**Practice Midterm Exam**

Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Part I\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_/25

Part II \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ /75

Total \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_/100

**Part I (25 points total): Write down the single best answer to each question in the space provided below. Each question is worth 25 points.**

Answer to 1\_\_\_\_\_\_\_\_\_\_\_

Answer to 2\_\_\_\_\_\_\_\_\_\_\_

Answer to 3\_\_\_\_\_\_\_\_\_\_\_

Answer to 4\_\_\_\_\_\_\_\_\_\_\_

Answer to 5\_\_\_\_\_\_\_\_\_\_\_

1. To study the weight distribution of newborns, a researcher randomly samples 3226 newborns. The weight distribution of these 3226 newborns is presented below:



The weight distribution of the newborns in this sample is\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Normally distributed because the histogram has a bell shape and is symmetric around the mean
2. Normally distributed because the two tails are flat.
3. Normally distributed because of the central limit theorem (N=3226>100)
4. Not enough information
5. A sample of 100 newborns has a normal distribution with an average of 3.5kg and the standard deviation of 0.5kg. The standard error is 0.05. A newborn named Richard is heavier than 84.13 percent of the newborns in the sample. Richard’s weight is\_\_\_\_\_\_\_
6. 3 kg
7. 3.45 kg
8. 3.55kg
9. 4kg
10. Not enough information.
11. Professor Pham surveys a random sample of 100 households in a community of 10,000 households. The average household income in this sample is $86,000. The sample standard deviation is $5000. The sampling distribution of the sample means has the mean of \_\_\_
12. $81,000
13. $85,500
14. $86,000
15. Not enough information
16. A company of 10,000 employees buys 100 raffle tickets for the company’s holiday party. The HR head checks the value of 16 random raffle tickets. She finds among the 16 raffle tickets that she checks, the average value is $22, and the standard deviation is $24. Assume that the sampling distribution of the sample means is normal and the population standard deviation equals sample standard deviation. The 95% confidence interval is ($16, $28).
	1. This confidence interval is not valid because N=16<100, so the central limit theorem does not apply.
	2. We are 95% confident that the average value of the 16 raffle tickets that the company buys is between $16 and $28.
	3. We are 95% confident that the average value of the 100 raffle tickets that the company buys is between $16 and $28.
	4. b&c.
	5. None of the above
17. .$H\_{0}:μ=12$, z-obtained= 2.5, significance level $α=0.05$. Assume that the sampling distribution is normal. We \_\_\_\_\_\_\_\_\_\_\_\_
18. Reject the null hypothesis because if the null hypothesis is true, we would be more likely to obtain a sample mean further away from 12.
19. Reject the null hypothesis because if the null hypothesis is true, we would be more likely to obtain a sample mean closer to 12.
20. Reject the null hypothesis because 2.5>0.05
21. Cannot decide because we do not know the sample mean.

**Part II (75 points total): Short Answers**

1. (10 points) The General Social Survey (GSS) has been administered to randomly selected samples of adults Americans since 1972 and explores a broad range of characteristics and issues, including confidence in the Supreme Court, attitudes about assisted suicide, number of siblings, and level of education. Please visit the GSS website for more information (<https://gss.norc.org/>).

The “GSS\_2018.dta” data on Blackboard is the GSS data in 2018.

Variable “*natrace*” in GSS\_2018 records the answer to the following question: “In your opinion, are we spending too much, too little, or about the right amount of money on improving the conditions of Blacks?” You can see part of the 2018 GSS questionnaire here: https://gss.norc.org/Documents/quex/GSS2018%20Ballot%201%20-%20English.pdf

Please summarize the opinions of the American adults in GSS\_2018 on the amount of money the US is spending on improving the conditions of Blacks from the *natrace* variable. You can present your finding in the format of a table, a graph, or both. Please be sure to also explain your finding in words.

**Important note about the data: On Monday Sept 28, I updated GSS\_2018 data on Blackboard to make the “age” variable easier to analyze. Please re-download the GSS\_2018.dta.**

1. a) (10 points) Examine the *age* variable in GSS\_2018.dta data on Blackboard. We want to test whether the average age of all American adults is 50 years old. How many standard deviations is the sample average age in the GSS2018 away from the hypothesized population mean of 50 years old? Show your work.

 b) (5 points) In class, what do we call the measurement that you just calculated?

1. In a large university, a researcher selected a random sample of 100 former PhD students in science, technology, engineering, and mathematics (STEM) from a list of PhD graduates at this university. She discovered that it had taken an average of 6 years for STEM graduates to finish their degrees. The sample standard deviation is 2 years. Assume that the population standard deviation equals the sample standard deviation.

In this university, non-STEM PhDs took about 5.5 years to finish their degrees.

1. (10 points) Test whether STEM PhDs, on average, take longer to finish their degrees than other PhD students at this university? Use a two-tailed z-test, **z-obtained method (not the confidence interval method)**, and a 5% significance level. Follow the five-step hypothesis-testing model.
2. (10 points) Interpret the significance level $α=0.05$ in this context.

Hint: $α=$ type I error. Think about what type I error means in this context.

Note: Please be concise. I won’t read beyond the fourth line.

1. (10 points) Draw a theoretical sampling distribution under the assumption that the null hypothesis in (a) was true. Please label where 6 and 5.5 are, significance level alpha=0.05, the z critical value at $α=0.05$, and the z-obtained in part (a). Shade the area of the p-value. You don’t need to calculate the p-value here.
* Hint: Think about how many standard deviations 6 and 5.5 are from the mean of the sampling distribution. Also, you only need to label what on the question. Don’t need to calculate more numbers to label.
* **Instruction:** If you do not use a computer software to draw the graph, you can hand draw it on a piece of paper, take a picture of the graph, and paste the picture on the Answer Sheet.
1. (10 points) What is the hypothesis testing conclusion in part (a) if $α=0.008$. Now use the confidence interval method with the appropriate confidence interval to the significance level alpha=0.008. Explain.

Note: You don’t need to show all the 5 steps of the hypothesis testing model. You only need to show the steps that give different results from part (a).

1. (5 points) Interpret the confidence interval you calculated you part (d )
2. (5 points) Use the confidence interval in part (d) to explain what null hypothesis values that we can reject.