Stat 6021: Module A Practice Questions

Topic A.2: Sampling Distributions

1. Statistical theory tells us the distribution of the sample means with a fixed sample size, under certain circumstances. The sampling distribution is an approximation of the density histogram of the sample means. We know the sample means vary from sample to sample. The sampling distribution tells us the expected value (mean) of the distribution, and the standard deviation of the sample means.

   (a) Suppose the variable X follows a normal distribution with mean μ and standard deviation σ. Consider taking random samples, each with size n, repeatedly. What is the sampling distribution of the sample mean, $\bar{x}$?

   (b) Suppose the variable X has an unknown distribution but known mean μ and known standard deviation σ. What is the name of the statistical theory that informs us that the sampling distribution of the sample mean, $\bar{x}$, can be well-approximated by a normal distribution?

2. An automatic machine in a manufacturing process produces subcomponents. The lengths of the subcomponents follow a normal distribution with a mean of 116 cm and a standard deviation of 4.8 cm.

   (a) Find the probability that one selected subcomponent is longer than 118cm.

   (b) Find the probability that if 3 subcomponents are randomly selected, their mean length exceeds 118cm.

Topic A.3: Confidence Intervals

3. What are the goals of constructing a confidence interval?

4. How does increasing the confidence level affect the margin of error and the width of the confidence interval? Hint: sketching the standard normal distribution will be helpful.

5. How does increasing the sample size affect the margin of error and the width of the confidence interval? Briefly explain.

6. Use R to find the value of the t-multiplier when constructing a confidence interval for the mean in the following situations:

   (a) 94% confidence interval with $n = 49$.

   (b) 86% confidence interval with $n = 82$.

   (c) 74% confidence interval with $n = 150$.

   Hint: you will need either the `qt()` or `pt()` function.

7. A random sample of 100 students had a mean grade point average (GPA) of 3.2 with a standard deviation of 0.2.

   (a) Calculate a 97% confidence interval for the mean GPA for all students.

   (b) What is the margin of error for the confidence interval found in the previous part? What is the margin of error telling us?

   (c) Based on this confidence interval, is it reasonable to say that the mean GPA of all students is 3.25 or greater?

Topic A.4: Hypothesis Testing

8. What is a goal of conducting a hypothesis test?

9. Hypothesis statements are always about the **population parameter / sample statistic** (choose one) of interest.

10. For each of the situations, state the appropriate null and alternative hypotheses, in symbols and in words. Sketch how you would find the p-value based on the calculated test statistic.

    (a) David's car averages 29 miles per gallon on the highway. He just switched to a new motor oil that is advertised as increasing gas mileage. He wants to investigate if the advertisement is accurate.

    (b) The diameter of a spindle in a small motor is supposed to be 4 millimeters. If the spindle is too small or too large, the motor will not function properly. The manufacturer wants to investigate whether the mean diameter is moved away from the target.

    (c) The average time in traffic between 2 points of a congested highway used to be 2 hours. The government invested money to improve travel times by building extra lanes and overpasses. Citizens want to access if travel times have improved, on average.

11. To have more evidence against the null hypothesis, our test statistic should be **larger / smaller** (choose one) in magnitude. Briefly explain.

12. How does increasing the difference between the sample mean and the population mean under the null hypothesis affect the test statistic and the evidence against the null hypothesis?

13. How does increasing the sample size affect the test statistic and the evidence against the null hypothesis?

14. Use R to obtain the critical values of the following hypothesis tests:

    (a) $H_0 : \mu = 3.5, H_a : \mu \neq 3.5$, with $\alpha = 0.08$ and $n = 96$.

    (b) $H_0 : \mu = 75, H_a : \mu < 75$, with $\alpha = 0.12$ and $n = 43$.

    (c) $H_0 : \mu = 10, H_a : \mu > 10$, with $\alpha = 0.045$ and $n = 132$.

    Hint: you will need either the `qt()` or `pt()` function.

15. Use R to obtain the p-values of the following hypothesis tests:

    (a) $H_0 : \mu = 48, H_a : \mu \neq 48$, with $t - stat = 2.14$ and $n = 50$.

    (b) $H_0 : \mu = 3, H_a : \mu > 3$, with $t - stat = 0.78$ and $n = 316$.

    (c) $H_0 : \mu = 12, H_a : \mu < 12$, with $t - stat = 1.57$ and $n = 34$.

    Hint: you will need either the `qt()` or `pt()` function.

16. The 10-year historical average yield of corn in the United States is 160 bushels per acre. A survey of 50 farmers this year gives a sample mean yield of 158.4 bushels per acre, with a standard deviation of 5 bushels per acre. Does this sample provide evidence that the yield of corn has decreased from the 10-year historical average? Conduct an appropriate hypothesis test.

    (a) State the null and alternative hypotheses.

    (b) Calculate the test-statistic.

    (c) Find the p-value and critical value.

    (d) State a conclusion in context.

    (e) How would you interpret the calculated p-value?

General Questions

17. Obtain the critical value of a hypothesis test where $H_0 : \mu = 145, H_a : \mu \neq 145$, with significance level $\alpha = 0.02$. Suppose the sample size is 50.

18. Obtain the t-multiplier for a 98% confidence interval. Suppose the sample size is 50.

19. Compare the critical value and the t-multiplier found in the previous 2 parts. What is an implication based on this comparison?

20. Suppose the hypothesis test in part 17 is carried out and the p-value is 0.043. Which of the following confidence intervals is/are possible?

    - (143.2, 144.5)
    - (151.3, 154.6)
    - (144.5, 163,5)

21. A random sample of 85 banded archerfish were collected, and their lengths were measured and recorded. Their average length was 20cm with a standard deviation of 3cm.

    (a) Construct a 95% confidence interval for the population mean length of banded archerfish.

    (b) Based on your confidence interval, is it plausible that the population mean length of banded archerfish is 21cm? Briefly explain.

    (c) Suppose you conduct the following hypothesis test. $H_0 : \mu = 21, H_a : \mu \neq 21$. Without actually performing any additional calculations, what do you expect the p-value of this hypothesis test will be? Briefly explain.

    - greater than 0.05
    - less than 0.05

    (d) Conduct the hypothesis test to verify your answer to the previous part.