

Problem Set 3

PO 7001 — Quantitative Methods I

100pts total

1. 8 students are asked to indicate how important grades are, on a scale from one to 10. We observe the following responses. Calculate the 95% confidence interval of the mean of this sample, and interpret this interval. How confident are you in your results and why? (you may use R if you want, but no canned “confidence interval” type of package) (8pts)

ID	response
1	2
2	5
3	9
4	1
5	6
6	6
7	8
8	2

2. Consider the following sample.

ID	response
1	10
2	20
3	5
4	45
5	12
6	73
7	45
8	2200

Calculate the 95% confidence interval of the mean of this sample, and interpret this interval. How confident are you in your results and why? (you may use R if you want, but no canned “confidence interval” type of package) (8pts)

3. You are told that the 95% confidence interval for a sample of size 100 is $(-3, 5)$. Can you then infer what the mean and standard deviation of that sample is? Why or why not? (8pts)
4. You are told that the 95% confidence interval for a sample X of size 100 is $(-3, 5)$, and that the 95% confidence interval for a sample Y of size 100 is $(3, 7)$. Can you conclude that the means of the two samples are significantly different? Justify your answer. (8pts)
5. What is wrong, if anything, in the following situations? (4pt each)
 - a. A random sample of size 20 is taken from a population that is assumed to have a standard deviation of 12. The standard deviation of the sample mean is $12/20$
 - b. A study says that the results are statistically significant at the 1% level and the P-value is 0.99.
6. A test of the null hypothesis $H_0 : \mu = \mu_0$ gives the test statistic $z = 1.34$. What is the p-value if the alternative is $H_a : \mu > \mu_0$? (5pts)
7. The P-value for a two-sided test of the null hypothesis $H_0 : \bar{x} = 30$ is 0.04.
 - (a) does the 95% confidence interval include the value 30? Why? (5pts)
 - (b) does the 90% confidence interval include the value 30? Why? (4pts)
8. A study found that the use of bed nets was associated with a lower prevalence of malarial infections in the Gambia. A report of the study states that the significance is $P < 0.001$. Explain what this means in a way that could be understood by someone who has not studied statistics. (6pts)
9. A research report described two results that both achieved statistical significance at the 5% level. The P-value for the first is 0.049; for the second it is 0.00002. Do the P-values add any useful information beyond that conveyed by the statement that both results are statistically significant? Write a short paragraph explaining your views on this question (6pts)
10. A P-value of 0.90 is reported for a significance test for a population mean. Interpret this result. (4pts)
11. Explain whether a test of significance can answer each of the following questions. (3pts each)
 - a. Is the sample or experiment properly designed?
 - b. Is the observed effect compatible with the null hypothesis?
 - c. Is the observed effect large?
12. A study with 7,500 subjects reported a result that was statistically significant at the 5% level. Explain why this result might not be particularly large or

important. (5pts)

13. Assume a sample size of $n = 20$.
- Use R to draw a picture of the distribution of the t statistic under the null hypothesis. (4pts)
 - What value of the test statistic would lead to a rejection of the null hypothesis at the 5% level for a two-sided alternative? (4pts)
14. Your software reports $\bar{x} = 15.3$ and $P = 0.04$ for a t test of $H_0 : \mu = 0$ versus $H_a : \mu \neq 0$. Based on prior knowledge, you can justify testing the alternative $H_a : \mu > 0$. What is the P-value for your significance test? (4pts)
15. Consider the following distribution of males/females in two cities (each cell reports the number of individuals):

	City 1	City 2
Men	38	68
Women	203	150

Calculate the χ^2 statistic both manually and with R. Report the associated p -value using R. What do you conclude? (8pts)