

MAS 250 Homework Assignment 8

Due: December 9 (Friday) 1 pm

Instruction: Turn in homework as a **single pdf file**.

1. We have the following variables:

- Name: University name
- SAT: average SAT score of new freshmen
- Top10: percentage of new freshmen in top 10
- Accept: percentage of applicants accepted
- Sfratio: Student-faculty ratio,
- Expenses: Estimated annual expenses (divided by 1000)
- Grad: Graduation rate.

Read **univ.txt**. Exclude Name from the analysis.

- (a) Interpret a scatter diagram matrix.
- (b) Run multiple regression (response: SAT, predictors: Top10, Accept, Sfratio, Expenses, Grad), Report R^2 , MSE, and interpret the overall F test and individual t tests.
- (c) Conduct residual analysis (Standardized residuals vs. Fitted, Histogram and Q-Q plot of residuals) with the model and assess the assumptions.

2. Four chemical plants, producing the same products and owned by the same company, discharge effluents into streams in the vicinity of their locations. To monitor the extent of pollution created by the effluent and to determine whether this differs from plant to plant, the company collected random samples of liquid waste, five specimens from each plant.

Group	Sample Size	Sample Mean	Sample standard deviation
A	5	1.568	0.1366
B	5	1.772	0.2160
C	5	1.546	0.1592
D	5	1.916	0.1689

- (a) (2 points) Write the statistical model and identify factor and response. State the null hypothesis.
- (b) (4 points) Complete the following ANOVA table. Specify the null hypothesis of the test and conduct an analysis of variance at $\alpha = 0.05$.

Source	D.F.	Sum of Squares	Mean of Squares	F	p -value
Model					0.0107
Error					
Total		0.94			

- (c) (2 points) Interpret the multiple testing results (at the overall significance level $\alpha = 0.05$) by Tukey.

Tukey multiple comparisons of means
95% family-wise confidence level

Fit: aov(formula = effluent ~ plant, data = anovaprob)

```
$plant
diff      lwr      upr      p adj
B-A  0.204 -0.10836258 0.51636258 0.2796961
C-A -0.022 -0.33436258 0.29036258 0.9969809
D-A  0.348  0.03563742 0.66036258 0.0264399
C-B -0.226 -0.53836258 0.08636258 0.2047920
D-B  0.144 -0.16836258 0.45636258 0.5647642
D-C  0.370  0.05763742 0.68236258 0.0176858
```

3. (**by R**) Suppose you want to determine whether the brand of laundry detergent used and the temperature affects the amount of dirt removed from your laundry. To this end, you buy two different brands of detergent (“Super” and “Best”) and choose three different temperature levels (“cold”, “warm”, and “hot”). With each combination 4 loads were washed and the dirt removed were recorded.
- Write a model with an interaction.
 - Draw a side-by-side boxplot and make comments.
 - Is there a difference in the mean of two detergents? Use $\alpha = 0.05$.
 - Is there a difference in the mean of three temperature levels? Use $\alpha = 0.05$.
 - Does there exist an interaction between detergent and temperature? Use $\alpha = 0.05$.
4. From the exercise problems in Chapter 10:
5 (R), 9, 13 (R)
5. Suggested problems (no submission):
1, 8, 14, 20, 22, 24