribution of the mean. The sample mean is an unbiased estimator of the population mean.

On average, a randomly selected sample has a mean equal to that in the population.

In hypothesis testing, we start by stating the null hypothesis. If the null hypothesis is true, then a random sample selected from a given population has a sample mean equal to the value stated .in the null hypothesis

Regardless of the distribution in the population, the sampling distribution of the sample mean is normally distributed. The probabilities of all other possible sample means we could select are normally distributed.

Using this distribution, we can state an alternative hypothesis to locate the probability of obtaining sample means with less than a 5% chance of being selected if the value stated in the null hypothesis is true.

Question

A researcher selects a sample of 50 employees to test the null hypothesis. The average employee works 40 hours per week. What is the mean for the sampling distribution for this population of interest if the null hypothesis is true?

Answer

The mean of the sampling distribution of the mean is the mean of the population from which the scores were sampled. Therefore, if a population has a mean μ, then the mean of the sampling distribution of the mean is also μ. The symbol μ_M is used to refer to the mean of the sampling distribution of the mean. Therefore, the formula for the mean of the sampling distribution of the mean can be written as: $\mu_M = \mu$

$$\mu_M = \mu = 40 \text{ hours}$$

Null hypothesis (H$_0$)

The null hypothesis states a population parameter is equal to a value. The null hypothesis is often an initial claim researchers specify using previous research or knowledge. In inferential statistics, the term "null hypothesis" is a general statement or default that there is no relation between two measured phenomena, or no association among groups. Rejecting or disproving the null hypothesis – and thus concluding there are grounds for believing there is a relation between two phenomena – is a central task in the modern practice of science; the field of statistics gives precise criteria for rejecting a null hypothesis.

The likelihood that a researcher rejects the null hypothesis decreases the closer the value of a sample mean is to the value stated by the null hypothesis. Similarly, the further the value of a sample mean is from the value stated in the null hypothesis, the greater is the risk that the null hypothesis could be rejected.

Alternative Hypothesis (H$_1$)

The alternative hypothesis states the population parameter is different than the value of the population parameter in the null hypothesis. The alternative hypothesis is what the researcher might believe to be true or hope to prove true.

Hypothesis to determine if a population mean, μ, is equal to some target value μ_0 include the following:

$$H_1: \mu = \mu_0: \text{NULL HYPOTHESIS}$$
$$H_1: \mu < \mu_0 \text{ (a lower-tailed test) or}$$
$$H_1: \mu > \mu_0 \text{ (an upper-tailed test) or}$$
$$H_1: \mu \neq \mu_0 \text{ (a two-tailed test)}$$

When $H_1: \mu = \mu_0$ the mean equals the target value. For instance, the mean working hours for educators is 30. If a sample mean gives 30 as a mean sample value, the sample mean equals the population mean. The null hypothesis is accepted.

A non-directional alternative hypothesis states the null hypothesis is wrong. A non-directional alternative hypothesis does not predict if the parameter of interest is larger or smaller than the reference value specified in the null hypothesis.

A directional alternative hypothesis states that the null hypothesis is wrong, and also specifies if the true value of the parameter is greater or less than the reference value in null hypothesis.

The advantage of using a directional hypothesis is increased power to detect the specific effect you are interested in.

The disadvantage is there is no power to detect an effect in the opposite direction.

Theoretical examples

Null hypothesis

Providing a revised salary package has little or no effect on the performance of employees in the public sector.

Alternative hypothesis

Providing a revised salary package is likely to affect the performance of employees in the public sector.

Null hypothesis

The division of work into smaller units has little or no effect on the reduction of production department employee strain.

Alternative hypothesis

The division of work into smaller units might reduce production department employee strain.

Null hypothesis

Implementing a clock card system will have little or no effect on absenteeism or lateness in the HR department.

Alternative hypothesis

Implementing a clock card system will have a positive effect on absenteeism or lateness in the HR department.

Non-directional

A researcher has results for a sample of employees who work in the administration department. She wants to know if the wages offered in this department differ from the national average of € 1000.

A non-directional alternative hypothesis is appropriate because the researcher is interested in determining if the scores are less or greater than national average.
H_0: μ = € 1000 vs. H_1: μ≠ € 1000

Directional

The same example could now be interpreted differently. A researcher has results for a sample of employees who work in the administration department. She wants to know if wages offered in this department differ from the national average of € 1000.

Researchers wants to know if administration staff earn above the national average of € 1000.

A directional alternative hypothesis can be used because the researcher is specifically hypothesizing that scores for trained students are greater than national average.
H_0: μ = € 1000. vs. H_1: μ > € 1000.

Z-tests in hypothesis testing

Z-tests are among the most straightforward of statistical hypothesis testing methods, and are often taught at an introductory level. These hold an important position in the dominion of applied statistical analysis.

The name Z-test comes from inference is made from the standard normal distribution, and 'Z' is the traditional symbol used to represent a standard normal random variable.

Z-tests were important because these gave researchers an easy way to do statistical hypothesis testing when computing power was limited. P-values and quantiles were easily obtained from standard normal tables.

Z-tests are used if the sample size is large or the population variance known. If the population variance is unknown and therefore has to be estimated from the sample itself.

and the sample size is not large (n < 30), the student's *t*-test may be more appropriate.

The z statistic formula is the sample mean minus the population mean stated in the null hypothesis, divided by the standard error of the mean.

$$\text{Z statistic : } Z = (X - \mu 0) / s\text{E}$$

Where

X is the sample mean

μ_0 is the mean of the population, and

S_E is the standard error of the mean

The Z statistic is an inferential statistic used to determine the number of standard deviations in a standard normal distribution that a sample mean deviates from the population mean stated in the null hypothesis.

Steps in Z-value calculation

Number of hours worked by a sample of employees in the production department was:

$$38, 39, 41, 40, 42, 44, 39, 36, 42$$

The population mean is 40 hours.

Standard deviation is 2,28.

Z score is -0.049.

Note a p-value also needs to be developed when doing such a test.

The Z-score

A z-score is the number of standard deviations from the mean a data point is. This measures how many standard deviations below or above the population mean a raw score is.

A z-score is also known as a standard score and can be placed on a normal distribution curve. Z-scores range from -3 standard deviations which would fall to the far left of the normal distribution curve up to +3 standard deviations which would fall to the far right of the normal distribution curve. To use a z-score, one needs to know the mean μ and also the population standard deviation σ.

Z-scores are a way to compare results from a test to a 'normal' population. Results from tests or surveys have many possible results and units.

A z-score can tell somebody where a value is compared to the average population's mean value.

In the example provided, the z-value was -0.049. The deviation was to the left but low. This was not statistically significant except values tended to be lower than the mean but not critical.

Some types of Z-tests

There are a few types of Z-tests done. All tests follow the same formula with a summation and square-rooted value. To facilitate the computation of such values, figures are input. The different tables provided in this section will explain how Z-tests are derived and how they could apply to an HR situation.

Example of one-sample z-test

The mean population age is 32 in the HR administrative department. In the sample, the mean value is computed as 37. The one-sample Z-test is performed.

	Population	**Sample**
Mean	32	37
Standard deviation	2.5	
Sample size	40	
Confidence Limit: 95%	Alpha : 0.05	95% confidence intervals
z-value	12.6491	Upper 32.7
two-tailed p-value	< 0.0001	Lower 31.2

Hypothesis test Formula:

$$z = \frac{\overline{x} - \Delta}{\frac{\sigma}{\sqrt{n}}}$$

where $\overline{x}$ is the sample mean, Δ is a specified value to be tested, σ is the population standard deviation, and n is the size of the sample. Please look up the significance level of the z-value in the standard normal table

Conclusion: Since the p-value is 0.0001 and is much lower than 0.05, this is critical. Hence, the two samples are statistically different.

Example of two-sample z-test

In this case, two samples each greater than 35 are tested. The mean for sample A is 40 and that for sample is 38. The SD values are provided as 10 and 12 respectively.

	Sample 1	Sample 2
Mean	40	38
Standard deviation	10	12
Sample size	35	35
Confidence Limit: 95%	Alpha : 0.05	95% confidence intervals
z-value	0.75	Upper 7.17
two-tailed p-value	0.448	Lower -3.17

Hypothesis test Formula:

$$z = \frac{\bar{x}_1 - \bar{x}_2 - \Delta}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

where $\bar{x}_1$ and $\bar{x}_2$ are the means of the two samples, Δ is the hypothesised difference between the population means (0 if testing for equal means), σ_1 and σ_2 are the standard deviations of the two populations, and n_1 and n_2 are the sizes of the two samples.

Conclusion: Since the p-value is 0.448 and is much higher than 0.05, this is not critical. Hence, the two samples are not statistically significant.

In practice, the two-sample *z*-test is not used often, as the two population standard deviations σ_1 and σ_2 are usually unknown. Instead, sample standard deviations and the *t*-distribution are used.

References

Easycalculation. (2017) Z-Score: Definition, Formula and Calculation, www.easycalculation.net.

Statistics How To. (2017) Z scores, http://www.statisticshowto.com/probability-and-statistics/z-score/

Houghton Mifflin Harcourt. (2016) Performing z-tests, www.cliffsnotes.com/study-guides/statistics.

Multiple-Choice Questions

1. Hypothesis tests
   A. observe if a prediction has worked or not.
   B. make a general observation.
   C. observe what should both occur in a test.
   D. always reject the null hypothesis wherever possible.

2. A general trend in descriptive human resource research is to choose
   A. a 95% confidence interval.
   B. a 90% confidence interval.
   C. a 99% confidence interval.
   D. a 70% confidence interval.

3. If the number of respondents exceeds 30, the test is chosen is
    A. a t-test.
    B. a paired t-test.
    C. a z-test.
    D. both t and z-test.

4. A z-test can be undertaken from a sample where the population's mean value is
    A. hidden.
    B. unknown.
    C. known.
    D. undisclosed.

5. A likely notation for a null hypothesis is
    A. Ho: $\mu_1 > \mu_0$
    B. Ho: $\mu_1 < \mu_0$
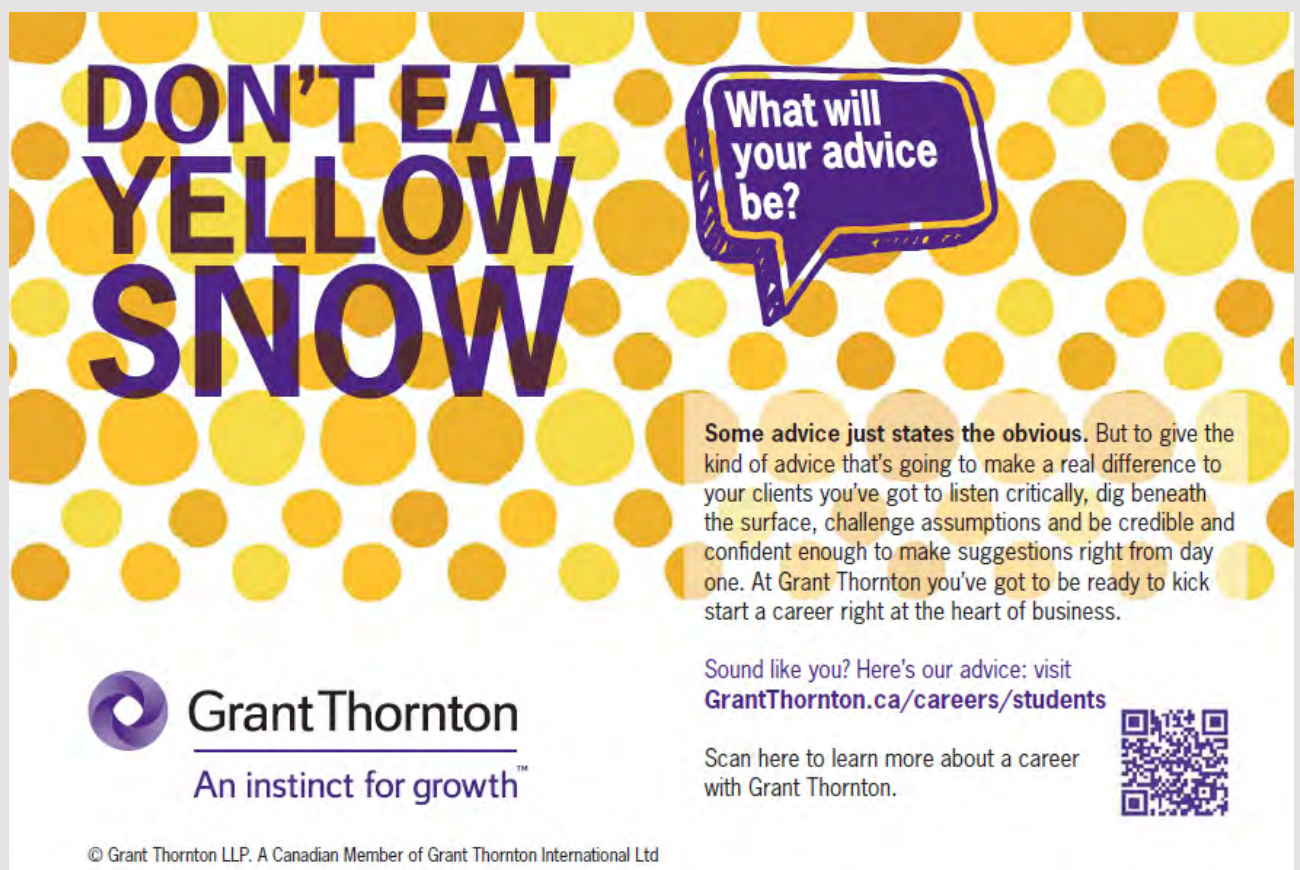   C. Ho: $\mu_1 \leq \mu_0$
    D. Ho: $\mu_1 = \mu_0$

6. The null hypothesis states a population parameter is…to a value.
    A. unequal.
    B. equal.
    C. proportional.
    D. inverse.

7. A value of 0.0001 illustrates
    A. an uncritical value.
    B. a critical value.
    C. an insignificant value.
    D. a meaningless value.

8. If the average earnings of employees are far lower than the mean value, the hypothesis test will be
    A. non directional.
    B. directional.
    C. transitional.
    D. optional.

9. In a two-sample z-test, a p-value of 0.5 is found out. This value is
   A. statistically significant.
   B. not statistically deviant.
   C. not statistically significant.
   D. statistically deviant.

10. In a two-tailed test, a lower value could be denoted by
   A. -2.25.
   B. +2.25.
   C. 0.
   D. 1.

Answers: 1A 2A 3C 4C 5D 6B 7B 8B 9C 10A

4   T-TESTS IN HRM

If the z-test offered researchers possibilities to compare actual values obtained in the findings compared to the mean value, the t-test remains an important variable in research. Most undergraduate HR and business students tend to give stress the t-test due of its practical, easy-to-use nature. If z-tests require samples of 30 and beyond, the t-test can be applied to any sample size less than 30.

A key part of the t-test is this compares two sets of values. This is interesting in so far students are capable of comparing values of two different sets of data. These could be the difference in responses between men and women. This could also be observations made before and after a set of observations, bringing researchers to choose what type of t-test each could undertake.

In statistical research, the t-test formula must be used while this could be claimed that it is more complex than the z-test formula. Researchers could feed inputs in an online calculator with responses displayed. These would tell if there are significant differences between the subjects under study. If t-test differences are significant, then these would have to confirm the hypothesis works. Alternatively, no significance implies that the null hypothesis must be accepted.

In this book, a few concepts of t-tests are described. The one-tailed t-test where the mean value should be known. In most of our HR cases, we illustrated 40 hours might be an acceptable mean for employees per week. Then, outcomes are compared, say, a most employees work over 40 hours. Then a one-tailed positive t-test might take place since the deviation moves to the right. If there are many employees crossing the 40-hour line, then the t-result might be significant.

Another t-test could be undertaken for two independent means. This could be the mean of two groups under study; Group A and Group B. Both groups are assessed and the data obtained are treated either manually or using an online calculator. Here, a first glance might give a general idea if values vary significantly or not. Else, this could be the tendencies of both groups, either similar or opposing. If these oppose much, then the value of the t-test might be significant. Little deviations would mean low significance and the null hypothesis could be accepted.

t-tests could apply either between groups and within groups. There is then the control and the experimental group. The control group might already provide a set of assumed values while the experimental group offers unexpected values. Once again differences will have to be seen and sorted out. It is also not to be forgotten that the p-value can be tested at a significance value of 1 or 5%.

One-sample t-test

This test looks at whether the mean of data from one group is different from a value you specify.

Example

A department's goal is to have a performance indicator which is significantly higher than the industry indexed standard of 100. The company's latest survey puts it at 110. Is an indicator of 110 significantly higher than the industry standard of 100?

Two-sample t-test

This test examines whether the means of two independent groups are significantly different from one another.

Example

A hypothesis states skilled workers are better than unskilled workers. The average indexed performance of skilled workers is 125, while the average score from unskilled workers is 90. Is 125 significantly different from 90?

Paired t-test

This test is for when you give one group of people the same survey twice. A paired t-test lets you know if the mean changed between the first and second survey.

Example

An HR manager surveyed the same group of employees twice: once in January and a second time in June, after each had been trained in safety for six months. Did the company's attitude change with regards to the safety training?

Note while t-tests can tell researchers if something is significantly different, it is up to each to determine whether such difference is meaningful. Small differences can generally have little significance in reasonable sample sizes but they might be statistically different if the sample size is large enough.

Statistical Significance of t-Test

A t-test's statistical significance indicates whether the difference between two groups' averages most likely reflects a 'real' difference in the population from which the groups were sampled.

Example

If researchers are interested in whether the average employee of firm X spends more time than the average employee in firm Y per month.

Researchers might ask 30 people from each sample about work hours. Each might observe a difference in those averages like 42 for the average form X worker and 38 for the average firm Y worker. But that difference is not *statistically significant*. This could be luck of which 30 people one randomly sampled makes one group appear to spend more hours working than the other.

If the researchers asked 200 firm X employees and 200 firm Y employees and see a big difference, this difference is less likely to be caused by the sample being unreliable.

The t-test's effect size complements statistical significance, describing the magnitude of the difference, whether the difference is statistically significant.

Definition

A *statistically significant* t-test result is a difference between two groups unlikely to have occur as the sample is atypical.

Statistical significance is determined by:

- the size of the difference between the group averages,
- the sample size, and-the standard deviations of the groups.

For practical purposes statistical significance suggests the two larger populations from which researcher sample differ.

The p-value in significance testing

Statistical significance is determined by looking at the p-value. The p-value gives the probability of observing the test results under the null hypothesis.

- The lower the p-value, the lower the probability of obtaining a result like the one observed if the null hypothesis is true. A low p-value indicates decreased support for the null hypothesis.
- The cutoff value for determining statistical significance is decided by the researcher, but often choosing a value of .05 or less. This corresponds to a 5% or lower chance of obtaining a result like the one observed if the null hypothesis was true.

One sample t-test procedure

The one sample t-test is a statistical procedure to determine if observations could have been generated by a process with a specific mean. Suppose one is interested in determining if an assembly line produces boots weighing 0.5 kilos.

To test this hypothesis, researchers could collect a sample of boots from the assembly line, weigh each, and compare the sample with a value of 0.5 kg using a one-sample t-test.

Hypothesis in one sample t-test

There are two hypotheses for a one sample *t*-test, the null hypothesis and the alternative hypothesis. The alternative hypothesis assumes some difference exists between the true mean (μ_1) and the comparison value (μ_0), whereas the null hypothesis assumes no difference exists.

The one sample t-test determines if the null hypothesis should be rejected, given sample data.

The alternative hypothesis can assume one of three methods depending on the question.

- If the goal is to measure difference, regardless of direction, please use a two-tailed hypothesis.
- If the direction of the difference between the sample mean and the comparison value matters, either an upper-tailed or lower-tailed hypothesis is used.

The null hypothesis remains the same for each type of one sample *t*-test. The hypotheses are formally defined below:

The null hypothesis H_0 assumes the difference between the true mean μ_1 and the comparison value μ_0 is equal to zero.

Mathematically, this is:

$$H_0: \mu_1 = \mu_0$$

The two-tailed alternative hypothesis H_1 assumes the difference between the true mean μ_1 and the comparison value μ_0 is not equal to zero.

Mathematically, this is:

$$H_0: \mu_1 \neq \mu_0$$

The upper-tailed alternative hypothesis H_1 assumes the true μ_1 of the sample is greater than the comparison value μ_0.

Mathematically, this is viewed as:

$$H_0: \mu_1 > \mu_0$$

The upper-tailed alternative hypothesis H_1 assumes the true μ_1 of the sample is greater than the comparison value μ_0.

Mathematically, this is viewed as:

$$H_0: \mu_1 < \mu_0$$

Assumptions of the one-sample t-test

As a parametric procedure-a procedure which estimates unknown parameters, the one sample t-test makes some assumptions. Although *t*-tests are quite rigorous, please evaluate the degree of deviation from these assumptions to assess the quality of the results.

The one sample *t*-test has four main assumptions:

- The observations are independent of one another.
- The dependent variable must be continuous.
- The dependent variable should be normally distributed.
- The dependent variable should not contain any outliers, values that are out of the range of observations.

Worked example

Step 1

State a null hypothesis statement H_0.

The accepted hypothesis is there is no difference in hours of work:

$$H_0: \mu = 40.$$

Step 2

Write an alternate hypothesis. This is the researcher tests. He thinks there *is* a difference in the mean hours increased

$$H_1: \mu > 40.$$

Step 3

Identify the following pieces of information researchers must calculate to the test statistic.

The question should give the researcher these items:

The sample mean($\bar{x}$). This is given 44 hours.
The population mean(μ). Given as 40.
The sample standard deviation, S= 10.
Number of observations(n) = 25.

Step 4

Insert the items from above into the t score formula.

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$$

$$t = (44 - 40) / ((10 / \sqrt{(25)})$$
$$t = (4 / 2) = 2.$$

This is the calculated t-value.

Step 5

Observed values

p-value and statistical significance:
The two-tailed p-value equals 0.0569.
By conventional criteria, this difference is considered to be not quite statistically significant.

Confidence interval:

The hypothetical mean is 44.00.

The actual mean is 40.00.

The difference between these two values is -4.00.

The 95% confidence interval of this difference: From -8.13 to 0.13.

Two sampled t-tests

When there is variance for population A is about the same as of population B, researchers can estimate the common variance by bringing together information from samples from population A and population B. An informal check for this is to compare the ratio of the two sample standard deviations. If these two are equal, this ratio would be 1. Since these are samples and therefore involve error, we cannot expect the ratio to be one.

In this examine a more formal test for equality of variances is examined.

Let N_1 be the sample size from population A, s_1 be the sample standard deviation of population 1.

Let N_2 be the sample size from population B, s_2 be the sample standard deviation of population 2.

The test statistic t is:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\left(\frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2}\right)\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}}$$

Where X_1 and X_2 averaged are the mean values of both samples.

Worked Example – Two sample t-test

Scores of students are evaluated in two courses to test significance. They are analysed below using an online calculating system. The test is unpaired as these test students in two independent tests.

Scores of HRM students in Managing Diversity Course	Scores of HRM students in Human Resource Management course
57	62
60	65
48	52
55	78
56	63
71	77
63	67
53	57

Students scored better in the HRM module than Managing Diversity course. There might be significance in the value.

Computed values

This is a paired test with 8 elements under study.
Degrees of freedom are 8+8-2=14.

Difference Scores Calculations

df_1=degrees of freedom: 8-1=7.

Managing Diversity Course students	Human Resource Management course
Treatment 1	*Treatment 2*
N_1: 8 $df_1 = N - 1 = 8 - 1 = 7$ M_1: 57.88 SS_1: 336.88 $s^2_1 = SS_1/(N - 1) = 336.88/(8-1) = 48.12$	N_2: 8 $df_2 = N - 1 = 8 - 1 = 7$ M_2: 65.12 SS_2: 562.88 $s^2_2 = SS_2/(N - 1) = 562.88/(8-1) = 80.41$
t-value calculation	
$s^2p = ((df_1/(df_1 + df_2)) \times s^2_1) + ((df_2/(df_2 + df_2)) \times s^2_2) = ((7/14) \times 48.12) + ((7/14) \times 80.41) = 64.27$ $s^2M1 = s^2p/N_1 = 64.27/8 = 8.03$ $s^2M2 = s^2p/N_2 = 64.27/8 = 8.03$ $t = (M_1 - M_2)/\sqrt{(s^2M1 + s^2M2)} = -7.25/\sqrt{16.07} = -1.81$	

The *t*-value is -1.81.
The *p*-value is .092.
The result is not significant at $p < .05$.
There is marked difference in performance before and after the training exercise.

Two sample paired t-test

The following are examples of performance units produced by employees before and after a training program.

Before training	After training
34	42
35	39
36	45
37	44
34	48
38	38
36	39
39	41

Difference Scores Calculations

df_1=degrees of freedom: 8-1=7.

Before training	After training
Treatment 1-Before training N_1: 8 $df_1 = N - 1 = 8 - 1 = 7$ *Mean*$_1$: 36.12 SS_1: 22.88 $s^2_1 = SS_1/(N - 1) = 22.88/(8\text{-}1) = 3.27$	*Treatment 2-After training* N_2: 8 $df_2 = N - 1 = 8 - 1 = 7$ *Mean* $_2$: 42 SS_2: 84 $s^2_2 = SS_2/(N - 1) = 84/(8\text{-}1) = 12$

t-value Calculation

$s^2p = ((df_1/(df_1 + df_2)) \square s^2_1) + ((df_2/(df_2 + df_2)) \square s^2_2) = ((7/14) \square 3.27) + ((7/14) \square 12) = 7.63$

$s^2M1 = s^2p/N_1 = 7.63/8 = 0.95$
$s^2M2 = s^2p/N_2 = 7.63/8 = 0.95$

$t = (M_1 - M_2)/\sqrt{(s^2M1 + s^2M2)} = -5.88/\sqrt{1.91} = -4.25$

The *t*-value is -4.25269.
The *p*-value is .000402.
The result is significant at $p < .05$.
There is marked difference in performance before and after the training exercise.

T-tests are useful for research purposes among students undertaking tests which will help improve findings. Such tests follow the z-test but offer the advantage of being tested upon fewer samples, often less than 30. The variety of t-tests, namely the one sample, the two sample, and the paired tests offer a fairly suitable range of possibilities to solve hypotheses developed by students. Familiarity with such concepts can only help in the achievement of suitable findings in research.

References

Horse, T. (2017) T-test, Definitions and Examples, Statistics how to.

Investopedia (2017) What is a t-test? http://www.investopedia.com.

Lucey, T. 91992) quantitative Techniques, DP Publications.

Statistics Solutions 92017) One sample t-test, http://www.statisticssolutions.com.

Trochim, W. (2006) The t-test, Research Methods Knowledge Base.

Multiple-Choice Questions

1. The t-test examines whether means of…are significantly different from each other.
   A. one independent group.
   B. one independent group.
   C. two dependent groups.
   D. two independent groups.

2. A/An….test is for when one gives a group of people the same survey twice.
   A. one-tail.
   B. two-tailed.
   C. unpaired.
   D. paired.

3. A…t-test result is one in which a difference between two groups is unlikely to have occurred as the sample is atypical.
   A. statistically non-significant.
   B. practically significant.
   C. statistically significant.
   D. practically non-significant.

4. If the goal is to measure any difference, regardless of direction,…is used.
   A. a one-tailed hypothesis.
   B. a positive-tailed hypothesis.
   C. a two-tailed hypothesis.
   D. a negative-tailed hypothesis.

5. If the direction of the difference between the sample mean and the comparison value matters,…hypothesis is used.
   A. an upper-tailed.
   B. either an upper-tailed or lower-tailed.
   C. a lower-tailed.
   D. neither an upper-tailed nor lower-tailed.

6. The upper-tailed alternative hypothesis H_1 assumes the true μ_1 of the sample is greater than the comparison value μ_0. Mathematically, this is viewed as

   A. H_0: $\mu_1 < \mu_0$
   B. H_0: $\mu_1 > \mu_0$
   C. H_0: $\mu_1 = \mu_0$
   D. H_0: $\mu_1 - \mu_{0 =1.}$

7. An assumption of a one sample two-tailed test is the
   A. the dependent variable should be normally distributed.
   B. the independent variable should be normally distributed.
   C. the dependent variable should be broadly distributed.
   D. the independent variable should be generally distributed.

8. There is a marked significant difference in a two-tailed test with a p-value of
   A. 0.92.
   B. 0.01.
   C. 0.055.
   D. 0.36.

9. In a two-sample t-test, a paired comparison could be
   A. after training only.
   B. during training.
   C. before and after training.
   D. at any time during training.

10. The…value for determining statistical significance is usually a value of .05 or less.
   A. cutoff.
   B. strategic.
   C. parametric.
   D. cut-throat.

Answers: 1D 2D 3C 4C 5B 6B 7A 8B 9C 10A

5   CHI-SQUARE TESTS IN HRM

A useful quantitative technique is the Chi-Square test. The meaning of the square means values must fit in a table and components are squared to fit in the formula. In this case, the Chi-Square is represented by X^2 and values are calculated using a complex calculation which this chapter explains. This test seeks to understand if variables are dependent upon each other with regards to choices or preferences. The condition is to separate respondents by identifying them differently like male/female, young/older workers, sample A or B… This differentiates choices and explains if each are significant. The Chi-Square test follows the t-test discussed in the previous chapter with the addition that this deals with more variables compared with the t-test.

The Chi-Square test of independence

The Chi-Square test of independence is used to determine if there is a significant relation between two nominal or categorical variables. The frequency of one nominal variable is compared with different values of the second nominal variable.

An R X C contingency table can display the data, where R is the row and C is the column.

For example, a researcher wants to examine the relation between gender (male v/s female) and attitude to shift work (high v/s low). This is based on the concept males would be more likely to work on shift compared with females who might have the desire to get more free time during the evenings and avoid shift work.

The chi-square test of independence can be used to examine this relation. If the null hypothesis is accepted there would be no relation between gender and shift work. If the null hypothesis is rejected the inference would be there is a relation between gender shift work.

– e.g. females tend to dislike shift work and males tend to accept shift work as a workable option.

Using the Chi-Square test

The test procedure described in this lesson is appropriate when meeting the following conditions.

- The sampling method is simple random sampling.
- The variables under study are each clear-cut.
- If sample data are in a contingency table, the expected frequency count for each table is at least five.

This approach consists of four steps.

- State the hypothesis,
- Formulate an analysis proposal,
- Analyse sample data, and
- Interpret results.

The steps explained

In this section, the four steps are explained and these will be followed by two examples to help understand how to do a Chi-Square test.

State the hypotheses

Suppose Variable X has r levels, and Variable Y has c levels. The null hypothesis states that knowing the level of Variable X does not help you predict the level of Variable Y. That is, the variables are independent.

H$_0$: Variable X and Variable Y are independent.
H$_1$: Variable X and Variable Y are not independent.

The alternative hypothesis is that knowing the level of Variable X can help the researcher predict the level of Variable Y.

Formulate an analysis proposal

The analysis plan describes how to use sample data to accept or reject the null hypothesis. The plan should specify the following elements.

- Significance level
  Often, researchers choose significance levels equal to 0.01, 0.05, or 0.10; but any value between 0 and 1 can be used. In this book, a significance level of 0.05 is sought for HR research.

- Test method
  The researcher will use the Chi-Square test for Independence to determine if there is a significant relationship between two categorical variables.

Analysis of sample data

Using sample data, researchers must find degrees of freedom, expected frequencies, test statistic, and the p-value associated with the test statistic. The approach described in this section is illustrated in the sample problem at the end of this lesson.

Degrees of freedom

Degrees of freedom (*df*) are equal to:

$$DF = (R - 1) \, X \, (C - 1)$$

where R is the number of levels for one categorical variable, and C is the number of levels for the other categorical variable.

Expected frequencies

The expected frequency counts are computed separately for each level of one categorical variable at each level of the other categorical variable. Compute R X C expected frequencies, according to the following formula.

$$E_{RC} = (n_R \, X \, n_C) \, / \, n$$

where

- E_{RC} is the expected frequency count for level R of Variable X and level C of Variable Y,
- n_R is the total number of sample observations at level R of Variable X,
- n_c is the total number of sample observations at level C of Variable Y, and
- n is the total sample size.

Test statistic

The test statistic is a chi-square random variable (X^2) defined by the following equation.

$$X^2 = \sum [O_{RC} - E_{RC})^2 / E_{RC}]$$

where

- O_{RC} is the observed frequency count at level R of Variable X and level C of Variable Y, and
- E_{RC} is the expected frequency count at level R of Variable X and level C of Variable Y.

P-value

The p-value is the probability of observing a sample statistic as extreme as the test statistic. Since the test statistic is a chi-square, please use the Chi-Square distribution calculator to assess the probability associated with the test statistic. And please use the degrees of freedom computed above.

Interpret Results

If the sample findings are unlikely, given the null hypothesis, researchers rejects the null hypothesis. This involves comparing the p-value to the significance level, and rejecting the null hypothesis when the p-value is less than the significance level.

Technically, if p is less than 0.05, the null hypothesis is rejected meaning there is no relation between the subject under investigation-preferably males and females and the choices under consideration like whether there is significance in difference between what males and females prefer.

Worked Example 1

There is an HR survey trying to find the significance between the choice of normal working hours for men and women in a firm. The data is displayed in a 2 × 2 table below.

	Normal Hours of work (9–4)	Shiftwork-3 tier system
Men	50	60
Women	70	35

We need to find out the Chi-Square value and interpret the results.

Manually, data could be interpreted as:

	Normal Hours of work (9–4) C_1	Shiftwork-3 tier system C_2	Total Rows $R_1 + R_2$
Men R_1	50	60	110
Women R_2	70	35	105
Total Columns $C_1 + C_2$	120	95	215

Analysing sample data

Degrees of freedom (df) = (R-1) X (C-1)
(2-1) X (2-1) = 1.

The expected frequency counts are computed separately for each level of one categorical variable at each level of the other categorical variable. Compute R X C expected frequencies, according to the following formula.

$$E_{RC} = (n_R \ X \ n_C) \ / \ n$$

$E_{Men \ 1,1}$ Row 1 × Column 1/total rows and columns: 110 × 120 / 215 = 61.4
$E_{Men1,2}$ Row 1 × column 2/total rows and columns: 110 × 95/215=48.6

$E_{Women \ 2,1}$ Row 2 × Column 1/total rows and columns: 105 × 120 / 215=58.6
$E_{Women1,2}$ Row 2 × column 2/total rows and columns: 105 × 95/215=46.4

Test statistic values

The test statistic is a chi-square random variable (X^2) defined by the following equation.

$$X^2 = \sum \ [O_{RC} - E_{RC})^2 \ / \ E_{RC}]$$

where

- O_{RC} is the observed frequency count at level *R* of Variable X and level *C* of Variable Y, and
- E_{RC} is the expected frequency count at level *R* of Variable X and level *C* of Variable Y.

X^2 is calculated as

$(50-61.4)^2/61.4 + (60-48.6)^2/48.6 + (70-58.6)^2/58.6 + (35-46.4)^2/46.4 = 9.803$

- The X^2 value is 9.803 or 9.8 to 1 d.p.
- The *p*-value is .001744.
- The result is significant at $p < .05$.

This means there are significant differences in terms of choices of men and women regarding shift work. Probably, the difference is caused because women might be less willing to do shift work than men due to family responsibilities.

Worked Example 2

Since the formulation of the Chi-Square model is already learnt, researcher can have more ease computing the statistics using an online calculator.

The following situation illustrates the views of full-time and part-time HR students on industrial placement, seminars, and guided tours for personal development.

	Industrial Placement	Seminars	Guided Tours
Full-time HR students	50	30	40
Part-Time HR students	40	60	20

To conduct a Chi-Square test, the total rows and columns must be added. These are shown in the amended table below.

	Industrial Placement	Seminars	Guided Tours	Total Rows
Full-time HR students	50	30	40	120
Part-Time HR students	40	60	20	120
Total Columns	90	90	60	240

Computed online and with lesser complication:

- The chi-square statistic is 17.8 to 2 d.p.
- The p-value is .000138.
- The result is significant at $p < .05$.

Conclusion:

There are significant observations between full and part-timers in HRM regarding three conditions imposed due to the significant value of 0.00014 to 5 d.p. Significance occurred in terms of proportions choosing seminars – twice as much for part-time and guided tours twice as much for full-time HR students.

Note closer figures would reduce the level of significance. There is also more difficulty of tabulation if more variables are included e.g. (5 × 5 table)

Precaution:

- In an HR test, the following might apply:
- 2×2 or 2×3 tables look interesting.
- The sampling method was simple random sampling.
- The variables under study were categorical.
- Significance is important although any result might occur.
- If values are too close in the tables, the test might not be relevant.
- Use an online Chi-Square calculator to do the test.
- Interpret the result convincingly.

References

Lucey T. (1992) Quantitative Techniques, DP Publications.

Stangroom, J. (2017) Statistics calculator, http://www.socscistatistics.com/tests/chisquare.

Statistics Solutions (2017) Chi-Square Tests http://www.statisticssolutions.com

Multiple-Choice Questions

1. In the Chi-Square test, degrees of freedom are calculated as
   A. Rows-1 × Columns-1.
   B. Rows-2 × Columns-1.
   C. Columns × Columns.
   D. Rows × Rows.

2. In a Chi-Square test, there are
   A. observed and expected frequencies.
   B. expected frequencies only.
   C. observed frequencies only.
   D. all observations possible.

3. The sampling method for a Chi-Square test is
   A. simple random sampling.
   B. stratified sampling.
   C. quota sampling.
   D. convenience sampling.

4. In a Chi-Square test, data is displayed in a row and column
   A. convenience table.
   B. congruence table.
   C. compatible table.
   D. contingency table.

5. The null hypothesis in a Chi-Square test would confirm
   A. Variable X and Variable Y are dependent.
   B. Variable X and Variable Y are interdependent.
   C. Variable X and Variable Y are semi-independent.
   D. Variable X and Variable Y are independent.

6. In a Chi-Square test, the following result is obtained: The X^2 value is 9.8 to 2 and the p-value is .001744. The Chi-Square test is
   A. not significant.
   B. insignificant.
   C. invaluable.
   D. significant.

7. In a Chi-Square test, if the sum of rows in 300, the sum of columns will be
   A. variable.
   B. 300.
   C. slightly lower.
   D. usually higher.

8. In a Chi-Square test, the frequency of one nominal variable is compared with… values of the second nominal variable.
   A. similar.
   B. different.
   C. same.
   D. unexplored.

9. The Chi-Square computation involves the sum of rows × sum of columns divided by the
   A. sum of total rows.
   B. sum of total columns.
   C. sum of total rows and columns.
   D. sum of individual rows and columns.

10. The Chi-Square test becomes complex when
    A. a few variables are tested.
    B. variables are modified.
    C. more variables are introduced.
    D. variables are being rationalised.

Answers: 1A 2A 3A 4D 5D 6D 7B 8B 9C 10C

6 ANOVA TESTS IN HRM

Analysis of variance (ANOVA) is an analysis tool used in statistics splitting the aggregate variability or inconsistency found inside a data set into two parts: regular factors and random factors. The regular or systematic factors have a statistical influence on the given data set, but the random factors do not. Analysts use the analysis of the variance test to determine the result independent variables have on the dependent variable in a regression study.

The analysis of variance test is the initial step in factors affecting a given data set. Once the analysis of variance test is finished, researchers perform additional testing on the methodical factors contributing to the data set's discrepancy. Researchers use the analysis of the variance test results in an F-test to generate additional data aligning with proposed regression models.

The test allows comparison of more than two groups at the same time to determine whether a relationship exists between them. The test analyses many groups to determine the types between and within samples.

ANOVA is a statistical technique assessing potential differences in a scale-level dependent variable by a nominal-level variable having 2 or more categories. Like Firm A against Firm B, Firm C…

The ANOVA, originated from Ronald Fisher (1918), is a consolidation of the t and the z test which have the problem of only allowing the nominal level variable to have just two categories. This test is also called the Fisher analysis of variance.

General Purpose of ANOVA

Researchers and students use ANOVA in many ways. The use of ANOVA depends on the research design. ANOVAs are used in three ways: One-way ANOVA, two-way ANOVA, and Multivariate ANOVA.

One-Way ANOVA

A one-way ANOVA refers to the number of independent variables-not the number of categories in each variable. A one-way ANOVA has just one independent variable. For example, difference in performance can be assessed by Firm A, and more firms in that variable.

The one-way ANOVA is used to determine if there are any statistically significant differences between the means of three or more independent or unrelated groups. The one-way ANOVA compares the means between the groups a researcher is interested in and determines if any of those means are statistically significantly different from each other. One-Way ANOVA is a parametric test.

The one-way ANOVA tests the null hypothesis:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k$$

where μ = group mean and k = number of groups. If, the one-way ANOVA yields a statistically significant result, researchers will accept the alternative hypothesis H_1, which implies that there are at least two group means that are statistically significantly different from each other.

At this point, it is important to understand that the one-way ANOVA is an omnibus test statistic and cannot tell the researcher which specific group/s was/were statistically significantly different from each other, only that at least two groups were.

This test has appellations like:

- One-Factor ANOVA.
- One-Way Analysis of Variance.
- Between Subjects ANOVA.

The variables used in this test are known as:

- Dependent variable.
- Independent variable also known as the grouping variable, or factor. This variable divides cases into two or more mutually exclusive levels, or groups.

The relevance of the F-test in ANOVA

F-tests are named after its test statistic, F, with reference to Ronald Fisher. The F-statistic is simply a ratio of two variances.

Variances are a measure of dispersion, or how far the data are dispersed from the mean. Larger values represent greater dispersion.

Variance is the square of the standard deviation.

F-statistics are based on the ratio of mean squares. The term mean squares may sound perplexing but it is simply an estimate of population variance that accounts for the degrees of freedom used to calculate that estimate.

Despite being a ratio of variances, a researcher can use F-tests in a wide variety of situations. Obviously, the F-test can assess the equality of variances. By changing the variances that are included in the ratio, the F-test becomes a very flexible test.

For example, a researcher can use F-statistics and F-tests to:

- test the overall significance for a regression model.
- to compare the fits of different models.
- to test specific regression terms, and to test the equality of means.

To use the F-test to determine whether group means are equal, please include the correct variances in the ratio.

In one-way ANOVA, the F-statistic is this ratio:

F = variation between sample means/variation within the samples

Uses and application of one-way ANOVA

The One-Way ANOVA is often used to analyze data from:

- Field studies.
- Experiments.
- Quasi-experiments.

The One-Way ANOVA is often used to test:

- Statistical differences among means of two or more groups.
- Statistical differences among means of two or more interventions.
- Statistical differences among means of two or more change scores.

Both the One-Way ANOVA and the independent samples t-test can compare the means for two groups. However, only the One-Way ANOVA can compare means across three or more groups.

Researchers often follow some rules of thumb for one-way ANOVA:

- Each group should have at least six subjects – ideally more. Inferences for the population will be more tenuous with too few subjects.
- Balanced designs – same number of subjects in each group are ideal; extremely unbalanced designs increase the possibility violating any of the requirements/ assumptions will threaten the validity of the ANOVA *F* test

Worked Example 1

The following example provides a performance indicator in three firms A, B, and C. The output refers to the number of hours worked by the employees. The selection is random and this follows the normal distribution.

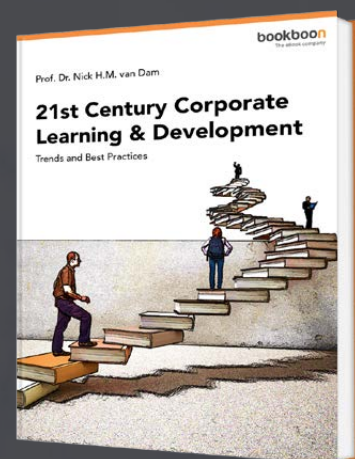| Firm A	Firm B	Firm C
40	44	37
43	44	35
42	40	35
38	45	36
45	46	34
37	46	32
39	48	37

Some trends are seen like Firm C with employees working less, Firm B with employees exceeding 40 hours and Firm A where 40 could be an estimated mean.

Information is fed on an online calculator.

Data Analysis

	Treatments			
	Firm A	**Firm B**	**Firm C**	**Total**
N	7	7	7	21
ΣX	284	313	246	843
Mean	40.5714	44.7143	35.1429	40.1429
ΣX²	11572	14033	8664	34269
Std.Dev.	2.8785	2.4976	1.7728	4.6291

Summary of data

Note:

ΣX equals the sum of all working hours in Firms A, B and C.
Mean is the total hours divided by 7.
Sum of means squared is like: 284 × 284.
Standard deviation is calculated by the SD formula.

Results:

Result details				
Source	*SS*	*df*	*MS*	
Between-treatments	322.5714	2	161.2857	$F = 27.38814$
Within-treatments	106	18	5.8889	P=0.00001
Total	428.5714	20		

Where:

SS is the sum of squares between and within treatments.
df: represents the degrees of freedom (3-1) = 2.

- The *f*-ratio value is 27.38814.
- The *p*-value is < .00001.
- The result is significant at $p < .05$.

The result:

$$[F(2, 20) = 27.38, p = 0.0001]$$

Two-Way ANOVA

A two-way ANOVA refers to an ANOVA using two independent variables. With the example above, a 2-way ANOVA can examine differences in performance scores known as the dependent variable by the independent variable (1) and category independent variable (2).

Two-way ANOVAs can be used to examine relations between two independent variables. Relations indicate differences are not uniform across all categories of the independent variables. For instance, technical level workers may have higher performance scores overall compared to semi-skilled workers, and are much greater in Firm A compared to Firms B and C.

Two-way ANOVAs are also called factorial ANOVAs. Factorial ANOVAs can be balanced by having the same number of participants in each group or unbalanced having different number of participants in each group. Unequal size groups can make appear there is an effect when this may not be the case. There are several procedures a researcher can do to solve this problem.

Assumptions of Two-way ANOVA

There some assumptions to do Two-way ANOVA and these are conditions for Two-way ANOVA.

Assumption 1

The *dependent variable* should be measured at the continuous level – these are interval or ratio variables.

Assumption 2

The *two independent variables* should consist of two or more categorical, independent groups.

Assumption 3

Researchers should be independent of observations, meaning there is no relationship between observations in each group or between groups.

Assumption 4

There should be no significant outliers. Outliers are data points in your data which do not follow the usual pattern.

Assumption 5

The dependent variable should be normally distributed for each combination of the two independent variables groups.

Assumption 6

There needs to be homogeneity of variances for each combination of groups of the two independent variables.

Worked Example 2

There are two broad categories under evaluation. There are administration and production employees under observation. Each are assessed on variables like: long breaks (1hour), medium breaks (30 minutes at the rate of 2 per day) and staggered breaks (4 short breaks of 10 minutes and one break of 20 minutes). Their choices are displayed below. Their Score are given between one and ten. Six employees are randomly selected from each group.

Administration	Long Break	Medium Break	Staggered breaks
A	5	4	7
B	6	4	7
C	6	5	8
D	6	6	4
E	4	4	8
F	6	4	8

Production	Long Break	Medium Break	Staggered breaks
G	8	6	5
H	8	6	5
I	10	5	4
J	7	6	4
K	7	6	5
L	8	6	5

There are differences in the tables. Administration looks more in favour of staggered breaks as work is of sedentary nature and this need more relaxation from sitting a long time in office.

At the production level, long break and medium break look more appealing compared to shorter breaks, given hard work and physical effort,

Table of Means: Cells, Rows, Columns, and Total

Mean Values	Long Break	Medium Break	Staggered breaks	Total Mean Rows
Administration	5.5	4.5	7	5.6
Production	8	5.8	4.6	6.1
Total Mean Column	6.75	5.1	5.8	5.9

TWO WAY ANOVA SUMMARY

Source	SS	df	MS	F	P
Between groups	55.58	5			
Rows	2.25	1	2.25	2.48	0.12
Columns	15.16	2	7.58	8.37	0.001
Rows × columns	38.16	2	19.08	21.07	0.000002
Within groups	27.16	30	0.9		
Total	82.75	35			

The table is redrafted as follows for interpretation.

Source	SS	df	MS	F	P
Between groups	55.58	5			
Administration and Production	2.25	1	2.25	2.48	0.12
Long, medium and staggered breaks	15.16	2	7.58	8.37	0.001
Departments × breaks	38.16	2	19.08	21.07	0.000002
Within groups	27.16	30	0.9		
Total	82.75	35			

$F_{(1,30)} = 2.48$, p0.12, results are significant; significant differences exist between administration and production department.

$F_{(2,30)} = 8.37$, p 0.001, results are significant, significant differences exist between the types of breaks available.

$F_{(2,30)} = 21.07$, p 0.00002, results are extremely significant in that there is high degree of variability between each department and each type pf break chosen.

This was also clearly seen such variability was important from the start and each case must be treated independently.

Worked Example 3

There is a need to find out variances regarding wages practised in the public and private sectors of a country. The wages are indexed in ratio terms for three classes of employees: executives, tactical or mid-management level, and operational level. The test aims to find out if there are significant variations in the two sectors. Employees are randomly selected with names Aa, Bb…. The information is in the table below.

Public	Executive	Tactical	Operational
Aa	50	14	6
Bb	60	13	7
Cc	60	22	10
Dd	100	26	12
Ee	110	23	12
Ff	75	18	8

Private	Executive	Tactical	Operational
Gg	80	12	5
Hh	180	15	8
Ii	210	25	7
Jj	170	26	6
Kk	150	16	6
Ll	80	15	8

Seen from the tables, some differences do exist. Executives tend to be better paid in the private sector than the public sector. At the tactical or operational level, public employees lead. The data is analysed in terms of mean values and results are then displayed.

Here is a two-way ANOVA summary. *Vassarstats* provides an online calculator which should be 'prepared' in terms of rows and columns (two rows and three columns) and then input data please. Note even means could be inputted into fields.

Two Way ANOVA Summary

Source	SS	df	MS	F	P
ROWS: Public/Private	4290.25	1	4290.25	7.22	0.0116
COLUMNS: Executive, Tactical, Operational	76105.56	2	38052.78	64.07	<.0001
Rows × columns	10084.66	2	5042.33	8.49	0.0012
Error	17817.17	30	593.91		
Total	108297.64	35			

Interpretation:

$F(1,35)=7.22$, p 0.0116 shows significant differences between public and private firms, $F(2,35)=64.07$, $p<0.0001$ shows very strong significance between the three occupations and $F(2,35)=8.49$, $p=0.0012$ shows very strong significance between the sum of organisations and grades. There are significant variations in the sample regarding worker grades in public and private sectors. Mean salaries differ as seen from the initial observation.

References

Lowry, R. (2007) Two-Way Analysis of Variance for Independent Samples, Vassar Stats.

Lucey, T. (1992) Quantitative Techniques, DP Publications.

Stepup analytics (2017) Two-way ANOVA calculation, http://www.stepupanalytics.com

Wessa.net (2007) Two-Way ANOVA – Free Statistics Software (Calculator), www.wessa.net

Multiple-Choice Questions

1. Analysis of variance (ANOVA) is a statistical tool splitting aggregate variability into two
   A. irregular factors and random factors.
   B. regular factors and random factors.
   C. regular factors and non-random factors.
   D. irregular factors and non-random factors.

2. The ANOVA test analyses many groups to determine types between and
   A. within samples.
   B. among samples.
   C. without samples.
   D. outside samples.

3. A one-way ANOVA has
   A. One dependent and one independent variable.
   B. no independent variable.
   C. two independent variables.
   D. one independent variable.

4. The F-statistic is a ratio of two variances measuring how far the data are
   A. dispersed from the mean.
   B. close to the mean.
   C. away from the mean.
   D. within limits from the mean.

5. Compared to the t-test, One-Way ANOVA can compare the means across
   A. two groups.
   B. one group.
   C. three or more groups.
   D. just three groups.

6. An ANOVA result like F (2, 20) = 27.38, p = 0.0001 states variations among groups are
   A. highly insignificant.
   B. highly significant.
   C. less significant.
   D. insignificant.

7. Two-way ANOVAs are also called
   A. variate ANOVAs.
   B. factorial ANOVAs.
   C. binary ANOVAs.
   D. multi-factorial ANOVAs.

8. In a Two-way ANOVA, there needs to be…for each combination of the groups of the two independent variables.
   A. inconsistent variances
   B. unequal variances
   C. heterogeneous variances
   D. homogeneous variances

9. A highly significant variability in a Two-Way ANOVA could be seen from
   A. F (2,10)= 20, p = 0.0001.
   B. F (2,10)= 2, p = 0.7.
   C. F (2,10)= 8, p = 0.06.
   D. F (2,10)= 6, p = 0.08.

10. A notation F (2,6) =3.57, p=0.57 denotes
    A. a high level of inconsistency between variables.
    B. a high level of significance between variables.
    C. a high level of homogeneity between variables.
    D. a high level of incongruity between variables.

Answers: 1B 2A 3D 4A 5C 6B 7B 8D 9A 10C

7 HRM PROBABILITY CONCEPTS

Probability refers to the likelihood of an event occurring or not. Managers can make such predictions. For instance, material availability and production use could depend on the correct assumption raw materials reached the organisation on time. Failure to do so means the outcome is unlikely to happen and the firm faces hardship.

Regarding HRM or other business functions, probability is a suitable tool to evaluate occurrences. In HRM, probabilities could range from recruiting qualified staff, expected performance level, and the benefit each might bring after recruitment. Another one could be the opportunity of training to employees and see how far this benefits each. A training offering no tangible benefits is one with a high probability of failing.

There are basic concepts regarding probability. Probabilities estimate occurrences but cannot determine like t-tests or other parametric tests if there are chances the occurrences might be critical to a firm. In probability, the values are either in fractions and decimals.

The general convention in probability is the value ranges between zero and one with varying degrees of progression. Zero is not counted as an occurrence but any value above this symbolises a likely occurrence. All combined probabilities sum to a maximum of one.

This chapter includes the factor tree as being an understandable way of doing probability test with the goal ideas develop from a node – an initial point of start – and emerge into a tree with varying options. Often these are as follows: likely; unlikely; most likely. There is also a formulation in probability which utilises the same concept but explains a direct way of combining probability values.

Since the HR field is concerned with business, there is also a section on expected values where each outcome is allocated a positive or a negative value. This activity ensures values could be computed and linked with the investment – usually a negative value-and seen as to if one or another outcome would be the most optimising one. This allows those of us who are business managers to see how wisely we can take a decision and find out an opportunity of doing things economically and rightly.

Probability is a mathematical concept helping HR managers determine which pathway to choose for reaching a satisfactory decision. And enforcing such managers to develop projections and possibilities for evaluating choices based on chance factors which are not dissociate from rational decisions.

Mutually exclusive events

Mutually exclusive events are those which cannot occur concurrently – where the occurrence of one event results in non-occurrence of the other event. Such events cannot be real at once. The occurrence of one event makes the occurrence of another event impossible. These are also known as disjoint or separate events.

- Mutually exclusive events are events when occurrence is not concurrent. When the occurrence of one event cannot control the occurrence of other, such events are independent events.
- In mutually exclusive events, the occurrence of one event will result in the non-occurrence of the other.
- Mutually exclusive events are represented mathematically as P (A and B) = 0 while independent events are represented as P (A and B) = P(A) P(B).

Examples

- Being a human resource manager and an operative are exclusive – one cannot do both at once.
- Tossing a coin: Heads and Tails are mutually exclusive.
- Undertaking paperwork in an office and doing physical activities are mutually exclusive.

Probability of an event happening=

Number of ways this can happen/Total number of outcomes

When two events A and B are exclusive each may never happen together:

P (A and B) = 0

The probability of A and B equals 0 or impossible

Example

Manager AND Operative

An employee cannot be a manager and an operative at the same time.
The probability of a manager and an operative is zero.

P (Manager AND Operative) = 0.

The probability of A or B is the sum of the individual probabilities:

$$P (A\ or\ B) = P(A) + P(B)$$

The probability of A or B equals the probability of A plus the probability of B.

Example

Manager OR Operative

In an interview for having a manager is, so P(manager)=1/20
the probability of an operative is also 16/20, so P(operative)=16/20

When both events combine:

The probability of a manager or an operative is (1/20) + (16/20) = 17/20:

$$P \text{ (Manager or Operative)} = (1/20) + (16/20) = 17/20$$

Independent events

Independent events are the events, in which the probability of one event does not control the probability of the occurrence of the second event. The occurrence or non-occurrence of such an event has no effect on the occurrence or non-occurrence of another event. The product of separate probabilities is equal to the probability both events occur.

There are a few things to note about this experiment. Choosing one option from the sample, replacing this, then choosing another option from the same sample is a compound event.

Since the first sample was replaced, choosing a sample on the first try has no effect on the probability of choosing a similar sample on the second try. These events are independent.

Example 1

For a training programme, there are 3 operatives, 5 technicians, 2 human resource officers and 6 trainees. They are known as samples.

A sample is chosen at random from the card names linked with each staff member. After replacing this, a second card name is chosen.

What is the probability of choosing:

- An operative and a trainee?
- A technician and a trainee?
- A human resource officer?

Solution

Assuming replacement occurs, the event is independent.

An operative and a trainee: 3/16 × 6/16 = 18/196 or 0.09

A technician and a trainee: 5/16 × 6/16 = 30/196 or 0.15.

A human resource officer: 2/16 or 0.125

Example 2

A nationwide survey showed 65% of workers dislike working Sundays. If four workers are chosen at random, what is the probability all four like working Sundays? Please round answer to the nearest percent.

Solution

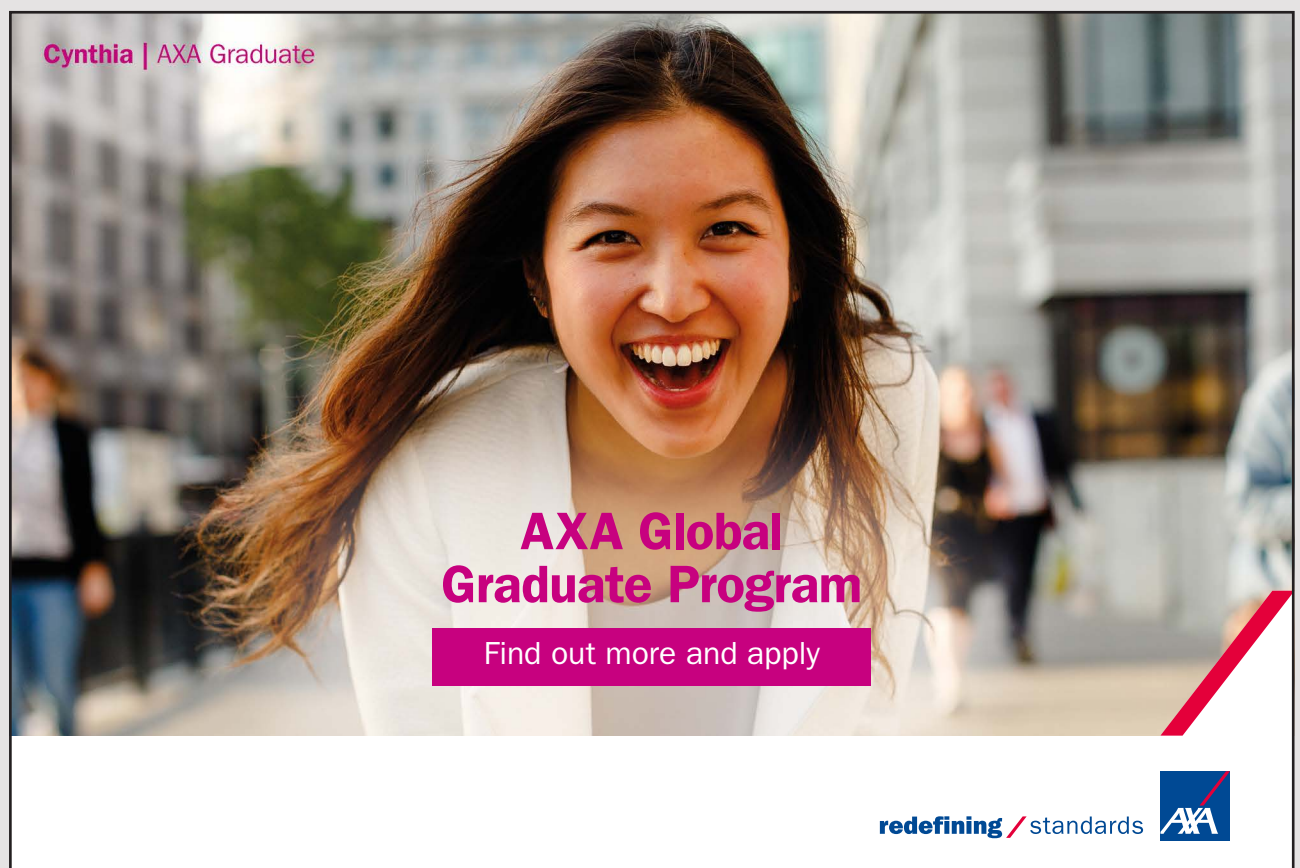65% dislike working on Sundays, then 35% like working.

4 workers are chosen.
Each worker's probability is 35/100.
All 4 probabilities are: $0.35 \times 0.35 \times 0.35 \times 0.35 = 0.015$

Tree Diagram

A tree diagram is a way of representing a sequence of events. Tree diagrams are useful in probability since each record all possible outcomes in a clear, simple manner.

Tree diagrams represent a series of independent events like coin tosses or conditional probabilities such as drawing cards without replacement. Each node on the diagram represents an event and associates with the probability of this event. The root node represents the certain event and has probability one. Each set of sibling nodes represents an exclusive and exhaustive partition of the parent event.

The probability associated with a node is the chance of this event occurring after the parent event occurs. The probability events leading to a particular node occurs is equal to the product of this node and parents' probabilities.

General conditions

- All probabilities combined provide a maximum value of one.
- Various possibilities exist to get a combined outcome.
- The most cost-effective solution is accepted.

Example

A manager recruits both in the administration and the finance department of a firm. The choice is 50% for both recruits.

In administration, the probability of recruiting an experienced graduate is 10%, the probability for recruiting a graduate without experience is 20% and the remaining applies for a trainee administrative clerk.

In finance, the probability of recruiting an experienced finance graduate is 20%, the probability for recruiting a graduate without experience is 50% and the remaining applies for a trainee finance clerk.

The outcomes are below:
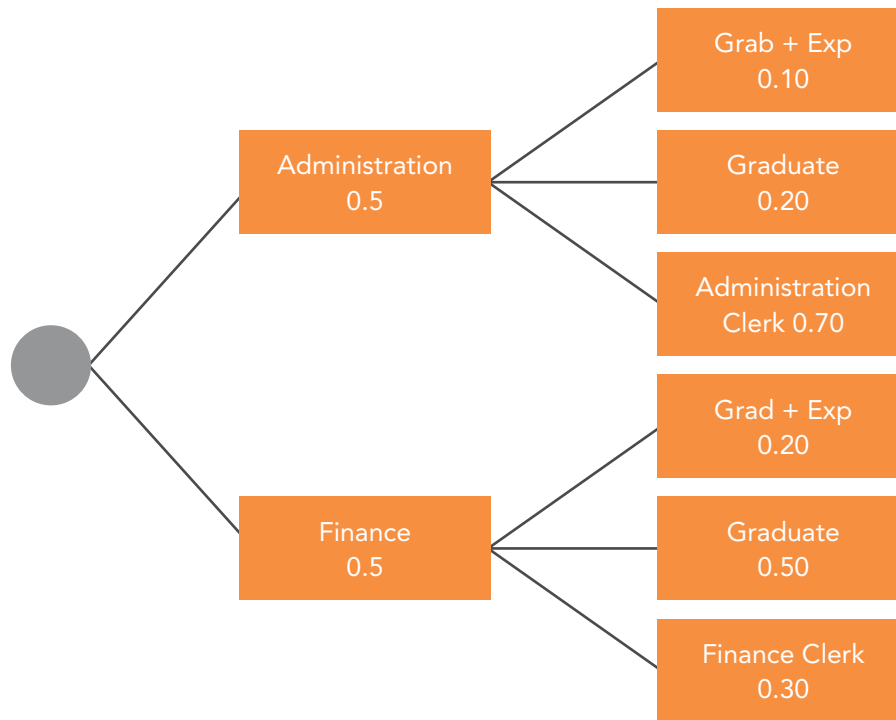
Options

Probability of having:

- An experienced, administration graduate.
- An experienced, finance graduate.
- A graduate in both areas without experience.
- Clerks in both areas.

Solution

A graduate in administration with experience: $(0.5 \times 0.5) = 0.25$

A graduate in finance with experience: $(0.5 \times 0.2) = 0.10$

A graduate in both areas without experience:

Administration: $(0.5 \times 0.2) = 0.10$
Finance: $(0.5 \times 0.5) = 0.25$
Combined: $0.10 + 0.25 = 0.35$

Clerks in both areas:

Administration: $0.5 \times 0.7 = 0.35$
Finance: $0.5 \times 0.3 = 0.15$
Combined: $0.35 + 0.15 = 0.5$

Hence, there is more chance have clerks in both areas than having, say, an experienced finance graduate.

Expected Value

The expected value (EV) is an anticipated value for a given investment. In statistics and probability analysis, the EV is calculated by multiplying each of the possible outcomes by the likelihood each outcome will occur, and summing all of those values. By calculating expected values, investors can choose the scenario most likely to give desired outcome.

Scenario analysis is one technique for calculating the EV of a business opportunity. This uses estimated probabilities with multivariate models, to examine possible outcomes for a proposed investment. Scenario analysis also helps human resource managers determine whether they are taking on an appropriate level of risk, given the likely outcome of the labour or financial investment.

The EV or expectation, mean, or first moment. EV can be calculated for single discreet variables, single continuous variables, multiple discreet variables, and multiple continuous variables. Please use integrals for continuous variable situations.

Example

An HR manager is offered two options: using the current pay system or a computerised pay system.

The current pay system which relies on paper has a fixed annual cost of €3000 and the computerised pay system needs an investment of €5000.

The options are as follows:

In the traditional pay system, the benefits are estimated at €7000 with a probability of 0.5 and the losses are estimated at €2000. In the computerised pay system, the benefits are around €20000 while losses are at €2000 with a probability of 0.3. Please present the situations on a tree diagram.
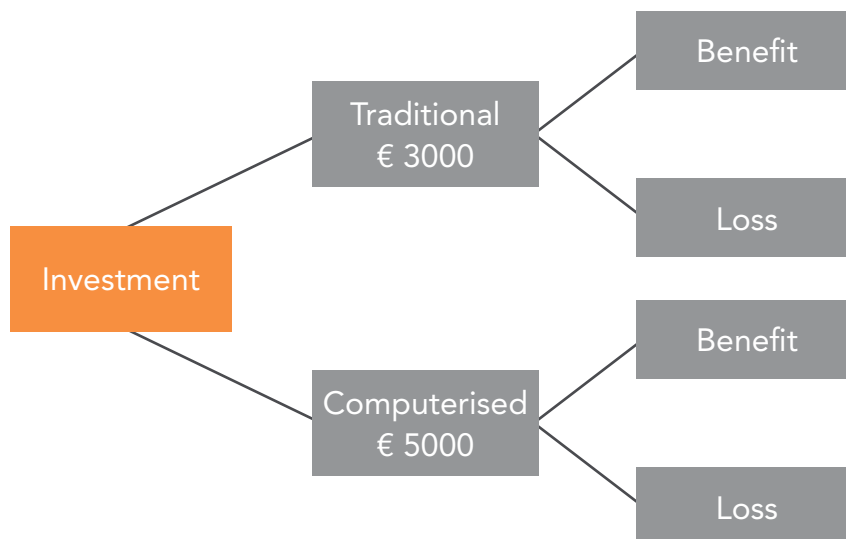
Options

Traditional:

Benefit: -€3000 + €7000 × 0.5 = +€500.

Loss: -€3000 -€2000 × 0.5 = -€4000.

New System:

Benefit: -€5000 + €20000 × 0.7 = €9000.

Loss: -€5000 − €2000 × 0.3 =-€5600.

The new system earns more benefits: €2000 compared to €500.

The combined outcomes: Profits and losses are:

Traditional: €500-€4000 = €4500 (-ve).

New system: €9000-€5600 = € 3400 (+ve).

The new system is considered as being better. Still the traditional system offers benefits but losses are more concerning.

Although this method works, the main drawback is forecasted values might not exact and probabilities might not be exact.

References

BBC. (2017) Maths – probability, http://www.bbc.co.uk/schools/gcsebitesize/maths/statistics/probability.

Lucey, T. (2002) Quantitative Techniques, Cengage Learning EMEA; 6th edition, ISBN: 978-1844801060

Maths is Fun. (2014) Probability, http://www.mathsisfun.com/data/probability.

Onlinemathlearning.com. (2017) Probability tree diagrams, http://www.onlinemathlearning.com/probability-tree-diagrams.

Multiple-Choice Questions

1. Probability refers to the…of an event in that this would occur.
   A. perception.
   B. challenge.
   C. forecast.
   D. likelihood.

2. Mutually exclusive events are those which…occur concurrently.
   A. should.
   B. never.
   C. always.
   D. might.

3. Independent events are events, in which the probability of one event never control the probability
   A. of the non-occurrence of the second event.
   B. of the occurrence of the same event.
   C. of the occurrence of the second event.
   D. of the non-occurrence of the same event.

4. A survey showed that 20% of all workers in a country like working part-time. If two workers are chosen at random, what is the probability all four like working Sundays?
   A. 0.04.
   B. 0.4.
   C. 0.2.
   D. 0.08.

5. Tree diagrams may represent
   A. conditional probabilities.
   B. unconditional probabilities.
   C. conditional chances.
   D. unrestricted probabilities

6. All probabilities combined provide a…value of one.
   A. maximum.
   B. minimum.
   C. given.
   D. forecasted.

7. The expected value is an
   A. unassured value for a given investment.
   B. unpredicted value for a given investment.
   C. unknown value for a given investment.
   D. anticipated value for a given investment.

8. In a scenario, an initial training investment is €2000 with a productivity gain of €5000 with a probability of 80%. The expected value is
   A. €500.
   B. €4000.
   C. €200.
   D. €2000.

9. The probability of recruiting a graduate in HRM is 0.7. The probability of having a skilled graduate is 0.3. The probability of having an unskilled graduate is
   A. 0.7.
   B. 0.3.
   C. 0.49.
   D. 0.21.

10. In independent probability, the condition is to…the sample
    A. remove.
    B. replace.
    C. change.
    D. eliminate.

Answers: 1D 2B 3C 4A 5A 6A 7D 8D 9C 10B

8 CORRELATION AND REGRESSION CONCEPTS IN HRM

An important concern for Human Resource Management students is the question of using correlation and regression as techniques applicable to quantitative research. Often data input on computer software working out the data available make sit a way to present all information with the values, degrees of significance, and if the data presented is of statistical significance or not.

Correlation and regression are two topics close and might be interpreted as one concept while there is a difference in these two terms. Correlation addresses two variables. One the independent variable and the dependent variable. Continued exposure to training – the independent variable – will positively influence employee output in quality or performance – the dependent variables. In correlation, the inputs must be continuous like the more the time spent on training, the higher will be the performance.

Correlation concepts include two variables only. This is also aligned with the straight line equation $y = mx + c$ where y is the output variable, m is the gradient or strength of the variable, x is the independent variable and c is the intercept or starting point to be considered. Correlation is a step ahead of this by including a formula to interpret the values input and derive a correlation coefficient.

Regression is a further step ahead of correlation where there should be two or more variables to be included. This takes the shape of $y = a + bx_1 + cx_2 + dx_3$. implying the relevance of several independent variables influencing output. This is formulated in a more complex manner with the R coefficient which could be the rank or Spearman coefficient.

With a regression formula, HR researchers can develop a basic econometric model. This would explain the interaction of each variable in the model and the strength per variable is found through the provision of values per independent variable. In this way, b, c, or d shown under the model $y = a + bx_1 + cx_2 + dx_3$ must be in mathematical terms.

The use of correlation and regression matter in research but please take to see the models are used and applied well. Observing will not give consistent results since these formulae work with add-on values like an increase in some independent variable. Successive data is collected in this way implying observations are made at different times. These apply to economic concepts scaled over years like inflation and compensation. A useful element in these concepts is the evaluation of the trend's strength or weakness. This could be observed then calculated to have a clearer, more logical trend estimate.

The linear equation

A linear equation made up of two expressions set equal to each other. A linear equation is distinct because:

- This has one or two variables.
- No variable in a linear equation is elevated to a power greater than 1 or used as the denominator of a fraction.

A linear equation in two variables describes a relation in which the value of one variables depends on the value another.

In a linear equation in *x* and *y*,

- x is called *x* is the independent variable and
- y depends on it. It is called the dependent variable.

If the variables have other names, yet do have a dependent relation, the independent variable is plotted along the horizontal axis. Most linear equations are functions that is, for every value of *x*, there is only one corresponding value of *y*.

When researchers assign a values to independent variables, *x*, we can compute the value of the dependent variable, *y*. We can plot the points named by each *x*, *y* pair on graph. The importance of emphasising graphing linear equations is we should already know any two points define a line, hence finding many pairs of values satisfying a linear equation is easy:

- Find two pairs of values, and
- draw a line through the points they describe.

All other points on the line provide values for *x* and *y* satisfying the equation.

Describing linear relationships

The graphs of linear equations are always lines. One important thing to remember about those lines is: Not every point on the line the equation describes necessarily is a solution to the problem the equation describes.

Examples

There is a linear relationship between the number of successive days employees worked and the production output. This is positive.

There is a linear relationship between an increase in the number of hours worked by employees and overheads. This is negative as overheads decrease.

There is a linear relationship between the number of successive days employees worked and the fixed cost. This is a neutral relation.

The first case is below.

Output
Dependent
variable

Successive days worked
Independent variable
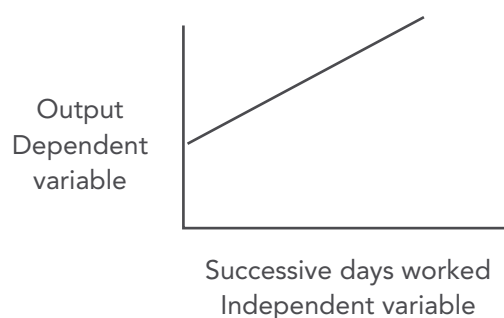
Note each value on the horizontal line increases to influence output. An intercept on the line expresses:

- Output is not necessarily zero if one starts on the first day. There might be a fixed cost factor like electricity or mechanical overhead affecting it.
- The line gradient remains an important determinant in the linear equation.

The line gradient

When a line slopes up from left to right, this has a positive slope. This means a positive change in y associates with a positive change in x. The steeper the slope, the greater the rate of change in y in relation to the change in x. When one deals with data points plotted on a coordinate plane, a positive slope indicates a positive correlation and the steeper the slope, the stronger the positive correlation.

The gradient is the change in the dependent factor in relation to the independent to the dependent variable.

Two suitable points are needed to be located to determine the gradient – a lower and upper limit in the graph.

Output
Dependent
variable

Δy

Δx

Successive days worked
Independent variable
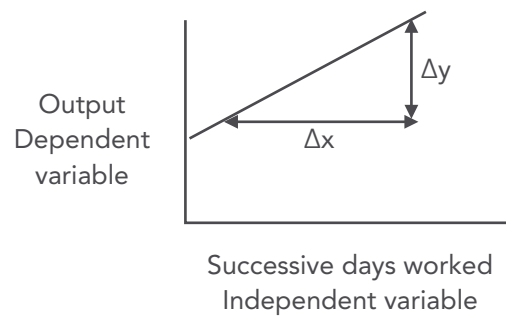
The change is interpreted as $\Delta y/\Delta x$

Example

The output changed from 100 to 120 units $\Delta y = 120-100 = 20$

The number of days worked were from Monday to Friday $\Delta x = 4$ days.

The gradient is $20/4 = 5$.

Similar trends can be drawn using a gradient. In a negative trend, as the slope goes downwards, the value will be negative. A negative gradient indicates a downwards slope.

The Correlation Concept

The correlation is one of the most common, useful statistics. A correlation is a single number describing the degree of relation between two variables.

Correlation is a statistical technique which can show whether and how strongly pairs of variables relate. Salary and qualifications are related – experienced, qualified people tend to earn more than new, unskilled people. The relation is imperfect. People of the same qualification vary in pay, and we can think of two employees we know where the unskilled one is paid less than the skilled. Younger workers' average wage (20–30) is less than the average weight of older ones (31–40), whose average wage is less than that of people (51–55). Correlation can tell researchers how much of the variation in employees' qualifications relate to wages.

Research data may contain unsuspected correlations. The researcher may suspect there are correlations, but might be unaware of which are strongest. An intelligent correlation analysis can lead to a greater understanding of data.

Pearson correlation coefficient

The Pearson correlation coefficient is a useful statistical formula measuring the strength between variables and relations. This formula is the Pearson R test. When conducting a statistical test between two variables, please conduct a Pearson correlation coefficient value to define how strong this relation is between those two variables.

To determine how robust, the relation between two variables is, a formula must be followed to produce what is the coefficient value.

The coefficient value can range between -1.00 and 1.00. If r is close to 0, this means there is no relation between variables. If r is positive, this means as one variable gets larger the other gets larger. If r is negative this means as one gets larger, the other gets smaller – inverse correlation.

Pearson coefficient formula

The Pearson coefficient formula is:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

This is normally done for a sample

where:

- x_i: × values
- y_i: y values
- $\bar{x}$: mean values of x
- dashed y mean values of y
  divided by:
- deviation scores of x and y as shown in the formula
- If the coefficient value is in the negative range, then this means the relation between variables is negatively correlated, or as one value increases, the other decreases.
- If the value is in the positive range, then this means the relation between variables correlates positively, or both values increase or decrease together.

Let us consider analysing the relation between workers' age and stated level of income. There might be a positive or negative relationship between workers' age and income level. After conducting the test, the Pearson correlation coefficient value is +.45.

There is a reasonable positive correlation between the two variables, and the strength of the relation is positive and strong. The researcher could conclude there is a strong relation and positive correlation between workers' age and income. As employees age, income increases.

Example

The years of service and the wage earnings of workers are provided below.

Required:

Calculation of the Pearson coefficient.

Years of service	Earnings in €
2	1000
3	1200
4	1200
5	1500
6	1600
7	1800
8	2000
9	2300
10	2700

Table of information

X Values $\sum = 54$, Mean = 6 $\sum(X - M_x)^2 = SS_x = 60$	*Y Values* $\sum = 15300$, Mean = 1700 $\sum(Y - M_y)^2 = SS_y = 2500000$	*X and Y Combined* $N = 9$ $\sum(X - M_x)(Y - M_y) = 12000$
R Calculation $r = \sum((X - M_y)(Y - M_x)) /$ $\sqrt{((SS_x)(SS_y))}$	$r = 12000 / \sqrt{((60)(2500000))}$ $= 0.9798$	*Pearson coefficient* $r = 0.9798$

The value of R is 0.9798. This is a strong positive correlation, which means that high X variable scores go with high Y variable scores. A very strong correlation between age of workers and the wages that they earn in an incremental way.

The value of R^2, the coefficient of determination, is 0.96.

A correlation report can also show a second result of each test – statistical significance. In this case, the significance level will tell you how likely the correlations reported may be due to chance in the form of random sampling error. If you are working with small sample sizes, choose a report format including the significance level. This format also reports the sample size.

Other examples of correlation

- Productivity index and skills of employees
- Workplace attendance and punctuality reinforcement measure
- Transparency of activities and yardstick like traceable documents in the HR office

Only two variables are needed in the correlation model as a convention. If added variables exist, then please use the regression equation.

Regression Analysis

In statistical modelling, regression analysis is a statistical process for estimating the relations among variables. This includes techniques for modelling and analysing several variables, when the focus is on the relation between a dependent variable and one or more independent variables.

Regression analysis helps understand how the dependent variable's typical value changes when any one of the independent variables is varied, while the other independent variables are held fixed.

Regression analysis estimates the conditional outcome of the dependent variable given the independent variables – that is, the average value of the dependent variable when the independent variables are fixed.

The estimation target is a function of the independent variables called the regression function. In regression analysis, this is of interest to characterise the dependent variable's variation around the regression function which can be described by a probability distribution.

Regression analysis is used in forecast and prediction where the use has large overlap with experimentation. Regression analysis is used to understand which among the independent variables relate to the dependent variable, and to explore the forms of these relations.

Assumptions of regression analysis

Classical assumptions for regression analysis include:

- The sample is representative of the population for the interpretation prediction.
- The error is a random variable with a mean of zero conditional on the explanatory variables.
- The independent variables are measured with no error.
- The independent variables are linearly independent, i.e. it is not possible to express any predictor as a linear combination of the others.

- The errors are uncorrelated, that is, the variance-covariance matrix of the errors is diagonal and each non-zero element is the variance of the error.
- The variance of the error is constant across observations. Otherwise, weighted square means or other methods might instead be used.

These are enough conditions for the least – squares estimator to have desirable properties. In particular, these assumptions imply the parameter estimates will be unbiased or consistent and efficient in the class of linear unbiased estimators. This is important to note actual data rarely satisfies assumptions. So the method is used even though assumptions are false. Variation from assumptions can sometimes be used to measure of how useless the model is. More advanced treatments may relax these assumptions. Reports of statistical analyses often include analyses of tests on the sample data and methodology for model's fit and usefulness.

Regression Analysis formula

In linear regression, the model specification is the dependent variable, y is a linear combination of the *parameters* but need not be linear in the *independent variables*. For a simple linear regression, the formula could be expressed as:

$$y = ß + ßx_1 + ßx_2 + ßx_3 + e$$

Simple linear regression

Simple linear regression is a way to describe a relation between two variables through an equation of a straight line, called line of best fit, which models this relation.

Linear Regression Formula

This calculator uses this formula to derive the equation for the line of best fit:

$$y = a + bx$$

where

$$b = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} \quad \text{and} \quad a = \bar{y} - b\bar{x}$$

Example

Hours – Independent x	Output Index Dependent y
35	100
36	120
37	110
38	125
39	142
40	145
41	150
42	152
43	155
44	155

Calculations

A statistical calculator works out:

Best-fit values	
Slope	6.067 ± 0.8621
Y-intercept	-104.5 ± 34.14
X-intercept	17.23
1/Slope	0.1648
95% Confidence Intervals	
Slope	4.079 to 8.055
Y-intercept	-183.3 to -25.80
X-intercept	6.316 to 22.79
Goodness of Fit	
R square	0.8609
Sy.x	7.830
Is slope significantly non-zero?	
F	49.52
DFn, DFd	1,8
P Value	0.0001
Deviation from horizontal?	Significant
Data	
Number of XY pairs	10
Equation	Y = 6.067X − 104.5 or Y = − 104.5 + 6.067X

The regression line including the line of best fit is below:
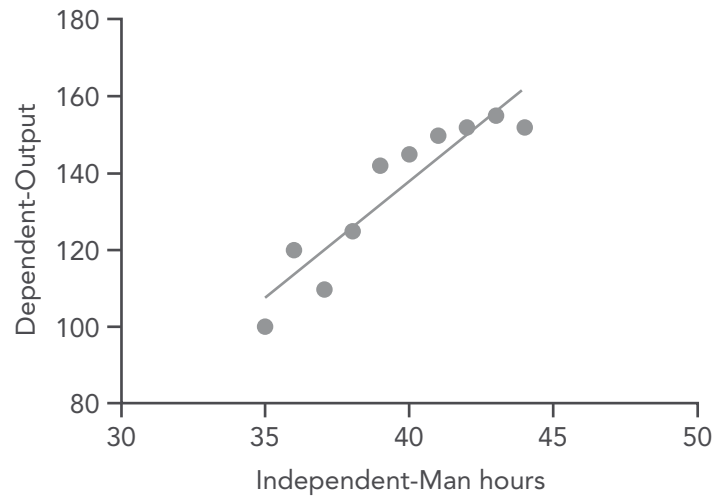
The p-value reads 0.0001 which is significant. In this case, there is a strong relation between variables and there is significance between hours and output.

Multiple Regression

Multiple regression is an extension of simple linear regression. This is used when the researcher wants to predict the value of a variable based on the value of two or more other variables. The variable the researcher wants to predict is the dependent variable or, the outcome, target, or criterion variable.

One could use multiple regression to understand if staff performance can be predicted based on training time, test, or short-exam anxiety, lecture attendance, and gender.

Multiple regression allows the researcher to determine the overall fit – variance explained – of the model and the relative contribution of each predictor to the total variance explained.

The researcher might want to know how much of the variation in staff performance can be explained by training time, test anxiety, lecture attendance and gender but also the relative contribution of each independent variable in explaining the variance.

The table is presented below for female employees.

Expected Output- Provided by computer to 2 d.p	Predicted Output Y	Training time between 28-38 hours X_1	Test/100 X_2	Attendance/10 X_3	Gender- Female-Age X_4
148.8	150	30	60	10	20
131	125	29	62	9	22
122.3	120	32	57	8	23
149.3	160	34	67	9	24
181.7	180	37	72	10	26
194	190	38	75	10	30
177.7	180	30	65	10	32

Computed formula:

$$y = -175.18 + 2.7343 \, X_1 + -0.072653 \, X_2 + 19.746 \, X_3 + 2.4429 \, X_4$$

to 2 d.p

$$y = -175.2 + 2.7 \, X_1 + -0.07 \, X_2 + 19.75 \, X_3 + 2.44 \, X_4$$

Slope variances per input are:

-175.2 as intercept
$2.7 \, X_1$ as training time reasonable value
$-0.07 \, X_2$ as test variances
$19.75 \, X_3$ as high consistent training time
$2.44 \, X_4$ as age gap of reasonable value

Total Output was predicted as: 1,105. Expected Output was: 1.104.8.

References

Easy Calculation (2017) Regression calculator, https://www.easycalculation.com/statistics/statistics.

Eduplace (2016) Linear Equations, Houghton Mufflin Company.

Graphpad Software (2017) Correlation and regression calculations, Graphpad Software Company.

Laerd Statitsics (2013) Multiple Regression Analysis using SPSS Statistics, Lund Research Ltd.

Maths is Fun (2016) Linear Equations, https://www.mathsisfun.com/algebra/linear-equations.

Research Methodology Net (2014), Correlation analysis, http://research-methodology.net/research-methods/quantitative-research/correlation-regression.

Soper, D. (2017) Online regression calculator, Free Statistics calculators 4.0, DanielSoper.com.

Multiple-Choice Questions

1. In a linear equation, $y = mx + c$, the m notation is known as
   A. independent variable.
   B. dependent variable.
   C. intercept.
   D. gradient.

2. An increase in overtime hours can boost productivity. The dependent variable is
   A. increase.
   B. boost.
   C. overtime.
   D. productivity.
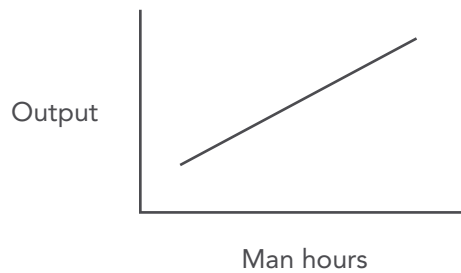
3. The diagram represents a correlation. This could be evaluated as

Output

Man hours

   A. a weak negative correlation

   B. a strong negative correlation

   C. a weak positive correlation

   D. a strong positive correlation.

4. There was a weak correlation between employee supervision and the decrease in absenteeism. An appropriate r value could be

   A. +0.9.

   B. -0.4.

   C. +0.01.

   D. -0.8.

5. In a correlation equation, the R^2 value is

   A. the correlation coefficient.

   B. the square of the correlation coefficient.

   C. the least square value of the correlation coefficient.

   D. the square of the regression coefficient.

6. A correct way of writing a regression equation is

   A. $y = ß + ßx_1 + e$.

   B. $y = ß + ßx_1 + ßx_2 + ßx_3 + e$.

   C. $y = ß + e$.

   D. $y = ß + ßx_1 + e$.

7. $y = ß + ßx_1 + ßx^2 + ßx_3 + e$

   A. unassured value for a given investment.

   B. unpredicted value for a given investment.

   C. unknown value for a given investment.

   D. anticipated value for a given investment.

8. The Pearson coefficient formula is stated as:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

The square rooted notations refer to

A. the standard deviation values.

B. the normal deviation values.

C. the normal distribution values.

D. the mean values.

9. A regression notation is represented as y = -10 + 2.7X_1 + 27X_2 -8X_3. The most influential factor in the formula would be

A. 10.

B. X_1.

C. X_2.

D. X_3.

10. The difficulty of relying on a regression line comes when relationships are

A. linear.

B. squared.

C. curvilinear.

D. hyperbolic.

Answers: 1D 2D 3D 4B 5B 6B 7A 8A 9C 10C

9 INDEX NUMBERS AND FISHBONE DIAGRAMS APPLIED TO HRM

On reaching the final chapters of this book, two short topics address HR students' needs with understanding concepts the students come across in studies and research. Indexing and fishbone concepts are developed in this lesson to help students be familiar with such concepts while use them as accessories in research.

The Indexing concept is key to grasp knowledge of trends affecting business. This comes from consumer price indexing, inflation, economic forecasts, and related areas better linked with economics. Labour economics do matter and are aligned with indexing concepts. Students could use the indexed concept to make comparisons with the recent past and interpreting the future through technical data interpolation. Bearing in mind research outputs are linked with forecasts, HR students might formulate ideas using the indexed variable as a research and forecasting tool.

Indexed numbers have a base value starting at 100 and successive values will simply rise or even go below such a standard. This value helps the student understand how variables fluctuate over time and how the trend is followed. The base year remains an important determinant for forecast because this could not remain the same for a long time. In economic forecasts, the indexed base year must change in relation to recent developments. A transition from a developing agricultural economy to an industrial economy changes the base year but there might also be recessions or crises making the researcher move to a new indexed base year.

The second part of the lesson considers the fish-bone diagram as a useful research toolkit. This cause and effect diagram can be used at the start of a research project to carve the research idea. For instance, various causes to a problem might be developed and effects might be also observed. The diagram could be used in the conclusion chapter to show how the causes for a problem were identified and what actions were taken to better address or solve the problem.

Ishikawa's fishbone diagram could improve quality systems in an organisation and this application to an HR situation might be relevant. This could apply to a project where the problems are in a company and attempts to solve by determining causes and effects.

Both index numbers and the fish-bone diagram could be improvised in research, especially when creativity improves research quality. By showing reviewers or panel members, a certain mastery of such concepts, the outcome of a presentation or research work could only be improved by stating such concept's applicability in research and showing an applied technique to HR research could have a positive impact.

The Laspeyres Index

German economist Étienne Laspeyres (1834–1913) proposed the Laspeyres Index for measuring prices or quantities relating to a base period.

A Laspeyres price index is computed by taking the ratio of the total cost of purchasing a specified group of commodities at current prices to the cost of that same group at base-period prices and multiplying by 100. The base-period index number is thus 100, and periods with higher price levels have index numbers over 100.

The distinctive feature of the Laspeyres index is this uses a group of commodities purchased in the base period as the basis for comparison. In computing the index, a commodity's relative price (the ratio of the current price to the base-period price) is weighted by the commodity's relative importance to all purchases during the base period.

The Laspeyres price index tends to overstate price increases because, as prices change, consumers alter purchasing decisions by choosing fewer products with large price increases while buying more products showing low or no price increases. If consumers can do this without being disatisfied, the use of base-period commodity selections tends to overstate declines in the standard of living. Similar to the price index, the Laspeyres quantity index uses base-period prices to compare aggregate production levels in two periods.

Example

The following table illustrates the rise in aggregate labour costs which could be of interest to an HR researcher. A progression of values is provided and a base year must be established.

Year	Aggregate labour cost in agriculture in €
2011	1500
2012	1575
2013	1620
2014	1640
2015	1700
2016	1650
2017	1800

Initial observation

The labour costs have kept rising over the years and a €300 rise has been evidenced in the agricultural sector. The first observation states labour costs have risen in industry and could be a reason production overheads rose and made the industry less competitive.

Regarding HR, there could be a plausible reason for the decline of labour in the industry and the increase in unemployment in the sector. What could be done with the data provided?

The Laspeyres Index could help declare 2011 as the base year. Hence €1500 would be the base year salary. To index this the value is considered as 100. The table is now improved as follows. The value of 100 is reached by dividing 1500 by 1500 and multiplying it by 100. The same should apply for all others, say 2012 equals to 1575/1500 × 100 = 105.

Year	Aggregate labour cost in agriculture in €	Index
2011	1500	100
2012	1575	105
2013	1620	108
2014	1640	109
2015	1700	113
2016	1650	110
2017	1800	120

A 20-point increase was noted in the industry which is a clearer observation to the aggregate labour figures. Over 6 years, the index moved by 20 with an average of 20/6 = 3.33 per annum. In 2018 and 2019 there could be a rise of 3.3% annually and the index might rise to 126.

The figures are now input on a chart.

Observations

Wages rose steadily from 2011 to 2015 and the indexed values were correctly interpreted. In year 2016, wages dropped a little which could be any causal factor like a lack of competitiveness, reduction in aggregate demand in industry, or adverse economic conditions.

In 2017, aggregate labour costs rose sharply with a kink in the curve. This could to some extent affect the indexed value and future evaluations or scenarios. This is where we must be careful.
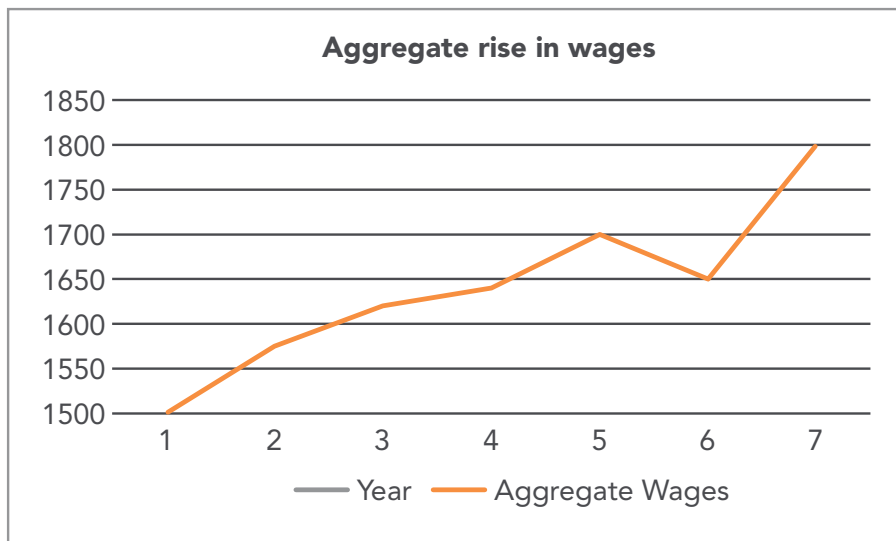
The indexed chart, similar to the one above is represented below. In this case, the understanding is clearer, more suitable to interpolate compared with the previous table which was useful but less efficient to gauge forecasting.
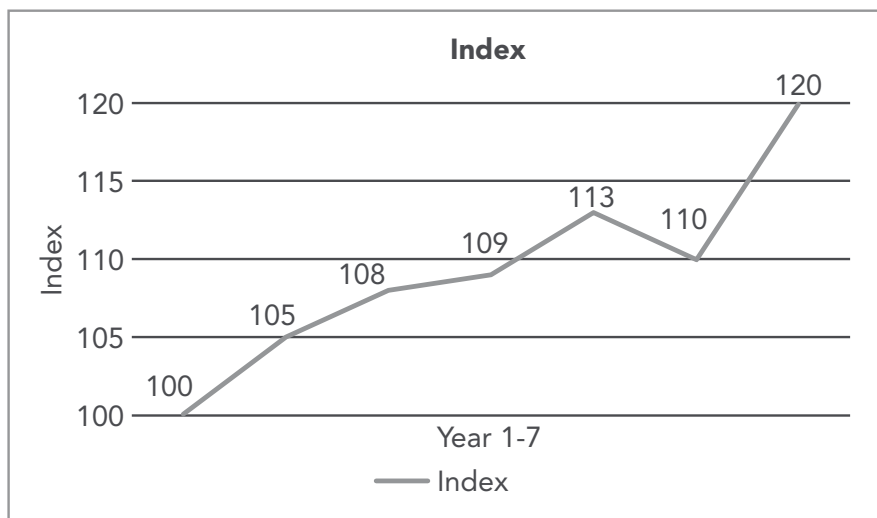
A realignment of the base year

A base year is the first of a series of years in an economic or financial index. A base year in an index set to an arbitrary level of 100. New, up-to-date base years are introduced to keep data current in a particular index. Any year can serve as a base year, but analysts often choose recent years.

There may be a possibility to realign the base year if new trends are observed. For instance, if wages are move from €1700 to € 2500, 2017 could be the base year.

In the United States, base years may extend up to 25 years before anew base year is created.

Cross tabulation of the index

A cross tabulation of the index becomes a useful tool in business and for HR managers. This allows for comparative analyses in an industry but also encourage managers and researchers to become more critical of figures. A five-year comparative cross tabulation of issues pertaining to HR is provided below.

Industry	2014	2015	2016	2017	2018
Mining	100	110	110	108	107
Agriculture	100	97	98	96	94
Manufacturing	100	103	105.5	107	108
Education	100	102	104.5	105	107
Tourism and Travel	100	104	106	108	108
Financial Services	100	105	107	109	113.5
Information and Communication Services	Not Available	100	108	112	116.5

To better gauge progressions the chart is amended to indices below, meaning digits less than ten.

| Industry	2014	2015	2016	2017	2018
Mining	10	+0.10	+0.10	+0.08	+0.07
Agriculture	10	-0.03	-0.02	-0.04	-0.06
Manufacturing	10	+0.03	+0.05	+0.07	+0.08
Education	10	+0.02	+0.04	+0.05	+0.07
Tourism and Travel	10	+0.04	+0.06	+0.08	+0.08
Financial Services	10	+0.05	+0.07	+0.9	+0.11
Information and Communication Services	Not Available	100 or base year	+0.08	+0.112	+0.116

Progression points observations

In the mining sector, there has been stagnation in the last three years with a three-point decline.
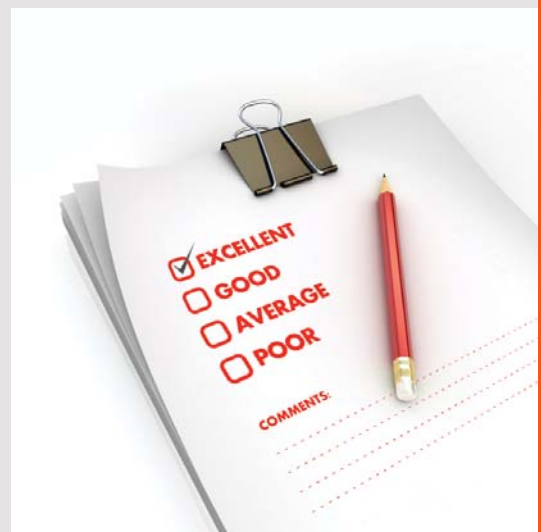
Agriculture's decline is steady reaching a record low of four points since 2014 which explains an industry downturn.

There has been constant growth in the manufacturing sector of over five points and this looks likely to progress in years to come.

Education looks to be a promising sector with constant growth over the years and this is likely to progress by two additional points in the next quarter.

Tourism and travel progressed up eight points but seem to stagnate in the past two years. This might be critical for labour growth in the sector.

Financial services grow steadily with a constant 2-point growth and an aggregate 11-point growth meaning comparatively high in growth if compared to agriculture with a 14-point comparative move.

ICT services keep growing with some stagnation in the last two years with an index move of 0.04 stating there is room for improvement. Labour in this sector is in growth but the sector needs revamping to attract young unemployed people.

Relevance of Index numbers

According to the Office of National statistics (UK), Index numbers measure changes and simplify comparisons. Economists are interested in how changes in the monetary value of economic transactions can affect changes in price to measure inflation and changes in quantity to measure sales volume or economic output. Index numbers measure these changes over time. But index numbers can also make other comparisons, such as between regions of a country.

Mathematically, an economic index number is an average of the many different changes in price or quantity of the goods and services included in the index. There are many different valid ways of determining this average and, consequently, many different possible estimates of average price and quantity changes. It is therefore important that index numbers are calculated in a way that is appropriate for their intended use. These do apply also to human resource managers particularly when this concerns finding out data relating to labour, performance indices, industry forecast, competitiveness and growth indices, etc. They also help synthesising a large amount of data in a simple, technical and more understandable manner.

The Cause and Effect Fishbone diagram

A cause and effect of fishbone diagram, or Ishikawa diagram, is a conception tool for sorting out potential causes of a problem to identify root causes. A fishbone diagram helps brainstorm sessions to focus discussion.

After the group brainstormed all possible causes for a problem, a facilitator helps the group rate potential causes according to importance and diagram a hierarchy. The diagram's design looks like a fish skeleton.

When using a team approach to problem solving, there are often many opinions as to the problem's origin. One way to capture these different ideas and stimulate the team's brainstorming on root causes is the cause and effect diagram, or fishbone.

The fishbone visually shows the potential causes for a problem or effect. This is useful in a group setting and for situations where little quantitative data is available for analysis.

Since people often like to decide what to do about a problem, fishbones can help bring out a thorough examination of issues behind the problem which might lead to a concrete solution.

To construct a fishbone, researchers state the problem as a question, such as:

- Why are employees absent on Mondays?
- What is the reason behind employee frustration at work?
- What explains a high turnover at work?

Framing as a question helps brainstorming, as each root cause idea should answer the question. The team should agree on the statement of the problem and then place this question in a box at the head of the fishbone.

The rest of the fishbone consists of one line drawn across the page, attached to the problem statement, and several lines, or bones, coming out from the main line. These branches feature different categories. Which categories to use depend on you. There are some standard choices:

Fishbone suggested categories

Service Industries	Manufacturing Industries	Process
Policies	Machines	Determine Customers
Procedures	Methods	Advertise Product
People	Materials	Create Purchase
Plant/Technology	Measurements	Sell Product
	Mother Nature(Environment)	Ship Product
	Manpower(People)	Upgrade

Simon, K. (2017) Cause and effect Diagram, iSixSigma.com

For the sake of HRM students, the following concepts could be individually developed although most of them could apply to the model. Illustrations are in the diagrams below.

Case One: Employee frustration at work that increases turnover

The first diagram states the problem and factors.
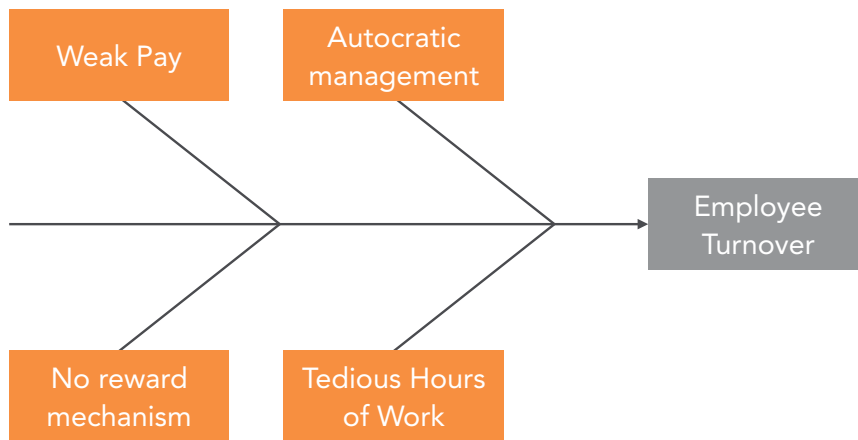
Next, the HR Manager will have to work out with her colleagues and determine causes for such a problem at work. She could analyse and understand reasons behind the cause and effect.

The cause and effect displays the problem differently. Above, all key ideas are forwarded while in the next diagram the cause and effect is explained in the top and bottom bones. Note the diagram is a creative effort and there is no rule to present this. This depends on the HR's creativity of the HR team including the researcher.

Cause and effect design

The second diagram helps in explaining the cause and effect in a more elaborate manner.
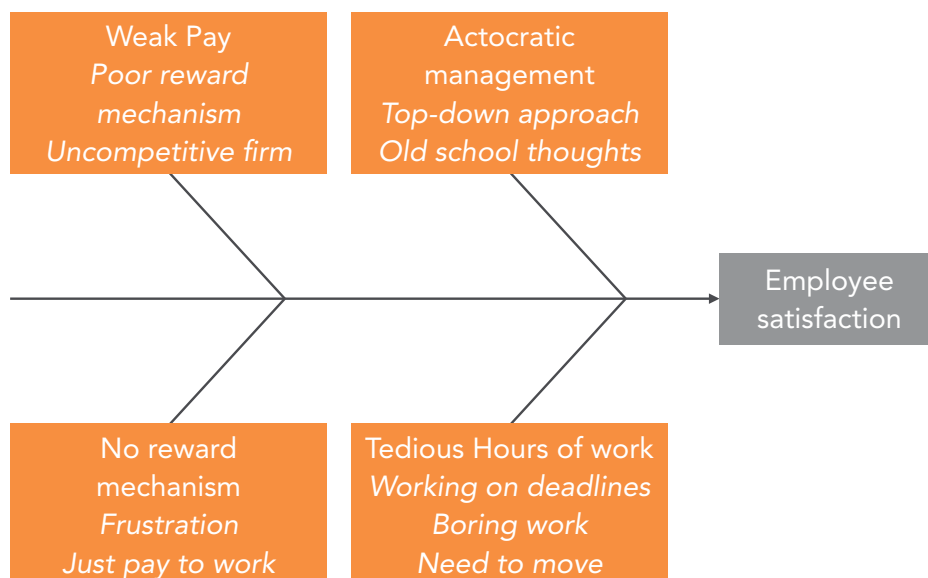
From the cause and effect illustrated, a brainstorming session would consider how to remedy the situation. Maybe the situation must be improved by changing the diagram. Or the problem could be addressed through consultation and meetings.
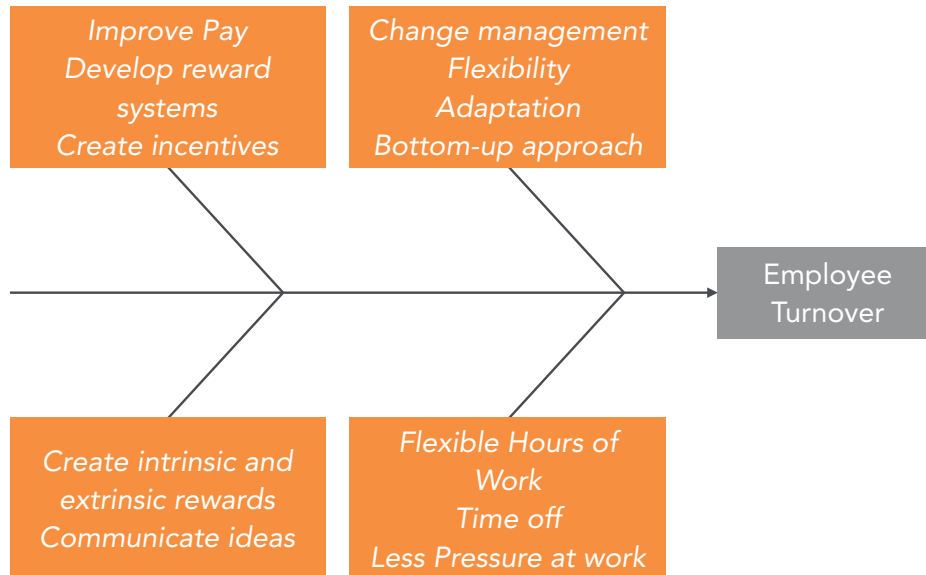
Precautions on using the Fishbone Concept

Below is a list of precautions that a fishbone diagrams never provide all solutions to problems. This list addresses but never fixes problems.

The fishbone diagram is not a problem solver.

It is clear that the Fishbone diagram is not a problem solver. Based on the idea behind the diagram, it identifies the cause and effect of a phenomenon. Ideally, it is good to highlight what were the main elements leading to an unsolved problem.

Such a diagram identifies causes and effects.

The main object of the Fishbone diagram is to assess causes and effects. Causes are the main reasons that lead to a problem or the creation of such a problem. Effects are the aftermaths of the problem. Example, a bottleneck in distribution creates delay in gaining the information right.

The fishbone could be a useful tool used by researchers to analyse a problem.

Researchers could analyse a problem using this technical approach. Very often, one hears of a manager stating 'let us go about brainstorming'. The session looks like all problems and ideas are developed in an unstructured manner and there needs to be coherence in the elements discussed. Poor structuring might not lead to a real identification of a problem.

For academic work, fishbone diagrams help sorting out or crafting a problem.

Sorting out a problem is important in research. For example, hypotheses used by researchers could be developed. The cause and effect is usually linked with the independent (cause) and dependent (effect) variable. Researchers could include this concept while formulating their hypotheses.

Solutions and proposals can be made in the diagram.

Although solutions are not directly obtained from the Fishbone diagram, there could be certain ways of stating the solution or finding out a potential solution. It must be borne in mind that the initial approach to a solution does not necessarily conclude the research. Rather, an improved diagram that states how the problem was addressed could be helpful in addressing the problem.

Fishbone diagrams have been used in japan to discuss problems faced at the management level but the simplicity of the model makes it a useful one even at the operational and tactical level to better address a problem and forward proposals to management.

References

A Dictionary of Statistical Terms. (2017), Laspeyres Index, Fourth Edition, International Statistical Institute, Longman Group, London.

Mindtools. (2017) Cause and Effect Analysis, Identifying the Likely Causes of Problems, https://www.mindtools.com.

Simon, K. (2017) Cause and effect Diagram, iSixSigma.com

Multiple Choice Questions

1. An index number helps
   A. scale data in a simple manner.
   B. upscale data in a simple manner.
   C. scale data in a complex manner.
   D. generate data in a random manner.

2. An index base is 100. A record low would be scaled at
   A. 98.
   B. 92.
   C. 101.
   D. 110.

3. A base year is reviewed after
   A. 25 years.
   B. a sudden fluctuation of data.
   C. a new data trend is observed.
   D. a couple of years.

4. Regarding labour prices, HR researcher would choose
   A. a CPI index.
   B. a household expenditure index.
   C. a GDP index.
   D. an employment index.

5. An analysis of information across industry would relate to
   A. a multi-variate or cross analysis.
   B. a bivariate analysis.
   C. an indexed observation.
   D. an economical analysis.

6. The Fishbone diagram refers to
   A. review and reasoning.
   B. cause and effect.
   C. analysis and decision.
   D. discussion and debate.

7. A cause for employee turnover could be
   A. extensive management support.
   B. two-way communication.
   C. low employee motivation.
   D. a fairly competitive work environment.

8. An effect in the Fishbone diagram regarding attrition at work could be
   A. high labour availability.
   B. young staff.
   C. motivated staff.
   D. skeletal staff.

9. A model which can be applied to HR strategy by using the Fishbone diagram is
   A. the Mc Kinsey 7 s model.
   B. the product portfolio matrix.
   C. Porter's value chain model.
   D. the traditional marketing mix.

10. The difficulty of relying on a Fishbone diagram is this
    A. is a problem identifier.
    B. is not an immediate problem solver.
    C. develops a problem mind map.
    D. is complex to use.

Answers: 1A 2B 3C 4D 5A 6B 7C 8D 9A 10B

10 MULTIPLE CHOICE ASSESSMENT

This test comprises all topics covered in this book. Students can go through the text and find questions in the topics. There is one answer per question. Answers are at the end of this test. This activity allows students to gauge knowledge on HRM research. Some questions are direct while others need thinking and evaluation. This book offers 200 objective tests for practice to students and researchers.

1. Much of a useful research project involves the researcher working alone, but high-level…skills are needed.
   A. communication.
   B. correlation.
   C. information.
   D. tactical.

2. One of the most significant external factors for HR is workforce growth. This refers to
   A. a threat.
   B. an opportunity.
   C. a weakness.
   D. a strength.

3. The first step in the process is to problematise or develop a
   A. research model.
   B. research outcome.
   C. research question.
   D. research experience.

4. A…is a diagram used to organise information visually.
   A. workbook.
   B. working plan.
   C. route map.
   D. mind map.

5. The aim of the…abstract is allowing readers to decide whether to read the whole dissertation or just dip into those chapters of particular interest.
   A. descriptive.
   B. long.
   C. short.
   D. ideal.

6. Well-written annotations help…key information in each article or book.
   A. develop.
   B. crystallise.
   C. expand.
   D. broaden.

7. Conducting a…review demonstrates an author's knowledge about a particular field of study.
   A. topic.
   B. book.
   C. literature.
   D. structural.

8. The…controls research using primary methods.
   A. external observer.
   B. reviewer.
   C. researcher.
   D. independent company.

9. …involve taking organised notes about occurrences in the world.
   A. Observations.
   B. Reviews.
   C. Analyses.
   D. Studies.

10. …refers to situations where the sample does not reflect the characteristics of the target population.
    A. Sampling methodology.
    B. Sampling bias.
    C. Sampling size.
    D. Sampling determination.

11. A…is often interpreted as the opposite of a sample as its intent is to count everyone in a population rather than a segment.
    A. subject of study.
    B. test.
    C. survey.
    D. census.

12. A longitudinal study follows the same sample over time and makes…observations.
    A. new.
    B. single.
    C. repeated.
    D. some.

13. A stacked chart can represent…of data in a mirror-like effect and these can clearly show contrasts between the data exposed
    A. one set.
    B. two sets.
    C. several sets.
    D. four sets.

14. The concept of…helps an HR manager divide a range of data by highlighting which one is lowest and which one is highest.
    A. quartile.
    B. upper quartile
    C. percentile.
    D. lower quartile.

15. In a normal distribution, the following concept applies.
    A. Mean< Mode> Median.
    B. Mean< Mode= Median.
    C. Mean=Mode> Median.
    D. Mean=Mode=Median.

16. An HR manager thinks the wage increase would be between 10% and 15% of the total wage. A 95% confidence interval would be written as
    A. 5% CI : 0.10≤ CI ≥0.15.
    B. 95% CI : 0.10≤ CI ≥0.15.
    C. 95% CI : 0.10≤ CI ≥0.15.
    D. 99% CI : 0.10≤ CI ≥0.15.

17. Z-tests are used if the sample size is…or the population variance known.
    A. small.
    B. large.
    C. less than 30.
    D. up to 15.

18. A…lets the researcher know if the mean changed between the first and second survey.
    A. paired t-test.
    B. single t-test.
    C. one sample t-test.
    D. directional t-test.

19. In a Chi-Square test, there are 3 rows and 3 columns, the degrees of freedom are
    A. 9.
    B. 4.
    C. 6.
    D. 2.

20. In a regression equation like $y = 20 + 3 X_1 - 10 X_2 + 5 X_3 + 4 X_4$, the most powerful variable is
    A. X_2.
    B. X_1.
    C. X_3.
    D. X_4.

Answers: 1A 2B 3C 4D 5A 6B 7C 8D 9A 10B

11 D 12 C 13 B 14 A 15 D 16 C 17 B 18 A 19 B 20 A

Rating:

16–20: Excellent – Great grasp of topic and knowledge.

12–15: Satisfactory – Well done, needs some refinement.

10–11: Fair – Acceptable standard of pass, can do better.

Less than 10: Needs revision and more attention to key concepts.

11 EXAMINATION PAPER ASSESSMENT

This section comprises two papers which could suit assessment. Suggestions are at the end of this section and these could become suitable answers. The question papers can be used both by the learner and the tutor for practice and revision as most of the topics covered are also dealt with in this textbook.

PAPER ONE

SECTION A: QUALITATIVE QUESTIONS

1. A company's human resource manager is conducting a situational analysis in her firm.
   a) What is the relevance of a situational analysis?
   b) Identify two threats in an organisation.
   c) Identify two opportunities in an organisation.
   d) Use the SWOT model and explain how a company aligns internal strengths with organisational goals.

2. The following notes are for you.
   - Talent management management.
   - Evidence of talent development and good performance.
   - Difficulties in defining talent.
   - Talent needs retention, development, and performance.
   - Survey respondents: 75
   - Good related values obtained.

   Please prepare a single paragraph abstract below 150 words using the notes. Please highlight three to four key words.

3. Here is information on human resource planning for you.
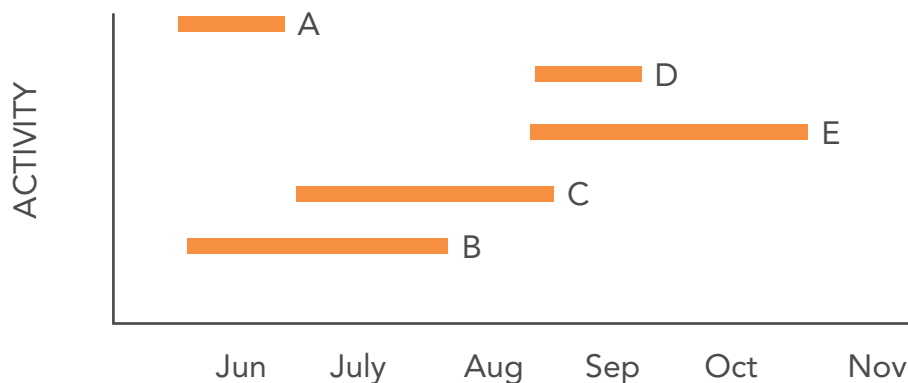
   The HR Manager designed the GANTT chart as she drafted ideas. Her chart needs labelling and amending. Identify activities in the following sequence: screening applications, selection test, interview, job offer, induction training. Draw and label an amended chart and state how the HR plan becomes proficient.

4. Your human resource manager informs you must prepare a report on the importance of training at work. You will present it at a meeting in June 2018.
   a) You want to introduce a literature review before gaining primary data. What are key items you will address in the literature review?
   b) Which method would you use to gain primary data on employee training in the firm?
   c) Write **two** research questions you want to find on employee training in your firm.

SECTION B: QUANTITATIVE QUESTIONS

5. The units produced by nine different employees of a firm are as follows:
   10, 12, 14, 10, 20, 14, 16, 16, 14

   a) What is the modal value in production units?
   b) What is the mean production unit?
   c) Calculate the median production unit?
   d) Using the mean value, calculate the variance using the formula $\sum[y-\bar{y}]^2$.
   e) The standard deviation is calculated as 3.16. What does this mean?
   f) Draw a chart to represent the units output. [3]
   g) Determine the lower and upper quartile. [2]

6. Read the following information.
   - The option of working overtime is YES with a probability of 0.7. and the opportunity cost is € 700,000.
   - The option of not working overtime has a negative opportunity cost is € 300,000.

   a) Draw a probability decision tree to illustrate the situation. [3]
   b) Calculate each option. [3]
   c) Find the combined estimate and determine the viability of the option. [4]

7. The situation below explains the choice of employees in the human resource department of a firm regarding vacations. Consider the situation below and answer the questions.

Employees on holidays

■ Easter ■ June vacations ■ December
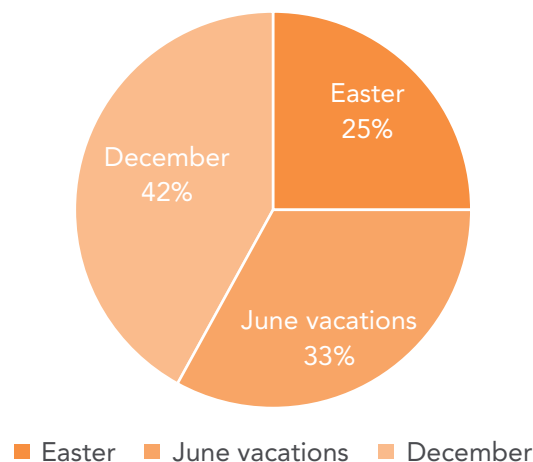
a) Calculate the angle regarding employees choose December vacations? Provide the value in the simplest terms.
b) 45 more employees choose December than July school vacations. Use the angles provided to explain the ratio December to July vacations.
c) How many employees work in this organisation?

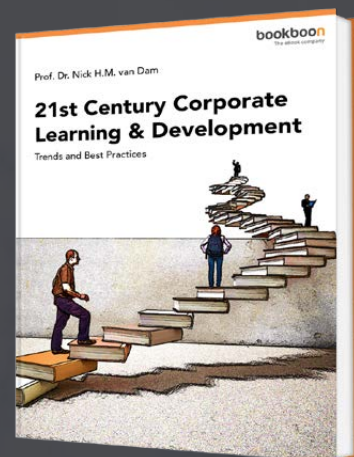A linear regression equation for productivity defined by the human resource manager is given by as follows:

$$Y = 1.2 + 2.8\,X$$

d) What is intercept on the line?
e) What is the gradient of the linear equation?
f) Sketch a curve to illustrate the situation.
g) Calculate the value if X_1 is 100 and X_2 is 200.
h) What is the limitation of this regression line?

PAPER TWO

SECTION A: QUALITATIVE QUESTIONS

1. A situational analysis exercise find reasons behind the position of a tertiary educational institution.
   a) What is the institution's relevance in a situational analysis?
   b) What might be the institution's strengths?
   c) What could be the environment's threats?
   d) Why is transforming threats into opportunities complex?
   e) How could the consolidation of strengths better the educational institution?

2. As an HR officer, you must write a structured abstract referring to the influence of foreign labour in the manufacturing sector of your country using the following guidelines.
   - Aim of the research.
   - Purpose of the study.
   - Main findings.
   - Benefits of the study.
   - Key words.

   You are advised to prepare a structured abstract of not more than 150 words. You also need to highlight 3–4 key words.

3. The following information is provided to you as HR Manager regarding human resource planning.
   Purchase of computers and software, recruitment of computer trainees, development of software, training of computer trainees, selection and retaining the best trainees, reward system for trainees.

   The following directives are for you:

   - The first three tasks should be completed within three months starting July 2017 in a sequenced manner.
   - Training will be carried out as from October.
   - Selection and retention will be as from mid-November.
   - Reward system will be effective as from December in the same year.

   Prepare a GANTT chart to illustrate the situation and explain the precautions that you have taken in doing so.

4. Your human resource manager informs you must prepare a report on sustainable employment at work. You will present this at a meeting in December 2017.
   a) You want to introduce a literature review on sustainable employment which means suitable jobs for all types of employees. Where will you derive information on this topic?
   b) Give **two** reasons why to get secondary data before primary data.
   c) Write **two** research questions you want to discuss on sustainable employment in your firm.
   d) Prepare a sample data to illustrate the type of employees and related characteristics needed for the needs of a questionnaire survey.

SECTION B: QUANTITATIVE QUESTIONS

5. The wages per day of eight employees of a firm are as follows:
   300, 310, 320, 300, 340, 350, 330, 360.

   The wages are in National Currency units (C) and depend on productivity with the basic wage of C 300.

   a) If C 300 are indexed as 100, calculate the indexed value for employees earning C 330 and C 360 respectively.
   b) Estimate the mean wage of the workers.
   c) State the modal wage of the workers.
   d) Rearrange the data and find out the median wage of the workers.
   e) If the lower-quartile is estimated at C 305, find out the upper-quartile.
   f) The standard deviation is 22.6. What does this imply?

6. Read the following information.
   Saxon HR Consultants aims at offering either direct consultancy or online tailor-made consultancy to companies needing expert HR advice. Observe the table below.

	Option A Direct Consultancy	Option B Online tailor-made consultancy
Cost of option	€ 500,000	€ 300,000
Probability of profit	0.6	0.8
Probability of loss	0.4	0.2
Expected outcome of profit	€1,000,000	€ 800,000
Expected outcome of loss	€ 450,000	€ 500,000

Draw a decision tree to explain the situation.

a) Costs of option will be deducted from final outcome. Provide a reason for this.
b) Calculate the total expected value of Option A.
c) State the net gain of option A.
d) Calculate the total expected value of option B.
e) State the net gain of option B.

7. The situation below explains the performance of employees in the production department of a firm. Consider each situation explained below and draw a labelled diagram for each one.
   a) There is a linear relationship between a rise in bonus and employee productivity index. This has a limit to some extent.
   b) There was strong negative correlation between an increase in working hours and employee productivity index.
   c) Short pauses included during long hours of work have the effect of dampening the negative correlation compared to example (b).
   d) What is the weakness of using correlation equations in a workplace situation?

Pepe tries out a hypothesis for his HR dissertation and obtains the following result.

In a t-test for comparison, Pepe obtains a p-value of 0.20 which is greater than 0.05.

e) How can you explain a hypothesis where Pepe wanted to find out the reaction of HR and Accounting students regarding the validity of a crash course before the exams?
f) Define the null hypothesis.
g) Write down the alternative hypothesis.
h) If the p-value was 0.02, which hypothesis would you formulate?

SOLVING PAPER 1

Question 1

a) A situational analysis helps in determining the existing situation in a firm and allow the organisation undertake an internal analysis with strengths and weaknesses and an external one with opportunities and threats.
b) Economic: Rise in costs of production due to imports
Technological: Replacement of manual labour by machines which is a threat.
c) Opportunities in an organisation:
Political: Government favouring deregulation hence opening up doors for competition.
Social: More people aiming to work at flexible times.
d) Aligning internal Strengths with organisational goals:
Provide training to employees.
Recruiting qualified staff to work.
Providing better facilities at work.
Compensation workers in a competitive manner.

Question 2

ABSTRACT:

This research understands talent management in a small enterprise. The research was conducted among 100 respondents out of which 75 responded positively. Effective talent development leads to good performance at work. This proved the hypothesis linking talent management with good performance is viable. Respondents mentioned difficulties in developing talent in the firm because of this vague term. The research revealed the development of talent depends on a systematic process included retention, development and performance. Scores obtained in the questionnaire proved the hypothesis and is clear talent development will have an impact on performance.

Key words: talent, performance, talent development, impact.

Question 3
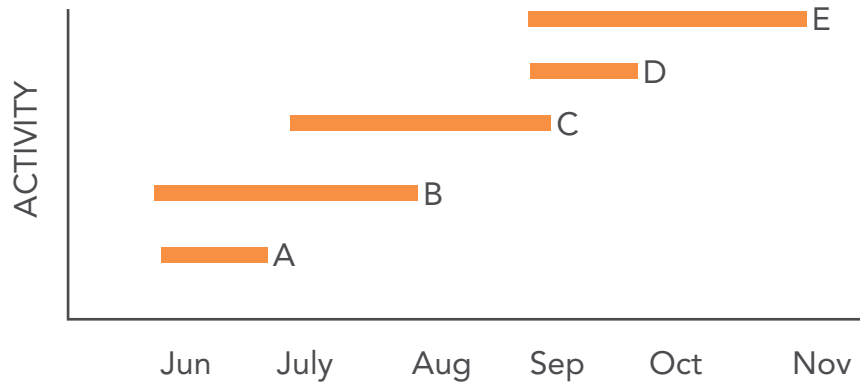
Activity A: Jun–July: Screening applications
Activity B: Jun–August: Selection test
Activity C: July–Sep: Interview
Activity D: Sep–Oct: Induction training
Activity E: Sep–Nov: Induction training

HR plan proficiency: Systematic, sequential. Activities overlap but progress smoothly.

Question 4

a) Literature Review:
   Relevance of training.
   Practice of training.
   Types of training available.
   Application to Mauritius.
   Importance of training in Mauritius.
   Selected practices of training.
   Literature gap and review.

b) Questionnaires seeking information on type of training needed.
   Observation: training methods practised at work.
   Interviewing: How training impacts at work.
   Case studies: Evaluating the benefit of training at work.

c) How does training impact organisational performance in XYZ Company? (Qualitative)
   What correlation is there between training strategy and employee productivity in XYZ Company? (Quantitative)
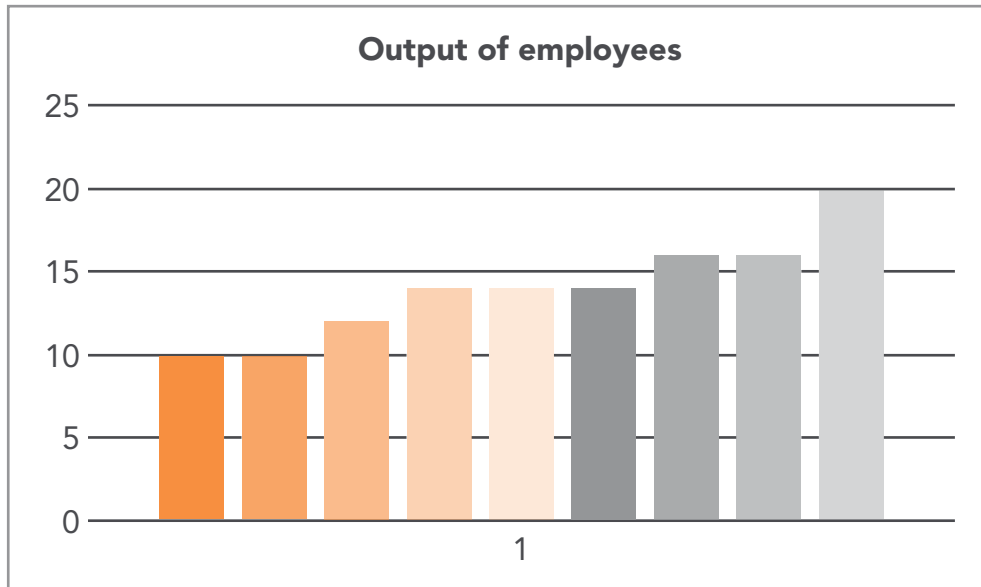
Question 5

a) Modal value is 14; it occurs thrice.

b) Mean value: 10+10+12+14+14+14+14+16+16+20/9 + =14.

c) Median: Middle value: 14.

d) Variance: $[10-14]^2 + [10-14]^2 + [16-14]^2 + [16-14]^2 + [20-14]^2$ divided by 9-1 = 8.88 to 2 d.p.

e) This illustrated the difference in value of the scores obtained from the mean. In general, they are quite large but acceptable.

f) Chart to represent situation.

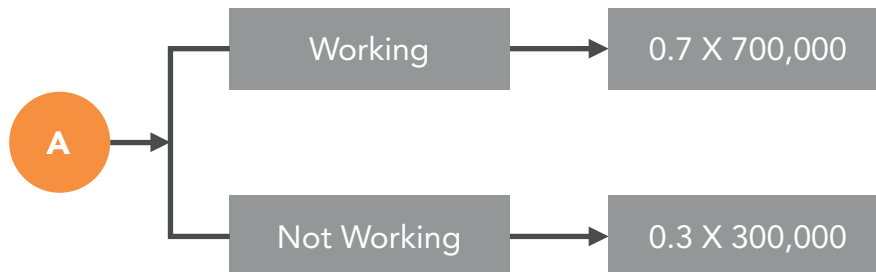**Output of employees**

g)

10,10, |12,14,1|4,14,16, |16,20

Median 14, Lower quartile: 10+12/2=11, Upper Quartile: 16=16/2 = 16.

Question 6

a) Diagram below

b) Option A: Working: 0.7 × 700,000 = € 490,000

Option B: Not working: 0.3 × 300,000 =€ 90,000

c) Combined estimate:

Option A-Option B: €490,000-90,000 = € 400,000.

The option remains viable. Working overtime adds profit to the firm.

Question 7 Part A
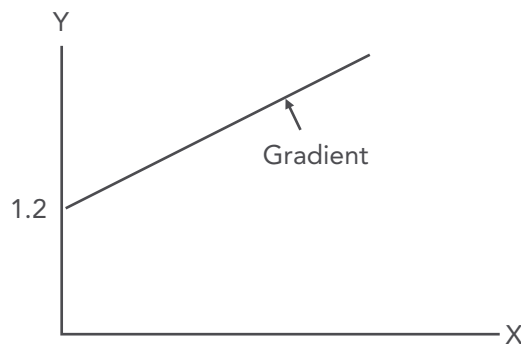
a) December vacations: = 42/100 × 360 = 1500

December vacations: (360-(90+120)0 = 1500

b) 30-degree difference means 45 employees.

c) 1500 equals: 45 × 5=225 employees

1200 equals: 45 × 4= 180 employees

900 equals: 45 × 3=135 employees.

Total employees: 225 + 180 + 135 =540.

Question 7 Part B

$$Y = 1.2 + 2.8\ X$$

a) Intercept is 1.2.

b) Gradient is 2.8

c) If X_1 equals 100, Y equals 1.2 + 2.8 (100) =29.2

d) If X_2 equals 20, Y equals 1.2 + 2.8 (20) =57.2

e) Limitation: Trend line stretches in a linear way up to a limit. Beyond that, it will tend to become curvy.

SOLVING PAPER 2

Question 1

a) Analyses the four key elements: Strengths; weaknesses; threats; and opportunities.

b) Teaching staff, well-equipped labs, infrastructure, location.

c) Competition, position in industry, funding from government.

d) Because the environment is competitive with no quick fix solution.

e) Make the institution still competitive through training of staff, research, improvement of logistics and facilities for staff and students.

Question 2

Aim of research: to see how foreign labour influences production in the manufacturing sector of a country.

Purpose of study: the study evaluates the contribution of foreign labour in selected industries in the manufacturing sector of a country s.

Main findings: there was evidence of positive contribution of foreign labour. it filled in gaps where local labour was missing. it contributed to maintaining output. it boosted productivity.

Benefits of the study: the study claims the need to maintain foreign labour in a country under the process of globalisation. this has been confirmed through hypotheses.

Key words: foreign labour, manufacturing sector, country, influence and benefits.

Question 3

A: Jul–Aug: Purchase of computers and software

B: Aug–Sep: recruitment of computer trainees

C: Sep–Oct: development of software

D: Oct–Nov: training of computer trainees,

E: Mid-Nov–End Nov: selection and retaining the best trainees

F: December: reward system for trainees.

Precaution: Sequential planning, the need for overlap, planning gaps, deadlines, etc.

Question 4

Government sources, gazettes, publications.

Literature review information.

Trends and interpretation from existing data.

First, secondary data is already available and proven information-depending on source.

Second, secondary data is the basis for future research.

What are the key elements of sustainable work in your country?

How far do employees undertake sustainable employment locally?

Types of employees: Young, middle-aged, older. Age bracket: 18–30, 31–45, 46–55, 55+,

Job type: Physical, sedentary, operational, managerial.

Question 5

a) C 300=100, therefore C 330 will be 110 and 360 will be 120 as they are divided by 3.

b) Mean wage: 300+300+310+320+330+340+350=360 divided by 8 = C 326.25

c) Most common: C 300, twice occurrence.

   300, 300, 310, 320 ,330, 340, 350, 360.

d) Since there are 8 numbers, mid value is 320+330/2= 325.

e) 300,**300,310**,320,330,**340,350**,360.

   Median Quartile: 320+330/2=325

   Lower Quartile: 300+310/2=305

   Upper Quartile: 340=350/2=345

f) Fair deviation but not too great from mean value: C 326.25. Salaries are within range of acceptability: + or – C 22.60.

Question 6

A decision tree explaining the situation.
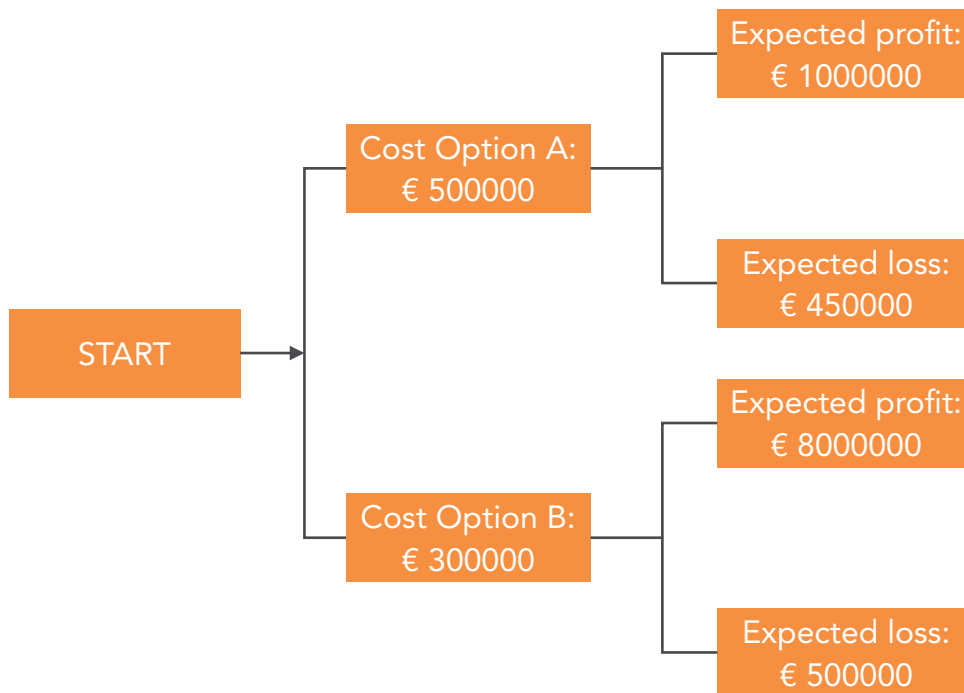
g) Total expected value of Option A.

   0.6×1000000 -0.4×450000 = € 440000 (GAIN)

h) Net gain of option A.

   €4400 000-500000= € -60000 (LOSS)

i) Total expected value of option B.
   0.8×800000-0.2×500000 = € 540 000 (GAIN)

j) Net gain of option B.
   Net gain: € 540 000-€ 300 000= €240 000(PROFIT)

k) Choose Option B

Question 7

a) There is a linear relation between a rise in bonus and employee productivity index. This has a limit to some extent.

Productivity

Bonus
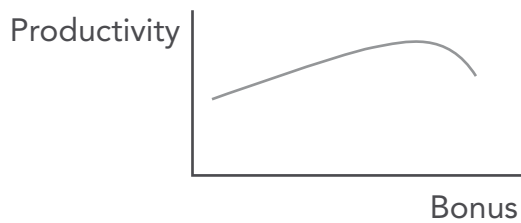
A linear rise followed by a start in decline-very soft curve or flattened one

b) There was strong negative correlation between an increase in working hours and employee productivity index.

As hours rise, productivity gets lower

c) Short pauses included during long work hours dampen the negative correlation compared to example (b).
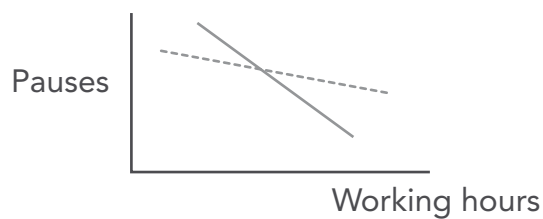
See the dampening effect with pauses, productivity can go down but not decline sharply

d) They are simplistic interpretations of a complex situation and variables are limited- only to two here.

Question 8

e) If the p-value is greater than the 5% confidence limit, it means that the crash course has had no effect on the performance of both Accounting and HR students.

f) H_O: There is no likely effect on performance whether students follow a crash course prior to the examinations or not.

g) H_1: There is a likely effect on performance when students follow a crash course prior to the examinations.

h) This is statistically significant. $p < 0.05$ at value 0.02 might mean that a crash course could have improved examination performance in this special case.

12 THE RESEARCH TOOLKIT

The research toolkit is the ultimate chapter of this book. I assume the researcher is preparing a report and is on way to work out the research findings. This is a demanding task when this part becomes the researcher's personal input. The literature review is based on what others have written on the topic while the findings section represents primary research. This part is substantive, regarding a project's failure or success.

A research toolkit below is based on an undergraduate's work or may be at a post-graduate level. Such a toolkit comprises tools and techniques in undertaking effective research. For instance, from a questionnaire, one can ask what findings to report and how to report them. The basic way might be to report information in a sequential way based from the information available and to reach a conclusion after analysing responses. The idea might seem good but there are other means of ensuring that the findings reflect upon the hypotheses developed by the researcher.

As such, the research toolkit is neither an end nor exhaustive. There are many possibilities which could be an alternative. Research students might fail crystallising information. This chapter provides students with techniques on how to better summarise information and present in a meaningful way.

There is no formula on to how to present data. Some argue diagrams are essential while others state tables and figures could be more suitable. While both arguments are good, a combination of both might improve displaying information, and sometimes, graphs and charts help readers skip the reading material and infer from diagrams.

In HRM, this book welcomes the mixed-method approach since this is a social science and borrows much from economics, psychology, and other disciplines. In this way, students can input information by using data and figures. Application software helps but can risk providing too much information beyond the scope and knowledge of the researcher. The toolkit comes in as a useful component helping in gauging the way in which data could be well transcribed in a meaningful, coherent, and appealing manner. This is also the requirement for a competent presentation of data making sense to those supervising the student or readers wishing to grasp information from the findings.

Finding the hypotheses
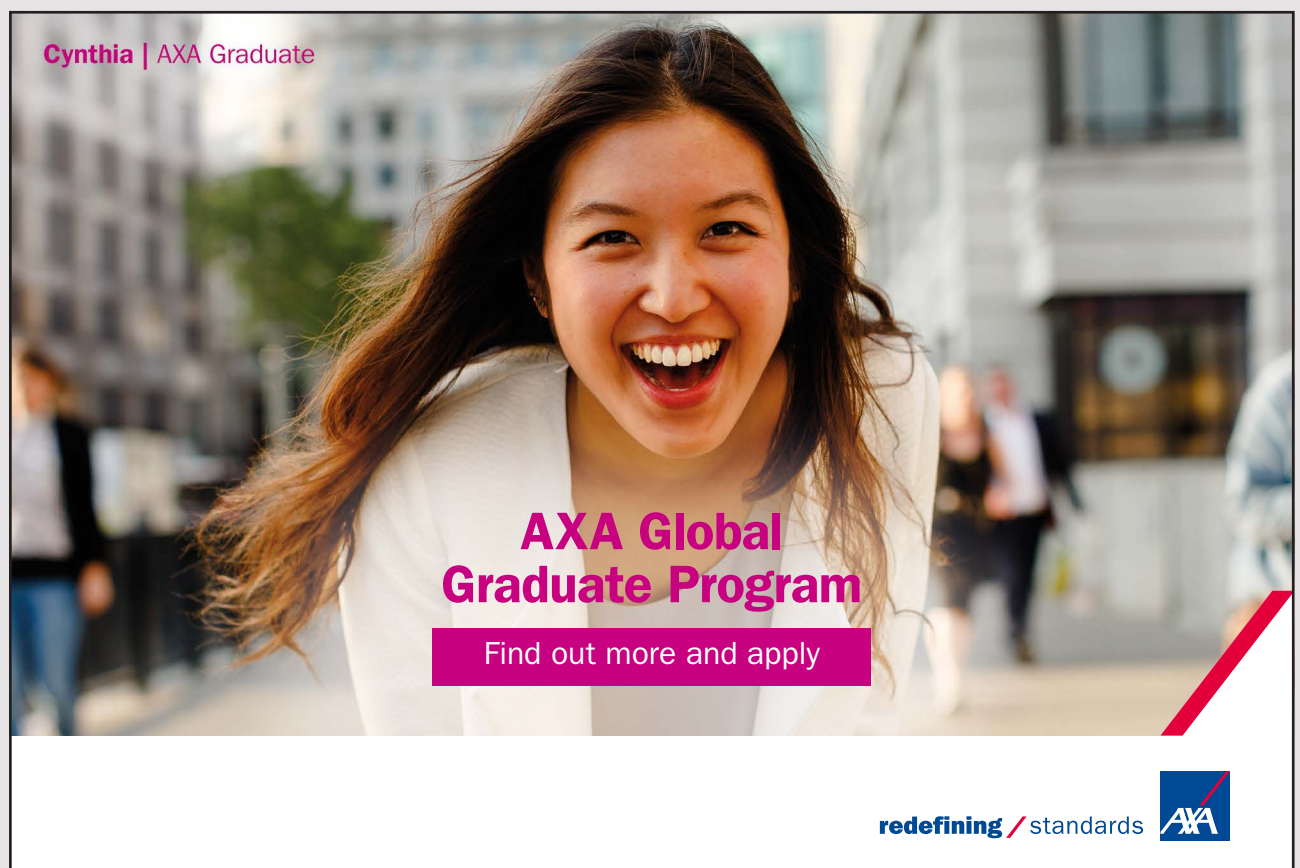
Let us suppose a questionnaire has ten questions. How can the information obtained be helpful in sorting out the hypotheses?

Certain conditions to be fulfilled.

- All questions must flow in sequence.
- Please assess hypotheses must be in questions.
- Please bundle the questions like one to three or hypothesis one.
- The main hypothesis should not be found in the questionnaire, or the purpose fails.
- There is no need to have all questions to answer the hypotheses.
- Some questions can be general, like age or marital status.
- Certain open-ended questions go more towards a conclusion.
- Certain findings can be supported by open-ended questions.
- Sometimes a panel of expert opinion can support or enhance the hypothesis findings.

Questionnaire

A questionnaire of some 13 questions is below to help the researcher look for hypotheses. Please develop around 20 questions. Over 20 questions might make the research vague, misleading, and hard to conceptualise.

Model:

The effect of reward on employee performance.

Key hypothesis:
- There is a relationship between effective reward and performance management in Company X.

Sub-hypotheses:
- Rewards must exist or must be created to impact performance.
- The attractiveness of rewards must be a determinant for good performance at work.
- Rewards have to be effective to influence performance.
- Reward must have an influence on performance.

Questionnaire design
- Does your company provide you with rewards?
- What type of rewards exist in your company?
- Classify the rewards in terms of intrinsic/personal or extrinsic/tangible.
- Are the rewards attractive?
- Which rewards are more likely to attract an employee in your firm?
- Are employees attracted by rewards? If so, how?
- How often employees are rewarded at work?
- What is the reaction of employees towards reward?
- Can you evidence better performance:
    Prior to reward?
    After the reward?
- Does your supervisor support you in the provision of reward?
- Is there any negative action or punishment for not attaining desired performance?
- How do you assess an effective reward?
- Is this possible in your organisation? If yes, how?
- What type of rewards help you better perform at work?
- Can you perform effectively without rewards?
- Does your company need to develop a reward culture for improving performance?
- If so, how?

Bundling the hypotheses

Although questions flow in some logical order, there is no need to say all the sub-hypotheses will be found there. They can be either in an orderly or less orderly manner but the researcher must explain where the hypotheses lie.

Recap of the questionnaire

Hypothesis H_1:
Rewards must exist or must be created to impact performance.

Rewording:
H_1: There is a need to create a reward mechanism in Company X.
Null Hypothesis: There is no apparent need to create a reward mechanism in Company X.

Bundle questions:

- *Does your company provide you with rewards?*
- *What type of rewards exist in your company?*
- *Classify the rewards in terms of intrinsic (personal) or extrinsic (tangible).*

Hypothesis H_2:
The attractiveness of rewards must be a determinant for good performance at work.

Rewording:
H_2: Rewards should be attractive to influence performance in company X.
Null Hypothesis: There is no reason for rewards to be attractive or not to influence performance in Company X.

Bundle questions:

- *Are the rewards attractive?*
- *Which rewards are more likely to attract an employee in your firm?*
- *Are employees attracted by rewards? If so, how?*

Hypothesis H_3:
Rewards have to be effective to influence performance.

Rewording:
H_3: There is a relation between effective reward and performance.

Null hypothesis: There is no relation between effective reward and performance.

Bundle questions:

- *How often employees are rewarded at work?*
- *What is employees' reaction towards reward?*
- *Can you evidence better performance:*
  - *Prior to reward?*
  - *After the reward?*
- *Does your supervisor support you in the provision of reward?*
- *Is there any negative action or punishment for not attaining desired performance?*
- *How do you assess an effective reward?*
- *Is this possible in your organisation? If yes, how?*

Hypothesis H_4:
Reward must influence performance.

Rewording:
H_4: Rewards, extrinsic or intrinsic, will have an influence on performance. Such performance might be improved.

Null hypothesis: Whether rewards are intrinsic or extrinsic, there will be no significant improvement Company X's performance.

Bundle questions:

- *What type of rewards help you better perform at work?*
- *Can you perform effectively without rewards?*
- *Does your company need to develop a reward culture for improving performance?*
- *If so, how?*

Main hypothesis: There is a relation between effective reward and performance management in Company X.

The main hypothesis will be based on:

H_1: There is a need to create a reward mechanism in Company X.

H_2: Rewards should be attractive to influence performance in company X.

H_3: There is a relation between effective reward and performance.

H_4: Rewards, extrinsic, or intrinsic, will have an influence on performance. Such performance might be improved.

See how all the sub-hypotheses converge to give reason to the main hypothesis developed. The results be congruent.

Qualitative Interpretation

This chapter will not deal too much with qualitative information although this is most welcome researchers develop the acumen for qualitative research. Questions are formatted in the same way as for quantitative questions. One would invite students dealing with psychology and related issues to do a qualitative survey.

Surveys in the qualitative area are analysed in terms of tendency. Like:

'A majority reported…'
'Most responses were in favour of…'
'In general, we saw….'
'We could synthesise….'

Along with the different styles to state an answer, percentages work. Say '67% of respondents stated that…'. Please note charts and diagrams with percentages and values count as qualitative research if the interpolation is basic, simple, and non-extraneous.

Quantitative Interpretation

The focus of this toolkit chapter is how to create quantitative data out of the questions. From a quick view, there is no chance to create quantitative questions like:

'What percentage of employees work in your department?'
'What is the average score of employees regarding their performance?'
'In what percentage terms could you suggest the likeliness of the success of a training programme in HRM?'

These questions, unfortunately, do not form part of a questionnaire nor are they likely to develop suitable answers and predictive values. These could remain as the responsibility of a forecaster, project manager, and even other managers who undertake individual projection. Employees as sample members cannot give such answers.

In this way, we can see how certain, not all, questions can develop quantitative answers. Likert scales are suitable examples in the creation of quantitative questions without much engaging the respondent or the interviewee.

Likert Scale Interpretation

Likert scales are very often used by students in research. Anybody can use between 4 or 5 scales ranging from one extreme to the other one. These could range from strongly disagree to strongly agree. Whether there are four or five scales, respondents provide answers.

Example

Managers' relations with employees

Out of a sample of 50 employees, let us see responses:

		Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
1.	The manager entertains sound relations with employees.	2	8	5	23	12
2.	Communication is a two-way process.	6	10	6	18	10
3.	Management aims to provide support to employees.	5	10	6	19	10
4.	Managers provides feedback to employees.	8	7	4	18	13

Points rating

Please provide a points rating. Like 1 for strongly disagree, 2 for disagree, 3 for neutral, 4 for agree, and 5 for strongly agree. Some others might prefer 0.2 for strongly disagree, 0.4 for disagree up to 1 for strongly agree.

Now with a scale between 1 and 5, the table reads a s follows:

Revised table

		Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
1.	The manager entertains sound relations with employees.	2	16	15	92	60
2.	Communication is a two-way process.	6	20	18	72	50
3.	Management aims to provide support to employees.	5	20	18	76	50
4.	Managers provides feedback to employees.	8	14	12	72	52

Next I compute mean values:

For item 1, it reads as:
2+16+15+92+60 divided by 50 =3.7, variance 1.25, Standard deviation (SD) =1.12 to 2 d.p.

The table is now as follows:

		Mean	Standard Deviation	Variance
1.	The manager entertains sound relations with employees.	3.7	1.12	1.25
2.	Communication is a two-way process.	3.32	1.32	1.74
3.	Management aims to provide support to employees.	3.38	1.28	1.63
4.	Managers provides feedback to employees.	3.42	1.41	2

Observations

All values exceed the median value 3 meaning that relations are fairly good. It scores higher in terms of manager and employee relations, and broadly similar regarding communication, support and feedback. Generally, employee relations look fair in this context.

To support a hypothesis, one could say there is good relation between management and employees. A compounded mean value can also be made by adding all mean values and divided them by the 4 items under the heading. This makes 3.7+3.32+3.38+3.42 divided by 4 =3.45. Note the scores are neither weak nor exceedingly high meaning there is room for improvement in the sample.

Standard deviation or the squared value (Variance) do help like deviations ranging 1.1 to 1.4 from the mean value. This dispersion is fairly low and reasonable meaning findings are consistent here.

Calculating a Z-value

Another exercise could be the calculation of a Z-value and check how consistent are the values between the population and the sample. This could be done by students making statistical interpretations of populations and mean.

Let us say for instance the mean scores for students in an HR examination is 62. This refers to a population.

A sample of 30 students is a requirement for the Z-test. The sample mean is 67. Data is input in a scientific calculator and reads as:

Population Mean (μ):	62
Population Variance (σ^2):	4
Sample Mean (M):	67
Sample Size (n):	30

Z Score Calculations

$$Z = (M - \mu) / \sqrt{(\sigma^2 / n)}$$
$$Z = (67 - 62) / \sqrt{(4 / 30)} \quad Z = 5 / 0.36515$$
$$Z = 13.69306$$

Assuming the test is two-tailed and the significance level is 0.05:

The Z-score is 13.69306. The p value is 0. The result is significant at p <0.05.

This means the sample performance matters. The students performed well in the HR test.

Z-score for 2 population proportions

General assumption: The size should be above 30.
Proportions should be close and not too dispersed.

Example

Proportion or total number of individuals from sample Population 1 who have the characteristic in question, for example, computer literate young employees between 18 and 35 in the HR department: 24 out of 30.

Total number of people in sample Population 1: 30

Proportion or total number of people from sample Population 2 who have the characteristic in question, for example, computer literate older employees who are above 36 in the HR department: 14 out of 25.

Total number of individuals in sample Population 2: 25

The significance level is 5% and as a convention a two-tailed test is done.

Result

The Z-Score is 1.9178.
The p-value is 0.05486.
The result is *not* significant at p <0.05.
The proportion of Yes or No responses for Observation 1 is 0.8.
The proportion for Observation 2 is 0.56.

Observation

Younger employees are more computer literate than older ones as this could be a trend with a p value close but no less than 0.05.

T-test calculation

For samples less than 30, a t-test applies. A t-test compares means of two groups.

One sample t-test

A one sample t-test compares the mean with a hypothetical value. In most cases, the hypothetical value comes from theory. The hypothetical value can also come from previous data. For example, compare if the mean hours of work in a firm differs from 36.5, a value determined in a study.

Working hours of 10 employees (maximum 40 hours)
38, 39, 37, 34, 32, 31, 34, 38, 37, 36.
Mean working hours: 36.5
Sample mean: 35.6 Standard deviation: 2.72 Standard Errors of mean: 0.86. Number: 10

P value and statistical significance:

The two-tailed P value equals 0.3221

By conventional criteria, this difference is considered to be not statistically significant.

Confidence interval:

The hypothetical mean is 36.50

The actual mean is 35.60

The difference between these two values is -0.90

The 95% confidence interval of this difference: From -2.84 to 1.04

Intermediate values used in calculations:

t = 1.0478

Degrees of freedom = 9

Standard error of difference = 0.859

Paired t-test

For instance, compare if performance differs between a control and test group, between men and women, or any other two groups. Maximum output is 20 units.

Once the information is input on an online t-test calculator, results are:

Men (Output)	Women (Output)
20, 18,18, 15, 16, 16, 15, 19, 20, 20	12, 13, 20, 15, 18, 19, 14, 15, 15, 18
Mean: 17.7	Mean: 15.9
SD: 2.06	SD: 2.69
Standard Error of Means: 0.65	Standard Error of Means: 0.85
N: 10	N: 10

Since this is a paired t-test,

P value and statistical significance:

The two-tailed p-value equals 0.1529.

By conventional criteria, this difference is considered to be not statistically significant.

Confidence interval: The mean of Male less Female equals 1.80.

95% confidence interval of this difference: From -0.81 to 4.41.

Intermediate values used in calculations: t = 1.5615

Degrees of freedom f = 9

Standard error of difference = 1.153.

Chi-Square test

This test evaluates differences student choice sat for the Research methods test and stated preferred tests:

	Qualitative test	Quantitative test	Marginal Row Totals
Group 1 students	70	30	100
Group 2 students	40	28	68
Marginal Column Totals	110	58	168 (Grand Total)

The amended table shows the sum of squares and differences attributed per cell after inputting the data in a scientific chi-square calculator.

	Qualitative test	Quantitative test	Marginal Row Totals
Group 1 students	70 (65.48) [0.31]	30 (34.52) [0.59]	100
Group 2 students	40 (44.52) [0.46]	28 (23.48) [0.87]	68
Marginal Column Totals	110	58	168 (Grand Total)

The chi-square statistic is 2.2367.
The *p*-value is .134769.
This result is *not* significant at $p < .05$.

Observation
There is no major difference between Group one and two regarding preferences for either tests. There is a common preference for qualitative tests.

Null Hypothesis
The null hypothesis asserts the independence of variables under consideration. For instance, Group and test choice are independent of each other.

Multiple Chi-Square Test (5 rows × 3 columns)

There is a model below on the choice of pensions for three categories of employees. The information is in a 5 × 3 chi-square table to find differences or dependences among variables.

Using the table, I observe:

Younger employees like more pension incentive and triple bonus.
Middle-aged employees are keener to consider state and company pension.
Older employees are likelier to stick to bank deposits, state pensions, and company pensions.

	State Pension	Company pension	Tailor-made pension	Bank deposit	Pension incentive with triple bonus	Row Totals
Employees 18–30	15	20	25	10	30	100
Employees 31–45	20	20	20	18	24	102
Employees 45 and above	25	25	10	20	15	95
Column Totals	60	65	55	48	69	297 (Grand Total)

Please input data on a chi-square calculator
The significance level is at 5%.

The worked out table:

	State Pension	Company pension	Tailor-made pension	Bank deposit	Pension incentive with triple bonus	Row Totals
Employees 18–30	15 (20.20) [1.34]	20 (21.89) [0.16]	25 (18.52) [2.27]	10 (16.16) [2.35]	30 (23.23) [1.97]	100
Employees 31–45	20 (20.61) [0.02]	20 (22.32) [0.24]	20 (18.89) [0.07]	18 (16.48) [0.14]	24 (23.70) [0.00]	102
Employees 45 and above	25 (19.19) [1.76]	25 (20.79) [0.85]	10 (17.59) [3.28]	20 (15.35) [1.41]	15 (22.07) [2.27]	95

The chi-square statistic is 18.117.
The p-value is .020366.
The result is significant at $p < .05$.

Observation

The hypothesis could mean there are differences in pension schemes among employee categories and these matter. The null hypothesis proving no dependence between pension schemes and employee category is rejected here.

Correlation test

Correlation tests should have values that progress. Example, a continuous observation of salary rise (indexed values) from year 1 to year 10.

The values are input below:

Years at work	Indexed wage	Residuals (Input by calculator itself)
1	100	2.54
2	110	4.09
3	110	-4.36
4	120	-2.81
5	130	-1.27
6	135	-4.72
7	150	1.82
8	160	3.36
9	170	4.9
10	170	-3.54

Data Summary

$\sum X = 55$	$\sum X^2 = 385$
$\sum Y = 1355$	$\sum Y^2 = 189625$
$\sum XY = 8150$	

R = 0.99	R2 = 0.979	Slope = 8.455	Y intercept = 89	Standard Error of Estimate = 3.96
t-value: 19.39	Degrees of freedom = 10-2 =8	p<.0001		

The equation reads as: Y = 8.455X + 89.00

The R^2 value is 0.97.

There is high strong correlation between wages and years of service in the said company.

Regression calculation

A researcher wants to find out a regression possibility using employee output units as variable Y which are dependent on age X_1, years of experience X_2, training X_3 and adaptability X_4.

The equation without calculation is written as

$$Y = ß_1X_1 + ß_2X_2 + ß_3X_2 + ß_4X_4 + e$$

Values could be added to the equation as years of activity and other independent variables progress. They are as follows:

Age X_1	Experience X_2	Training X_3	Adaptability X_4	Y	Predicted Y values
10	10	20	30	100	106.9
12	12	25	35	120	123.5
14	14	30	40	150	140.1
15	17	40	45	160	158.8
16	19	50	45	170	170.8
18	19	55	50	180	178.1
20	21	60	55	200	194.7
22	23	70	55	210	209.4
24	25	70	65	220	227.9
26	26	75	70	240	239.9

Note that all the values are indexed.
The computer generates predicted Y values.

Part of the residual analysis is summarised:

R^2 = 0.98 to 2 d.p.	F value = 85.46
Mean = -0.022	Variance = 36.34
Normality condition: Evidence against normality	

The regression equation is:

$Y = 21.84 + 2.637X_1 + 4.648X_2 + 0.009358X_3 + 0.3994X_4$

or in a simpler way:

$Y = 21.84 + 2.6X_1 + 4.6X_2 + 0.01X_3 + 0.4X_4$

implying the importance given to age and years of experience in this equation.

One-Way ANOVA

Students' performance in three tests are assessed. They are all evaluated on 20 marks. The three tests:

- Oral language skills.
- Writing competence.
- Mathematics.

Five students' scores:

Oral /20	Written /20	Mathematics/20
17	18	18
16	16	16
15	15	13
16	14	17
17	13	18

The results are in a statistical ANOVA calculator to check results. They are displayed here:

Treatments				
N	1	2	3	Total
	5	5	5	15
$\sum X$	81	76	82	239
Mean	16.2	15.2	16.4	15.9333
$\sum X^2$	1315	1170	1362	3847
Std.Dev.	0.8367	1.9235	2.0736	1.6676

The table is further analysed:

Result details				
Source	SS	df	MS	
Between-treatments	4.1333	2	2.0667	F = 0.71264
Within-treatments	34.8	12	2.9	
Total	38.9333	14		

The f-ratio value is 0.71264.
The p-value is .509972.
The result is *not* significant at $p < .05$.

Observation

Student performance is generally similar and do not have a significant relation between independent and dependent variables. The independent variables are the different examinations proposed and these influence the final result of the exams. Performances were not critical in Mathematics, for instance, compared with other courses.

Two-way ANOVA calculation

In a two-way ANOVA, two variables might be input like Young and Elder workers along with independent variables like relations with peers, subordinates, and superiors, a bit in line with Herzberg's hygiene factors.

To make ANOVA useful, set similar scales like an index of 1–10 or in the previous example, an examination score limit of 20. These could be obtained from scores input by respondents in the questionnaire. Even mean values could be input but, for simplicity, each pair comprises the response of five employees.

	Relationship with peers/10	Relationship with superiors/10	Relationship with subordinates/10
Employees (18–30)	10 9 8 9 8	6 7 6 6 7	9 9 8 8 9
Employees (31–50)	9 8 7 7 6	6 6 6 8 6	9 6 8 9 9

		Relationship with peers	Relationship with superiors	Relationship with subordinates	Total
Employees (18–30)	N	5	5	5	15
	ΣX	44	32	43	119
	Mean	8.8	6.4	8.6	7.9333
	ΣX^2	390	206	371	967
	V σ^2	0.7	0.3	0.3	1.64
	SD σ	0.84	0.55	0.55	1.28
	SE	0.37	0.24	0.24	0.33
Employees (31–50)	N	5	5	5	15
	ΣX	37	32	41	110
	Mean	7.4	6.4	8.2	7.3333
	ΣX^2	279	208	343	830
	V σ^2	1.3	0.8	1.7	1.67
	SD σ	1.14	0.89	1.3	1.29
	SE	0.51	0.4	0.58	0.33
Total	N	10	10	10	30
	ΣX	81	64	84	229
	Mean	8.1	6.4	8.4	7.6333
	ΣX^2	669	414	714	1797
	V σ^2	1.43	0.49	0.93	1.69
	SD σ	1.2	0.7	0.97	1.3
	SE	0.38	0.22	0.31	0.24

Where (V) σ^2 represents variance and σ represents the standard deviation.

The ANOVA table is summarised below:

Source	SS	Degrees of freedom	MS	F	P
Rows	2.7	1	2.7	3.18	0.0872
Columns	23.27	2	11.64	13.69	0.0001
rows × columns	2.6	2	1.3	1.53	0.2369
Error	20.4	24	0.85		
Total	48.97	29			

Since the p value is 0.23, the result is not significant.

The Tukey test will state whether the values were significant or not.

Tukey test values	HSD [0.05]	HSD [0.01]
Rows [2]	0.69	0.94
Columns [3]	1.03	1.33
Cells [6]	1.8	2.22

HSD = the absolute [unsigned] difference between any two means (row means, column means, or cell means) required for significance at the designated level:

HSD[.05] for the .05 level,

HSD[.01] for the .01 level.

The HSD test between row means can be meaningfully performed only if the row effect is significant; between column means, only if the column effect is significant; and between cell means, only if the interaction effect is significant.

Observation

There is insignificant employee dependence with categories 918-30) and 30+ aged employees. Scores differed and were insignificant. Both categories of employees have insignificant dependence or influence on the three types of relations. But there are plausible reason relations with superiors were poorer in each case.

$$[F (2, 29) = 1.53, p = 0.24]$$

Additional notes

Once researchers master quantitative concepts and can use them with confidence, please follow these precautionary efforts:

- Write clear, direct hypotheses. Like a significant relation between X and Y.
- Spell out the dependent and the independent variable.
- Develop a null hypothesis.
- State straightforward, cogent hypothesis. Like the relation between motivation and performance, performance and reward, reward and motivation, reward and performance.
- Develop sub hypotheses which align with the main hypothesis.
- Use a few tests depending upon relevance.
- Never include all tests to impress people like t-test and chi-square test or ANOVA.
- State what you wish to find.
- Synthesise information.
- Keep ideas flowing.
- Check consistency of figures like t, p, or Z values.
- Make the research interesting.
- Show how much you enjoyed researching.

Best of luck!

ABOUT THE AUTHOR

Dr Nirmal Kumar Betchoo is Dean of Faculty of Business and Management at Mauritius' Université des Mascareignes. He has been in education since 1986 and taught at the tertiary level for 22 years. He specialises in the social sciences, human resource management, business, and strategy. He has so far published ten books, written over 45 peer-reviewed international research articles, and took part in over 15 local and international conferences. He, moreover, writes for the national press on culture and economy since 2002 with over 125 papers published including occasional exclusive reports for leading Mauritian papers 'Le Mauricien', 'L'Express' and 'Business Magazine'.

Dr Betchoo is an associate editor of the European Scientific Journal and the external editor for the United States' Journal of Mass Communication and Journalism and Indian Novelty Journal. He reviewed over 35 research papers including the Inderscience Journal. He is panel specialist for Mauritius' Tertiary Education Commission on accrediting undergraduate and post-graduate business programmes. He researched with each of Mauritius' public institutions and universities. He moderated for Mauritius' University of Technology (UTM), set papers for promotional examinations at the Public Service Commission with the Mauritius Examinations Syndicate, and moderated papers with the Mauritius Institute of Training and Development since 2011. Dr Betchoo supervises PhD students at the Université des Mascareignes and coaches post-graduate students vocationally. – Cyrus. K. Jones