Public Policy 529: Midterm Practice Questions

1. MULTIPLE CHOICE. For a large random sample, which of the following are TRUE? Select ALL that apply. (3 points)

a. The sample mean is an unbiased estimator of the population mean.

b. The sample range is an unbiased estimator of the population range. **FALSE: the sample max is negatively biased and the sample min is positively biased – so the sample range is negatively biased.**

c. Sampling error always leads to biased estimates**. FALSE. Sampling error can fluctuate around zero, and does not in of itself lead to bias.**

2. For a certain professional qualifying exam, individuals must take a series of two tests and pass at least one in order to obtain certification. For each exam, there is a 70 percent chance of failure.

a. What is the probability that an individual test-taker will obtain certification? (4 points)

There are two ways to do this.
First method:
P(PP)+P(PF)+P(FP) = 0.3*0.7+0.7*0.3+0.3*0.3 = 0.21+0.21+0.09 = .51

Alternate method:
1-P(FF) = 1-0.7*0.7 = 1-0.49 = .51

b. In a group of six individuals, what is the probability that at least one person obtains certification? (3 points)

= 1 – P(no one obtains certification)
We know from part (i) that P(individual does not obtain certification) = 0.49.
So, answer = 1 – (0.49^6) = .986

3. A researcher is interested in the possible association between: classrooms with high rates of children below the poverty line, and teachers with master's degrees. She knows the following facts. 40 percent of teachers have an MA degree. 20 percent of classrooms have a high poverty rate among students. Of the classrooms with a low poverty rate among students, 49 percent of teachers have an MA degree.

a. Find the probability that a given classroom has a high rate of students below the poverty line, conditional on the teacher having an MA degree. (4 points)

P(low and MA) = P(MA | low) * P(low) = 0.49 * 0.8 = 0.392

Cell probabilities:

	MA	**~MA**	**Total**
High pov.	**0.008**	**0.192**	**0.2**
Low pov.	**0.392**	**0.408**	**0.8**
Total	**0.4**	**0.6**	**1.0**

P(pov | MA) = 0.008/0.4 = 0.02

b. Find the probability that a given classroom has a teacher with an MA degree, conditional on having a high rate of students below the poverty line. (4 points)

=P(MA | high pov) = P(MA and high pov)/P(pov) = 0.008/0.2 = 0.04

c. Which of the following are TRUE? Select ALL that apply. (4 points)

1. Since the probabilities in (a) and (b) are equal, we can see that the variables are independent.

   The probability in (i) is P(high poverty rate | MA).
   The probability in (ii) is P(MA | high poverty rate).
   These tell us nothing about independence.

2. Since the probabilities in (a) and (b) are not equal, we can see that the variables are independent.

   The probability in (i) is P(high poverty rate | MA).
   The probability in (ii) is P(MA | high poverty rate).
   These tell us nothing about independence.

3. Since the probabilities in (a) is not equal to 20 percent, we can see that the variables are not independent.

   The probability in (i) is P(high poverty rate | MA) = 0.02.
   The unconditional probability P(high poverty rate) = 0.2.
   So this is TRUE.

4. Since the probabilities in (b) is not equal to 40 percent, we can see that the variables are not independent.

The probability in (ii) is P(MA | high poverty rate) = 0.04.
The unconditional probability P(A) = 0.4.
So this is TRUE.

4. A report indicates that teacher's annual salaries in Ontario have a mean in the population of $50,000 and standard deviation in the population of $10,000. Suppose the distribution is approximately normal. Give an interval of values that contains about 72% of salaries. (4 points)

For an interval containing 72%, we want to look at the z-score corresponding to a right-tail probability of 14%. From table A, that z-score is about 1.08 (or 1.09).
Common mistake: using z=0.58, which corresponds to the right-tail probability of 28%.

So we need two raw scores now, corresponding to z scores of 1.08 and -1.08.

For 1.08, raw score = 10000*1.08+50000=60,800.
For -1.08, raw score = 39,200.

The interval is from $39,200 to $60,800.

5. A researcher is interested in understanding work schedules. She collects data on a random sample of adults in 2010, asking "how many hours do you work in an average week?" She produces the following summary statistics:

```
                  number of hours worked last week
-------------------------------------------------------------
      Percentiles      Smallest
 1%            0              0
 5%            0              0
10%            0              0       Obs                1234
25%           29              0       Sum of Wgt.        1234

50%           40                      Mean           36.26094
                        Largest       Std. Dev.      19.20424
75%           48             89
90%           56             89       Variance       368.8029
95%           64             89       Skewness      -.4293289
99%           84             89       Kurtosis        3.06609
```

a. Calculate a 95% confidence interval for the mean number of hours worked. (5 points)

CI = mean +/- z*(s.e.) = mean +/- z*s/(n^0.5) = 36.26094 +/- 1.96 * 19.20424/(1234^0.5)
(35.2, 37.3)

Note that n^0.5 = is just a way to write the square root of n. Likewise, 1234^0.5 = √1234.

b. What is the margin of error for the 95% confidence interval in part (i)? (2 points)

= 1.96 * 19.20424/(1234^0.5)
=1.07

c. She also counts that 15% of the observations fall at or below one hour. If this variable were normally distributed, what percentage of observations in the sample would be expected to report one or fewer hours worked? (4 points)

z = (1 – 36.26094)/19.20426 = -1.84
From Table A, 3.3%.

d. Which of the following are TRUE (mark ALL that apply)? (4 points)

1. This variable appears normally distributed, so the confidence interval technique in part (i) is appropriate.

2. This variable does not appear normally distributed, so the confidence interval technique in part (i) is not appropriate.

3. This variable does not appear normally distributed, but the confidence interval technique in part (i) is still appropriate.

4. We do not have sufficient information to know if this variable is approximately normal.

6. In a survey with a sample of 400, a researcher finds that the mean number of hours per week spent on the internet is 10 with a standard deviation of 4. Suppose the null hypothesis is that the true population mean is 9.5 hours. Can the researcher reject the null hypothesis at the .05 level of significance (α=.05).

H_0: μ = 9.5
H_A: μ ≠ 9.5

To calculate the t-statistic, we first need to find the standard error.

$$\hat{\sigma}_{\bar{y}} = \frac{4}{\sqrt{400}} = .2$$

The t-statistic is then:

$$t = \frac{10 - 9.5}{.2} = \frac{.5}{.2} = 2.5$$

We can compare this to the critical value of t with 400 – 1 = 399 degrees of freedom. From the t-table, we can see that the critical value would fall between 1.984 and 1.96. Our calculated t-statistic is much larger in magnitude. The p-value would be less than .02. Using a Z-statistic would work fine.