112 Final Exam Answer Sheet

NAME: ____________________

1. ____________________

2. ____________________

3. ____________________

4. ____________________

5(a). ____________________

5(b). ____________________

6(a).

Source	df	SS	MS	F
Treatments				
Error				
Total				

6(b). ____________________

7. ____________________

8. ____________________

22. ______________________________

23. ______________________________

24. ______________________________

25. ______________________________

26. ______________________________

27. ______________________________

28. ______________________________

29. ______________________________

30. Test statistic: ______________________________

Result: (Circle one) Reject H_0 Do not reject H_0

31. ______________________________

32. ______________________________

33. Circle one: Deterministic Probabilistic

Circle one: First-order Second-order

34. ______________________________

Directions: Calculators are allowed for this exam; no other books, notes, or assistance is allowed. **Put all answers on the answer sheet provided.** All numeric answers should be to three decimal places. Exam is conducted under the Tulane honor code.

For problems 1 and 2: The athletic director of a college decided to conduct an analysis of which students attended football games. She randomly selected 500 male and 400 female students. Of the male students, 365 said they had been to a football game within the past year and of the female students, 175 had attended a game. She would like to see if the proportion of males attending games differs from the proportion of females.

1 pt 1) Give the test statistic for the appropriate test.

1 pt 2) Suppose the $p-$value were .035 (not necessarily its actual value.) What would be the correct conclusion at $\alpha = .05$? (Choose one)

(a) The evidence is *insufficient* to prove the proportions of men and women attending football games differ.

(b) The evidence is *insufficient* to prove the proportions of men and women attending football games are the same.

(c) The evidence is *sufficient* to prove the proportions of men and women attending football games differ.

(d) The evidence is *sufficient* to prove the proportions of men and women attending football games are the same.

1 pt 3) Are Japanese managers more motivated than American managers? Random independent samples of each were administered and their scores on the SSATL exam, measuring motivation for upward mobility, are summarized below. Find the $p-$value of this test.

	American	Japanese
Sample size	211	100
Mean SSATL score	65.75	67.83
Standard deviation	11.07	6.41

For problems 7 and 8: Two new drugs, drug A and drug B, have been developed to treat glaucoma in dogs. Ten dogs of each of five different breeds of dog– a Yorkshire terrier, a Labrador retriever, a basset hound, an English bulldog, and a Great Dane– are chosen for this test. To test the drugs' effectiveness, a drop of drug A is placed in one eye of each dog (which eye it is placed in is chosen randomly) and a drop of drug B in the other. After putting in the drops, the researchers measure the eye pressure in each eye.

$\boxed{\text{1 pt}}$ 7) What is the response variable for this experiment? (Choose one)

(a) Type of dog

(b) Drugs A and B

(c) Eye pressure

$\boxed{\text{1 pt}}$ 8) Give an example of a treatment in this experiment.

For problems 9–13: In fitting a least squares line to $n = 11$ data points, the following quantities are computed: $SS_{xx} = 926$, $SS_{yy} = 538$, $SS_{xy} = -406$, $\overline{x} = 27$, and $\overline{y} = 29$.

$\boxed{\text{1 pt}}$ 9) Find the slope of the least squares line.

$\boxed{\text{1 pt}}$ 10) Find the y–intercept of the least squares line.

$\boxed{\text{1 pt}}$ 11) Calculate s^2.

$\boxed{\text{1 pt}}$ 12) Calculate r, the coefficient of determination.

$\boxed{\text{1 pt}}$ 13) Find the sum of errors SE for the least squares line.

For problems 17–19: Regression analysis is used to investigate the "survival size" y of nonprofit hospitals. Ten states are chosen at random and the survival size for all nonprofit hospitals for each state is recorded, as are the following data:

x_1 = the percentage of beds that are for-profit hospitals
x_2 = ratio of number of people in HMO's versus number with hospital insurance
x_3 = state population (in thousands)
x_4 = percent of state that is urban

A printout fitting the model $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$ to this data is given below:

Dep Variable: Y

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	4	246537.05939	61634.26485	28.180	0.0001
Error	15	32807.94061	2187.19604		
C Total	19	279345.00000			

Root MSE	46.76747	R-square	0.8826
Dep Mean	360.50000	Adj R-sq	0.8512
C.V.	12.97295		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	295.327091	40.17888737	7.350	0.0001
X1	1	-480.837576	150.39050364	-3.197	0.0060
X2	1	-829.464955	196.47303539	-4.222	0.0007
X3	1	0.007934	0.00355335	2.233	0.0412
X4	1	2.360769	0.76150774	3.100	0.0073

1 pt 17) Find the regression standard deviation s.

1 pt 18) Find the test statistic to test the overall utility of this model.

2 pts 19) It was hypothesized prior to this study that increases in the percentage of for-profit hospital beds would decrease the survival size of nonprofit hospitals. You want to test this at $\alpha = .05$.

(a) Find the test statistic for this test.

(b) Find the rejection region for this test.

For problems 21–23: Retail price data for $n = 60$ hard drives were recently extracted from a computer magazine. Three of the variables recorded for each hard drive were

y = retail price

x_1 = microprocessor speed (values in sample range from 10 to 40 megahertz)

x_2 = chip size (values in sample range from 286 to 486 CPU)

The data were used to fit a regression model; the printout is below:

Analysis of Variance

SOURCE	DF	SS	MS	F VALUE	PROB > F
MODEL	2	34593103.008	17296051.504	19.018	0.0001
ERROR	57	51840202.926	909477.24431		
C TOTAL	59	86432305.933			

ROOT MSE	953.66516	R-SQUARE	0.4002
DEP MEAN	3197.96667	ADJ R-SQ	0.3792
C.V.	29.82099		

Parameter Estimates

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR 0: PARAMETER = 0	PROB > |T|
INTERCEPT	1	-373.526392	1258.1242296	-0.297	0.7676
SPEED	1	104.838940	22.36298195	4.688	0.0001
CHIP	1	3.571850	3.89422935	0.917	0.3629

OBS	SPEED	CHI	Dep Var PRICE	Predict Value	Std Err Predict	Lower 95% Predict	Upper 95 Predic	Residual
1	33	38	5099.0	4464.9	260.768	3942.7	4987.	634.1

1 pt 21) What assumptions are needed for any inferences derived from the regression analysis to be valid? (Choose one)

(a) Normally distributed retail prices.

(b) Independent random samples taken from populations with equal variances.

(c) Independent, normally distributed random errors with mean 0 and constant variance.

(d) None, because the sample size exceeds 30.

For problems 24 and 25: Teenage alcoholism is a problem in the United States. To discover why teenagers are turning to alcohol, a survey was conducted to find out if the family's status has any relationship to the teen's alcohol consumption. A random sample of 200 teenagers were questioned, with results summarized below:

		Alcohol	
	None	Occasional	Frequent
Upper class	4	16	10
Upper middle class	11	40	24
Lower middle class	9	47	9
Lower class	6	17	7

1 pt 24) If alcohol consumption and family status are independent, find the expected number of upper class teens who occasionally drink.

1 pt 25) In this problem, not all assumptions have been satisfied to run a valid hypothesis test. Specifically, what assumption is violated?

(Test Bank 13.22)

1 pt 26) A national publication has determined that 10% of all U.S. colleges are "high" status, 20% are "moderate" status, and 70% are "low" status. A researcher wants to study the relationship between a school's status and its faculty's publishing activity. She theorizes that if publishing activity does not depend on school status, the number of authors in each status category should be the same as the percentage of schools in that category. Of 200 papers randomly selected, 25 were from high status schools, 30 from moderate status schools, and 145 from low status schools. Find the test statistic used to test her claim.

1 pt 27) Weevils cause tremendous crop damage each year. Three chemicals (A, B, and C) designed to control weevil populations were applied, one to each of three cotton fields. After 3 months, patches were taken from each field and the percentage of cotton plants with weevil damage was recorded for each. Calculate the test statistic for the Kruskal-Wallis H-test using the data below:

A		B		C	
10.8	9.8	22.3	20.4	9.8	10.8
15.6	16.7	19.5	23.6	12.3	12.2
19.2	19.0	18.6	21.2	16.2	17.3

1 pt 31) In working an ANOVA problem you have 4 samples with 7 items per sample. How many degrees of freedom should be used for the denominator of the test statistic?

1 pt 32) Which of the following best describes SSE? (Choose one)

(a) It is a measure of the variance *between* the treatments.

(b) It is a measure of the variance *within* the treatments.

(c) It is a measure of the total variance.

(d) It is a measure of the reliability of the hypothesis test.

1 pt 33) The model $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon$ is: (Choose one from each pair)

A (deterministic/probabilistic) model that is (first-order/second-order.)

1 pt 34) Which test can be thought of as a non-parametric version of a paired differences test? (Choose one)

(a) Sign test.

(b) Wilcoxon rank sum test.

(c) Wilcoxon signed rank test.

(d) Kruskal-Wallis H-test.

Tables

TABLE XII Critical Values of T_L and T_U for the Wilcoxon Rank Sum Test: Independent Samples

Test statistic is the rank sum associated with the smaller sample (if equal sample sizes, either rank sum can be used).

a. α = .025 one-tailed; α = .05 two-tailed

n_2 \ n_1	3		4		5		6		7		8		9		10	
	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U
3	5	16	6	18	6	21	7	23	7	26	8	28	8	31	9	33
4	6	18	11	25	12	28	12	32	13	35	14	38	15	41	16	44
5	6	21	12	28	18	37	19	41	20	45	21	49	22	53	24	56
6	7	23	12	32	19	41	26	52	28	56	29	61	31	65	32	70
7	7	26	13	35	20	45	28	56	37	68	39	73	41	78	43	83
8	8	28	14	38	21	49	29	61	39	73	49	87	51	93	54	98
9	8	31	15	41	22	53	31	65	41	78	51	93	63	108	66	114
10	9	33	16	44	24	56	32	70	43	83	54	98	66	114	79	131

b. α = .05 one-tailed; α = .10 two-tailed

n_2 \ n_1	3		4		5		6		7		8		9		10	
	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U
3	6	15	7	17	7	20	8	22	9	24	9	27	10	29	11	31
4	7	17	12	24	13	27	14	30	15	33	16	36	17	39	18	42
5	7	20	13	27	19	36	20	40	22	43	24	46	25	50	26	54
6	8	22	14	30	20	40	28	50	30	54	32	58	33	63	35	67
7	9	24	15	33	22	43	30	54	39	66	41	71	43	76	46	80
8	9	27	16	36	24	46	32	58	41	71	52	84	54	90	57	95
9	10	29	17	39	25	50	33	63	43	76	54	90	66	105	69	111
10	11	31	18	42	26	54	35	67	46	80	57	95	69	111	83	127

Source: From F. Wilcoxon and R. A. Wilcox, "Some Rapid Approximate Statistical Procedures," 1964, 20–23.

TABLE XIV Critical Values of Spearman's Rank Correlation Coefficient

The α values correspond to a one-tailed test of H_0: $\rho = 0$. The value should be doubled for two-tailed tests.

n	α = .05	α = .025	α = .01	α = .005	n	α = .05	α = .025	α = .01	α = .005
5	.900		—	—	18	.399	.476	.564	.625
6	.829	.886	.943	—	19	.388	.462	.549	.608
7	.714	.786	.893	—	20	.377	.450	.534	.591
8	.643	.738	.833	.881	21	.368	.438	.521	.576
9	.600	.683	.783	.833	22	.359	.428	.508	.562
10	.564	.648	.745	.794	23	.351	.418	.496	.549
11	.523	.623	.736	.818	24	.343	.409	.485	.537
12	.497	.591	.703	.780	25	.336	.400	.475	.526
13	.475	.566	.673	.745	26	.329	.392	.465	.515
14	.457	.545	.646	.716	27	.323	.385	.456	.505
15	.441	.525	.623	.689	28	.317	.377	.448	.496
16	.425	.507	.601	.666	29	.311	.370	.440	.487
17	.412	.490	.582	.645	30	.305	.364	.432	.478

Source: From E. G. Olds, "Distribution of Sums of Squares of Rank Differences for Small Samples," *Annals of Mathematical Statistics*, 1938, 9.

TABLE XIII Critical Values of T_0 in the Wilcoxon Paired Difference Signed Ranks Test

One-Tailed	Two-Tailed	$n = 5$	$n = 6$	$n = 7$	$n = 8$	$n = 9$	$n = 10$
$\alpha = .05$	$\alpha = .10$	1	2	4	6	8	11
$\alpha = .025$	$\alpha = .05$		1	2	4	6	8
$\alpha = .01$	$\alpha = .02$			0	2	3	5
$\alpha = .005$	$\alpha = .01$				0	2	3
		$n = 11$	$n = 12$	$n = 13$	$n = 14$	$n = 15$	$n = 16$
$\alpha = .05$	$\alpha = .10$	14	17	21	26	30	36
$\alpha = .025$	$\alpha = .05$	11	14	17	21	25	30
$\alpha = .01$	$\alpha = .02$	7	10	13	16	20	24
$\alpha = .005$	$\alpha = .01$	5	7	10	13	16	19
		$n = 17$	$n = 18$	$n = 19$	$n = 20$	$n = 21$	$n = 22$
$\alpha = .05$	$\alpha = .10$	41	47	54	60	68	75
$\alpha = .025$	$\alpha = .05$	35	40	46	52	59	66
$\alpha = .01$	$\alpha = .02$	28	33	38	43	49	56
$\alpha = .005$	$\alpha = .01$	23	28	32	37	43	49
		$n = 23$	$n = 24$	$n = 25$	$n = 26$	$n = 27$	$n = 28$
$\alpha = .05$	$\alpha = .10$	83	92	101	110	120	130
$\alpha = .025$	$\alpha = .05$	73	81	90	98	107	117
$\alpha = .01$	$\alpha = .02$	62	69	77	85	93	102
$\alpha = .005$	$\alpha = .01$	55	61	68	76	84	92
		$n = 29$	$n = 30$	$n = 31$	$n = 32$	$n = 33$	$n = 34$
$\alpha = .05$	$\alpha = .10$	141	152	163	175	188	201
$\alpha = .025$	$\alpha = .05$	127	137	148	159	171	183
$\alpha = .01$	$\alpha = .02$	111	120	130	141	151	162
$\alpha = .005$	$\alpha = .01$	100	109	118	128	138	149
		$n = 35$	$n = 36$	$n = 37$	$n = 38$	$n = 39$	
$\alpha = .05$	$\alpha = .10$	214	228	242	256	271	
$\alpha = .025$	$\alpha = .05$	195	208	222	235	250	
$\alpha = .01$	$\alpha = .02$	174	186	198	211	224	
$\alpha = .005$	$\alpha = .01$	160	171	183	195	208	
		$n = 40$	$n = 41$	$n = 42$	$n = 43$	$n = 44$	$n = 45$
$\alpha = .05$	$\alpha = .10$	287	303	319	336	353	371
$\alpha = .025$	$\alpha = .05$	264	279	295	311	327	344
$\alpha = .01$	$\alpha = .02$	238	252	267	281	297	313
$\alpha = .005$	$\alpha = .01$	221	234	248	262	277	292
		$n = 46$	$n = 47$	$n = 48$	$n = 49$	$n = 50$	
$\alpha = .05$	$\alpha = .10$	389	408	427	446	466	
$\alpha = .025$	$\alpha = .05$	361	379	397	415	434	
$\alpha = .01$	$\alpha = .02$	329	345	362	380	398	
$\alpha = .005$	$\alpha = .01$	307	323	339	356	373	

Source: From F. Wilcoxon and R. A. Wilcox, "Some Rapid Approximate Statistical Procedures," 1964, p. 28.