Fall 2021 Applied Statistics Qualifier

Problem 1

The yield of a chemical process (Y) depends on Temperature ($Temp$) in degrees C and Pressure ($Press$) in kg/cm^3. The following model is applicable to this process:

$$Y = \gamma_0 Temp^{\gamma_1} Press^{\gamma_2}$$

The data set contains n=18 tests to study the influence of Temperature and Pressure on the process yield.

a. Using an appropriate method, obtain the starting values for $\gamma_0,\ \gamma_1,\ and\ \gamma_2$. Provide your program and output to demonstrate how you arrived at these values.

b. Using the starting values obtained in part a, find the least squares estimates of the parameters. Continue evaluating the model, adjusting the predictors if necessary, and state the final equation here. Provide evidence of a valid model.

c. Based on the model found in part b, determine if Temperature and Pressure are statistically significant predictors.

d. With 95% Confidence, verify by hand the SAS output estimating the influence of Temperature on process yield.

Problem #2

A local health clinic sent fliers to its clients to encourage everyone, but especially persons at high risk, to get a flu shot in time f or protection against an outbreak. In a small pilot study, n=159 clients were randomly selected and asked if they had actually received a flu shot. The data is summarized as follows:

Shot: 1=Yes, 0=No for receiving the flu shot.

Age: The age of the subject in the study.

Aware: A healthscore based on a series of survey questions. The higher the score, the healthier the client.

Gender: 1= Male, 0= Female.

a. Using the data set provided, create a valid regression model that can be used to predict the probability that a client gets a flu shot. Demonstrate (diagnostics, p-values, etc.) that you have created a valid model. State you final model with the coefficients.

b. Discuss how each significant predictor influences the probability (or odds) of getting the shot. Use a point estimate rather than a confidence interval.

c. Find the probability, and also the odds, of a Male aged 65 with a Health Awareness of 60 getting a flu shot.

d. Assume that a cutoff of p=.35 has been decided. In other words, if the model predicts a probability greater than or equal to .35, we conclude that the client will get the flu shot. If the model predicts a probability less than .35, we conclude that the client will NOT get the flu shot. Based on this situation, determine the misclassification rate for this model by indicating the number of incorrect predictions.

Problem #3

A computer software firm was encountering difficulties in forecasting the amount of time needed for large scale upgrades of their system. As part of a study to address these difficulties, n=24 programmers were classified into equal groups by type of experience.

(Factor A: 1=Small systems and 2=Large Systems) and years of experience.
(Factor B: 1=under 5 years, 2= 5 to 10 years, 3=more than 10 years experience).

Each programmer was then given an upgrade assignment and asked to predict the number of days necessary to complete the project. After the actual project was completed, the prediction error (forecast time minus actual time) became the dependent variable (Y)..

Consider this a two-factor completely crossed Analysis of Variance study. The overall objective of this study is to investigate the influence both type of experience and years of experience have on Y(prediction error). We will also employ a simultaneous comparison procedure with the primary objective being the pairwise comparison of factor level means.

a. Create an appropriate and valid statistical model that can examine the pairwise comparison of factor means. Justify the validity of your model in terms of diagnostics using software output and your comments. If any modifications to the data are required, discuss your modifications then demonstrate the validity of your modified model. Assuming a valid model is created, you can then use your model to answer parts b and c.

b. Discuss the results of your model. Specifically:

1. Is there evidence of any statistically significant interaction between the factors? Discuss why or why not. What does the possibility or absence of interaction between the factors actually indicate within the context of this analysis.
2. Is there any evidence of a difference between factor level means (cell means)? Discuss why or why not.

c. As previously stated, the research objective is the pairwise comparison of the factor level means. We wish to understand the influence of the simple main or main effects on the dependent variable. Using an appropriate multiple comparison procedure, provide simultaneous comparisons (underlining) of the treatment means. Justify the multiple comparison procedure you have chosen. Interpret your findings.