

STA201 Assignment 3

Problem 1

The owner of a second-hand car dealership wants to study the relationship between the age of a car and its selling price. Listed below is a random sample of 12 used cars sold at the dealership during the last year.

Age (years)	9	7	11	12	8	7	8	11	10	12	6	6
Price (thousand \$)	8.1	6	3.6	4	5	10	7.6	8	8	6	8.6	8

- Draw a scatter diagram and comment the relation between the age of the car and its selling price.
- Determine the Pearson correlation coefficient and the coefficient of determination and interpret it.

Problem 2

At a business case competition, two judges were tasked with scoring (out of 1000) the case reports submitted by 12 teams. Using the following data, we would like to examine the association between the scores of the two judges.

Judge 1	650	760	740	700	590	620	700	690	900	500	610	700
Judge 2	900	720	690	850	920	800	890	920	1000	690	700	760

- Compute the Spearman rank correlation between the scores of the two judges.
- Comment on the association between the scores of the two judges.

Problem 3

An electric company is studying the relationship between the energy consumption (in thousand kilowatt-hours) and the number of rooms in a private single-family residence. A random sample of 10 homes yielded the following.

Number of Rooms	12	9	14	6	10	8	10	10	5	7
Energy Consumption (thousand kWh)	9	7	10	5	8	6	8	10	4	7

- Determine the regression equation of energy consumption on the number of rooms.
- Interpret the model.
- What is the predicted energy consumption, in thousand kWh, for a six-room house.
- Comment on the goodness of fit of the model.

Problem 4

Designers of backpacks use exotic material to make packs that fit comfortably and distribute weight to eliminate pressure points. For fitting a regression model of price of backpack on the capacity (cubic inches) and comfort rating of backpacks, a data for 10 backpacks are used. Comfort was measured using a rating from 1 to 5, with a rating of 1 denoting average comfort and a rating of 5 denoting excellent comfort. The output of the regression model is as follows on the next page:

```

Residuals:
    Min       1Q   Median       3Q      Max
-84.12 -27.18  10.61  36.90  48.26

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  356.12083   197.17401    1.806 0.113859
x1           -0.09874    0.04588   -2.152 0.068372 .
x2           122.86721    21.79975    5.636 0.000786 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 51.14 on 7 degrees of freedom
Multiple R-squared:  0.8318,    Adjusted R-squared:  0.7838
F-statistic: 17.31 on 2 and 7 DF,  p-value: 0.00195

```

- Determine the estimated regression equation that can be used to predict the price of a backpack given the capacity and the comfort rating.
- Interpret the model.
- Predict the price for a backpack with a capacity of 4500 cubic inches and a comfort rating of 4.
- Comment on goodness of fit of the model.

Problem 5

Suppose that we are working with some doctors on heart attack patients. The dependent variable is whether the patient has had a second heart attack within 1 year (yes=1). We have two independent variables, one is age of the patient and the other is a score on anxiety scale (a higher score means more anxious). After applying logistic regression model, we have the following output:

```

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.064    0.000    0.000    0.000    1.446

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -471.441  223186.509   -0.002    0.998
Age           6.394   3057.349    0.002    0.998
Anxiety       1.347    611.470    0.002    0.998

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 27.7259  on 19  degrees of freedom
Residual deviance:  3.7087  on 17  degrees of freedom
AIC: 9.7087

Number of Fisher Scoring iterations: 23

```

- Determine the estimated logistic regression equation.
- Calculate the odds ratio and interpret.