STAT 8020 R Lab 1: Simple Linear Regression

Whitney Huang, Clemson University

May 04, 2026

Contents

Leaning Tower of Pisa

The dataset `PisaTower.csv` provides annual measurements of the lean (the difference between where a point on the tower would be if the tower were straight and where it actually is) from 1975 to 1987. We would like to characterize lean over time by fitting a simple linear regression.

Load the dataset

Code:

```
PisaTower <- read.csv("PisaTower.csv")
head(PisaTower)
```

```
##     lean year
## 1 2.9642 1975
## 2 2.9644 1976
## 3 2.9656 1977
## 4 2.9667 1978
## 5 2.9673 1979
## 6 2.9688 1980
```

Descriptive analysis

Numerical summary

Compute numerical summaries for both variables (`lean` and `year`):

- Minimum, maximum, mean, median, and quartiles
- Standard deviation of `lean`
- Correlation between `lean` and `year`

Interpretation:

- Comment on the center and variability of lean
- Interpret the correlation in terms of direction and strength

Code:

Answer:

Graphical summary

Create the following plots:

- A boxplot of lean
- A scatterplot of lean vs. year

Interpretation: Describe the relationship between `lean` and `year` in terms of:

- Direction (positive/negative)
- Strength (weak/moderate/strong)
- Form (linear/nonlinear)

Code:

Answer:

Simple Linear Regression

1. Identify the response variable, the predictor variable, and the sample size.

Answer:

2. Fit a simple linear regression model and interpret the output.

Using `summary(lmfit)`, answer the following:

- What is the estimated slope ($\hat{\beta}_1$)?
- Interpret the slope in context (be specific about the units).
- Is the slope statistically significant?
  - Report the p-value.
  - State your conclusion in words (e.g., is there evidence of a linear relationship?).
- What does R^2 tell you about the model fit?
  - Interpret it in context.

Code:

3. Write down the fitted linear regression model.

Answer:

4. What is $\hat{\sigma}$, the estimate of σ? Briefly explain what this quantity represents.

Answer:

5. Construct a 95% confidence interval for β_1.

Code:

Answer:

6. Test the following hypothesis: $H_0 : \beta_1 = 0$ vs. $H_a : \beta_1 \neq 0$ with $\alpha = 0.05$

Answer:

7. Construct a 90% confidence interval for E[`lean`] in year 1984. What this interval represents (mean response vs. individual prediction)?

Code:

8. Use residual plots to check model assumptions. Specifically,

For each plot, comment on:

- Residuals vs Fitted: linearity and constant variance
- Normal Q-Q plot: normality
- Scale-Location: homoscedasticity
- Residuals vs Leverage: influential points

Do the regression assumptions appear reasonable?

Code:

Answer:

9. Would it be a good idea to use the fitted linear regression equation to predict `lean` in year 2010? Explain your answer.

Answer: