Class Exercise

1. **Teabag Biodiversity**

A recent study found that a single teabag can contain traces of up to 400 different kinds of insect DNA, reflecting the biodiversity of the fields where the tea leaves are grown. A food safety researcher believes that for certain popular brands, the true average number of insect DNA types per teabag is actually less than 400.

To investigate this claim, the researcher randomly samples 30 teabags from a major tea manufacturer. DNA barcoding of each teabag reveals an average of $\overline{x} = 390$ distinct insect DNA types, with a sample standard deviation of 28.

(a) At the 2.5% level of significance, is there sufficient evidence to support the claim that the average number of insect DNA types per teabag is lower than 400? State the hypotheses clearly, conduct the test, report the estimated P-value, and write a conclusion in the context of the problem.

Solution:

- **Hypotheses:** $H_0 : \mu = 400$ (The average number of insect DNA types is 400)
$H_1 : \mu < 400$ (The average number is less than 400)
- **Test Statistic:** Since the population standard deviation is unknown, we use a t-test.
$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{390 - 400}{28/\sqrt{30}} = \frac{-10}{5.112} \approx -1.956$$
- **P-value:** With degrees of freedom $df = n - 1 = 29$, the P-value for a left-tailed test is $P(t_{29} < -1.956)$. Using a t-table, we find that $-2.045 < -1.956 < -1.701$, which corresponds to a P-value between 0.025 and 0.05. Exact computation gives P-value ≈ 0.030.
- **Conclusion:** Since the P-value (0.030) is greater than $\alpha = 0.025$ (alternatively, our test statistic -1.956 is not less than the critical value -2.045), we fail to reject the null hypothesis. There is not sufficient evidence at the 2.5% level of significance to support the claim that the average number of insect DNA types per teabag is lower than 400.

(b) Construct and interpret an appropriate confidence bound for the true mean number of insect DNA types in teabags produced by this manufacturer. Explain how this bound can be used to support the conclusion obtained in part (a).

Solution: Since we conducted a left-tailed hypothesis test at $\alpha = 0.025$, the appropriate confidence bound is a 97.5% upper confidence bound for the mean.

$$\text{Upper Bound} = \overline{x} + t_{0.025,29}\left(\frac{s}{\sqrt{n}}\right)$$

$$\text{Upper Bound} = 390 + 2.045\left(\frac{28}{\sqrt{30}}\right) = 390 + 2.045(5.112) = 390 + 10.45 = 400.45$$

Interpretation: We are 97.5% confident that the true average number of insect DNA types per teabag is at most 400.45.
Support for (a): The upper confidence bound (400.45) includes values that are 400 and slightly above. Because the hypothesized mean of 400 is less than our upper bound, it falls within the plausible range of values for μ. Therefore, we cannot reject $H_0 : \mu = 400$ at the 2.5% significance level.

2. **Life Spans**

A random sample of 100 recorded deaths in the US during the past year showed an average life span of 71.8 years. Assuming a normal population and a population standard deviation of 8.9 years, does this seems to indicate that the mean life span today is greater than 70 years?

Hypotheses:
$H_0 : \mu = 70$
$H_1 : \mu > 70$
Test Statistic: Since σ is known, we use a z-test.

$$z = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{71.8 - 70}{8.9/\sqrt{100}} = \frac{1.8}{0.89} \approx 2.022$$

(a) Use 0.05 level of significance.

Solution: For a right-tailed test at $\alpha = 0.05$, the critical value is $z_{0.05} = 1.645$. Since our test statistic $z = 2.022 > 1.645$, we reject H_0. There is sufficient evidence to indicate that the mean life span is greater than 70 years.

(b) 0.01 level of significance.

Solution: For a right-tailed test at $\alpha = 0.01$, the critical value is $z_{0.01} = 2.326$. Since our test statistic $z = 2.022 < 2.326$, we fail to reject H_0. There is not sufficient evidence to indicate that the mean life span is greater than 70 years.

(c) What is the p-value of this test. What is your final conclusion?

Solution: The P-value is the area to the right of our test statistic:

$$P\text{-value} = P(Z > 2.022) = 1 - 0.9784 = 0.0216$$

Final Conclusion: A P-value of 0.0216 provides moderate evidence against the null hypothesis. It is small enough to reject H_0 at the 5% level, but not small enough to reject it at the more stringent 1% level. Therefore, the conclusion depends strictly on the chosen level of risk (α). Larger sample is needed to settle the question.

3. **Punctuation Cushioning**

Because ending an Internet comment with a punctuation mark can sometimes seem too abrupt or aggressive, many users have begun adopting "punctuation cushioning": inserting a space before a full stop, question mark, or exclamation mark to soften the tone. A linguist believes that more than 15% of online commenters now use punctuation cushioning regularly. A random sample of 250 Internet comments revealed that 48 used punctuation cushioning.

a. Does this sample provide evidence to support the linguist's claim? Use $\alpha = 0.05$ level of significance. Include a conclusion in the context of the problem and report the $P-$value.

Solution:

- **Hypotheses:** $H_0 : p = 0.15$
  $H_1 : p > 0.15$
- **Sample Proportion:** $\hat{p} = \frac{48}{250} = 0.192$
- **Test Statistic:**

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.192 - 0.15}{\sqrt{\frac{0.15(0.85)}{250}}} = \frac{0.042}{\sqrt{0.00051}} \approx 1.86$$

- **P-value:** $P\text{-value} = P(Z > 1.86) = 1 - 0.9686 = 0.0314$
- **Conclusion:** Since the P-value (0.0314) is less than $\alpha = 0.05$, we reject the null hypothesis. There is sufficient evidence at the 5% level of significance to support the linguist's claim that more than 15% of online commenters use punctuation cushioning regularly.

b. Construct and interpret an appropriate confidence bound for the true proportion of online comments that use punctuation cushioning. Explain how this confidence bound could be used to support the conclusion in part (a).

Solution: Since we conducted a right-tailed test at $\alpha = 0.05$, the appropriate bound is a 95% lower confidence bound.

$$\text{Lower Bound} = \hat{p} - z_{0.05}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$\text{Lower Bound} = 0.192 - 1.645\sqrt{\frac{0.192(0.808)}{250}} = 0.192 - 1.645(0.0249) = 0.192 - 0.0410 = 0.1510$$

Interpretation: We are 95% confident that the true proportion of online commenters who use punctuation cushioning is at least 15.10%.
Support for (a): The entire 95% confidence interval for the true proportion lies above 15% (starting at 15.10%). Because the hypothesized value of 15% is strictly below our lower bound, it is not a plausible value, reinforcing our decision to reject $H_0 : p = 0.15$ in favor of $p > 0.15$.

c. State the Type I and Type II errors in the context of this problem.

Solution:

- **Type I Error (α):** Concluding that more than 15% of commenters use punctuation cushioning when, in reality, the true proportion is 15% (or less).
- **Type II Error (β):** Failing to conclude that more than 15% of commenters use punctuation cushioning when, in reality, the true proportion actually is greater than 15%.