Descriptive statistics (m3c)

Report The report presented on the effectiveness of the new drug against pro cancer, while having good intentions at heart, unfortunately contains errors in the interpretation of the statistical information it contains based on the following data: Number of weeks the patient lived after taking the drug 3 5 6 6 8 8 9 9 9 10 11 45 With this data having a mean of 10. 75 and a standard deviation of 11. 02, it would seem that the drug performed well since it is higher than the mean number of weeks that a prostate cancer patient lives after receiving a confirmed diagnosis of being in stage 4 is 9. 6, with a standard deviation of 3. 2. However, due to the nature of the data, it was inappropriate to use the mean as an accurate measure of central tendency due to the existence of an extreme data value, thus pulling the entire data towards one direction (Berg & Latin, 2008). Furthermore, the large standard deviation should also provide an indication of how variable the data set is, since the standard deviation measures the average distance of the data set values from the mean. Thus, the results could be interpreted as having an average distance of 11. 02 weeks from the mean of 10. 75 weeks, which is very disturbing because the standard deviation is even higher than the mean itself. This further emphasizes the variability of the data set. In this particular data set which includes an extreme data value, the more appropriate measure of central tendency would be the median or the “ middle value” of the data set. The median is the most appropriate measure of central tendency in this case because it is not affected by the existence of extreme values (Kazmier, 2004). Rather, it locates the middle value, or the mean of the two middle values, among the values of the data set and considers this appropriately. Thus, if the data set is reevaluated, it would be more appropriate to conclude that the “ average” number of weeks

the patient lived after taking the drug is 8. 5 which is the arithmetic mean of the two middle values, 8 and 9. Another possible option would be to get rid of the extreme value before taking the mean and standard deviation of the data set (Bluman, 2004). This would ensure that the values of the data set will not be pulled towards one direction. If this were done, then the value of “ 45” may be taken away and this would result to a mean of 7. 64 and a standard deviation of 2. 4, which significantly changes the initial results into better representations of the behavior of the data set. References Berg, K., & Latin, R. (2008). Essentials of research methods in health, physical education, exercise science and recreation, (3rd ed.). Baltimore, MD: Lippincott Williams & Williams. Bluman, A. (2004). Elementary statistics: A step by step approach, 5th ed. McGraw-Hill. Kazmier, L. (2004). Schaum's Outlines: Theory and Problems on Business Statistics (4th ed.). USA: McGraw-Hill.