**STAT 512**

**HW 1**

**Name: Haoran Zhang**

**Section: 010**

**Time: 1:30-2:45 pm Tuesday/Thursday**

1. (13 pts.) A regression analysis relating test scores (Y) to training hours (X) produced the following fitted equation: Ŷ = 20 – 0.6X. Be sure to show your work for all parts in this question.
   1. (2) What is the fitted value of the response variable corresponding to X = 4?

When

* 1. (2) What is the residual corresponding to the data point with X = 5 and Y = 21?
  2. (2) If X decreases by 3 units, how does Ŷ change?
  3. (2) An additional test score is to obtained for a new observation at X = 7. Would the test score for the new observation necessarily be 15.8? Explain.

No, the equation gives an estimated mean of Y as 15.8 for , but the residual cannot be neglected.

* 1. (2) The error of sums of squares (SSE) for this model was found to be 15.3. If there were n = 20 observations, provide the best estimate for σ2.

Provided that MSE is the unbiased estimator of (), so

* 1. (3) Rewrite the regression equation in terms of X\* where X\* is training time measured in minutes. Show that your answer makes sense, i.e., gives the same predictions as the original equation (one example is sufficient).

Check: when ,

1. (13 pts.) The director of admissions of a small college selected 10 students at random from the new freshman class in a study to determine whether a student’s GOA at the end of the freshman year can be predicted from the ACT test score. The data is GPA.csv.
   1. (4) Compute a linear regression to predict GPA based on the ACT test score.

Therefore, the fitted equation is:

* 1. (2) Compute the residual standard deviation,
  2. (2) Give a point estimate and a 95% confidence interval for the slope and intercept and interpret each of these in words. (*Point estimate* is another word for parameter estimate.) Why would someone be interested if 0 is included in the interval for β1?

are point estimate for slope and intercept separately, there are the best choices, since they are unbiased and have minimum variance

It’s 95 percent confident that the slope will be in the range of .

It’s 95 percent confident that the intercept will be in the range of .

Some people are interested in testing if 0 is included in the interval for , because when , there is no linear association between the X and Y.

* 1. (5) Perform a hypothesis test on whether ACT test score is associated with GPA.

Choosing ,

According to the T table, p-value

Since, p value >0.05, cannot reject the null hypothesis.

3. (3 pts) When conducting statistical tests concerning the parameter, why is the T test more versatile than the F test?

The T test can be used for both one-sided and two-sided tests, but the F test can be only used for two-sided test.

4. (3 pts) For each of the following questions, explain whether a confidence interval for a mean response or a prediction interval for a new observation is appropriate.

a). What will be the humidity level in this greenhouse tomorrow when we set the temperature level at 31 °C?

The prediction interval is appropriate, because it is to predict the humidity level on next day when temperature is to be set at 31 °C.

b). How much do families whose disposable income is $23,500 spend, on the average, for meals away from home?

The confidence interval for a mean response is appropriate, because it is asked to calculate the confidence interval of the average families spending on meals.

c). How many kilowatt-hours of electricity will be consumed next month by commercial and industrial users in the Twin Cities service area, given that the index of business activity for area remains at its present level?

The prediction interval is appropriate, because it is needed to predict the next consumption of electricity at given level.

5. (8 pts) A substance used in biological and medical research is shipped by air fright to users in cartons of 1,000 ampules. The data below, involving 10 shipments, were collected on the number of times the carton was transferred from one aircraft to another over shipment route (X) and the number of ampules found to be broken upon arrival (Y). Assume a simple linear regression is appropriate. Data: Airfreight.csv

a) (2) Verity that the fitted regression line goes through the point (

Therefore, the fitted equation is:

Check:

b) (2) Because of changes in airline routes, shipments may have to be transferred more frequently than in the past. Estimate the mean breakage for the following numbers of transfers: X = 2. Use separate 99 percent confidence intervals. Interpret your results

When , it is 99 percent confident that the mean number of breakages is in the range of (15.976, 20.424).

c) (2) Next shipment will entail two transfers. Obtain a 99 percent prediction interval for the number of broken ampules for this shipment. Interpret your prediction interval.

When next shipment will entail 2 transfers, it is 99 percent confident that the number of breakages will drop into the range of (12.748, 23.652).

1. (2) In the next several days, three independent shipments will be made, each entailing two transfers. Obtain a 99 percent prediction interval for the mean number of ampules broken in three shipments.

6. (10) Write the name of the following notations, related formulas (if there is one) and describe what they measure.

1. dependent variable (response variable)

It is the measurement of one variable of one set of observations.

1. mean of Y

It measures the average of the dependent variable.

1. estimated value of Y according to the linear regression model

It is the value of the estimated regression function at the level X of the predictor variable

1. (Epsilon) error

It is the value of independent and normally distributed random errors with mean 0 and variance

1. residual

The residual is the difference between the observed value and the corresponding fitted value.

1. slope

It is the slope of the linear regression model

1. point estimate of

Under the linear regression model, is the least squares estimator of slope.

1. error standard deviation

it is the standard deviation of error term.

1. estimation of error standard deviation
2. error mean squared error

It is an unbiased estimator of .

1. point estimator of

it stands for the estimate of the standard deviation of for linear regression model.

1. point estimator of

Under normal error regression model, it predicts the standard deviation of single estimated value of response value

1. point estimator of standard deviation

It predicts the standard deviation of next single observation on Y for a given level of predictor variables.

1. point estimator of standard deviation

It predicts the standard deviation of m new observations on Y for a given level of predictor variables.