Multiple Regression (Review)

1. We have a dataset measuring the price (\$), size (ft^2), number of bedrooms, and age (years) of 518 houses in Easton, Pennsylvania. We fit a regression model to explain price in terms of the other variables.

```
The regression equation is
PRICE = 25875 + 39.2 SIZE - 1145 BEDROOM - 354 AGE

Predictor    Coef  SE Coef      T      P
Constant    25875     3555   7.28  0.000
SIZE       39.196    2.138  18.34  0.000
BEDROOM     -1145     1153  -0.99  0.321
AGE        -353.8    266.9  -1.33  0.186

S = 12612.2   R-Sq = 51.0%   R-Sq(adj) = 50.7%


Analysis of Variance

Source           DF           SS           MS       F      P
Regression        3  85029785549  28343261850  178.18  0.000
Residual Error  514  81760176401    159066491
Total           517  1.66790E+11
```

   (a) Interpret the estimated coefficient of Bedroom in the context of the fitted regression model.

   (b) What does the result of the t test on the coefficient of Size indicate?

   (c) What does the result of the t test on the coefficient of Bedroom indicate?

   (d) What does the result of the F test indicate?

Multiple Regression with Qualitative Predictors

2. We asked 46 NYU students how much time they spend on social media, and what their primary computer is (Mac or PC). We are going to use regression to find out if one type of computer associated is with more social media usage. We have the response variable

   $$\text{Social} = \text{amount of time (in minutes per week) using social media}$$

   We would like to use "OS" as a predictor variable, which is a categorical (qualitative) variable taking values in the set $\{\text{Mac}, \text{PC}\}$.

   (a) Why does the model $\text{Social} = \beta_0 + \beta_1 \text{OS} + \varepsilon$ not make sense?

   (b) Give two different models to explain Social in terms of OS.

   (c) Consider the model from part (b) involving the dummy variable "PC". What is the interpretation of β_0?

   (d) Again, consider the model from part (b) involving the dummy variable "PC". What is the interpretation of β_1?

3. Using the data from problem 2, we fit the regression model in Minitab, and got the following output.

```
The regression equation is
Social = 295 - 132 PC

Predictor     Coef  SE Coef      T      P
Constant    295.20    57.09   5.17  0.000
PC         -132.34    84.49  -1.57  0.124

S = 285.436   R-Sq = 5.3%   R-Sq(adj) = 3.1%
```

   (a) What is the estimated mean social usage for Mac users?

   (b) What is the estimated mean social usage for PC users?

   (c) What is the interpretation of the p-value for the test on the coefficient of PC?

4. We use the same data as in the previous problem, but now we are interested in whether or not texting behavior differs by cell phone type (Blackberry, iPhone, other smart phone, or standard cell phone).

   (a) Introduce dummy variables to encode cell phone type.

   (b) Using the variables you defined in part (a), devise a regression model which explains text usage in terms of cell phone type.

   (c) What is the interpretation of β_0, the intercept?

   (d) What are the interpretations of the other coefficients in your model?

5. We fit a model that explains Text in terms of cell phone type using dummy variables for cell phone type.

```
The regression equation is
Text = 132 + 91 Blackberry + 349 iPhone + 68 Smartphone

Predictor    Coef  SE Coef     T      P
Constant    131.7    316.9  0.42  0.680
Blackberry   90.8    501.0  0.18  0.857
iPhone      349.0    354.2  0.99  0.330
Smartphone   68.3    388.1  0.18  0.861

S = 776.121   R-Sq = 3.9%   R-Sq(adj) = 0.0%

Analysis of Variance

Source          DF        SS      MS     F      P
Regression       3   1025437  341812  0.57  0.640
Residual Error  42  25299274  602364
Total           45  26324711
```

(a) What is the estimated mean Text usage for people without smart phones?

(b) What is the estimated mean Text usage for people with iPhones?

(c) Is there statistically significant evidence that people with iPhones exhibit different texting behavior (volume) than people without smart phones?

(d) Is cell phone type useful for predicting Text?