Name: Key

Standard One: Exam Period One – Reasons for Using Statistical Methods?

It is very hard to collect population measures even when the population is relatively small. Note that if you were to just take thirty seconds to survey each person in the university student body, which is roughly 32,972, it would take about eleven and a half days.

In order to get around this another instructor and I put together a lengthy, anonymous and voluntary survey. In this survey, which one hundred and thirty six students filled out, we asked a variety of questions. With the collected data we can summarize the information provided by the data and make inference on the entire student body of the University of South Carolina.

In this sense why is statistics an important discipline? Particularly, what is an example of a question that we can answer using statistics? Use the information above and one of the methods we learned this semester to provide a good example.

Answers may vary, but an example can be seen by comparing two variables, like gender and whether or not someone has cheated on their significant other. In this case we saw that there was little difference across gender which is very interesting.

Standard Two: Exam Period One – Experiments: Samples vs. Populations, and Sampling

In the context of this survey, identify the following:

Sampling Method: Convenience or Volunteer Sample

Sample: 136 responders

Population: 32,972 students

Name: _____

Standard Three: Exam Period One –Types of Data

In the context of this survey, identify the data type, i.e. Discrete, Continuous or Categorical, of the following:

Gender:

Please choose your gender: Categorical
- Male
- Female

Self-rating:

How many alcoholic drinks have you had this week? (Fri 8th through Fri 15th) Discrete

Car Accidents:

How many car accidents have you been in, in which you were a driver? Discrete

GPA:

What is your GPA? Continuous / [Discretely Measured]

Standard Four: Exam Period One – Graphical Summaries

In the context of this survey, identify an appropriate graphical summary of the following:

Gender: Pie chart / Bar chart

Self-rating: Dot Plot / Bar chart Histogram also acceptable

Car Accidents: Same as above.

GPA: Histogram / Boxplot

Name: _____

Below is the sample data for the self-attractiveness rating of survey respondents from the survey mentioned previously.

2, 5, 6, 6, 6, 6, 6, 6, 6, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 8, 8, 8, 8, 8, 9, 10 $n=26$

Note: $\sum x = 178$

$\sum(x - \bar{x})^2 = 53.3846$

$variance = \frac{\sum(x-\bar{x})^2}{n-1}$

$Position\ of\ the\ pth\ percentile = \left(\frac{p}{100}\right) * (n+1)$ **[Hint: Use for quartiles]**

Standard Five: Exam Period One – Measures of Center

Mean: $\overline{X} = \frac{178}{26} = 6.8462$

Median: $\left(\frac{50}{100}\right)(26+1) = 13.5^{th}$ position $\longrightarrow$ Median $= \frac{7+7}{2} = 7$

Mode: 7

Standard Six: Exam Period One – Measures of Variability

Range: $10 - 2 = 8$

Variance: $s^2 = \frac{53.3846}{26-1} = 2.1354$

Standard Deviation: $s = \sqrt{2.1354} = 1.4613$

Name: _____

Standard Seven: Exam Period One – Percentile and Quartiles

Q_1: $\frac{25}{100}(26+1) = 6.75 \longrightarrow Q_1 = .75(6) + .25(6) = 6$

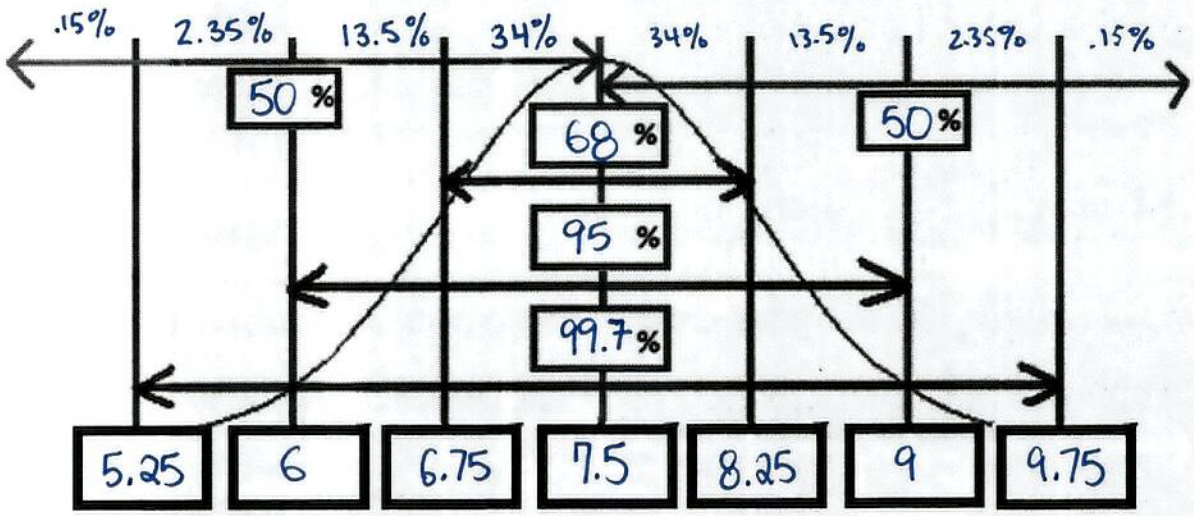
Q_3: $\frac{75}{100}(26+1) = 20.25 \longrightarrow Q_3 = .25(8) + .75(8) = 8$

IQR: $Q_3 - Q_1 = 8 - 6 = 2$

Name: _____

Standard Eight: Exam Period One – The Empirical Rule

Assume the population mean self-attractiveness rating is 7.5 with a standard deviation of 0.75. Fill in the following chart using those values and answer parts a through c.



a. Approximately 16% of students rated themselves lower than what value?

6.75

b. Approximately what percent of the data lies above that value in part a?

100-16 = 84%

c. Calculate the Z-Score for the student that rated themselves a 2. Is this an outlier?

$$z = \frac{observation - mean}{standard\ deviation} = \frac{2-7.5}{.75} = \frac{-5.5}{.75} = -7.\overline{33}$$

$-7.\overline{33} < -3$ → outlier!

Name: _____

Standard Nine: Exam Period One – Variable Association – Contingency Tables

Use the preceding tables to explain whether or not you believe there is an association between gender(rows) and whether or not they would be willing to marry someone they didn't love for whatever reason.(cols).

Take the count table and use the values within to create the conditional percentage table to justify whether or not you believe there is an association between the two variables. Is this result surprising to you?

COUNT	No	Yes	Total
Female	76	14	**90**
Male	31	13	**44**
	107	**27**	**134**

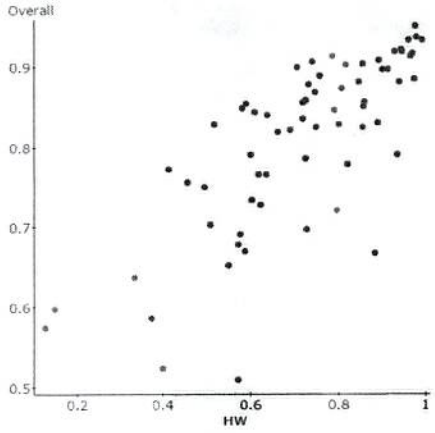
Conditional %	No	Yes	Total
Female	84.44%	15.55%	100%
Male	70.45%	29.54%	100%

Here the count shows we have about the same number of Males and females would marry someone they didn't love but after putting it into context by calculating the conditional percentage it's clear that males have a higher proportion in that category suggesting men might be more likely to be willing to marry someone they didn't love.

Standard Ten: Quiz Thirteen – Scatterplots and Correlation Coefficient

Many students complain about the homework so I decided to investigate whether or not any of you had a case when saying it was "completely useless!" I decided to create a scatterplot so I could investigate the association between students' homework grades and their overall grades.

Given the following scatterplot, which I created using the results of all students in this course last semester, can you draw a conclusion about whether or not the homework is useful? Use arguments about the strength and direction, or lack thereof, of the association to support your answer.



Here, there appears to be a weak to moderate, positive association between HW and overall grade which suggests it's of some importance.

Name: _____

Standard Eleven: Quiz Thirteen – Linear Regression

Completing a simple linear regression for predicting overall grade with a student's homework grade I found the regression equation to be $\widehat{\boldsymbol{Overall}} = .\mathbf{5102} + .\mathbf{4157} * \boldsymbol{HW}$ with $R^2 = .5889$.

a. About what percent of variation of the overall grades is explained by homework?

About 58.89% of the variation in overall grades is explained by HW.

b. If I have 100% on my homework, what can I expect my overall grade to be?

$\hat{y} = .5102 + .4157(1) = .9259$

If you have a 100% on the homework we can expect a 92.59% overall, on average.

c. Can you interpret the intercept of this model? If so, please do.

$\hat{y} = .5102 + .4157(0) = .5102$

If you have a 0% on the homework we can expect a 51.02% overall, on average.

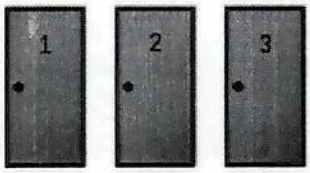
d. Can you use this model to predict your homework grade in another class using your homework grade from that class?

It would not make sense to use this model as it is designed only using data from students that took my Stat 201 class.

Name: _____

Standard 12: Exam Period One – Basic Probability Results

Let's say you're on the famous game show, Let's Make a Deal and you were lucky enough to be selected to play for the big deal of the day! The setup of the game is this – there are three doors and there is a very expensive prize behind one of the doors, like a new car; the other two doors yield you nothing!



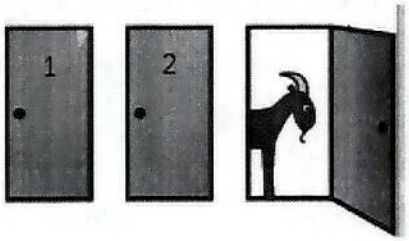
You have three choices to choose from, let A denote the event that you choose the winning door. Find the following probabilities for **A = You chose the winning door:**

a. $P(A) = \frac{1}{3} = .\overline{33}$

b. $P(A^c) = 1 - \frac{1}{3} = .\overline{66}$

Standard 13: Exam Period One – Probability – Conditional Probability

Now, regardless if your initial guess was correct the host will show you that there is a goat behind one of the other doors. For example, if you chose door one the host might show you door three which tells you that the attractive prize is definitely behind door 1 or door 2.

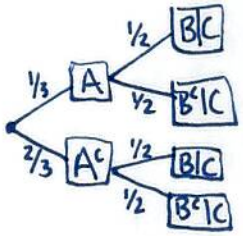


Now, he will ask you if you want to change your mind – you now have two options: you can stick with your initial gut instinct or you can choose to switch to the other door. Let B denote the event that you choose the winning door this time. Find the following probabilities for **B = You chose the winning door and C = door 3 is a loser**

$P(B|C) = 1/2$

$P(B^c|C) = 1/2$

Note



B must occur for you to win

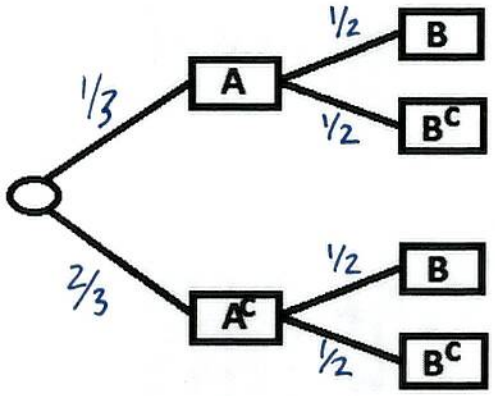
$P(A \cap B) = 1/3 \cdot 1/2 = 1/6$

$P(A^c \cap B) = 2/3 \cdot 1/2 = 2/6 = 1/3$ ← Better to assume you were wrong.

Name: _____

Standard 14: Exam Period One – Probability – Disjoint vs. Independent

a) Use standards 12 and 13 to assign the correct probability to the correct branch on the tree diagram. Note: you need to assume **independence** of events A and B.



b) Find the following probabilities and choose the best strategy for winning the big deal of the day, i.e. is it better to assume your first answer was wrong and switch or assume your first answer was right and keep it?

$P(A \cap B) =$ $1/3 \cdot 1/2 = 1/6$

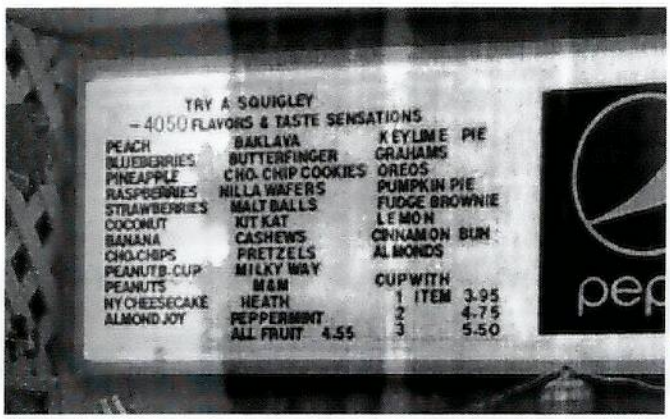
$P(A^C \cap B) =$ $2/3 \cdot 1/2 = 2/6 = 1/3$

You have twice the chance to win if you assume you were wrong first.

Standard 17: Exam Period One – Counting Techniques

While at a conference near Carolina Beach I got ice cream at Squigley's Ice Cream & Treats. I noticed this lovely math problem (it's a mathstery) on their menu board and I worked out the answer on the back of my receipt.

Considering the following ice cream item where they mix in 1, 2 or 3 items into their ice cream – they have 32 options for mix-ins or you can also get it plain, or with all the fruit. How many combinations are really possible; the hint is that this sign is actually incorrect. (Just write out the formula)



1 plain
1 all the fruit
$\binom{32}{1} = {}_{32}C_1$ one mixin
$\binom{32}{2} = {}_{32}C_2$ two mixins
$+ \binom{32}{3} = {}_{32}C_3$ three mixins

5490