Statistics 515 – Fall 2000 - Exam 3 ***(modified a bit)***

Part I: Answer seven of the following eight questions. If you complete more than eight, I will grade only the first eight. Five points each.

1) Define what is meant by the *p-value* (or the *observed significance level*) of a test.

2) Consider the following partial ANOVA table for a one-way analysis of variance. How many different treatment groups were there in this experiment?

Source	SS	DF
Treatments	2794.39	3
Error	762.30	36
Total	3556.69	39

3) A one-way analysis of variance has five different treatment groups. The population means for these five groups are μ_1, μ_2, μ_3, μ_4, and μ_5. In terms of these parameters, specify the null and alternate hypotheses that are tested by the p-value in the ANOVA table.

4) In performing a linear regression to predict *y* from *x*, what four assumptions must be satisfied?

5) The basic regression equation is $y = \beta_0 + \beta_1 x + \varepsilon$.
Identify which term in the equation is the:
dependent variable_____ slope _____
independent variable_____ intercept _____
error _____

6) PROC INSIGHT produces three plots when performing linear regression or a one-way ANOVA. The **scatter plot** of the independent and dependent variable, the **residual vs. predicted plot**, and the **q-q plot of the residuals**. Which assumption(s) do you check by looking at the residual vs. predicted plot?

7) (Circle the correct answer). In simple linear regression, the values of β_0 and β_1 are chosen so that they minimize the **SSR / SSE / TSS / β_0 / β_1**.

8) (Circle the correct answers) If the assumptions of a regression model for predicting *y* from *x* are met, and we do not reject the null hypothesis that $\beta_1=0$, then we conclude that *x* **can /cannot** be used to predict *y*. If we do reject the null hypothesis that $\beta_1=0$ then we **may/may not** conclude that *x* causes *y*.

Part II: Answer every part of the next two problems. Read each problem carefully, and show your work for full credit. Twenty points each.

1) In a 1929 paper, Edwin Hubble reported a relationship between the estimated distance of a nebula from the earth and its estimated velocity moving away from the earth. The attached output includes a regression on part of that data set. The variables used are

DISTANCE = distance to the nebula in megaparsecs (30.9 million trillion km)
VELOCITY = velocity away from the earth in kilometers/second

a) Assume that the assumptions of the regression model are met and that we accept the equation of the regression line that was found. An increase in velocity of 10 km/sec would correspond to how many more megaparsecs of distance?

b) At α=0.05 do we accept or reject H_0: β_1=0?

c) Assume that the assumptions of the regression model are met and that we accept the equation of the regression line that was found. What percent of the variation in the estimated distance is explained by the estimated velocity?

d) Assume that the assumptions of the regression model are met and that we accept the equation of the regression line that was found. If a new nebula with a velocity of 500 km/sec is observed, give a 95% prediction interval for its distance.

e) Comment on each of the assumptions that can be checked from the two residual plots. Say if each of these assumptions seems to hold or not, and how you can tell.

2) The following is the incomplete work for a linear regression problem.

SS_{xx} = 10.000 average x = 5.000
SS_{yy} = 6.000 average y = 2.000
SS_{xy} = 7.000

Source	SS	DF	MS	F	Prob>F
Regression	4.900	1	_____	_____	0.0354
Error	1.100	3	_____		
Total	_____	_____			

a) Complete the table above by writing in the missing values.

b) Determine the estimated regression equation.

c) What was the original sample size?

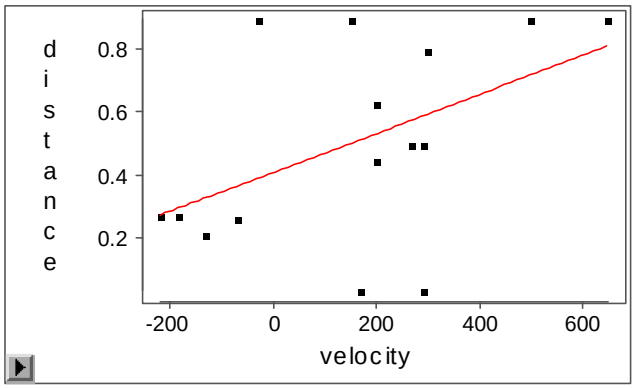
d) Determine a 90% confidence interval for the slope β_1.

```
DATA nebulae;
INPUT velocity distance;
CARDS;
 170    0.032
 290    0.034
-130    0.214
 -70    0.263
-185    0.275
-220    0.275
 200    0.450
 290    0.500
 270    0.500
 200    0.630
 300    0.800
 -30    0.900
 650    0.900
 150    0.900
 500    0.900
;
PROC INSIGHT;
OPEN nebulae;
FIT distance=velocity;
RUN;

PROC REG DATA=nebulae;
MODEL distance=velocity / ALPHA=0.05 CLI CLM;
RUN;
```

Model Equation

distance = 0.4065 + 0.0006 velocity



Summary of Fit

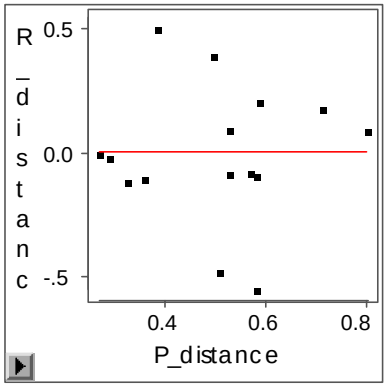
Mean of Response	0.5049	R-Square	0.2312
Root MSE	0.2902	Adj R-Sq	0.1721

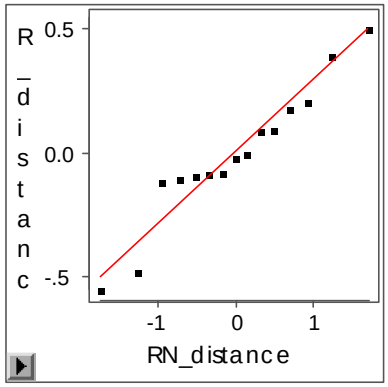
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
Model	1	0.3294	0.3294	3.91	0.0696
Error	13	1.0950	0.0842		
C Total	14	1.4244			

Parameter Estimates

Variable	DF	Estimate	Std Error	t Stat	Pr >|t|	Tolerance	Var Inflation
Intercept	1	0.4065	0.0899	4.52	0.0006	.	0
velocity	1	0.0006	0.0003	1.98	0.0696	1.0000	1.0000





5					
	lnt	lnt	lnt	lnt	lnt
15	velocity	distance	R_distance	P_distance	RN_distance
1	170	0.032	-0.4797	0.5117	-1.2450
2	290	0.034	-0.5519	0.5859	-1.7394
3	-130	0.214	-0.1121	0.3261	-0.9458
4	-70	0.263	-0.1002	0.3632	-0.7137
5	-185	0.275	-0.0171	0.2921	0.0000
6	-220	0.275	0.0045	0.2705	0.1651
7	200	0.450	-0.0802	0.5302	-0.3349
8	290	0.500	-0.0859	0.5859	-0.5150
9	270	0.500	-0.0735	0.5735	-0.1651
10	200	0.630	0.0998	0.5302	0.5150
11	300	0.800	0.2079	0.5921	0.9458
12	-30	0.900	0.5120	0.3880	1.7394
13	650	0.900	0.0915	0.8085	0.3349
14	150	0.900	0.4007	0.4993	1.2450
15	500	0.900	0.1842	0.7158	0.7137

The REG Procedure
Model: MODEL1
Dependent Variable: distance

Output Statistics

Obs	Dep Var distance	Predicted Value	Std Error Mean Predict	95% CL Mean		95% CL Predict		Residual
1	0.0320	0.5117	0.0750	0.3496	0.6737	-0.1359	1.1593	-0.4797
2	0.0340	0.5859	0.0854	0.4014	0.7704	-0.0677	1.2395	-0.5519
3	0.2140	0.3261	0.1174	0.0725	0.5798	-0.3502	1.0025	-0.1121
4	0.2630	0.3632	0.1037	0.1393	0.5872	-0.3026	1.0290	-0.1002
5	0.2750	0.2921	0.1311	0.008849	0.5754	-0.3959	0.9801	-0.0171
6	0.2750	0.2705	0.1402	-0.0325	0.5734	-0.4259	0.9668	0.004541
7	0.4500	0.5302	0.0760	0.3660	0.6945	-0.1179	1.1784	-0.0802
8	0.5000	0.5859	0.0854	0.4014	0.7704	-0.0677	1.2395	-0.0859
9	0.5000	0.5735	0.0826	0.3951	0.7519	-0.0784	1.2254	-0.0735
10	0.6300	0.5302	0.0760	0.3660	0.6945	-0.1179	1.1784	0.0998
11	0.8000	0.5921	0.0870	0.4042	0.7799	-0.0625	1.2466	0.2079
12	0.9000	0.3880	0.0954	0.1818	0.5942	-0.2721	1.0480	0.5120
13	0.9000	0.8085	0.1709	0.4394	1.1777	0.0809	1.5362	0.0915
14	0.9000	0.4993	0.0750	0.3373	0.6613	-0.1483	1.1469	0.4007
15	0.9000	0.7158	0.1303	0.4342	0.9974	0.0284	1.4031	0.1842

Sum of Residuals	0
Sum of Squared Residuals	1.09504
Predicted Residual SS (PRESS)	1.33408