STAT 516 Fall 2004 Quiz 1-3 Answers

QUIZ 1

1) Under certain assumptions, the quotient of **2 means** / **2 variances** follows a **t** / χ^2 / ***F*** distribution.

2) A *t* with df=*n*-1 is related to an *F* with df=1 and *n*-1 by the relationship ***t*/2 = *F*** / $t^2=F$ / $\sqrt{t} = F$.

A linear regression is performed to predict housing prices (in $) from the size (in square feet). Assume that all of the assumptions for a regression were met. The regression returned the following information:

Price = 58,000 + 32 Area MSE = 21,270,000 Square root of MSE = 4,612 p-value = 0.06

3) What price does this equation estimate for a 1,000 sq. ft. house? **58,000+32(1,000)=$90,000**

4) For each additional 100 square feet, how much does the price change by? **32(100)=$3,200**

5) Would you be surprised if the real price was off by more than $15,000 from the predicted price? Why or why not? **Yes we should be surprised. The root-MSE is $4,612, so a house that was $15,000 off would be more than three-standard deviations away. If normality holds this would happen less than 0.3% of the time.**

6) At α=0.05 would we accept or reject H_0:β_1=0? **With a p-value of 0.06 we would accept (that is, fail to reject H_0). The p-value is not less than α.**

QUIZ 2

1) Complete the following partial ANOVA table for a regression that had 92 observations.

Source	SS	df	MS	F	p-value
Regression	1673142	1	1673142	____C___	0.000
Error	___A____	90	1075		
Total	1769888	B			

A = Either 1,769,888-1,673,142=96,746 or 90*1075=96,750 (the difference is rounding error in the MS)
B= Either 1+90=91 or 92-1=91
C= 1673142/1075=1556.411

Questions 2-5 concern the attached SAS output for predicting the amount of protein in minnow larvae from the concentration of metals in the water.

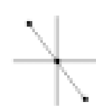
2) For each extra unit of metals in the water, what is the predicted change to the protein level? **slope=-0.2912**

3) At α=0.05, do you accept or reject the null hypothesis that β_1=0? **Reject as p-value .0001 < α=0.05.**

4) On average, how far do you expect each of the observations to be from the regression line? **Root-MSE=22.2899**

5) What percentage of the variation in the protein levels of the minnows is explained by the amount of metals in the water?

6) What is the correlation coefficient for the figure at right? **-1**



R-square=.5106=51.06%

QUIZ 3

1. State the four assumptions the errors must satisfy in order for a regression analysis to be valid. **The errors must: have mean zero (linear form is appropriate), constant variance, be normally distributed, and be independent.**

2. Define what is meant by the *p-value*. **The probability of observing a test-statistic at least as extreme as the one observed given that the null hypothesis is true.**

3. Ninety-five percent of all future observations should fall within the bands of a 95% **prediction interval** / **confidence interval for the mean** . When comparing these two intervals, the confidence interval for the mean is the **narrower** / **wider** of the two.

Questions 4-6 use the SAS output described on the attached page.

4. State what H_0 is being tested by the Type I test p-value of 0.3673 for Area. **β_{area}=0 given non-native is included in the model.**

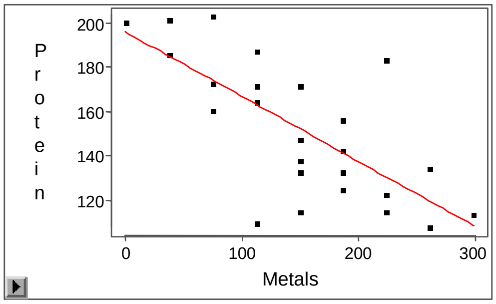
5. State what H_0 is being tested by the Type III test p-value of 0.3026 for Area. **β_{area}=0 given non-native, elevation, distnear, and distsc are included in the model.**

6. Circle all of the models that would be acceptable choices to predict the number of native species. Put a star next to the one that would be the best if you wanted to use the simplest possible model. **Those in bold are definite, those in italics are borderline.**

Protein	=	Metals
Response Distribution:		Normal
Link Function:		Identity

Model Equation

Protein = 195.880 - 0.2912 Metals



Parametric Regression Fit

Curve	Degree(Polynomial)	Model DF	Model Mean Square	Error DF	Error Mean Square	R-Square	F Stat	Pr > F
	1	1	11924.6400	23	496.8417	0.5106	24.00	<.0001

Summary of Fit

Mean of Response	152.2000	R-Square	0.5106
Root MSE	22.2899	Adj R-Sq	0.4894

Analysis of Variance

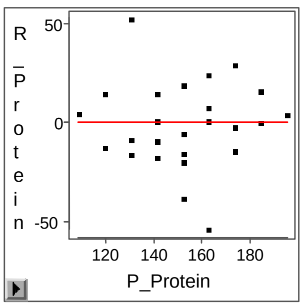
Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
Model	1	11924.6400	11924.6400	24.00	<.0001
Error	23	11427.3600	496.8417		
C Total	24	23352.0000			

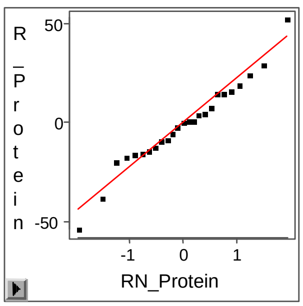
Type III Tests

Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
Metals	1	11924.6400	11924.6400	24.00	<.0001

Parameter Estimates

Variable	DF	Estimate	Std Error	t Stat	Pr >|t|	Tolerance	Var Inflation
Intercept	1	195.8800	9.9684	19.65	<.0001	.	0
Metals	1	-0.2912	0.0594	-4.90	<.0001	1.0000	1.0000





The data set *Galapagos* is from an article in the journal *Science* by Johnson and Raven. It concerns the number of native species on the various Galapagos Islands based on the number of non-native species (`NonNative`), the area of the island in km^2 (`Area`) , the elevation of the island in m (`Elev`), the distance from the nearest other island in km (`DistNear`), and the distance from Santa Cruz in km (`DistSC`). The accompanying SAS output was generated from this data set, and you may assume the assumptions for performing regression are met.

Model Equation

Native = 6.7437 + 0.2409 NonNative - 0.0025 Area + 0.0184 Elev + 0.0982 DistNear - 0.0265 DistSC

Type I Tests

Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
NonNative	1	19339.7233	19339.7233	382.89	<.0001
Area	1	42.6551	42.6551	0.84	0.3673
Elev	1	524.2878	524.2878	10.38	0.0036
DistNear	1	2.5511	2.5511	0.05	0.8241
DistSC	1	54.0551	54.0551	1.07	0.3112

Type III Tests

Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
NonNative	1	5831.6001	5831.6001	115.46	<.0001
Area	1	56.0507	56.0507	1.11	0.3026
Elev	1	554.4144	554.4144	10.98	0.0029
DistNear	1	34.0055	34.0055	0.67	0.4200
DistSC	1	54.0551	54.0551	1.07	0.3112

R-Square Selection Method

Number in Model	R-Square	Adjusted R-Square	C(p)	Variables in Model
1	0.9133	0.9102	10.3452	NonNative
1	0.6253	0.6119	131.0884	Elev
1	0.3836	0.3616	232.4104	Area
1	0.0276	-.0071	381.6643	DistSC
1	0.0000	-.0357	393.2369	DistNear
2	**0.9372**	**0.9326**	**2.3191**	**Elev NonNative************
2	0.9153	0.9090	11.5007	Area NonNative
2	0.9137	0.9073	12.1833	DistNear NonNative
2	0.9133	0.9069	12.3417	DistSC NonNative
2	0.6489	0.6229	123.2110	DistSC Elev
2	0.6265	0.5989	132.5713	Area Elev
2	0.6254	0.5977	133.0412	DistNear Elev
2	0.3945	0.3496	229.8550	Area DistSC
2	0.3887	0.3434	232.2985	Area DistNear
2	0.0450	-.0257	376.3713	DistNear DistSC
3	**0.9401**	**0.9332**	**3.1207**	**Area Elev NonNative**
3	*0.9379*	*0.9308*	*4.0141*	*DistSC Elev NonNative*
3	*0.9376*	*0.9304*	*4.1650*	*DistNear Elev NonNative*
3	0.9160	0.9063	13.2280	Area DistNear NonNative
3	0.9153	0.9056	13.4967	Area DistSC NonNative
3	0.9141	0.9041	14.0277	DistNear DistSC NonNative
3	0.6666	0.6282	117.7638	DistNear DistSC Elev
3	0.6491	0.6086	125.1227	Area DistSC Elev
3	0.6268	0.5837	134.4600	Area DistNear Elev
3	0.4234	0.3569	219.7193	Area DistNear DistSC
4	**0.9411**	**0.9317**	**4.6733**	**Area DistSC Elev NonNative**
4	*0.9402*	*0.9306*	*5.0702*	*Area DistNear Elev NonNative*
4	*0.9401*	*0.9305*	*5.1097*	*DistNear DistSC Elev NonNative*
4	0.9166	0.9032	14.9764	Area DistNear DistSC NonNative
4	0.6674	0.6141	119.4555	Area DistNear DistSC Elev
5	**0.9428**	**0.9308**	**6.0000**	**Area DistNear DistSC Elev NonNative**