USC Viterbi
School of Engineering
Daniel J. Epstein
Department of Industrial
and Systems Engineering

ISE529 Predictive Analytics
Instructor: Dr. Tao Ma
2025 Spring

Homework 2
Due by: **Feb. 20**, 2025, 11:59 PM

Instructions:

1. Print your First and Last name and NetID on your answer sheets
2. Submit all your answers including Python scripts and report in a single Jupyter Lab file (.ipynb) or along with a single PDF to Brightspace by due date. No other file formats will be graded. No late submission will be accepted.
3. Total 5 problems. Total points: 100

---

1. (20 points)
Consider the following computer output.

The regression equation is Y = 254 + 2.77 x1 – 3.58 x2

Predictor	Coef	SE Coef	T
Constant	253.810	4.781	?
x1	2.7738	0.1846	15.02
x2	–3.5753	0.1526	?

S = 5.05756 R-Sq = ? R-Sq (adj) = 98.4%

Analysis of Variance

Source	DF	SS	MS	F
Regression	2	22784	11392	?
Residual error	?	?	?	
Total	14	23091		

(a) Fill in the missing quantities.
(b) What conclusions can you draw about the significance of regression?
(c) What conclusions can you draw about the contributions of the individual regressors to the model?

Note: check the critical value in the F-distribution or t-distribution table.

2. (20 points)

A study was performed on wear of a bearing and its relationship to x_1 = oil viscosity and x_2 = load. The data can be found in attached file *bearingdata.csv*.

(a) Fit a multiple linear regression model in the form of $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$. Write out the estimated model.

(b) Estimate σ^2 and compute the t-statistics for each regression coefficient. Using $\alpha = 0.05$, what conclusions can you draw?

(c) Test for significance of overall regression using $\alpha = 0.05$. What is the P-value for this test? What are your conclusions?

(d) Use the model to predict wear when $x_1 = 25$ and $x_2 = 1000$.

(e) Use the extra sum of squares method to investigate the usefulness of adding x_2 = load to a model that already contains x_1 = oil viscosity. Use $\alpha = 0.05$.

(f) Refit the model with an interaction term. Test for significance of regression using $\alpha = 0.05$.

(g) Use the extra sum of squares method to determine whether the interaction term contributes significantly to the model. Use $\alpha = 0.05$.

3. (20 points)

We have used a sample of 30 observations to fit a regression model. The full model has 9 regressors, the variance estimate is $\hat{\sigma}^2 = MS_E = 100$, and $R^2 = 0.92$.

(a) Calculate the F-statistic for testing significance of regression. Using $\alpha = 0.05$, what would you conclude?

(b) Suppose that we fit another model using only four of the original regressors and that the error sum of squares for this new model is 2200. Find the estimate of σ^2 for this new reduced model. Would you conclude that the reduced model is superior to the old one? Why?

(c) Find the value of C_p for the reduced model in part (b). Would you conclude that the reduced model is better than the old model?

4. (20 points)

Use the Carseats data set (attached Carseats.csv) to answer the following questions.

(a) Fit a multiple regression model to predict Sales using Price, Urban, and US.

(b) Provide an interpretation of each coefficient in the model.

(c) Write out the model in equation form, show the qualitative variables properly.

(d) For which of the predictors can you reject the null hypothesis $H_0\colon \beta_j = 0$?

(e) On the basis of your response to the previous question, fit a smaller model that only uses the predictors which is statistically significant. Compare it to the model in (a), which one is a better model?

(f) Using the model from (e), obtain 95 % confidence intervals for the coefficient(s).

5. (20 points)
Perform the following Python code to generate simulated data, and answer the following questions:

```
rng = np.random.default_rng (10)
x1 = rng.uniform(0, 1, size=100)
x2 = 0.5 * x1 + rng.normal(size=100) / 10
y = 2 + 2 * x1 + 0.3 * x2 + rng.normal(size=100)
```

(a) Write out the form of the underlying true linear model. What are the regression coefficients?
(b) Use function *corr*() to calculate the correlation between x_1 and x_2? Create a scatterplot matrix displaying the relationship between the variables.
(c) Using this data, fit a linear regression model to predict y using x_1 and x_2. Describe the results obtained. Can you reject the null hypothesis H_0: $\beta_1 = 0$ and/or null hypothesis H_0: $\beta_2 = 0$?
(d) Now fit a least squares regression to predict y using only x_1 or using only x_2 respectively. Comment on your results. Can you reject the null hypothesis H_0: $\beta_1 = 0$? It is observed that x_1 and x_2 cannot be simultaneously significant in the model in (c). What is this implied?