Quantitative Political Methods
Midterm Review

Practice problem set:

1. Identify each variable as nominal/ordinal/interval and discrete/continuous:

   a. Type of car driven
   Nominal, discrete

   b. General health (poor, reasonably good, excellent)
   Ordinal, discrete

   c. College tuition
   Interval, continuous

   d. Number of political parties in a country
   Interval, discrete

   e. Religious affiliation
   Nominal, discrete

   f. Distance between home and work
   Interval, continuous

2. Create a boxplot for the following dataset: 4, 8, 32, 41, 16, 11, 1, 38, 3, 27, 48, 19
   Is this data skewed? If so, in which direction?

   Numbers in order: 1, 3, 4, 8, 11, 16, 19, 27, 32, 38, 41, 48
   Median = (16+19)/2 = 17.5
   Q1 = (4+8)/2 = 6
   Q3 = (32+38)/2 = 35
   IQR = 35-6 = 29
   Outliers: 29(1.5)=43.5, so outliers would be less than -37.5 (6-43.5) and greater than 78.5 (35+43.5). Since none of our numbers are outliers, our boxplot has a box from 6 to 35 with a median of 17.5 and whiskers out to the lowest and highest numbers, 1 and 48.

   Skew: mean=20.667, median=17.5. Because median<mean, the data is positively skewed—but not by very much.

3. The following table contains the GDP per capita (in thousands of international dollars) for four European countries.

Belgium	Germany	France	Luxembourg
38	38	35	90

a. Find the mean.

(38+38+35+90)/4 = 50.25

b. Find the median.

Numbers in order: 35, 38, 38, 90. Median is 38.

c. Find the mode.

38

d. Find the range.

90-35 = 55

e. Find the variance.

$$\mathbf{S^2 = \Sigma(y_i - \bar{y})^2/(n-1)}$$

$$\mathbf{S^2 = \frac{(35-50.25)^2 + (38-50.25)^2 + (38-50.25)^2 + (90-50.25)^2}{4-1}}$$

$$\mathbf{S^2 = 2112.75/3 = 704.25}$$

f. Find the standard deviation.
$S=\sqrt{704.25} = 26.538$

g. Would you say that one of these observations is an outlier?

Yes. From looking at the data, it seems obvious that 90 is an outlier. By the IQR(1.5) rule, 90 is not technically an outlier. Q1=36.5 and Q3=64. IQR=27.5. 27.5(1.5)=41.25. 64+41.25=105.25, but because there are so few numbers the argument can easily be made that 90 is an outlier.

4. The distribution of math SAT scores for students that get into Washington University in St. Louis follows a normal distribution with a mean of 740 and a standard deviation of 23. If we poll a random sample of 51 students, what is the probability that the mean score of your sample will be 746 or below?

$$Z=\frac{y-\mu}{\sigma}$$

Because we want a probability for the sampling distribution (out of all the samples, what percentage would have a score at 746 or below), we use standard error as our σ.
$SE=\sigma/\sqrt{n}=23/\sqrt{51}=3.22$
$Z=(746-740)/3.22=1.863$
Probability from z-table = .0314

But we want the probability that the mean is at or below 746, so we need 1-.0314=.9686.

5. The Freshman Fifteen is an expression that commonly refers to an amount (somewhat arbitrarily set at fifteen pounds) of weight often gained during a student's first year at college. You decide to test whether this expression holds true for Washington University in St. Louis. You randomly select 16 sophomores and gather data on how much weight (in pounds) they gained the previous year. The mean of your data is 14.5 lbs. and the sample standard deviation is 0.8 lbs.

a. Identify the population for this study.

The population is sophomores at Washington University in St. Louis

b. Describe the sample distribution for this study.

The sample distribution is a distribution with a mean at 14.5 lbs and a standard deviation of 0.8 lbs.

c. Describe the sampling distribution for this study as precisely as possible.

Standard error=s/√n=0.8/√16=0.2
The sampling distribution approximates a normal distribution with a mean at 14.5 and a standard error of 0.2.

d. Calculate the point estimate and a 95% confidence interval for the population mean. Explain what your confidence interval means.

Point estimate=$\bar{y}$=14.5
T for a 95% confidence interval with 15 degrees of freedom=2.131
$\bar{y}$±t(se)= 14.5±2.131(0.2)= 14.5±.426.
[14.074, 14.926]. With repeated random sampling of WUSTL sophomores of a sample size of 16, the mean weight gain will fall between 14.074 and 14.926 95% of the time—it is unlikely that the "freshman fifteen" is truly the "freshman fifteen" at Wash U

6. Apple claimed that iPhone 5 is "the biggest thing to happen to iPhone since iPhone." Among other improvements, the iPhone 5 claimed improved battery life over the old versions. For example, the standby time has been improved to 225 hours (a 25-hour improvement over iPhone 4S). To test this claim, you collect a sample of battery longevity from 100 randomly selected owners of the iPhone 5. Among these 100 owners, you find that the battery life in the new iPhone 5 is 217 hours with a standard deviation of 40 hours. Test the research hypothesis that *the batteries in the new iPhone 5 differ from the 225 hours claimed by Apple.* Use a 0.05 level of significance.

1. **Make assumptions. Data is interval and continuous, the sample was randomly selected, etc.**
2. **Identify null and alternative hypotheses.**
   H_0: μ=225
   H_a: $\mu \neq$225
3. **Calculate a test statistic.**
   $\bar{y}$=217
   se=s/√n=40/√100=4
   Z=($\bar{y}$-μ_0)/se
   Z=(217-225)/4=-2.
4. **Find a p-value. Z=-2, so p=.0228. Since this is a 2-sided test, we use 0.0228(2)=.0456.**
5. **Conclusion. Our significance level was 0.05. Since 0.0456<0.05, we can reject our null hypothesis-the probability of getting a sample mean as extreme as 217 in either direction if the null hypothesis were true is 0.0456, so the batteries in the new iphone 5 do differ from the 225 hours claimed by Apple.**

7. You roll a 6-sided die.
   a. What's the probability of rolling a four three times in a row?

   Let's first calculate the probability of rolling three 1s in a row.

   On binomial table, n=3, r=3, p=1/6=.1667

   The probability of rolling three 4s (r=3) in three rolls (n=3) is about .0046 (between .003 and .008) on the table.

   b. Assume I roll the dice four times and get the same number each time. I am suspicious and suspect that the die is unfair. Test my hypothesis using a 0.05 level of significance.

Following the same method as above, I would get that the probability of rolling a die four times and getting the same number each time is about .0014.

However, this calculating is conducted under the assumption of a fair die (i.e., the probability of any given number on one roll is 1/6).

So, the probability of observing this phenomenon (a die rolled four times, giving the same number each time) is p=.0014. This is my p-value.

Since p<alpha, I would reject the null hypothesis.

8. Imagine you are interested in the different patterns of support for the Spanish government among citizens of Catalan population. You decide to conduct a survey asking people "Do you have confidence in the national government?" Possible answers include Yes or No. Your were able to poll 243 Catalans. Of these 243 respondents, 86 said, "Yes."

   a. Provide a point estimate for the percent of Catalans that have confidence in the government.

   86/243=.354

   from our sample, a point estimate for the percent of Greeks who have confidence in the government is 35.4%

   b. Identify the sampling distribution of this study. Be precise.

   Se=$\sqrt{\pi(1-\pi)/n}$ = $\sqrt{.354(1-.354)/243}$ = .0307

   The sampling distribution approximates a normal distribution with π=.354 and a standard error of .0307

   c. Construct a 92% confidence interval of the percent of Catalans that have confidence in the government.

   $\pi \pm z(se)$
   Z for a confidence level of 92% has a probability of (1-.92)/2=.04. Z≈1.75

   .354±1.75(.0307)=.354±.0537

   [.3003, .4077]
   With a random sampling of 243 Greeks, 92% of proportions of Greeks who support the government will fall between .3003 and .4077.

d. Test the theory that less than 40% of Catalans support the government using a 0.05 significance level.
H_0: π=.40
H_a: π<.40

Test statistic: $Z=(\pi_{bar}-\pi_0)/se$
π_{bar}=.354
π_0=.40
$se=\sqrt{\pi_0(1-\pi_0)/n} = \sqrt{.40(1-.40)/243} = .0314$

$Z=(.354-.40)/.0314 = -1.465$

$P(Z>1.465)\approx.0715$

Because .0715>.05, we fail to reject the null hypothesis. From our sample, it could be true that less than 40% of Greeks support the government.

9. Suppose a random sample is taken of 200 Jedi Knights. The mean number of droids killed by a Jedi Knight is 19, with a standard deviation of 2. Construct and interpret a 92% confidence interval for the mean number of droids killed.

$\bar{y}\pm z(se)$
z for a 92% confidence level (p=.04) = 1.75
$se=s/\sqrt{n} = 2/\sqrt{200} = .141$

$19\pm1.75(.141)=19\pm.247$
[18.753, 19.247]
With repeated random sampling of 200 Jedi Knights, the mean number of droids killed by a Jedi would be between 18.753 and 19.247 92% of the time.

10. The distribution of Quantitative GRE scores for graduate students who drop out of graduate school has a mean of 550 and standard deviation of 30. These scores are distributed normally.

   a. If I select one graduate school dropout at random, what is the probability their score was 574 or below?

If the scores are distributed normally, $z=(\bar{y}-\mu)/\sigma$ = (574-550)/30 = 0.8. p at 0.8 =.2119, but we want the probability of their score being 574 or BELOW, so we need 1-.2119 = .7881

   b. If you take a sample of 100 graduate school dropouts, what is the probability that the average Quantitative GRE score of your sample is 574 or below? Explain why this number is different from the answer you put down for part (a).

   Se=s/√n = 30/√100 = 3
   Z=(574-550)/3 = 8. The p-value for z=8 is very tiny and we can't even get it from our z-table. Since we want the probability of scores being 574 or below, it will be 1-somethingverytiny= a number very close to 1. This answer is different from our answer in (a) because in (a) we were just looking for the probability that one person had a score like that. When we are asking about the probability of getting a mean like that with a large sample size, we're asking what proportion of the sampling distribution falls below 574. This is why we had to use the standard error instead of simply the standard deviation.

   c. Assume that you do **not** know the true population mean and standard deviation. You take a random sample of 24 graduate school dropouts and calculate a mean of 570 and a standard deviation of 16. Construct a 95% interval for the true population mean.

   $\bar{y}$± t(se)
   t score for 95% confidence and 23 degrees of freedom = 2.069
   se=s/√n = 16/√24 = 3.266

   570±2.069(3.266) = 570±6.757
   [563.243, 576.757]
   With repeated random sampling of 24 grad school dropouts, you will find a mean GRE score between 563.243 and 576.757 95% of the time.

d. The average GRE scores for all graduate students is 582.5. Using only the sample described in part (c), test the research hypothesis that graduate school dropouts have lower GRE scores than average. Specify all five steps for conducting a hypothesis test. Choose your own level of significance.

1. **Assumptions: interval data, random samples, we can assume the population is distributed normally.**
2. **Null and alternative hypotheses:**
   H_0: μ=582.5
   H_a: μ<582.5
3. **Test statistic.**
   $\bar{y}$=570
   se=3.266
   t=($\bar{y}$-μ_0)/se
   t=(570-582.5)/3.266 = -3.827
4. **P-value. For 23 degrees of freedom, the p-value of a t-score of 3.827 is <.002.**
5. **Let's use a significance level of 0.01. Since .002<.01, we can reject our null hypothesis. With a sample mean of 570, the probability that grad school dropouts' GRE scores were the same as GRE scores for all graduate students is very small.**