**STA3001 Linear Models**

Homework 1

**Due on Oct. 9, 2023**

- You should work out this Homework individually. Group works or discussions are not acceptable.
- No late Homework will be accepted.

**Exercise 1** Determine the 95 and 99th percentiles of

a. The normal distribution with mean 10 and standard deviation 3;

b. The $t$ distributions with 10 and 25 degrees of freedom;

c. The chi-square distributions with 1,4, and 10 degrees of freedom;

d. The $F$ distributions with 2 and 10, and 4 and 10 degrees of freedom.

**Exercise 2** It is a fact that two distributions are the same if (all) their percentiles are identical.

a. Convince yourself, by looking up several percentiles, that the square of a standard normal distribution is the same as a chi-square distribution with one degree of freedom. Determine the percentile of the $\chi_1^2$ and the percentile of the square of a standard normal distribution, $Z^2$, and show that they are the same. Use the fact that $P(Z^2 \leq z) = P(-\sqrt{z} \leq Z \leq \sqrt{z})$. Hence, for example, the 95th percentile of $Z^2$ is the same as the 97.5th percentile of $Z$.

b. Convince yourself, by looking up several percentiles, that the square of a $t$ distribution with $\nu$ degrees of freedom is the same as the $F(1, \nu)$ distribution.

**Exercise 3** A car dealer is interested in modeling the relationship between the weekly number of cars sold and the daily average number of salespeople who work on the showroom floor during that week. The dealer believes that the relationship between the two variables can be described by a straight line. The following data were supplied by the car dealer:

| Week of | No. of Cars Sold $y$ | Average.No. of Sales People on Duty $x$ |
|---|---|---|
| January 30 | 20 | 6 |
| June 29 | 18 | 6 |
| March 2 | 10 | 4 |
| October 26 | 6 | 2 |
| February 7 | 11 | 3 |

a. Construct a scatter plot ($y$ vs $x$) for the data.

b. Assuming that the relationship between the variables is described by straight line, use the method of least squares to estimate the $y$ intercept and the slope of the line.

c. Plot the least squares line on your scatter plot.

d. According to your least squares line, approximately how many cars should the dealer expect to sell in a week if an average of five salespeople are kept on the showroom floor each day?

e. Calculate the fitted value $\hat{\mu}$ for each observed $x$ value. Use the fitted values to calculate the corresponding residuals. Plot the residuals against the fitted values. Are you satisfied with the fit?

f. Calculate an estimate of $\sigma^2$.

g. Construct a 95% confidence interval for $\beta_1$ and use it to access the hypothesis that $\beta_1 = 0$.

h. Given the results of (a) - (g), what conclusions are you prepared to draw about the relationship between sales and number of salespeople on duty.

i. Would you be willing to use this model to help determine the number of salespeople to have on duty next year?

**Exercise 4** Use R for Exercise 2.4. Check your hand calculations with the results from these programs.

**Exercise 5** Grade point averages of 12 graduating MBA students, GPA, and their GMAT scores taken before entering the MBA program are given below. Use the GMAT scores as a predictor of GPA, and conduct a regression of GPA on GMAT scores.

| $x$ = GMAT | $y$ = GPA |
|---|---|
| 560 | 3.20 |
| 540 | 3.44 |
| 520 | 3.70 |
| 580 | 3.10 |
| 520 | 3.00 |
| 620 | 4.00 |
| 660 | 3.38 |
| 630 | 3.83 |
| 550 | 2.67 |
| 550 | 2.75 |
| 600 | 2.33 |
| 537 | 3.75 |

a. Obtain and interpret the coefficient of determination $R^2$.

b. Calculate the fitted value for the second person.

c. Test whether GMAT is an important predictor variable (use significance level 0.05).

**Exercise 6** Occasionally, a model is considered in which the intercept is known to be zero a priori. Such a model is given by

$$y_i = \beta_1 x_i + \epsilon_i, i = 1, 2, \ldots, n$$

where the errors $\epsilon_i$ follow the usual assumptions.

a. Obtain the LSEs $(\hat{\beta}_1, s^2)$ of $(\beta_1, \sigma^2)$.

b. Define $e_i = y_i - \hat{\beta}_1 x_i$. Is it still true that $\sum_{i=1}^n e_i = 0$? Why or why not?

c. Show that $V(\hat{\beta}_1) = \sigma^2 / \sum_{i=1}^n x_i^2$.

**Exercise 7** An investigation involving five factors has singled out temperature as having the greatest impact on the accelerated lifetime of a special type of heater. On the advice of the progress engineer, temperatures 1,520, 1,620, 1,660, and 1,702°$F$ were chosen.

Twenty-four heaters were selected at random from the current production and split randomly among the four temperatures. The life times of these heaters are given below.

| Temperature $T$ | Lifetime $y$ (Hours) | | | | | |
|---|---|---|---|---|---|---|
| 1,520 | 1,953 | 2,135 | 2,471 | 4,727 | 6,143 | 6,314 |
| 1,620 | 1,190 | 1,286 | 1,550 | 2,125 | 2,557 | 2,845 |
| 1,660 | 651 | 837 | 848 | 1,038 | 1,361 | 1,543 |
| 1,708 | 511 | 651 | 651 | 652 | 688 | 729 |

a. Plot the data and summarize the important features of the relationship.

b. Transform the $y$'s to LY = ln $y$ and replot the data. Comment on the functional relationship.

c. Fit the model

$$\text{LY} = \beta_0 + \beta_1 T + \epsilon$$

   i. Assess the fit by adding the fitted line to the scatter plot.

   ii. If you are not satisfied with the fit, state why. What other approach might you take to get a better fitting model?