Stat 6021: Module A Practice Questions Solutions

Topic A.2: Sampling Distributions

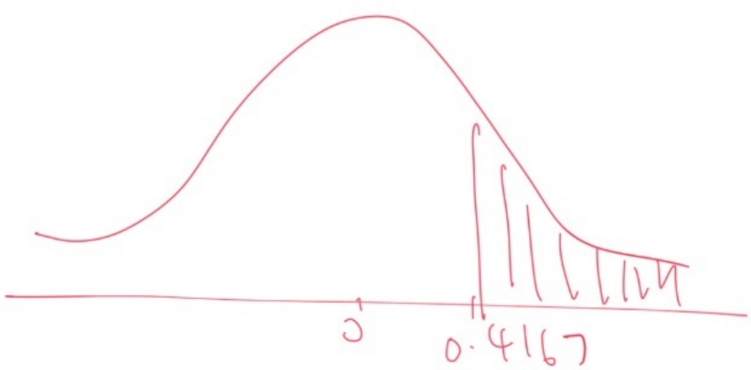
1. (a) Since $X \sim N(\mu, \sigma)$, then $\bar{x} \sim N(\mu, \frac{\sigma}{\sqrt{n}}$.

   (b) The central limit theorem.

2. (a) Let X denote the length of a subcomponent. Then $X \sim N(116, 4.8)$. Therefore,

$$\begin{aligned} P(X > 118) &= P(Z > \frac{118 - 116}{4.8} \\ &= P(Z > 0.4167) \\ &= 0.3385 \end{aligned}$$

   The corresponding sketch of the standard normal distribution is

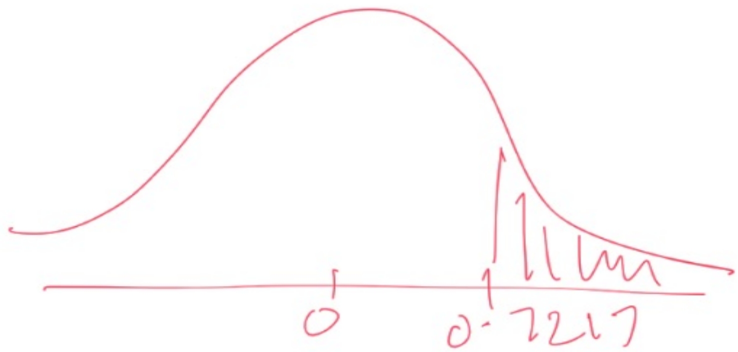


   so the probability is found by typing `1-pnorm(0.4167)`.

   (b) For the sample mean of 3 subcomponents, $\bar{x} \sim N(116, \frac{4.8}{\sqrt{3}}$. Therefore,

$$\begin{aligned} P(\bar{x} > 118) &= P(Z > \frac{118 - 116}{4.8/\sqrt{3}} \\ &= P(Z > 0.7217) \\ &= 0.2352 \end{aligned}$$

   The corresponding sketch of the standard normal distribution is



so the probability is found by typing `1-pnorm(0.7217)`.

Topic A.3: Confidence Intervals

3. - Provide an estimate for the unknown parameter of interest
   - Provide a range of plausible values for the unknown parameter of interest
   - Provide a measure of uncertainty

4. As the level of confidence increases, the multiplier increases in magnitude. Therefore, the margin of error increases.

5. As the sample size increases, the standard error of the sample mean decreases, since $se(\bar{x} = s/\sqrt{n}$. Therefore, the margin of error decreases.

6. To find the t multiplier for a $(1-\alpha)\times 100\%$ confidence interval, we use `qt(1-alpha/2, df)` where $df = n - 1$ for a confidence interval for the mean.

   (a) `qt(0.97,48)` which gives 1.9263

   (b) `qt(0.93,81)` which gives 1.4904

   (c) `qt(0.87,149)` which gives 1.1307

7. (a) We have $\bar{x} = 3.2, s = 0.2, n = 100$. Since the confidence level is 97%, the t-multiplier is found using `qt(0.985, 99)`, which gives 2.202. Therefore, the 97% confidence interval is

$$\begin{aligned} \bar{x} &\pm t_{1-\alpha/2,df} \times \frac{s}{\sqrt{n}} \\ 3.2 &\pm 2.202 \times \frac{0.2}{\sqrt{100}} \end{aligned}$$

   which gives (3.156, 3.244).

(b) The margin of error is

$$t_{1-\alpha/2,df} \times \frac{s}{\sqrt{n}} = 2.202 \times \frac{0.2}{\sqrt{100}} = 0.044$$

This means that in 97% of random samples of students, each with 100 students, the difference between its sample mean and the true population mean will be at most 0.044.

(c) No, it is not reasonable to say the mean GPA of all students is 3.25 or greater since the entire confidence interval is below 3.25. In fact, it will be reasonable to say that the mean GPA of all students is below 3.25, since the entire confidence interval is below 3.25.

Topic A.4: Hypothesis Testing

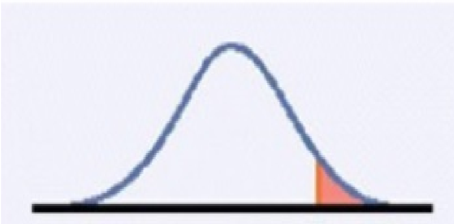
8. To distinguish patterns seen in data between those that are due to chance and those that reflect a real feature of the phenomenon under study.

9. Hypothesis statements are always about the **population parameter** of interest.

10. (a) $H_0 : \mu = 29$. The average gas mileage with the new motor oil is 29 mpg.

    $Ha : \mu > 29$. The average gas mileage with the new motor oil is greater than 29 mpg (i.e. the gas mileage has increased as claimed)

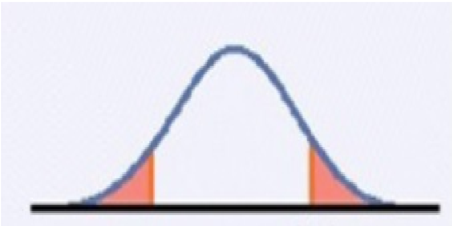
    Since our alternative hypothesis is a "greater than" statement, the p-value is the area to the right of the t statistic.

    (b) $H_0 : \mu = 4$. The average diameter of the spindles is 4 millimeters.

    $Ha : \mu \neq 4$. The average diameter of the spindles is different from 4 millimeters.

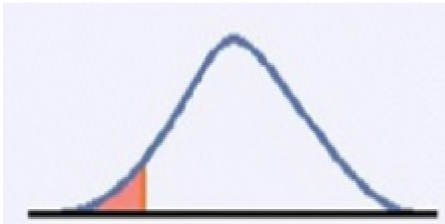
    Since our alternative hypothesis is a "different from" statement (2-sided), the p-value is the areas at the 2 tail ends of the t distribution.

(c) $H_0 : \mu = 2$. The average travel time is 2 hours.

$Ha : \mu < 2$. The average travel time is less than 2 hours (travel time improved).

Since our alternative hypothesis is a "less than" statement, the p-value is the area to the left of the t statistic.



11. To have more evidence against the null hypothesis, our test statistic should be **larger** in magnitude. The test statistic measures how dissimilar our sample is with the value under the null hypothesis.

12. Test statistic increases, which leads to more evidence against the null hypothesis.

13. Test statistic increases as sample size increases. More evidence against the null hypothesis.

14. For a two-sided test, the critical value for a hypothesis test at α is found using `qt(1-alpha/2, df)`. For a 1-sided test, the critical value is found using `qt(1-alpha, df)`

    (a) `qt(0.96,95)` which gives 1.7696

    (b) `qt(0.88,42)` which gives 1.1919

    (c) `qt(0.955,131)` which gives 1.7080

15. (a) `(1-pt(2.14,49))*2` which gives 0.0374. Alternatively, `(pt(-2.14,49))*2`

    (b) `1-pt(0.78,315)` which gives 0.2180.

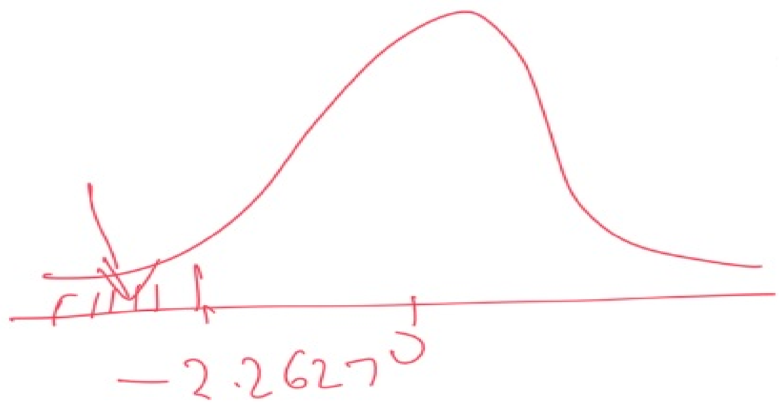
    (c) `pt(1.57,33)` which gives 0.9370.

16. (a) $H_0 : \mu = 160, H_a : \mu < 160$.

    (b)

$$\begin{aligned} t &= \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \\ &= \frac{158.4 - 160}{5/sqrt50} \\ &= -2.2627 \end{aligned}$$

    (c) The p-value is 0.0141, found by `pt(-2.2627,49)`. The sketch below shows how the p-value is found from a t distribution.



The critical value is found by `qt(0.95,49)` which gives 1.6766.

(d) Since our p-value is 0.0141 which is less than 0.05 (or since the magnitude of our test statistic is 2.2627 which is greater than the critical value of 1.6766), we reject the null hypothesis. Our data support the claim that the average yield this year is less than 160 bushels per acre.

(e) If the average yield this year is truly 160 bushels per acre, the probability of obtaining a sample of 50 farms with an average yield of 158.4 bushels per acre or less, is 0.0141.

General Questions

17. The critical value is `qt(0.99,49)` which gives 2.4049.

18. The t-multiplier is `qt(0.99,49)` which gives 2.4049.

19. They are the same values, numerically. Conclusions from a 2-sided hypothesis test conducted at significance level α will be consistent with conclusions from a $(1-\alpha)\times 100\%$ confidence interval.

20. Only (144.5, 163,5) is possible. At $\alpha = 0.02$, we fail to reject the null hypothesis. So the null value of 145 will be inside the corresponding 98% confidence interval.

21. (a)

$$\begin{aligned} \bar{x} &\pm t_{1-\alpha/2,df} \times \frac{s}{\sqrt{n}} \\ 20 &\pm 1.9886 \times \frac{3}{\sqrt{85}} \end{aligned}$$

which gives (19.3529, 20.6471).

(b) No, since 21 lies outside the confidence interval.

(c) The p-value will be less than the significance level 0.05, since the confidence level is at 95%. We will reject the null hypothesis since the null value of 21 lies outside the 95% confidence interval.

(d)

$$\begin{aligned} t &= \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \\ &= \frac{20 - 21}{3/\sqrt{85}} \\ &= -3.073181 \end{aligned}$$

The p-value is found using `2*pt(-3.073181, 84)` which gives 0.0029. The critical value is found using `qt(0.975,84)` which gives 1.98861.

Since the p-value is less than 0.05 (or since the magnitude of the test statistic is greater than the critical value), we reject the null hypothesis. Our data support the claim that the average length of banded archerfish is different from 21cm.