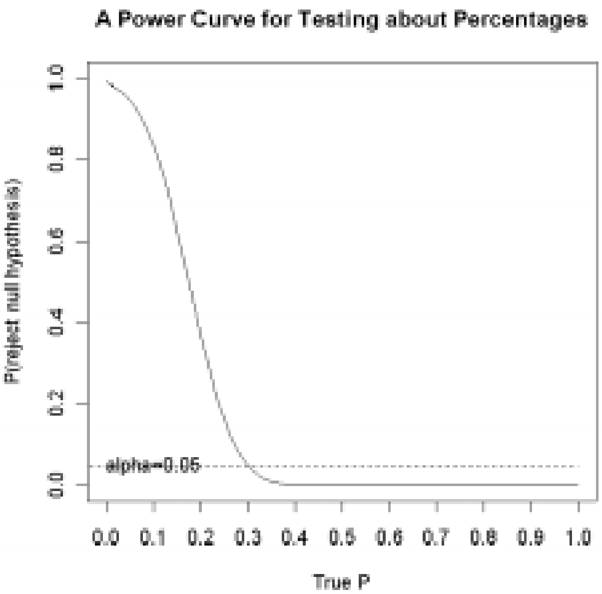
Statistics 515 Practice Exam 3 ***(Based on the Fall 2001 Exam)***

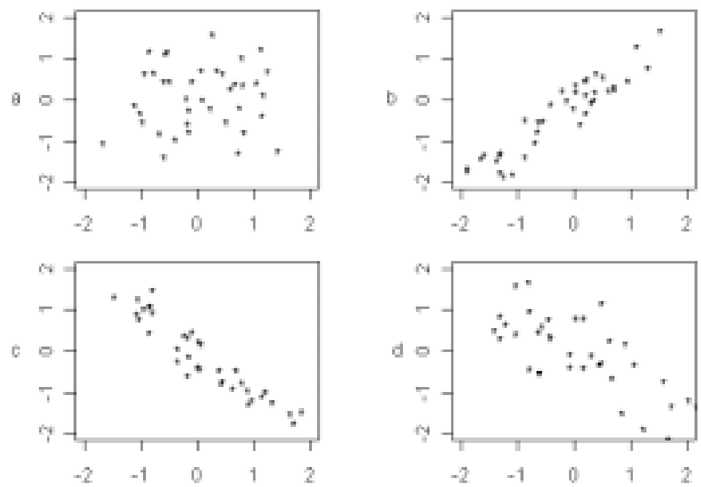
Part I: Answer seven of the following eight questions. If you complete more than seven, I will grade only the first seven. Five points each.

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

2) In performing a simple linear regression or a one-way ANOVA, what four assumptions must be satisfied?

3) A power curve is shown below for a test of hypothesis about a population percentage p. The power curve is for a test with α=0.05. What null and alternate hypotheses is the curve for?





4) For the graphs given above, identify which of them comes from a population with correlation coefficient:
r = -0.95 _____ r = -0.6 _____ r = 0.0 _____ r = 0.95 _____

5) (Circle the correct answer) When simple linear regression is performed, the
$\sqrt{MSR}$ / $\sqrt{MSE}$ / MSR / r^2 / r is the estimate of the standard deviation (σ) of the errors.

Questions 6-8 refer to the following partial ANOVA table for a one-way analysis of variance.

Source	SS	df	MS	F	p-value
Treatments	529.11	8	66.14	1336.85	0.0001
Error	0.89	18	0.05		
Total	530.00	26			

6) In the experiment conducted to gather the above data there were _____ different treatments and _____ total observations.

7) What null and alternate hypothesis are being tested by the p-value given above? Identify any parameters that you use in stating the hypothesis (e.g. β_0 is the intercept, μ_1 is the first mean).

8) To test the assumptions for this one-way ANOVA we could construct a **Q-Q plot** for each of the treatment groups and also make **side by side boxplots** of the observations in each treatment group. What assumption is checked by constructing a Q-Q plot for each of the treatment groups?

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) The attached data set `bears` can be found in *Sports Afield*, (September, 1981). The data concerns several bears that were captured and released. For obvious reasons it would be useful if one could estimate the weight of a bear by simply calculating their length from a photograph (instead of trying to get it to step on a scale.) Note that, in particular, the first bear has a length of 78 inches and a weight of 334 pounds.

a) Assume the assumptions of the regression model are met. If the we observe a bear of length 28 inches, its estimated weight would be -111.43 pounds!?! Why should we not be concerned about this?

b) Assume the assumptions of the regression model are met. Construct a 95% confidence interval for the slope β_1.

c) Assume the assumptions of the regression model are met. If a new bear with a length of 78 inches is observed, what range are you 95% certain the new bear's weight will fall in?

d) Assume the assumptions of the regression model are met. What percent of the variation in the bears weight is explained by their length?

e) Which one of the four assumptions can we tell is violated because the residual vs. predicted plot makes a funnel shape (is wider on the right than the left)?

2) The following is the incomplete work for a simple linear regression for predicting a variable y from another variable x.

SS_{xx} =	10.0	average x = 3.0
SS_{yy} =	7.2	average y = 3.6
SS_{xy} =	-7.0	n = 5

Source	SS	DF	MS	F	Prob>F
Regression	______	_____	______	6.3913	0.08557
Error	2.3000	3	_______		
Total	______	_____			

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

b) Determine the estimated regression equation.

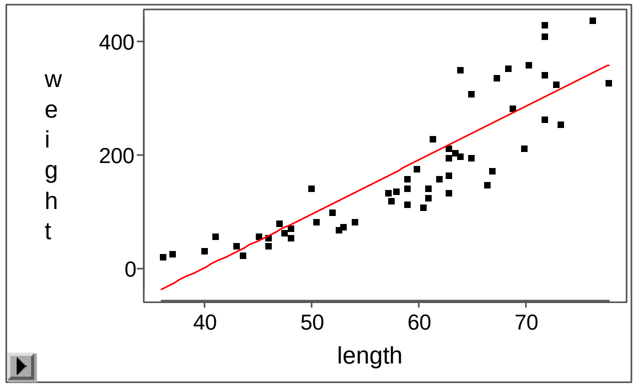
c) What null and alternate hypothesis are being tested by the p-value given above? Identify any parameters that you use in stating the hypothesis (e.g. β_0 is the intercept, μ_1 is the first mean).

d) Do we accept or reject that null hypothesis at an α-level of 0.05?

```
DATA bears;
INPUT length weight @@;
CARDS;
78      334   59      120   69      289
43.5    29    72      416   62      166
45      65    50.5    90    57.3    140
60      182   50      148   47      86
47.5    70    72      436   76.5    446
61      150   57.5    125   67      180
72      348   46      62    59      150
63      172   61.5    236   68.5    360
58      144   61      132   72      270
53      80    54      90    63      140
52.5    76    63.5    212   65      202
64      356   40      40    52      105
73.5    262   63      220   59      166
67.5    344   48      60    36      26
65      316   43      46    63      202
73      332   64      204   64      204
46      48    41      64    70.5    365
70      220   66.5    154   48      79
37      34    60.5    116
;

PROC INSIGHT;
OPEN bears;
FIT weight=length;
RUN;

PROC GLM DATA=bears;
MODEL weight=length / ALPHA=0.05 CLI;
RUN;

PROC GLM DATA=bears;
MODEL weight=length / ALPHA=0.05 CLM;
RUN;
```

Model Equation

weight = - 375.540 + 9.4324 length



Summary of Fit

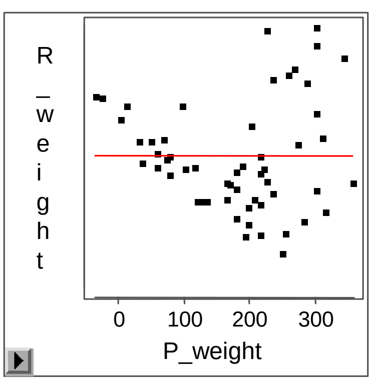
Mean of Response	180.5179	R-Square	0.7671
Root MSE	55.7013	Adj R-Sq	0.7628

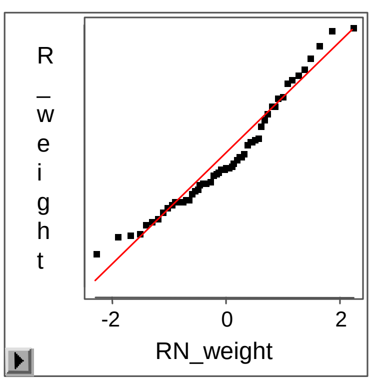
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
Model	1	551729.687	551729.687	177.83	<.0001
Error	54	167542.295	3102.6351		
C Total	55	719271.982			

Parameter Estimates

Variable	DF	Estimate	Std Error	t Stat	Pr >|t|	Tolerance	Var Inflation
Intercept	1	-375.5403	42.3578	-8.87	<.0001	.	0
length	1	9.4324	0.7073	13.34	<.0001	1.0000	1.0000





The GLM Procedure

Observation	Observed	Predicted	Residual	95% Confidence Limits for Individual Predicted Value	
1	334.0000000	360.1886485	-26.1886485	244.3285869	476.0487100
2	120.0000000	180.9726346	-60.9726346	68.3055456	293.6397237
3	289.0000000	275.2968524	13.7031476	161.7322487	388.8614562
4	29.0000000	34.7700970	-5.7700970	-80.0080662	149.5482602
5	416.0000000	303.5941178	112.4058822	189.4176542	417.7705814
6	166.0000000	209.2699000	-43.2699000	96.5199361	322.0198639
7	65.0000000	48.9187297	16.0812703	-65.4723878	163.3098472
8	90.0000000	100.7970495	-10.7970495	-12.5057502	214.0998492
9	140.0000000	164.9375176	-24.9375176	52.2461014	277.6289337
10	182.0000000	190.4050564	-8.4050564	77.7281823	303.0819305
11	148.0000000	96.0808386	51.9191614	-17.2991611	209.4608382
12	86.0000000	67.7835732	18.2164268	-46.1512325	181.7183790
13	70.0000000	72.4997841	-2.4997841	-41.3317003	186.3312686
14	436.0000000	303.5941178	132.4058822	189.4176542	417.7705814
15	446.0000000	346.0400158	99.9599842	230.6573616	461.4226700
16	150.0000000	199.8374782	-49.8374782	87.1329748	312.5419816
17	125.0000000	166.8240019	-41.8240019	54.1381245	279.5098794
18	180.0000000	256.4320089	-76.4320089	143.1883210	369.6756968
19	348.0000000	303.5941178	44.4058822	189.4176542	417.7705814
20	62.0000000	58.3511515	3.6488485	-55.8032300	172.505532
etc...					

The GLM Procedure

Observation	Observed	Predicted	Residual	95% Confidence Limits for Mean Predicted Value	
1	334.0000000	360.1886485	-26.1886485	329.3279124	391.0493845
2	120.0000000	180.9726346	-60.9726346	166.0493602	195.8959090
3	289.0000000	275.2968524	13.7031476	254.6631328	295.9305720
4	29.0000000	34.7700970	-5.7700970	8.2586296	61.2815644
5	416.0000000	303.5941178	112.4058822	279.8223698	327.3658658
6	166.0000000	209.2699000	-43.2699000	193.7333133	224.8064866
7	65.0000000	48.9187297	16.0812703	24.1364995	73.7009599
8	90.0000000	100.7970495	-10.7970495	81.6566320	119.9374670
9	140.0000000	164.9375176	-24.9375176	149.8316764	180.0433588
10	182.0000000	190.4050564	-8.4050564	175.4080863	205.4020265
11	148.0000000	96.0808386	51.9191614	76.4886084	115.6730688
12	86.0000000	67.7835732	18.2164268	45.2010328	90.3661137
13	70.0000000	72.4997841	-2.4997841	50.4444442	94.5551241
14	436.0000000	303.5941178	132.4058822	279.8223698	327.3658658
15	446.0000000	346.0400158	99.9599842	317.0229855	375.0570461
16	150.0000000	199.8374782	-49.8374782	184.6343124	215.0406439
17	125.0000000	166.8240019	-41.8240019	151.7595358	181.8884681
18	180.0000000	256.4320089	-76.4320089	237.6446725	275.2193453
19	348.0000000	303.5941178	44.4058822	279.8223698	327.3658658
20	62.0000000	58.3511515	3.6488485	34.6856922	82.016610
etc...					