



---

Theses and Dissertations

---


2005-11-28

## Ipsative Score Distortion on Affinity 2.0

Alec J. Brown

*Brigham Young University - Provo*

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>

 Part of the [Counseling Psychology Commons](#), and the [Special Education and Teaching Commons](#)

---

### BYU ScholarsArchive Citation

Brown, Alec J., "Ipsative Score Distortion on Affinity 2.0" (2005). *Theses and Dissertations*. 706.  
<https://scholarsarchive.byu.edu/etd/706>

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact [scholarsarchive@byu.edu](mailto:scholarsarchive@byu.edu), [ellen\\_amatangelo@byu.edu](mailto:ellen_amatangelo@byu.edu).

IPSATIVE SCORE DISTORTION ON AFFINITY 2.0

by

Alec J. Brown

A thesis submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Educational Specialist

Department of Counseling Psychology and Special Education

Brigham Young University

December, 2005

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a dissertation submitted by

Alec J. Brown

This dissertation has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

\_\_\_\_\_  
Date

\_\_\_\_\_  
Lane Fischer, Chair

\_\_\_\_\_  
Date

\_\_\_\_\_  
Melissa A. Heath

\_\_\_\_\_  
Date

\_\_\_\_\_  
Ellie Young

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the dissertation of Alec J. Brown in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

---

Date

---

Lane Fischer  
Chair, Graduate Committee

Accepted for the Department

---

Melissa A. Heath  
Area Coordinator  
Counseling Psychology & Special Education

Accepted for the College

---

Barbara Culatta  
David O. McKay School of Education

## ABSTRACT

### IPSATIVE SCORE DISTORTION ON AFFINITY 2.0

Alec J. Brown

Department of Counseling Psychology and Special Education

Educational Specialist

This study investigated distortion that occurs when raw scores are converted to ipsative scores on *Affinity 2.0*, a relatively new instrument for assessing sexual interest. Using a sample of 146 non-offending, heterosexual females, this study examined the characteristics of distorted ipsative score profiles and attempted to develop an algorithm to identify such distortions. A method was developed for defining distortion objectively. Of the 146 profiles, 125 were found to contain some degree of distortion. Several hypotheses were formulated as to variables that might be related to distortion. These relationships were examined using Pearson Product Moment Correlations. Several statistically significant, but weak, correlations were found. An interaction effect was calculated for four of these variables, and was found to have a moderately strong correlation with distortion ( $r = .530, p < .01$ ). An algorithm for identifying distortion was developed using this interaction effect. Several cut-off scores were tested. The most effective cut-off only correctly identified 42.9% of the significantly distorted profiles. Implications and limitations of the results are discussed, and directions for future research are provided.

## ACKNOWLEDGEMENTS

I would like to express gratitude to Dr. Lane Fischer for acting as committee chair, assisting me with my questions, and guiding me through the thesis process. I would also like to thank Dr. Melissa A. Heath and Dr. Ellie Young for providing important feedback and suggesting revisions to the transcript.

## TABLE OF CONTENTS

Abstract .....	iv
Acknowledgements .....	v
Table of Contents .....	vi
List of Tables .....	viii
List of Figures .....	ix
Introduction .....	1
Statement of Problem .....	2
Statement of Purpose .....	2
Review of Literature .....	3
Methods for Evaluating Deviant Sexual Interest .....	3
Penile Plethysmography .....	4
Viewing Time .....	5
The Abel Assessment of Sexual Interest .....	8
Affinity 2.0 .....	10
Assessment Procedure .....	10
Data Interpretation .....	11
Reliability and Validity .....	12
Problems with Ipsative Scores .....	13
Statement of Problem .....	16
Statement of Purpose .....	17
Methods .....	18

Participants .....	18
Materials .....	18
Procedure .....	20
Data Analysis .....	21
Results .....	25
Discussion .....	29
References .....	33
Appendix .....	38



## LIST OF TABLES

1. Example of Method Used to Calculate Distortion Score by Comparing Rank Orders .....	22
2. Explanation of Variables Hypothesized to be Related to Distortion .....	24
3. Frequency of Distortion Score .....	27

## LIST OF FIGURES

1. Scatterplot of Distortion Scores and Interaction Effect Values  
for Entire Sample ..... 26

## Introduction

Assessment of deviant sexual interest can play an important role in the treatment and relapse prevention of sexual offenders. Assessment aids in the identification of preferred sexual stimuli and targets of deviant sexual behavior (Fischer, 2000). Once preferred stimuli are known, this information can be used to develop treatment plans directed at avoiding situations that may be high-risk, and preventing the initiation of an offense chain leading to perpetration by sexual offenders (Ward, Louden, Hudson, & Marshall, 1995).

The assessment of deviant sexual interest is complicated, and this field is an evolving science. Penile plethysmography (PPG) has been one of the most commonly used methods for evaluating deviant sexual interest, but this procedure seems to have serious limitations. For example, PPG involves exposing the participant to sexually explicit images, a procedure that may not be ethical in some cases. It is also an expensive, labor intensive and highly invasive procedure (Laws & Gress, 2004), but perhaps the most serious drawback is the lack of standardization. According to Laws and Gress, PPG does not have a standardized procedure for administration, scoring and interpretation of data—raising serious questions about the reliability and validity of the results.

In response to the limitations of PPG, researchers have tried to develop alternative methods for evaluating deviant sexual interest. One such method is viewing time assessment. Research has indicated that people spend more time viewing preferred sexual stimuli (Quinsey, Ketsetzis, Earls, & Karamanoukian, 1996; Wright & Adams, 1994; Zamansky, 1956). Therefore, in viewing time assessment the participant is exposed to images of various potential sexual stimuli, while viewing time is implicitly measured.

Two viewing time assessments are available commercially—the *Abel Assessment for Sexual Interest (AASI)* and *Affinity*. There appear to be serious questions about the reliability and validity of the AASI (Fischer, 2000; Smith & Fischer, 1999), and while there is some evidence for the reliability and validity of *Affinity* (Glasgow, 2003; Glasgow, Croxen, & Osborne, 2003) this evidence still needs further replication and support.

### *Statement of Problem*

Currently several researchers at Brigham Young University are gathering data on samples of non-offending, heterosexual males and females in an attempt to further examine the validity and reliability of the *Affinity* assessment. In doing so they have noticed that in some cases the conversion from raw viewing times to ipsative scores can result in distortion, where the ipsative score profile indicates a pattern of sexual attraction different from that of the raw viewing time profile. At the present time, no method exists for identifying cases where an ipsative score profile is distorted. Such a procedure would be valuable to researchers, by allowing them to eliminate invalid profiles from their analyses. It would also be valuable in the practical use of the *Affinity* instrument. Notifying clinicians of the potential invalidity of a profile would decrease the probability of arriving at erroneous conclusions regarding an examinee, and possibly indicate cases where the examinee was trying to obscure his or her true sexual preferences.

### *Statement of Purpose*

This thesis attempted to define ipsative score distortion objectively, examine the characteristics of distorted ipsative score profiles, and develop an algorithm to identify such distortions. Using a sample of non-offending heterosexual females, distorted profiles

were identified, and the attributes of these profiles were examined. Possible statistical methods for identifying the distortions were explored. The implications for researchers and clinicians are discussed.

### Review of Literature

Ward et al. (1995) proposed the idea of an offense chain, or a set of stages which sexual offenders progress through, leading to perpetration. These stages are distal planning, contact with potential victims, proximal planning, offending, cognitive restructuring, and future resolution. Theoretically this offense chain could be broken in the initial stages by structuring a sexual offender's environment in such a way that the offender is able to avoid potentially high-risk situations.

In this way, understanding a sexual offender's deviant sexual interests is key to the development of treatment and relapse prevention plans. Knowledge of deviant sexual interest can be used to identify likely victim groups. As a result it is possible to identify stimuli and environments which are potential triggers of the offense chain, and the offender's environment can be structured in such a way that exposure to these triggers is minimized (Fischer, 2000).

#### *Methods for Evaluating Deviant Sexual Interest*

Generally the assessment of sexual preference involves clinical interviews, analysis of records, self-report, and penile plethysmography (Laws, 1989; Marshall, 1996). Clinical interviews and self reports, "are somewhat subjective and may be compromised by dissimulation" (Fischer, 2000, p. 304). Therefore, other assessment tools, which are objective and less vulnerable to deceitful patterns of responding, are needed. PPG and viewing time (VT) assessment are two such methods currently in use.

*Penile plethysmography (PPG)*. It has been hypothesized that there are several stages of male sexual behavior including, sustained attention to the given stimuli, movement toward the stimuli, and physiological response—including genital response (Singer, 1984). PPG measures sexual interest at the final stage—sexual arousal/penile engorgement. In PPG a measurement device is attached to the individual’s penis. The individual is then exposed to audio and/or visual depictions of various sexual objects or behaviors, while penile engorgement is recorded (Fischer, 2000, p. 305).

PPG seems to be more objective and less vulnerable to dissimulation than clinical interviews. Unfortunately, this assessment procedure has a different set of drawbacks. PPG involves exposing the participant to sexually explicit images, which may not be ethical with some individuals—most notably adolescents. According to Fischer (2000), exposure to sexually deviant imagery, “may inadvertently introduce, stimulate, or ratify [sexual] behavior” (p. 305). The highly invasive nature of PPG raises additional ethical concerns, or at least makes the process labor intensive and uncomfortable. PPG also requires expensive equipment, and can only be used with males (Laws & Gress, 2004). Studies on vaginal plethysmography have indicated the procedure is invalid, because there is little correlation between women’s physiological response and their subjective experience of sexual arousal (Heiman, 1980; Hoon, 1984; Singer, 1984; Wincze, Hoon & Hoon, 1978).

These concerns are sufficient to underscore the need for alternate assessments, but another serious drawback to PPG exists. According to Laws and Gress (2004), PPG does not have a standardized procedure for the administration, scoring or interpretation of data--seriously limiting the value of obtained data. Without standardized procedures it is

virtually impossible to make inter-individual comparisons about sexually deviant interest. Lack of standardization also makes it impossible to establish the reliability of PPG and because reliability supersedes validity, validity is likewise not established.

*Viewing time.* The need for alternatives to penile plethysmography has led to the development of several other assessment methods. These assessment procedures measure sexual interest at stage one of Singer's (1984) model—sustained attention to preferred sexual stimuli. These tests, the *Abel Assessment for Sexual Interest* and *Affinity 2.0*, assess sexual interest by implicit measurement of time spent viewing potential sexual stimuli.

Research on sustained visual attention as an indicator of sexual interest began in 1942 when Rosenzweig demonstrated that sustained visual attention to sexual stimuli could be used to differentiate individuals with high sexual interest from those with low sexual interest (Rosenzweig, 1942). Rosenzweig devised an instrument he called a “photoscope,” and this study was conducted to validate the instrument. The “photoscope” contained 10 sexual and 10 nonsexual photographs placed on cards in a Rolodex-type device. Participants controlled the time spent viewing each slide, while an observer recorded viewing time from behind one-way glass.

Participants in this study were 20 inpatients with schizophrenia. Ten of these patients had been classified as having high sexual interest and ten had been classified as having low sexual interest, based on staff observations of overt sexual behavior. Both groups viewed the nonsexual photographs for an average of 19 seconds, but the groups differed significantly in the time spent viewing the sexual photographs. The high sexual interest group viewed sexual photographs for an average of 40 seconds, while the low

sexual interest group only viewed these slides for an average of 13 seconds (Rosenzweig, 1942).

Zamansky continued the research on viewing time, by examining whether VT could be used to discriminate homosexual males from heterosexual males (1956). In this study 20 heterosexual and 20 homosexual males were shown slides of male, female or neutral content. Once again, the results supported VT as an indicator of sexual preferences. On average, homosexual males viewed the pictures of men 3.55 seconds longer than heterosexual males, and heterosexual males viewed the pictures of women 2.37 seconds longer than homosexual males. These differences were significant at the  $p < .001$  level. Zamansky concluded that sexual preference, “will manifest itself in the pattern of an individual’s visual fixations, if these fixations can be measured without his awareness” (1956, p. 446).

Ware, Brown, Amoroso, Pilkey, and Pruesse (1972) measured sustained visual attention to pornographic images as a part of their study of the semantic meaning of these images. Participants in this study were college males, who viewed 15 slides, which varied in sexual explicitness from a clothed heterosexual couple to explicit depictions of various sexual acts. The authors found that viewing time increased as pictures became more sexually explicit. They also found that the best predictor of viewing time was a semantic scale labeled activity. The correlation between VT and activity was 0.93, indicating a strong relationship. Activity was a measure of the participants’ ratings of how stimulating, active and hot each image was. To put it simply, participants spent more time viewing pictures they found to be sexually stimulating.

Quinsey, Rice, Harris, and Reid (1993) measured heterosexual male and female’s



viewing time of nude, but non-sexually explicit images. This study found very little variation in viewing time across different ages and genders of the stimuli. The only significant difference was that the pictures of adults and pubescents of the preferred gender were viewed longer than all categories on the non-preferred gender. These results seem to indicate VT may not be a strong indicator of sexual preference when nonexplicit stimuli are used.

A number of other studies conducted in the 1990's lend further support to the use of VT as an indicator of sexual preference/attraction. Wright and Adams (1994) measured VT of male and female images from *Playgirl* and *Playboy* magazines. Participants included adult, heterosexual males and females, as well as adult, homosexual males and females. Significant differences in VT were found between groups, and in all cases slides of preferred sexual stimuli were viewed longest (Wright & Adams, 1994). In another study, child molester and normal heterosexual males' VT responses to nude male and female children, pubescents and adults were compared (Harris, Rice, Quinsey, & Chaplin, 1996). As expected, normal heterosexual males showed the greatest sustained visual attention to adult females. VT decreased with the age of female stimuli, and was lowest for all male stimuli. Interestingly, the child molesters' VT profile was very different. The average viewing time for men with a history of molesting was low, and the variability was restricted. Finally, Quinsey et al. (1996) conducted two studies of VT using normal heterosexual males and females. The studies produced consistent results, providing evidence of a pattern of sustained visual attention to preferred sexual stimuli among heterosexuals. As indicated by the Harris et al. (1996) study, participants had the longest viewing times for adult models of the preferred gender, with VT decreasing with

age. VT for the nonpreferred gender was generally flat across ages (Quinsey et al., 1996).

In summary, it appears viewing time can be used to discriminate between various groups of individuals based on their sexual preferences or levels of sexual interest (Rosenzweig, 1942, Zamansky, 1956, Harris et al., 1996). Also, there appear to be a consistent, positively correlated relationship between viewing time and preferred sexual objects (Wright & Adams, 1994; Quinsey et al., 1996). There are some cautions; for example, stimuli that are nonexplicit may result in less variable and less distinct VT responses, and sexual offenders seem to show flatter, more restricted VT patterns. However, this research supports the potential for viewing time assessment as a reasonable and meaningful alternative method for the evaluation of sexual interest.

#### *The Abel Assessment of Sexual Interest*

The *Abel Assessment of Sexual Interest (AASI)* is a psychometric instrument that measures sexual attraction to 22 categories of possible sexual stimuli. During administration the participant views 160 slides of individuals of various ages and genders. Some of the pictures portray various paraphilias, but all subjects are clothed. When viewing these slides, the subject is asked to rate how sexually arousing each slide is using a 7-point Likert-type scale. Sustained attention to each slide is implicitly measured throughout this process (Fischer, 2000).

On the surface this assessment instrument appears to be very promising. It claims to be able to assess attraction to a broad number of stimuli. It does not use images that are sexually explicit, avoiding some ethical issues. It is not physically intrusive, and it is less susceptible to faking because subjects are not aware of the variable being measured.

However, a number of questions have been raised regarding the validity and reliability of the AASI.

Up to this point attempts to substantiate the reliability and validity have produced discouraging results. Several researchers have examined the test-retest reliability of the AASI. Smith and Fischer (1999) found an average test-retest reliability of +0.63, Kaufman, Rogers, and Daleiden (1998) found average coefficients ranging from +0.55 to +0.67. In both cases the reliability did not fall in the desired range of  $r > 0.80$  (Anastasi, 1988).

Research has found mixed results regarding the validity of the AASI. Smith and Annon (1998) compared the AASI to PPG and found no significant correlations between viewing time and admitted sexual behavior. Smith and Fischer (1999) examined the validity with adolescent participants and found similar results. In the Smith and Fischer study the AASI was only able to discriminate offenders from nonoffenders slightly better than what could be done by chance, with reported hit rates ranging from 52-59%. On the other hand, some researchers have found the Abel Screen is able to identify offenders with accuracy comparable to PPG (Gray, 1999, Johnson & Listiak, 1999, Letourneau, 1999). Because of these mixed results, it is questionable whether the Abel Screen should be used in the assessment of sexual offenders at this time. Such results led Fischer and Smith (1999) to conclude that the AASI, “is a promising instrument based on an interesting concept. However, the evidence of its reliability and validity for use with adults is weak as yet. Further refinement is necessary so that its use may become a reliable and valid means to promote appropriate treatment of sexual offenders and possibly also a means to protect potential victims” (pp. 203-204).

### *Affinity 2.0*

*Affinity 2.0* is a computer program that, like the Abel Screen, assesses sexual interest through a combination of overt self-report and covert measurement of sustained visual attention to stimuli (Glasgow, 2003). The *Affinity* program was initially developed to assess the sexual interest of men with mild mental retardation. However, the manual for the most recent version of the program, version 2.0, states the program can be used for individual assessment with all adult male offenders and can be used for research purposes with adult male nonoffenders, juvenile male offenders, and female offenders (Glasgow, 2003).

*Assessment procedure.* The *Affinity* program consists of two main tasks—a ranking procedure and a rating procedure (Glasgow, 2003). For the ranking task the subject is shown eight simple sketches of people, representing different sexes and stages of development. Because the sketches are simple, they can only be discriminated based on clothing type, size and ratio of body parts. Subjects are prompted to rank these images starting with most sexually attractive. Once they have ranked all the images they find sexually attractive, they then rank the remaining images starting with the most sexually unattractive. This task allows subjects to self-report the gender and age groups they find sexually attractive. Also, the examiner can choose to have the program display the assessment results in this rank order, making discrepancies between self-reported interest and actual viewing time immediately apparent visually.

The second part of *Affinity* is a rating task. During this portion of the assessment the subject is presented with seven images from each of eight stimulus categories (Glasgow, 2003). Each category depicts individuals of a certain gender and age group

(adult, juvenile, prejuvenile and small child). The examiner can choose to present these 56 images in random order, or a predetermined sequence. Typically a random administration is utilized. As the subject is exposed to each image, he is asked to rate how sexually attractive or unattractive he finds the individual depicted in the picture on a likert-type scale. Possible ratings range from -7 to +7.

The subject's ratings are recorded by the computer, while two types of viewing time are covertly measured—On Task Latency (OTL) and Post Task Latency (PTL). OTL is a measurement of viewing time, beginning with presentation of the image and ending when the subject makes a rating. PTL is a measurement of the time that elapses between the subject making a rating and then clicking “next image” (Glasgow, 2003).

*Data interpretation.* The results of *Affinity* can be presented in several formats including raw data—average viewing time for each category—and mean rank data. Mean ranks are the primary method for reporting results, and are obtained by assigning each image a rank corresponding to its raw score. For OTL the image that was viewed longest would receive a ranking of 56, while the image viewed for the shortest duration would receive a ranking of 1. Once all images have been ranked, a mean rank is computed for each category. Then the results are presented in a line graph (Glasgow, 2003).

The manual for *Affinity* lists two reasons for using mean rank, or ipsative data. First, it makes it possible to chart all data on a single graph. Second, it helps minimize the effect of outliers on mean scores (Glasgow, 2003). On the other hand, the manual acknowledges that:

A disadvantage of using mean ranks is that all correspondence with actual values in the raw data is lost, as is most sense of the distribution of scores. There is no

“calibration” of latencies, so the absolute difference between latencies regarding two categories is only expressed relative to all the other scores, not to absolute duration. It is wise therefore to refer back and forth between raw data and mean rank charts in order to properly analyze the results. (Glasgow, 2003, p. 53).

According to the test developer, the preferred method of data interpretation is visual analysis, with attention to the real world significance of patterns to the data (Glasgow, 2003). In an article published in 2003 (Glasgow, Croxen, & Osborne, 2003), the test developer gives several examples of data interpretation. Generally this procedure, as described in the article, involves visually analyzing the results in a line chart form. The practitioner considers the individual’s history of sexual offense and sexual behavior, examines his self-reported attraction to certain stimulus categories, and considers how this information fits with his pattern of viewing time. Particular emphasis is given to the mean rank, viewing time information because this information is considered to provide insight that is unaffected by dissimulation. If there is an inconsistency between an individual’s pattern of viewing time and his history or self-reported sexual interests, it would be considered justifiable cause to further examine the individual’s self-reports, and consider the possibility of deceitfulness on his part.

*Reliability and validity.* Research examining the reliability and validity of *Affinity* is sparse to date. Glasgow et al. (2003) conducted a pilot study of *Affinity*, utilizing single case methodology. They concluded *Affinity* was able to produce valuable information as a part of an assessment of sexual interest. However, little evidence of reliability and validity was gained from this study, because the single case methodology does not lend itself to generalization.

A second study by Glasgow and Croxen (2003) provided more information about validity and reliability. The sample for this study consisted of 31 pedophile offenders and 31 non-offenders. Internal consistency for each of the stimulus categories was analyzed, and the resulting Cronbach Alphas ranged between 0.76 and 0.93.

Validity was then examined in two ways. First, the researchers computed the correlations between rated sexual interest and viewing time. This analysis resulted in a median correlation of 0.57 for the non-offenders and 0.35 for the offenders. The authors accounted for this difference by pointing out that there is a greater likelihood of “denied pedophile interest in the offender sample” (Glasgow & Croxen, 2003, p. 6). Second, they examined the discriminate validity of *Affinity*, and found that the test was able to correctly identify pedophiles with 96% accuracy. On the other hand, 23% of the non-offenders were falsely identified as offenders. In the authors’ opinion, these results were sufficient to conclude that *Affinity* can be of practical use. As a pilot study these results are encouraging, but clearly the reliability and validity of *Affinity* must be further examined.

#### *Problems with Ipsative Scores*

Raw scores and normative or standard scores are commonly used in psychological assessment. As such, practitioners are generally familiar with the basic characteristics and uses of such data. However, the mean rank scores employed by *Affinity* are rather uncommon. These scores are considered to be ipsative. By definition the term ipsative means, “using yourself as the norm against which to measure something.” (Encarta, n.d.). While normative scores describe a person’s performance in comparison to others, ipsative scores describe a person’s performance in comparison to himself.

A technical definition of ipsative scores is, “any score matrix, which has the property that the sum of the scores over the attributes for each of the entities is constant.” (Clemans, 1956, p. 4). This means that ipsative scores always sum to a constant. If one score is higher on an ipsative scale, then one or more other scores must be lower. In other words, “each score for an individual is dependent on his scores on other variables” (Clemans, p. 1).

Because ipsative scores are used to make intra-individual comparisons it is first necessary to convert all scores to a common, standard form. Failing to make such conversions before comparing data would be like measuring one person in feet and another in meters and then trying to compare their heights without accounting for the differing scales of measurement. Converting raw scores to a standard form is typically done by converting them to standard scores. However, in the case of *Affinity* this conversion is not necessary, because all scores are already in a common form—viewing time in seconds. Once the different scores have been converted to a standard form, intra-individual comparison’s can be made, but the units are not considered to be ipsative until a further conversion has been performed.

The next step in converting raw scores to ipsative scores is to make the scores sum to a constant. This typically involves adding a constant number to each of the standard scores, such that the resulting sum of scores is 100. However, in the case of *Affinity* this is accomplished by converting raw scores to rank scores. Scores are ordered by viewing time, and the longest viewing time is assigned a rank of 56. The slides are ranked in order down until the last slide, which is given a rank of 1. Next, mean ranks for



each category are computed. As a result of this ranking procedure, the mean ranks always sum to 228, making them ipsative.

The practical value of ipsative scores is that they allow quick comparison of several of an individual's attributes. However, a number of drawbacks exist. First, it is important to note that all of the foundational research connecting viewing time to sexual interests relied on raw scores. The use of ipsatives was introduced by Abel (1996) and in many ways was disconnected from the literature. Second, by making all of an individual's scores sum to a constant, any relation to the absolute value of the attribute is lost. Many of the weaknesses of ipsative scores are tied to this loss.

One weakness of ipsative scores is that they cannot be used to make inter-individual comparisons regarding specific attributes, because all relation to the absolute value of an attribute was lost when the conversion to ipsative scores was made. Any attempts at inter-individual comparisons would be like concluding that because my right leg is longer than my left leg, my right leg is also longer than your left leg.

Another weakness is that conclusions about the degree of a trait possessed by an individual cannot be made based on ipsative scores. A high ipsative score does not equate to a high degree of an attribute in actual value—only in comparison to the other attributes the individual possesses. For example, if ipsative scores were used to rank an individual's personality traits, and he obtained his highest score on the extroversion scale, it would not necessarily be accurate to conclude he is an extrovert. If this individual's raw scores were low in all of the measured personality traits, but he scored slightly higher on extroversion, his ipsative score for extroversion would be his highest score, but relative to the general population it would still be a low score.

The same principles that complicate the interpretation of ipsative scores can also be responsible for distortion of ipsative scores. When raw scores on *Affinity* are converted to ipsative mean ranks, the natural variability in scores is changed to an artificial ranking, with even intervals between each score. In this conversion large differences between scores can be minimized and minor differences can be exaggerated. For example, when an individual has a flat viewing time profile, minor, random differences in VT scores could cause one category to be consistently ranked higher than another. The resulting mean rank profile would indicate the individual has a sexual preference for that category, when in reality no meaningful differences in VT exist. In such cases, if a practitioner relies solely on the mean rank scores for interpretation of data, it is likely they will reach false conclusions about the subject's sexual preferences.

#### *Statement of Problem*

When raw scores on *Affinity* are converted to ipsative scores they are changed from ratio to ordinal data. In doing so the data's relation to actual viewing times is lost. Theoretically, certain distortions can occur as a result of this conversion—causing insignificant differences in viewing time to appear meaningful, and possibly even minimizing meaningful differences.

Currently several researchers at Brigham Young University are gathering data on samples of non-offending, heterosexual males and females in an attempt to further examine the validity and reliability of the *Affinity* assessment. In doing so they have observed cases where the conversion to ipsative scores has indeed caused distortions in an individual's score profile. At the present time, no method exists for identifying cases where an ipsative score profile is invalid. Such a procedure would be valuable to

researchers, by allowing them to eliminate invalid profiles from their analyses. It would also be valuable in the practical use of the Affinity instrument. Notifying clinicians of the potential invalidity of a profile would decrease the probability of arriving at erroneous conclusions regarding a subject, and possibly indicate cases where the subject was trying to obscure his true sexual preferences.

#### *Statement of Purpose*

The purpose of this study was to define ipsative score distortion objectively, examine the characteristics of distorted ipsative score profiles, and attempt to develop an algorithm to identify such distortions. Using a sample of non-offending heterosexual females, distorted profiles were identified, and the attributes of these profiles were examined. Possible statistical methods for identifying the distortions were explored. The implications for researchers and clinicians are discussed.

## Methods

### *Participants*

The present study used data gathered by Kara Harmon as a part of her doctoral dissertation. Her study examined the reliability and validity of *Affinity 2.0* scores with non-offending, heterosexual females. Participants were 154 females with a minimum age of 18. All participants were undergraduate students at Brigham Young University and were recruited using short presentations given by researchers to undergraduate psychology classes. In these presentations potential participants were informed that the purpose of the study was to test a new device measuring sexual interest. They were also informed that participation would involve viewing several pictures of fully-clothed models, rating those images on their attractiveness or unattractiveness and then completing a short questionnaire. Participants were compensated with movie tickets.

Only participants who reported exclusive heterosexual interest, with no history of pedophilia, were included in the experimental groups. Participants were screened for these variables through the informed consent procedure and a self-report questionnaire that was administered at the end of the evaluation. Eight participants were excluded from the study because they did not report exclusively heterosexual interest, leaving a sample of 146 participants. These participants were compensated like the other participants.

### *Materials*

All participants included in the experimental group completed *Affinity 2.0* at test and retest. *Affinity* is a computer program that assesses sexual interest through a combination of overt self-report and covert measurement of sustained visual attention to various stimuli (Glasgow, 2003). The *Affinity* program was initially developed to assess

the sexual interest of men with mild mental retardation. However, the manual for the most recent version of the program, version 2.0, states the program can be used for individual assessment with all adult male offenders, and can be used for research and evaluation with adult male nonoffenders, juvenile male offenders, and female offenders (Glasgow, 2003).

The *Affinity* program consists of 2 main tasks—a ranking procedure and a rating procedure (Glasgow, 2003). The ranking task allows participant to self-report the stimulus groups they find sexually attractive. During the ranking task the participant is exposed to images depicting individuals of varying ages and genders. The participant is asked to rate how sexually attractive or unattractive he finds the individual depicted in the picture on a Likert-type scale. The participant's ratings are recorded by the computer, while viewing time is measured (Glasgow, 2003).

After taking *Affinity 2.0* for the second time, participants were given the Demographics, Social Desirability and Sexual Interest Questionnaire (DDSQ). The DDSQ is composed of three sections. In the first section the participants reported demographic variables such as age, ethnicity, year in school, and marital status. The second section was a measure of social desirability called the M-C 2(10). The M-C 2(10) is a condensed form of the Marlowe-Crowne Social Desirability Scale (Strahan & Gebrasi, 1972). The M-C 2(10) was included to help indicate whether social desirability may have influenced a participant's viewing times.

The third section of the DDSQ was a question about sexual orientation. The question was an adaptation of the Kinsey Heterosexual-Homosexual Scale (Kinsey, Pomeroy, & Martin, 1998). The second question asked whether the participant had a

history of pedophilia. These questions were used as exclusionary variables. Of the original 154 participants, 8 were excluded based on their responses to these items, leaving the 146 participants used in the present study.

### *Procedure*

Assessing sexual interest can be an intrusive and uncomfortable experience for many individuals. Some individuals may be hesitant to respond to questions honestly unless a comfortable and confidential environment is provided. Brigham Young University is a private religious school that requires students to comply with a strict code of conduct. This code of conduct includes prohibitions on sexual activity outside of marriage including: homosexual activity, pedophilia and extra-marital sexual relations. Violations of this code can result in university sanctions, expulsion, and even criminal prosecution (Brigham Young University, n.d.). Consequently, participants with homosexual or pedophilic tendencies may be extremely hesitant to respond honestly on research instruments.

To address this issue all participants were required to read and sign an informed consent document. Care was taken to provide participants with a comfortable, confidential assessment environment. The informed consent document explained the purpose of the study and expectations for participants. It also discussed in detail the methods for protecting the confidentiality of the identity of all potential participants. This included assigning numbers to each participant, deleting the names, only keeping one master list of the names, and ensuring participants that no names would be used in the study or reported to the university. The master list was kept in a secure, locked file.

The informed consent procedure was also used to screen for any history of pedophilia. A negative history of pedophilia and an age of 18 or older were the initial inclusion criteria. Participants were asked to confirm that they met those criteria before engaging in the assessment.

After reading and signing the informed consent document, participants were taken to a private room, containing a computer with the *Affinity 2.0* program. In each case the researcher instructed the participant in how to begin the program and then exited the room. The researcher informed each participant that she would wait outside the room until the participant completed the assessment, to make sure they were not disrupted. After completing the assessment, the researcher answered any questions asked by the participant and scheduled a time to retest. At retest the same procedure was followed, except that the DDSQ was administered following *Affinity 2.0*.

#### *Data Analysis*

Data analysis involved examining the characteristics of distorted profiles, and attempting to develop an algorithm that could be used to identify distorted profiles. Before either of these steps could be completed it was necessary to define distortion objectively. The primary concern with distorted profiles is that raw viewing times may indicate one pattern of sexual interest, while ipsative score profiles indicate a different pattern of sexual interest. For example, there were several instances in the study sample where a participant's highest mean raw viewing time was in the Adult Female category, but on the ipsative profile the Adult Male category was ranked the highest. This type of inconsistency could lead to erroneous conclusions about sexual interest, so in the present

study distortion is defined in terms of inconsistent rank ordering between raw VT and ipsative scores. This type of inconsistency was defined by creating a distortion score.

Distortion scores were obtained by first ranking the ipsative scores from 1 to 8, with 1 being the category with the highest ipsative score and 8 being the category with the lowest score. Raw viewing times were also ranked in the same way, with the category with the longest viewing time being ranked 1, and the category with the shortest viewing time being ranked 8. Once the scores had been ranked, the total number of rank position movements between raw viewing time and ipsative scores were totaled. Each value was converted to an absolute value of movement. One point was assigned for each rank position moved. The points for all categories were summed to obtain a total distortion score for that participant. An example of this procedure is shown in Table 1.

Table 1

*Example of Method Used to Calculate Distortion Score by Comparing Rank Orders*

	ADF	ADM	JUF	JUM	PJF	PJM	SCF	SCM	$\Sigma$
Raw VT Rank	1	2	4	3	5	6	7	8	—
Ipsative Rank	2	4	1	3	5	6	7	8	—
Positions Moved	-1	-2	+3	0	0	0	0	0	—
Absolute Value of Positions Moved	1	2	3						6

Note: ADF = Adult Female, ADM = Adult Male, JUF = Juvenile Female, JUM = Juvenile Male, PJF = Prejuvenile Female, PJM = Prejuvenile Male, SCF = Small Child Female, SCM = Small Child Male.



The result is a distortion score with possible values ranging from 0 (no difference in rank order between raw viewing time and ipsative scores) to 32 (the score that would be obtained if all categories moved the maximum amount possible).

Once distortion scores were obtained, the characteristics of distorted profiles were explored. First, the profiles were examined visually. This was done using Microsoft Excel to graph the ipsative score profiles and the raw viewing time profiles for all of the participants. The profiles were then compared for each participant. The purpose of this visual examination was to identify characteristics of distorted or undistorted profiles, which could be further examined statistically. These graphs are contained in the appendix. This visual analysis did not identify any characteristics of distorted profiles that were present in all cases. However, several hypotheses were developed based on this examination. First, distorted profiles appeared to be more likely to have greater variance in the raw viewing times. Second, some distorted profiles had one or more categories with significant skew. Third, distortion seemed more likely to occur in cases where there was less variation between mean raw viewing time for each category. Fourth, distortion seemed more likely when there was a greater standard deviation of skew scores.

Once several hypotheses regarding distortion were developed, these hypotheses were tested using Pearson Product Moment Correlations. It was assumed that the variables that contribute to distortion would have stronger correlations with the distortion score. Variable labels and descriptions of variables are contained in Table 2.

Table 2

*Explanation of Variables Hypothesized to be Related to Distortion*

Variable Label	Description
STM	<p>The standard deviation of mean raw viewing times for each subject. This variable was calculated as a percent of the mean raw VT to correct for differences in mean raw VT from subject to subject.</p> $STM = (SD_{\text{rawVT}}/M_{\text{rawVT}})*100$
AVGSD	<p>The average raw viewing time standard deviation across all eight stimulus categories. This variable was calculated as a percent of the mean raw VT to correct for differences in mean raw VT from subject to subject.</p> $AVGSD = [\sum(SD_{\text{ADF}}, SD_{\text{ADM}}, \dots)/8]/M_{\text{rawVT}}*100$
SUMSKEW	<p>The sum of raw viewing time skew scores for all eight stimulus categories.</p> $SUMSKEW = \sum(\text{SKEW}_{\text{ADF}}, \text{SKEW}_{\text{ADM}}, \dots)$
SDSKEW	<p>The standard deviation of raw viewing time skew scores for a subject.</p>

After correlation coefficients had been calculated, interaction effects were considered. An algorithm for identifying distorted profiles was developed using the significant correlations. Different cut-offs were tested, and the cut-off yielding the best hit rate was identified.

## Results

The method described in the data analysis section was used to calculate distortion scores for each subject. When this method was applied to the sample, 21 participants had a distortion score of 0. For the remaining 125 participants distortion scores ranged from 2 to 16. Frequencies of distortion scores are outlined in Table 3.

Next, Pearson Product Moment Correlations were calculated for the variables described in Table 2. The correlation between distortion and STM was not significant ( $r = -.162, p > .05$ ). However, the remaining correlations were weak, but statistically significant. A positive correlation was found between distortion and AVGSD ( $r = .405, p < .01$ ), distortion and SUMSKEW ( $r = .321, p < .01$ ), and distortion and SDSKEW ( $r = .265, p < .01$ ).

Although three of these variables appear to be related to distortion, none is a strong predictor of distortion by itself. In fact, the variable most strongly correlated with distortion, AVGSD, can only explain 16.4% of the variability in distortion scores ( $r^2 = .164$ ). Because these correlations were weak, other relationships were examined. It was reasonable that some interaction of these variables might best predict distortion, so several interaction effects were calculated using different combinations of the variables. The strongest correlation was found when all four variables were used to calculate an interaction effect. This interaction effect (INT) was calculated using the following formula:

$$\text{INT} = (\text{AVGSD} * \text{SUMSKEW} * \text{SDSKEW}) / \text{STM}$$

When a Pearson correlation coefficient was calculated for the variable INT and distortion, a moderately strong relationship was found ( $r = .530, p < .01$ ). Although a

moderately strong relationship is present, these variables still only account for 28.1% of the variability in distortion ( $r^2 = .281$ ). The relationship between these variables and distortion is graphically represented in Figure 1.

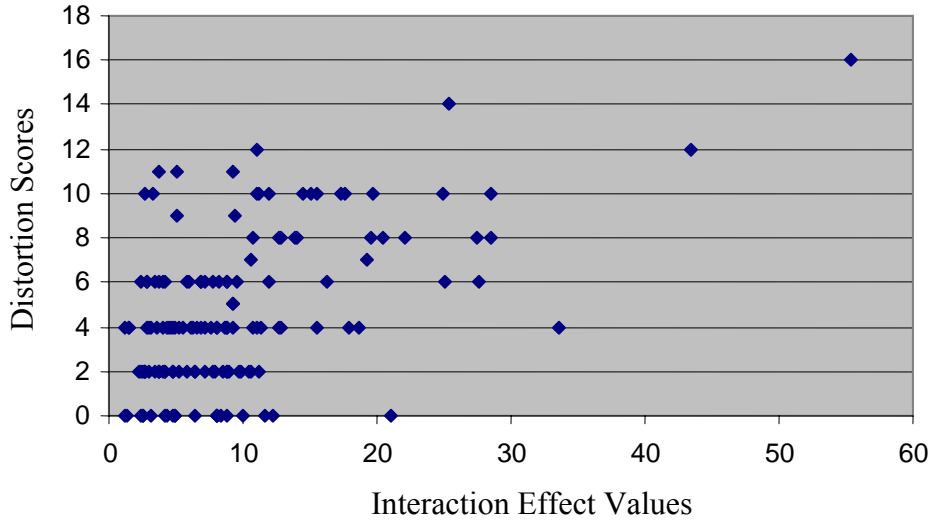


Figure 1. Scatterplot of distortion scores and interaction effect values for entire sample.

The correlation between the variable INT and distortion was the most significant correlation discovered; therefore, INT was the variable used in the development of an algorithm for identifying distortion. For the purposes of the present study distortion scores greater than 10 were considered significantly distorted. This cut-off was chosen because it provided a reasonable range for identifying distorted scores, with any score from 11 to 32 being considered distorted, but it did not label an excessive percentage of profiles as significantly distorted. As shown in Table 3, only 4.9% of the study sample received a distortion score greater than 10.

Table 3

*Frequency of Distortion Scores*

Distortion Score	Frequency	Percent
0	21	14.4
2	32	21.9
4	36	24.7
5	2	1.4
6	21	14.4
7	2	1.4
8	10	6.8
9	2	1.4
10	13	8.9
11	3	2.1
12	2	1.4
14	1	.7
16	1	.7

Once this definition of significant distortion had been decided upon, different values of INT were tested as cut-off points, and hit rates were examined. The best hit rate was obtained when the cut-off of 25 was used (INT scores greater than 25 were considered to indicate significant distortion). Using this cut-off 42.9% of the significantly distorted profiles were correctly identified as such. On the other hand, 4.3% of the undistorted profiles were falsely identified as distorted. Although this formula did not incorrectly identify a large percentage of profiles as distorted, it did a poor job of identifying the significantly distorted profiles.

## Discussion

The primary purpose of the present study was to develop an algorithm for the identification of distorted profiles on *Affinity 2.0*. Although several variables related to distortion were identified, the correlations were weak, and even when interaction effects were calculated, only moderately strong relationships were found. Without identifying a variable strongly correlated with distortion, an algorithm that predicted distortion with any meaningful level of accuracy could not be developed. The hit rates obtained by using the proposed formula were not accurate enough to be of any practical value.

Although a useful formula for identifying significant distortion was not developed, the variables found to be related to distortion give us some insight into causes of distortion. The individual variable most strongly correlated with distortion was AVGSD (see Table 2). This positive correlation indicates distortion is more likely in cases where the raw VT standard deviations for stimulus categories are larger. The positive correlations between SUMSKEW and distortion, and SDSKEW and distortion also suggest distortion is more likely when raw VT for stimulus categories is more skewed, or when there is greater variance in skew scores for stimulus categories. To summarize this information, we would expect greater probability of distortion in cases where there is greater variability in the time a person spends viewing each slide, and where the individual views one or two slides for a significantly longer time period than most other slides (causing skew).

This information makes logical sense when we consider how the rank order scores are obtained. If a participant views one slide in a stimulus category for a particularly long time, and views the other slides in the category for short durations, the mean raw viewing

time for that category would be high. However, when the scores were converted to rank scores only the one outlier would receive a high rank, but the others would probably receive lower rank scores. Therefore, another category with scores generally higher than all the scores except the outlier could end up with a higher mean rank score. In other words, the categories could flip-flop.

Although the correlation coefficients were significant, and there seems to be a logical explanation why these variables are related to distortion, these relationships should not be overstated. These variables only appear to account for a small portion of the variability in distortion scores, and the conclusion must be that there are other factors, yet undiscovered, that are more significant contributors to distortion.

As described in the results section, for the purposes of the present study significant distortion was defined as a distortion score greater than 10. This score was selected because it identified a reasonable percentage of the sample (approximately 5%), but still left ample possibilities for significant distortion scores (11 to 32). Obviously this cut-off is rather arbitrary, and logical arguments could be made for various cut-off points. However, this leads to an interesting finding in the present study. As shown in Table 3, only 21 of the 146 participants, or 14.4 percent, had zero distortion between their raw viewing time profiles and their ipsative score profiles. The remaining 125 participants (85.6%) had at least some ipsative score distortion.

Although scores of 10 and below were not considered to be significantly distorted in this study, the argument could be made that all distortion is meaningful to the practitioner. For example, if a participant spent the most time viewing slides in the Adult Female Category, followed by the Adult Male Category, but these category rankings



were inverted on the ipsative score profile, this would only be a distortion score of 2 (assuming all other category placements were consistent). Although this is the lowest possible distortion score, even this minor distortion could lead to contrary conclusions about the client's preferred sexual stimuli (i.e. she is more attracted to adult males than adult females).

Based on this argument, the entire use of ipsative scores is questionable, because they may lead to conclusions that are inconsistent with the raw viewing time (VT) profile. As mentioned previously, over 85% of the profiles in the present study were distorted to some degree, which means the ipsative profiles inaccurately represent sustained visual attention in the majority of cases. If research supports sustained visual attention as an indicator of sexual interest, but ipsative scores do not accurately represent sustained visual attention, how much confidence can be placed in the conclusions drawn from ipsative scores? It seems that this is a fundamental problem with the instrument that seriously threatens its validity.

While some might argue that this is an over-exaggeration of the flaws, the fact remains that distortion of varying degrees is the rule, not the exception, with *Affinity 2.0*. That is not to say that the underlying theory behind the instrument is unsound, or that the instrument should be thrown out all together, but if a reliable formula for identifying distortions cannot be found, and distortion seems to be too prevalent, alternative methods of scoring should be examined. The present study seems to underscore the value of such exploration. If a scoring method that was less susceptible to distortion, and more closely tied to raw VT could be developed this would be ideal, since all of the foundational research used raw VT, not ipsative scores.

If alternate methods for scoring *Affinity 2.0* can be found, it may be unnecessary to “flag” distorted profiles. In the meantime, the question of how to identify distorted profiles remains. Although an algorithm was not discovered, it is still possible to identify and “flag” distorted profiles by using a distortion score, such as the one used in the present study. This method circumvents the complicated process of discovering and developing a mathematical algorithm, and identifies distortion in practical, straightforward terms—inconsistency between raw viewing times and ipsative scores. As described previously, a cut-off score of ten is recommended. However, this is only a preliminary recommendation.

There are several important limitations to the present study. First, the sample only includes females, and it remains to be seen whether a similar pattern of distortion is also present in a male population. Second, the sample did not include any offenders, and it would also be important to know whether distortion rates are similar in that population. It should be noted that this is only a preliminary study. Before cut-off points are adopted or the conclusions of this study are generalized, the findings should be replicated and supported with other populations.

Because *Affinity 2.0* is a promising, but relatively unsubstantiated instrument, efforts to establish its reliability and validity are ongoing. These efforts are critical, but the question of ipsative score distortion may supersede them. If the scores these researchers are studying are distorted in the first place, it may be extremely difficult to establish reliability and validity, and it may be wise to begin considering alternatives to the use of ipsative scores.

## References

- Abel, G. G. (1996). *A new objective test for youthful offenders: The Abel Assessment for Interest in Paraphilias*. Atlanta, GA: Abel Screening, Inc.
- Anastasi, A. (1988). *Psychology testing* (6<sup>th</sup> ed.). New York: Macmillan.
- Brigham Young University. (n.d.). Church educational system honor code. Retrieved June 27, 2005 from <http://campuslife.byu.edu/HONORCODE/chaste.htm>
- Clemans, W. V. (1956). An analytical and empirical examination of some properties of ipsative measures. *Psychometric Monographs*, 14.
- Encarta dictionary. (n.d.). Retrieved June 27, 2005, from <http://encarta.msn.com/encnet/features/dictionary/DictionaryResults.aspx?search=ipsative>
- Fischer, L. (2000). The Abel Screen: A non-intrusive alternative? In D. R. Laws, S. M. Hudson & T. Wards (Eds.), *Remaking relapse prevention with sex offenders: A sourcebook* (pp. 303-318). Thousands Oaks, CA: Sage Publications.
- Fischer, L. & Smith, G. (1999). Statistical adequacy of the Abel assessment for interest in paraphilias. *Sexual Abuse: A Journal of Research and Treatment*, 11 (3), 195-205.
- Glasgow, D. (2003). Affinity 2.0 [Software and Instruction], David Glasglow: Database Right.
- Glasgow, D. V., & Croxen, J. D. (2003). *Viewing time as a measure of paedophile sexual interest: Empirical evidence, the potential of social cognitions as a theoretical model, and some possible implications for assessment and intervention*. Manuscript submitted for publication.

- Glasgow, D. V., Croxen, J., & Osborne, A. (2003). An assessment tool for investigating pedophile sexual interest using viewing time: An application of single case methodology. *British Journal of Learning Disability, 31*, 96-102.
- Gray, S. (1999, September). Outcomes of the Abel assessment and the penile plethysmograph in a sample of sex offenders in outpatient treatment. In S. Johnson (Chair), *Current research comparing plethysmography with the Abel Assessment for Sexual Interest*. Symposium conducted at the annual meeting of the Association for the Treatment of Sexual Abusers, Orlando, FL.
- Harris, G. T., Rice, M. E., Quinsey, V. L., & Chaplin, T. C. (1996). Viewing time as a measure of sexual interest among child molesters and normal heterosexual men. *Behaviour Research and Therapy, 34* (4), 389-394.
- Heiman, J. R. (1980). Female sexual response patterns: Interactions of physiological, affective, and contextual cues. *Archives of General Psychiatry, 37*, 1311-1316.
- Hoon, P. W. (1984). Physiologic assessment of sexual response in women: The unfulfilled promise. *Clinical Obstetrics and Gynecology, 27* (3), 767-780.
- Johnson, S. A., & Listiak, A. (1999, September). Comparing outcomes of plethysmographic assessment with the Abel assessment in a prison based sex offender sample. In S. Johnson (Chair), *Current research comparing plethysmography with the Abel Assessment for Sexual Interest*. Symposium conducted at the annual meeting of the Association for the Treatment of Sexual Abusers, Orlando, FL.

- Kaufman, K. L., Rogers, D., & Daleiden, E. (1998, October). *Sexual interest in incarcerated adolescent offenders: A test-re-test analysis*. Poster session presented at the annual meeting of the Association for the Treatment of Sexual Abusers, Vancouver, BC.
- Kinsey, A. C., Pomeroy, W. B., & Martin, C. E. (1998). *Sexual behavior in the human male*. Bloomington, IN: Indiana University Press.
- Laws, D. R. (1989). *Relapse prevention with sexual offenders*. New York: Guilford.
- Laws, D. R., & Gress, C. (2004). Seeing things differently: The viewing time alternative to penile plethysmography. *Legal and Criminological Psychology, 9* (2). 183-196.
- Letourneau, E. J. (1999, September). A comparison of the penile plethysmograph with the Abel Assessment for Sexual Interest on incarcerated military sex offenders. In S. Johnson (Chair), *Current research comparing plethysmography with the Abel Assessment for Sexual Interest*. Symposium conducted at the annual meeting of the Association for the Treatment of Sexual Abusers, Orlando, FL.
- Marshall, W. L. (1996). Assessment, treatment, and theorizing about sex offenders: Developments during the past twenty years and future directions. *Criminal Justice and Behavior, 23* (1), 162-199.
- Quinsey, V. L., Rice, M. E., Harris, G. T., & Reid, K. S. (1993). Conceptual and measurement issues in the phylogenetic and ontogenetic development of sexual age preference in males. In H. E. Barbaree, W. L. Marshall, & S. M. Hudson (Eds.), *The juvenile sex offender* (pp. 143-163). New York: Guilford.
- Quinsey, V. L., Ketsetzis, M., Earls, C., & Karamanoukian, A. (1996). Viewing time as a measure of sexual interest. *Ethology and Sociobiology, 17* (5), 341-354.

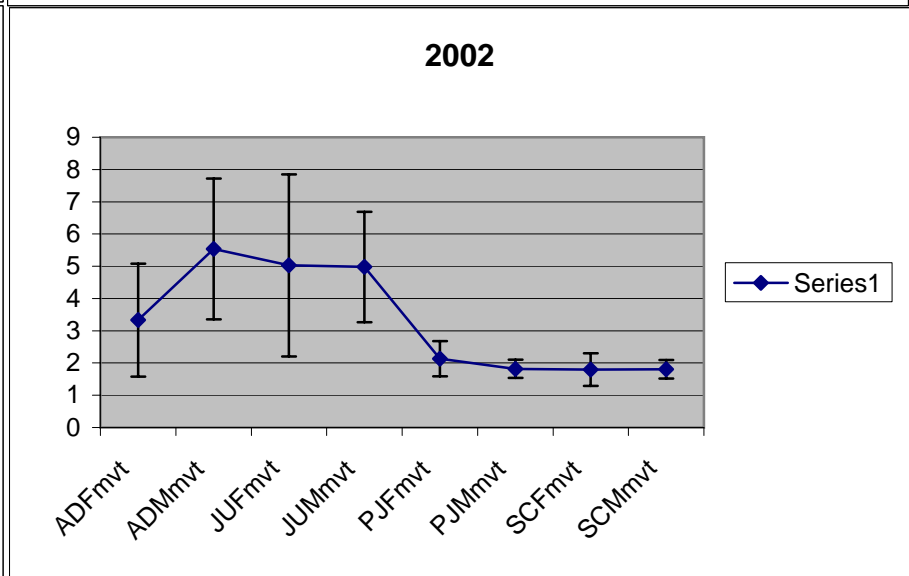
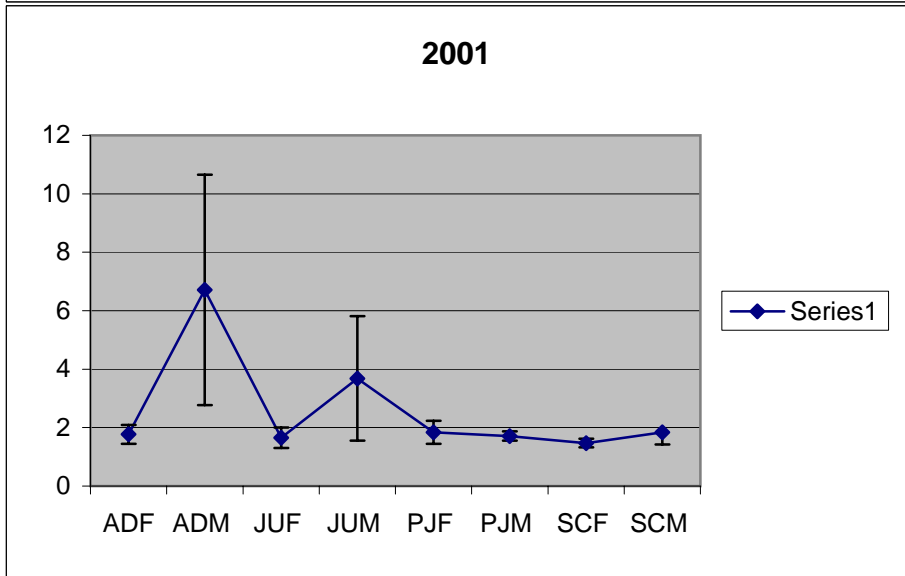
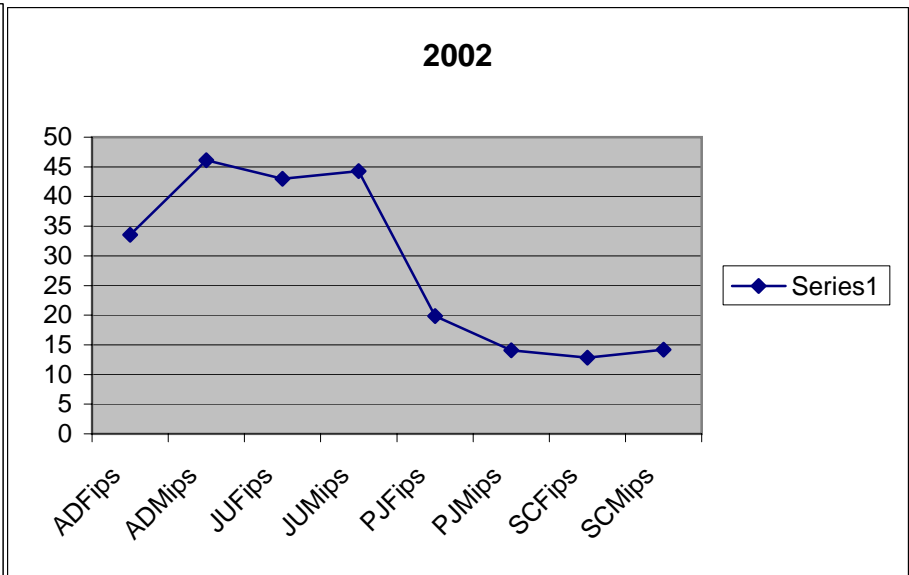
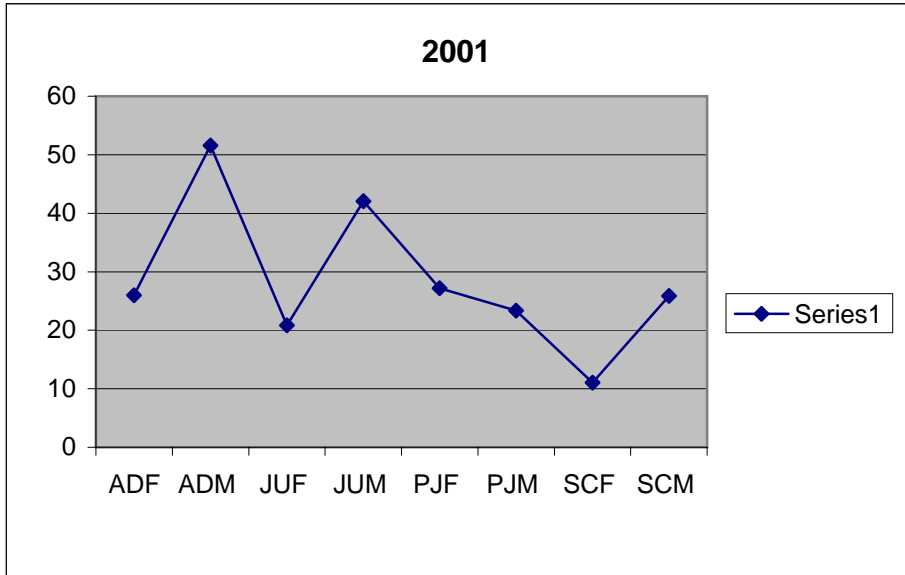
- Rosenzweig, S. (1942). The photoscope as an object device for evaluating sexual interest. *Psychosomatic Medicine*, 4, 150-158.
- Singer, B. (1984). Conceptualizing sexual arousal and attraction. *The Journal of Sex Research*, 20 (3), 230-240.
- Smith, B., & Annon, J. (1998, October). *Abel assessment and penile plethysmography: A comparison of instrument sensitivity*. Poster session presented at the annual meeting of the Association for the Treatment of Sexual Abusers, Vancouver, British Columbia.
- Smith, G., & Fischer, L. (1999). Assessment of juvenile sex offenders: Reliability and validity of the Abel Assessment for interest in paraphilias. *Sexual Abuse: A Journal of Research and Treatment*, 1 (3), 207-216.
- Strahan, R., & Gerbasi, K. C. (1972). Short, homogenous versions of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology*, 28, 191-193.
- Ward, T., Louden, K., Hudson, S. M., & Marshall, W. L. (1995). A descriptive model of the offense chain for child molesters. *Journal of Interpersonal Violence*, 10, 342-372.
- Ware, E. E., Brown, M., Amoroso, D. M., Pilkey, D. W., & Pruesse, M. (1972). The semantic meaning of pornographic stimuli for college males. *Canadian Journal of Behavioral Science*, 4, 204-209.
- Wincze, J. P., Hoon, E. F., & Hoon, P. W. (1978). Multiple measure analysis of women experiencing low sexual arousal. *Behavioral Research and Therapy*, 16, 43-49.

Wright, L. W., & Adams, H. E. (1994). Assessment of sexual preference using a choice reaction time task. *Journal of Psychopathology and Behavioral Assessment*, 16 (3), 221-231.

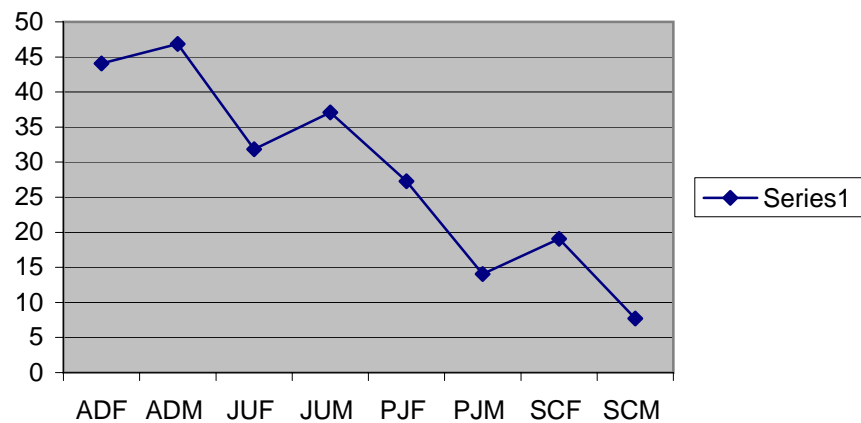
Zamansky, H. S. (1956). A technique for assessing homosexual tendencies. *Journal of Personality*, 24, 436-448.

## Appendix

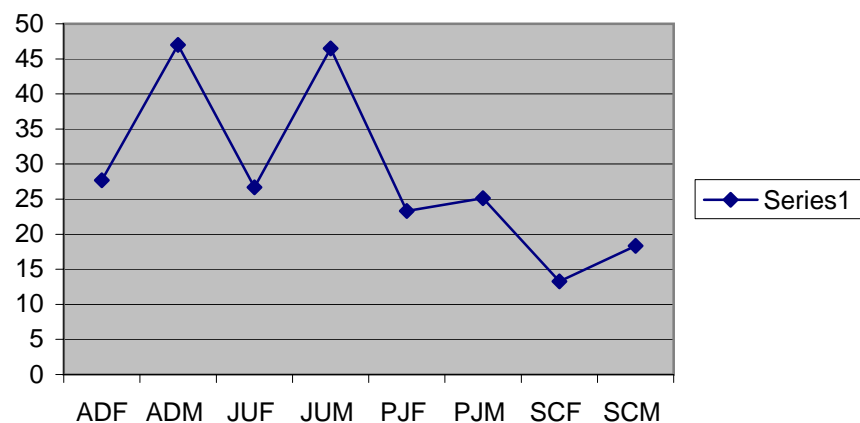




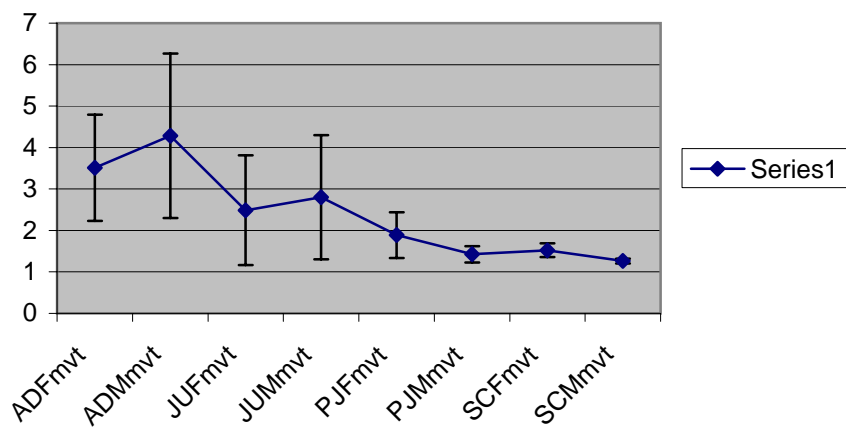
### 2003 Ipsative



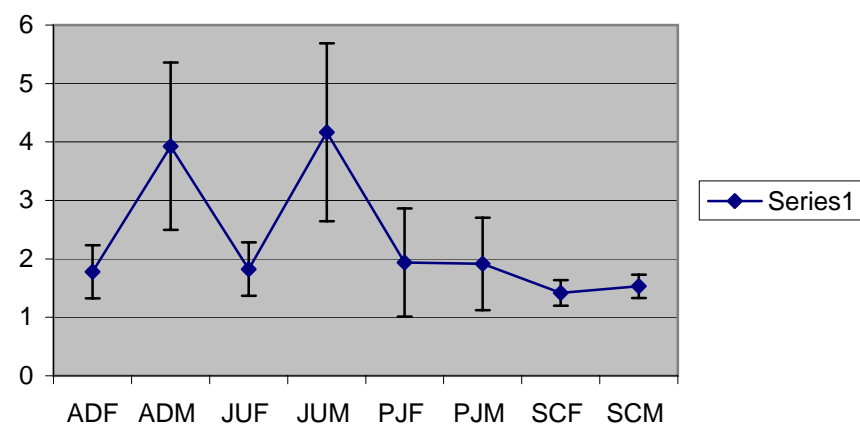
### 2004 Ipsative



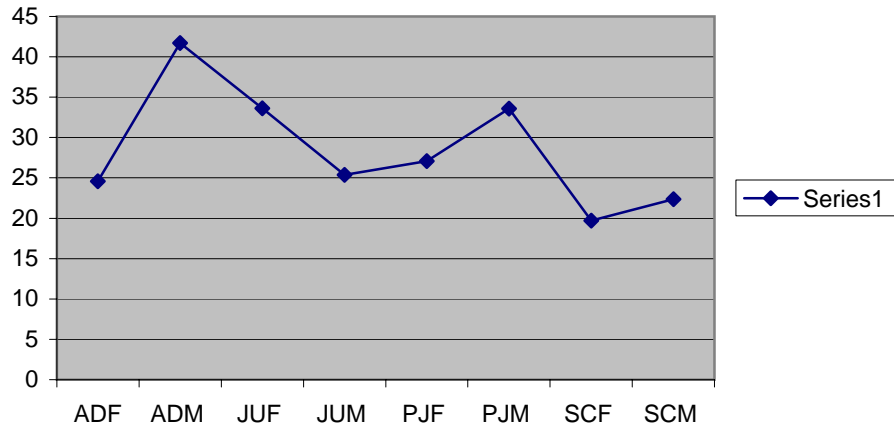
### 2003 VT



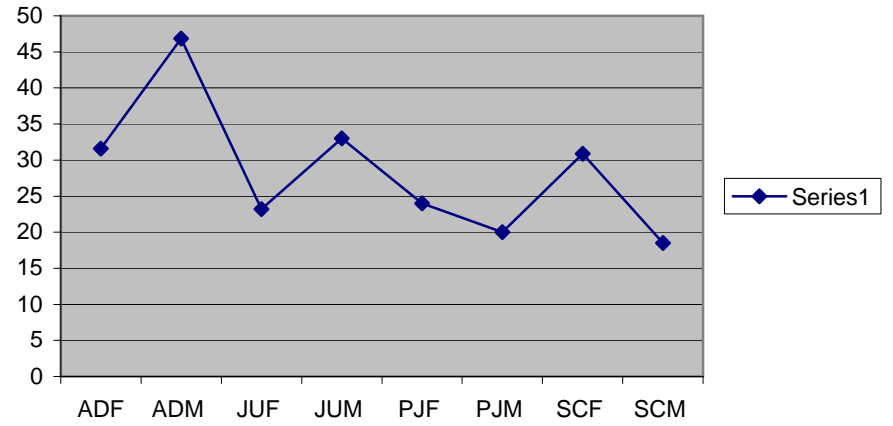
### 2004 VT



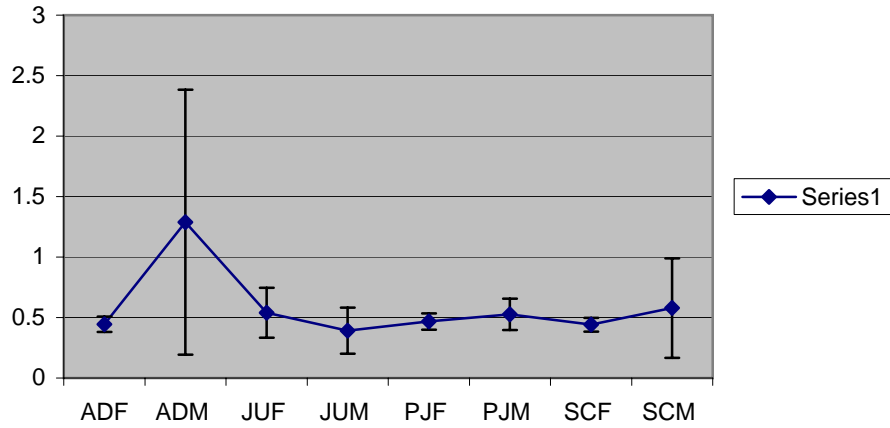
2005 Ipsative



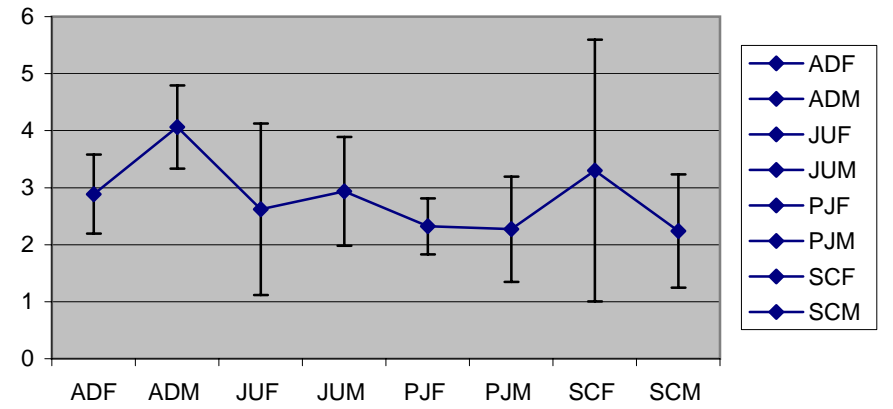
2006 Ipsative



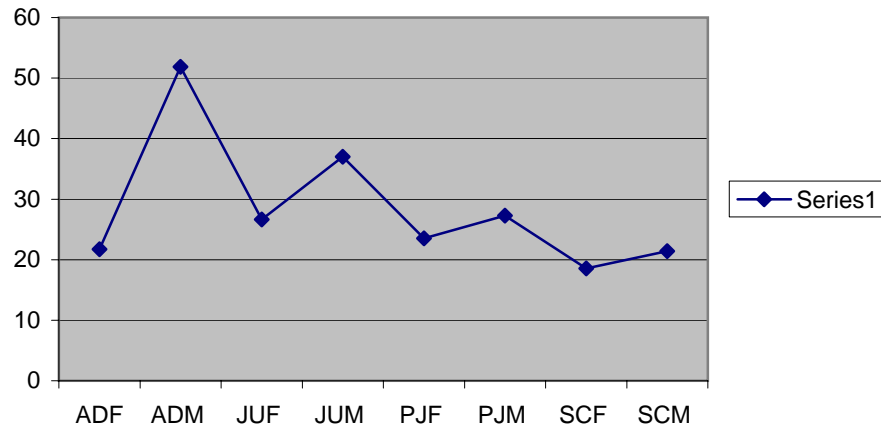
2005 Raw Vt



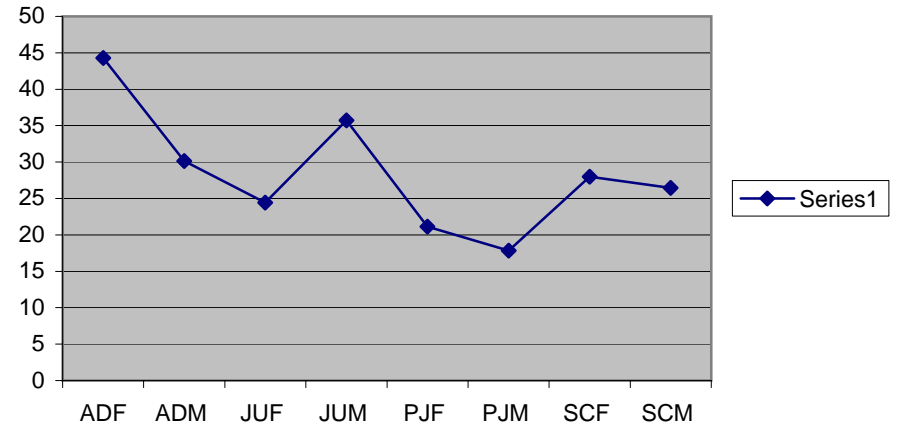
2006 Raw Vt



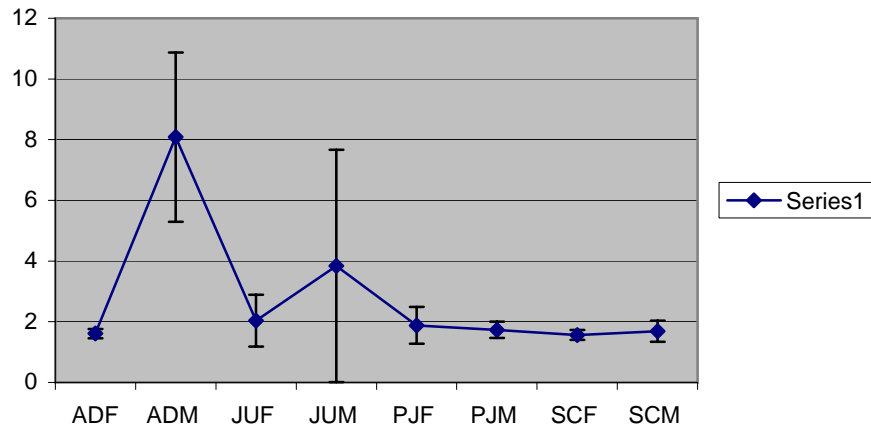
2007 Ipsative



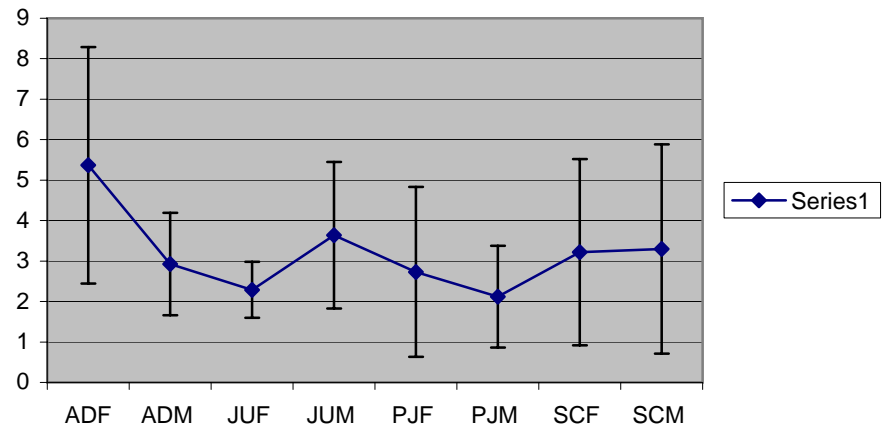
2009 Ipsative



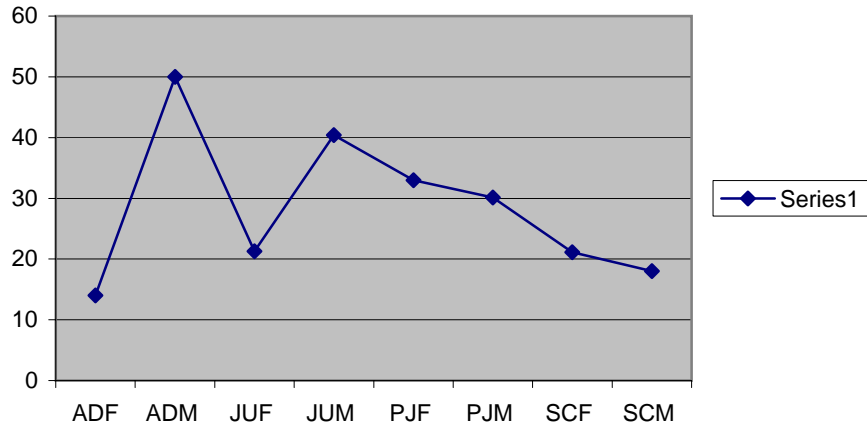
2007 Raw Vt



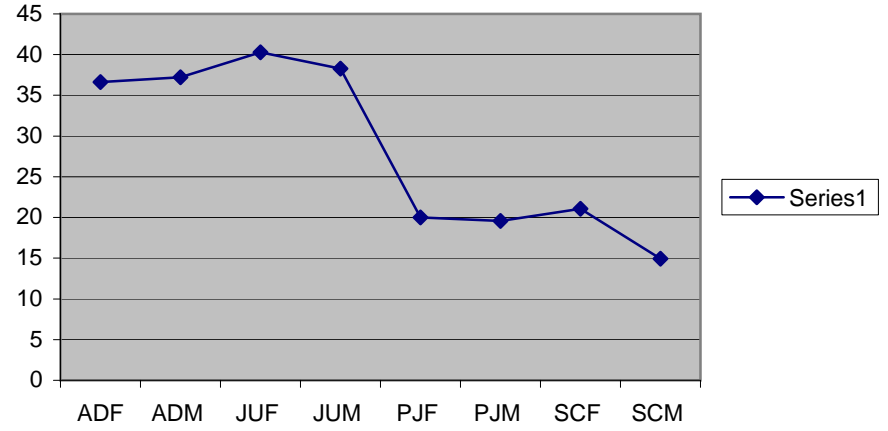
2009 Raw Vt



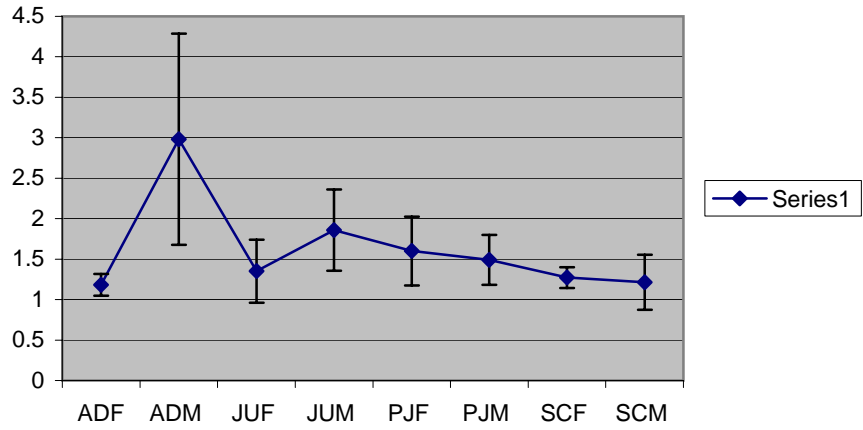
2010 Ipsative



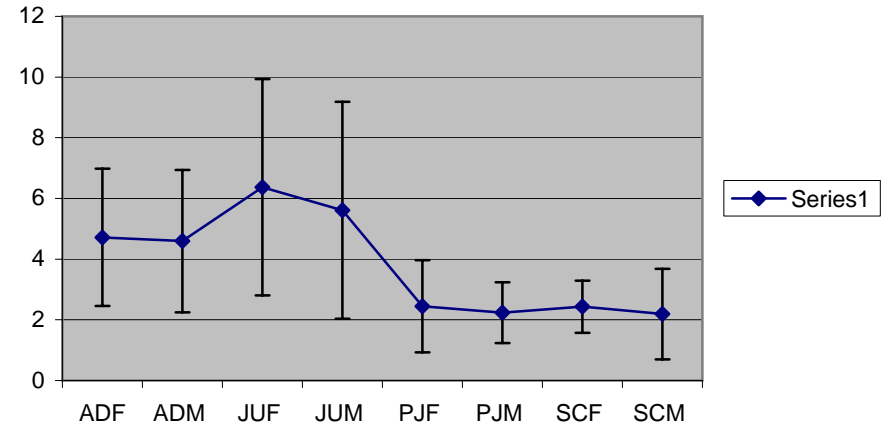
2011 Ipsative



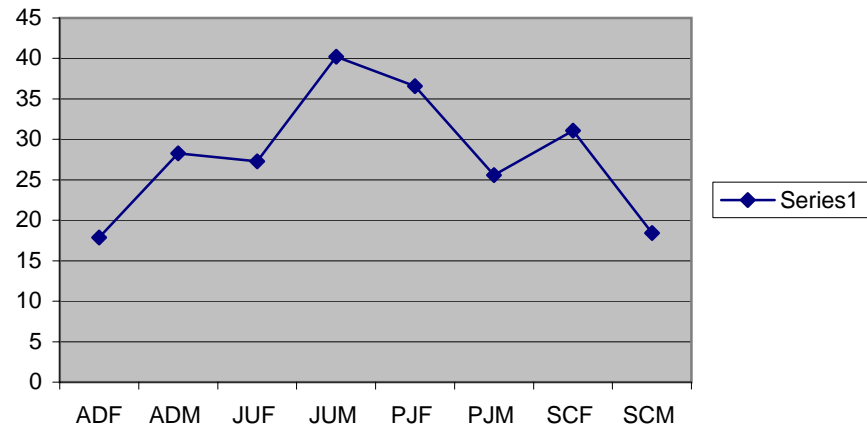
2010 Raw Vt



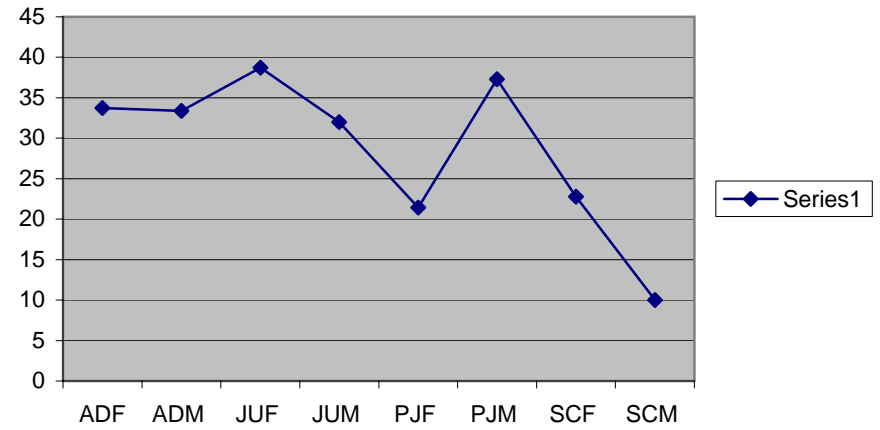
2011 Raw Vt



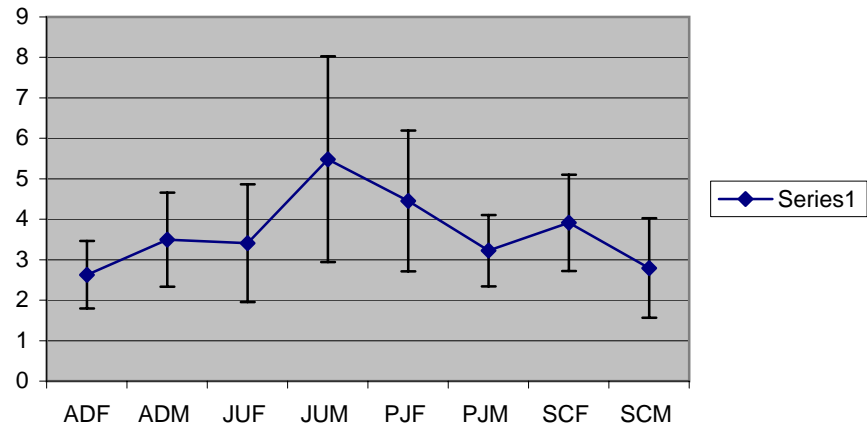
2012 Ipsative



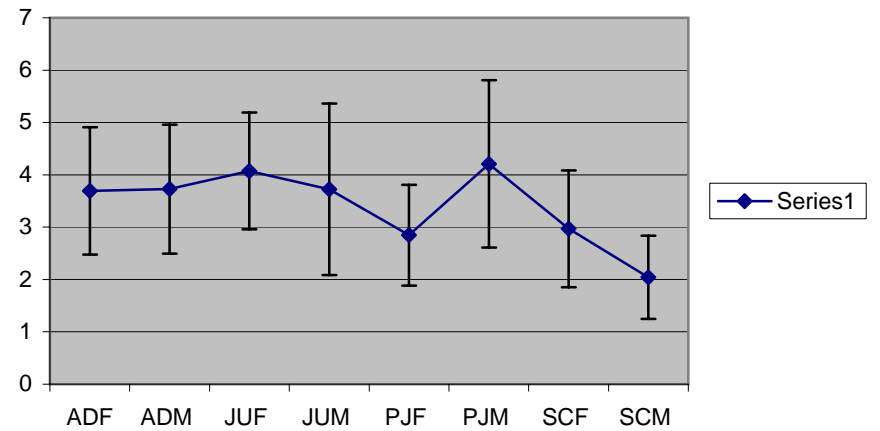
2013 Ipsative



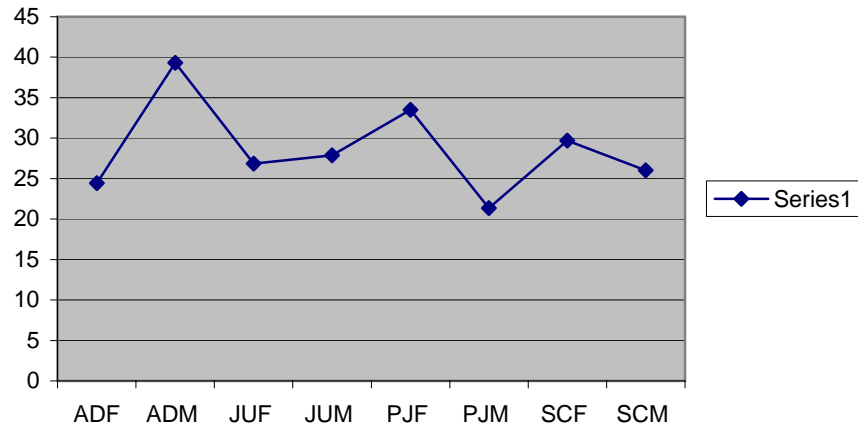
2012 Raw Vt



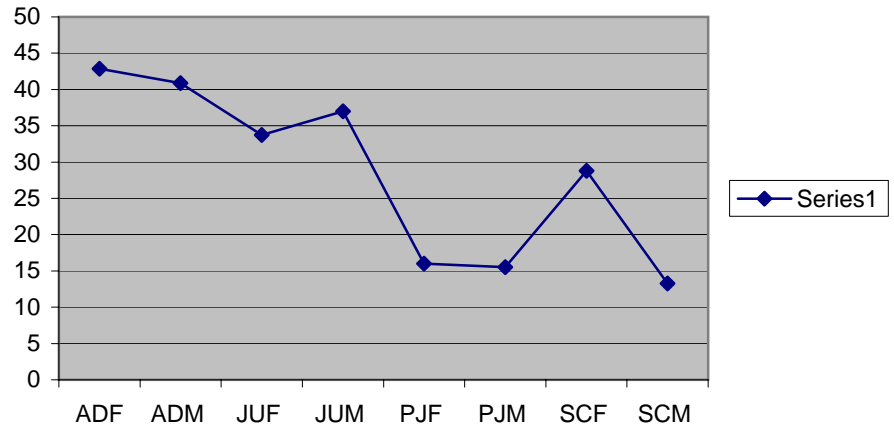
2013 Raw Vt



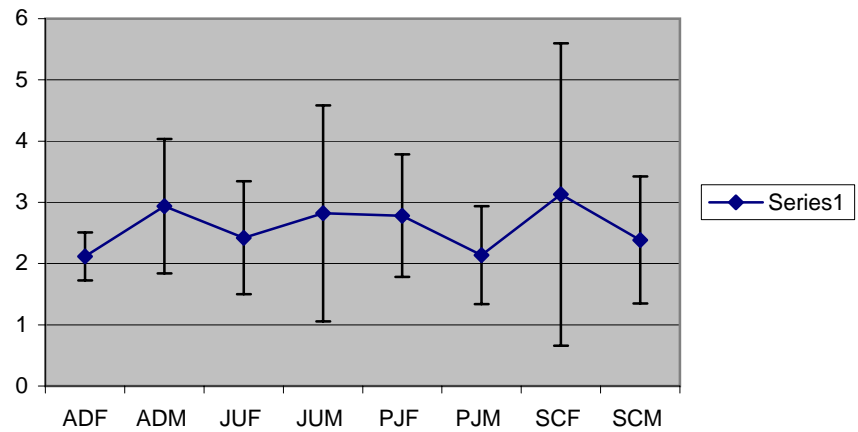
2014 Ipsative



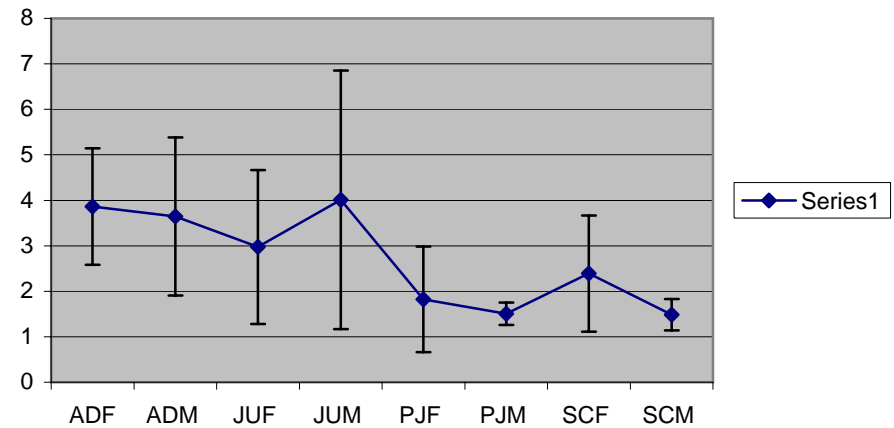
2015 Ipsative



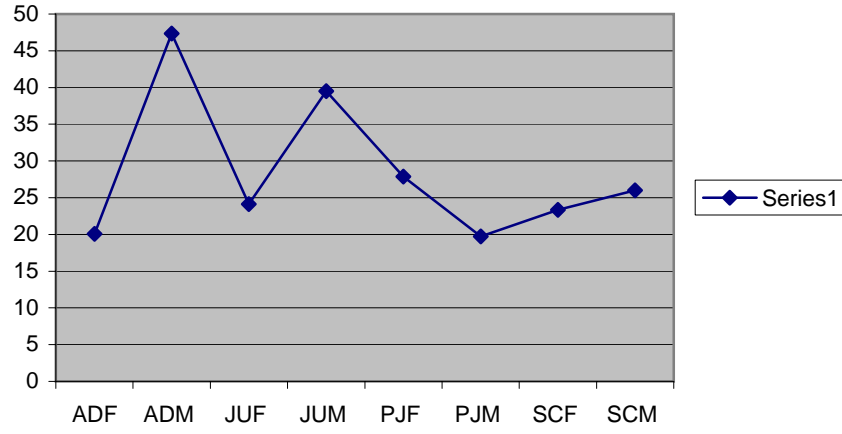
2014 Raw Vt



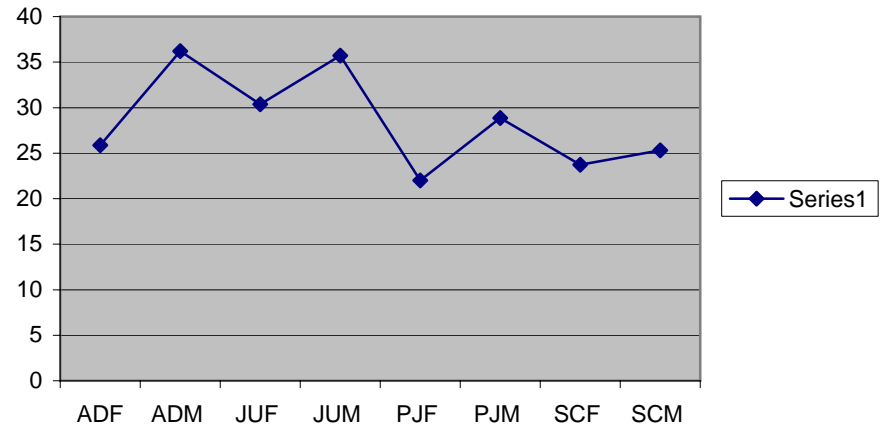
2015 Raw Vt



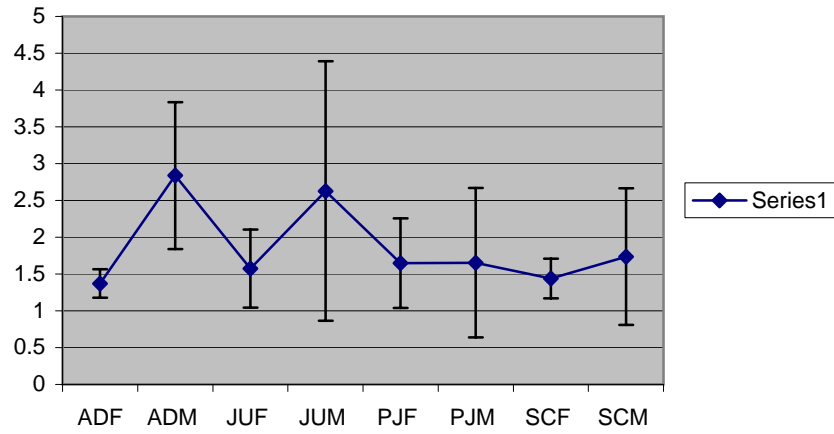
2016 Ipsative



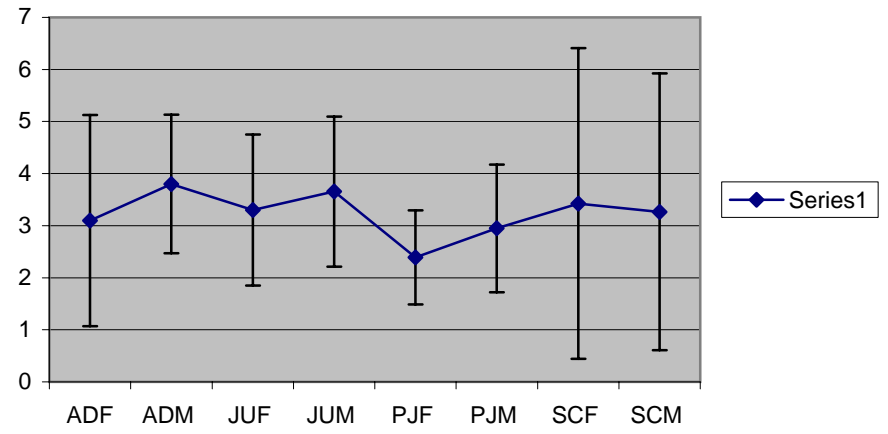
2017 Ipsative



2016 Raw Vt

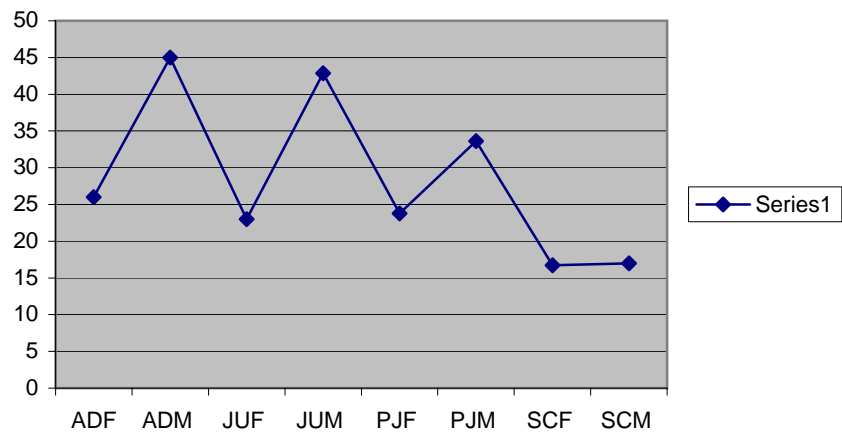


2017 Raw Vt

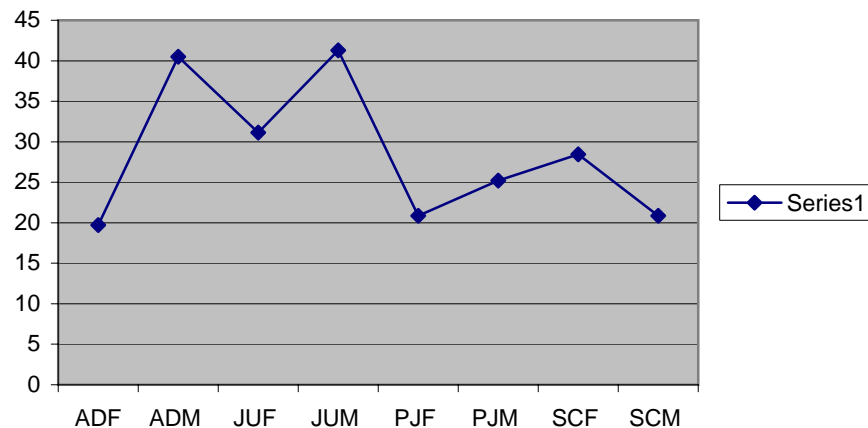




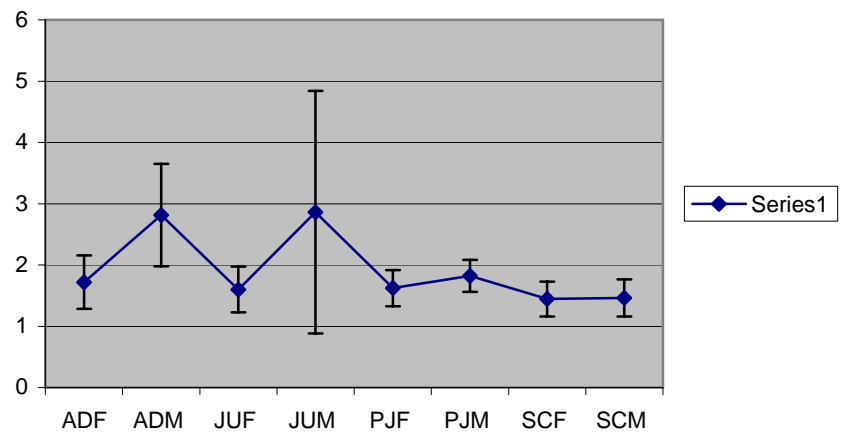
2018 Ipsative



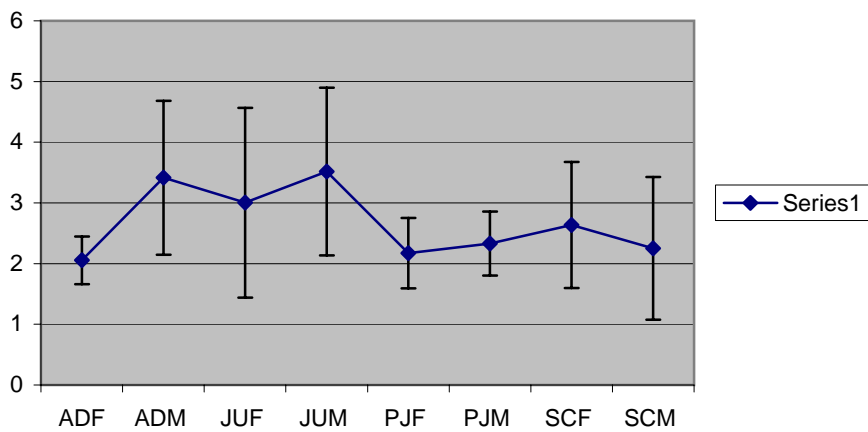
2019 Ipsative



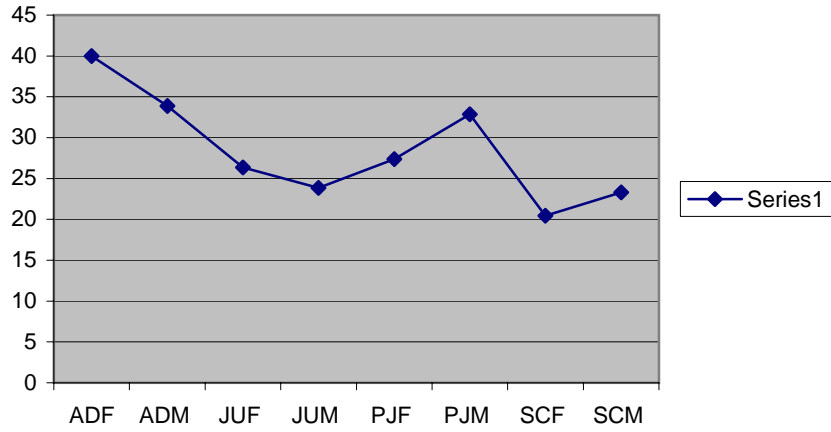
2018 Raw Vt



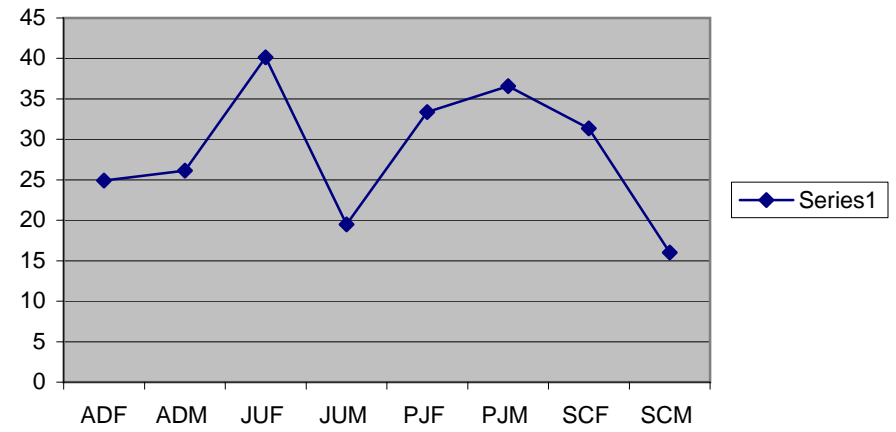
2019 Raw Vt



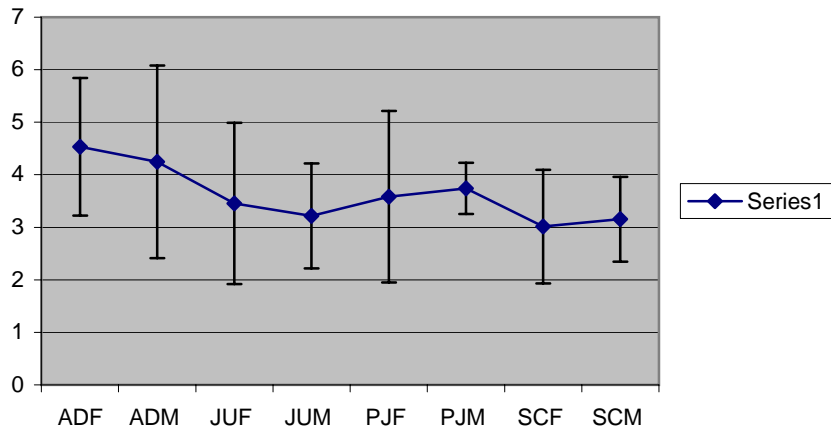
2020 Ipsative



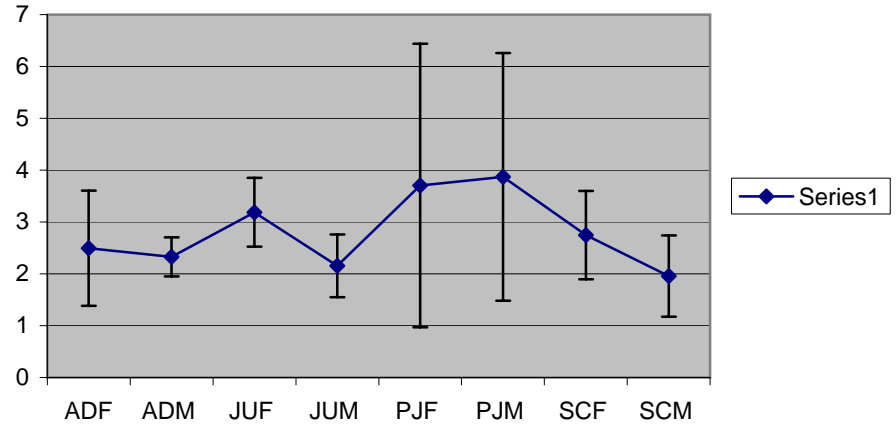
2021 Ipsative



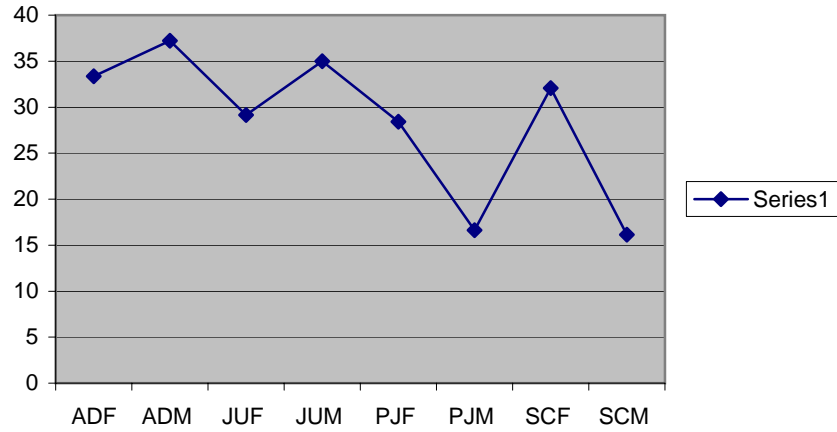
2020 Raw Vt



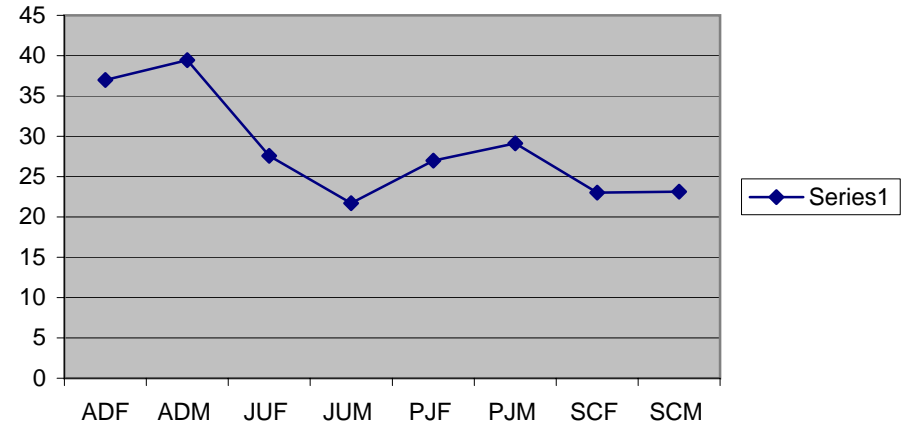
2021 Raw Vt



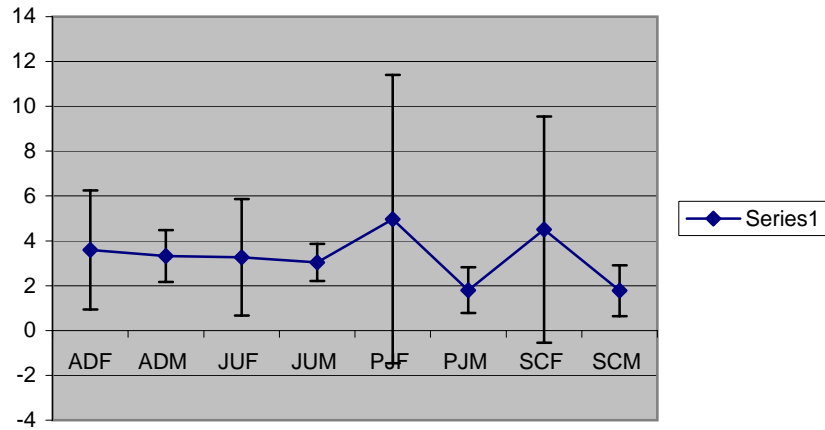
2022 Ipsative



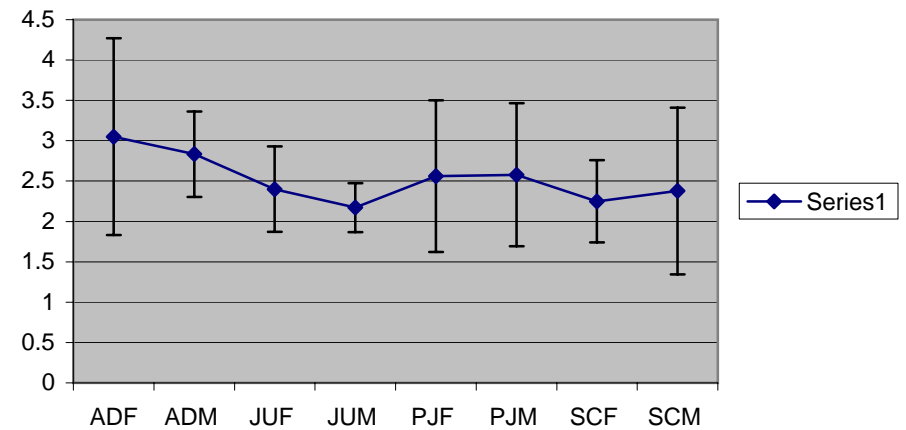
2023 Ipsative



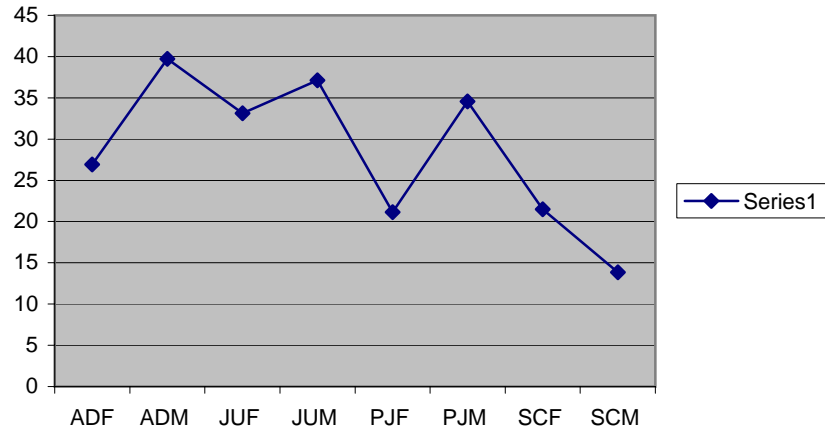
2022 Raw Vt



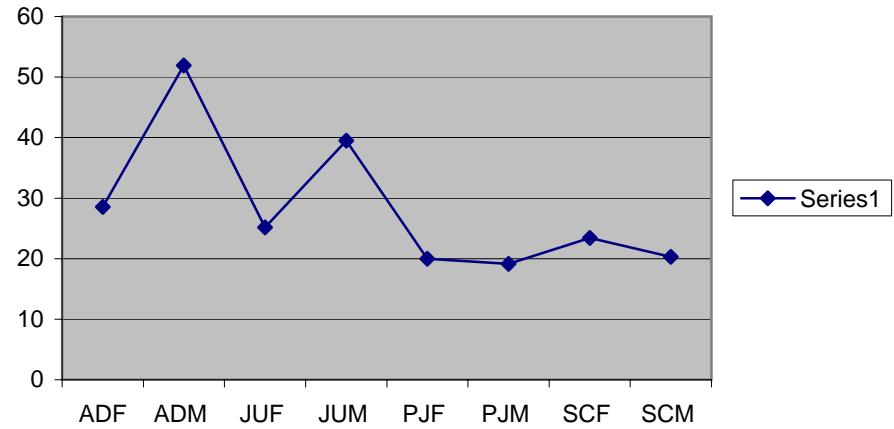
2023 Raw Vt



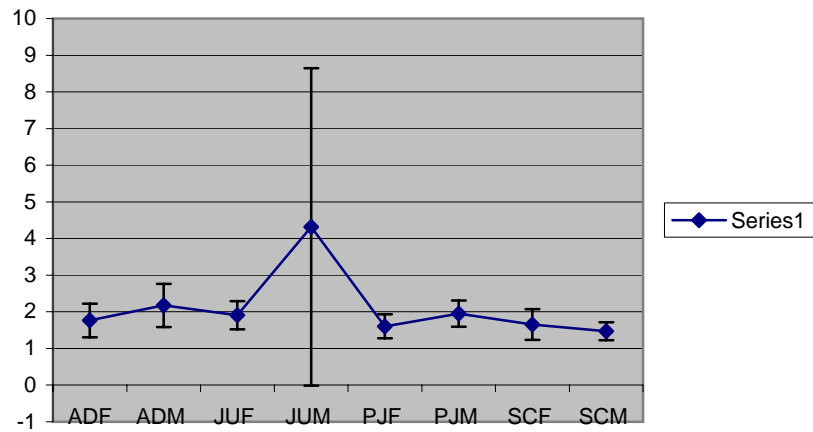
2024 Ipsative



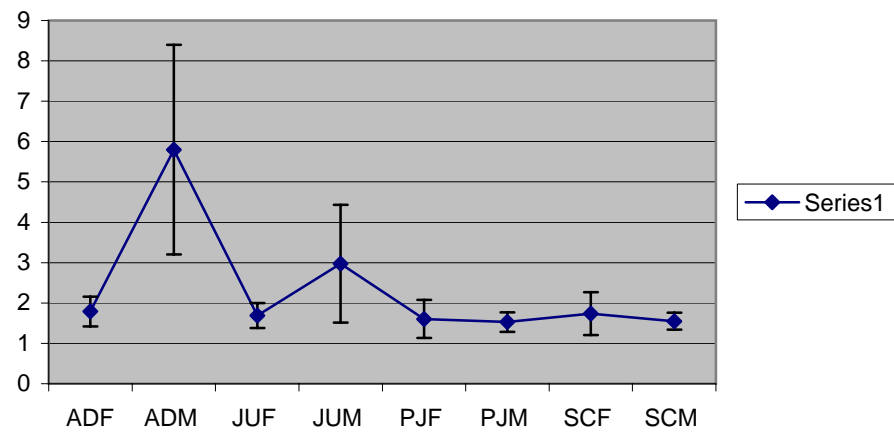
2025 Ipsative



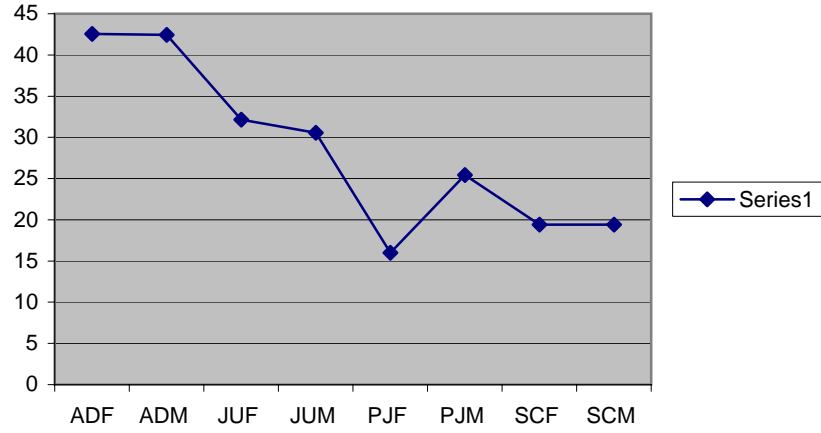
2024 Raw Vt



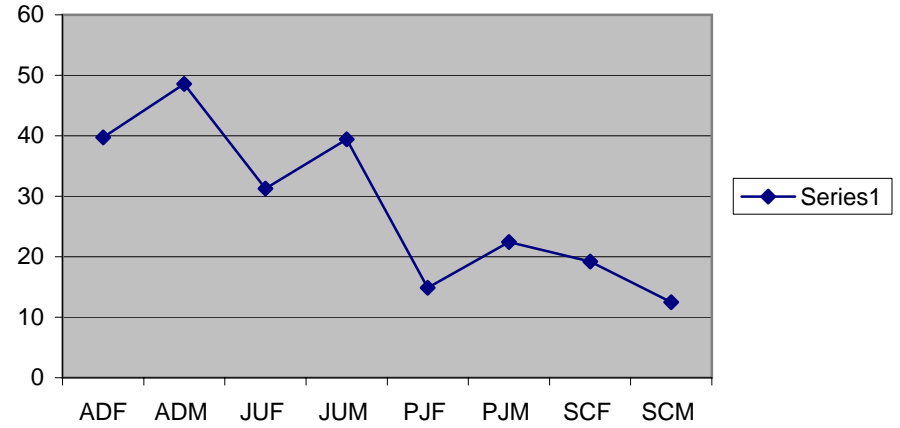
2025 Raw Vt



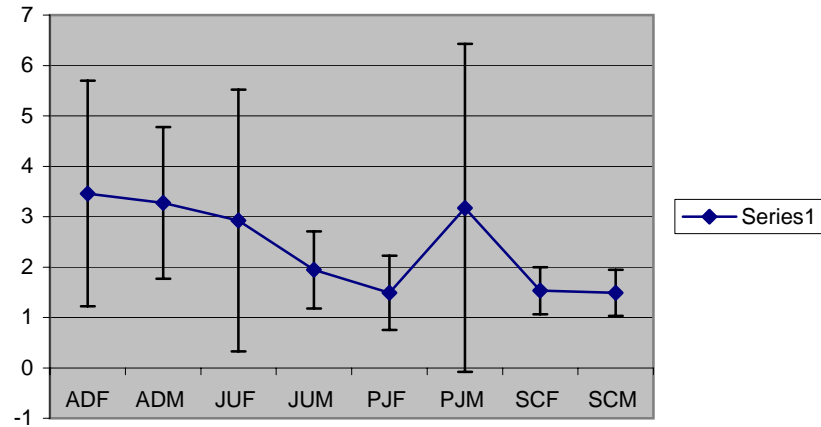
2026 Ipsative



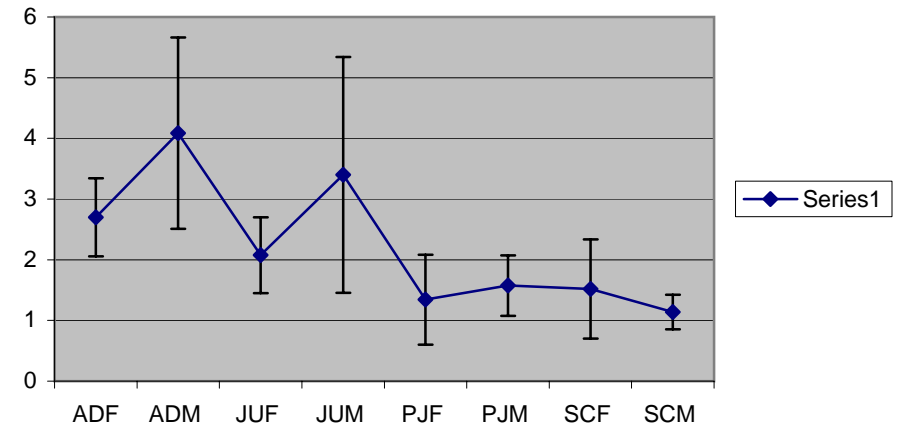
2027 Ipsative



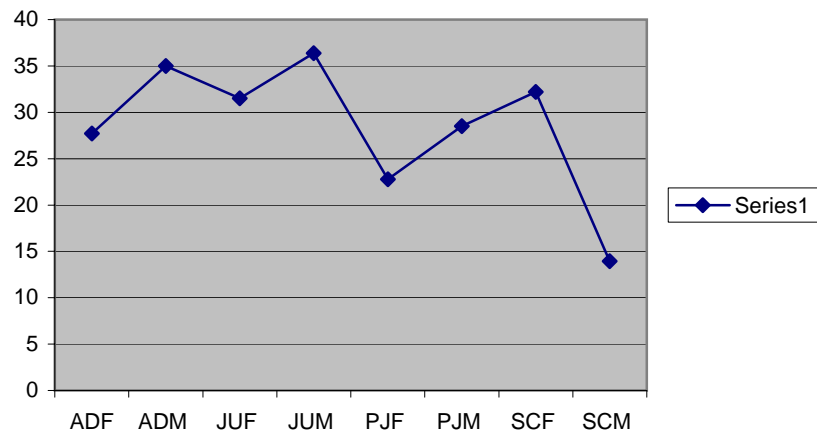
2026 Raw Vt



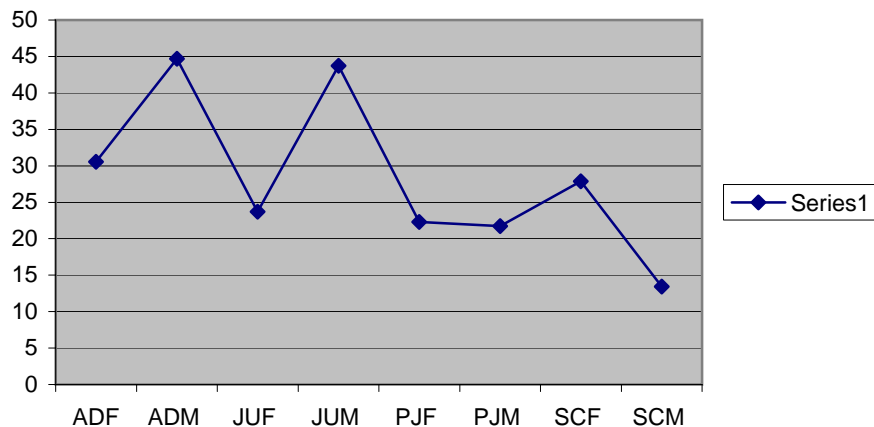
2027 Raw Vt



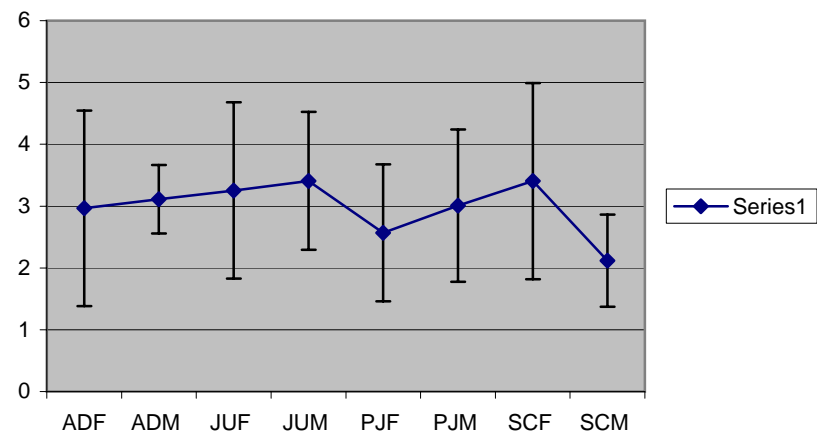
2028 Ipsative



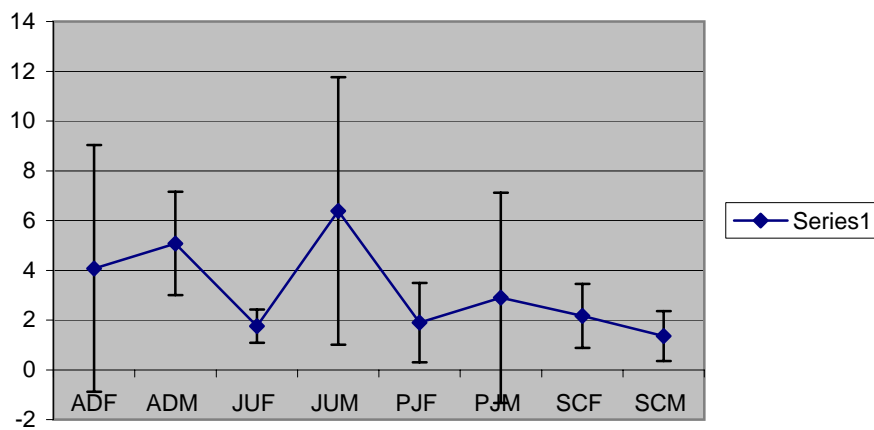
2029 Ipsative



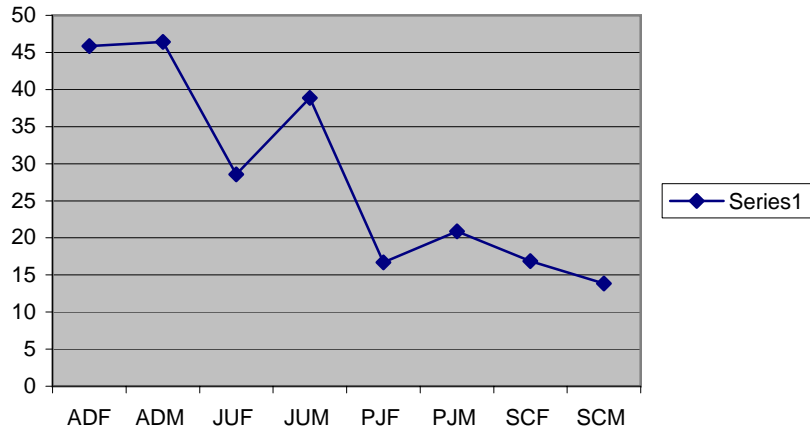
2028 Raw Vt



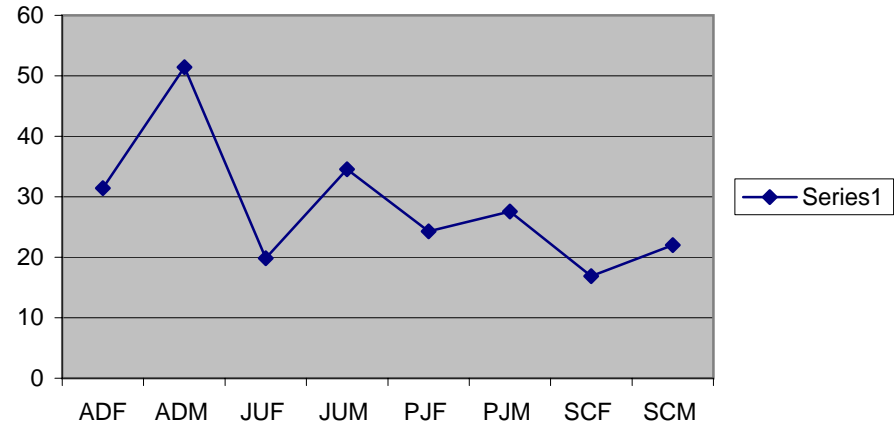
2029 Raw Vt



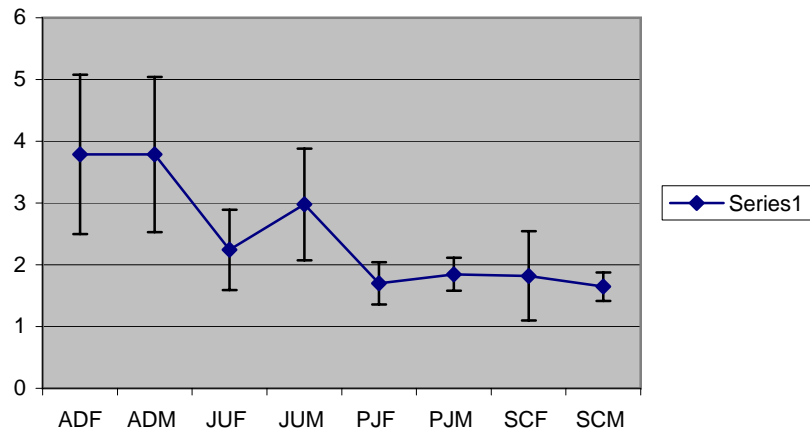
2031 Ipsative



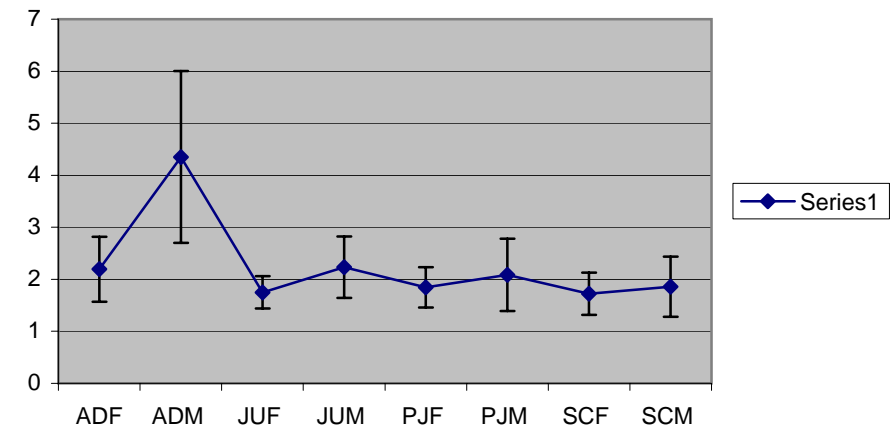
2032 Ipsative



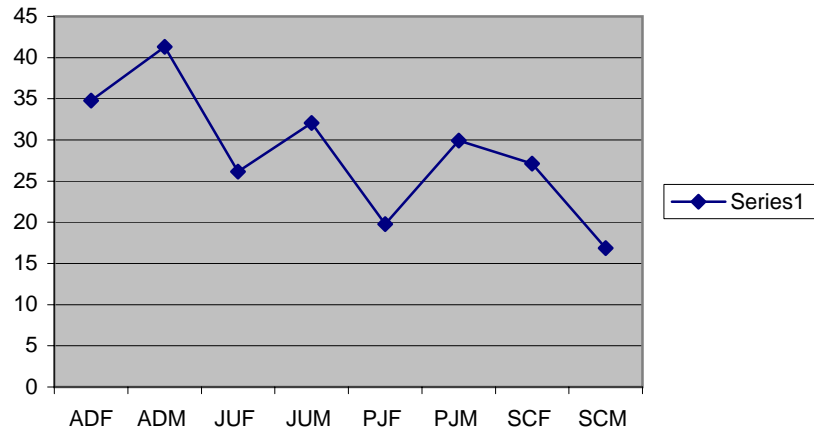
2031 Raw Vt



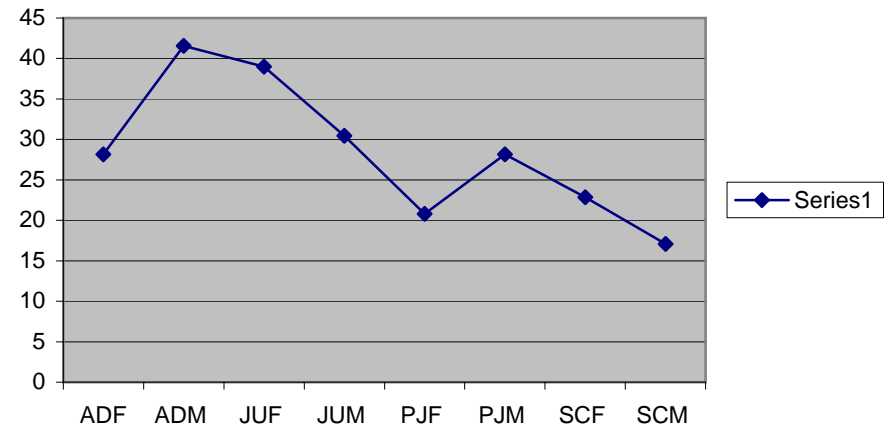
2032 Raw Vt



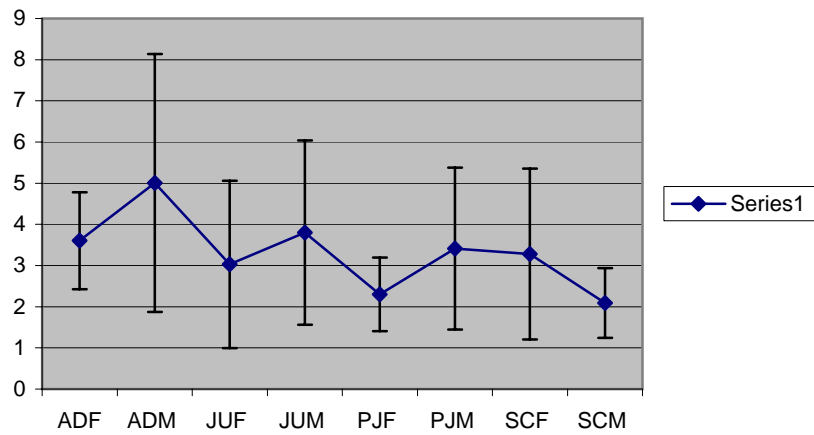
2033 Ipsative



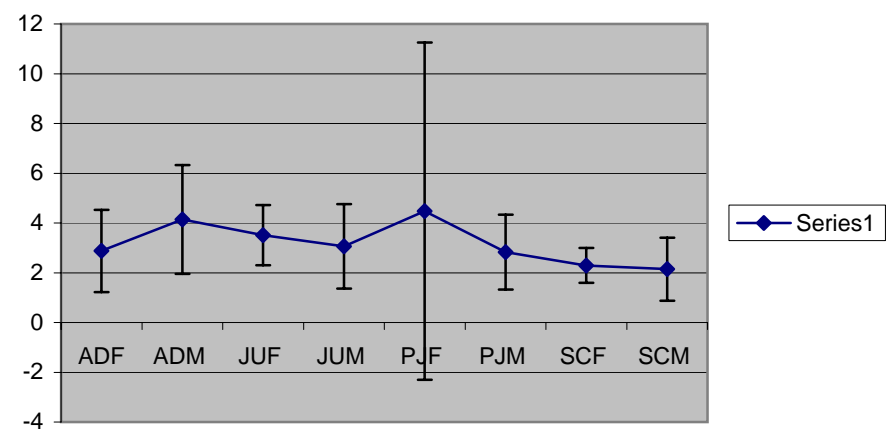
2034 Ipsative



2033 Raw Vt

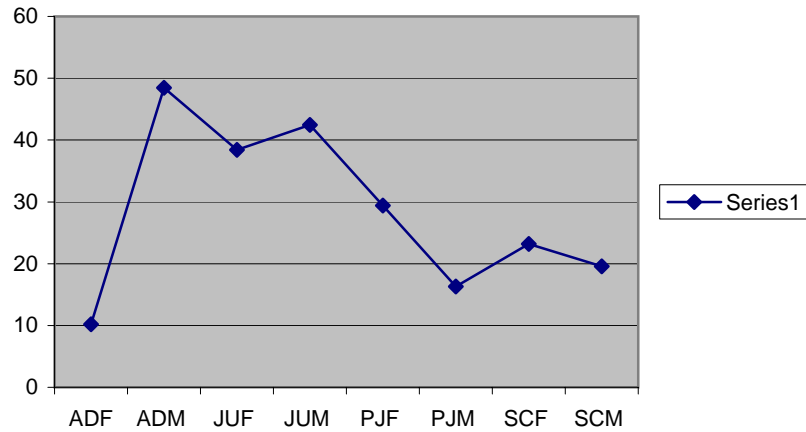


2034 Raw Vt

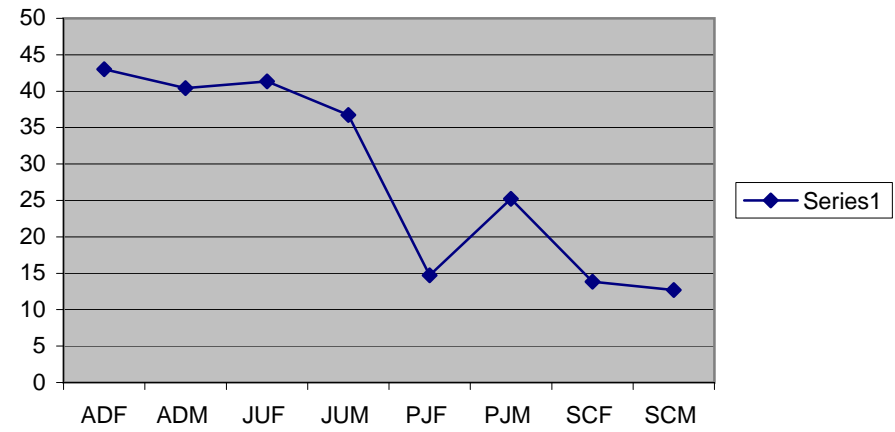




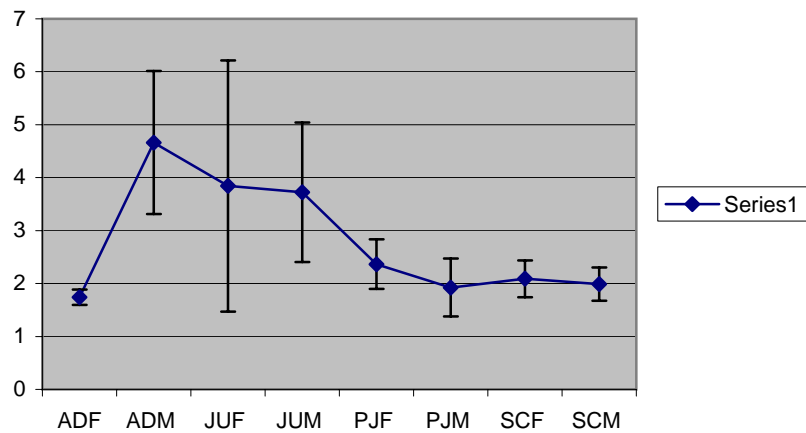
2035 Ipsative



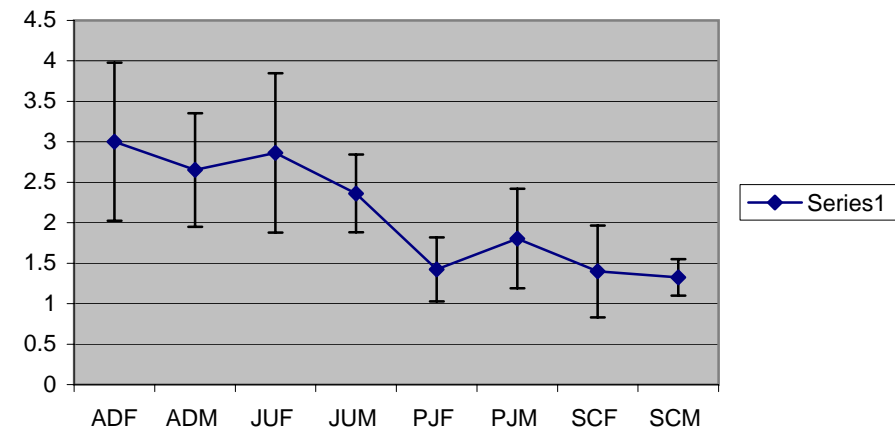
2036 Ipsative



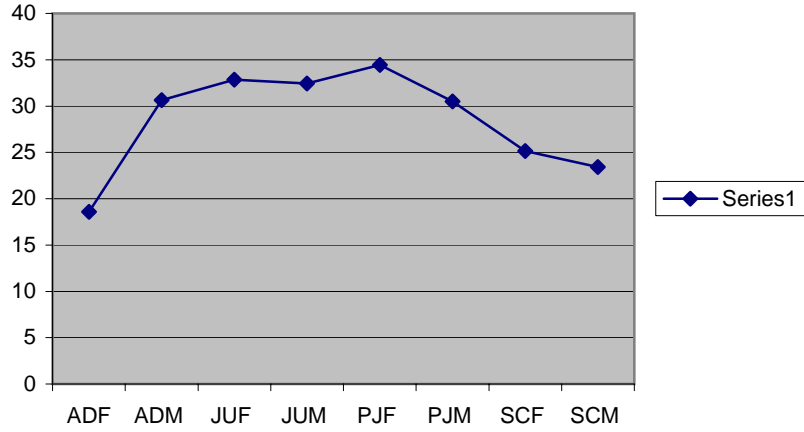
2035 Raw Vt



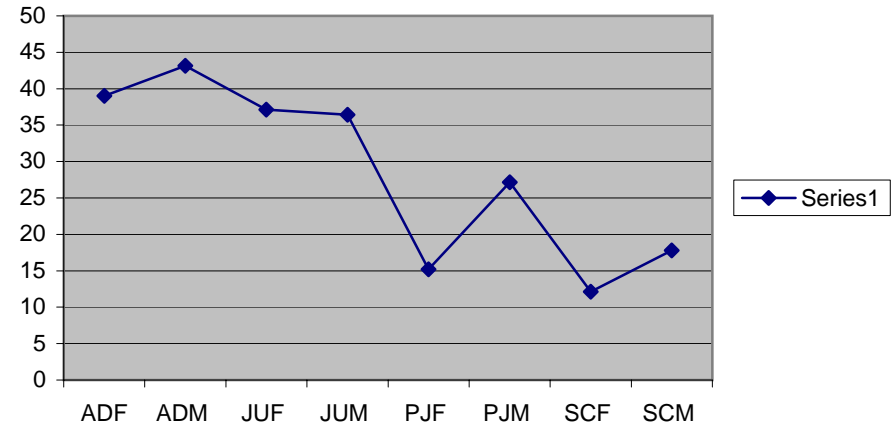
2036 Raw Vt



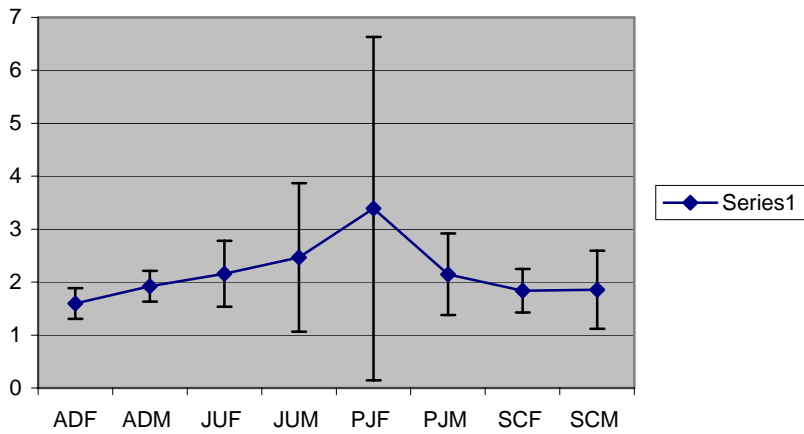
2037 Ipsative



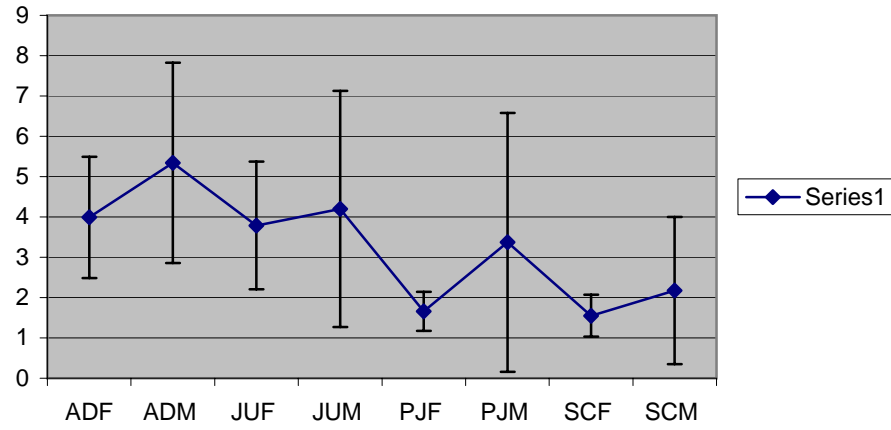
2038 Ipsative



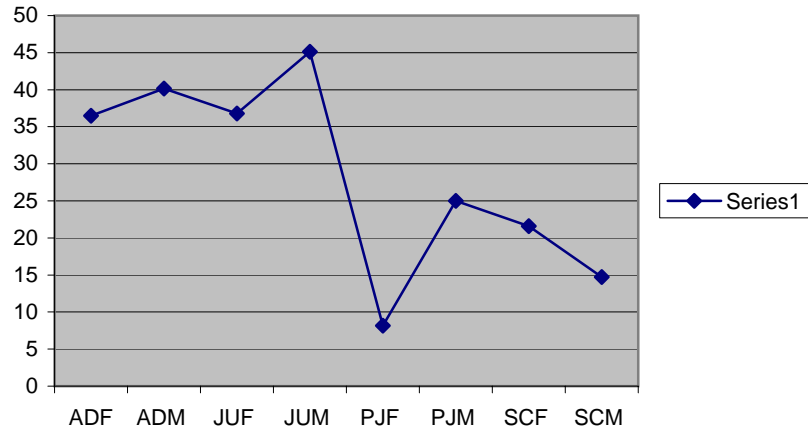
2037 Raw Vt



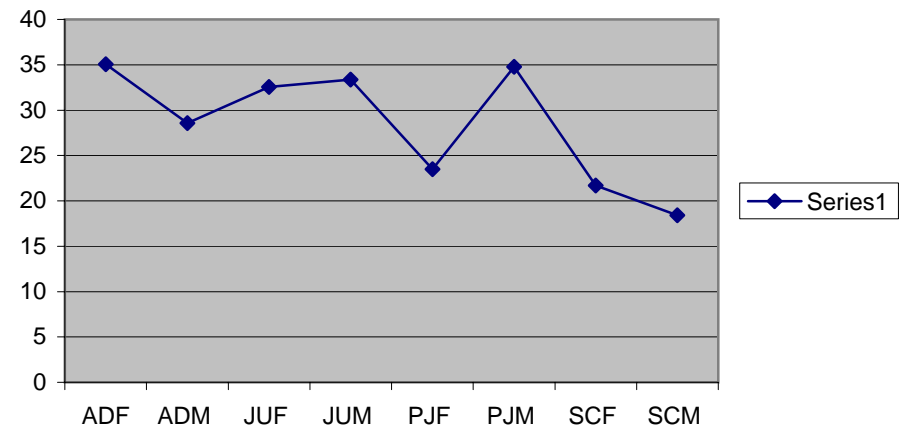
2038 Raw Vt



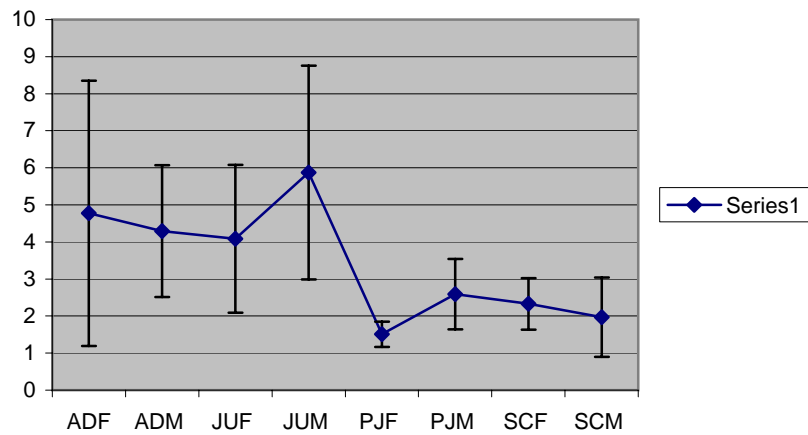
2040 Ipsative



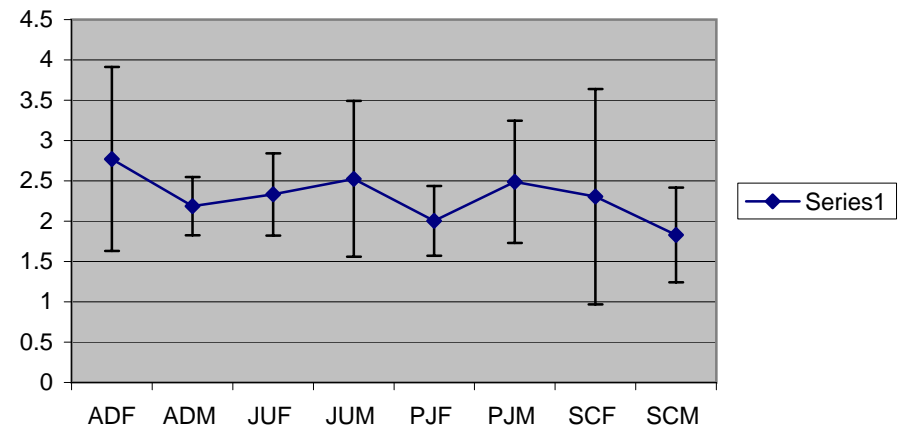
2041 Ipsative



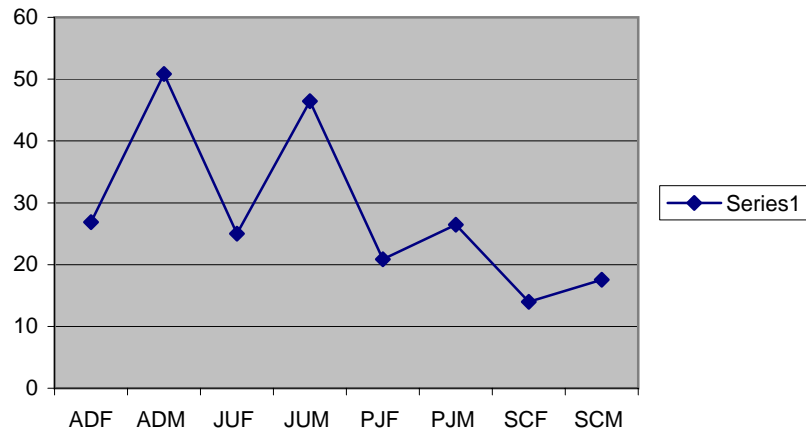
2040 Raw Vt



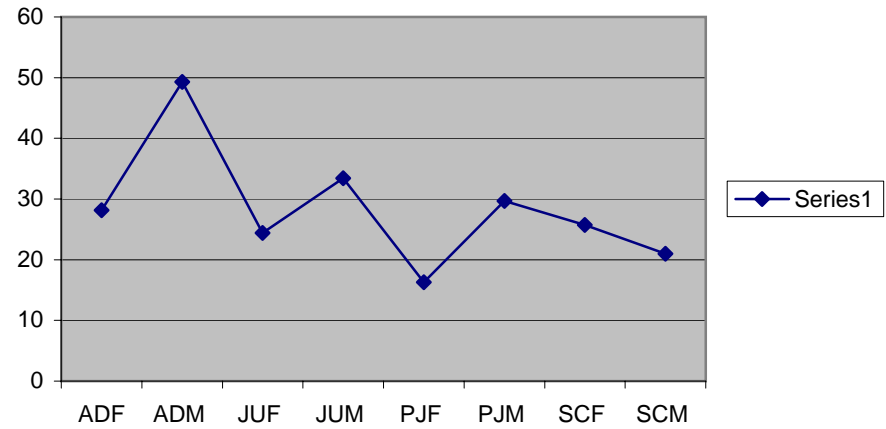
2041 Raw Vt



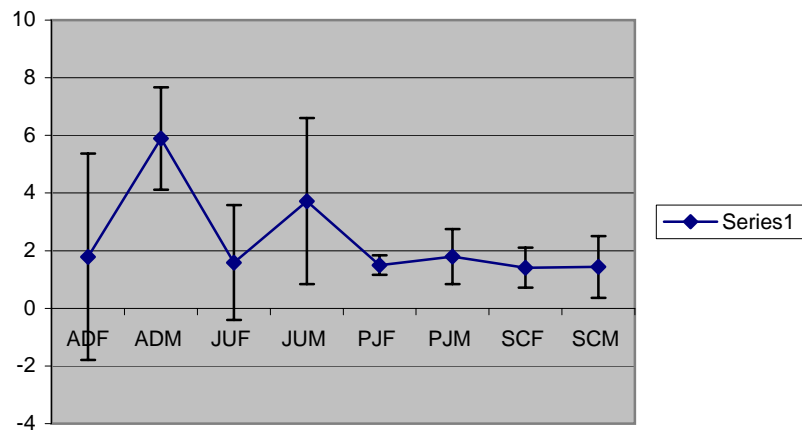
2042 Ipsative



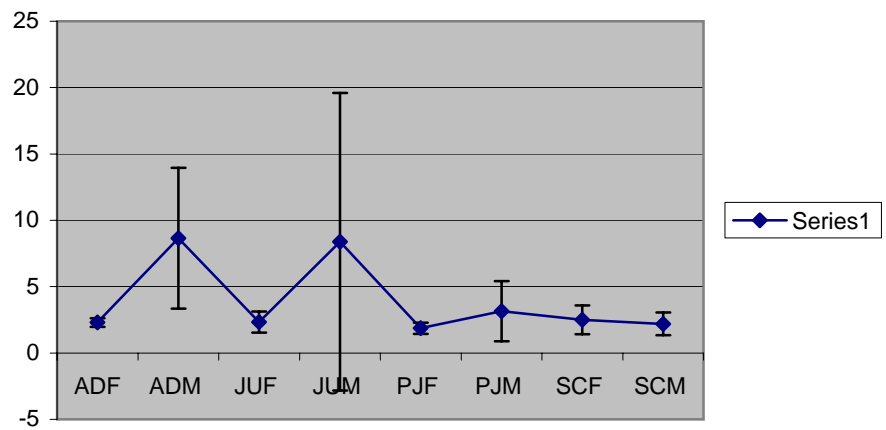
2043 Ipsative



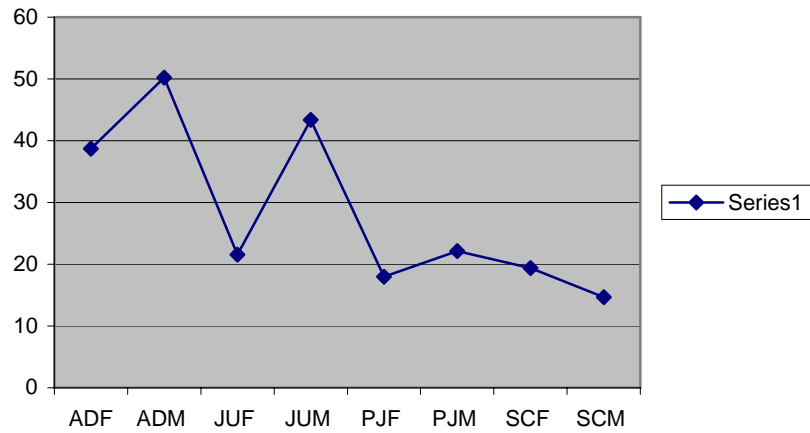
2042 Raw Vt



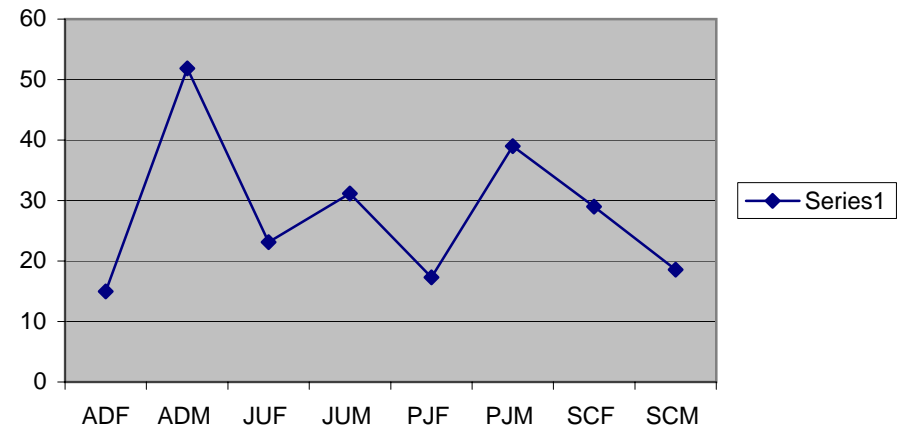
2043 Raw Vt



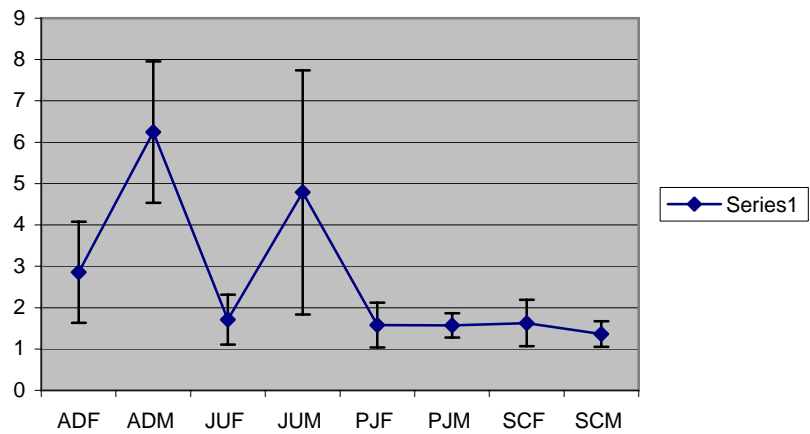
2044 Ipsative



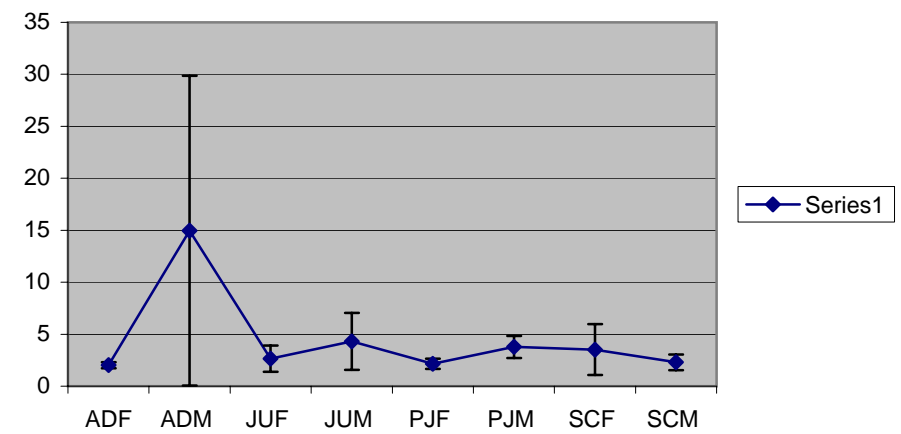
2046 Ipsative



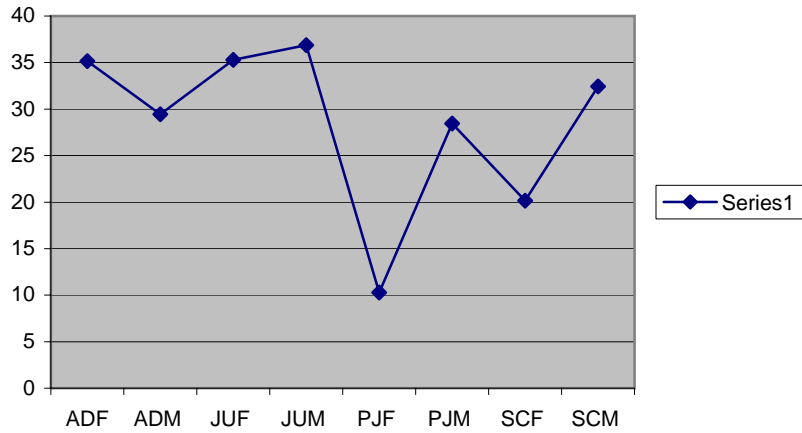
2044 Raw Vt



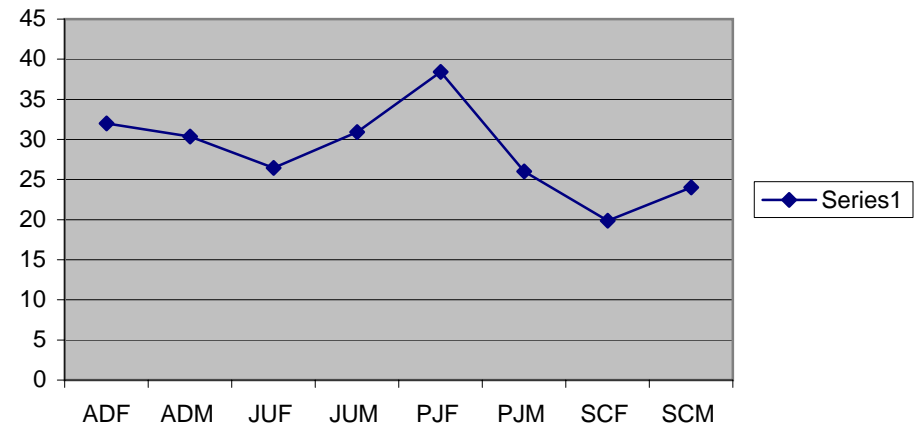
2046 Raw Vt



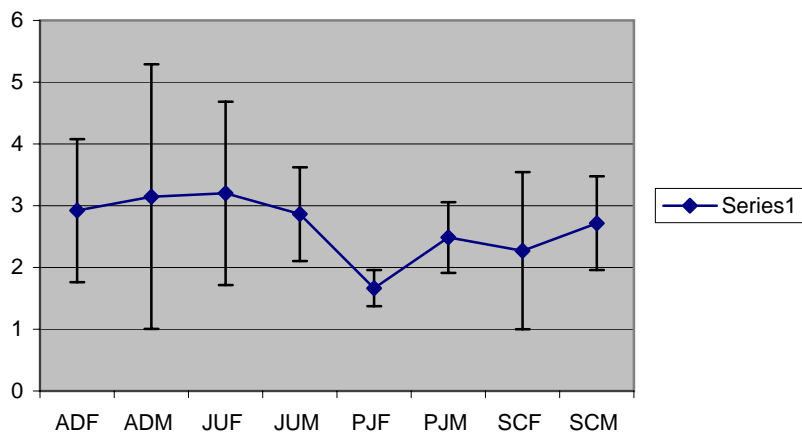
2048 Ipsative



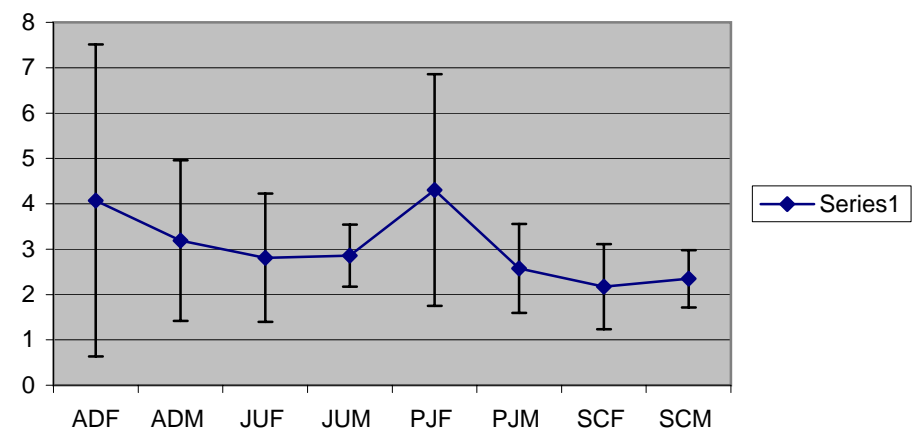
2049 Ipsative



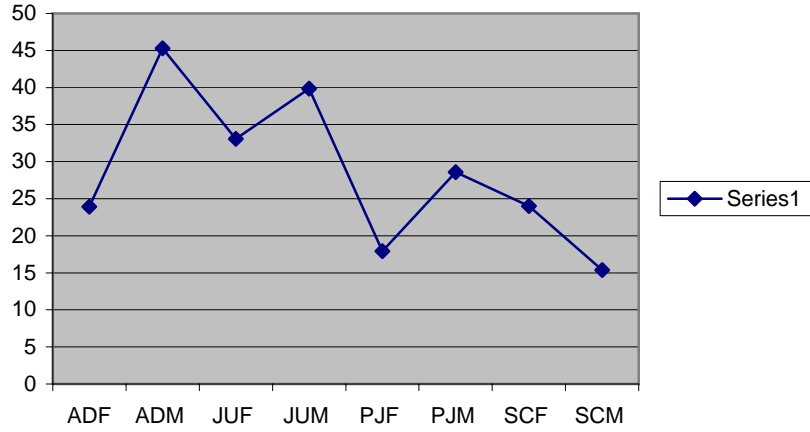
2048 Raw Vt



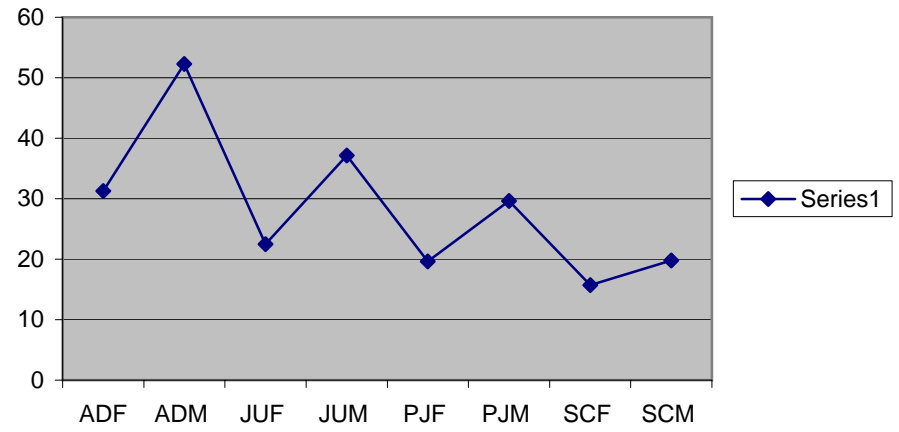
2049 Raw Vt



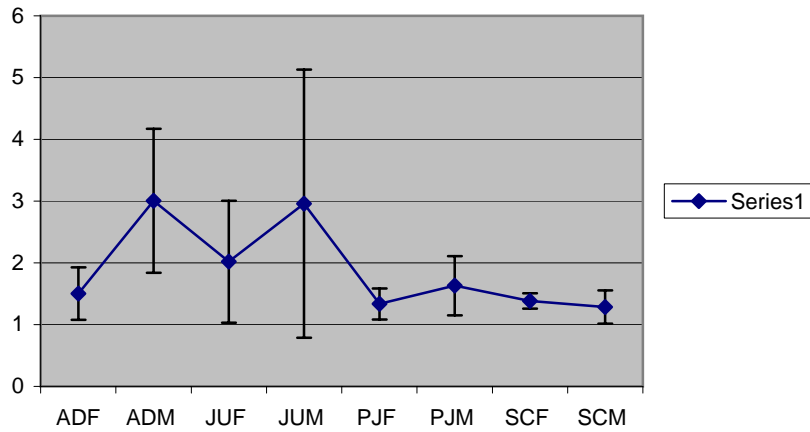
2050 Ipsative



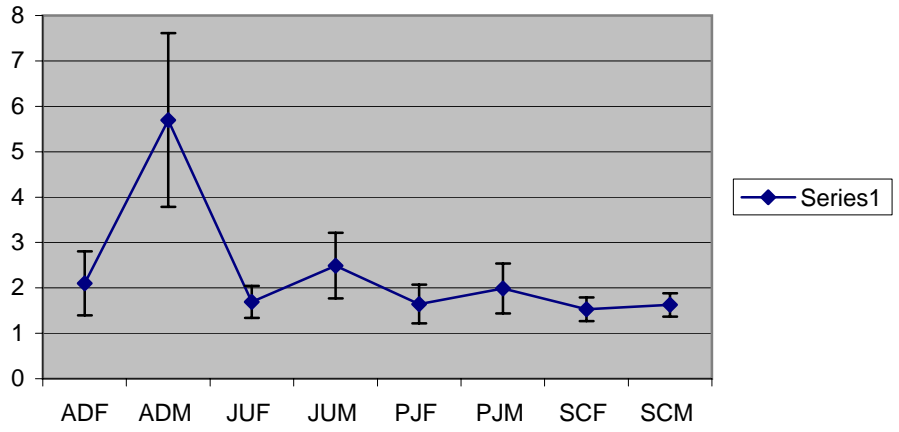
2052 Ipsative



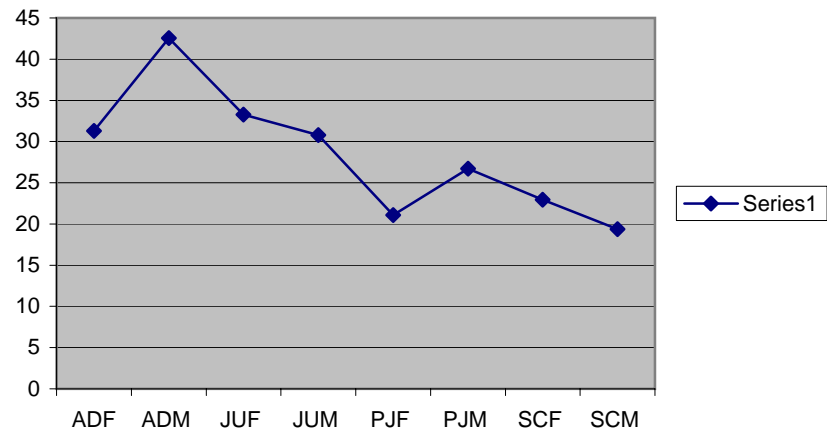
2050 Raw Vt



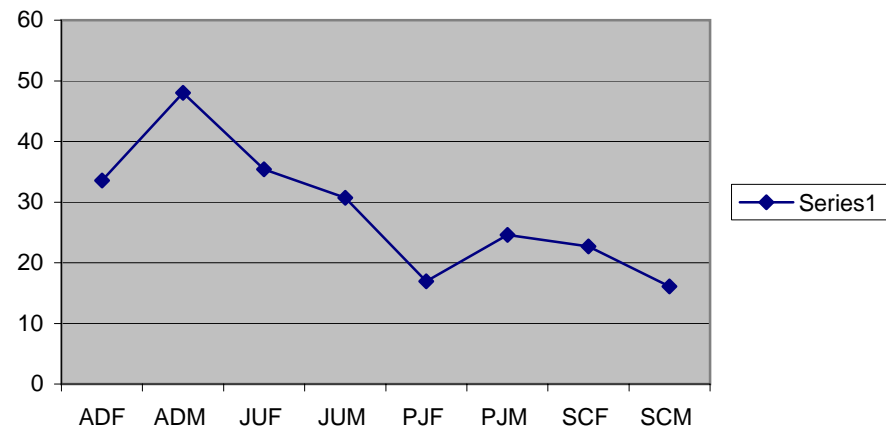
2052 Raw Vt



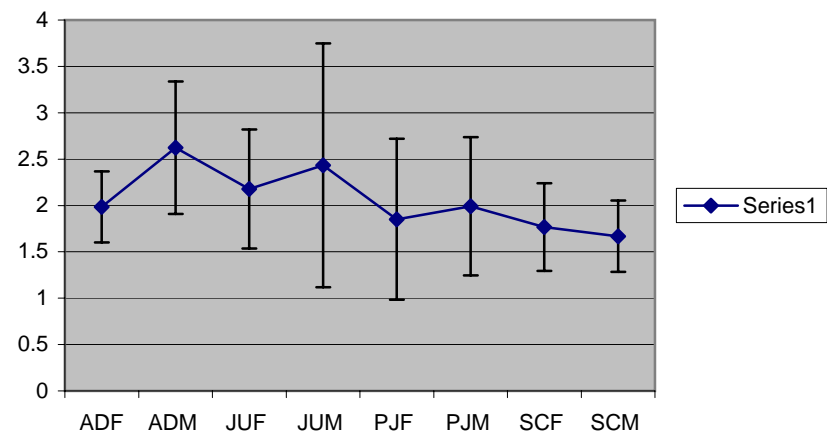
2053 Ipsative



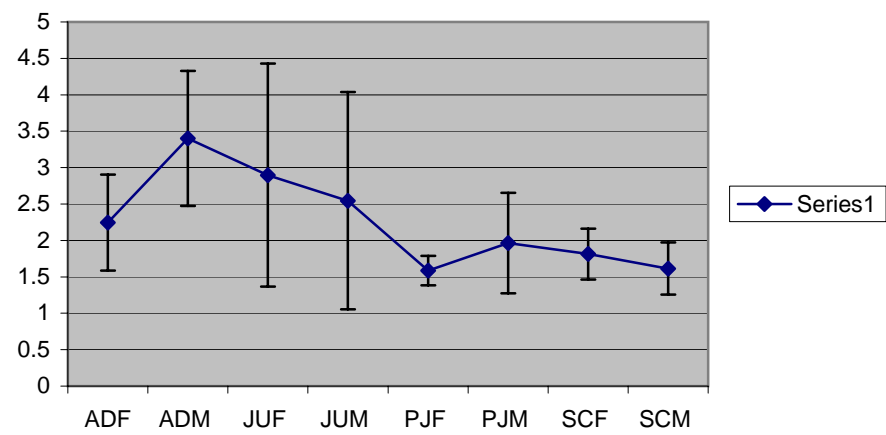
2054 Ipsative



2053 Raw Vt

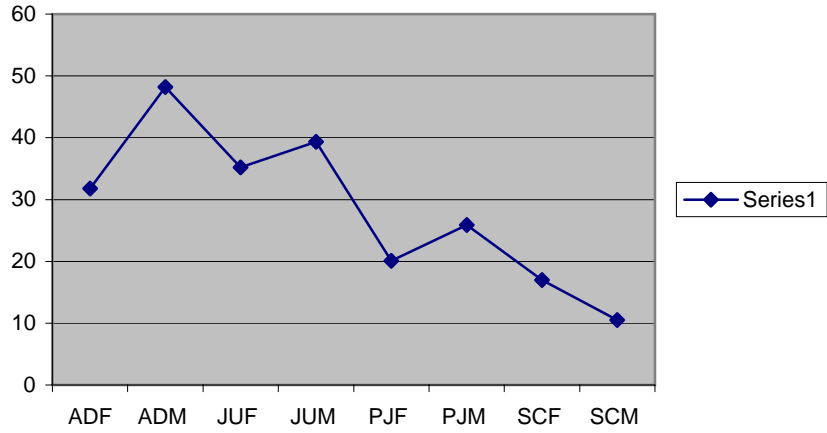


2054 Raw Vt

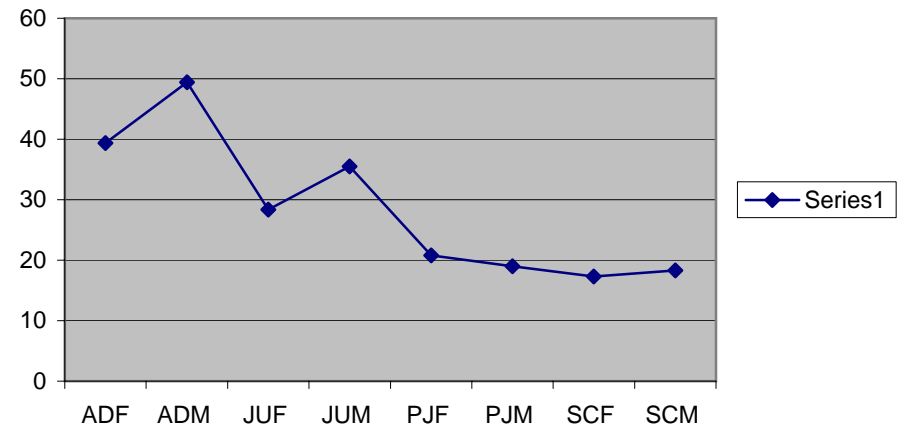




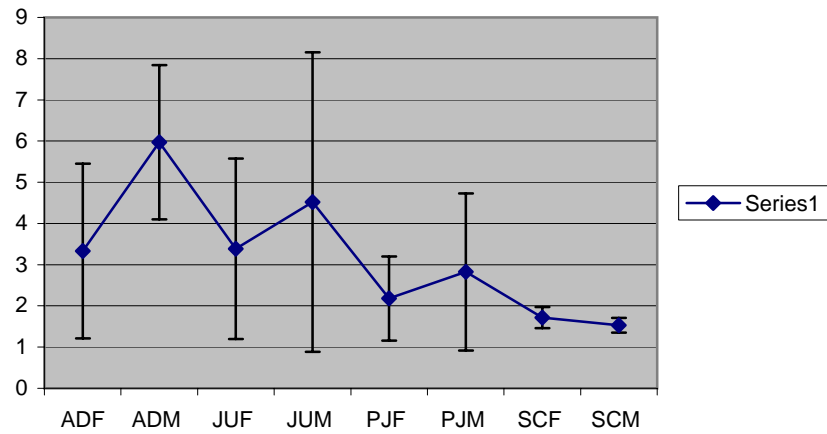
2055 Ipsative



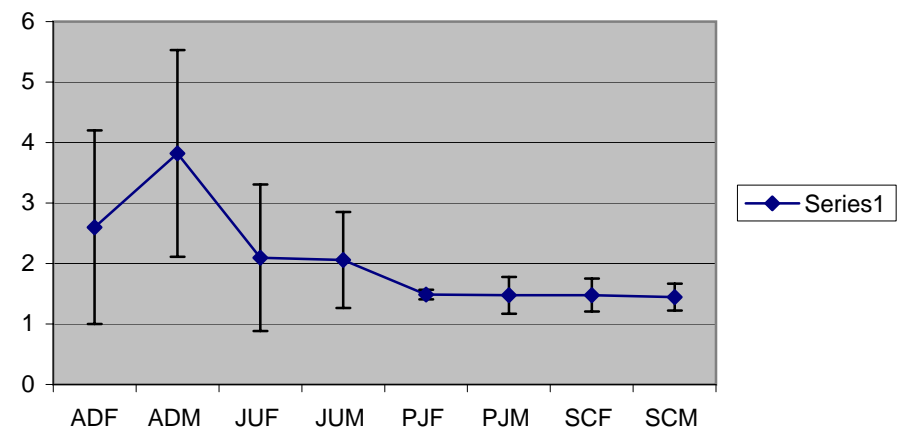
2056 Ipsative



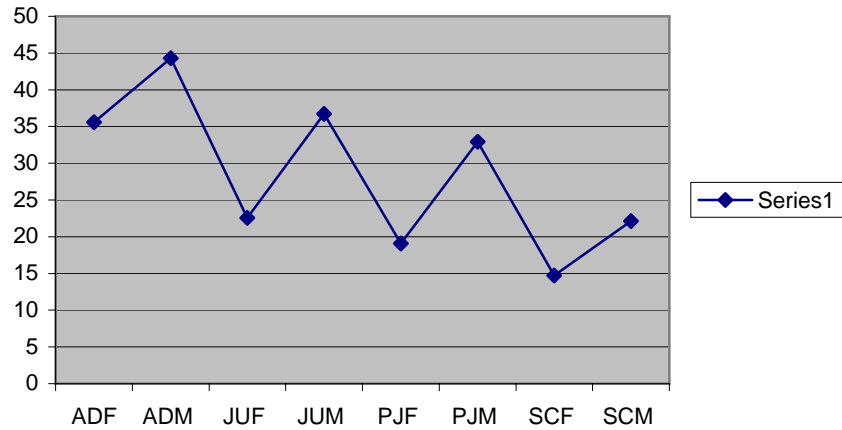
2055 Raw vt



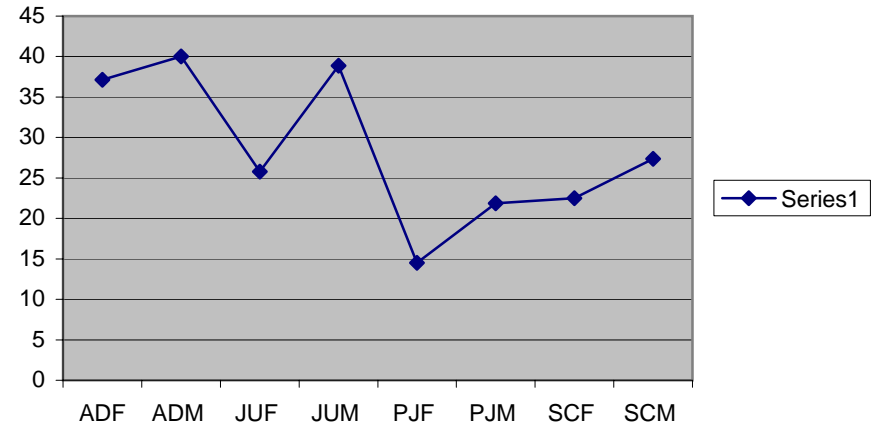
2056 Raw Vt



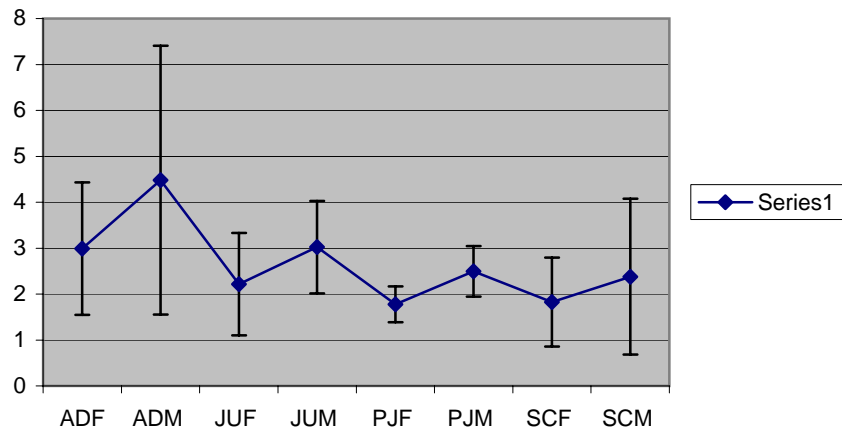
2057 Ipsative



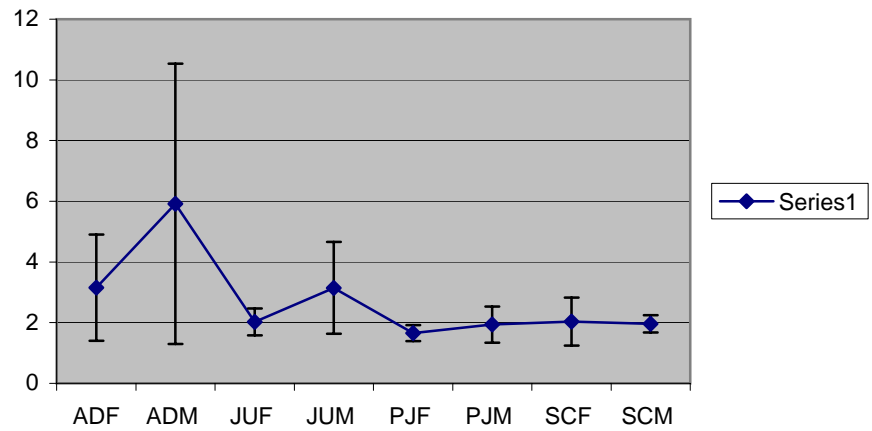
2058 Ipsative



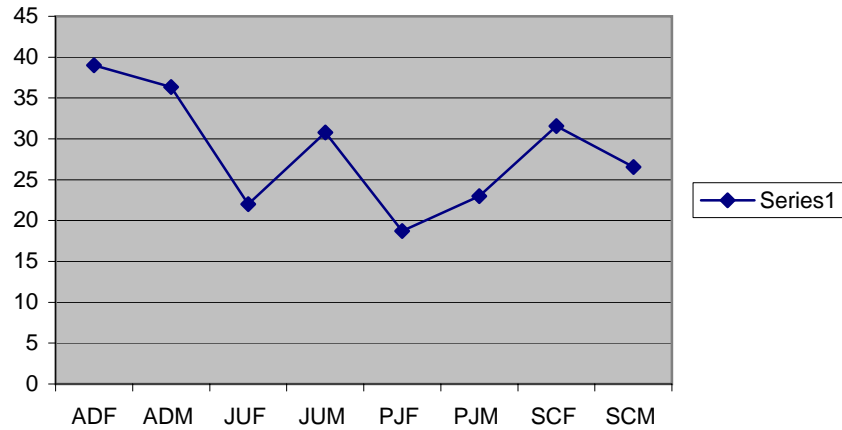
2057 Raw Vt



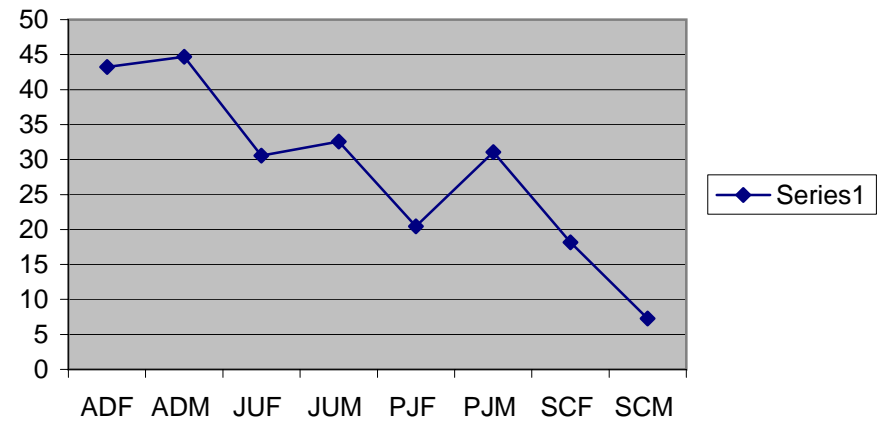
2058 Raw Vt



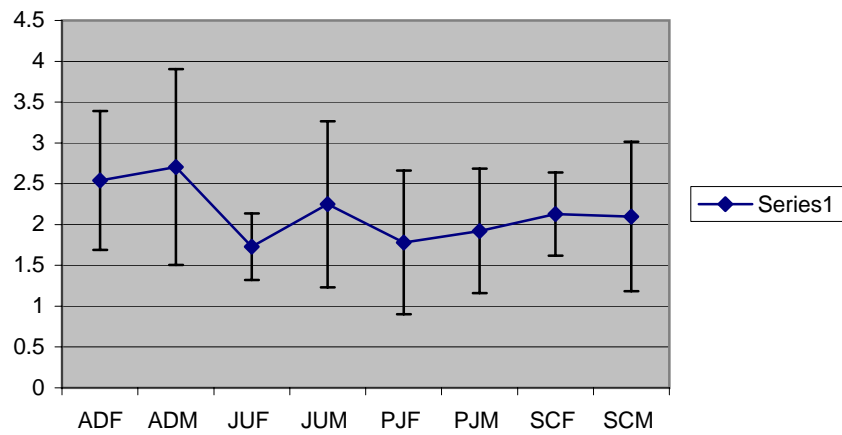
2059 Ipsative



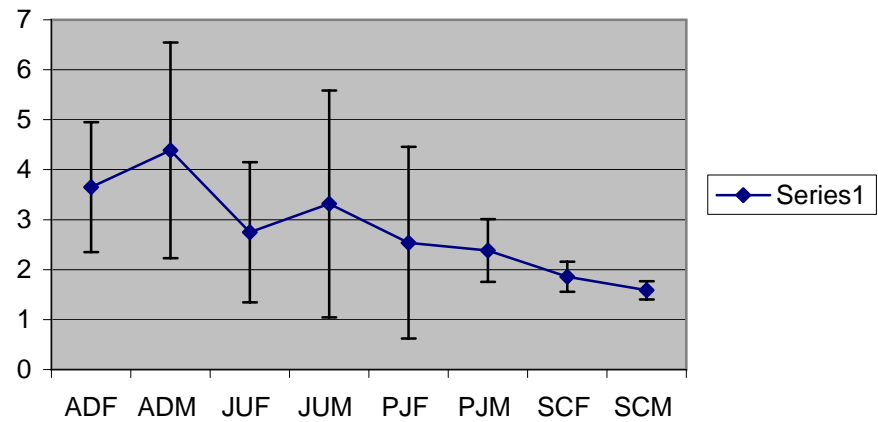
2061 Ipsative



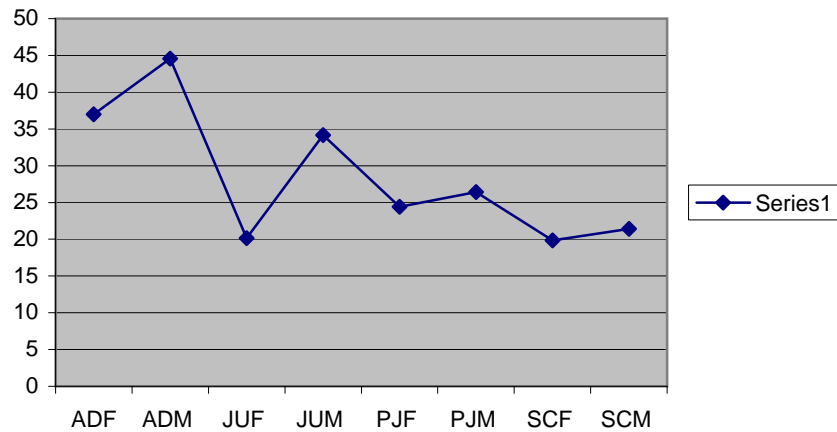
2059 Raw Vt



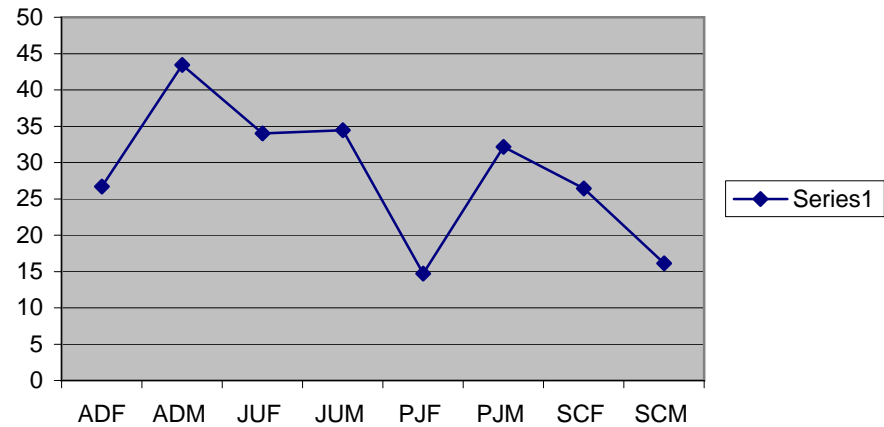
2061 Raw Vt



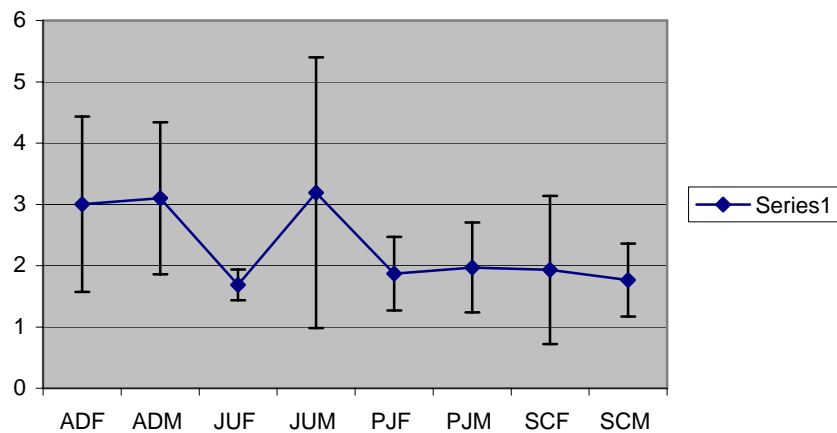
2062 Ipsative



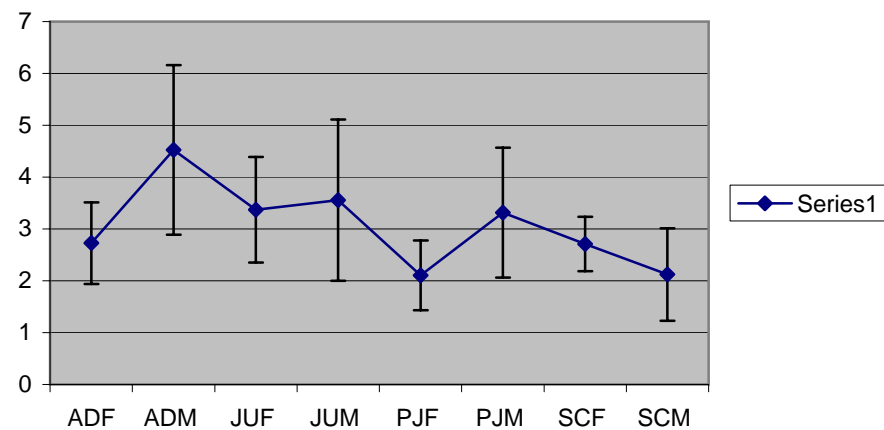
2063 Ipsative



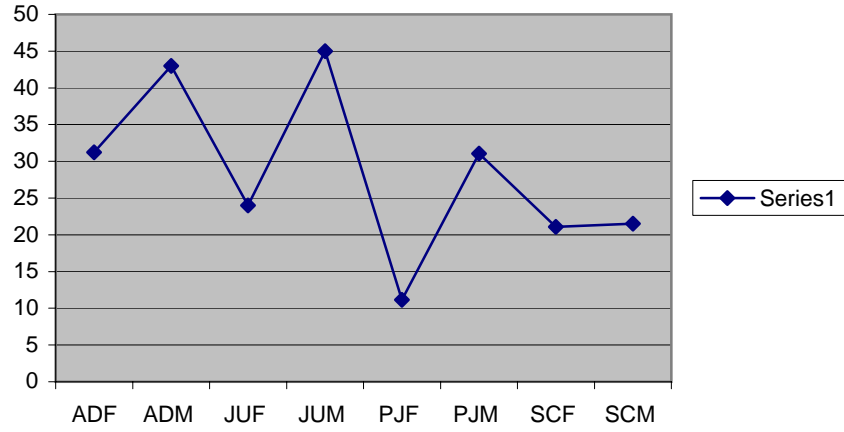
2062 Raw Vt



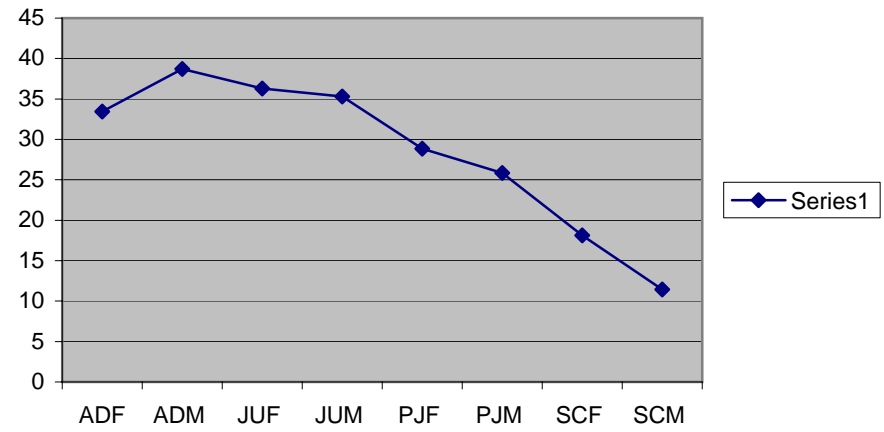
2063 Raw Vt



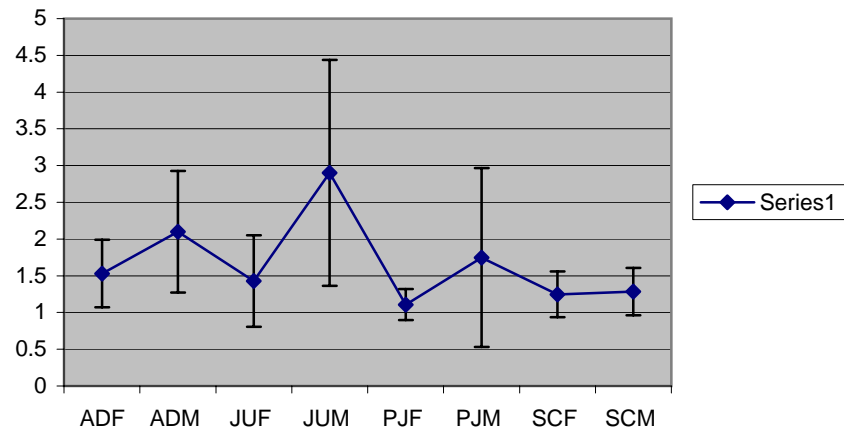
2064 Ipsative



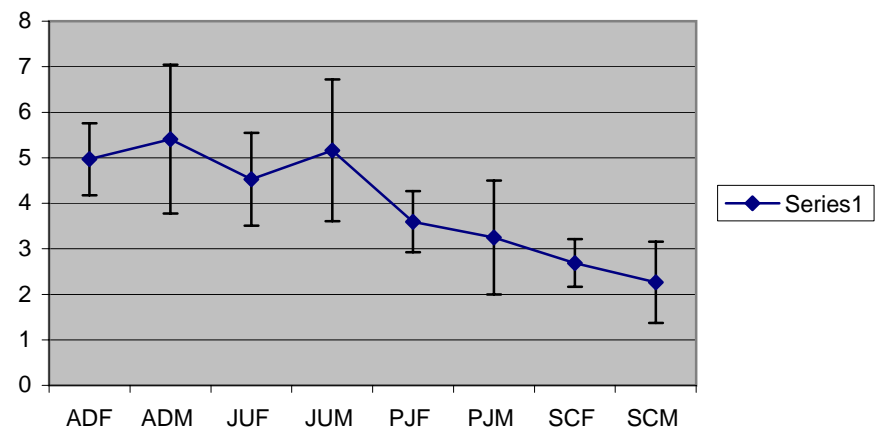
2065 Ipsative



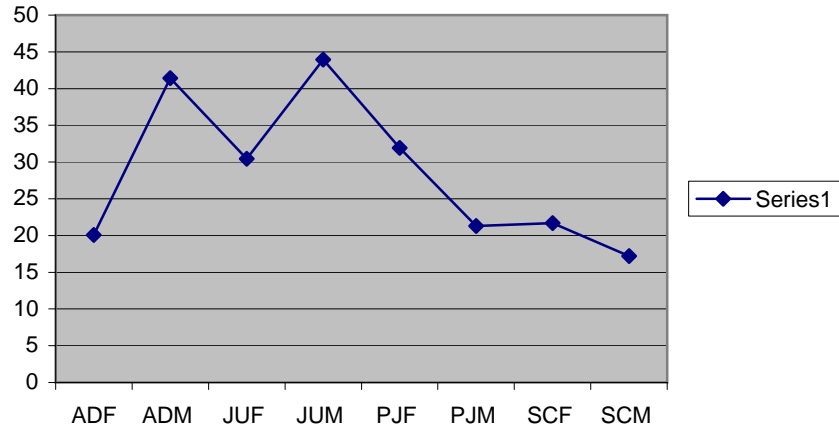
2064 Raw Vt



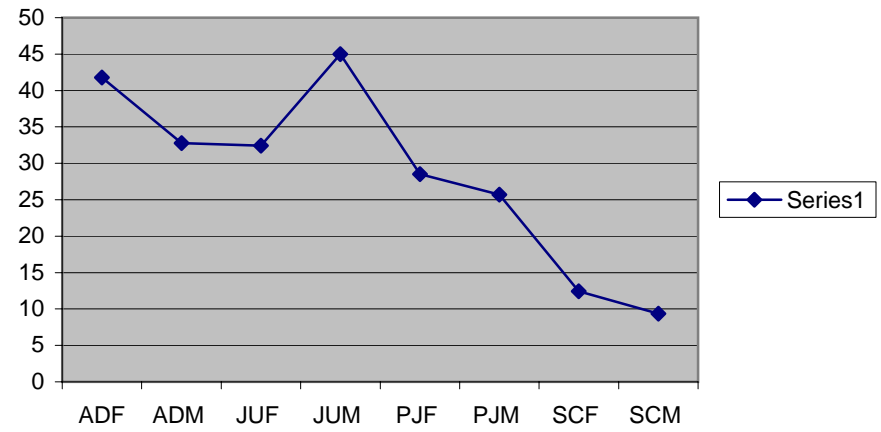
2065 Raw Vt



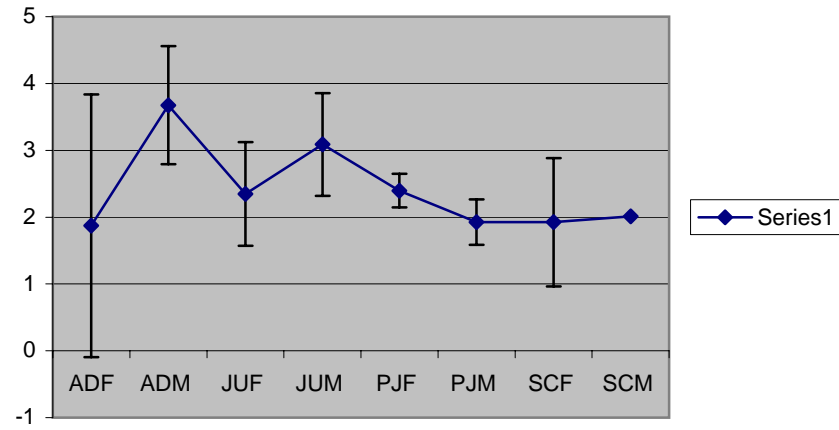
2066 Ipsative



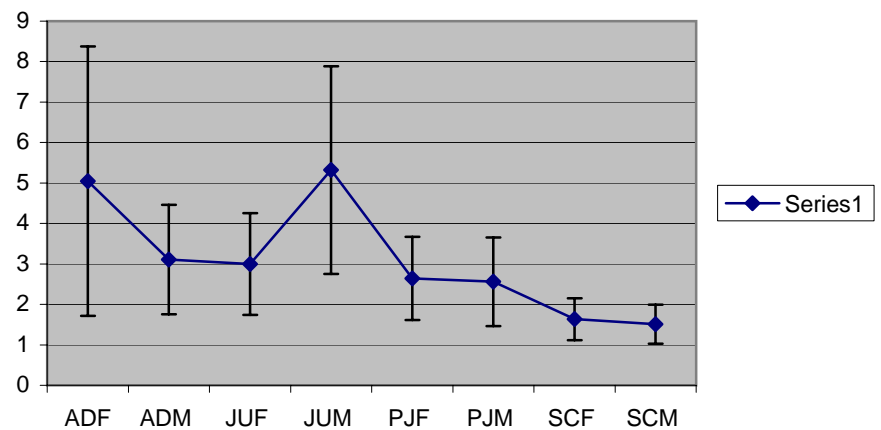
2067 Ipsative



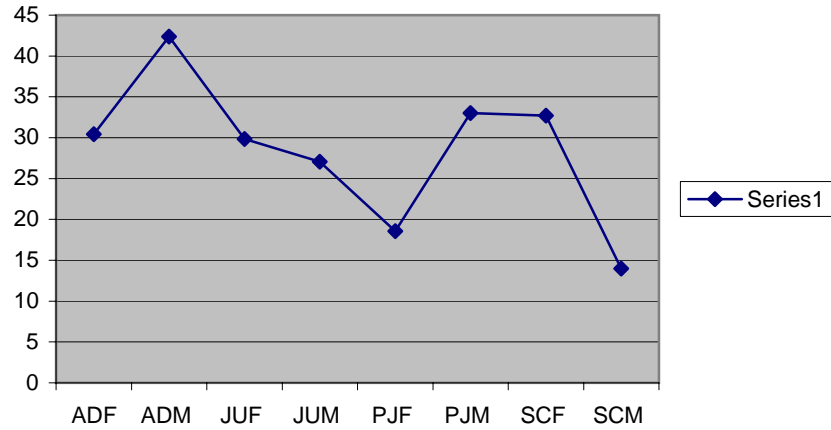
2066 Raw Vt



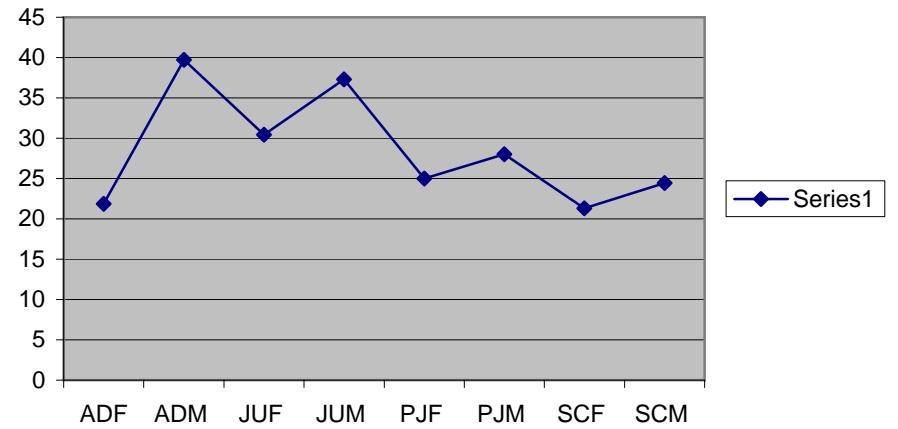
2067 Raw Vt



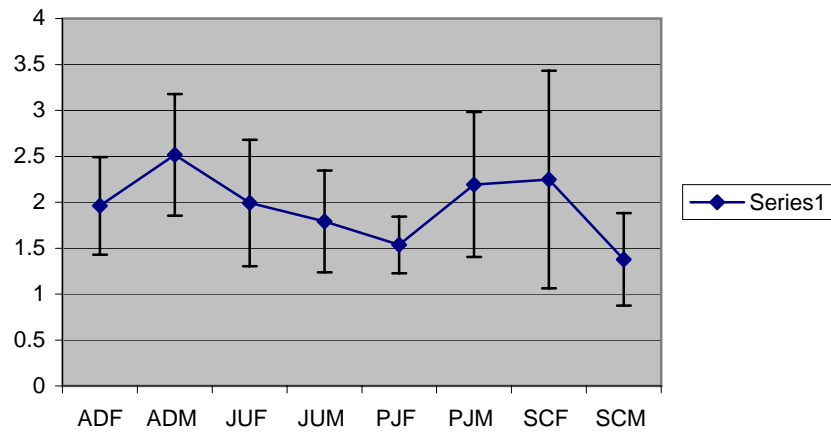
2068 Ipsative



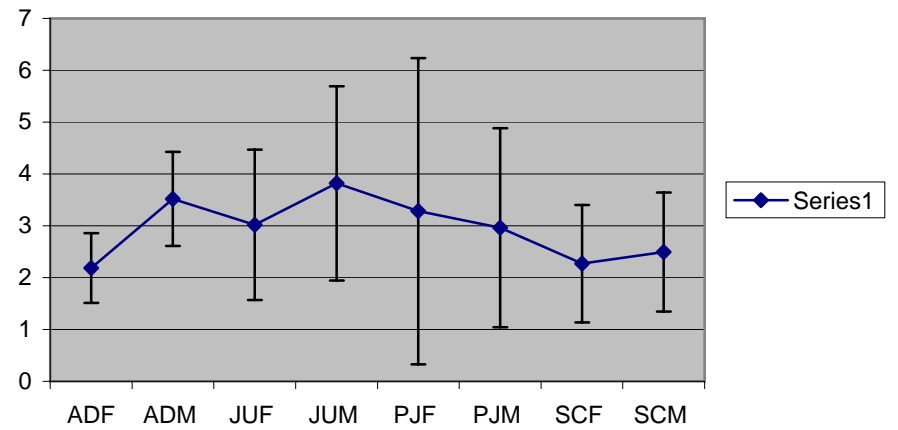
2069 Ipsative



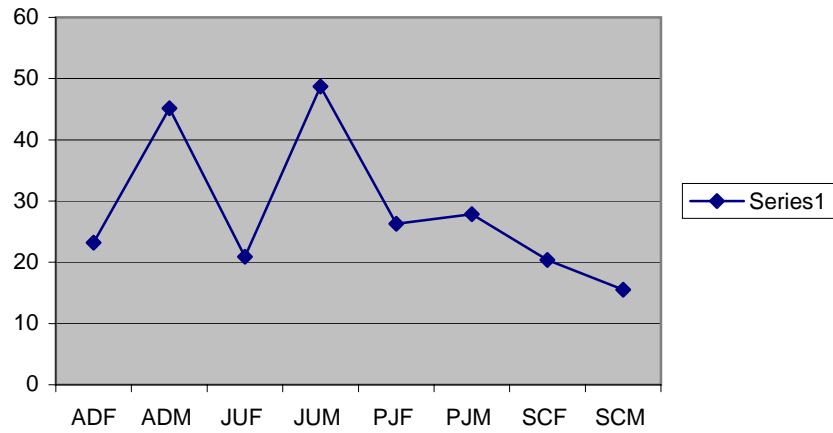
2068 Raw Vt



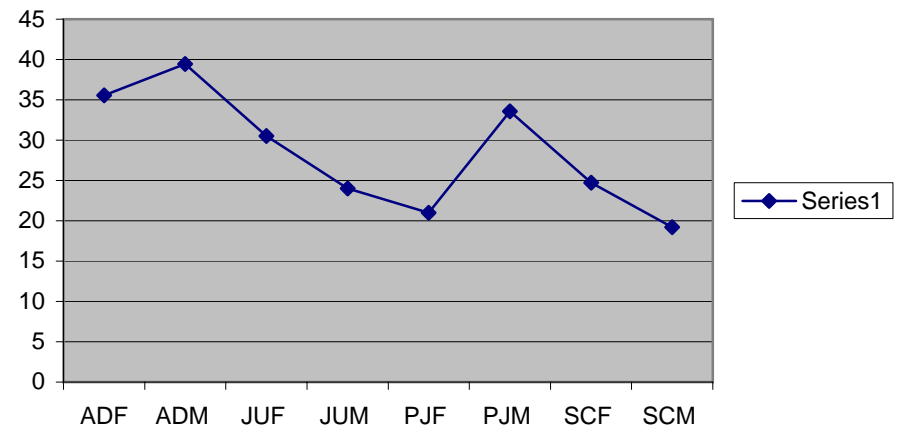
2069 Raw Vt



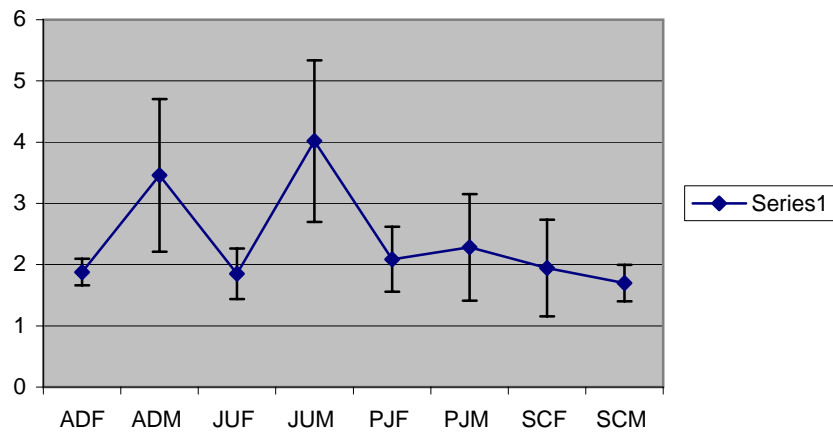
2070 Ipsative



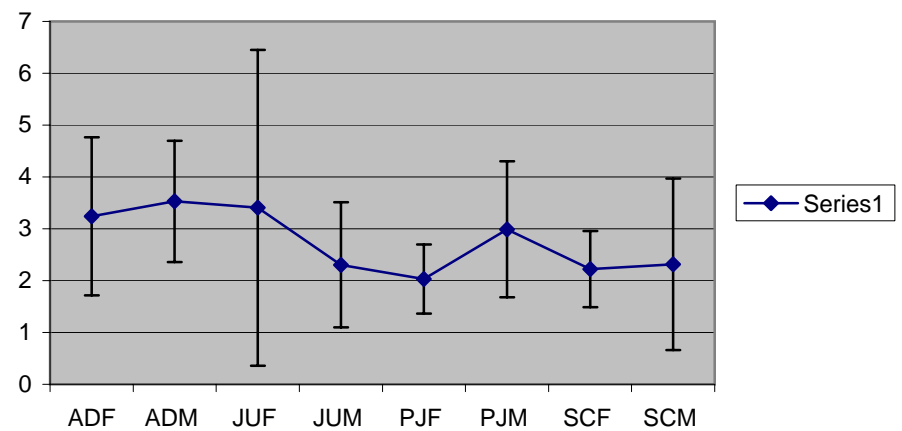
2071 Ipsative



2070 Raw Vt

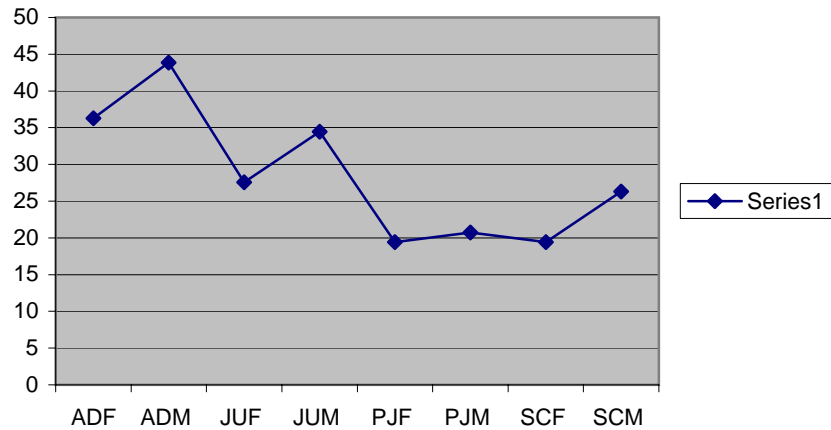


2071 Raw Vt

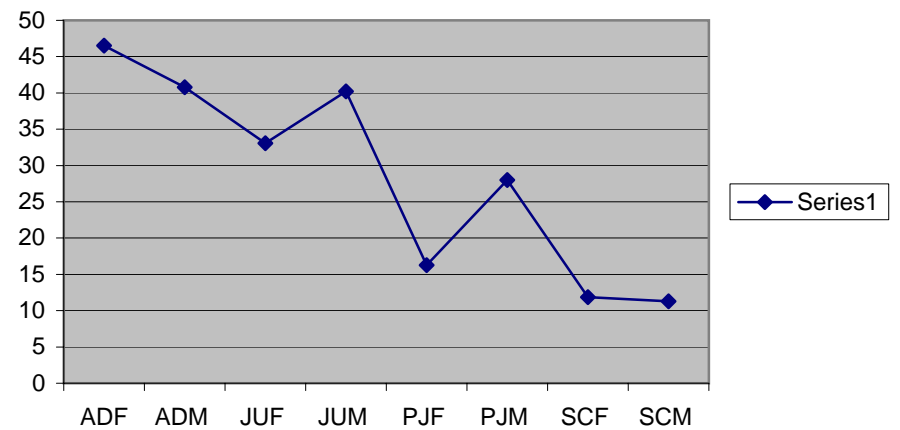




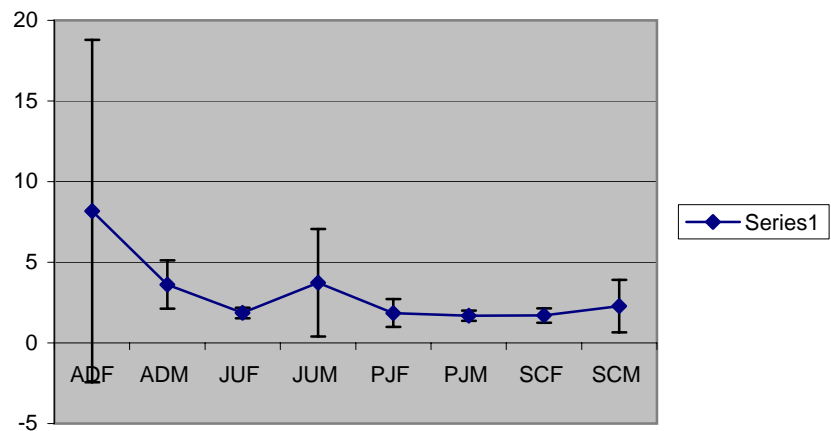
2072 Ipsative



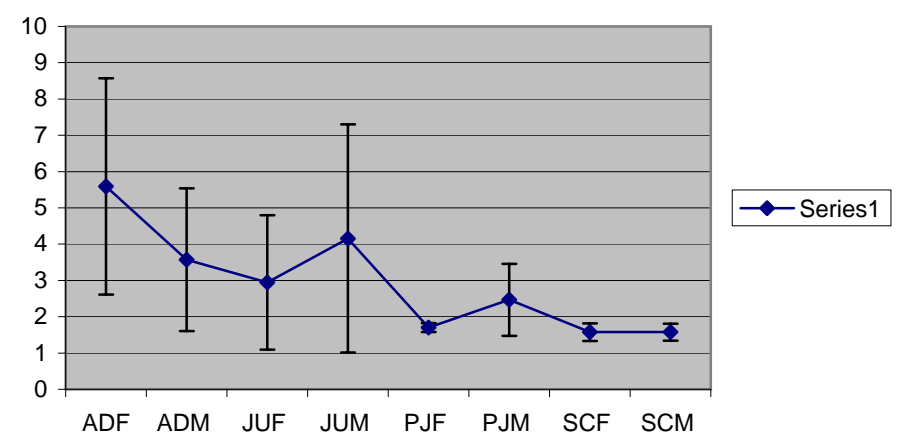
2073 Ipsative



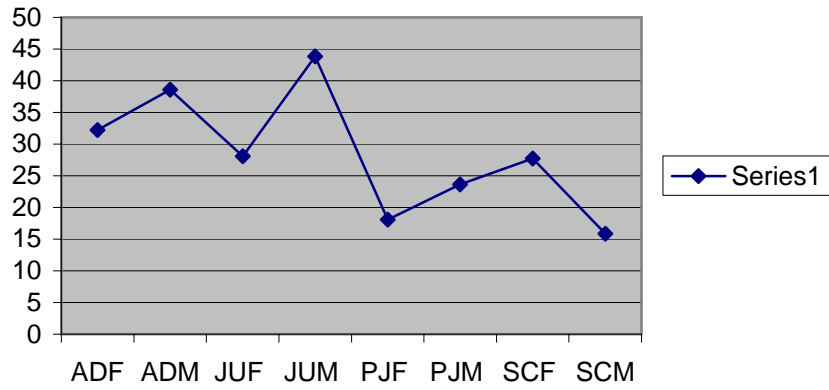
2072 Raw Vt



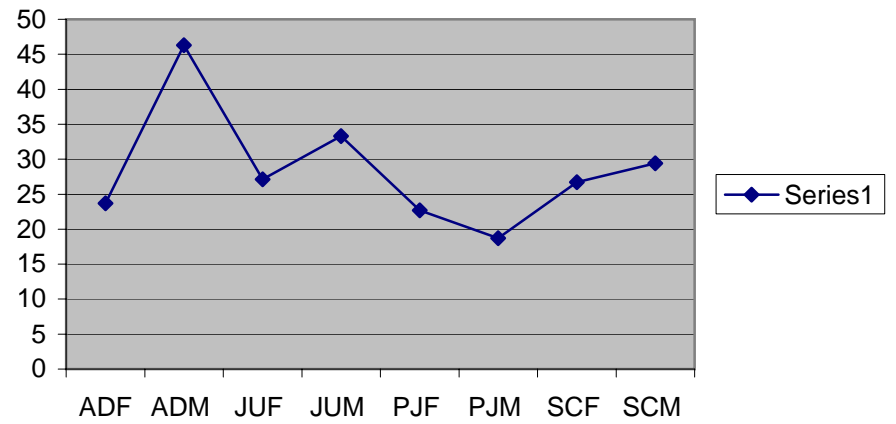
2073 Raw Vt



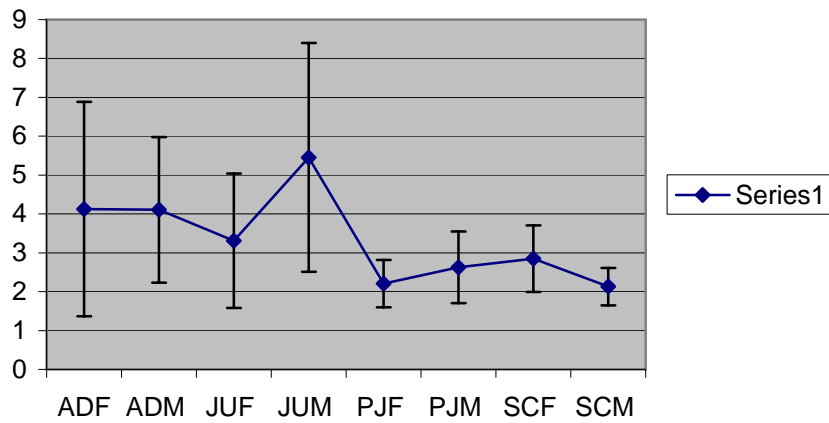
2074 Ipsative



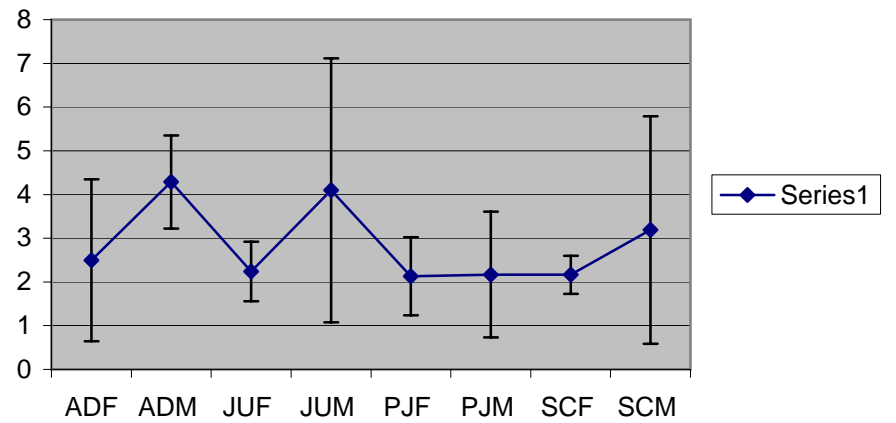
2075 Ipsative



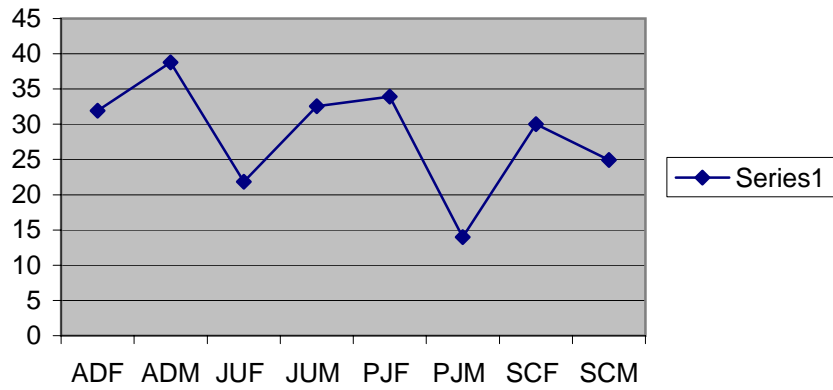
2074 Raw Vt



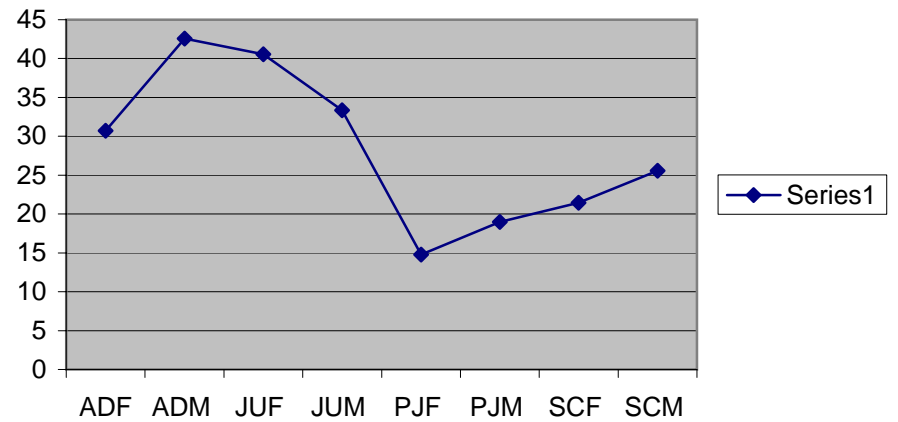
2075 Raw Vt



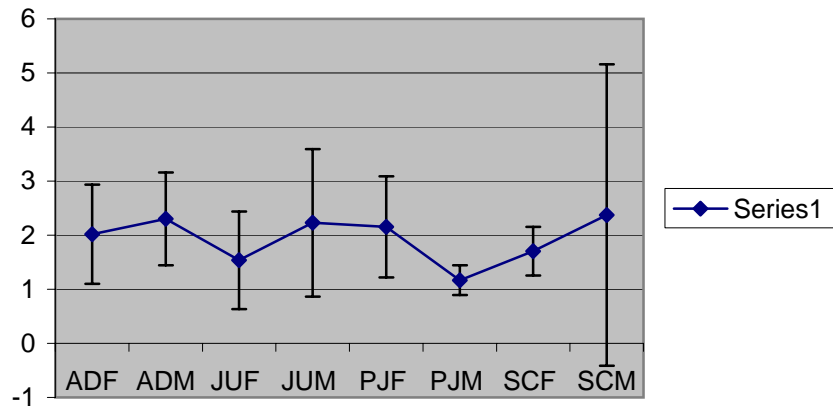
### 2076 Ipsative



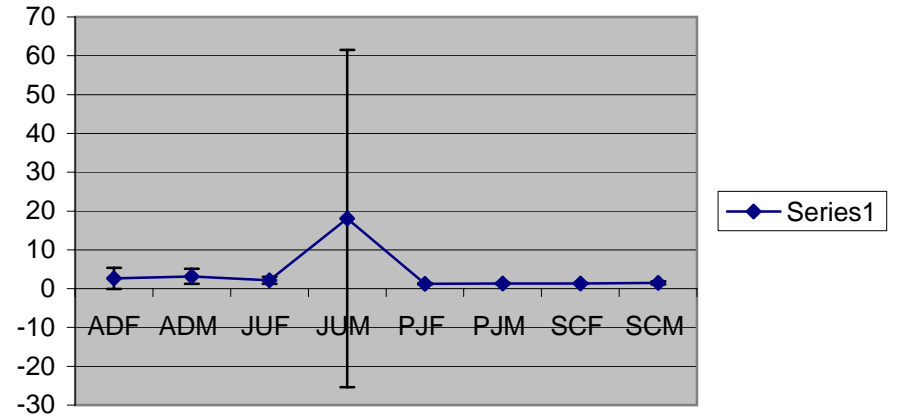
### 2077 Ipsative



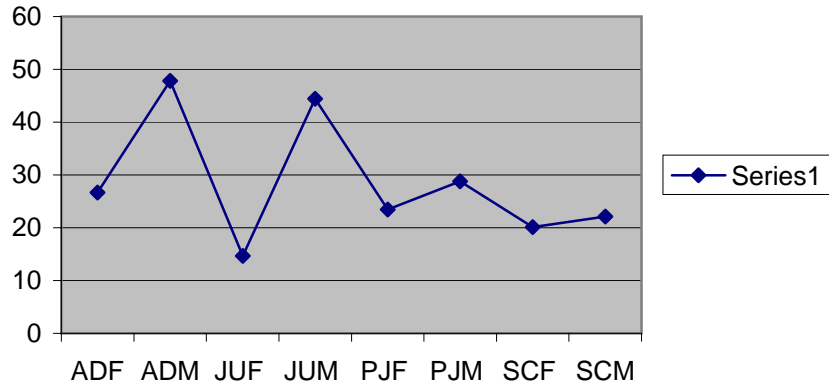
### 2076 Raw Vt



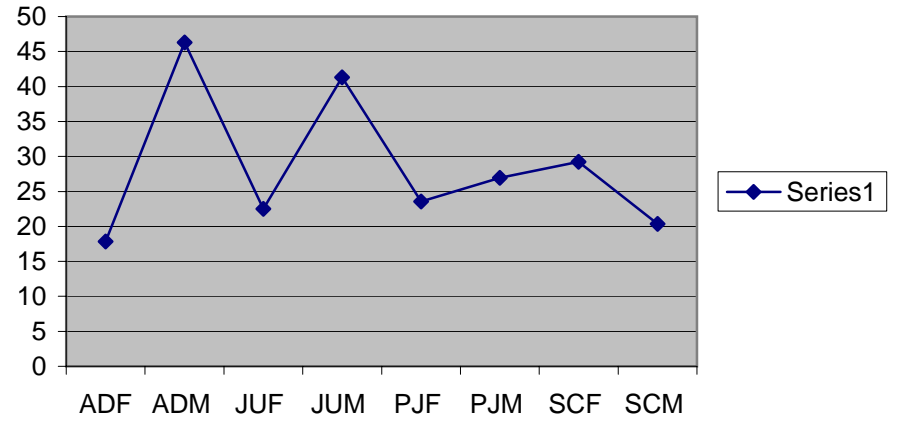
### 2077 Raw Vt



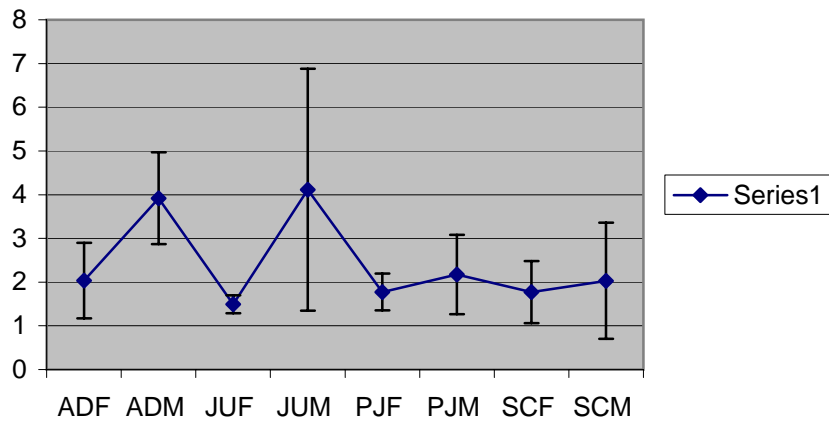
2078 Ipsative



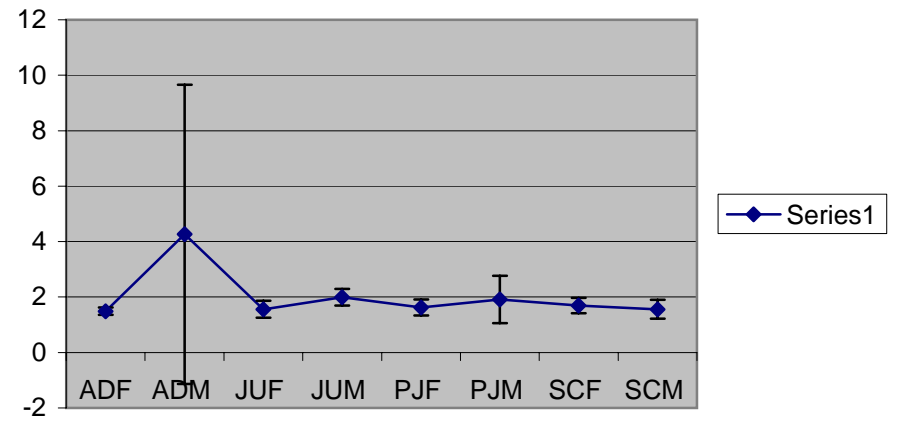
2079 Ipsative



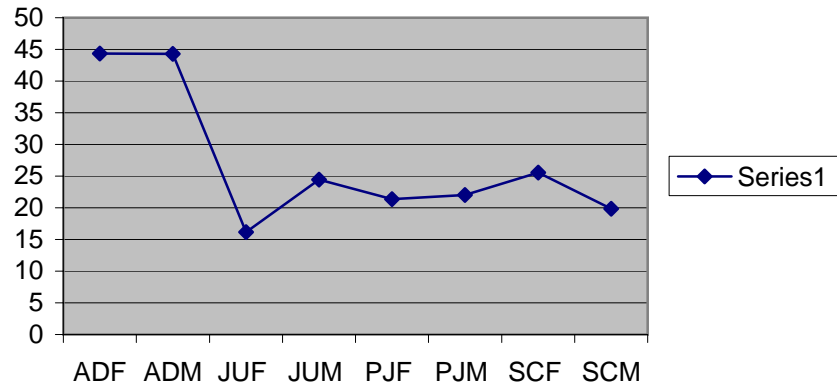
2078 Raw Vt



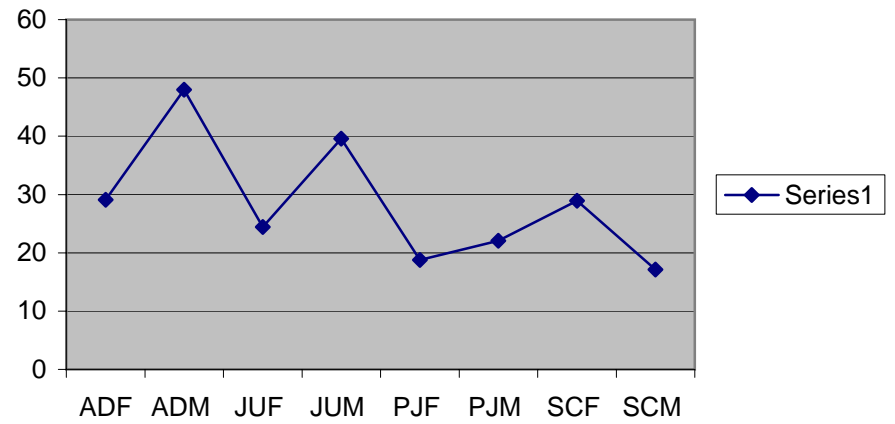
2079 Raw Vt



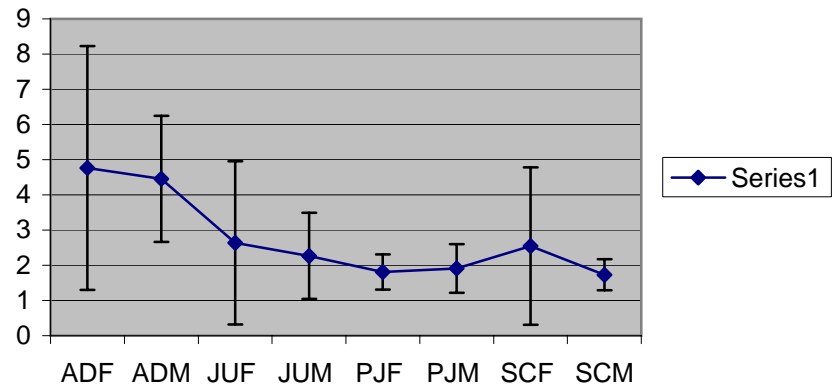
2080 Ipsative



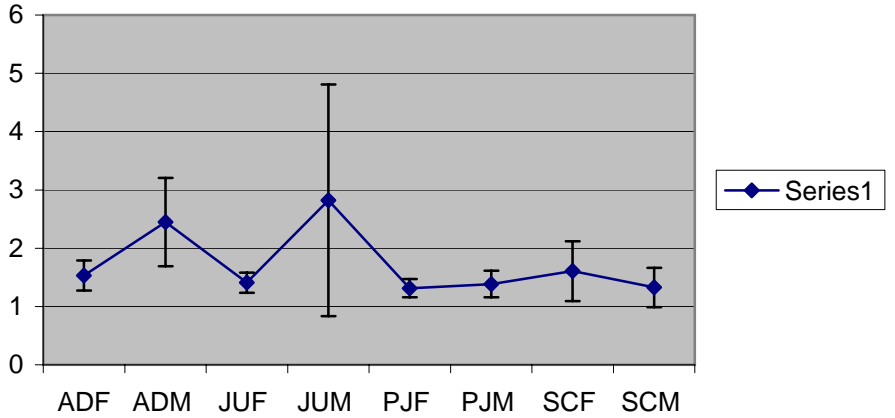
2081 Ipsative



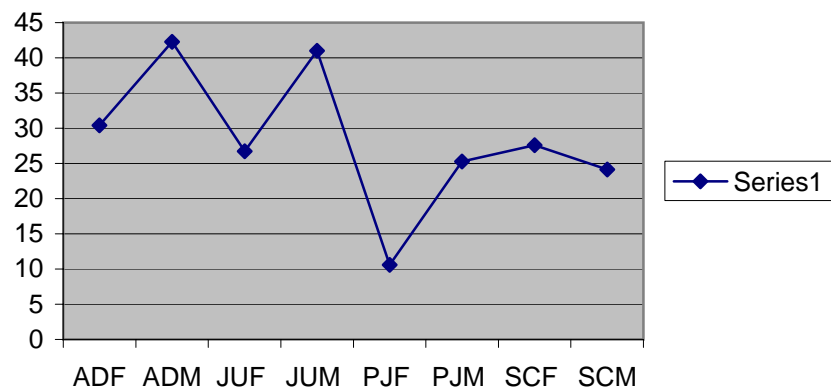
2080 Raw Vt



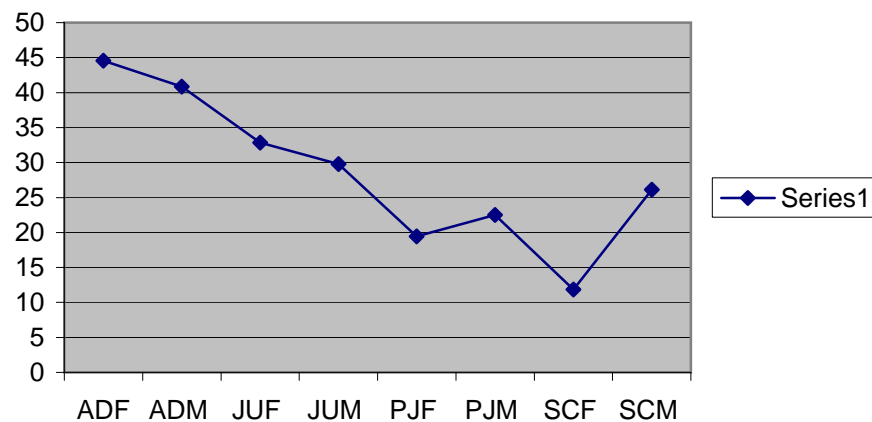
2081 Raw Vt



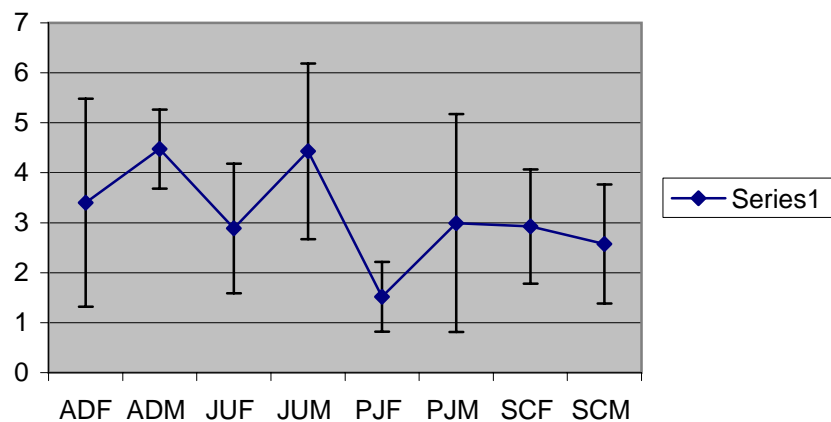
2082 Ipsative



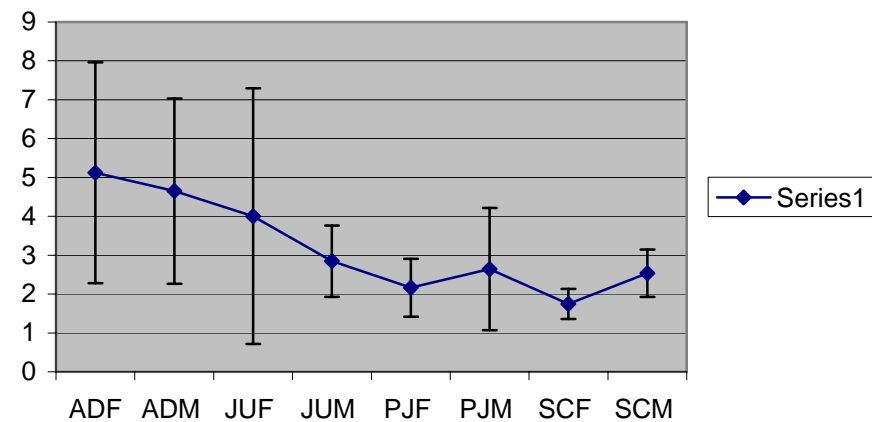
2083 Ipsative



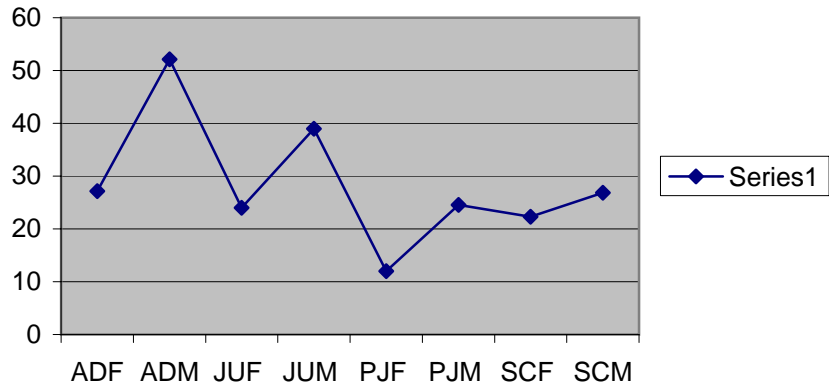
2082 Raw Vt



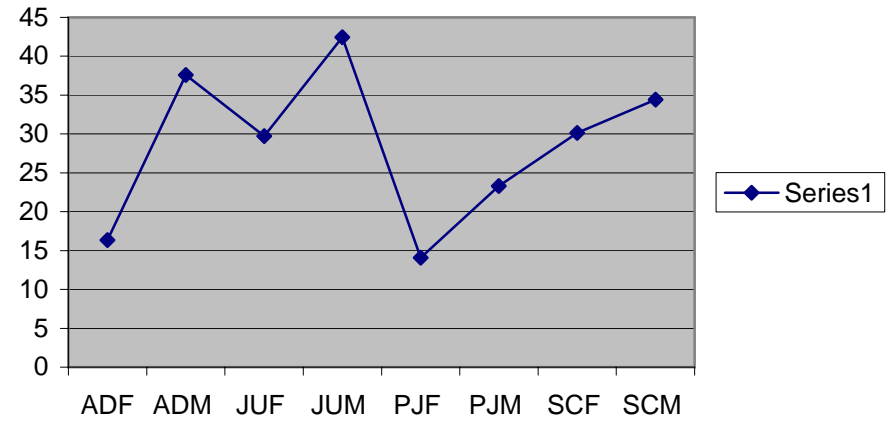
2083 Raw Vt



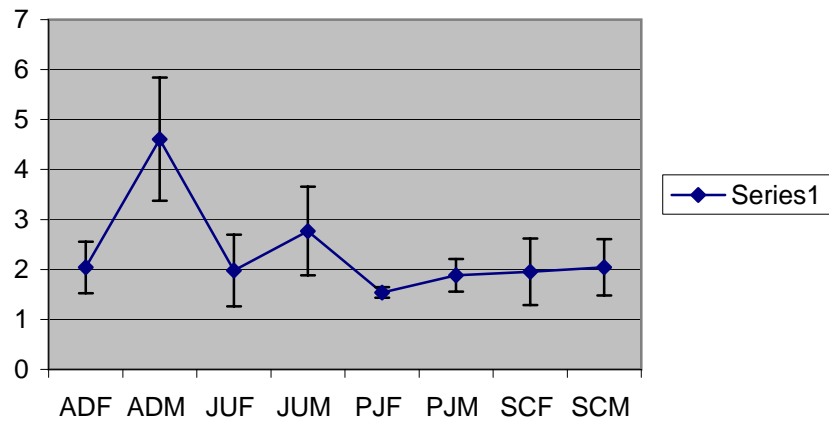
### 2084 Ipsative



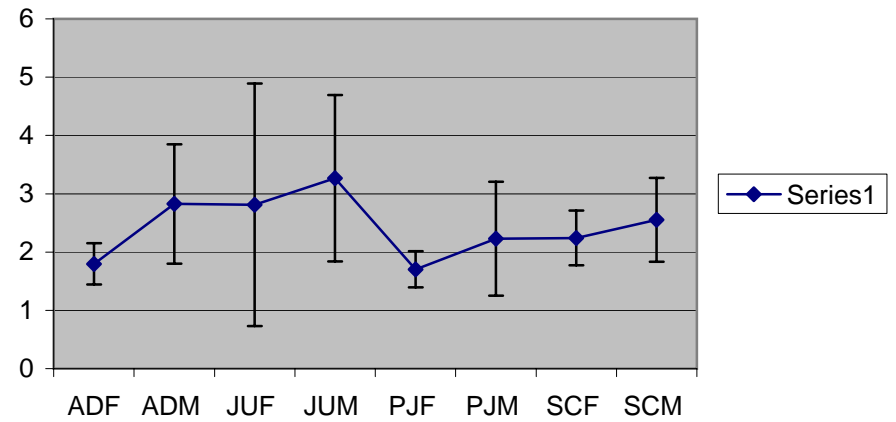
### 2085 Ipsative



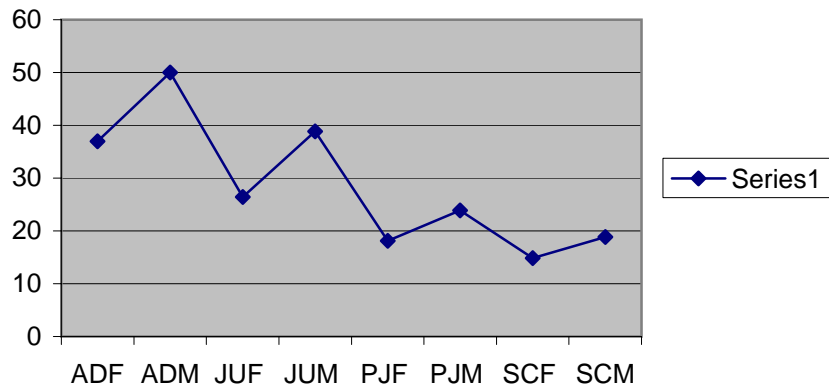
### 2084 Raw Vt



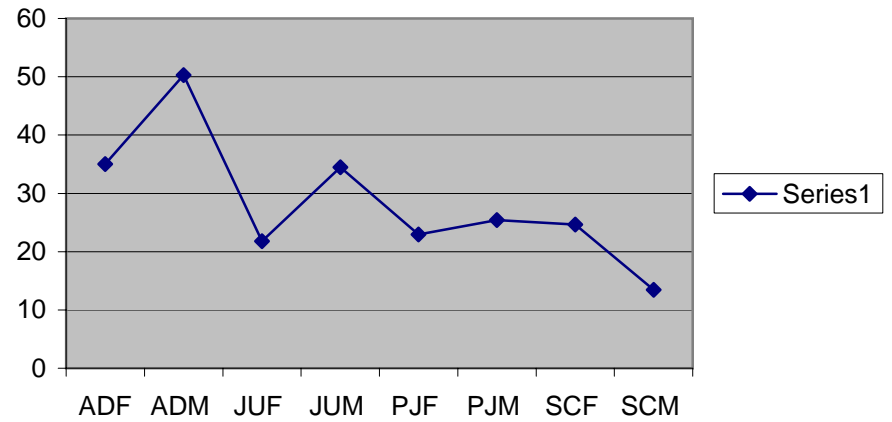
### 2085 Raw Vt



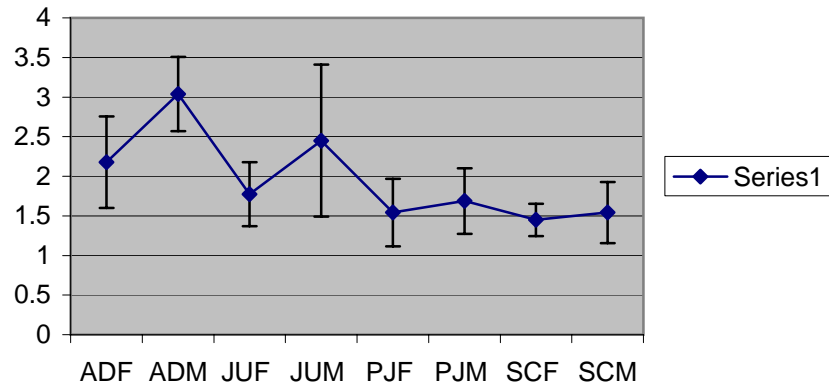
2086 Ipsative



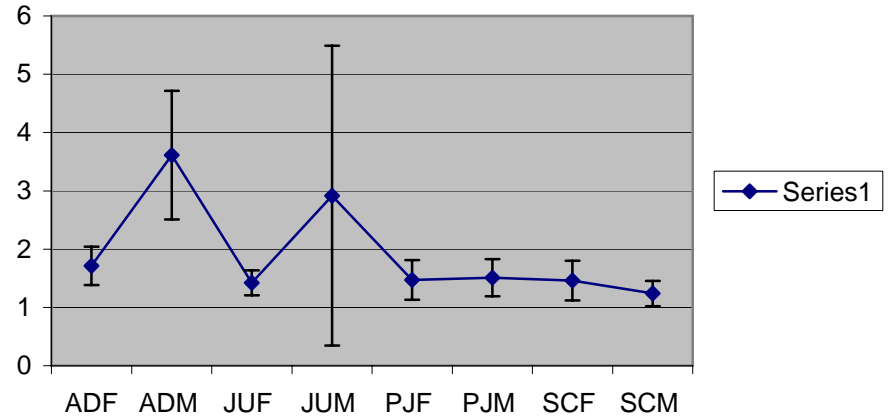
2087 Ipsative



2086 Raw Vt

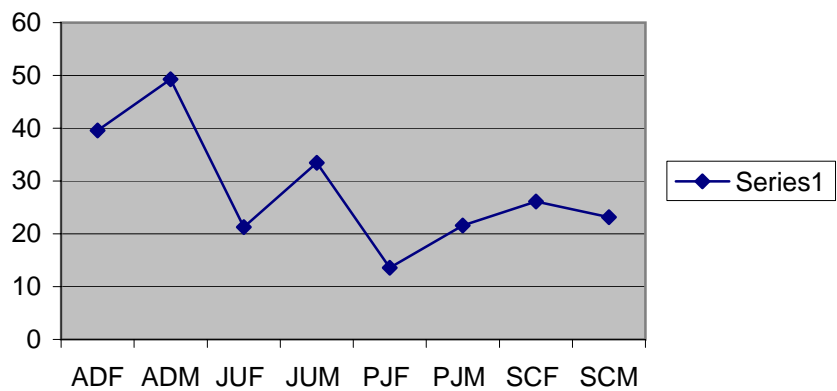


2087 Raw Vt

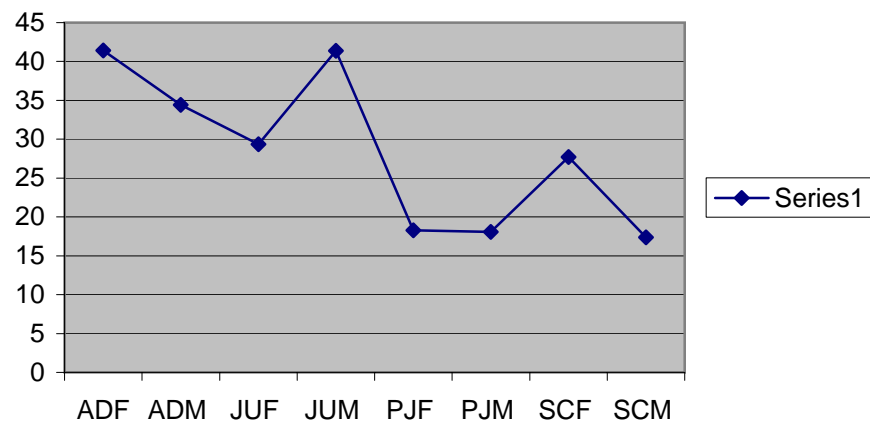




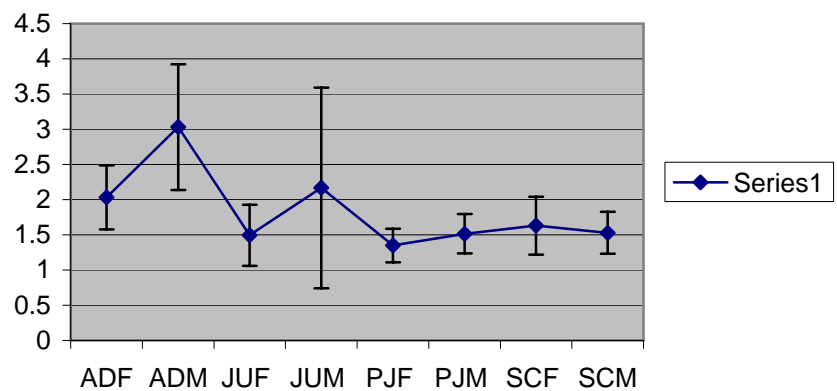
2088 Ipsative



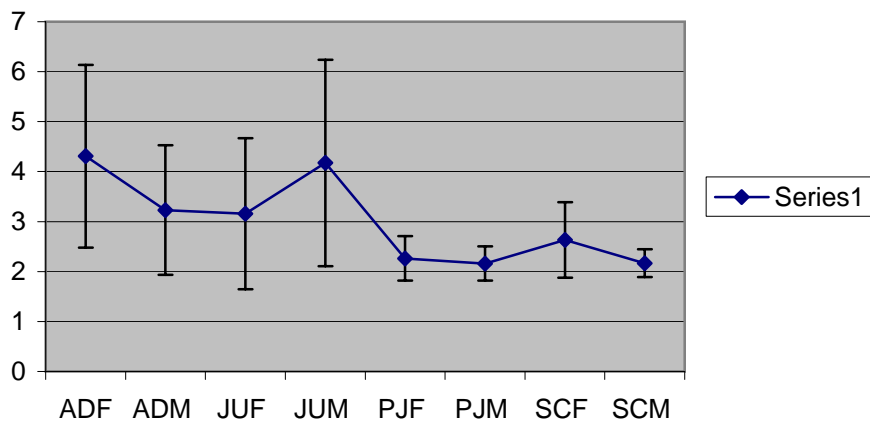
2089 Ipsative



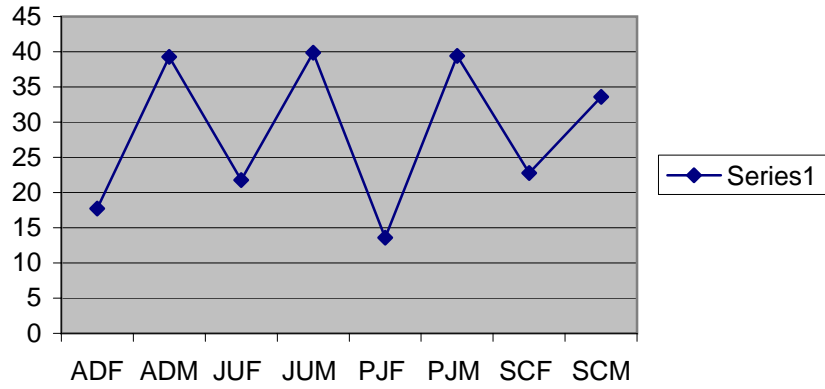
2088 Raw Vt



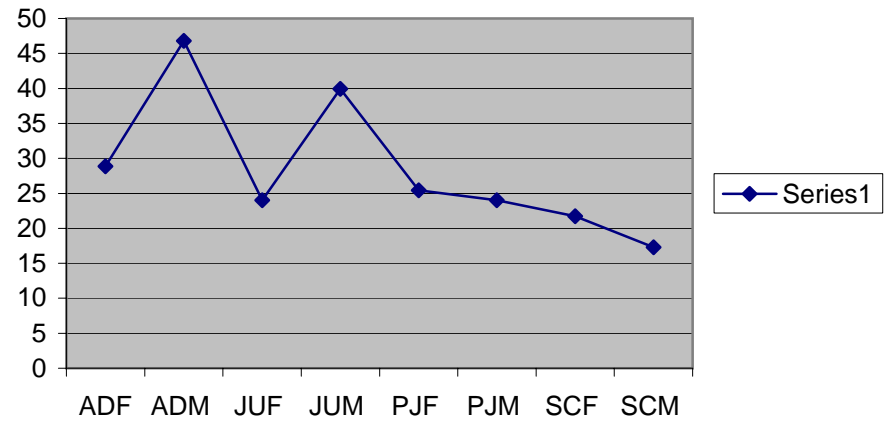
2089 Raw Vt



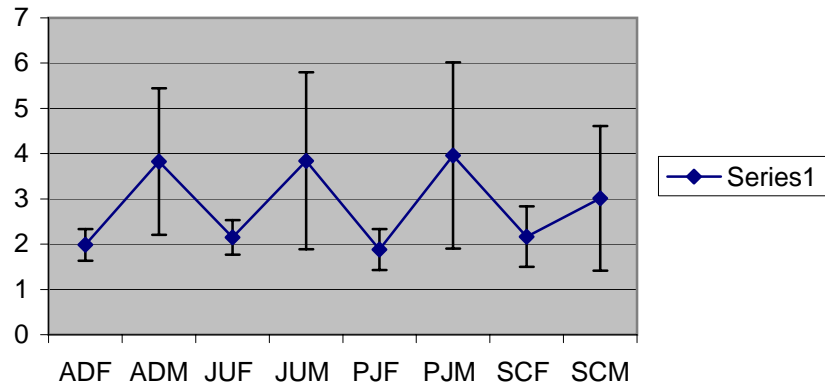
### 2090 Ipsative



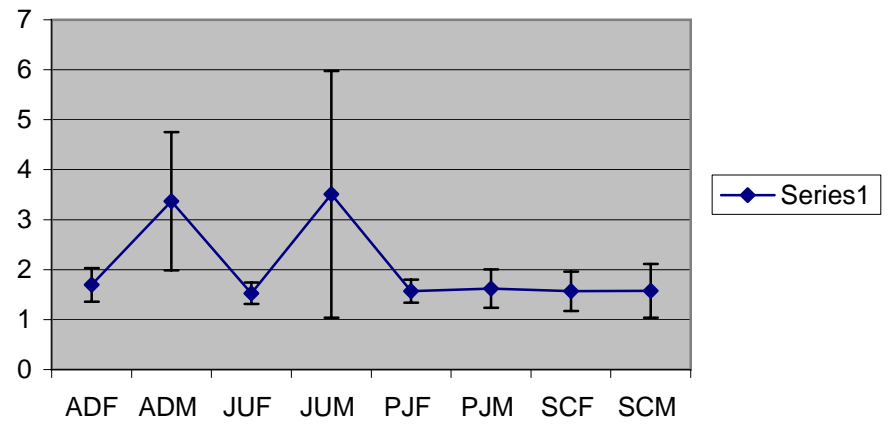
### 2091 Ipsative



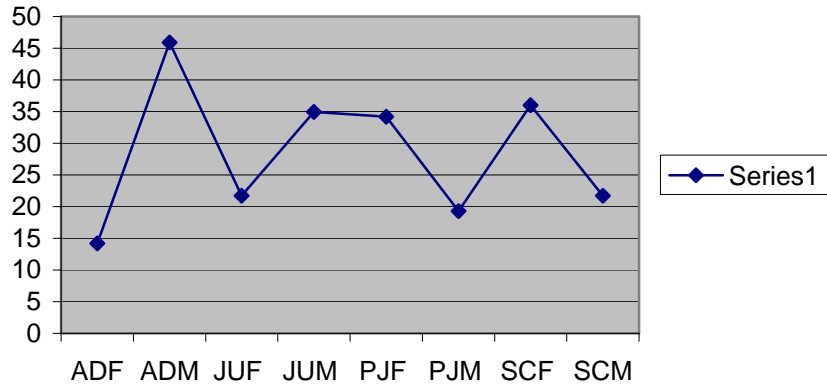
### 2090 Raw Vt



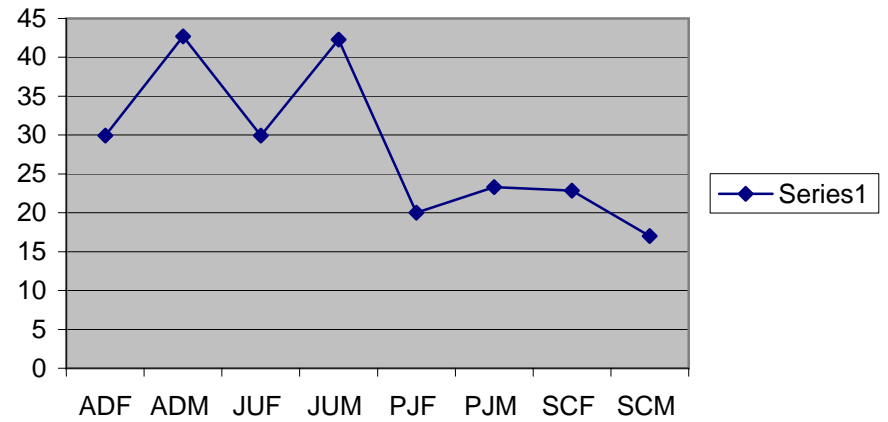
### 2091 Raw Vt



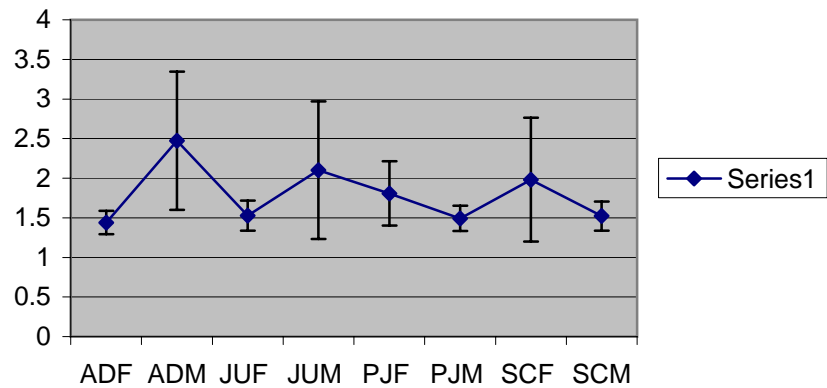
### 2092 Ipsative



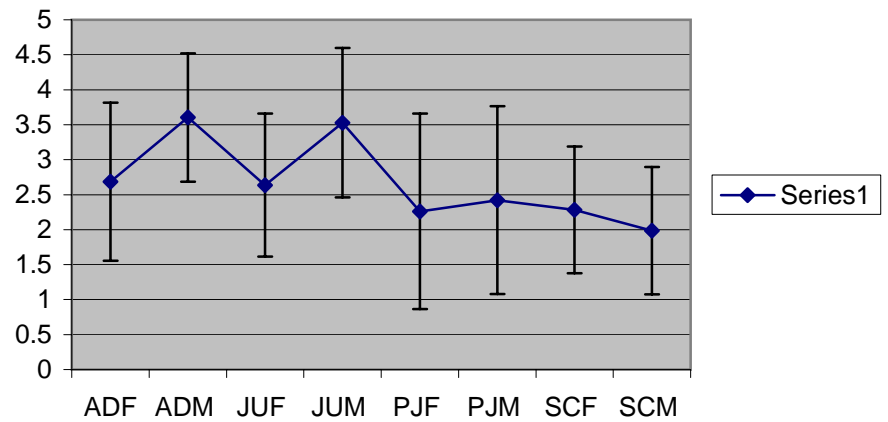
### 2093 Ipsative



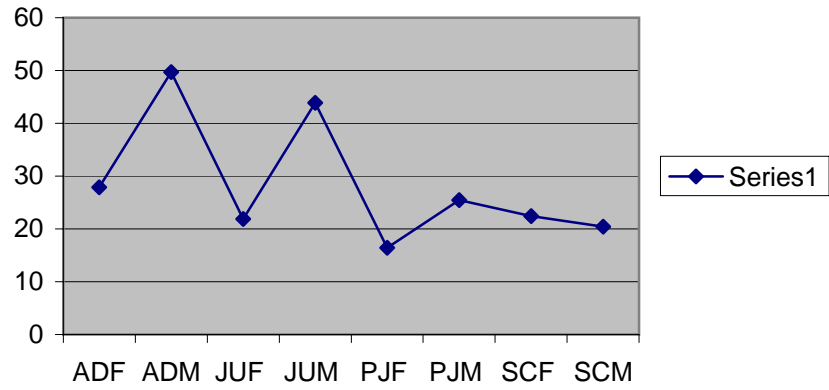
### 2092 Raw Vt



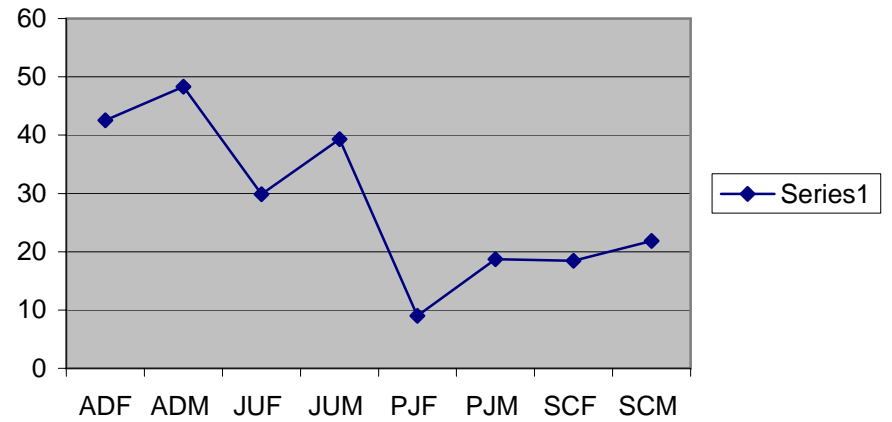
### 2093 Raw Vt



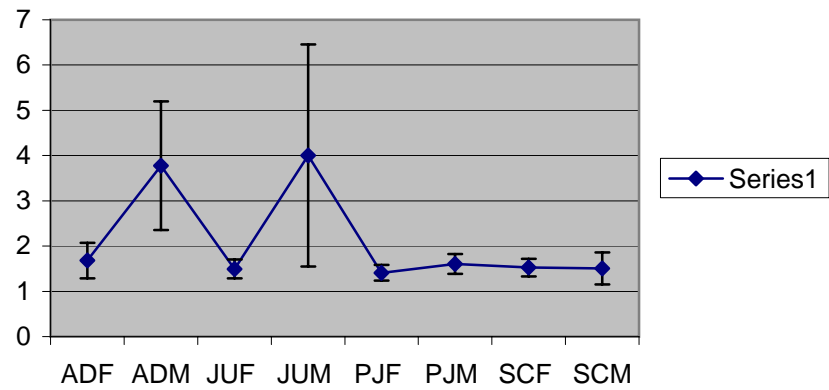
2094 Ipsative



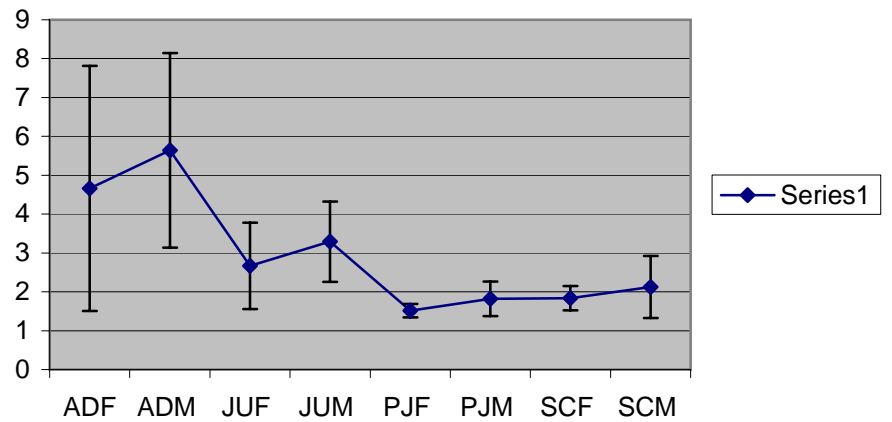
2095 Ipsative



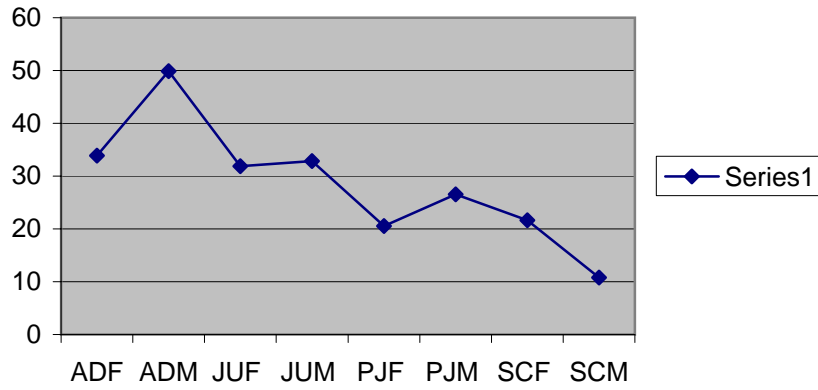
2094 Raw Vt



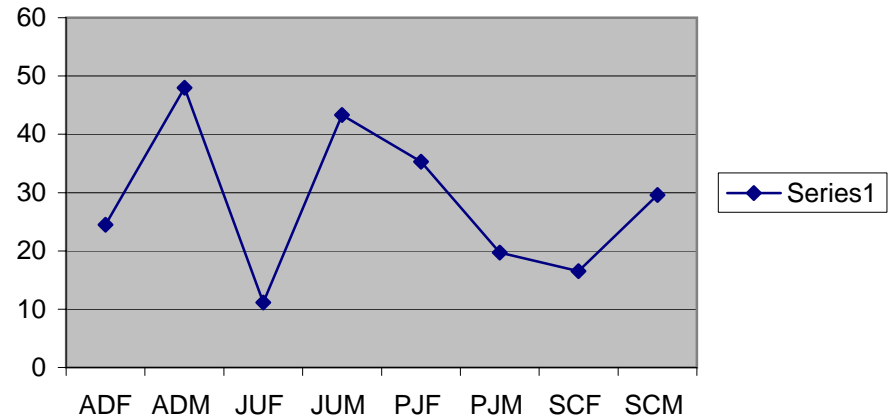
2095 Raw Vt



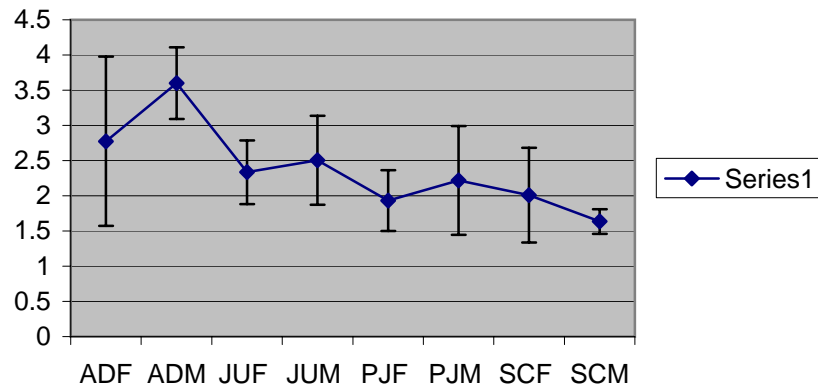
### 2096 Ipsative



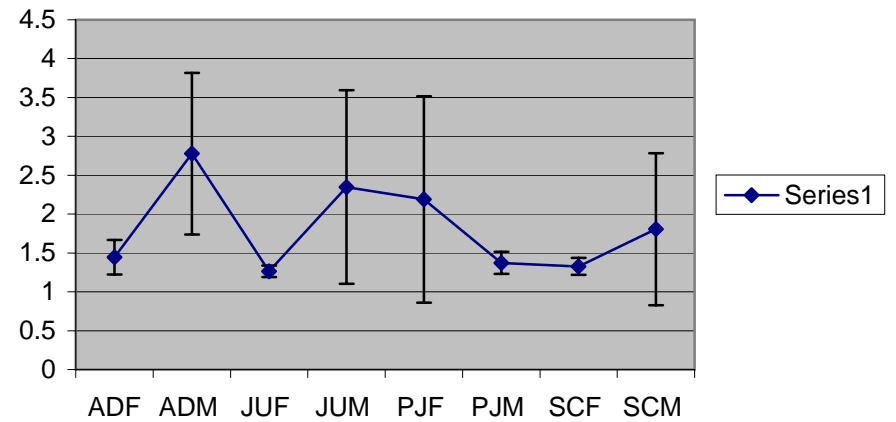
### 2097 Ipsative



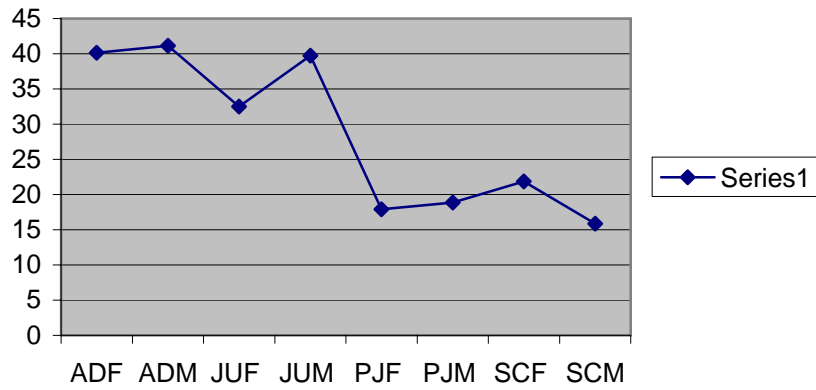
### 2096 Raw Vt



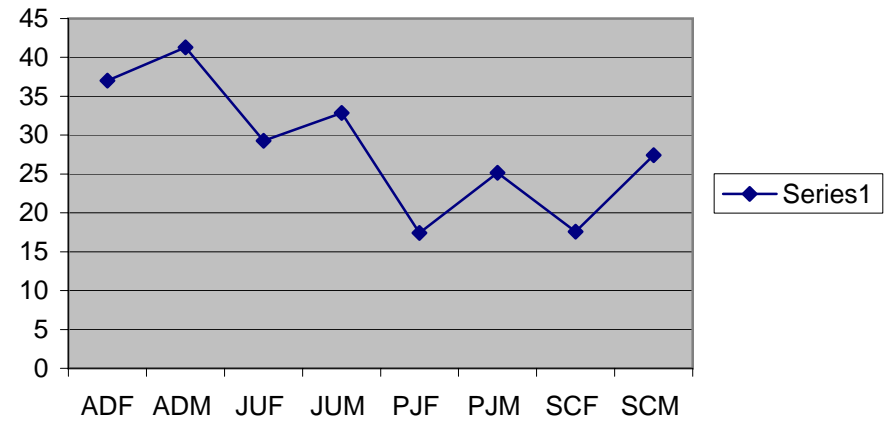
### 2097 Raw Vt



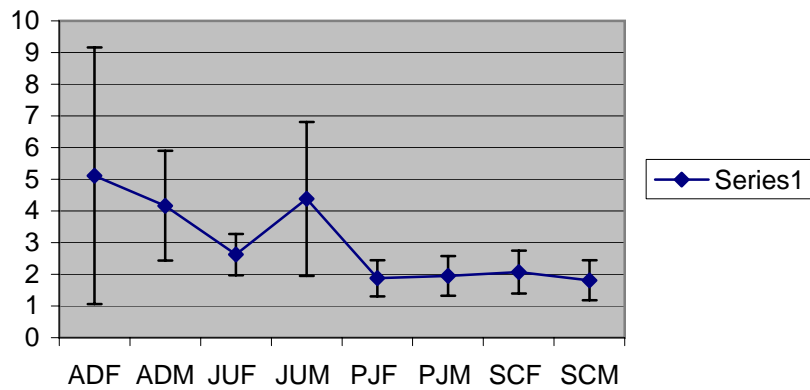
2098 Ipsative



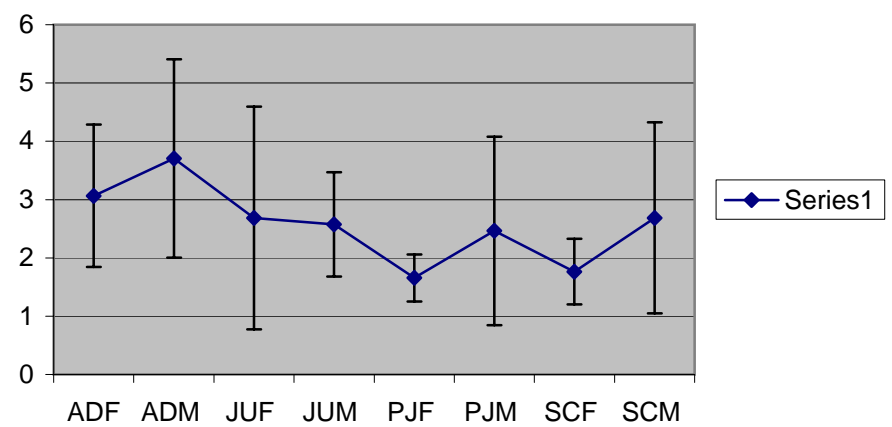
2099 Ipsative



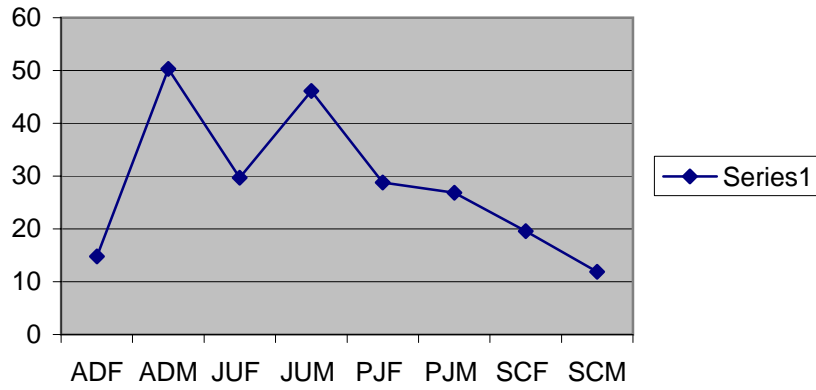
2098 Raw Vt



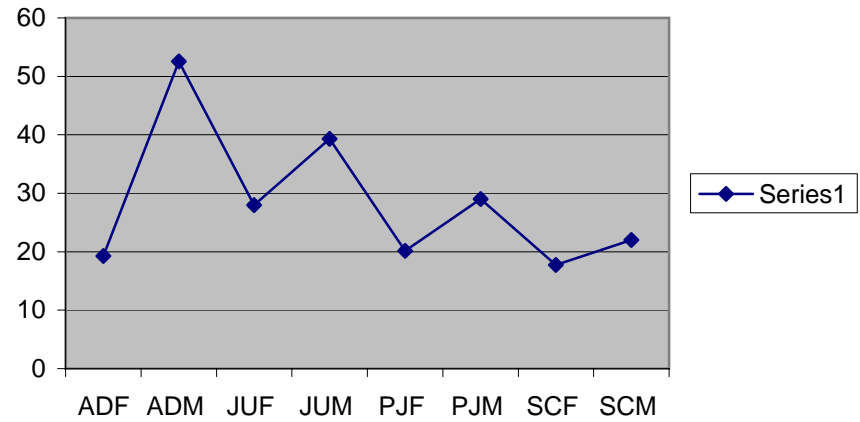
2099 Raw Vt



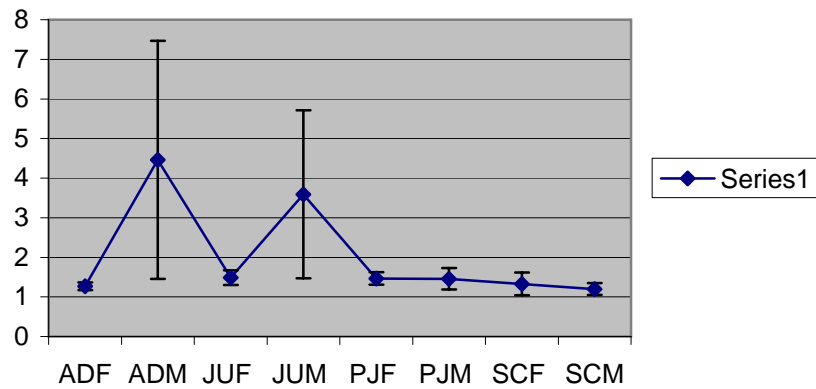
### 2100 Ipsative



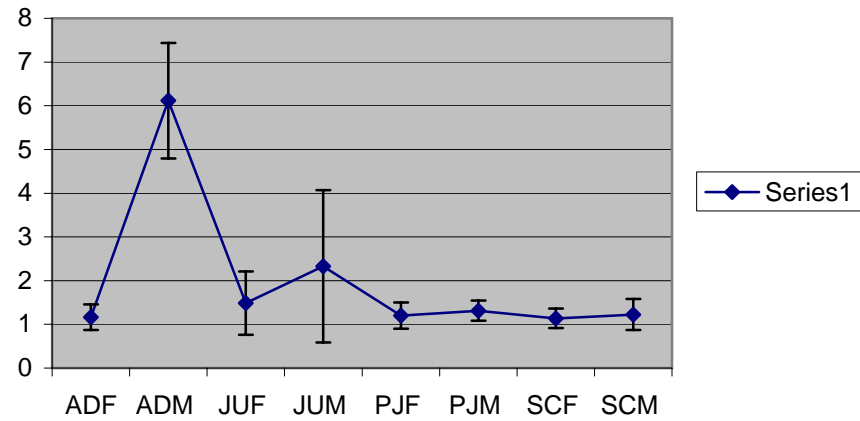
### 2101 Ipsative



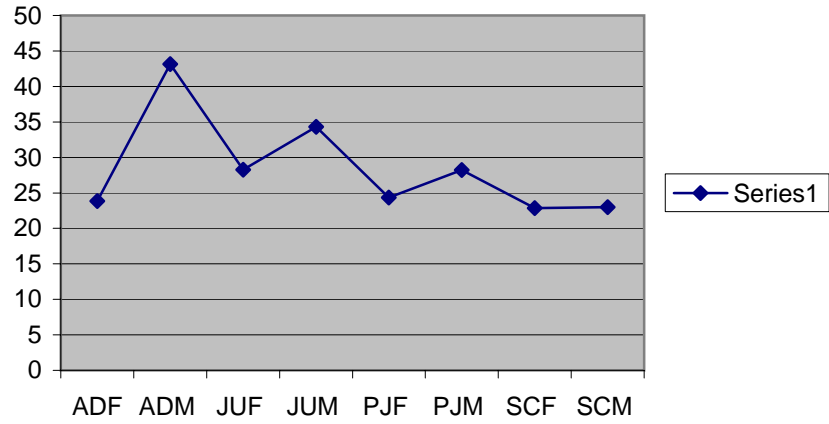
### 2100 Raw Vt



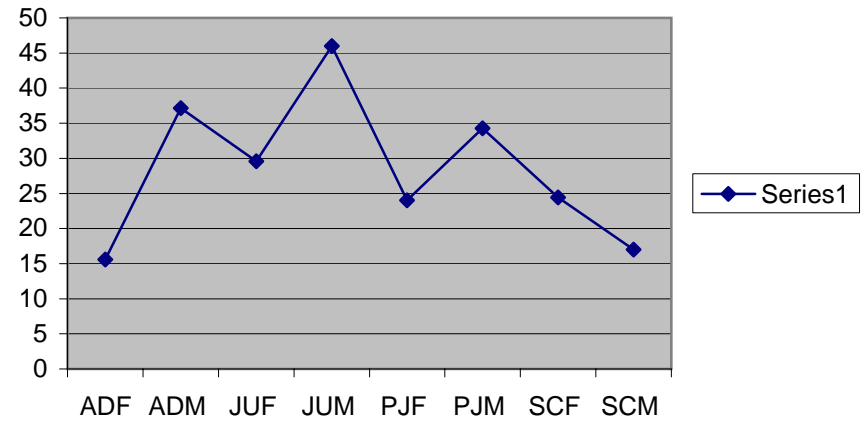
### 2101 Raw Vt



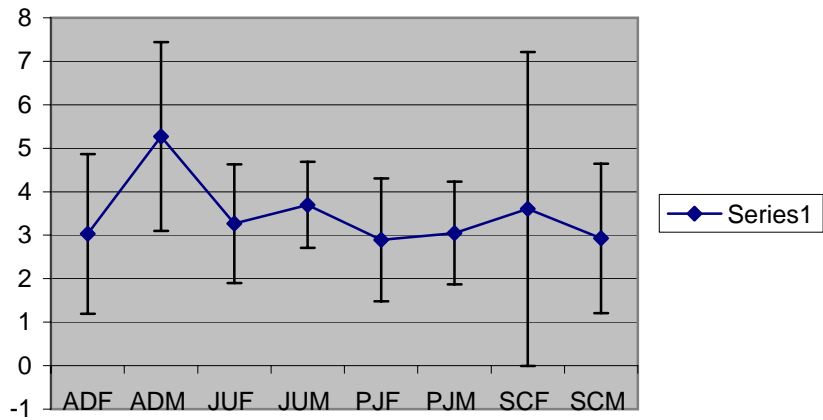
### 2102 Ipsative



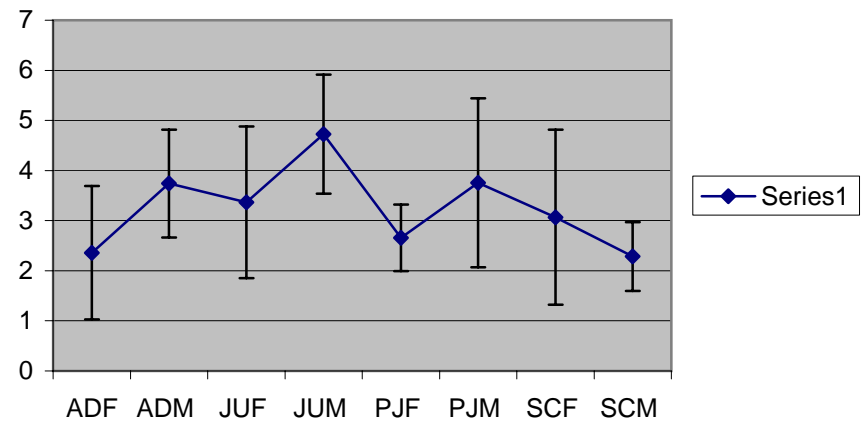
### 2103 Ipsative



### 2102 Raw Vt

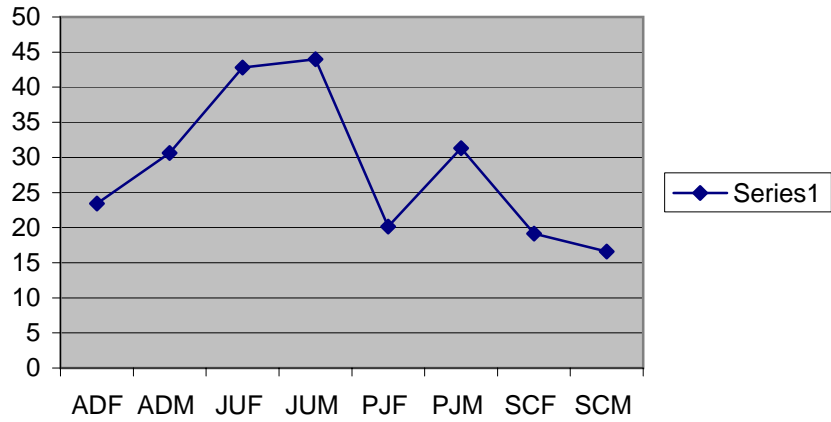


### 2103 Raw Vt

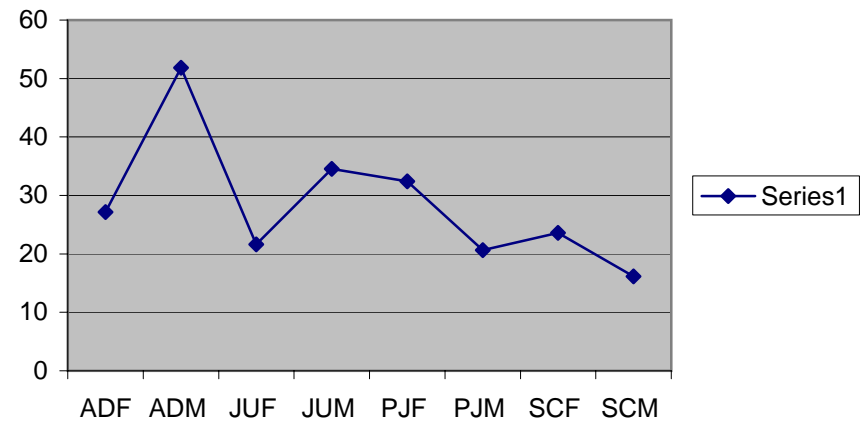




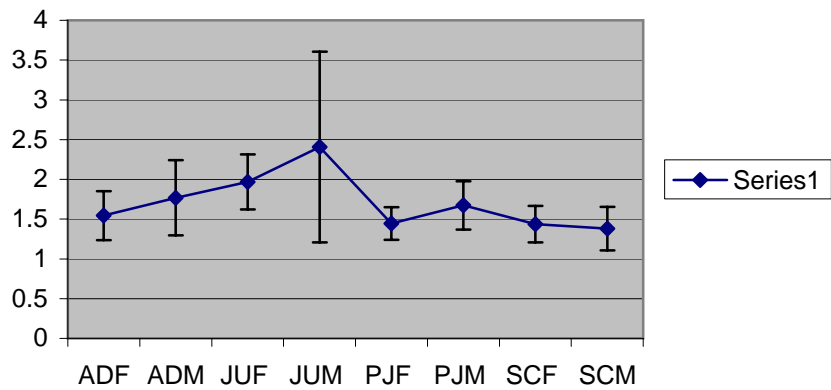
### 2104 Ipsative



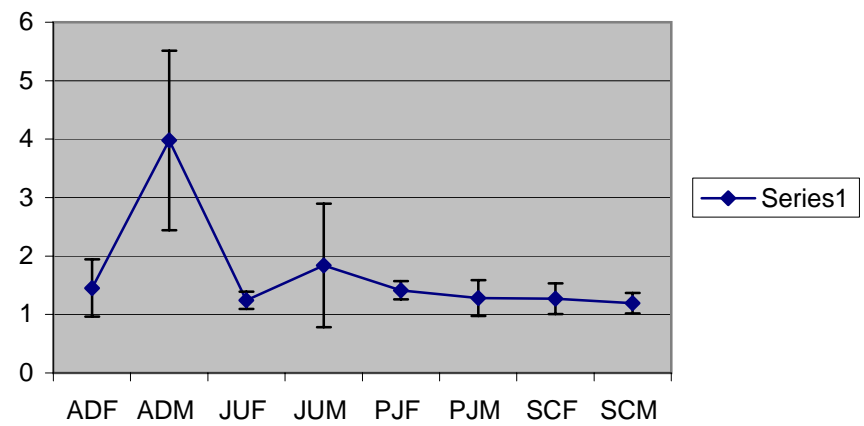
### 2105 Ipsative



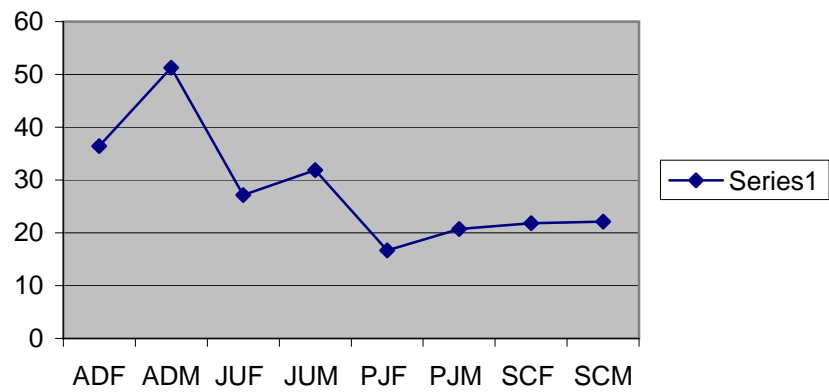
### 2104 Raw Vt



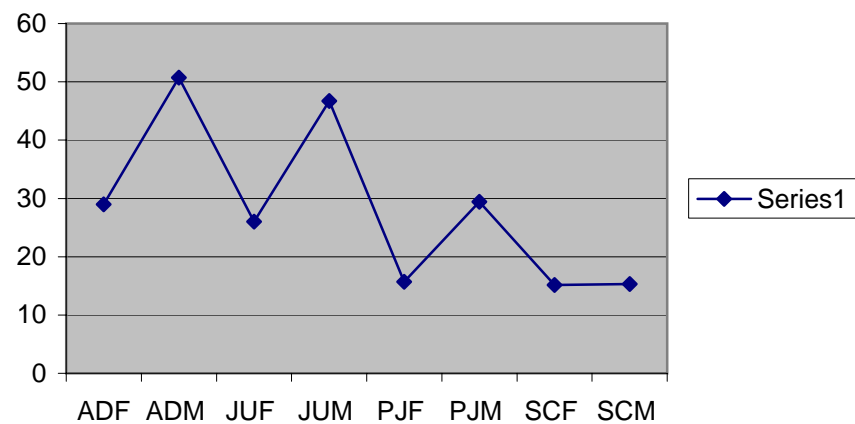
### 2105 Raw Vt



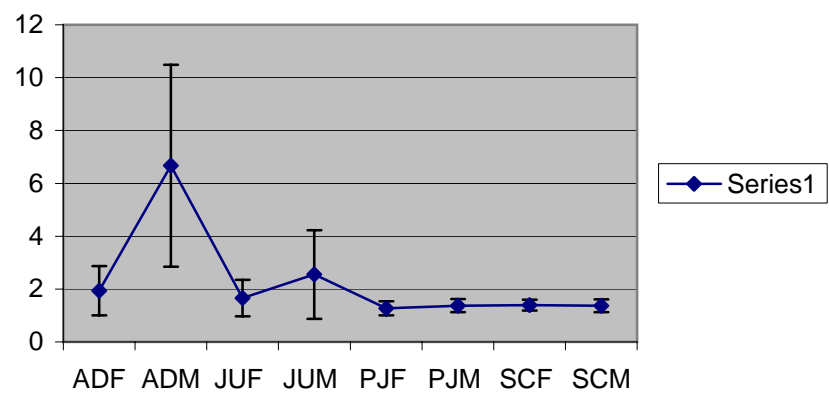
### 2106 Ipsative



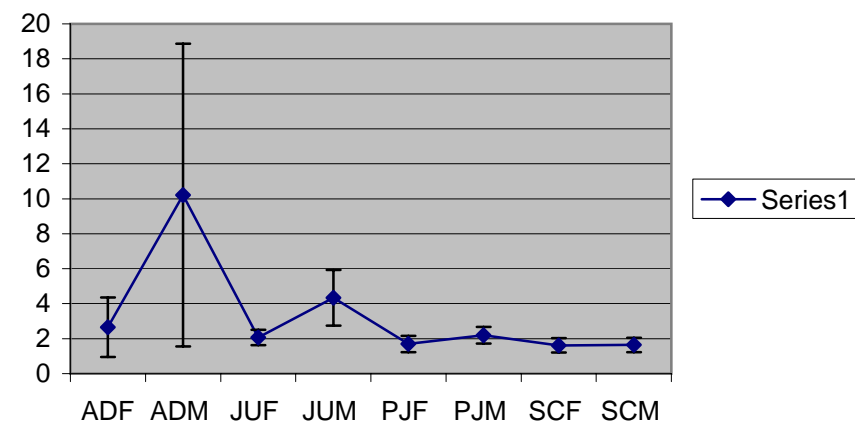
### 2107 Ipsative



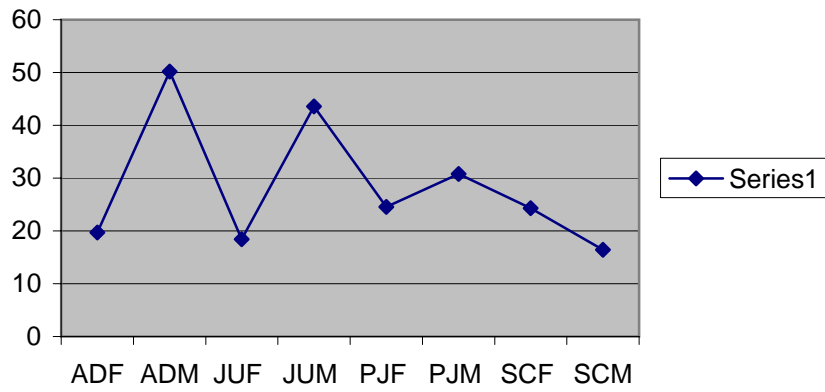
### 2106 Raw Vt



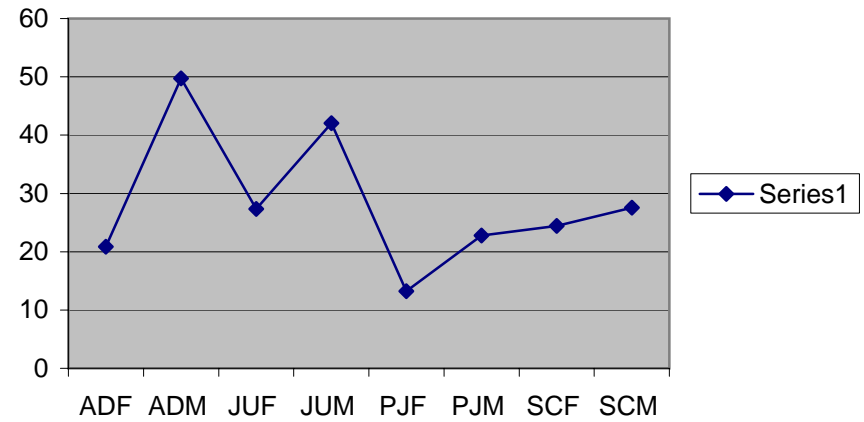
### 2107 Raw Vt



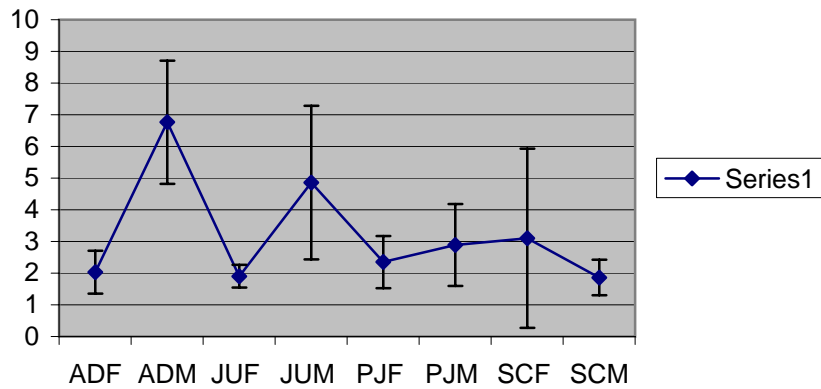
### 2108 Ipsative



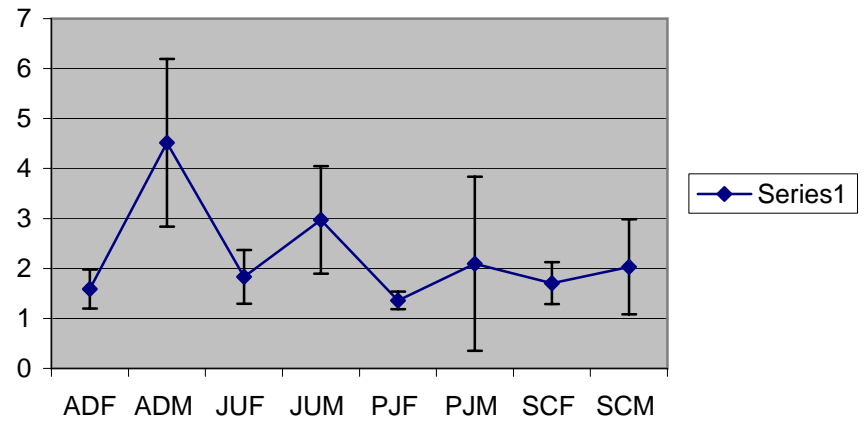
### 2109 Ipsative



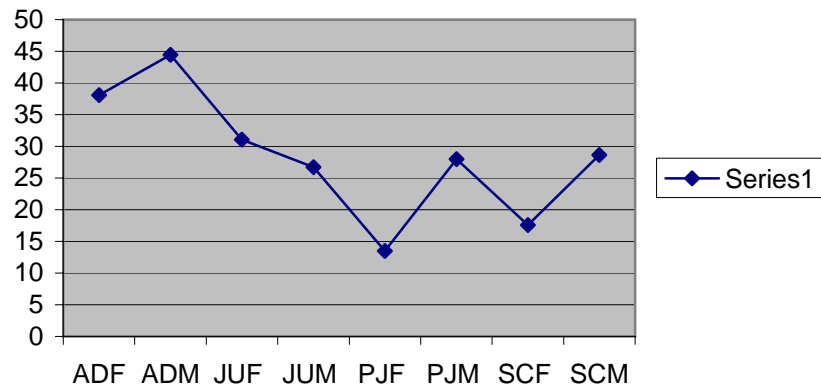
### 2108 Raw Vt



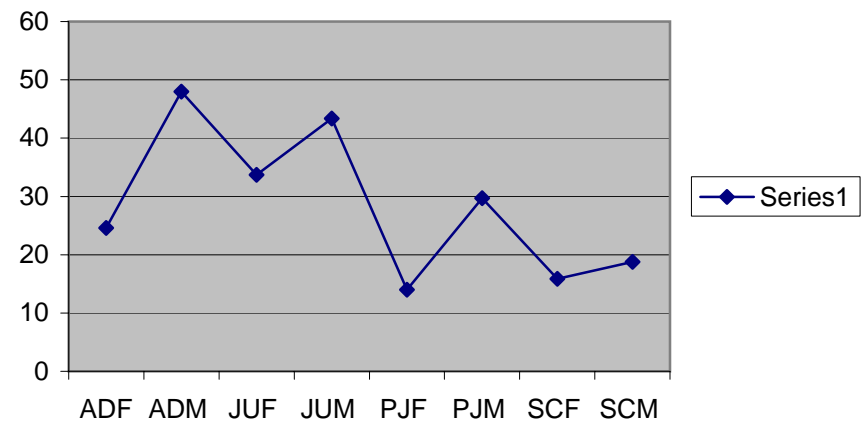
### 2109 Raw Vt



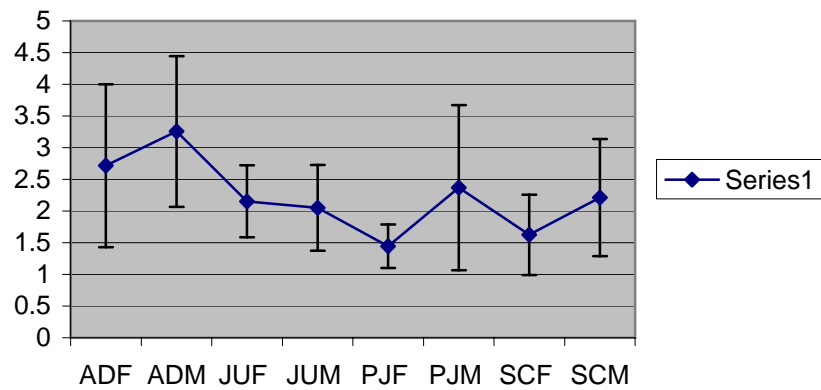
### 2110 Ipsative



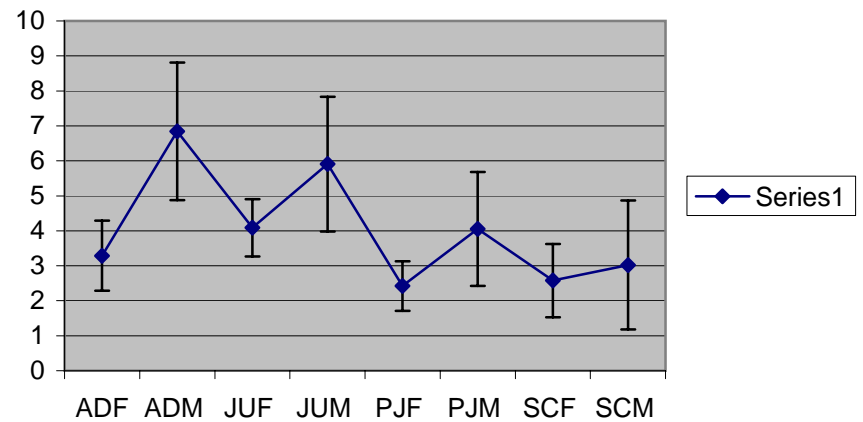
### 2111 Raw Vt



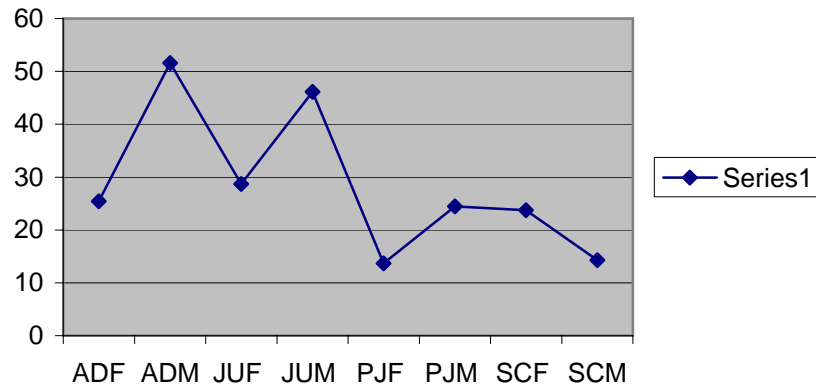
### 2110 Raw Vt



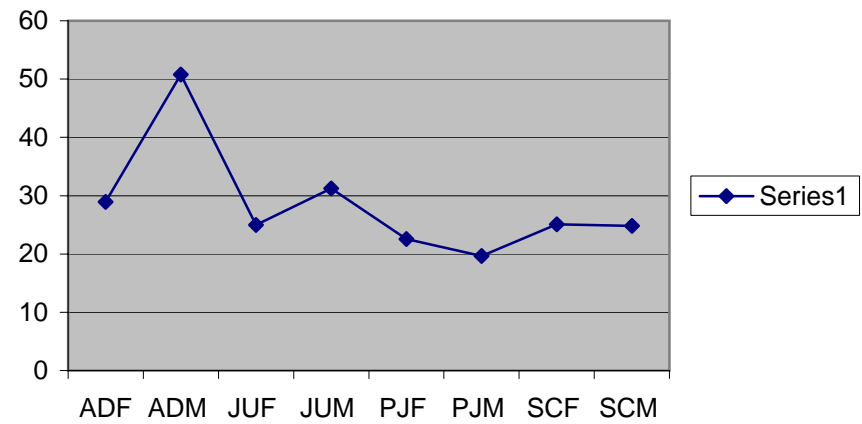
### 2111 Raw Vt



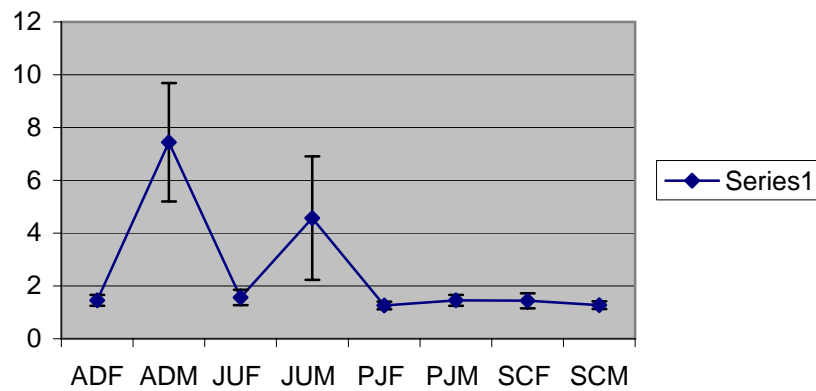
### 2112 Ipsative



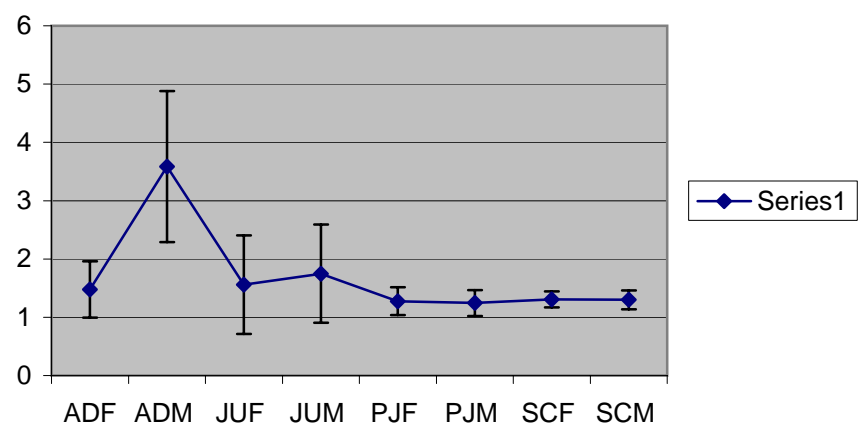
### 2113 Ipsative



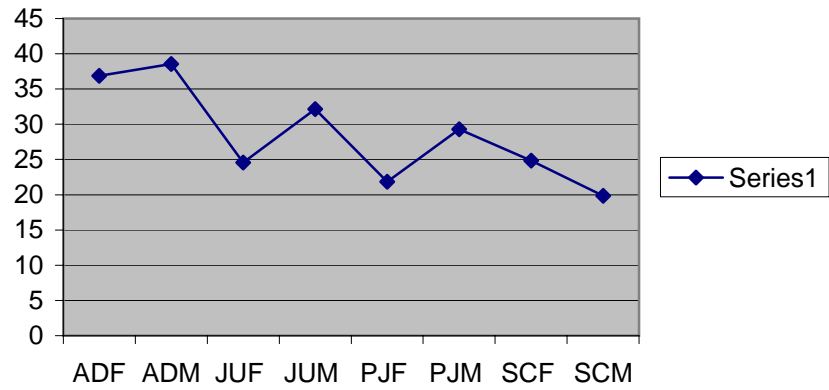
### 2112 Raw Vt



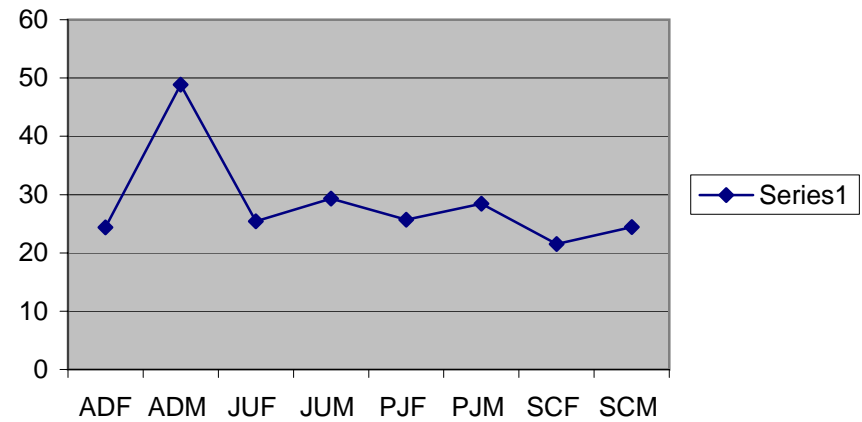
### 2113 Raw Vt



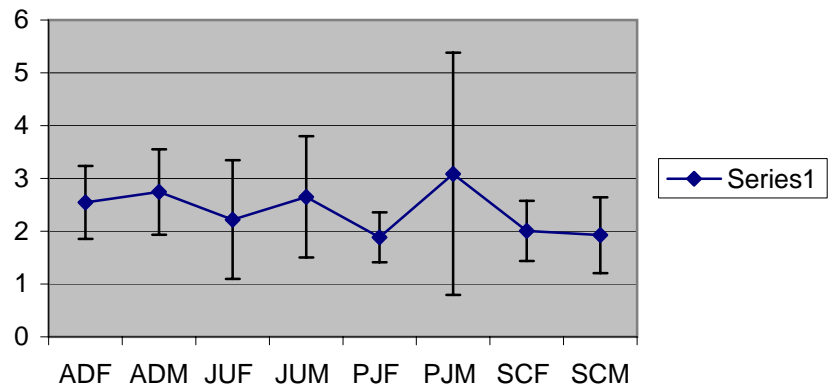
### 2114 Ipsative



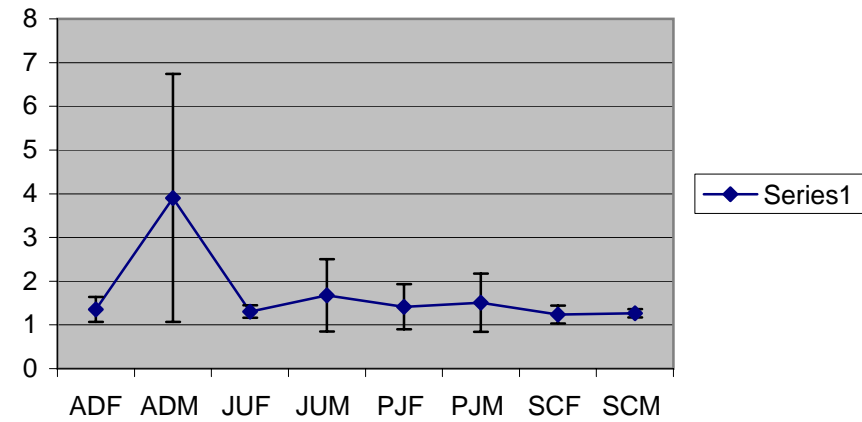
### 2115 Ipsative



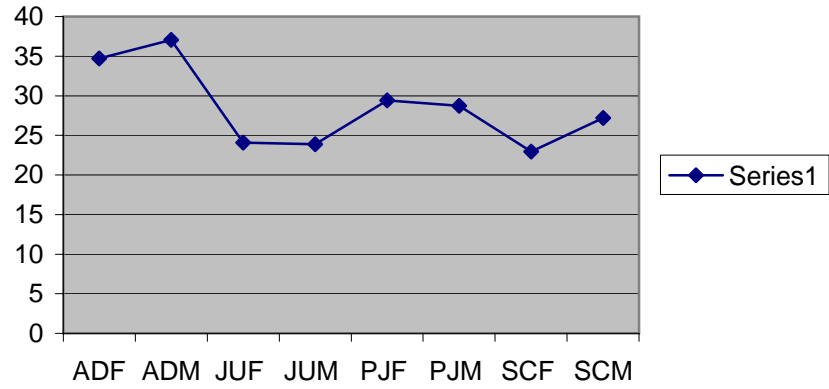
### 2114 Raw Vt



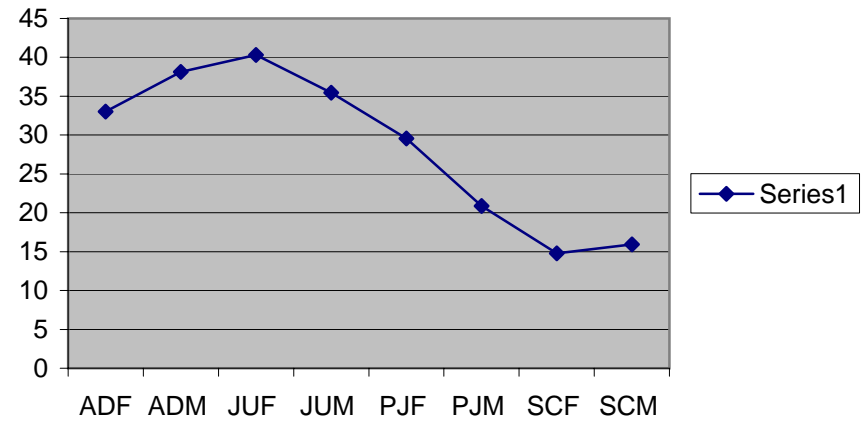
### 2115 Raw Vt



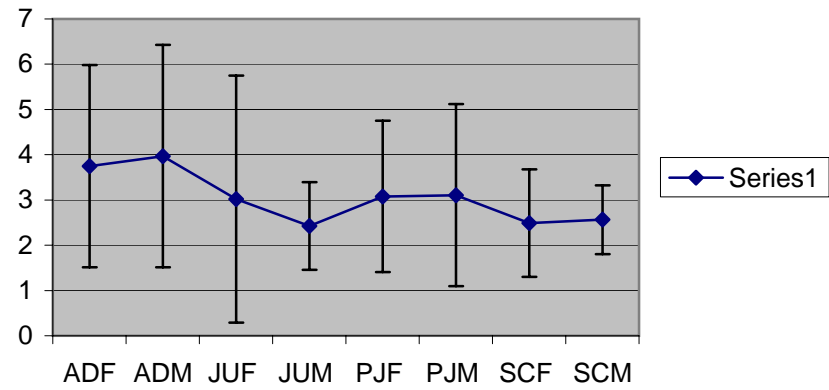
2116 Ipsative



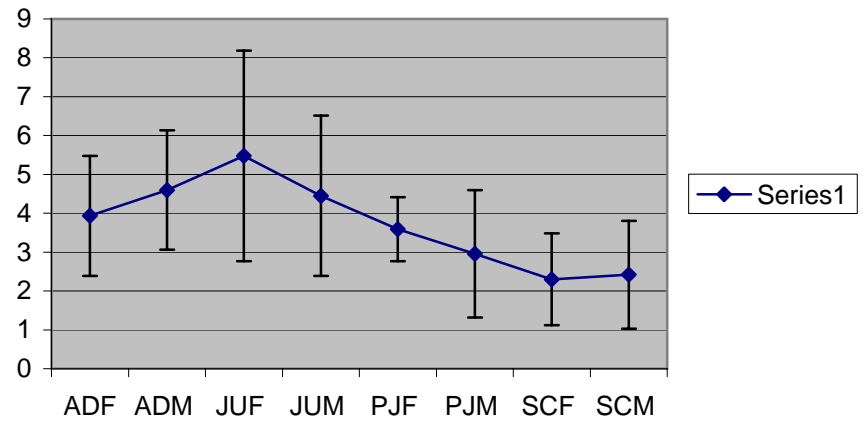
2117 Ipsative



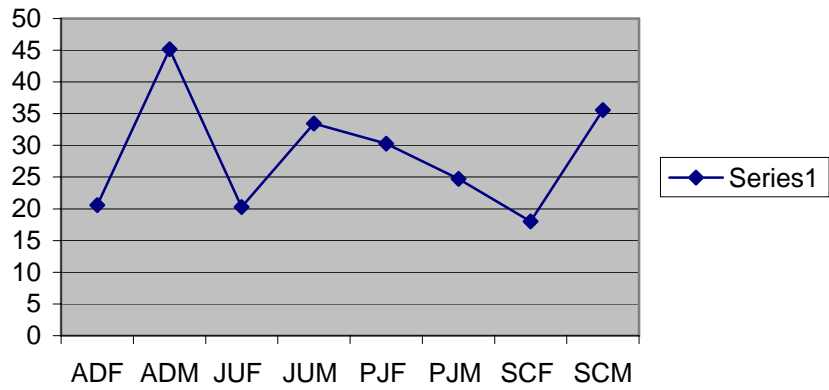
2116 Raw Vt



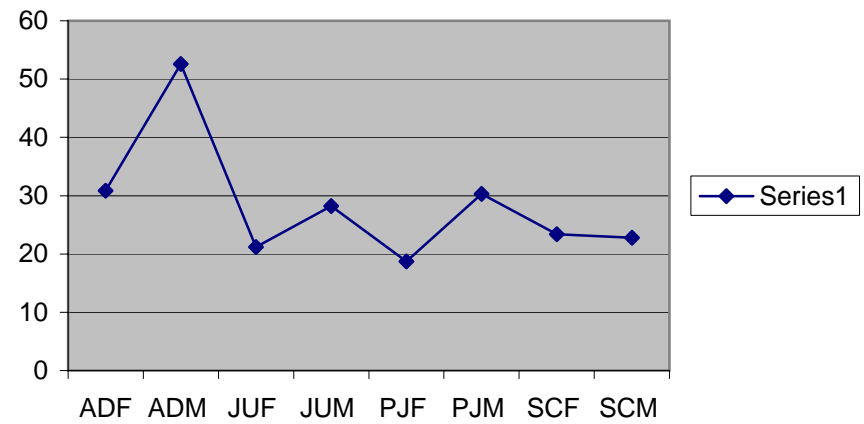
2117 Raw Vt



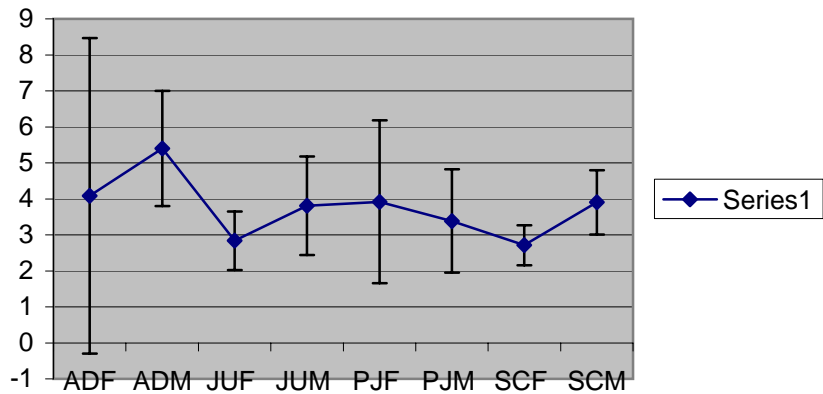
### 2118 Ipsative



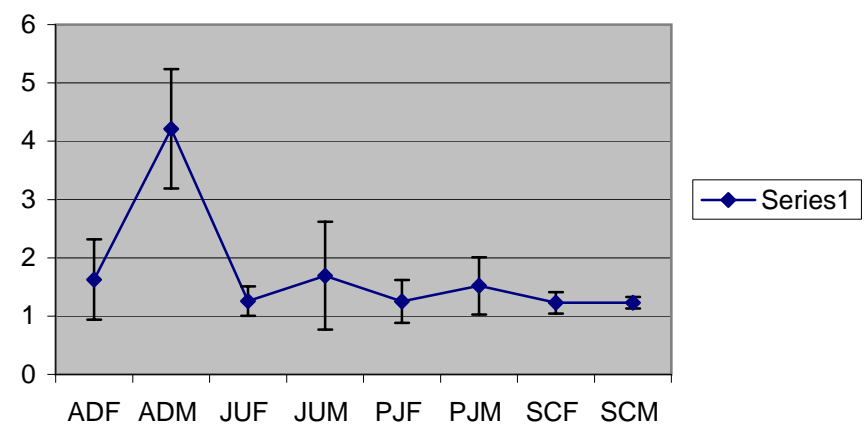
### 2119 Ipsative



### 2118 Raw Vt

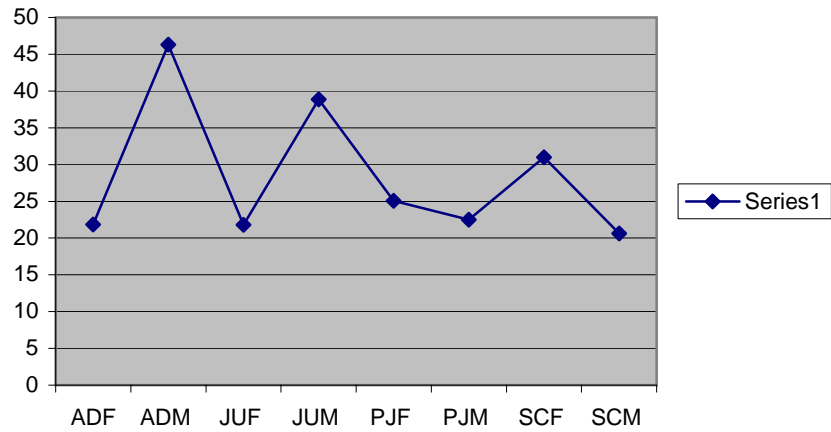


### 2119 Raw Vt

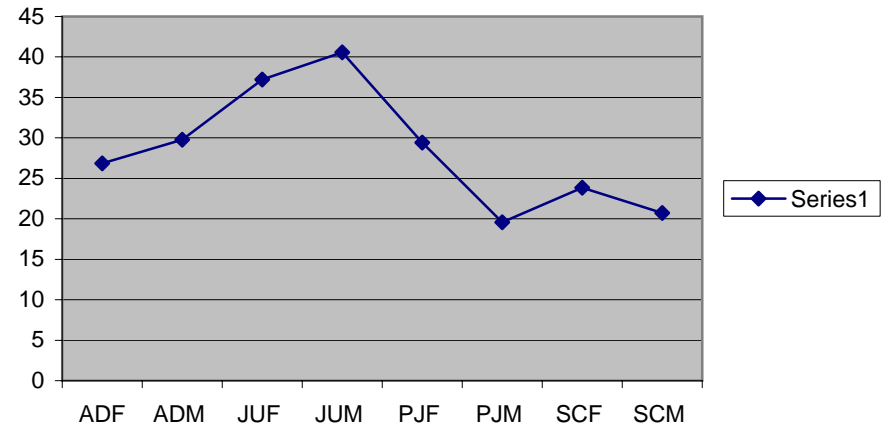




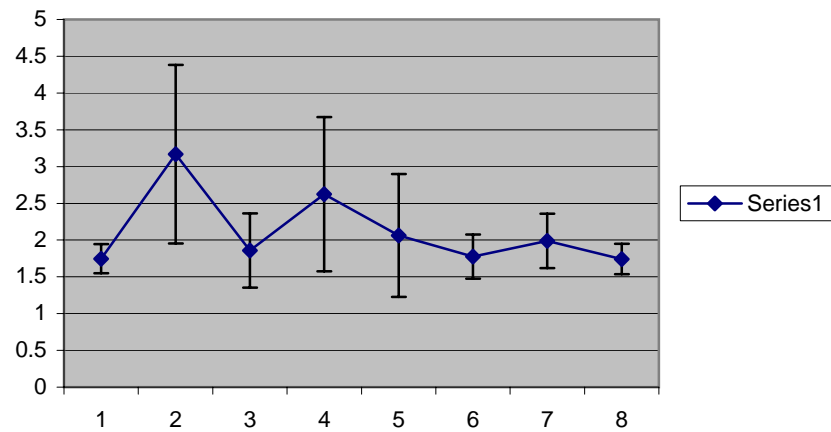
2120 Ipsative



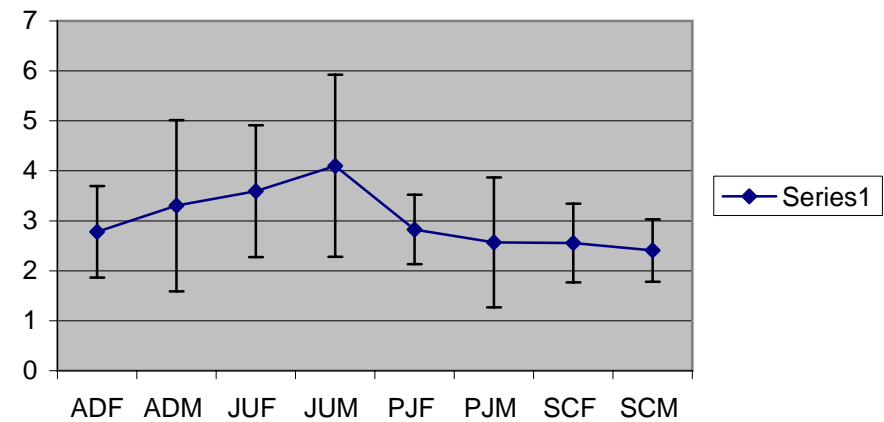
2121 Ipsative



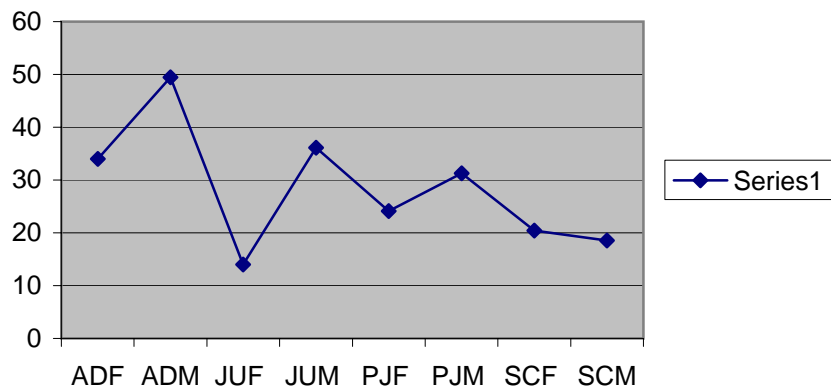
2120 Raw Vt



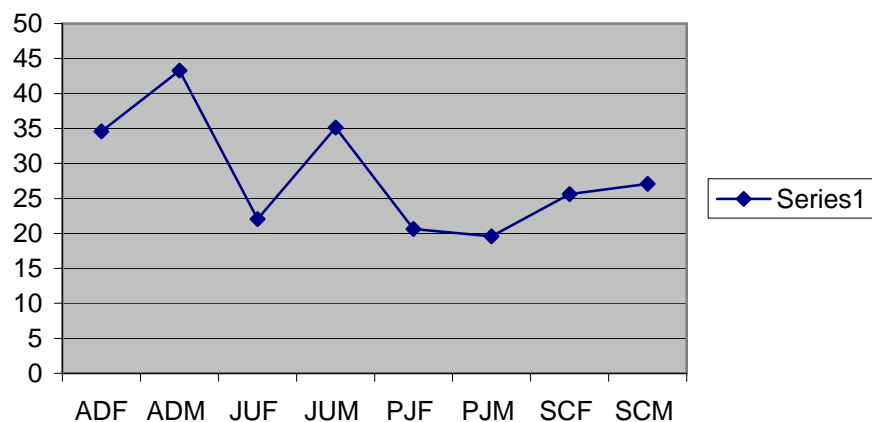
2121 Raw Vt



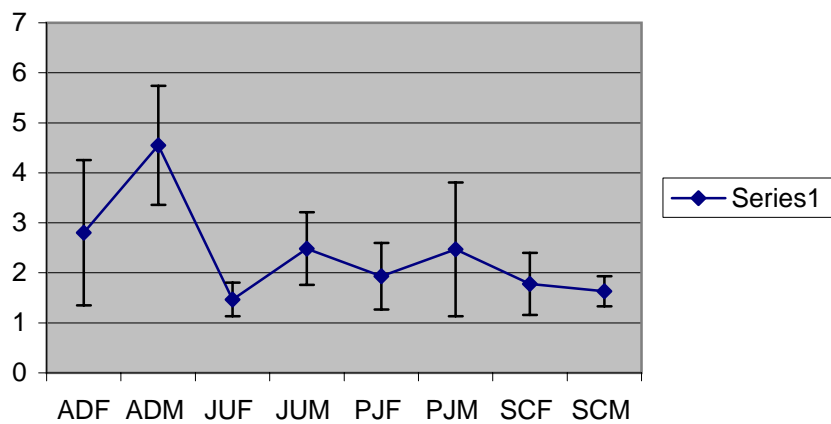
2122 Ipsative



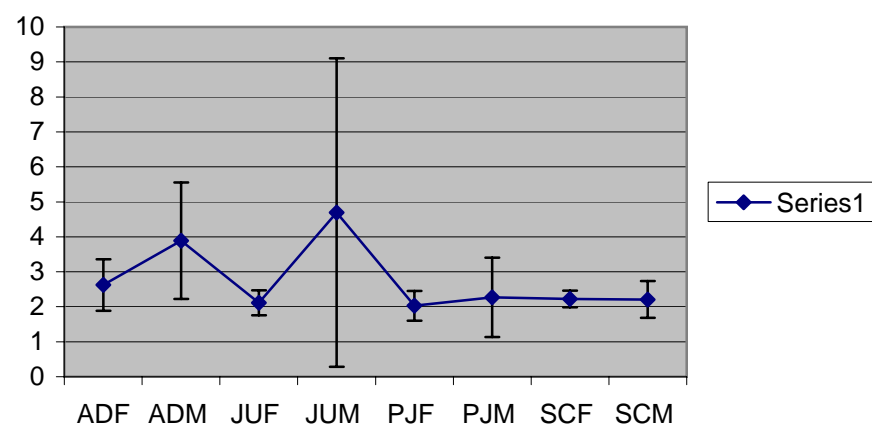
2123 Ipsative



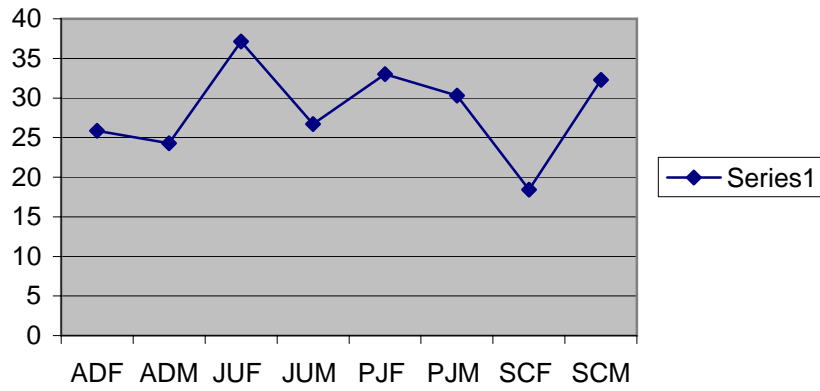
2122 Raw Vt



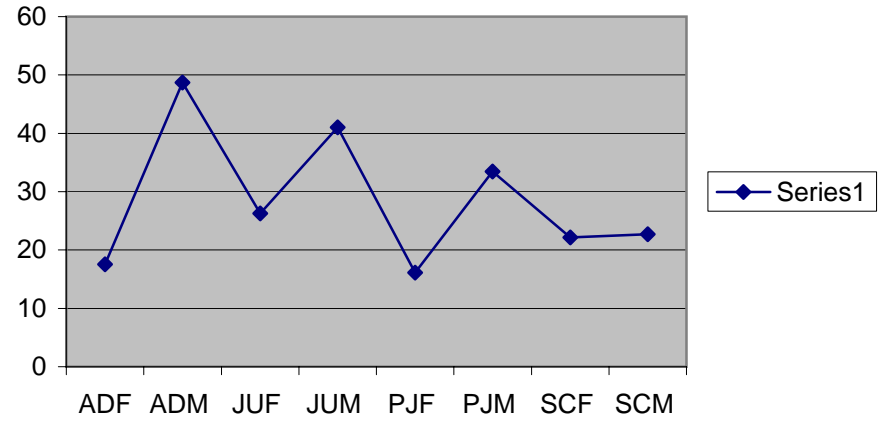
2123 Raw Vt



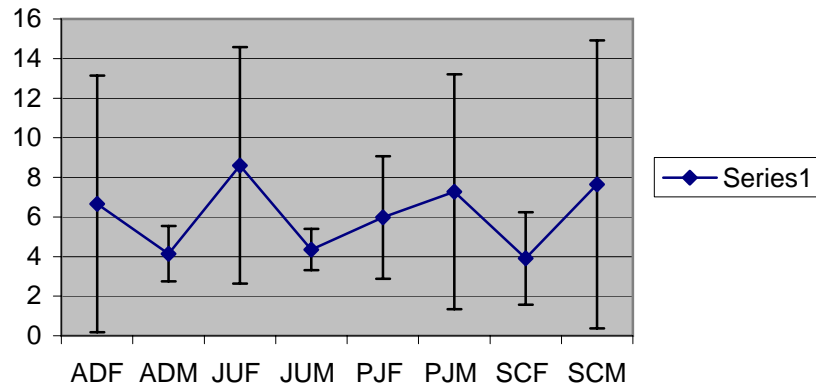
### 2124 Ipsative



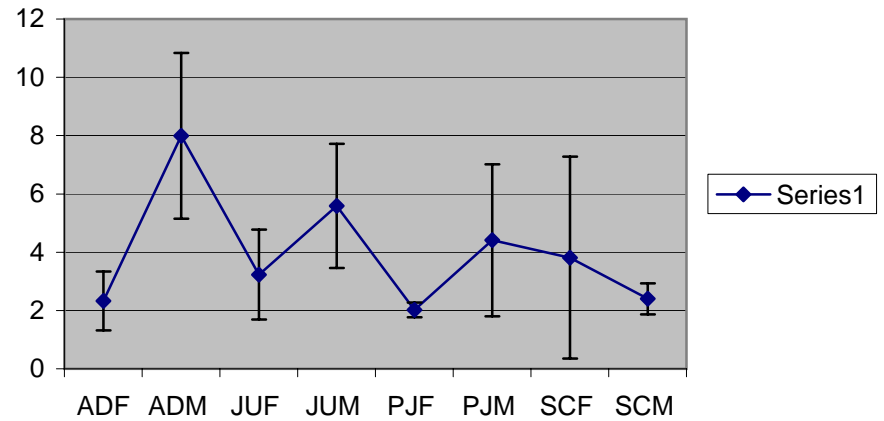
### 2125 Ipsative



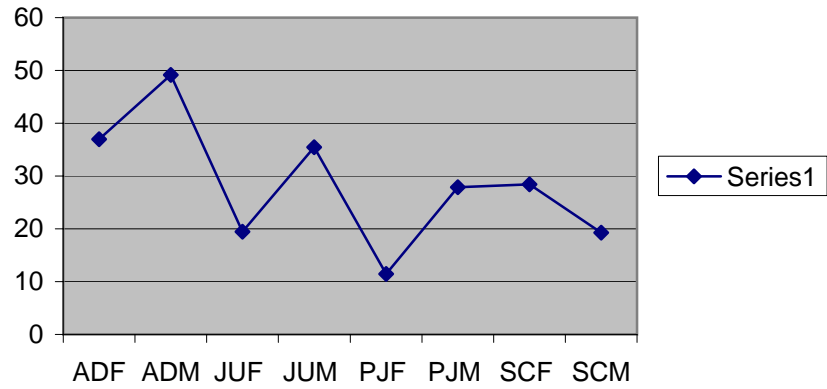
### 2124 Raw Vt



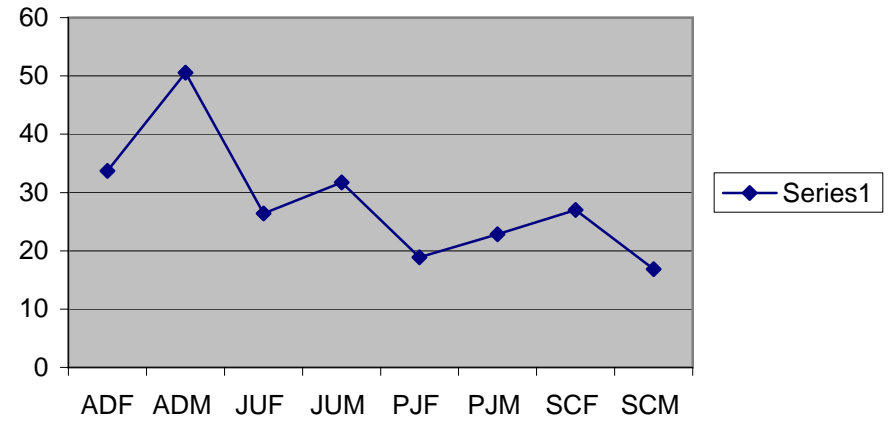
### 2125 Raw Vt



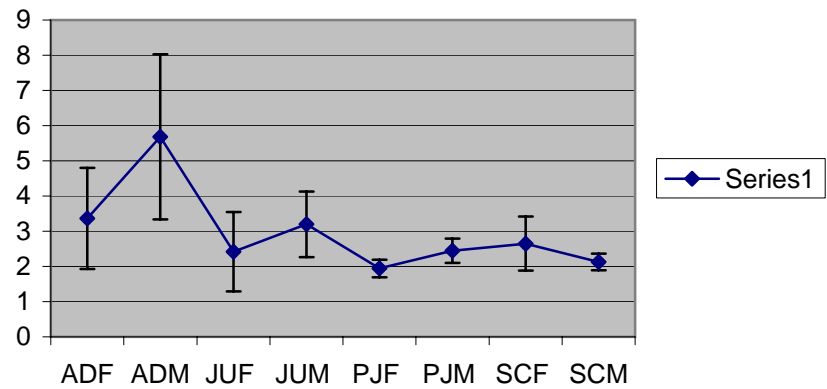
2126 Ipsative



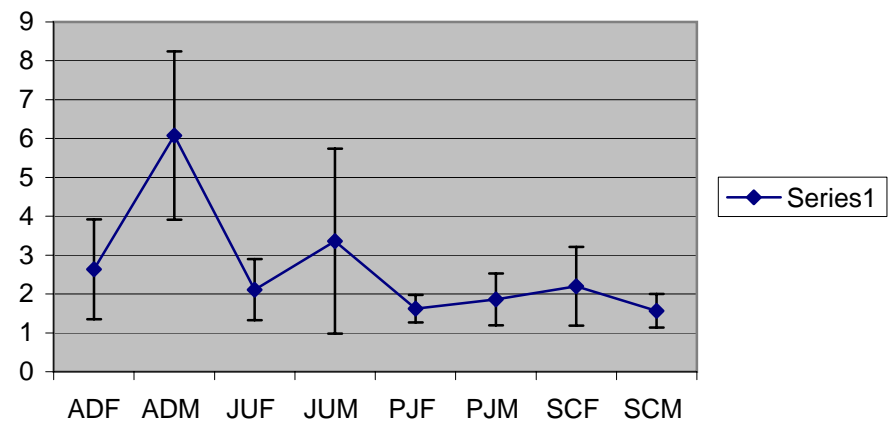
2127 Ipsative



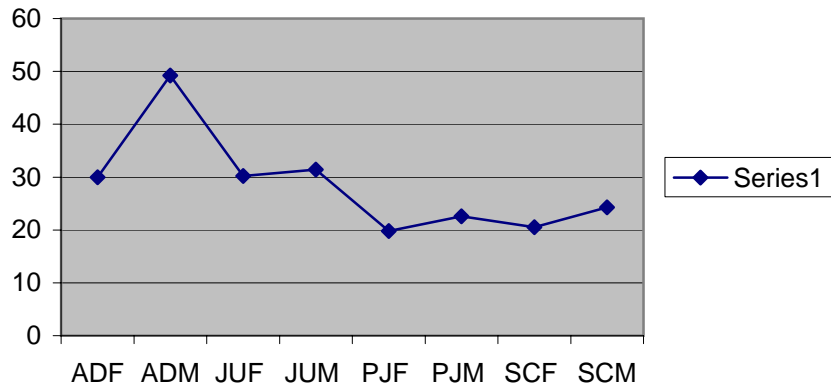
2126 Raw Vt



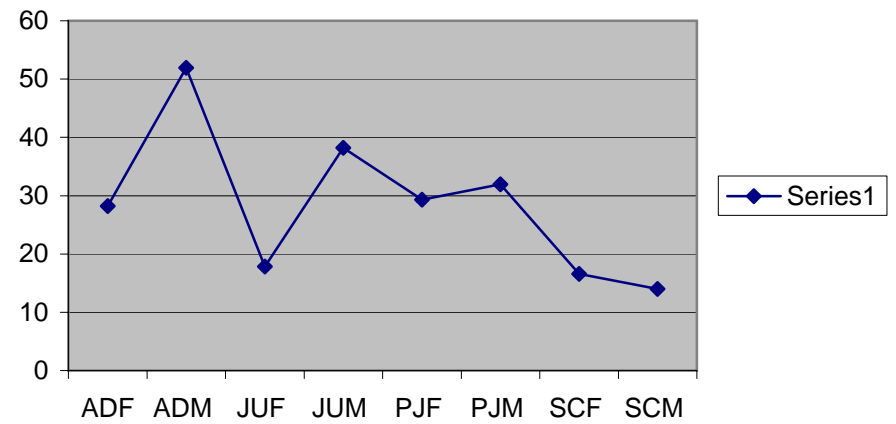
2127 Raw Vt



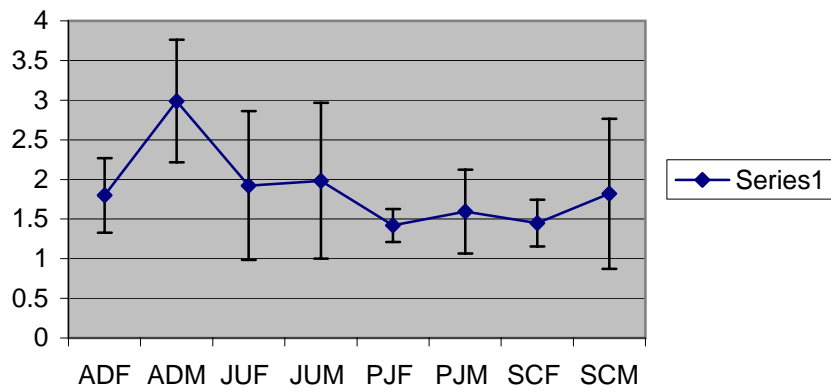
### 2128 Ipsative



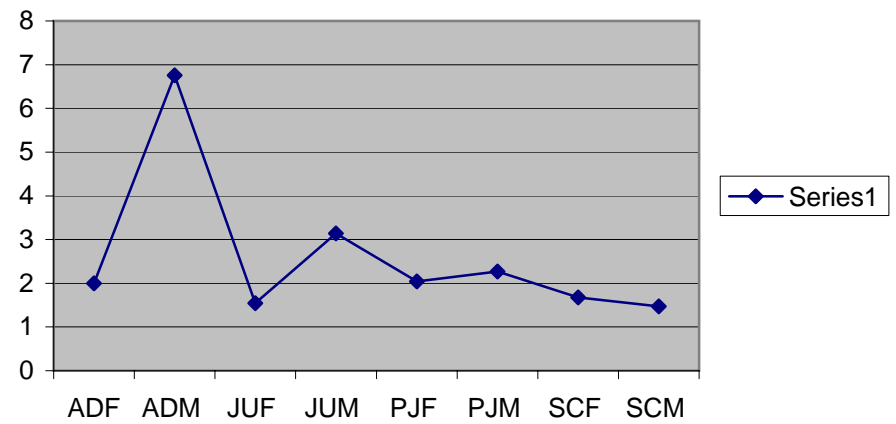
### 2129 Ipsative



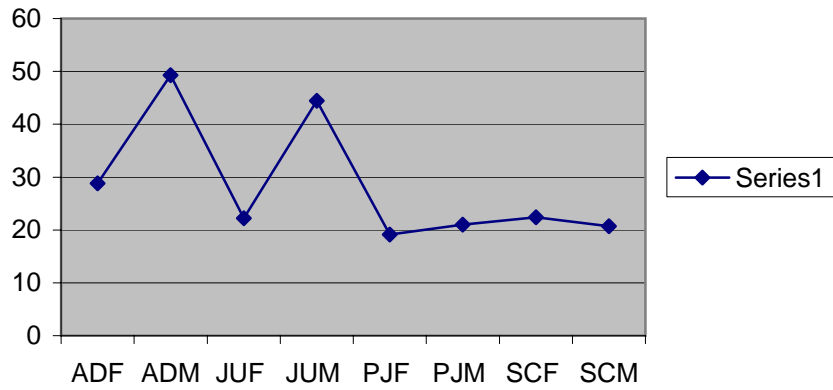
### 2128 Raw Vt



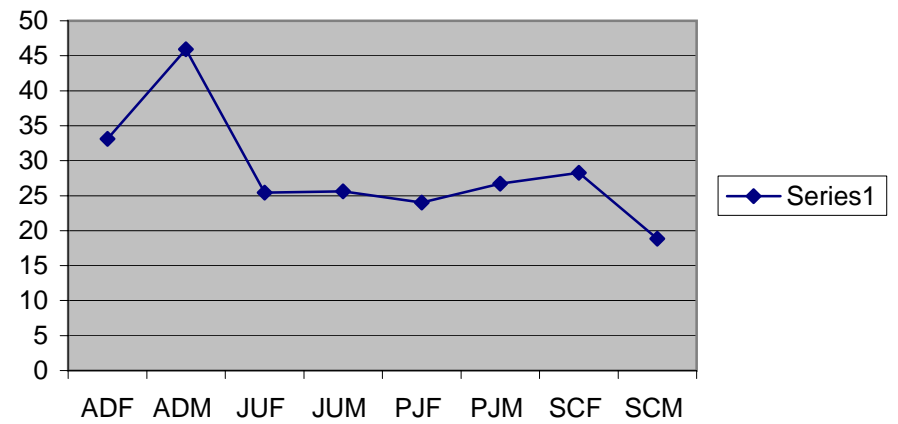
### 2129 Raw Vt



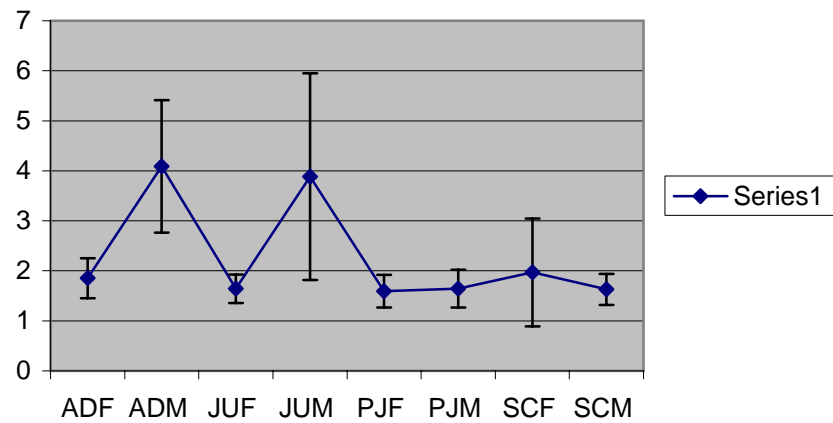
### 2130 Ipsative



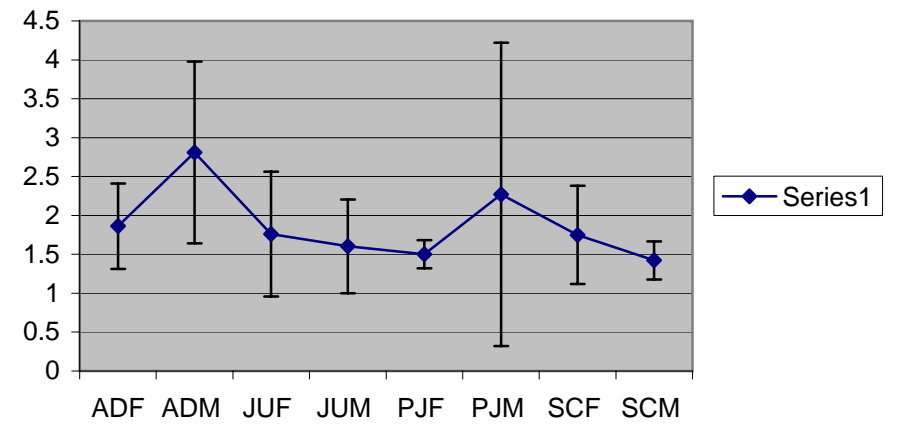
### 2131 Ipsative



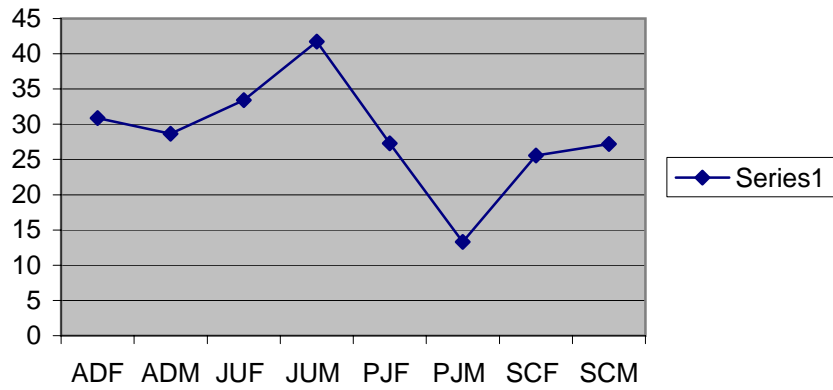
### 2130 Raw Vt



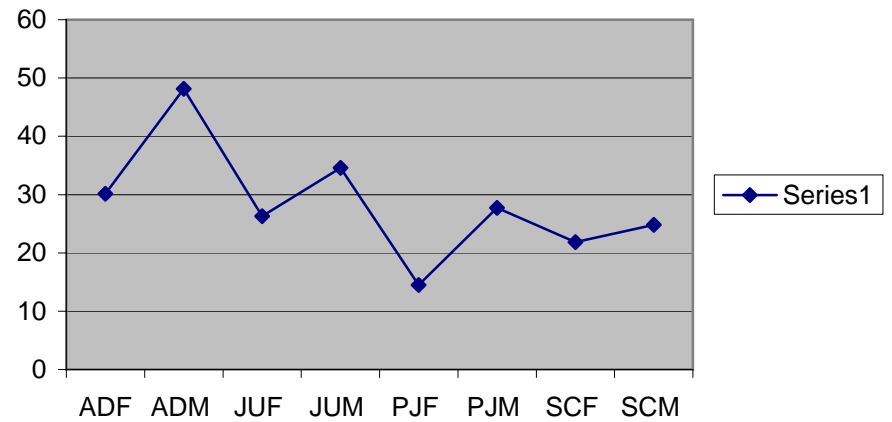
### 2131 Raw Vt



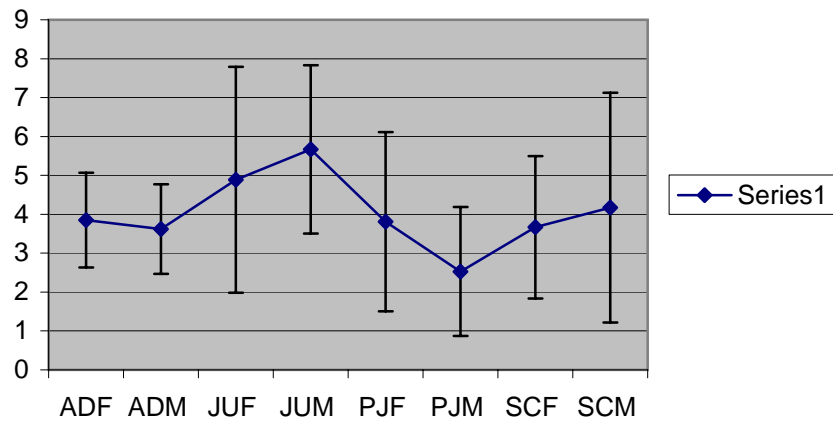
### 2132 Ipsative



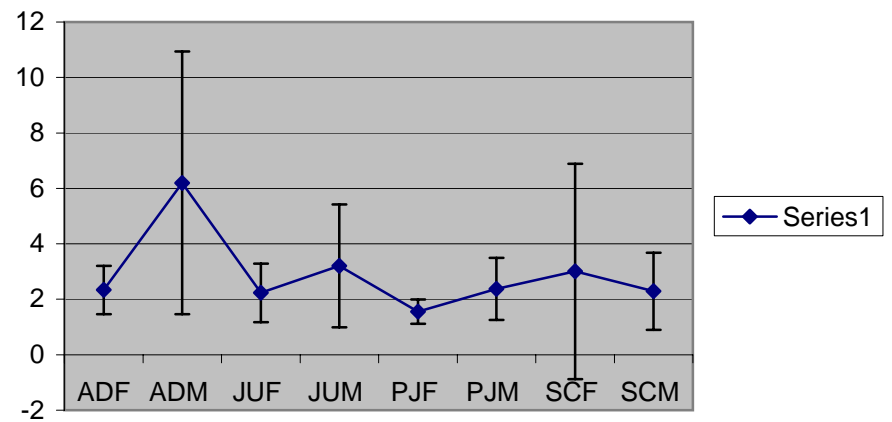
### 2133 Ipsative



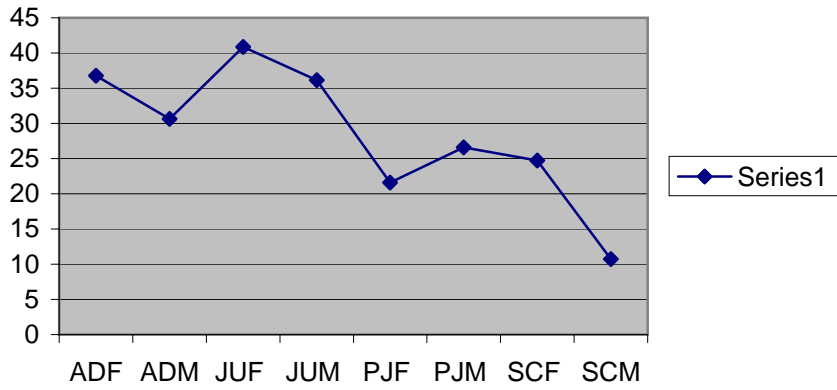
### 2132 Raw Vt



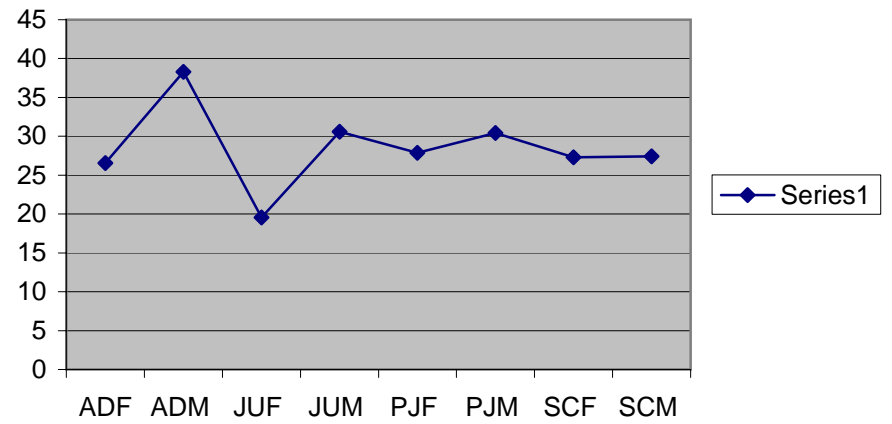
### 2133 Raw Vt



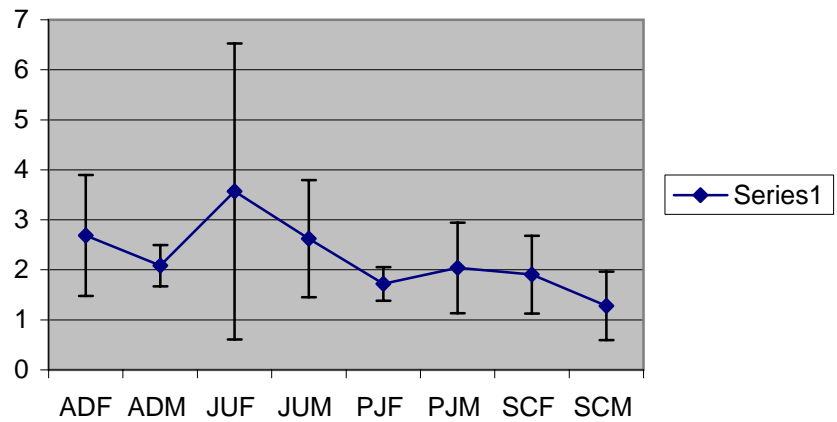
2134 Ipsative



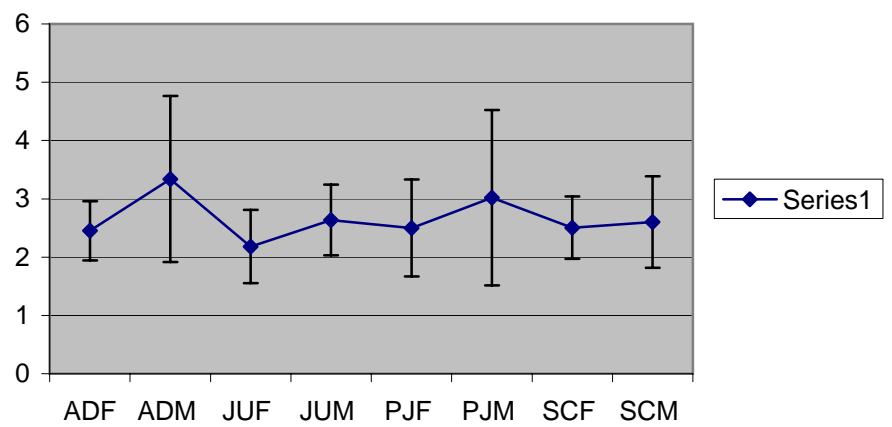
2135 Ipsative



2134 Raw Vt

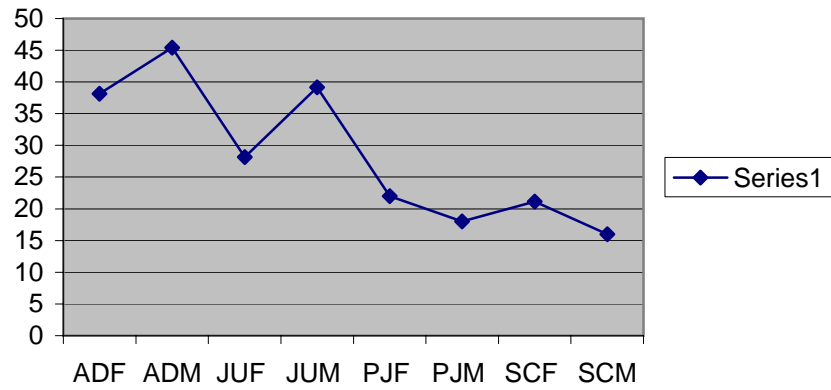


2135 Raw Vt

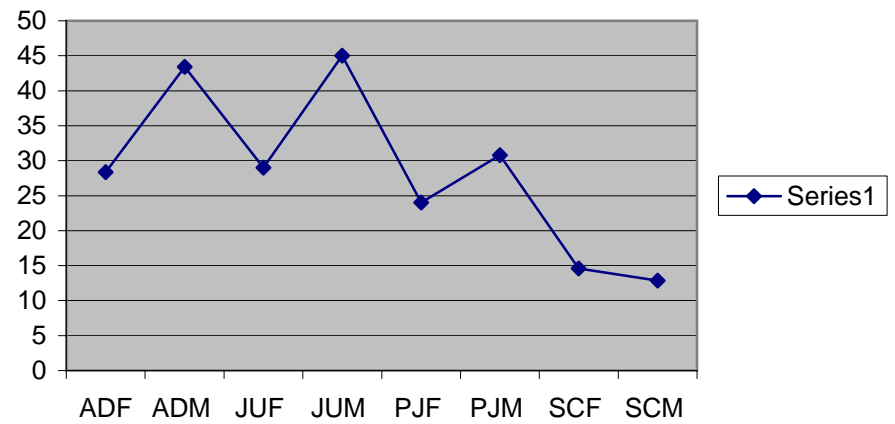




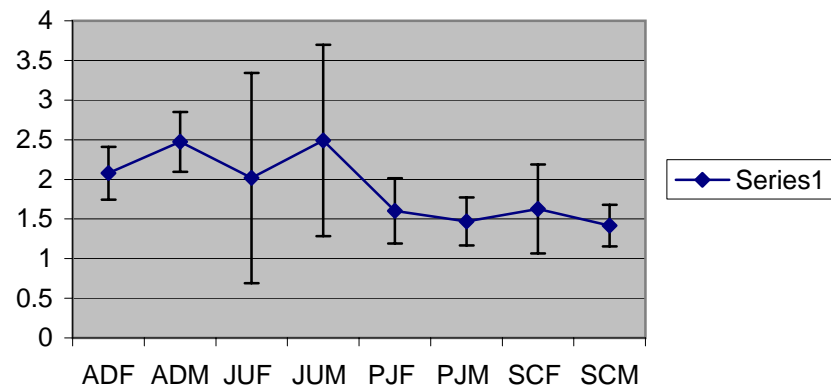
### 2137 Ipsative



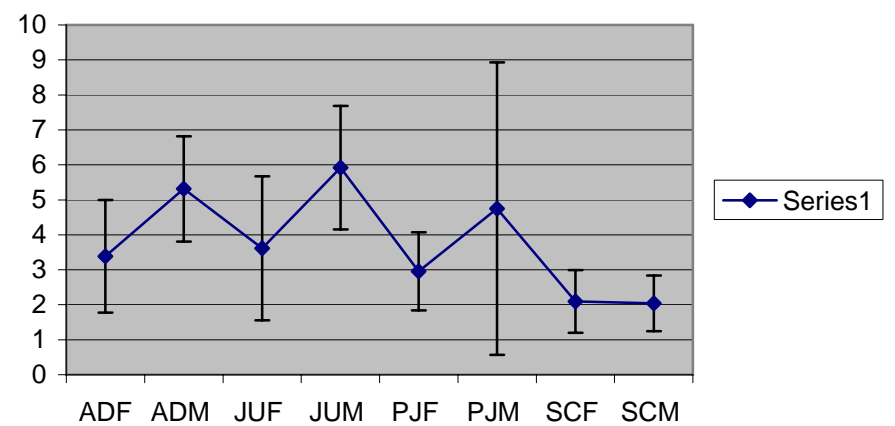
### 2138 Ipsative



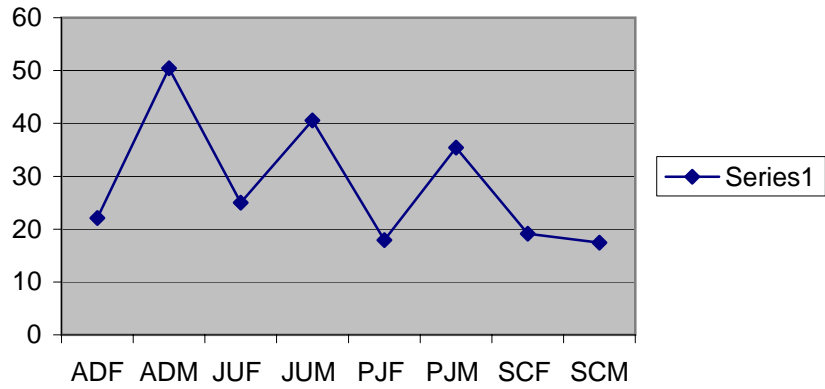
### 2137 Raw Vt



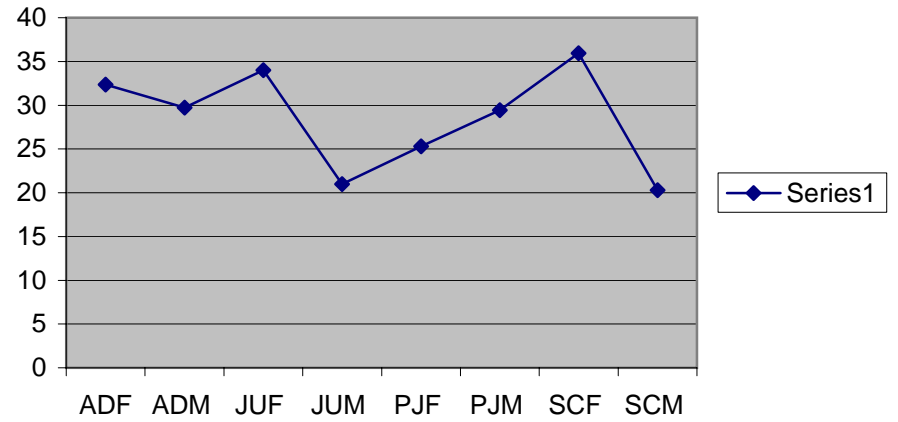
### 2138 Raw Vt



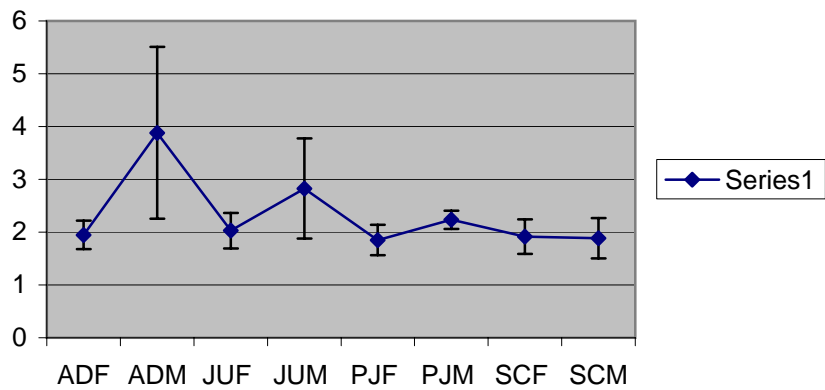
### 2139 Ipsative



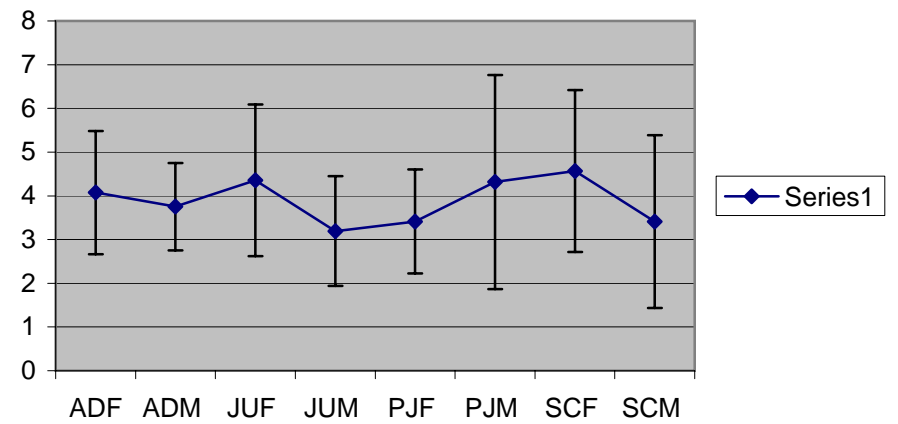
### 2140 Ipsative



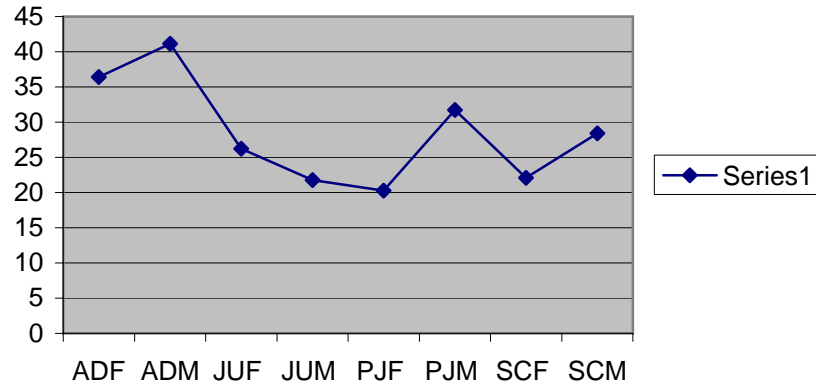
### 2139 Raw Vt



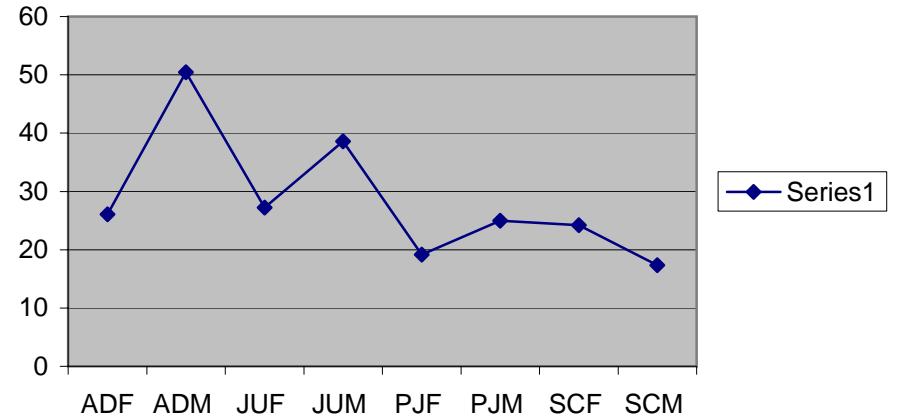
### 2140 Raw Vt



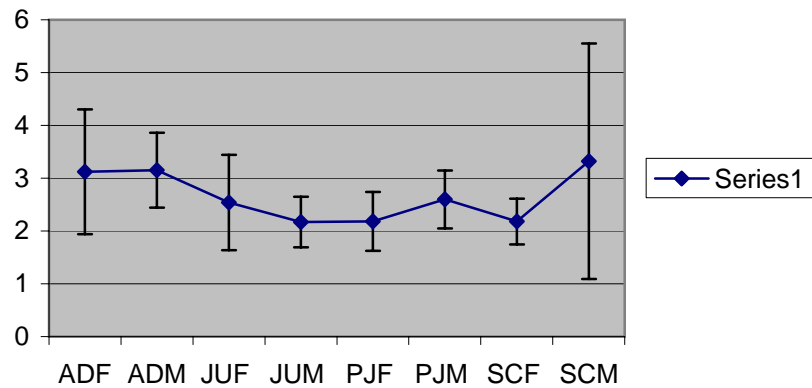
### 2141 Ipsative



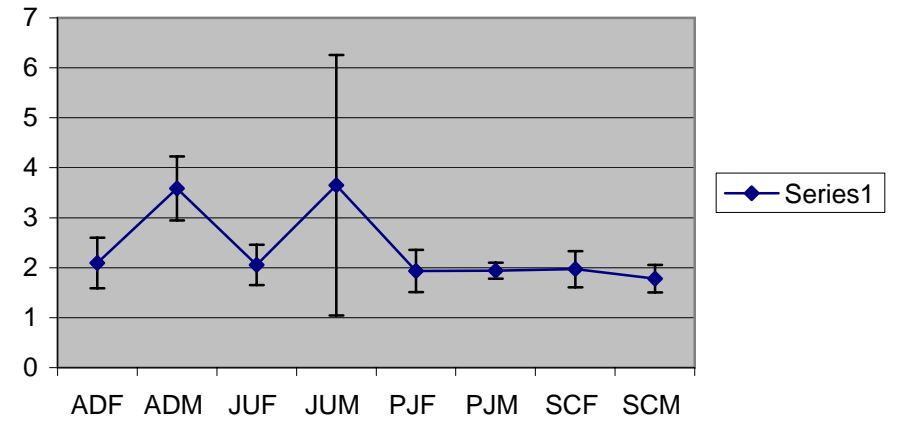
### 2142 Ipsative



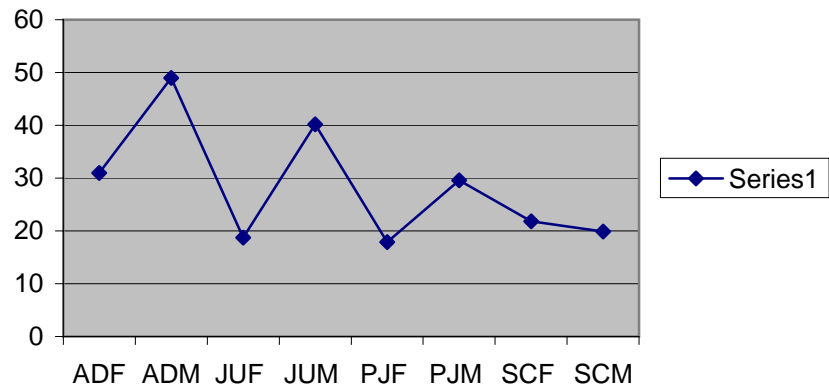
### 2141 Raw Vt



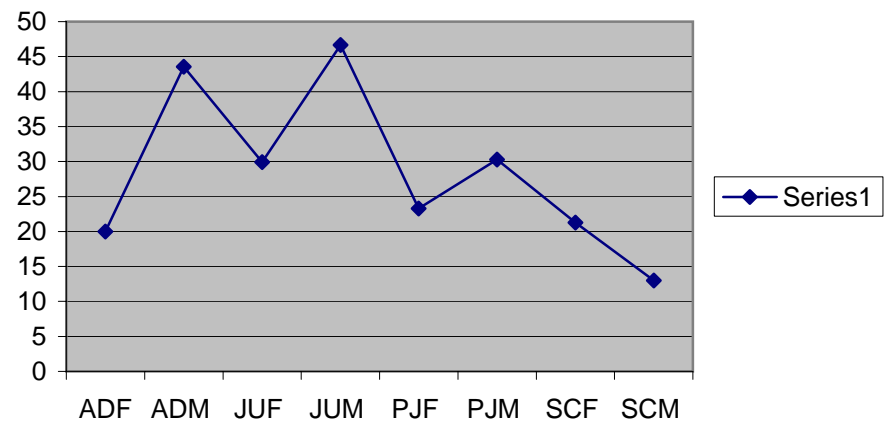
### 2142 Raw Vt



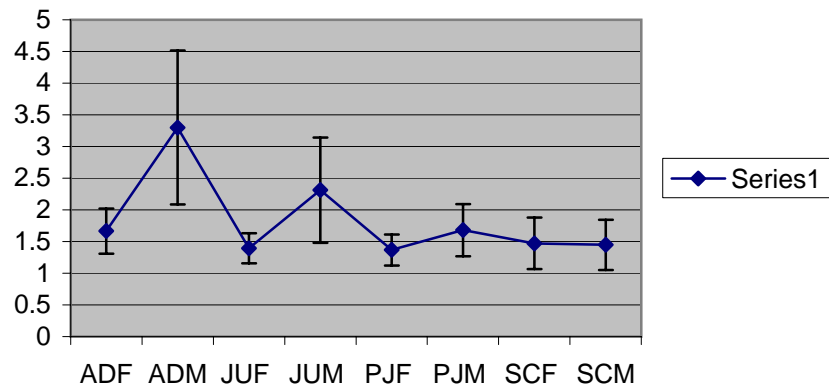
### 2143 Ipsative



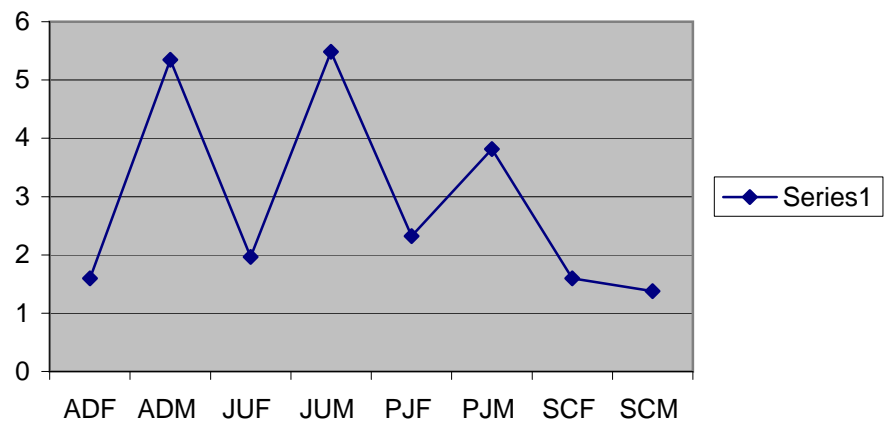
### 2144 Ipsative



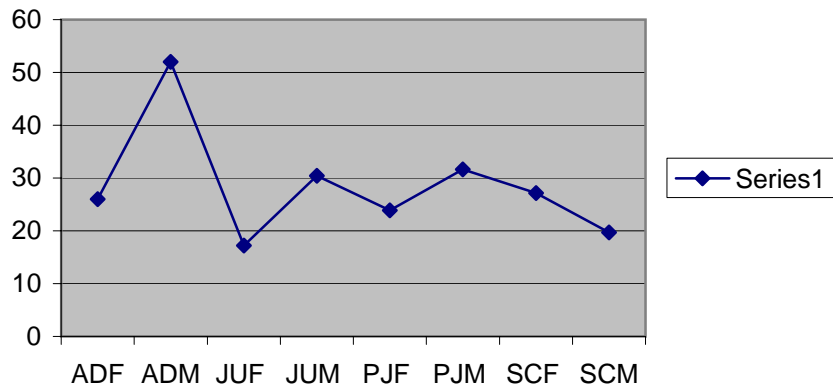
### 2143 Raw Vt



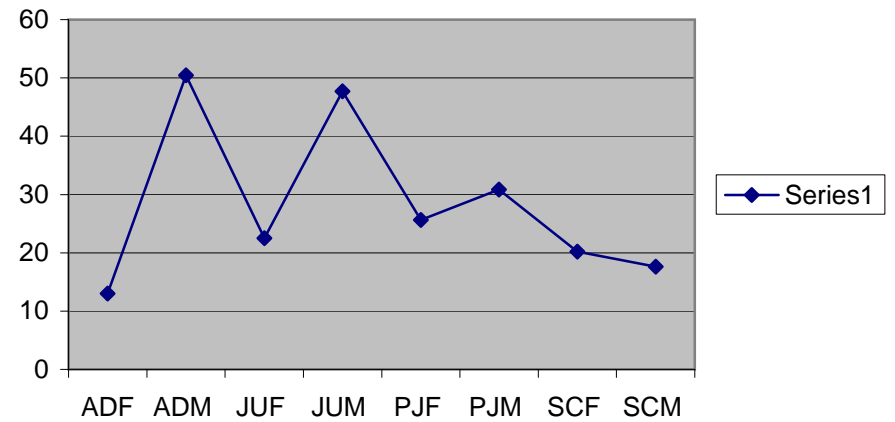
### 2144 Raw Vt]



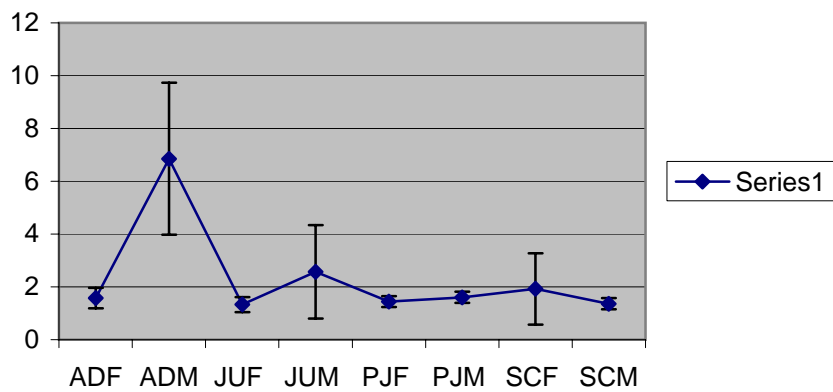
### 2145 Ipsative



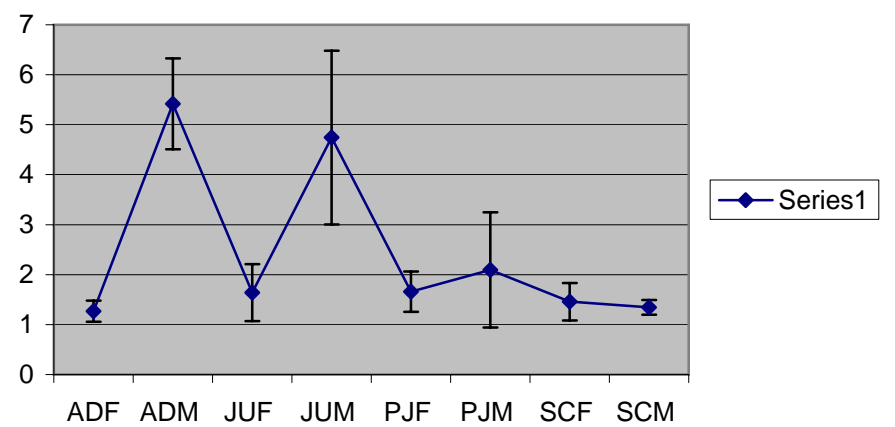
### 2146 Ipsative



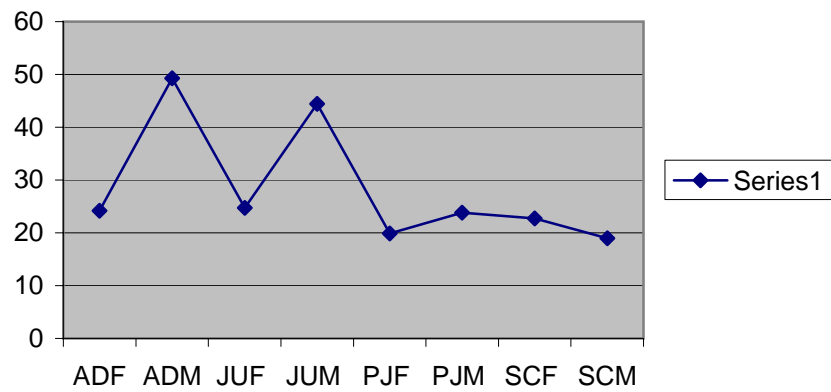
### 2145 Raw Vt



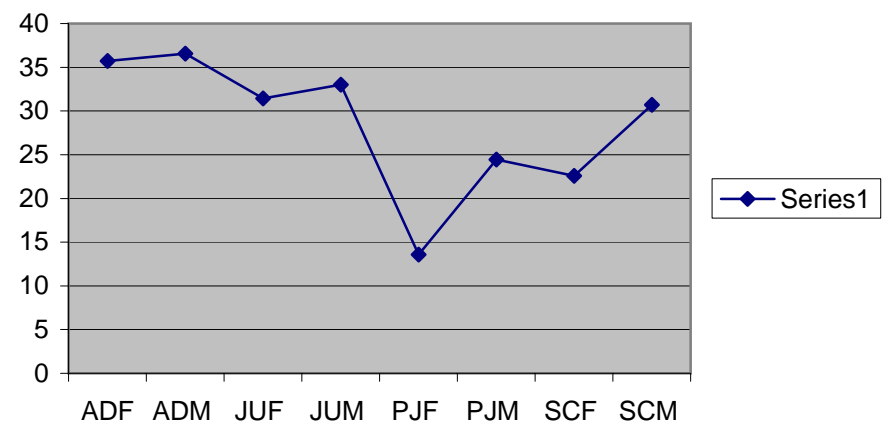
### 2146 Raw Vt



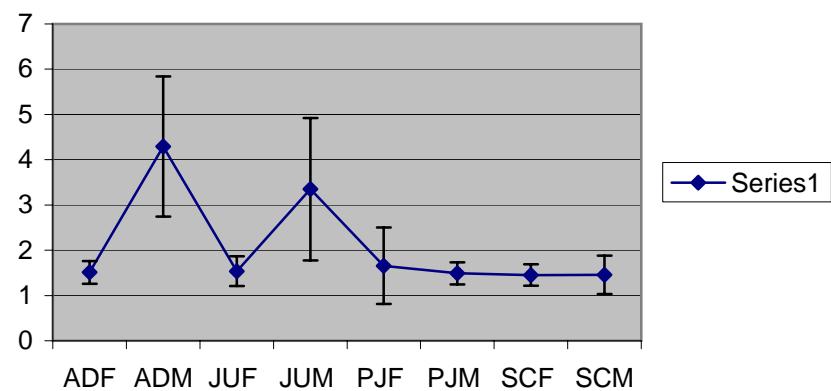
### 2147 Ipsative



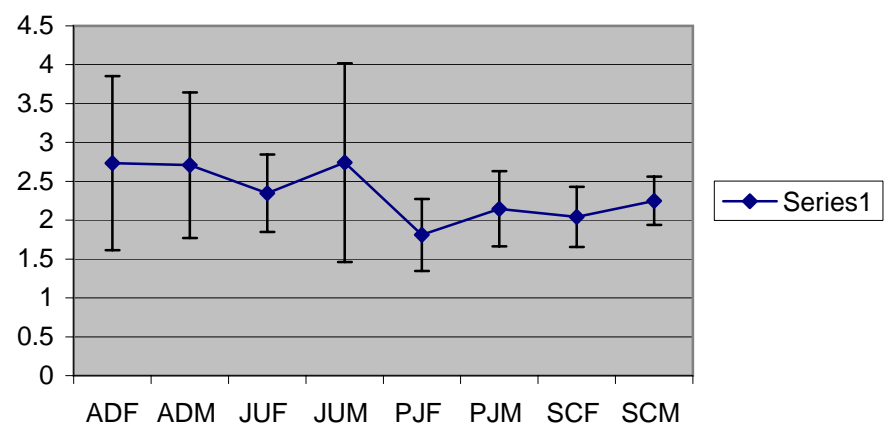
### 2148 Ipsative



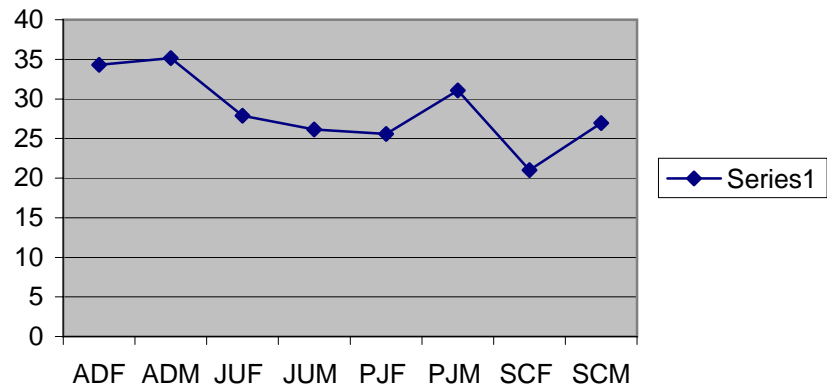
### 2147 Raw Vt



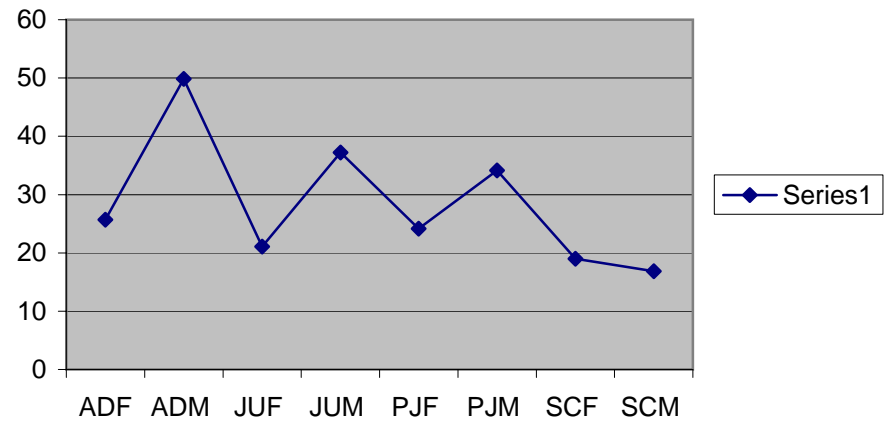
### 2148 Raw Vt



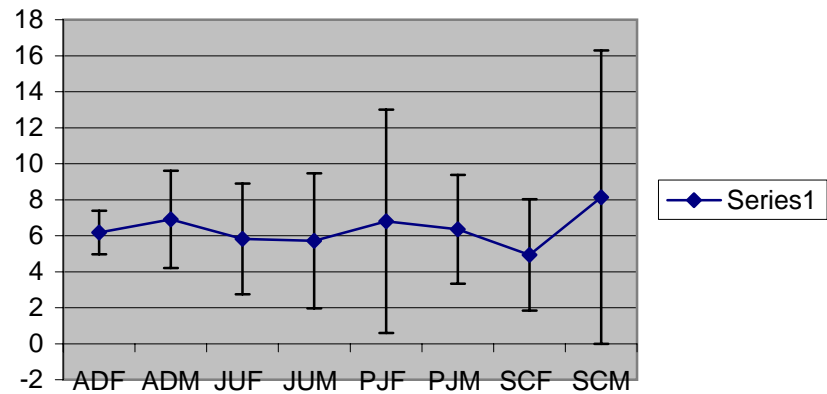
### 2149 Ipsative



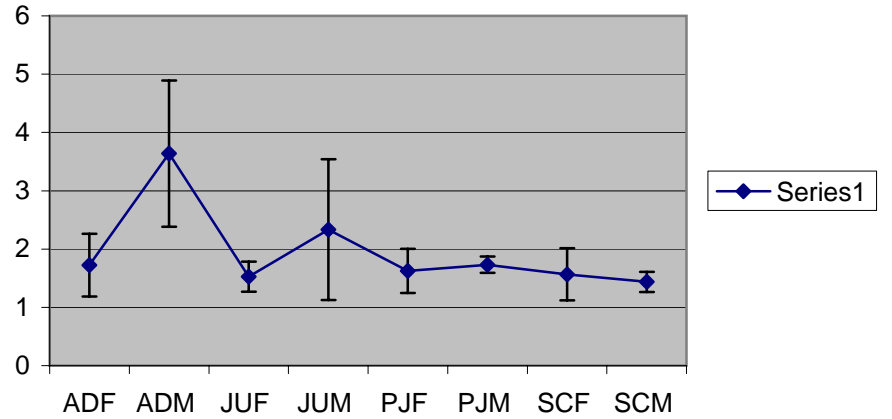
### 2150 Ipsative



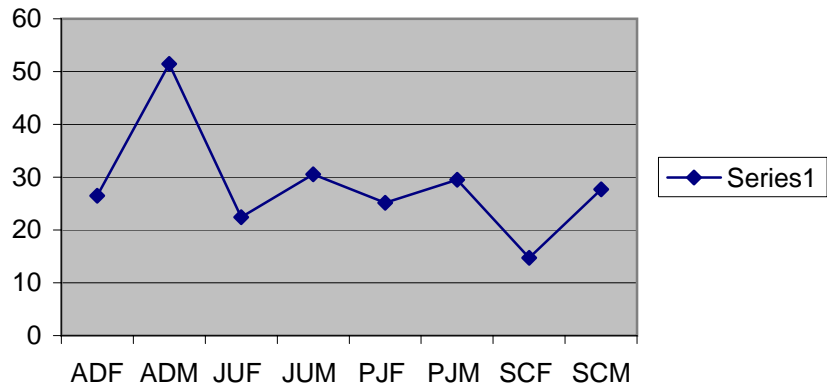
### 2149 Raw Vt



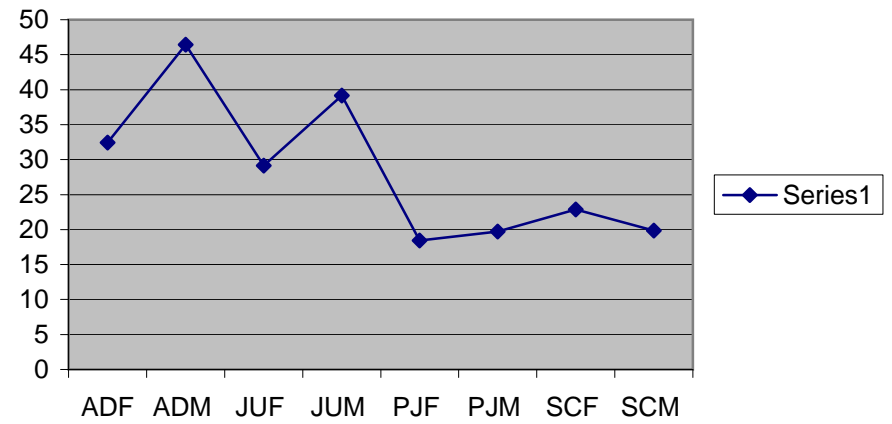
### 2150 Raw Vt



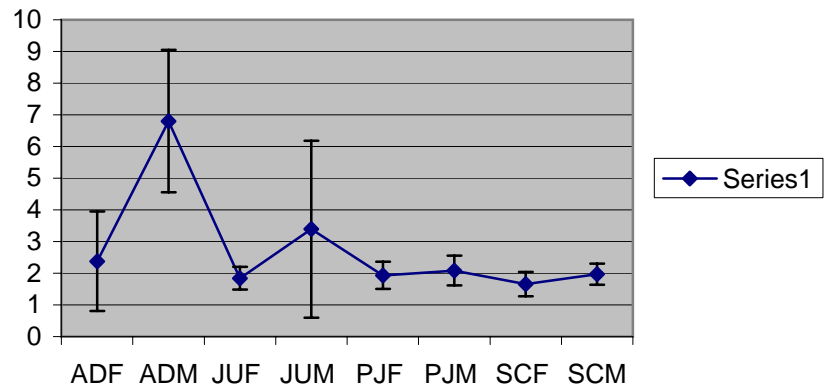
2151 Ipsative



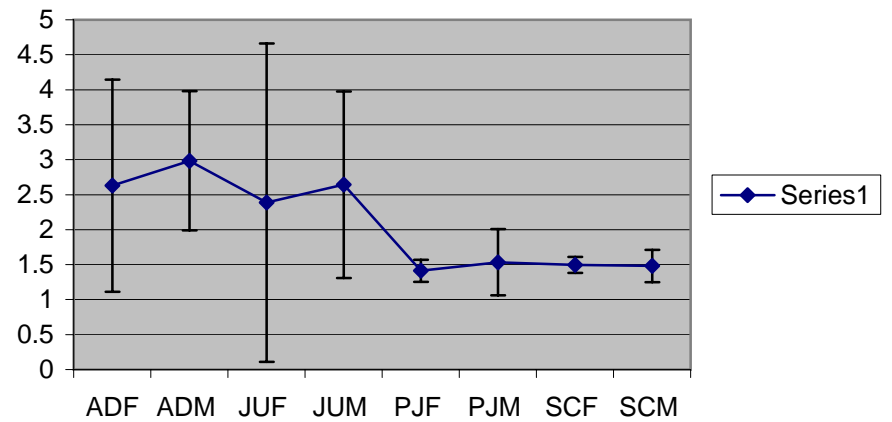
2152 Ipsative



2151 Raw Vt

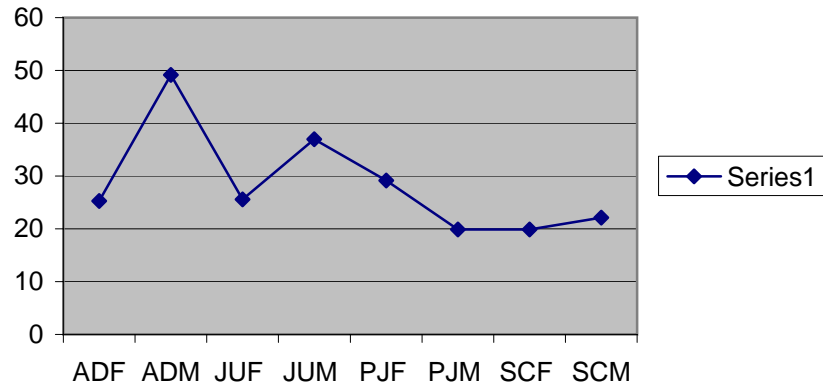


2152 Raw Vt

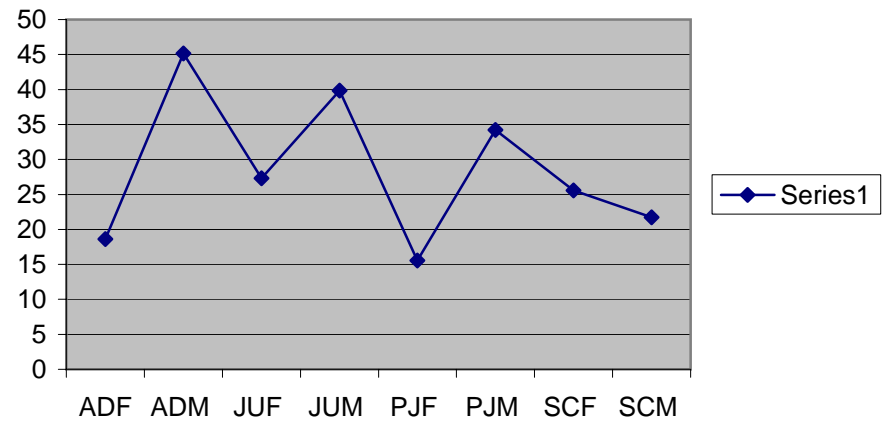




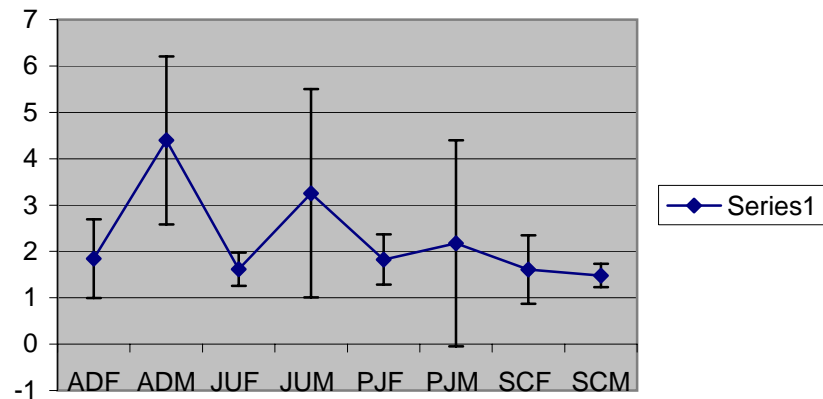
### 2153 Ipsative



### 2154 Ipsative



### 2153 Raw Vt



### 2154 Raw Vt

