**Department of Statistics**

**Quaid-i-Azam University, Islamabad**

**Linear Models Zahid Asghar**

Note: Attempt all the questions. Question No.1 has 15 points.

Question #1

1. In the expression Sales ≈ f(TV, Radio, Newspaper), "Sales" is the:
2. Response
3. Training Data
4. Independent Variable Feature
5. While doing a homework assignment, you fit a Linear Model to your data set. You are thinking about changing the Linear Model to a Quadratic one. Which of the following is most likely true:
6. Using the Quadratic Model will decrease your Irreducible Error.
7. Using the Quadratic Model will decrease the Bias of your model.
8. Using the Quadratic Model will decrease the Variance of your model
9. Using the Quadratic Model will decrease your Reducible Error
10. A hypercube with side length 1 in d dimensions is defined to be the set of points (x1, x2, ..., xd) such that0≤*xj*≤1 for all j = 1, 2, ..., d. Define the boundary region of the hypercube to be the set of all points such that there exists a j for which 0≤*xj*≤.05 or .95≤*xj*≤1 (i.e., it is the set of all points that have at least one dimension in the most extreme 10% of possible values). What proportion of the volume of a hypercube of dimension 50 is in the boundary region? Please give your answer as a value between 0 and 1 with 3 significant digits. If you think the answer is 50.52%, you should say 0.505: 0.995 (2 points)
11. True or False: A fitted model with more predictors will necessarily have a lower Training Set Error than a model with fewer predictors.
12. Why is linear regression important to understand? Select all that apply:
13. Top of Form

a) The linear model is often true

1. Linear regression is very extensible and can be used to capture nonlinear effects
2. Simple methods can outperform more complex ones if the data are noisy
3. Understanding simpler methods sheds light on more complex ones
4. Which of the following are true statements? Select all that apply:
5. Top of Form
   1. A 95% confidence interval is a random interval that contains the true parameter 95% of the time
   2. If I perform a linear regression and get confidence interval from 0.4 to 0.5, then there is a 95% probability that the true parameter is between 0.4 and 0.5.
   3. The true parameter (unknown to me) is 0.5. If I sample data and construct a 95% confidence interval, the interval will contain 0.5 95% of the time.
6. Which of the following indicates a fairly strong relationship between X and Y?
7. R2=0.9
8. The p-value for the null hypothesis β1=0 is 0.0001
9. The t-statistic for the null hypothesis β1=0 is 30
10. What is the difference between lm(y ~ x\*z) and lm(y ~ I(x\*z)), when x and z are both numeric variables?
11. The first one includes an interaction term between x and z, whereas the second uses the product of x and z as a predictor in the model.
12. The second one includes an interaction term between x and z, whereas the first uses the product of x and z as a predictor in the model.
13. The first includes only an interaction term for x and z, while the second includes both interaction effects and main effects.
14. The second includes only an interaction term for x and z, while the first includes both interaction effects and main effects.

9. What is the difference between lm(y ~ x\*z) and lm(y ~ I(x\*z)), when x and z are both numeric variables?

a) The first one includes an interaction term between x and z, whereas the second uses the product of x and z as a predictor in the model.

b) The second one includes an interaction term between x and z, whereas the first uses the product of x and z as a predictor in the model.

c) The first includes only an interaction term for x and z, while the second includes both interaction effects and main effects.

d) The second includes only an interaction term for x and z, while the first includes both interaction effects and main effects.

10. Suppose we collect data for a group of students in a statistics class with variables X1= hours studied, X2= undergrad GPA, and Y= receive an A. We fit a logistic regression and produce estimated coefficients β^o=−6, β^1=0.05,β^2=1.Estimate the probability that a student who studies for 40h and has an undergrad GPA of 3.5 gets an A in the class:

11. Which of the following is NOT a linear function in x:

1. f(x)=a+b2x
2. The discriminant function from LDA
3. δk(x)=xμkσ2−μ2k2σ2+log(πk)
4. logit(P(y=1|x)) where P(y=1|x) is as in logistic regression
5. P(y=1|x) from logistic regression

**Question #2 a)** Describe the differences between a parametric and a non-parametric statistical learning approach. What are the advantages of a parametric approach to regression or classification (as opposed to a nonparametric approach)? What are its disadvantages?

*Parametric approaches do not have as much of an issue with overfitting. They are more resitrictive, but as a result fewer observations of underlying data are required to fit them fairly well. Nonparametric methods are sometimes required when the underlying data is very different from what can be fit with parametric techniques, or when the underlying distribution is unknown but obviously not normally distributed.* (5 Points)

1. Carefully explain the differences between the KNN classifier and KNN regression methods.(5 Pts)

The KNN clasifier took a single observation as input, and yielded an output that is the average of the K nearest inputs. KNN regression on the other hand takes an abstract point as input, and treating it as if it were a real observation, estimates the output it would predict if it were real, it becomes a function that outputs a prediction for any input, and can be used for determining prediction boundaries, etc.

**Question#3** We now examine the differences between LDA and QDA. (10 Points)

1. If the Bayes decision boundary is linear, do we expect LDA or QDA to perform better on the training set? On the test set?

*If the Bayes decision boundary is linear, we expect QDA to perform better on the training set because it's higher flexiblity will yield a closer fit. On the test set, we expect LDA to perform better than QDA because QDA could overfit the linearity of the Bayes decision boundary.*

1. If the Bayes decision boundary is non-linear, do we expect LDA or QDA to perform better on the training set? On the test set?

*If the Bayes decision bounary is non-linear, we expect QDA to perform better both on the training and test sets.*

1. In general, as the sample size n increases, do we expect the test prediction accuracy of QDA relative to LDA to improve, decline, or be unchanged? Why?

*We expect the test prediction accuracy of QDA relative to LDA to improve, in general, as the the sample size $n$ increases because a more flexibile method will yield a better fit as more samples can be fit and variance is offset by the larger sample sizes.*

**Best of Luck**