Homework 8

Han Nguyen - TXN200004

11/25/2025

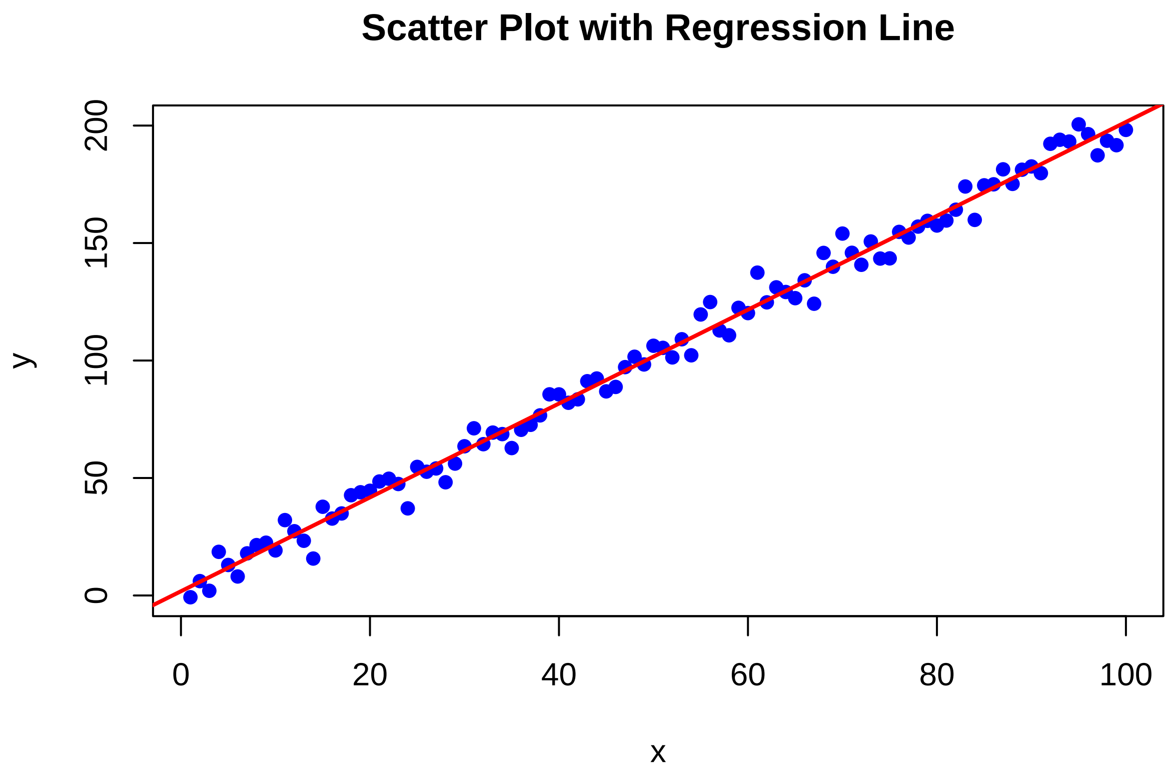
Problem 1

Simulate data from simple linear model

```r
# Set seed for reproducibility
set.seed(1)

# Create x values (arithmetic sequence from 1 to 100)
x <- 1:100

# Simulate errors from N(0, 6^2)
epsilon <- rnorm(100, mean = 0, sd = 6)

# Generate y values: Y = 1 + 2x + epsilon
y <- 1 + 2*x + epsilon
```

Scatter plot with regression line

```r
# Fit linear regression model
model <- lm(y ~ x)

# Create scatter plot with regression line
plot(x, y,
     main = "Scatter Plot with Regression Line",
     xlab = "x",
     ylab = "y",
     pch = 16,
     col = "blue")
abline(model, col = "red", lwd = 2)
```



Hypothesis Test (7-Step Procedure)

Step 1: State the null and alternative hypotheses

- $H_0 : \beta_1 = 2$
- $H_a : \beta_1 \neq 2$

Step 2: Choose the significance level

$\alpha = 0.05$

Step 3: Determine the appropriate test statistic

```
# Get model summary
summary_model <- summary(model)
beta1_hat <- coef(model)[2]
se_beta1 <- summary_model$coefficients[2, 2]
n <- length(y)
df <- n - 2
```

The appropriate test statistic for testing the slope coefficient is:

$$t = \frac{\hat{\beta}_1 - \beta_{1,0}}{SE(\hat{\beta}_1)}$$

which follows a t-distribution with $n - 2 = 98$ degrees of freedom under H_0.

Step 4: Formulate the decision rule

```r
# Calculate critical value
t_critical <- qt(0.975, df)
```

Decision rule: Reject H_0 if $|t| > t_{\alpha/2,n-2} = 1.9845$ or if p-value $< \alpha = 0.05$.

Step 5: Collect data and calculate the test statistic

```r
# Calculate test statistic
t_stat <- (beta1_hat - 2) / se_beta1
```

Using our simulated sample data:

- $\hat{\beta}_1 = 1.9973$
- $SE(\hat{\beta}_1) = 0.0188$
- $t = \frac{1.9973-2}{0.0188} = -0.1442$

Step 6: Make a decision

```r
# Calculate p-value
p_value <- 2 * pt(abs(t_stat), df, lower.tail = FALSE)
decision <- ifelse(abs(t_stat) > t_critical, "reject", "fail to reject")
```

Since $|t| = 0.1442$ is less than the critical value 1.9845, and the p-value is 0.8856 which is greater than $\alpha = 0.05$, we **fail to reject** H_0.

Step 7: Draw a conclusion

We fail to reject H_0. At the 5% significance level, there is insufficient evidence to conclude that the true slope β_1 is different from 2.

Problem 2

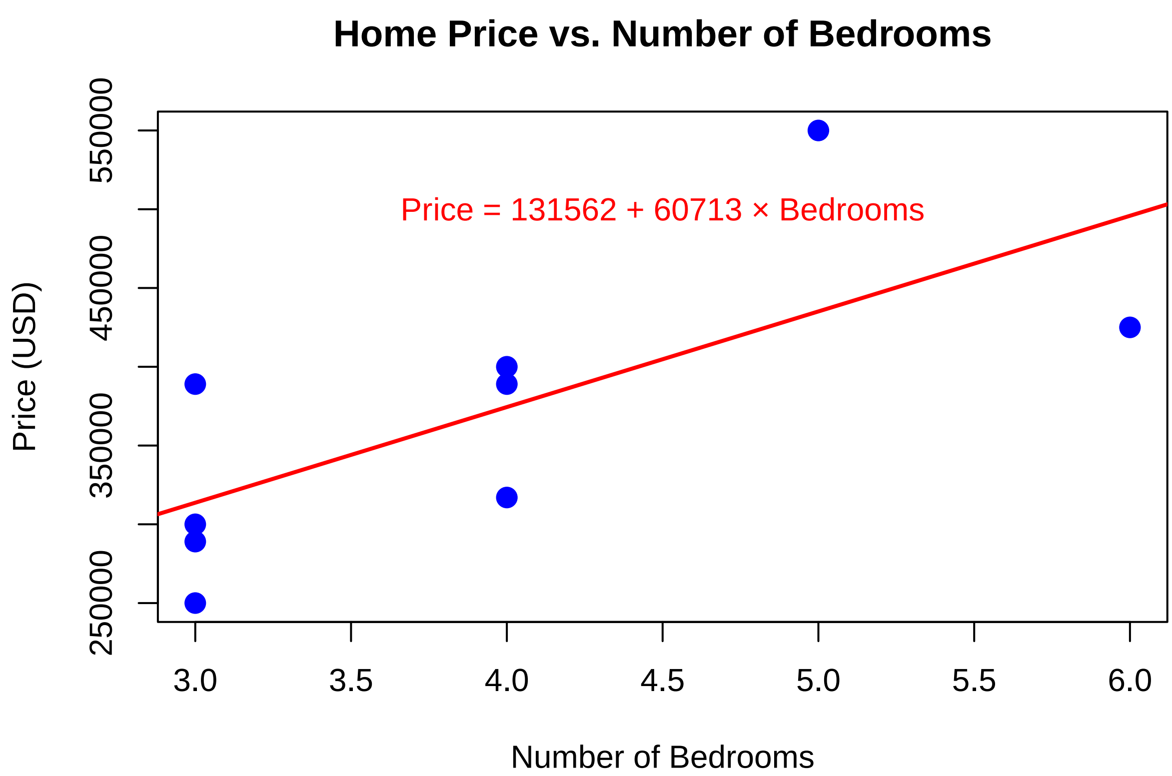
Create the Dallas homes dataset

```r
# Create the dataset
price <- c(300000, 250000, 400000, 550000, 317000, 389000, 425000, 289000, 389000)
bedrooms <- c(3, 3, 4, 5, 4, 3, 6, 3, 4)
```

Scatter plot with regression line

```r
# Fit linear regression model
model2 <- lm(price ~ bedrooms)

# Create scatter plot with regression line
plot(bedrooms, price,
     main = "Home Price vs. Number of Bedrooms",
     xlab = "Number of Bedrooms",
     ylab = "Price (USD)",
     pch = 16,
     col = "blue",
     cex = 1.5)
abline(model2, col = "red", lwd = 2)

# Display regression equation
coeffs <- coef(model2)
text(4.5, 500000,
```

```r
  paste0("Price = ", round(coeffs[1], 0), " + ", round(coeffs[2], 0), " × Bedrooms"),
  col = "red")
```



Confidence intervals for mean price

```r
# Create new data for bedrooms 2 to 8
new_data <- data.frame(bedrooms = 2:8)

# Compute confidence intervals for mean price
predictions <- predict(model2, newdata = new_data, interval = "confidence", level = 0.95)

# Combine with bedroom numbers for display
results <- cbind(new_data, predictions)
colnames(results) <- c("Bedrooms", "Predicted Price", "Lower 95% CI", "Upper 95% CI")

# Display results
knitr::kable(results, format = "markdown", digits = 0)
```

Bedrooms	Predicted Price	Lower 95% CI	Upper 95% CI
2	252987	136928	369047
3	313700	241198	386202
4	374412	320036	428789
5	435125	354065	516185
6	495838	368959	622716
7	556550	378957	734143

Bedrooms	Predicted Price	Lower 95% CI	Upper 95% CI
8	617262	387276	847249

The table above shows the predicted mean price and 95% confidence intervals for homes with 2 to 8 bedrooms in Dallas.

Problem 3

Load data and fit regression model

```r
# Load the UsingR package
library(UsingR)
```

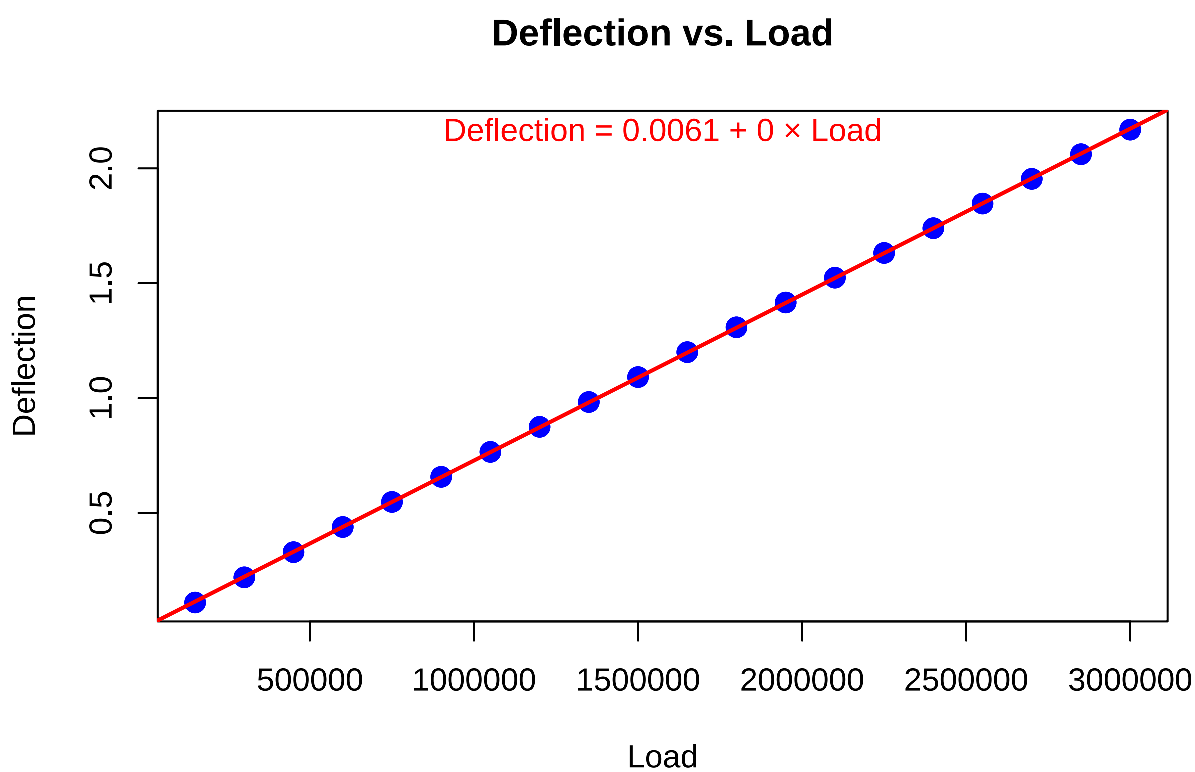
```
## Loading required package: MASS

## Loading required package: HistData

## Loading required package: Hmisc

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```r
# Load the deflection dataset
data(deflection)

# Fit linear regression model: Deflection ~ Load
model3 <- lm(Deflection ~ Load, data = deflection)

# Display model summary
summary(model3)
```

```
##
## Call:
## lm(formula = Deflection ~ Load, data = deflection)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0042751 -0.0016308  0.0005818  0.0018932  0.0024211
##
## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept) 6.150e-03  7.132e-04    8.623 1.77e-10 ***
## Load        7.221e-07  3.969e-10 1819.289  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.002171 on 38 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:      1
## F-statistic: 3.31e+06 on 1 and 38 DF,  p-value: < 2.2e-16
```

Scatter plot with regression line

```r
# Create scatter plot with regression line
plot(deflection$Load, deflection$Deflection,
     main = "Deflection vs. Load",
     xlab = "Load",
     ylab = "Deflection",
     pch = 16,
     col = "blue",
     cex = 1.5)
abline(model3, col = "red", lwd = 2)

# Add regression equation to plot
coeffs3 <- coef(model3)
text(x = mean(deflection$Load),
     y = max(deflection$Deflection),
     paste0("Deflection = ", round(coeffs3[1], 4), " + ", round(coeffs3[2], 4), " × Load"),
     col = "red")
```



95% Confidence Intervals for β_0 and β_1

```r
# Compute 95% confidence intervals for regression coefficients
ci <- confint(model3, level = 0.95)

# Display confidence intervals
ci
```

```
##                    2.5 %       97.5 %
## (Intercept) 4.705876e-03 7.593493e-03
## Load        7.212991e-07 7.229061e-07
```

The 95% confidence intervals are:

- β_0 **(Intercept)**: [0.004706, 0.007593]
- β_1 **(Slope)**: [7.213e-07, 7.229e-07]

This means we are 95% confident that the true intercept β_0 lies between 0.004706 and 0.007593, and the true slope β_1 lies between 7.213e-07 and 7.229e-07. The slope is very small, indicating that each unit increase in Load corresponds to a very small increase in Deflection.