Daniel Prusinski

401 Midterm Exam

1) Briefly describe the four levels of measures, providing an example of each (5 points).
   1) Nominal: refers to naming, usually done through assigning the object of measurement to a category representing a different attribute.
      Example: An example of a nominal measurement would be assigning subjects based on religious preference, perhaps Muslim, Buddhist, and Bahia.
   2) Ordinal: refers to a quantifiable ranking from most to least or some other logical sequence or ordering of a variable's categories. The variable being measured is not only measured by a different attribute but also amounts are implied.
      Example: An example of ordinal measurement would be ranking subjects from tallest to shortest.

Ricky Bobby	Tallest
Davey Jones	Second Tallest
Jethro Bodine	Penultimate Tallest
Floyd Christmas	Shortest

   3) Interval: refers to the subject being measured receiving a numerical score rather than a ranking. Zero is an randomly chosen point rather than lack of what is being measured, such that scores can be above and below zero.
      Example: An example of interval level measurement would be the table below showing the temperatures for the first week in January for MN.

Day of Week	Temperature: Fahrenheit (Interval Measurement)
Monday	24
Tuesday	21
Wednesday	18
Thursday	10
Friday	0
Saturday	-15
Sunday	-18

   4) Ratio: refers to the same definition of Interval measurement, <u>except</u> zero equals a complete lack of the variable being measured.
      Example: An example of ratio level measurement would be temperatures measured on the Kelvin scale, where zero is the absolute coldest and nothing can get colder.

Day of Week	Temperature: Kelvin (Ratio Level)
Monday	80
Tuesday	76
Wednesday	45
Thursday	44
Friday	40
Saturday	35
Sunday	32

2) Select a problem in your field that you believe needs to be examined empirically. Develop a hypothesis for such a study. Explain the independent and dependent variables as well as the units of analysis that would need to be tested (10 points).

1) Problem: Somali Refugee Students School attendance drops when the holiday Ramadan is being celebrated.

Hypothesis: There is a relationship between school attendance for Somali refugee students and Ramadan, such that when Ramadan is being celebrated school attendance drops as compared to when Ramadan is not being celebrated.

1. Independent Variable: Ramadan is the independent variable in that it is doing the "causing" in regard to attendance.
2. Dependent Variable: School Attendance is the dependent variable because it is the cause; such that as the date of Ramadan changes so does the attendance for the particular month.
3. Units of Analysis: Somali Refugee Students

3) It can be said that a set of scores can be reliable yet not valid; however, scores from an instrument cannot be valid unless they are reliable. What does this mean? Please use an example from your personal or professional experience when answering this question. (10 points)

1) Twice a year my refugee program called "Baro" conducts a community education seminar. After each seminar, attendees fill out a questionnaire or survey on their perception of the overall seminar. Scores are collected on the survey and are a reliable means of the attendee's perception of the event. But, if I were trying to conclude the satisfaction from my volunteers that served during the seminar, of which did not fill out a survey because they are <u>not</u> attendee's, from the reliable scores of the actual attendees the results would not be valid. This is because it is not possible to measure my volunteers' satisfaction from the attendee's survey; hence the scores are not valid. Validity is based on the extent to which the desired concept I am measuring is actually being measured by a score, scale, or index. If face and content validity were applied to my assessment of volunteer satisfaction, the results would be rendered invalid.

In the above example, the instrument of the survey is the Questionnaire Document (QD) that has the questions for each attendee to answer. Each QD must be reliable, meaning it must measure the attendee's perception of the event. If the QD asks questions such as the weather outside, favorite color, or other irrelevant questions the QD is not reliable and thus invalid. Each questions on the QD must be relate to what is actually being measured in order to be reliable. Another aspect of reliability is consistency. The QD is reliable if it generates the same results over time from the same participant. If I were to give my QD to an attendee one hour, one day and one week after the seminar each time the QD should have the same relative score in order for it to be reliable.

4) When might you use the median rather than the mean as the measure of central tendency? Please use an example from your personal or professional experience when answering this question. (5 points)

1) Management recently tasked me with analyzing the attendee age breakdown for each church service. The Gathering is our Sunday night church service which attracts a younger crowd, but some elderly people attend as well. When trying to accurately portray who is attending the service, the mean is not necessarily accurate given the polarity in age ranges. Thus, using the median age is far more accurate when portraying the greatest age represented. The actual results had quite a difference between Mean and Median such that the mean was 32 and the median was 25. Upon further analysis, it was found that 27% of attenders were ages 45-60 with 55% being between 15-25, this only validated that a mean would have not been accurate and that a median was the right tool to use.

5) What do measures of dispersion allow you to understand that measures of central tendency do not? (10 points)

1) I think of Central Tendency (CT) as viewing a TV in black and white. CT gives one a picture of data but it is rather plain such that it primarily tells one the location of the data through tools like median, mode, and mode.

A measure of Dispersion or (MD) lets one see in color. MD let's one see the spread and shape of data, presenting a full colorful picture of data. This picture that is now seen in color, MD, directly relates to how scores or points of datum differ or vary from each other and the mean. The MD lets one understand the range and variance of data in regard to other data points and the data as a whole. This is not possible when viewing data through CT.

6) The leadership of a company is concerned that its hiring practices are discriminatory. The human resources department provided the following information about job applicants for positions within the past year.

Hired	Non-white			White		
	Male	Female	Total	Male	Female	Total
Yes	48	34	82	107	62	169
No	72	34	106	57	26	83
Total	120	68	188	164	88	252

a. Which group of people (non-white males, non-white females, white males or white females) was most likely to obtain a position? (5 points)

Take yes's for each group divided by each correlating groups total to figure out probability of attaining a position.

Hired	Non-white			White		
	Male	Female		Male	Female	Total
Yes	48	34		107	62	
Total	120	68		164	88	
Yes/Total	.40	.5		.65	.70	

The group most likely to obtain a position is a White Female with a probability of 70%.

b. Were men or women more likely to obtain a position? (5 points)

Hired	Non-white			White		
	Male	Female	Total	Male	Female	Total
Yes	48	34	82	107	62	169
No	72	34	106	57	26	83
Total	120	68	188	164	88	252

My table

Hired	Non-white/White			Non-White / White		
	Male	Male	Total	Female	Female	Total
Yes	48	107	155	34	62	96
Total	120	164	284	68	88	156
Yes /Total			.55			.62

Women are more likely to obtain a position with a probability of 62% compared to men with a 55%.

c. Were non-whites or whites more likely to obtain positions? (5 points)

Hired	Non-white			White		
	Male	Female	Total	Male	Female	Total
Yes	48	34	82	107	62	169
No	72	34	106	57	26	83
Total	120	68	188	164	88	252

My Table

Hired	Non-white			White		
	Male	Female	Total	Male	Female	Total
Yes	48	34	82	107	62	169
Total	120	68	188	164	88	252
Yes/Total			.44			.67

Whites are more likely to obtain positions with a 67% probability compared to a 44% probability for non-whites.

d. Do these data show evidence of discrimination in hiring? Why or why not? (5 points)

The variance displayed in hiring is such that White Females are most likely to be hired followed by white males, non-white females, and lastly non-white males. Also, it is more probable to that a female will be hired than a male from the data. That being said, the definition of discrimination is "The unjust or prejudicial treatment of different categories of people or things, esp. on the grounds of race, age, or sex (Google Dictionary)." The data does show differences between race and gender, but it does not reveal normative data in regard to what denotes unjust or prejudicial treatment. While the data does show differences in hiring percentages, from the information given one cannot absolutely conclude discrimination is taking place. Plus, in our justice system, one is innocent until proven guilty, thus on cannot conclude discrimination simply from the above data. But, the data presented with valid testimony of discriminatory practices would lend itself very convincing to a jury demonstrating discriminatory practices. It should also be noted that more white candidates were interviewed than non-whites which would support a bias hiring argument.

7) A company is interested in exploring whether charging consumers a small fee to register on its website results in fewer consumer purchases from a certain advertising partner. The consumer website registration charge took effect in 2005.

Year	Registrations	Purchases
1998	447,000	13
1999	460,000	21
2000	481,000	24
2001	498,000	16
2002	513,000	24
2003	512,000	20
2004	526,000	15
2005	559,000	34
2006	585,000	33
2007	614,000	33
2008	645,000	39
2009	675,000	43
2010	711,000	50
2011	719,000	47

Create an SPSS data file for the data provided above. Calculate the following for *purchases* only, for all years of the data. Copy the Variable View, Data View, and Output Panel screens to your exam.

a. Frequency table (3 points)

Purchases					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	13.00	1	7.1	7.1	7.1
	15.00	1	7.1	7.1	14.3
	16.00	1	7.1	7.1	21.4
	20.00	1	7.1	7.1	28.6
	21.00	1	7.1	7.1	35.7
	24.00	2	14.3	14.3	50.0

	33.00	2	14.3	14.3	64.3
	34.00	1	7.1	7.1	71.4
	39.00	1	7.1	7.1	78.6
	43.00	1	7.1	7.1	85.7
	47.00	1	7.1	7.1	92.9
	50.00	1	7.1	7.1	100.0
	Total	14	100.0	100.0	

b. Quartiles (3 points)

Statistics		
Purchases		
N	Valid	14
	Missing	0
Percentiles	25	19.0000
	50	28.5000
	75	40.0000

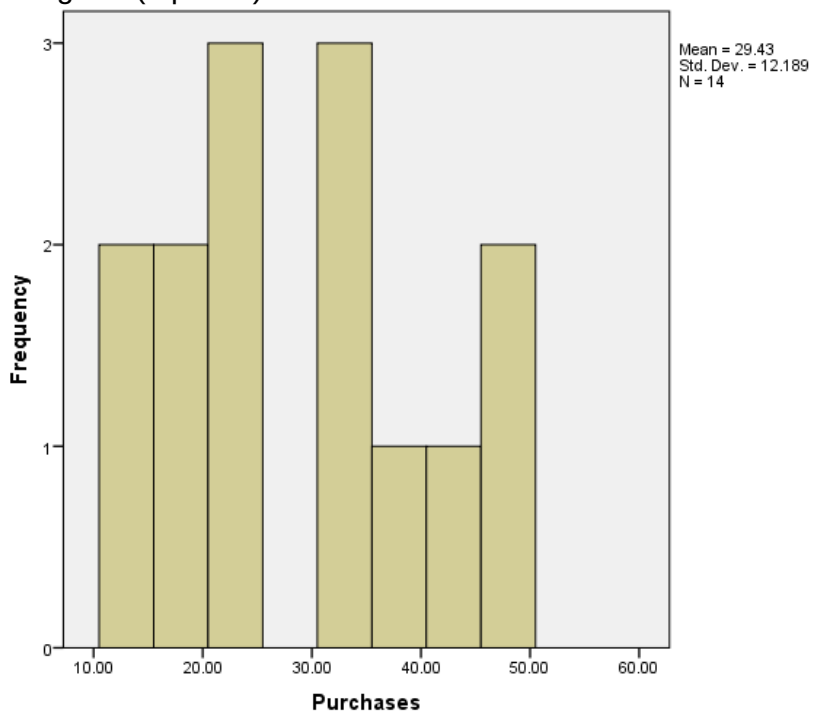
c. Mean, median, mode (3 points)

Statistics		
Purchases		
N	Valid	14
	Missing	0
Mean		29.4286
Median		28.5000
Mode		24.00a
a. Multiple modes exist. The smallest value is shown		

d. Range, variance, standard deviation (3 points)

Statistics		
Purchases		
N	Valid	14
	Missing	0
Std. Deviation		12.18899
Variance		148.571
Range		37.00

Histogram (3 points)



e. In a paragraph or two, describe your SPSS output (5 points).

I have no previous programming experience. Now that I am using SPSS on a regular basis, I view the software as a bridge from the typical point and click Graphical User Interface such as Excel to prepare us for Statistical Analysis Software such as SAS. At the heart of this “bridging” is the output panel for SPSS. The first aspect to understand with the SPSS output is the programing that takes place for each function. It is very helpful that the SPSS output lets one see the actual programming language SPSS is following to reach the desired function inputted by the user.

Following the code is usually the table or graph displaying the information the user inputted. What is unique about the code, tables, and graphs is the ability to further analyze and tweak the information in the output panel. One can double click on a specific displayed item and change it from the output panel. Graphs can be organized differently by unit size…, fonts and graphics can be enlarged in tables, and code can be added to or taken from, this all can be done in the output pane of SPSS. From a neophyte’s perspective, the output view is where one sees the computations and work displayed from the raw data being inputted from the data and variable input screens.

f. Without using SPSS, calculate the mean and standard deviation for *purchases* before and after the implementation of the consumer fee. Show your work. How could you interpret these results? (5 points)
Please Note: This work was done in Excel.

	Year	Registrations	Purchases			As a %		
1								
2	1998	447,000	13			2.91E-05		
3	1999	460,000	21			4.57E-05		
4	2000	481,000	24			4.99E-05		
5	2001	498,000	16			3.21E-05		
6	2002	513,000	24			4.68E-05		
7	2003	512,000	20			3.91E-05		
8	2004	526,000	15			2.85E-05		
9	Mean	133div7	19					
10	STDV		4.070801957					
11								
12								
13	2005	559,000	34			6.08E-05		
14	2006	585,000	33			5.64E-05		
15	2007	614,000	33			5.37E-05		
16	2008	645,000	39			6.05E-05		
17	2009	675,000	43			6.37E-05		
18	2010	711,000	50			7.03E-05		
19	2011	719,000	47			6.54E-05		
20	Mean	279/7	39.85714286					
21	STDV		6.468132242					
22								
23								
24	Work for STDV before 2005							
25	X	XBAR	X-Xbar	(x-Xbar) Squ		Sum of (X-Xbar)Squ	/n	Square Rt
26	13	19	-6	36			n=7	
27	21	19	2	4				
28	24	19	5	25				
29	16	19	-3	9				
30	24	19	5	25				
31	20	19	1	1				
32	15	19	-4	16				
33						116	16.5714	4.0708
34								
35	Work for STDV 2005 through 2011							
36	X	XBAR	X-Xbar	(x-Xbar) Squ		Sum of (X-Xbar)Squ	/n	Square Rt
37	34	39.8571	-5.85714286	34.3061			n=7	
38	33	39.8571	-6.85714286	47.0204				
39	33	39.8571	-6.85714286	47.0204				
40	39	39.8571	-0.85714286	0.73469				
41	43	39.8571	3.142857143	9.87755				
42	50	39.8571	10.14285714	102.878				
43	47	39.8571	7.142857143	51.0204				
44						292.8571429	41.8367	6.46813
45								

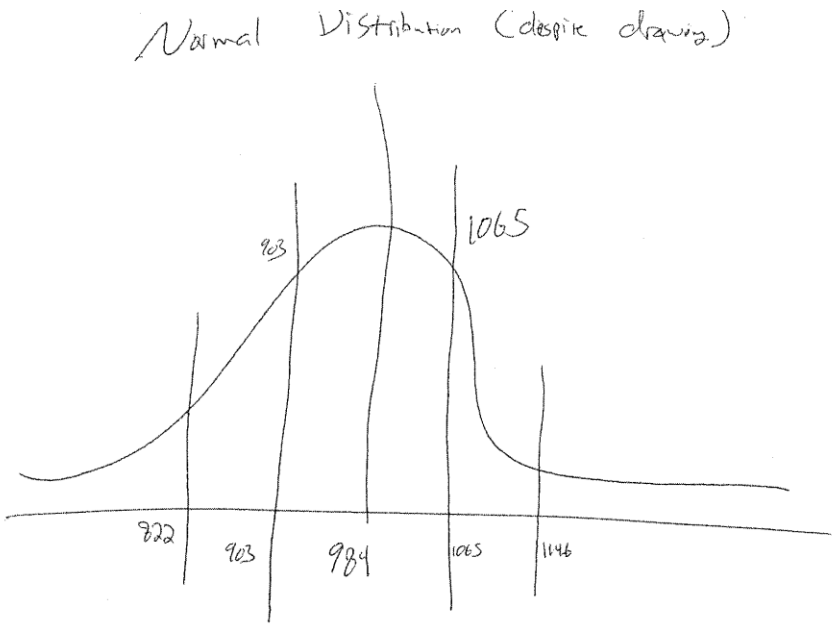
I would interpret these results as the following:

· After the consumer fee the average amount of purchases increased.

· There is a greater variance of purchases after the fee than before the fee, but even with the variances there are more purchases than before the fee was implemented.

· Also, analyzing purchases per registration one finds that after the fee there are more purchases per registration.
· Overall, even with the greater variance in purchases after the fee was introduced, more purchases were made and it was a good decision to implement the fee.

8) The director of the MBA program at Acme University wants to raise standards for admission. As a first cut, she is thinking about rejecting all applicants who had combined scores on the verbal and quantitative portions of the Graduate Record Examination (GRE) that fell into the bottom 70% of all who took the GRE in the previous year. Last year, the combined scores on the GRE were normally distributed with a mean of 984 and a standard deviation of 81. Given her criterion, at what combined score should she set the standard for further admission consideration? Show your work. (15 points)



984 = Mean

81 = Stdv

70% + lower No good

71% + up are good = Look for a z-score that Cover .29% + up

ie .29 = .3879 with a z score of .56

She should set the score at 1029.36 or higher for admission consideration.

$$.56 = \frac{x - 984}{81}$$

$$\frac{x - 784}{81} = .56$$

$$x = (.56 \cdot 81) + 984$$

$$x = 1029.36$$ Score

Work through

1029.36 score

$$\frac{1029.36 - 984}{81} = .56$$

$Z = .56$

Z score of .56 = .2877 or .29 + greater

Thus, those with a score 1029.36 + greater should be considered for admissions.