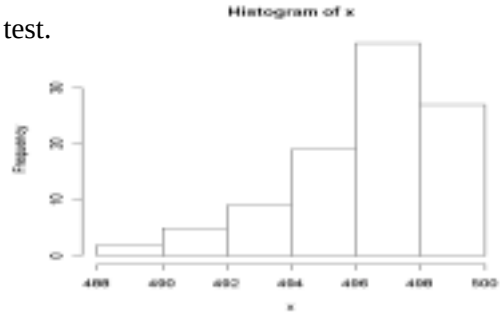
STAT 515 – Fall 2001 – Final Exam *(for practice in Fall 2002)*

Part I: Answer three of the following four questions. If you answer more than three I will grade only the first three. Five points each.

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

2) (Circle the correct answers) The histogram at the right is **skewed left / symmetric /skewed right**. We would expect that its mean would be **larger than / smaller than /equal to** its median.



3) A fair coin (probability of a head on one flip = 0.5) is flipped 10 times. What is the probability of observing exactly 5 heads?

4) A fair coin (probability of a head on one flip = 0.5) is flipped 1,000 times. Using the normal approximation to the binomial, approximately what is the probability of observing at least 512 heads?

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

1) A study in the *Journal of Head Trauma Rehabilitation* (Apr. 1995) attempted to see if a psychological exam called the SCL-90-R could be used to diagnose mild to moderate brain trauma. "Normal" individuals are known to have a score of about 40. A sample of 23 patients diagnosed with mild to moderate traumatic brain injury had a mean score of 48.43 and a standard deviation of 20.76.

A) State the appropriate null and alternate hypothesis to test if a high score on the exam seems to indicate mild to moderate traumatic brain injury. Identify any parameters you use in stating the hypotheses.

B) Test the hypothesis in part A. Report your conclusion at α=0.05 (e.g. Do we accept or reject H_0? Is a high test score associated with traumatic brain injury, or not?)

C) Construct a 95% confidence interval for the mean test score of patients who suffered mild to moderate traumatic brain injury.

D) What assumptions need to be satisfied in order to trust the results you obtained in parts B and C?

2) The attached data set includes the number of games won and the batting average for the 14 teams in the American League for the 1998 season. (Note that the 0.300 batting average is not a real observation and is missing the wins value.)

A) Note that four values have been deleted from the ANOVA table. What values should they have?

DF for total	_____	MS for Error	_____
MS for model	_____	F Statistic	_____

B) Assuming the assumptions for predicting this linear regression are met, perform the test of hypotheses for testing whether batting average predicts wins. What is the p-value, and what is your conclusion at α=0.05?

C) Assuming the assumptions for this simple linear regression are met, identify which of the following statements are true or false.

T / **F** 58.8% of a team's wins are explained by batting average.
T / **F** 58.8% of the variation in the number of wins a team has is explained by the batting average.
T / **F** The correlation coefficient for this regression is 0.588.
T / **F** The coefficient of determination for this regressions is 0.588.

D) Assuming the assumptions for this simple linear regression are met, give the estimated range that the model says 95% of the teams batting 0.300 should have win totals between.

E) Why can't we trust the interval in D?

3) The data below is from a 1992 study in the *Journal of Marketing*. One of the purposes of the study was to compare how men and women remembered products where the advertisement featured a male spokesperson. Samples of 150 male and 150 female viewers were used.

	Male Viewer	Female Viewer
Identified Product	95	41
Could Not Identify Product	55	109

A) Would this data be analyzed by using a test of independence, a test of homogeneity, or a goodness of fit test?

B) Write out the tables of expected values for conducting this test.

C) Give the formula for X^2 for this problem (plugging the values in, but not needing to simplify).

D) What is the rejection region (critical region) for conducting this test at α=0.05?

E) Why is, or why isn't, the sample size of this experiment large enough for performing this hypothesis test?

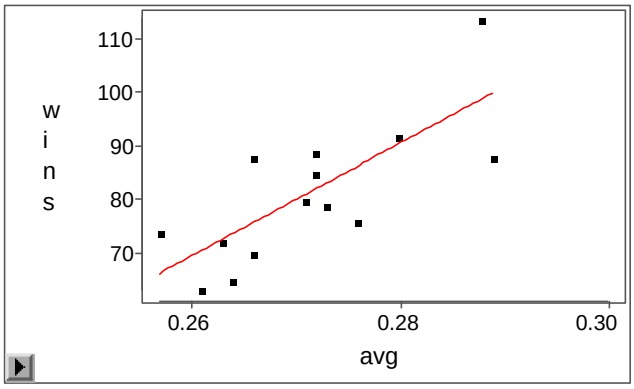
```
DATA alwins;
INPUT team $ wins avg;
CARDS;
NY          114         .288
Tor          88         .266
Balt         79         .273
Bos          92         .280
TB           63         .261
Clev         89         .272
Det          65         .264
Chic         80         .271
KC           72         .263
Minn         70         .266
Ana          85         .272
Tex          88         .289
Sea          76         .276
Oak          74         .257
NewTeam       .         .300
;

PROC INSIGHT;
OPEN alwins;
FIT wins=avg;
RUN;

PROC REG DATA=alwins;
MODEL wins=avg / ALPHA=0.05 CLI CLM;
RUN;
```

Model Equation

wins = - 205.777 + 1057.37 avg



Summary of Fit

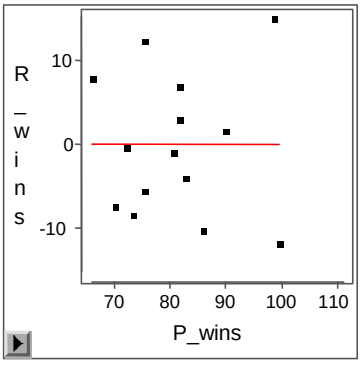
Mean of Response	81.0714	R-Square	0.5880
Root MSE	8.7867	Adj R-Sq	0.5537

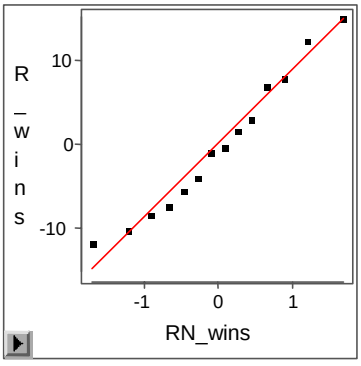
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
Model	1	1322.4642	_____	_____	0.0014
Error	12	926.4644	_____		
C Total	____	2248.9286			

Parameter Estimates

Variable	DF	Estimate	Std Error	t Stat	Pr >|t|	Tolerance	Var Inflation
Intercept	1	-205.7772	69.3480	-2.97	0.0118	.	0
avg	1	1057.3671	255.4804	4.14	0.0014	1.0000	1.0000





		Nom	Int	Int	Int	Int	Int
15		team	wins	avg	R_wins	P_wins	RN_wins
■	1	NY	114	0.288	15.2554	98.7446	1.7076
■	2	Tor	88	0.266	12.5175	75.4825	1.2053
■	3	Balt	79	0.273	-3.8841	82.8841	-0.2670
■	4	Bos	92	0.280	1.7144	90.2856	0.2670
■	5	TB	63	0.261	-7.1957	70.1957	-0.6608
■	6	Clev	89	0.272	7.1733	81.8267	0.6608
■	7	Det	65	0.264	-8.3678	73.3678	-0.8994
■	8	Chic	80	0.271	-0.7693	80.7693	-0.0881
■	9	KC	72	0.263	-0.3104	72.3104	0.0881
■	10	Minn	70	0.266	-5.4825	75.4825	-0.4550
■	11	Ana	85	0.272	3.1733	81.8267	0.4550
■	12	Tex	88	0.289	-11.8019	99.8019	-1.7076
■	13	Sea	76	0.276	-10.0562	86.0562	-1.2053
■	14	Oak	74	0.257	8.0338	65.9662	0.8994
■	15	NewTeam	.	0.300	.	111.4330	.

The REG Procedure
Model: MODEL1
Dependent Variable: wins

Output Statistics

Obs	Dep Var wins	Predicted Value	Std Error Mean Predict	95% CL Mean		95% CL Predict		Residual
1	114.0000	98.7446	4.8733	88.1266	109.3626	76.8527	120.6364	15.2554
2	88.0000	75.4825	2.7089	69.5803	81.3847	55.4488	95.5161	12.5175
3	79.0000	82.8841	2.3888	77.6793	88.0889	63.0447	102.7234	-3.8841
4	92.0000	90.2856	3.2359	83.2352	97.3361	69.8841	110.6871	1.7144
5	63.0000	70.1957	3.5242	62.5171	77.8742	49.5687	90.8226	-7.1957
6	89.0000	81.8267	2.3554	76.6947	86.9587	62.0063	101.6471	7.1733
7	65.0000	73.3678	2.9966	66.8388	79.8967	53.1406	93.5949	-8.3678
8	80.0000	80.7693	2.3495	75.6503	85.8884	60.9523	100.5864	-0.7693
9	72.0000	72.3104	3.1616	65.4219	79.1989	51.9643	92.6565	-0.3104
10	70.0000	75.4825	2.7089	69.5803	81.3847	55.4488	95.5161	-5.4825
11	85.0000	81.8267	2.3554	76.6947	86.9587	62.0063	101.6471	3.1733
12	88.0000	99.8019	5.0986	88.6929	110.9109	77.6678	121.9361	-11.8019
13	76.0000	86.0562	2.6392	80.3059	91.8064	66.0667	106.0456	-10.0562
14	74.0000	65.9662	4.3399	56.5103	75.4221	44.6138	87.3186	8.0338
15	.	111.4330	7.7026	94.6504	128.2156	85.9738	136.8921	.

Sum of Residuals	0
Sum of Squared Residuals	926.46437
Predicted Residual SS (PRESS)	1522.13052