LAB Work 8

STAT451: Applied Statistics for Engineers and Scientists I

Le Dai King - 23125010 - 23TT2

Contents

1 Problem 33 - Page 292

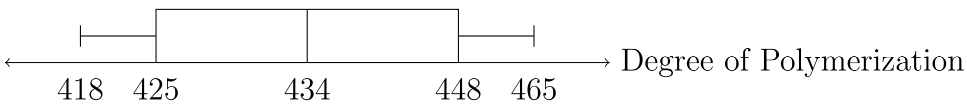
(a)

First, let's organize the data in ascending order:

$$418, 421, 421, 422, 425, 427, 431, 434, 437, 439, 446, 447, 448, 453, 454, 463, 465$$

Key statistics for the boxplot:

- Sample size $(n) = 17$
- Minimum $= 418$
- First quartile $(Q1) = 425$ (5th value in the ordered list)
- Median $(Q2) = 434$ (9th value)
- Third quartile $(Q3) = 448$ (13th value)
- Maximum $= 465$
- Interquartile Range (IQR) $= Q3 - Q1 = 448 - 425 = 23$



Interesting features:

- The data appears slightly left-skewed (longer tail on the left side).
- No outliers are present as all points fall within $1.5 \times IQR$ of the quartiles.
- The median is closer to $Q3$ than to $Q1$, indicating skewness.

(b)

To assess whether the sample comes from a normal distribution:

- **Visual inspection:** The boxplot shows slight left-skewness, which may deviate from normality.

- **Normal probability plot:** If constructed, the points should roughly follow a straight line. Given the skewness, some deviation is expected.

- **Statistical tests:** Shapiro-Wilk or Anderson-Darling tests could be used for formal testing, but with $n = 17$, the power to detect non-normality is limited.

Conclusion: While the sample size is small, the slight skewness suggests some deviation from normality. However, for many practical purposes, the distribution might be considered approximately normal.

(c)

Given the sample data:

- Sample mean ($\bar{x}$):

$$\bar{x} = \frac{418 + 421 + \cdots + 465}{17} = \frac{7392}{17} \approx 434.82$$

- Sample standard deviation (s):

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} \approx 15.14$$

- For a 95% confidence interval with $n - 1 = 16$ degrees of freedom, the t-critical value ($t_{0.025,16}$) ≈ 2.120.

The confidence interval is calculated as:

$$\bar{x} \pm t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}} = 434.82 \pm 2.120 \cdot \frac{15.14}{\sqrt{17}} \approx 434.82 \pm 7.78$$

$$\Rightarrow (427.04, 442.60)$$

2 Problem 35 - page 293

(a)

Given:

- Sample size (n) = 15
- Sample mean ($\bar{x}$) = 25.0%
- Sample standard deviation (s) = 3.5%

Assuming a normal distribution for failure strain, we construct a 95% confidence interval for the true average strain (μ):

$$\text{CI} = \bar{x} \pm t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}}$$

Where:

- $t_{0.025,14} \approx 2.145$ (from t-distribution table)
- Standard error = $\frac{3.5}{\sqrt{15}} \approx 0.9037$

Calculation:

$$25.0 \pm 2.145 \times 0.9037 \approx 25.0 \pm 1.938$$

$$\Rightarrow (23.062\%, 26.938\%)$$

Interpretation: We are 95% confident that the true average failure strain for the population lies between 23.062% and 26.938%.

(b)

For predicting the strain of a single adult, we use a prediction interval:

$$\text{PI} = \bar{x} \pm t_{\alpha/2,n-1} \cdot s \cdot \sqrt{1 + \frac{1}{n}}$$

$$25.0 \pm 2.145 \times 3.5 \times \sqrt{1 + \frac{1}{15}} \approx 25.0 \pm 2.145 \times 3.5 \times 1.032$$
$$\approx 25.0 \pm 7.744$$

$$\Rightarrow (17.256\%, 32.744\%)$$

Comparison:

- The prediction interval (17.256%, 32.744%) is much wider than the confidence interval (23.062%, 26.938%).
- This reflects the greater uncertainty when predicting a single observation compared to estimating the population mean.
- The prediction interval accounts for both the variability in estimating the mean and the natural variability between individuals.

Interval Type	**Range (%)**
95% Confidence Interval	(23.06, 26.94)
95% Prediction Interval	(17.26, 32.74)

3 Problem 37 - page 293

(a)

Given:

- Sample size (n) = 20
- Sample mean ($\bar{x}$) = 0.9255 strides/sec
- Sample standard deviation (s) = 0.0809 strides/sec
- Standard error of the mean (SEM) = 0.0181 strides/sec

Assuming normal distribution, we calculate the 95% confidence interval:

$$\text{CI} = \bar{x} \pm t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}}$$

Where:

- $t_{0.025,19} \approx 2.093$ (from t-distribution table)
- $\frac{s}{\sqrt{n}} = 0.0181$ (given as SEM)

Calculation:

$$0.9255 \pm 2.093 \times 0.0181 \approx 0.9255 \pm 0.0379$$

$$\Rightarrow (0.8876, 0.9634) \text{ strides/sec}$$

Interpretation: We are 95% confident that the true population mean cadence lies between 0.8876 and 0.9634 strides per second.

(b)

For predicting cadence of a single individual:

$$\text{PI} = \bar{x} \pm t_{\alpha/2,n-1} \cdot s \cdot \sqrt{1 + \frac{1}{n}}$$

Calculation:

$$0.9255 \pm 2.093 \times 0.0809 \times \sqrt{1.05} \approx 0.9255 \pm 2.093 \times 0.0809 \times 1.0247$$

$$\approx 0.9255 \pm 0.1734$$

$$\Rightarrow (0.7521, 1.0989) \text{ strides/sec}$$

Interpretation: We predict with 95% confidence that a randomly selected healthy male's cadence will fall between 0.7521 and 1.0989 strides per second.

(c)

To include at least 99% of the population values with 95% confidence:

$$\text{TI} = \bar{x} \pm k \cdot s$$

Where k is the tolerance factor for 99% coverage with 95% confidence:

- For $n = 20$, from tolerance factor tables: $k \approx 3.168$

Calculation:

$$0.9255 \pm 3.168 \times 0.0809 \approx 0.9255 \pm 0.2563$$

$$\Rightarrow (0.6692, 1.1818) \text{ strides/sec}$$

Interpretation: We are 95% confident that at least 99% of all healthy males have cadence values between 0.6692 and 1.1818 strides per second.

4 Problem 20 - page 321

Given Data

- Sample size (n) = 50 bulbs
- Sample mean ($\bar{x}$) = 738.44 hours
- Sample standard deviation (s) = 38.20 hours
- Standard error of the mean (SEM) = 5.40 hours
- Test statistic (Z) = -2.14
- p-value = 0.016

Hypothesis Test Setup

- Null hypothesis (H_0): $\mu = 750$ hours
- Alternative hypothesis (H_a): $\mu < 750$ hours (one-tailed test)
- Significance levels to consider: $\alpha = 0.05$ and $\alpha = 0.01$

Conclusions

1. For $\alpha = 0.05$:
   - p-value (0.016) < 0.05
   - **Reject** H_0
   - Conclude that there is sufficient evidence at the 5% significance level that the true average lifetime is less than 750 hours
2. For $\alpha = 0.01$:
   - p-value (0.016) > 0.01
   - **Fail to reject** H_0

  - Conclude that there is not sufficient evidence at the 1% significance level to claim that the true average lifetime is less than 750 hours

Recommended Significance Level and Conclusion

- The choice of significance level depends on the consequences of Type I and Type II errors:
  - Type I error (false positive): Rejecting H_0 when bulbs actually meet specifications
  - Type II error (false negative): Failing to reject H_0 when bulbs are inferior
- For this consumer decision:
  - A 5% significance level provides reasonable protection against accepting inferior bulbs
  - The 1% level might be too strict, potentially leading to accepting bulbs that don't meet specifications
- **Recommendation:**
  - Use $\alpha = 0.05$ significance level
  - Conclude that the bulbs have statistically significant shorter lifetime than advertised
  - Consider not proceeding with the purchase arrangement

5 Problem 22b - page 321

Given Data

- Sample size (n) = 30 pipes
- Sample mean ($\bar{x}$) = 206.73
- Sample standard deviation (s) = 6.35
- Standard error of the mean (SEM) = 1.16

- Production standard (μ_0) = 200

Hypothesis Test Setup

- Null hypothesis (H_0): $\mu = 200$ lb
- Alternative hypothesis (H_a): $\mu \neq 200$ lb (two-tailed test)
- Significance level: $\alpha = 0.05$ (typical industrial standard)

Test Procedure Given the large sample size ($n = 30$) and approximately normal data (from normal probability plot), we use a t-test:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{206.73 - 200}{6.35/\sqrt{30}} \approx \frac{6.73}{1.159} \approx 5.807$$

Degrees of freedom $= n - 1 = 29$

Critical Value Approach For $\alpha = 0.05$ and df = 29:

- Critical t-values: $\pm t_{0.025,29} \approx \pm 2.045$
- Calculated t-value (5.807) ¿ Critical value (2.045)
- Decision: Reject H_0

p-value Approach

- p-value < 0.0001 (for t = 5.807 with df = 29)
- Since p-value $< \alpha$ (0.05)
- Decision: Reject H_0

Conclusion There is strong evidence ($t_{29} = 5.807$, p < 0.0001) to conclude that the true average coating weight differs significantly from the production standard of 200.

The sample mean of 206.73 suggests that the process is applying *more* coating than specified by the production standards.

6 Problem 24 - page 321

Given Data The sample observations on stabilized viscosity (in Pa s):

$$2781,\ 2900,\ 3013,\ 2856,\ 2888$$

Sample Statistics

- Sample size $(n) = 5$
- Sample mean $(\bar{x})$:

$$\bar{x} = \frac{2781 + 2900 + 3013 + 2856 + 2888}{5} = \frac{14438}{5} = 2887.6\,\mathrm{Pa\,s}$$

- Sample standard deviation (s):

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{(2781 - 2887.6)^2 + \cdots + (2888 - 2887.6)^2}{4}} \approx 84.85\,\mathrm{Pa\,s}$$

- Standard error of the mean:

$$\mathrm{SEM} = \frac{s}{\sqrt{n}} = \frac{84.85}{\sqrt{5}} \approx 37.95\,\mathrm{Pa\,s}$$

Hypothesis Test

- Null hypothesis (H_0): $\mu = 3000\,\mathrm{Pa\,s}$
- Alternative hypothesis (H_a): $\mu \neq 3000\,\mathrm{Pa\,s}$ (two-tailed test)
- Significance level: $\alpha = 0.05$

t-Test Calculation

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{2887.6 - 3000}{84.85/\sqrt{5}} = \frac{-112.4}{37.95} \approx -2.962$$

Degrees of freedom $= n - 1 = 4$ **Critical Value Approach**

- Critical t-values: $\pm t_{0.025,4} = \pm 2.776$

- Calculated t-value (-2.962) ¡ -2.776
- Decision: Reject H_0

p-value Approach

- p-value for $t = -2.962$ with df = 4: 0.0208 (two-tailed)
- Since p-value (0.0208) ¡ α (0.05)
- Decision: Reject H_0

Conclusion There is statistically significant evidence ($t_4 = -2.962$, p = 0.0208) to conclude that the true average viscosity differs from the required $3000\,\mathrm{Pa\,s}$.

The sample mean of $2887.6\,\mathrm{Pa\,s}$ suggests that the average viscosity is *lower* than the required specification.