Public Policy 529
Midterm Exam W17

Student ID number (8-digits): ____________________

1. Suppose that the number of hours people sleep has a normal distribution with a mean of 7 and a standard deviation of 1.2. (20 points; 5 points each part)

   (a) What percentage of people sleep less than 7.5 hours?

   $$z = \frac{7.5-7}{1.2} = \frac{0.5}{1.2} = 0.42$$

   The area in the upper tail for z=0.42 is .3372, so the area below is: 1 - .3372 = .6628. So the answer is 66.28%.

   Rounding to nearest percentage is fine.

   (b) What percentage of people sleep less than 5.5 hours?

   $$z = \frac{5.5-7}{1.2} = -\frac{1.5}{1.2} = -1.25$$

   The area in the upper tail for z=1.25 is .1056. By symmetry, this is the same size as the area below z=-1.25. So the answer is 10.56%.

   (c) What percentage of people sleep 6 to 8 hours?

   It is helpful to notice that 6 and 8 are equally distant from the mean, so the area above 8 is equal to the area below 6. If we find the z-score for 8, we can identify these areas.

   $$z = \frac{8-7}{1.2} = \frac{1}{1.2} = 0.83$$

   The area in the upper tail for z=0.83 is .2033 or 20.33%. The area below 6, with its z-score of -0.83, is of the same size. If we subtract these two areas in the tails from 1.00, we can find the area in the range of 6 to 8.

   $$1.00 - .2033 - .2033 = .5934$$

   So the answer is 59.34%.

   (d) If you were to take repeated samples of size n=200 from this population, calculating the sample mean each time, in what range would 95% of the sample means fall?

   From the Central Limit Theorem, we know that these sample means will be normally distributed with a mean of 7 and a standard deviation (i.e. standard error) of:

   $$\frac{\sigma}{\sqrt{n}} = \frac{1.2}{\sqrt{200}} = .08485$$

From what we know of normal distributions, 95% of the data will be within 1.96 standard deviations of the mean. So, we expect 95% of sample means to be in the range:

$$7 \pm 1.96(.08485) = 7 \pm .0166$$

This equates to a range of 6.83 to 7.17.

2. In a recent survey of people who voted for Donald Trump (n=712), 51% of respondents say that the Bowling Green Massacre shows why Trump's immigration policy is needed.

   (a) Calculate a 99% confidence interval for this estimate. (5 points)

   The estimated proportion is .51. The z-statistic used to produce a 99% confidence interval is 2.58 (allow 2.57 or 2.576). According to the formula, the interval is:

$$\begin{aligned} \hat{\pi} \pm z \cdot \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}} &= .51 \pm 2.58 \cdot \sqrt{\frac{.51(1-.51)}{712}} \\ &= .51 \pm 2.58(.0187) \\ &= .51 \pm .048 \end{aligned}$$

   The confidence interval ranges from .462 to .558. Rounding to 2 decimal places is okay.

   (b) Interpret this confidence interval. (5 points)

   There are a few possibilities here:

   "99% percent of confidence intervals constructed following these procedures will contain the true population proportion."

   "There is a 99% percent chance that a confidence interval constructed in this way will contain the true population proportion."

   "Assuming the true population proportion is .51, then 99% of samples of size 712 will produce sample proportions that fall into the range .462 to .588."

3. The table below is the joint probability distribution for two variables in the American population. One variable indicates whether a person was born in this country (yes, no). The other variable shows opinions on how important it is to speak English (very, fairly, not very, not at all).

Importance of	Born in this Country?		
Speaking English	Yes	No	Total
Very	.62	.11	.73
Fairly	.18	.03	.21
Not Very	.04	.00	.04
Not at All	.02	.00	.02
Total	.86	.14	1.00

(a) What are the measurement levels of these two variables?

The Born in Country variable is nominal; the Importance of Speaking English variable is ordinal.

(b) What is P(Yes or Fairly)? (4 points)

P(Yes) + P(Fairly) - P(Yes and Fairly) = .86 + .21 - .18 = .89

(c) What is P(No and Very)? (4 points)

P(No and Very) = .11

(d) What is P(Very | Yes)? What is P(Very | No)? (8 points; 4 each)

$$P(\text{Very}|\text{Yes}) = \frac{P(\text{Very and Yes})}{P(\text{Yes})} = \frac{.62}{.86} = .721$$

$$P(\text{Very}|\text{No}) = \frac{P(\text{Very and No})}{P(\text{No})} = \frac{.11}{.14} = .786$$

(e) Are these two variables independent of each other? How do you know? (4 points)

No, they are not independent. The two conditional probabilities calculated in part (d) are not equal to the unconditional probability: P(Very) = .73.

4. In random sample survey, the mean amount of time spent commuting to work or school was 25 minutes (s=15; n=900).

(a) Construct a 95% confidence interval for this mean. (5 points)

The z-statistic used to produce a 95% confidence interval is 1.96. According to the formula, the interval is:

$$\begin{aligned} \bar{y} \pm z \cdot \hat{\sigma}_{\bar{y}} &= 25 \pm 1.96 \cdot \frac{15}{\sqrt{900}} \\ &= 25 \pm 1.96(.5) \\ &= 25 \pm .98 \end{aligned}$$

The confidence interval ranges from 24.02 to 25.98. Rounding to integers is okay.

(b) Interpret this confidence interval. (5 points)

Use same style of answers as in 2(b).

5. Answer true or false to the following questions. (10 points; 2.5 points each)

   (a) According to the Central Limit Theorem, the sampling distribution of the sample mean is always normal even if the population distribution is highly skewed.

   False. For this condition to hold, our sample size must be at least 30.

   (b) A survey researcher calls only during daytime hours, producing a sample with a high proportion of retirees. The responses to the survey questions thus contain response bias.

   False. This is a selection bias problem.

   (c) The larger the standard deviation of a variable in the population, the lower the level of precision when estimating the population mean with a sample mean.

   True.

   (d) If a measurement strategy has validity, but low reliability, the consequence is bias.

   False. Bias comes from a validity problem (i.e. systematic measurement errors). Reliability is just about consistency.

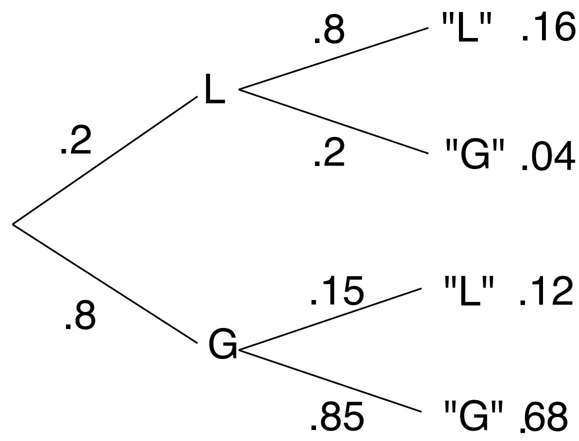
6. When people buy a used car that turns out to have lots of mechanical problems, they call it a "lemon." Suppose that 20% of used cars on the market are lemons. Before buying, you can take a car to a mechanic to check it out. If the car is a lemon, there is an 80% chance that the mechanic will identify it as a lemon. On the other hand, the mechanic will wrongly identify 15% of good cars as lemons.

   (a) Make a joint probability distribution table or probability tree to represent this situation. (9 points)

   Here is a joint probability distribution:

	Called Lemon	Called Good	Total
Lemon	.16	.04	.20
Good	.12	.68	.80
Total	.28	.72	1.00

   Here is a probability tree:



(b) What is the probability that a car taken to the mechanic will be called a lemon? (3 points)

P("Lemon") = .16 + .12 = .28

(c) What is the probability that a car is actually a lemon if the mechanic says it is okay? (3 points)

$$P(\text{Lemon}|\text{“Good”}) = \frac{P(\text{Lemon and “Good”})}{P(\text{“Good”})} = \frac{.04}{.72} = .056$$

7. The table below shows answers to a survey question that asks how often the respondent trusts other people.

How Often Trust Others?	Freq.	Percent	Cum.
1. Always Trust	37	2.99	2.99
2. Usually Trust	427	34.49	37.48
3. Usually Do Not Trust	628	50.73	88.21
4. Never Trust	146	11.79	100.00
Total	1,238	100.00	

(a) What is the measurement level of this variable? (2.5 points)

This is an ordinal variable.

(b) Calculate all appropriate measures of central tendency. (5 points; 2.5 points each)

The median is 3 ("Usually Do Not Trust"). The mode is also 3 ("Usually Do Not Trust").

(c) What proportion of people say that they always or usually trust others? (2.5 points)

.299 + .3449 = .3748

(d) Find an 80% confidence interval for the proportion that you found in part (c). (5 points)

The estimated proportion is .375. The z-statistic used to produce an 80% confidence interval is 1.28. According to the formula, the interval is:

$$\begin{aligned}
\hat{\pi} \pm z \cdot \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}} &= .375 \pm 1.28 \cdot \sqrt{\frac{.375(1-.375)}{1238}} \\
&= .375 \pm 1.28(.0138) \\
&= .375 \pm .0176
\end{aligned}$$

The confidence interval ranges from .357 to .393. Rounding to 2 decimal places is okay.