Public Policy 529: PRACTICE Midterm Exam

Student id number (8-digit): _____ANSWER KEY_____

Instructions: Read each question carefully and be sure you answer every part of every question. **Show all your work.** If you have difficulty with a question early on, my advice is to skip to the next question and come back to the former when you have completed the other questions. Good luck!

1. MULTIPLE CHOICE. Which of the following is implied by the Central Limit Theorem (select ALL that apply)? (4 points)

a. All variables have normal-shaped sample distributions if a random sample contains at least around 30 observations. **FALSE**
b. Population distributions are approximately normal whenever the population size is large. **FALSE**
c. For large random samples, the sampling distribution of the sample mean is approximately normal, regardless of the shape of the population distribution. **TRUE**
d. The standard error of the sample mean decreases as the sample size increases. **TRUE**

2. Suppose there is a room with 4 people. Show how you would calculate the probability that somebody in the room shares a birthday. You do not need to simplify the expression. (4 points)

It's easier to calculate the probability of the complement.
1 – prob(no one shares a birthday) = 1 – 365/365 * 364/365 * 363/365 * 362/365

3. When polled in 2006 and asked whether Turkey should be included in the European Union if it met all conditions set by the EU, the percentage who said yes was 42% in the UK. The sample size was 1,312.

i. Calculate a 95% confidence interval for the portion of the UK population that believes Turkey should be included in the EU. (5 points)

Since n>1,000, the Z- and t-statistics will be identical. The z-score for a 95% confidence interval is 1.96.
CI = 0.42 +/- 1.96 * (0.42 * 0.58 / 1312) ^ 0.5
= 0.42 +/- 1.96 * 0.014
= (0.39, 0.45)

ii. MULTIPLE CHOICE. Which of the following is NOT a correct interpretation of the confidence interval (select ONE)? (3 points)

a. We arrived at these numbers using a method that gives an interval including the true proportion 95% of the time.
b. We are 95% confident that the true sample proportion lies in this range. **THIS IS NOT CORRECT.**
c. 95% of samples of this size will produce confidence intervals that capture the true proportion.
d. We are 95% certain that this range includes the true population parameter.

iii. What is the margin of error for the confidence interval in part (i)? (3 points)

margin of error = z * s.e. = 1.96 * (0.42 * 0.58 / 1312) ^ 0.5
0.027
(You do not need to show work, if you showed it above. Note also that ^.05 means to take the square root of the number in the parentheses.)

iv. What is the probability that a random sample was selected such that the confidence interval in part (ii) does not contain the population mean? (3 points)

= 1 – 0.95 = 0.05

4. Suppose there's a genetic mutation present in 10% of the population. There's a test that, if you have the mutation, will give you an accurate result (i.e. test positive) 90% of the time. If you do not have the mutation, the test will give you an accurate result (i.e. test negative) 95% of the time.

i. Write out a contingency table using cell probabilities (with the probability of having the disease along one dimension, and the probability of testing positive along the other dimension). (5 points)

	Test Result		
Disease Status	**Positive**	**Negative**	**Total**
Has disease	**0.90 *0.10 =0.09**	**0.10 *0.10 =0.01**	**0.10**
Does not have	**0.05 *0.90 =0.045**	**0.95 *0.90 =0.855**	**0.90**
Total	**0.135**	**0.865**	**1.00**

ii. Find the probability of having the disease, conditional on testing positive. (4 points)

=0.09/0.135 = 67%

iii. MULTIPLE CHOICE. Compare your answer in part (ii) to the unconditional probability of having the disease, and select the correct interpretation (choose one). (3 points)

a. The conditional probability is higher than the unconditional probability because taking the test has added information. **THIS IS CORRECT.**
b. The conditional probability is lower than the unconditional probability because taking the test has added information.
c. The conditional probability and the unconditional probability are equal, because obtaining a test result and having the disease are independent.
d. The conditional probability is higher than the unconditional probability, even though obtaining a test result and having the disease are independent.
e. The conditional probability is lower than the unconditional probability, because the test result is frequently incorrect.

5. Assume the population distribution of the scores on the Psychomotor Development Index (PDI), a scale of infant development, is normal with mean 110 and standard deviation 16.

i. What is the probability of a PDI below 97? (4 points)

We use Z-statistics here because the question refers to the population, not a sample. Also, the standard deviation is known.

z = 97 – 110 / 16 = -0.8125
From Table A (and using symmetry), probability is 20%.

ii. A study uses a random sample of 81 infants and finds a mean of 109 and a standard deviation of 19. Find the 98% confidence interval for the mean. (5 points)

Given the sample size, a t-statistic should be our first choice, though the Z will offer a close approximation. Either method would be considered correct.

The t-statistic for a 98% confidence interval: From Table B, 2.374.

CI = 109 +/- t*19/(81^0.5) = 109 +/- 2.37 * 19/9 = (104.0, 114.0)

The z-score for a 98% confidence interval: From Table A, 2.32 or 2.33 (rounding error is fine).

CI = 109 +/- z * 19/(81^0.5) = 109 +/- 2.33 * 19/9 = (104.1, 113.9)

iii. EXTRA CREDIT. Comparing your answers to part (i) and part (ii), which of the following is correct (select one)? (2 points)

a. The confidence interval in part (ii) does not include 97. This makes sense, because the sample is biased. **FALSE**.

b. The confidence interval in part (ii) does not include 97. This makes sense, because it is a 98% confidence interval whereas in part (i) I calculate a smaller probability. **FALSE**.

c. The confidence interval in part (ii) does contain 97, by the Central Limit Theorem. **FALSE**.

d. Part (i) refers to the population distribution, whereas part (ii) refers to the sampling distribution of the mean. **TRUE**.