DOCUMENT RESUME

ED 461 473 SE 059 480

AUTHOR	Hayden, Bob; Roberts, Bill
TITLE	Using Birthday Data To Integrate Statistics into the K-12 Mathematics Curriculum.
PUB DATE	1995-00-00
NOTE	5p.; Published three times per year by the American Statistical Association and National Council of Teachers of Mathematics Joint Committee on the Curriculum in Statistics and Probability.
AVAILABLE FROM	American Statistical Association, 1429 Duke Street, Alexandria, VA 22314-3402. Tel: 703-684-1221.
PUB TYPE	Guides - Non-Classroom (055) -- Journal Articles (080)
JOURNAL CIT	Statistics Teacher Network; n38 Win 1995
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Elementary Secondary Education; *Learning Activities; Mathematical Applications; Mathematics Activities; *Mathematics Instruction; Relevance (Education); *Statistics; Student Interests; Student Motivation; Thematic Approach
IDENTIFIERS	*Birthdays

ABSTRACT

In response to the recommendation that teachers incorporate more statistics into their mathematics classes, this paper offers several ways in which students can work with real data. Among the ideas included are their birthdays. Having the students generate a data set that includes their month of birth, exact time of birth, day of the month, and cumulative day of the year--something which concerns themselves--encourages participation and interest. In the first grade, when students are exposed to number lines, the children can make dotplots of the day of the month. By the end of first grade children are doing single-digit subtractions. Current lessons on plotting numbers on a number line, comparing sizes of numbers, and simple subtraction can be combined into a lesson where those same skills are practiced in the course of describing some data. The above exercise can be repeated at higher grade levels as other kinds of numbers, such as decimals or negative numbers, are introduced. Other activities described include stem-and-leaf plots, calculating a mean, residuals, order statistics, categorical data, and relations between variables. Suggestions for broadening the database and making a bridge to calculus are also included. (PVD)

USING BIRTHDAY DATA TO INTEGRATE STATISTICS INTO THE K-12 MATHEMATICS CURRICULUM

by

Bob Hayden

and

Bill Roberts

Using Birthday Data to Integrate Statistics into the K-12 Mathematics Curriculum

There have been more recommendations for changes in the teaching of mathematics in the past few years than at any time since the "new math" era of the 1960s. One of those recommendations is that teachers incorporate more statistics into their classes. We suggest that one way to fit in new topics is to integrate them with old topics, so that a single lesson may serve a double purpose. We will illustrate this with examples. Since one of the strongest recommendations is to have students work with real data, we start by discussing one data set that your students can gather and use.

Young children are usually very interested in birthdays—especially their own! Even college students take pride in being able to report the exact time of their birth—and a surprising number are able to do so. The idea of having the students generate a data set that concerns themselves encourages participation and interest. Thus, you might collect birthday data from your class. The birth year will not prove interesting, as the range will be small for most K-12 classes. A class data set that includes month of birth, exact time of birth, day of the month, and cumulative day of the year will be more interesting. The cumulative day of the year may be obtained by counting, by using a calendar that includes that information, or by entering the dates into a spreadsheet. (Entering a constant year value for each birthday into a spreadsheet, you can use the Julian Date function to obtain the numerical day of year values.) If 2/29 is one of the birthdays, you could decide to designate the cumulative day of the year for this entry as zero (0), so that the number representing March first would be the same whether or not you are looking at a leap year. The remaining birthdays could be represented using values of one through three-hundred sixty-five.

Even in the first grade, textbooks use a number line to illustrate the addition, subtraction, or comparison of two numbers. The textbooks we consulted did not have the students doing much of this themselves, but we think it is important for children to *do* as well as see. If we have them put more than two numbers on a single number line, we are on our way to a dotplot. Having students make this display gives them practice in using a number line, something we wanted them to know about anyway, before we started trying to teach them statistics. For the birthday data, we could make dotplots of the day of the month.

Once we have the numbers on a line, we have also sorted the numbers. Dealing with the relative sizes of numbers helps students to grasp the meanings of numbers in a way that computation does not. There is a whole branch of statistics called order statistics that is based on ordering or sorting numbers more than on calculating with them. For example, the maximum and minimum values of a set of numbers would be examples of order statistics. So would the median (middle number in the sorted data), and other things such as quartiles, percentiles, and deciles.

By the end of first grade, children are doing single-digit subtractions. The difference between the maximum and minimum values in the data is called the *range*. It gives us an idea of how spread out the numbers are. Our suggestion is to combine current lessons on plotting numbers on a number line, comparing sizes of numbers, and simple subtraction, into a lesson where those same skills are practiced in the course of describing some data. For the birthday data, students should be able to estimate the range of most of the variables. If the months are treated as integers between 1 and 12, the range will usually be 11. For the day of the year, we can estimate the range, but not as accurately.

The above exercise can be repeated at higher grade levels as other kinds of numbers are introduced. Even some college students think that $0.2 < 0.05$ or $-2 < -5$, and are quite unsure as to how 4/7 compares to 11/19. They may also have difficulty finding 3.426 on a number line, so the extra practice may be worthwhile. The time of birth is the one variable in the birthday data that is not an integer. If times are quoted in hours and minutes, we can consider them as fractions with a denominator of 60. We can change these to decimal fractions of an hour. We can also look at this as a units change. In every case, we get rational numbers.

In addition to using dotplots, we found that stem-and-leaf plots usually prove quite interesting, not only for studying the pictorial characteristics of the data, but also the concepts of place value and digit truncation. For a stem-and-leaf of the day of the month, we pick stem units of 0, 1, 2 or 3 (representing the tens place) and leaf units of 0 through 9 (representing the units digits). The plot, if done by hand, will have the units values listed to the right of the appropriate tens value. Stem-and-leaf plots can be generated by many of the common statistical software packages as well.

When it comes to statistics that you calculate, the mean is probably the best known. Virtually all the college students we see can calculate a mean, so we figure it must be being taught, and taught well. Children can begin doing this as soon as they have the component skills, which are addition of a column of figures and division. We would like to see the mean not only calculated, but placed on a number line with the data. This helps to give insight into what a mean is, and also provides an error check. For the birthday data, we should also be able to estimate what the mean will be, since the variables all have an approximately uniform distribution. This provides practice in the art of estimating, but students should also see some examples where there is an element of surprise. There will be a bit of surprise in the birthday data for many students; unless you have a very large class, you are likely to observe significant departures from expectation in statistics like the mean or even the range. These can lead to a discussion of the idea of sampling error.

Once the mean is placed on the dotplot, it becomes possible to ask questions like "How many points below average was the 3?" or "Which observation was farthest from the mean?" This reinforces subtraction ideas and leads to the concept of a *residual*. The difference between an actually observed data value and a summary of the data (such as the mean) is called a residual. This can be used as an introduction to negative numbers, since no prior knowledge of these is required to put a "-" in front of the residuals for points that are below average. If the students have prior exposure to operations on negative numbers, they can add up the residuals and find that they sum to zero. Encourage them to check this out for other data sets and other summaries to see if it always happens. One way to generate negative numbers from the birthday data is to have the class (or each student) pick a "hero" or famous figure whose birthday they would like to find out. (We use "hero" here to refer to an admired person of either gender.) Let the students research the birthday, and then define for each child a variable that is the day of the year they were born minus the day of the year their hero was born. Can you estimate what the mean of this variable might be? Another calculated variable of interest might be the number of days until ones birthday. This could be a non-negative variable, or we could define it as number of days to your **nearest** birthday, with negative numbers indicating observations where the last birthday is closer to today's date than the next is.

If your students know how to multiply negative numbers, then they can square the residuals and add up the results. Dividing the sum by one less than the number of observations gives the variance. The square root of this is the standard deviation. The interpretation is that the standard deviation is *a typical value for the residuals*. Thus the standard deviation measures how variable or "spread out" the data are. We have some reservations about teaching young children about standard deviations because the computation is lengthy and not very intuitive. However, we expect that the topic is likely to be taught simply because lots of teachers already know about standard deviations. We think that some of the more visual topics such as dotplots are probably a better choice for children. If standard deviations **are** taught, we strongly prefer that the children use the method shown above that includes the residuals. They should learn to interpret the residuals in terms of a dotplot, and learn to interpret the standard deviation in terms of the residuals. In particular, we advocate abstinence from the various "computational" formulae that are floating about, as these tend to hide the residuals as well as the meaning of the process.

There are some more intuitive measures of variability based on order statistics. We have already mentioned the maximum and minimum as order statistics, and their difference, the range, is another measure of variability. Another common order statistic is the median. This is based on sorting the data and then selecting the middle value. We can take this one step further and find the first and third quartiles, which are just the medians of the upper and lower halves of the data. The difference between the third and first quartiles is called the *interquartile range* (IQR). It is yet another measure of variability. You should be aware that there are many inconsistent definitions of quartiles in use. Pick one you like and stick to it.

It is important to note that order statistics typically involve little or no computation. They are based rather on sorting the data and counting to find mileposts along our trip through the (sorted) data. The mathematics topics that they reinforce are ideas of order and magnitude rather than computational skills. Lest they seem to be of interest only for the early elementary grades, consider making a dotplot and finding order statistics for the following sets of data: 3, $\sqrt{10}$, π, 2.718, 3.14141414..., 22/7, 10/$\sqrt{10}$, 10/π or 0.005, 0.011, 0.01, 0.1, 1/20, 1/19, 1/9, 1/11.

So far we have considered data that are primarily numerical. Since month is a categorical variable, one could report the total number of observations and the percentage of the total represented by each category (month). A study of these percentages is usually interesting. A bar chart is the typical choice for displaying categorical data, but a dotplot differs only in artistic detail. Note that the distribution of the months may not be very uniform for small data sets. Further work for younger students might include exercises dealing with the calendar such as the order of the month names, or the length (in days) of particular months.

So far we have considered one variable at a time. We can also look at relations between variables by plotting one against the other. You might try day of month versus day of year, and day of month versus month, both of which generate interesting patterns that may appear to be identical or to be no pattern at all until you get an appropriate scale. If your class were to choose a single hero to compare their birthdays to, these data can be plotted against day of year.

In precalculus, we often compare the graph of f(x) with the graph of f(x+a), af(x), or f(ax). We can do the same with data. We can shift the day of year data by taking something other than 1 January as our starting point—say the first day of school or the first day of summer vacation. We can multiply I(y) or I(x) by a constant by changing units—we could use time of birth in hours or minutes or we could express day of year as decimal fractions of a 365-day year. At least one of the patterns should be found to be approximately periodic. It might make a good example for a trigonometry class, especially as it looks quite different from any trig function.

For all the things we have suggested, the apparent pattern may change as we increase the number of observations. You can increase your database by pooling birthdays from past classes or by pooling the data from several different current classes. You can also open it up to family members, Presidents of the United States, or all the teams in the NFL—whatever appeals to your class. To make a bridge to calculus and college, you can ask students to discuss how some of the graphs might look if you let the sample size increase without bound.

Bob Hayden and Bill Roberts
Plymouth State College
Plymouth, New Hampshire

059480

ERIC

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Using Birthday Data to Integrate Statistics into The K-12 Mathematics Curriculum

Author(s): Bob Hayden and Bill Roberts

Corporate Source: Plymouth State College, Plymouth New Hampshire

Publication Date: STN 38 Winter 95 issue

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following two options and sign at the bottom of the page.

[✓] Check here **For Level 1 Release:** Permitting reproduction in microfiche (4" x 6" film) or other ERIC archival media (e.g., electronic or optical) ***and*** paper copy.

The sample sticker shown below will be affixed to all **Level 1** documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Level 1

The sample sticker shown below will be affixed to all **Level 2** documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Level 2

[ ] Check here **For Level 2 Release:** Permitting reproduction in microfiche (4" x 6" film) or other ERIC archival media (e.g., electronic or optical), but *not* in paper copy.

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level **1**.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Sign here→ please

Signature: Jerry Moreno, STN Editor / Cathy Crocker	Printed Name/Position/Title: JERRY MORENO / CATHY CROCKER
Organization/Address: Dept. Mathematics, John Carroll University, University Hts OH 44118 / ASA, 1429 DUKE ST, ALEXANDRIA, VA 22317	Telephone: 2163974681 / 703-684-1221
	FAX: 2163973033 / 703-684-2037
	E-Mail Address: moreno@jcvaxa.jcu.edu / cathyc@amstat.org
	Date: 2/5/97 / 1/13/97

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

ERIC/CSMEE
1929 Kenny Road
Columbus, OH 43210-1080

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2d Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com