

PHC 6091 HW#3

1. HW3Q01.sas7bdat gives the systolic blood pressure (SBP), body size (QUET), dichotomized age (AGE_BINARY =1 if age \geq 60; AGE_BINARY =0 if Age <60), and smoking history (SMK = 0 if a nonsmoker, SMK = 1 if a current or previous smoker) for a hypothetical sample of 32 white males over 40 years old from the town of Angina.
 - a. State the appropriate regression model to use for comparing the mean blood pressure between the two smoking status (smokers versus non-smokers), controlling both AGE_BINARY and QUET (hint: check the 1st order and 2nd order interaction terms and determine which interaction term(s) should be included in the final regression model)
 - b. Based on the model you obtained in the previous section, determine the adjusted SBP means for smokers and nonsmokers.

Hint: First, find the overall sample mean of QUET and the mode of AGE_BINARY. You can use SAS Proc Freq to get the mode of AGE_BINARY. Remember mode is defined as the level of AGE_BINARY with the highest frequency.

```
Proc Freq data=HW3Q01;
```

```
table AGE_BINARY;
```

```
run;
```

Then, report the adjusted SBP means for smokers and non-smokers at the mean value of QUET and mode of AGE.

- c. Test whether the true adjusted mean SBP- QUET relationship is the parallel between the two age groups within each smoking status.
2. HW3Q02.sas7bdat presents the starting annual salaries (SAL) of a group of 30 college graduates who have recently entered the job market, along with their cumulative grade-point averages (CGPA). For the simple linear regression model of SAL (Y) on CGPA (X),
 - a. Examine a plot of the jackknife residuals versus the predicted values. Are any regression assumption violations apparent? If so, suggest possible remedies.
 - b. Examine whether the normality assumption violated. If violated, suggest possible remedies.
 - c. Examine outlier diagnostics, including leverage, jackknife residuals, and Cook's distance, and identify any potential outliers.
 - d. Find the least-square estimated equations for a quadratic model of SAL (Y) on CGPA (X).
 - e. Is there a collinearity problem for the model in part (d) based on VIF? If yes, remove the collinearity and find the estimated equations for the quadratic model.
 - f. Test for the significance of the addition of the X^2 term to the model in part (e).
 - g. Can we use lack-of-fit test to check the adequacy of fit of the quadratic model? If yes, perform the lack-of-fit test; if not, explain why.