**MVIS 5301 Statistical Applications for Visualization**

Problem Set #1 **Answer Key**

1. **What is the level of measurement for each of these variables (from the 2002** [**General Social Survey**](http://www3.norc.org/gss+website/)**) (nominal, ordinal, interval/ratio)? *Briefly* justify your answer.   
   *Note:* Ignore the response categories of *NAP, DK, and NA* for purposes of this question.**

CHILDS – ordinal – the first eight responses (0-7) are meaningfully ordered (increasing) in units that have equal intervals, and the answer zero has meaning as the absence of what is being measured (no children), perhaps making us think this is an interval/ratio variable. But the response category of “eight or more” (unequal interval from a response of 7 children to a response, e.g., of 12 children) makes this variable ordinal.

EDUC – interval/ratio – years of schooling are units that can be subdivided (continuous), are ordered in units that have equal intervals, and have a true zero point indicating that no schooling has taken place. Because the top category is not “20 or more,” it looks like the highest observed value is actually 20. thus, we classify this as an interval/ratio variable.

1. A student discovers that his grade on a recent test was in the 72nd percentile. If 90 students took the test, then approximately how many students received a higher grade than he did?

**Solution:** 90\*.28=25.2

1. A sample of 7 underweight babies was fed a special diet and the following weight gains (lbs) were observed at the end of three months.

6.7 2.7 2.5 3.6 3.4 4.1 4.8

Find the mean and standard deviation for these 7 babies and show your calculations (credit will not be given if you simply provide the answers).

**Solution:** Mean = (6.7 + 2.7 + 2.5 + 3.6 + 3.4 + 4.1 + 4.8)/7 = 27.8/7 = 3.97

Standard Deviation = sqrt{[(6.7-3.97)2 + (2.7-3.97)2 + …. ]/(7-1)} = sqrt{[12.39]/6} = 1.44

(other accepted answer is to divide by N, not N-1): = sqrt{[12.39]/7} = 1.33

1. The Nielsen Company publishes information on the TV-viewing habits of Americans in *Nielsen Report on Television.* A sample of 20 people yielded the weekly viewing times, in hours, in the table below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 25 | 41 | 27 | 32 | 43 |
| 66 | 35 | 31 | 15 | 5 |
| 34 | 26 | 32 | 38 | 16 |
| 30 | 38 | 30 | 20 | 21 |

**Solution:**

1. Determine the median of these data.

First, arrange the data set in increasing order:

5 15 16 20 21 25 26 27 30 **30 31**  32 32 34 35 38 38 41 43 66

There are 20 observations in this data set, so the position is right at the middle (position 10.5), halfway between the tenth and eleventh observations (shown in boldface). Thus, the median of the data set is (30+31)/2 = 30.5.

1. Determine the quartiles of these data.

Because the median of the entire data set is 30.5, the part of the entire data set that lies at or below the median is:

5 15 16 20 **21 25** 26 27 30 30

This data set has 10 observations, so its median is at position (10+1)/2 = 5.5, halfway between the fifth and sixth observations (in boldface) in the list. Thus the median of this data set—and hence the first quartile—is (21 + 25)/2 = 23; that is, Q1 = 23.

The second quartile is the median: Q2 = 30.5.

The part of the data set that lies at or above the median is:

31 32 32 34 35 38 38 41 43 66

This data set has 10 observations, so its median is at position (10+1)/2 = 5.5, halfway between the fifth and sixth observations (in boldface) in the list. Thus the median of this data set—and hence the first quartile—is (35 + 38)/2 = 36.5; that is, Q3 = 36.5.

In sum, Q1 = 23 hours; Q2 = 30.5 hours; Q3 = 36.5

1. Obtain the lower and upper limits.

Lower limit: Q1 – 1.5 × IQR

Upper limit: Q3 + 1.5 × IQR

IQR = Q3 – Q1

Thus: IQR = 36.5 – 23 = 13.5

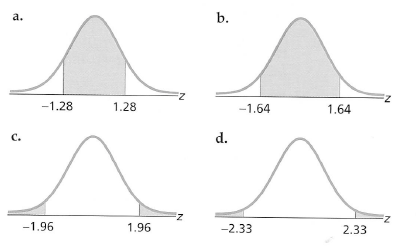
Lower limit: Q1 – 1.5 × IQR = 23 – 1.5 × 13.5 = 2.75 hours

Upper limit: Q3 – 1.5 × IQR = 36.5 + 1.5 × 13.5 = 56.75 hours

1. Determine potential outliers, if any.

The ordered list in part (1) reveals one observations—66—that lies outside the lower and upper limits—specifically, above the upper limit. Consequently, 66 is a potential outlier.

1. Use the *Area under the standard normal curve* table to obtain the shaded area under the standard normal curves below.



a. 0.7994 b. 0.8990 c. 0.0500 d. 0.0198