Public Policy 529
Midterm Exam

Student ID number (8-digits): ________________________

1. In the United States, the cost of visits to hospital emergency rooms has a normal distribution with a mean \$1,200 and a standard deviation of 300. (15 points; 5 points each part)

   (a) What percentage of visits cost more than \$1,000?

   $$z = \frac{1000 - 1200}{300} = -\frac{200}{300} = -0.67$$

   The area in the upper tail for z=0.67 is .2514, so the area below is: 1 - .2514 = .7486. By symmetry, this is the same size as the area above z=-0.67. So the answer is 74.86%.

   (b) What percentage of visits cost less than \$800?

   $$z = \frac{800 - 1200}{300} = -\frac{400}{300} = -1.33$$

   The area in the upper tail for z=1.33 is .0918. By symmetry, this is the same size as the area below z=-1.33. So the answer is 9.18%. Rounding to nearest percentage is ok.

   (c) For what percentage of visits are costs between \$1,200 to \$1,500?

   \$1,200 is the mean of the distribution, so 50% of the area lies above/below that value. It is not necessary to calculate the z-score to make this determination, but the z-score would be 0. We now find the z-score for \$1,500.

   $$z = \frac{1500 - 1200}{300} = -\frac{300}{300} = 1.00.$$

   The area in the upper tail for z=1.00 is .1587 or 15.87%. Subtracting this amount from 50% produces about 34%:

   $$50.00 - 15.87 = 34.13$$

2. True or False? The p-value from a significance test is calculated using the critical value of the test statistic. Explain your answer. (3 points)

   False. The critical value of the test-statistic reflects α, the level of significance. We construct the p-value from the value of the test statistic that we calculate using our sample statistic.

3. In a recent survey of supporters of presidential candidate Donald Trump (n=766), 81% said that they agree with Trump's statement that the election could be stolen from him as a result of voter fraud. The other supporters did not agree with the statement.

(a) Calculate a 95% confidence interval for this estimate. (5 points)

The estimated proportion is .81. The formula for the confidence interval is:

$$\begin{aligned}
\hat{\pi} \pm z \cdot \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}} &= .81 \pm 1.96 \cdot \sqrt{\frac{.81(1-.81)}{766}} \\
&= .81 \pm 1.96(.014) \\
&= .81 \pm .028
\end{aligned}$$

The confidence interval ranges from .782 to .838.

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

There are a few possibilities here:

"95% percent of confidence intervals constructed following these procedures will contain the true population proportion."

"There is a 95% percent chance that a confidence interval constructed in this way will contain the true population proportion."

"Assuming the true population proportion is .81, then 95% of samples of size 766 will produce sample proportions that fall into the range .782 to .838."

(c) Perform a significance test in which the null hypothesis is 84% Trump supporters agree with the statement ($\alpha = .05$). Report the test statistic, critical value of test statistic, and p-value. (10 points)

Hypotheses:

H_0: $\pi = .84$

H_a: $\pi \neq .84$

We use a z-statistic for difference of proportions test in which there is at least 15 cases in each category, which is the case here. With $\alpha = .05$, the critical value of z needed to reject the null hypothesis is 1.96 in absolute value.

The value of z for this significance test is:

$$z = \frac{.81 - .84}{\sqrt{\frac{.84(1-.84)}{766}}} = \frac{.03}{.0132} = -2.26$$

Since $|z| > 1.96$, we can reject the null hypothesis that the true proportion is .84. The p-value is $2 \times .0119 = .0238$ (using the z-table for 2.26).

4. The variable `SpendingCategory` comes from a public opinion survey that asked respondents which federal budget category has the lowest spending: foreign aid, national defense, or Social Security. The survey also asked whether the respondent voted in the 2012 election (`Vote2012`). The resulting data are presented in the table below. The cells contain frequencies.

Spending Category	Did Not Vote	Voted	Total
Foreign Aid	130	178	308
National Defense	57	78	135
Social Security	123	168	291
Total	310	424	734

(a) What is P(Foreign Aid or Voted)? (5 points)

$$\begin{aligned} P(\text{Foreign Aid or Voted}) &= P(\text{Foreign Aid}) + P(\text{Voted}) - P(\text{Foreign Aid and Voted}) \\ &= \frac{308}{734} + \frac{424}{734} - \frac{178}{734} \\ &= .4196 + .5578 - .2425 \\ &= .7548 \end{aligned}$$

Alternatively, the following method will work:

$$\begin{aligned} P(\text{Foreign Aid or Voted}) &= \frac{308 + 424 - 178}{734} \\ &= \frac{554}{734} \\ &= .7548 \end{aligned}$$

(b) Are the variables `SpendingCategory` and `Vote2012` independent of each other? Demonstrate mathematically (round all probabilities to 2 decimal places). (7 points)

Yes, these two variables are independent of each other. We can see that because all of the conditional probabilities are equal to the unconditional probabilities.

To prove the point, we need to check at least enough conditional probabilities to be sure that all are equal for one of the two columns or two of the three rows. This means at least two conditional probabilities must match their associated unconditional probability. Checking just one conditional probability vs. one unconditional probability is not enough to be sure (see more below).

$$
\begin{aligned}
P(\text{Did Not Vote }) &= \frac{310}{734} = .42 \\
P(\text{Did Not Vote} \mid \text{Foreign Aid}) &= \frac{130}{308} = .42 \\
P(\text{Did Not Vote} \mid \text{National Defense}) &= \frac{57}{135} = .42 \\
P(\text{Did Note Vote} \mid \text{Social Security}) &= \frac{123}{291} = .42
\end{aligned}
$$

$$
\begin{aligned}
P(\text{Voted}) &= \frac{424}{734} = .58 \\
P(\text{Voted} \mid \text{Foreign Aid}) &= \frac{178}{308} = .58 \\
P(\text{Voted} \mid \text{National Defense}) &= \frac{78}{135} = .58 \\
P(\text{Voted} \mid \text{Social Security}) &= \frac{168}{291} = .58
\end{aligned}
$$

$$
\begin{aligned}
P(\text{Foreign Aid}) &= \frac{308}{734} = .42 \\
P(\text{Foreign Aid} \mid \text{Did Not Vote}) &= \frac{130}{310} = .42 \\
P(\text{Foreign Aid} \mid \text{Voted}) &= \frac{178}{424} = .42
\end{aligned}
$$

$$
\begin{aligned}
P(\text{National Defense}) &= \frac{135}{734} = .18 \\
P(\text{National Defense} \mid \text{Did Not Vote}) &= \frac{57}{310} = .18 \\
P(\text{National Defense} \mid \text{Voted}) &= \frac{78}{424} = .18
\end{aligned}
$$

$$
\begin{aligned}
P(\text{Social Security}) &= \frac{291}{734} = .40 \\
P(\text{Social Security} \mid \text{Did Not Vote}) &= \frac{123}{310} = .40 \\
P(\text{Social Security} \mid \text{Voted}) &= \frac{168}{424} = .40
\end{aligned}
$$

For example, if we focus on the `Vote2012` variable, we need P(Did Not Vote) to equal any two conditional probabilities of the form P(Did Not Vote | [spending category]). Mathematically, it then will be true that the third conditional probability will also equal P(Did Note Vote). Additionally, it would then be true that P(Voted) is equal to all three conditional probabilities of P(Voted | [spending category]).

Alternatively, we could look at any two spending categories. For example, if P(Foreign Aid) =P(Foreign Aid | Did not Vote) and P(National Defense) = P(National Defense | Did Not Vote), then the remaining unconditional probabilities will equal their associated conditional probabilities.

Example: it could be that P(Foreign Aid) = P(Foreign Aid | Did Not Vote). It is still possible that the allocation of voters vs. non-voters in the other two spending categories makes P(National Defense ≠ P(National Defense | Did Not Vote) and P(Social Security) ≠ P(Social Security | Did Not Vote). Once we have determined either one these two equations holds with equality, however, the third must as well.

5. In a sample of 25 people who completed a job training program, the mean amount of time needed to find a job was 45 days with a standard deviation of 15. Can we reject the null hypothesis that the true mean time is 40 days ($\alpha = .05$)? Report your test statistic, the critical value of the test statistic, and p-value (approximate if necessary).

   In this case, we are performing a significance test that involves a mean, and our sample size is less than 30. We should use a t-statistic. We have $n-1$=24 degrees of freedom. At $\alpha = .05$, the critical value of the t-statistic will be 2.064, which comes from the column on the t-table where the $t_{.025}$ column intersects the row for 24 degrees of freedom.

   H_0: $\mu = 40$
   H_A: $\mu \neq 40$

$$t = \frac{45-40}{\frac{15}{\sqrt{25}}} = \frac{5}{3} = 1.67$$

   Since $|t| < 2.064$, we fail to reject the null hypothesis.

   Based upon the table, the p-value associated with a t-statistic of 1.67 at 24 degrees of freedom is in the range: $2\times.05 < p < 2\times.1$, which is $.1 < p < .2$. Since $p > .05$, we fail reject the null hypothesis.

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

   (a) A researcher has two samples from the same population with $\sigma = 10$. In Sample A, n=1,200. In Sample B, n=200. According to the Central Limit Theorem, the mean from Sample A is closer to the population mean than is the mean from Sample B.

   False. The Central Limit Theorem just says that sample means will be closer on average when sample sizes are larger. The mean from a specific small sample could very well be closer than the mean from a specific large sample.

   (b) A t-statistic is an appropriate test statistic in all scenarios for which a z-statistic is appropriate, but the reverse is not true.

   False. We cannot use t-statistics for tests involving proportions.

   (c) A measurement strategy can be biased but still perfectly reliable.

   True. Reliability is about consistency. If the bias is consistent, the measure is still reliable.

(d) Random measurement errors affect the reliability of a measurement strategy but do not affect its validity.

False. Validity is about about aiming at the right target with accuracy. Large measurement errors, even if random, reduce validity.

7. Suppose that, in the population, the mean feeling thermometer score for Barack Obama is 50 with a standard deviation of 30. You will take a random sample of size 800. In what range will 80% of the possible sample means fall? (5 points)

From the central limit theorem, we can apply the following formula to obtain the sampling distribution for the sample mean.

$$\mu_y \pm z \cdot \frac{\sigma}{\sqrt{n}}$$

We need to find the value of z that will provide an 80% confidence interval. This value would leave 10% of the area in each tail. From the z table, we see that z=1.28 puts an area of .1003 in the right-hand tail.

$$\begin{aligned}
\mu_y \pm 1.28 \cdot \frac{\sigma}{\sqrt{n}} &= 50 \pm 1.28 \cdot \frac{30}{\sqrt{800}} \\
&= 50 \pm 1.28 \cdot 1.06 \\
&= 14 \pm 1.36
\end{aligned}$$

Accordingly, 80% of sample means should be in the range of 48.64 to 51.36. Allow for rounding.

8. Suppose that 20% of the population has been exposed to a virus, and people can be tested for whether they have been exposed (test positive or negative). The test has a false positive rate of 10% (one out of every ten people who *have not* been exposed will test positive). The test has a false negative rate of 30% (three of of every ten people who *have been* exposed will test negative).

It may help to make a Venn diagram, joint probability distribution table, or probability tree. Here is a joint probability distribution:

	Tests Positive	Tests Negative	Total
Exposed	.14	.06	.20
Not exposed	.08	.72	.80
Total	.22	.78	1.00

(a) What is P(Not Exposed and Tests Positive)? (5 points)

$$\begin{aligned}
P(\text{Not Exposed and Positive}) &= P(\text{Not Exposed}) \times P(\text{Positive} \mid \text{Not Exposed}) \\
&= .8 \times .1 = .08
\end{aligned}$$

(b) What is P(Exposed and Tests Positive)? (5 points)

$$\begin{aligned}
P(\text{Exposed and Positive}) &= P(\text{Exposed}) \times P(\text{Positive} \mid \text{Exposed}) \\
&= .2 \times .7 = .14
\end{aligned}$$

(c) What is P(Exposed | Positive)? (5 points)

$$\begin{aligned} P(\text{Exposed} \mid \text{Positive}) &= \frac{P(\text{Exposed and Positive})}{P(\text{Positive})} \\ &= \frac{.14}{.08 + .14} \\ &= .636 \end{aligned}$$