**Project 2 – Part 2**

1. Write a macro to invoke PROC GLMSELECT five times on the **SalePrice** variable regressing on the interval variables. For each, request STEPWISE selection with the SELECTION= option and include DETAILS=STEPS to obtain step information and the selection summary table. Use 0.05 as the significance level for entry into and staying in the model. Call to macro to run SELECT for the options SL, AIC, BIC, AICC, and SBC and compare the selected models from the output. Does the significance level for entry into and staying in the model have any impact when you use options other than SL? Which variables stay in the model for each 5 options? Which selection methods and criteria would you recommend?
2. Invoke PROC REG with the plots option using rsquare adjrsq cp to produce a regression of **SalePrice** on all the other interval variables in the data set. Use selection = rsquare adjrsq cp. Which model you would suggest, and why? Hint: compare rsquare adjrsq cp.
3. Invoke PROC FREQ and create one‑way frequency tables for the variables **Bonus**, **Fireplaces**, and **Lot\_Shape\_2** and create two‑way frequency tables for the variables **Bonus** by **Fireplaces**, and **Bonus** by **Lot\_Shape\_2**. For the continuous variable, **Basement\_Area**, create histograms for each level of **Bonus**. Use a CLASS statement in PROC UNIVARIATE. Use the FORMAT procedure to format the values of **Bonus**.

Are there any unusual data values that could be due to coding errors for any of the categorical variables?

Examine the distribution of **Bonus** at each value of the predictors. What associations do you see?

1. Use the FREQ procedure to test for an association between the variables **Lot\_Shape\_2**   
   and **Bonus** as well as **Fireplaces** and **Bonus**. Generate the expected cell frequencies   
   and the cell’s contribution to the total chi‑square statistic.

Is there is evidence of an association between **Lot\_Shape\_2** and **Bonus**?

Does the Odds Ratio and Relative Risk table show a measure of strength of association? Explain.

What’s the Odds Ratio? How would you interpret that result?

1. Use PROC FREQ to test whether an ordinal association exists between **Bonus** and **Fireplaces**.
   1. DOES **Bonus** and **Fireplaces** have a significant ordinal association (use Mantel‑Haenszel chi‑square test)?
   2. For the Spearman correlation statistic, is the relationship is significant at the 0.05 significance level?
2. Fit a binary logistic regression model in PROC LOGISTIC. Select **Bonus** as the outcome variable and **VARIABLE assigned to your team** as the predictor variable. Use the EVENT= option to model the probability of being bonus eligible and request profile likelihood confidence intervals around the estimated odds ratios. Use the ALPHA=.10 option in the MODEL statement
3. The model should be based on the probability of being bonus eligible (**Bonus**=1)
4. The Testing Global Null Hypothesis: BETA=0 table. What does BETA=0 mean?
5. Is there any evidence for at least one of the regression coefficients for an explanatory variable is significantly different from 0? Explain.
6. What’s the logistic regression equation?
7. Is the *p*‑value for the **VARIABLE** significant at the 0.10 significance level?
8. What does the odds ratio for the **VARIABLE** indicate?
9. Interpret the value of the *c* (concordance) statistic.