STP 231 Statistics

Print Name:_____

Test 4

Honor Statement:

By signing below you confirm that you have neither given nor received any unauthorized assistance on this exam. This includes any use of a graphing calculator beyond those uses specifically authorized by the Mathematics Department and your instructor. Furthermore, you agree not to discuss this exam with anyone until the exam testing period is over. In addition, your calculator's memory and menus may be checked at any time and cleared by any testing center proctor or Mathematics Department instructor.

_____ _____

Signature **Date**

Instructions:

- The exam is worth a total of 100 points; please make sure your exam has all pages before you begin. Problems 1-3 are marked. Problems 4-7 are worth 6 points each and problems 8-12 are worth 4 points each. Make sure that you have all 10 pages to your exam. If you do not have 11 pages, then contact the Instructor administering the exam.
- Show all work in detail or your answer will not receive any credit. Include appropriate units on all questions that apply. Write neatly and box all answers.
- Please ask the student worker at the check out desk if you need scratch paper.
- **No calculators or computers that do symbolic algebra, like the Casio FX-2, TI-89, or TI- 92, may be used.**
- The formulas and tables are shown on the last page. You may take these pages off to help you during the exam
- **Part 1 of the exam is for free response. Show your work or explain the process for each problem to receive credit.**
- Part 2 is the multiple choice part of the exam. A table is given on page three for you to write in the letter for the correct answer. You may take the first page off to help you write the correct letter in for each answer.
- You may have one 4 by 6 note card. It must not contain any worked out problems. It must be attached to the back of your test when the exam is turned in.

1.(20) In the article "Television Viewing and Physical Fitness in Adults," the author wishes to determine whether time spent watching television is associated with cardiovascular fitness. Subjects were asked about their daily television viewing time and were classified as either physically fit or not physically fit depending upon how they performed on a step test. The data is shown in the table below and the expected frequencies are in parenthesis.

Television Viewing and Fitness Level

TV Viewing Time in hours	Physically Fit	Not Physically Fit	**Total**
0	35 (25.5)	147(156.5)	**182**
1-2	101 (102.2)	629 (627.8)	**730**
3-4	28 (35.0)	222 (215.0)	**250**
5 or more	4 (5.3)	34 (32.7)	**38**
Total	**168**	**1032**	**1200**

Let the significance level be 0.01, from this study is there an association between fitness level and tv viewing time. State the null and alternative hypothesis. Calculate the test statistic and p-value. State your conclusion in context of the problem.

H_0: there is no association between fitness level and t.v. viewing time

H_A: there is an association between fitness level and t.v. viewing time.

$$\chi^2_S = \frac{(35-25.5)^2}{25.5} + \frac{(147-156.5)^2}{156.5} + \frac{(101-102.2)^2}{102.2} + \frac{(629-627.8)^2}{627.8} +$$

$$\frac{(28-35)^2}{35} + \frac{(222-215)^2}{215} + \frac{(4-5.3)^2}{5.3} + \frac{(34-32.7)^2}{32.7}$$

$$\chi^2_S = 3.539 + 0.577 + 0.0141 + 0.0023 + 1.4 + 0.228 + 0.319 + 0.0517 = 6.131$$

$DF = (4-1)(2-1) = 3$

$\chi^2_{0.9} = 0.584 \qquad \chi^2_{0.1} = 6.251$

$.1 < p\text{-value} < .9 \not< .01$; therefore do not reject H_0

With a p-value between (0.1, 0.9), there is not sufficient evidence to support that there is an association between t.v. viewing time and physical fitness.

2. (15) From a General Social Survey married respondents were asked to characterize their feelings about being married. Out of 366 males, 242 labeled themselves as "very happy." Out of 423 females 257 labeled themselves as "very happy". Construct a 95% confidence interval for the difference between the two proportions p_1 representing the proportion of men and p_2 representing the proportion of women. Interpret your conclusion in context of the problem.

$$\tilde{p}_1 = \frac{242+1}{366+2} \; ; \; \tilde{p}_2 = \frac{257+1}{423+2}$$

$$\tilde{p}_1 = 0.6603 \quad \tilde{p}_2 = 0.60706$$

$$SE(\tilde{p}_1 - \tilde{p}_2) = \sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{n_1+2} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{n_2+1}}$$

$$= 0.03422$$

$$(\tilde{p}_1 - \tilde{p}_2) \pm 1.96 \times SE(\tilde{p}_1 - \tilde{p}_2) =$$

$$(0.6603 - 0.60706) \pm 1.96(0.03422)$$

$$0.053267 \pm 0.06706$$

$$[-0.0138, 0.1203]$$

The difference in the proportion of married men that label themselves as "very happy" and the proportion of married women that label themselves as "very happy" is as small as -0.0138 and as large as 0.1203 with 95% confidence.

3. (20) Tests for adverse reactions to a new drug yielded the results given in the table. At the 0.05 significance level, does the data provide sufficient evidence to conclude that $\Pr(\text{Headache} \mid \text{Drug})$ is greater than the $\Pr(\text{Headache} \mid \text{Placebo})$ State the null and alternative hypotheses. Calculate the test statistic and p-value. State your conclusion in context of the problem.

Reaction	~~Drug~~ Placebo	~~Placebo~~ Drug	Total
Headache	2 (5.022)	9 (5.978)	11
No Headache	82 (78.978)	91 (94.022)	173
Total	84	100	184

H_0: The probability of having a headache is the same for those that took the drug and the placebo

H_A: The probability of having a headache is greater for those that took the drug

$$\chi^2_S = \frac{(2-5.022)^2}{5.022} + \frac{(9-5.978)^2}{5.978} + \frac{(82-78.978)^2}{78.978} + \frac{(91-94.022)^2}{94.022}$$

$$= 1.8184 + 1.5277 + 0.1156 + 0.0971$$

$$\chi^2 = 3.559$$

$df = 1$

$$\frac{0.05}{2} < \text{p-value} < \frac{0.10}{2}$$

$0.025 < \text{p-value} < 0.05$ $\boxed{<}$ $\alpha = 0.05$

$\therefore$ reject H_0

With a p-value between 0.025 and 0.05, there is sufficient evidence to support that the probability of having a headache given the drug is greater than the probability of a headache given the placebo

For Part II – Multiple Choice: Circle your answer choice on the exam **AND** fill in the answer with the letter of the answer that you believe is the correct answer.

Problem Number	Letter of Answer	Problem Number	Letter of Answer	Problem Number	Letter of Answer
4		7		10.	
5.		8.		11.	
6.		9.		12.	

4. The data below shows the temperature (X) and the amount a plant grew(Y) in millimeters, for each of nine randomly selected days. Calculate the coefficient of determination. From the value of the coefficient of determination, is the proportion of variance in the amount that the plant grew explained by the linear relationship between the temperature and the amount that the plant grew?

$\Sigma(Z_x \cdot Z_Y) = 1.564$

i	X	Y	$Z_x = \frac{x-\bar{x}}{s_x}$	$Z_Y = \frac{y-\bar{y}}{s_Y}$	$Z_x \cdot Z_Y$
1	62	36	0.23	0.62	0.14
2	76	39	1.27	0.83	1.05
3	50	50	-0.66	1.59	-1.05
4	51	13	-0.59	-0.96	0.57
5	71	33	0.90	0.41	0.37
6	46	33	-0.96	0.41	-0.40
7	51	17	-0.59	-0.69	0.40
8	44	6	-1.11	-1.45	1.60
9	79	16	1.50	-0.76	-1.13

$r^2 = r$

$r^2 = \left[\frac{1}{n-1} \cdot \Sigma(z_x \cdot z_y)\right]^2$

$r^2 = \left(\frac{1}{8} \cdot 1.564\right)^2$

$r^2 = 0.0382$

A. 0.196; Yes B. 0.038; Yes C. 0.196; No

D. 0.038; No E. Not enough information

5. Two separate tests X and Y are designed to measure a student's ability to solve problems. Several students are randomly selected to take both tests and their results are shown below: Determine the regression equation for this data set. If
$\bar{x} = 48.667, s_x = 6.819$, $\bar{y} = 64.667$, $s_y = 7.314$, $n = 9$ and $r = 0.8671$

i	X	Y
1	48	73
2	52	67
3	58	73
4	44	59
5	43	58
6	43	56
7	40	58
8	51	64
9	59	74

$b_1 = r \cdot \frac{s_y}{s_x}$

$= 0.8671 \cdot \frac{7.314}{6.819}$

$b_1 = 0.9300$

$b_0 = \bar{y} - b_1 \bar{x}$

$b_0 = 64.667 - 0.9300 \cdot 48.667$

$b_0 = 19.404$

(A.) $\hat{y} = 0.93x + 19.40$ B. $\hat{y} = 19.4x + 0.93$ C. $\hat{y} = -0.63x + 67.83$
D. $\hat{y} = 0.067x + 48.33$ E. Not enough information

6. The data below is a random sample from a MAT 117 class that shows the percentage of class attendance (X), the final grade in the course (Y) and the predicted grade in the course $\hat{Y}$. Calculate the residual standard deviation and determine the percentage of observations that are within $\pm 1 \cdot s_e$ The Sum of Squares Residuals, SS(resid), is 869.1720

i	X	Y	$\hat{Y}$	$Y - \hat{Y}$	
1	50	65	60.75	4.25	✓
2	100	100	106.83	-6.83	✓
3	75	90	83.79	6.21	✓
4	55	62	65.36	-3.36	✓
5	0	6	14.68	-8.68	✓
6	50	78	60.75	17.25	X
7	75	76	83.79	-7.79	✓
8	55	78	65.36	12.64	X
9	60	72	69.97	2.03	✓
10	60	58	69.97	-11.97	X
11	50	57	60.75	-3.75	✓

$s_e = \sqrt{\frac{SS_{resid}}{n-2}}$

$= \sqrt{\frac{869.1720}{9}}$

$= 9.827$

$\frac{8}{11} \cdot 100\% = 72.7\%$

(A.) 9.83; 72.7 % B. 9.83; 81.8 % C. 9.32; 81.8%
D. 9.32; 72.7 % E. 8.89, 72.7- % F. 8.89, 81.8%

7. The data below shows the temperature (X) and the amount a plant grew(Y) in millimeters, for each of nine randomly selected days. Calculate the correlation coefficient, r. Perform the hypothesis test to determine if $\rho = 0$ and determine the p-value. If $\hat{y} = 1.905x + 5.055$, and the $\Sigma(Z_x \cdot Z_Y) = 8.875$

i	X	Y	$Z_x = \frac{x - \bar{x}}{s_x}$	$Z_Y = \frac{y - \bar{y}}{s_Y}$	$Z_x \cdot Z_Y$
1	23	49	-0.76	-0.74	0.56
2	25	53	-0.37	-0.33	0.12
3	28	59	0.21	0.27	0.06
4	21	42	-1.15	-1.44	1.66
5	21	47	-1.15	-0.94	1.08
6	25	53	-0.37	-0.33	0.12
7	26	55	-0.18	-0.13	0.02
8	30	63	0.60	0.68	0.41
9	34	67	1.38	1.08	1.49
10	36	75	1.77	1.89	3.35

A. 0.0.972; p-value < 0.02

B. 0.972; p-value < 0.01

C. Not enough information

D. 0.986; p-value < 0.02

E. 0.986; p-value < 0.01

8. Suppose that you wish to perform a chi-square test of independence. The two variables under consideration are sex and blood type. True or false, if the two variables are not associated, we would expect that the proportion of women in the sample with a given blood type would be roughly equal to the proportion of men in the sample with the same blood type.

A. True B. False C. Not enough information to determine

9. Which of the following are not conditions for performing a chi-square test?
   A. All expected frequencies are 1 or greater.
   B. Samples are random
   C. At most 20% of the expected frequencies are greater than 5.
   D. Samples are independent

10. When conducting a chi-square test, the p-value exceeds the specified significance level. What can be concluded?
   A. No conclusion can be made regarding H_O at the given significance level.
   B. Do not reject H_O at the given significance level.
   C. Reject H_O at the given significance level.

11. Which of the following would be true of scatterplots with a correlation of 0.85.

A. The points would fall generally around a line with a positive slope.
B. The points would not have any discernible form.
C. The points would fall generally around a line with a negative slope
D. The points would have less strength than another scatterplot with a correlation of 0.50

12. Suppose that people who smoke marijuana at a higher rate tend to measure with a lower IQ. Would the relationship between rate of smoking marijuana and IQ be positively or negatively correlated?

A. positively correlated B. negatively correlated

Formula Sheet

95% Confidence Interval for the difference between two probabilities

$$\tilde{p} = \frac{y+1}{n+2} \qquad SE_{(\tilde{p}_1 - \tilde{p}_2)} = \sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{n_1+2} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{n_2+2}} \qquad (\tilde{p}_1 - \tilde{p}_2) \pm 1.96 \times SE_{(\tilde{p}_1 - \tilde{p}_2)}$$

Chi-Square Test

$$E = \frac{(Row\ total)\cdot(Column\ total)}{Grand\ Total}, \qquad df = (r\text{-}1)(k-1)$$

$$\chi_s^2 = \sum \frac{(o_i - e_i)^2}{e_i}$$

Correlation Coefficient

$$r = \frac{1}{n-1}\sum\left(\frac{x-\bar{x}}{s_x}\right)\left(\frac{y-\bar{y}}{s_y}\right) = \frac{1}{n-1}\sum(Z_X \cdot Z_Y) \qquad t_s = r\sqrt{\frac{n-2}{1-r^2}} \quad \text{with } df = n\text{-}2$$

The Coefficient of Determination is approximately $r^2 \approx \frac{s_y^2 - s_e^2}{s_y^2} = 1 - \frac{s_e^2}{s_y^2}$

The Coefficient of Determination is exactly $r^2 = (r)^2$

The Fitted Regression line

$$\hat{y} = b_1 x + b_0 \qquad b_1 = r \times \left(\frac{s_y}{s_x}\right) \qquad b_0 = \bar{y} - b_1\bar{x} \qquad e_i = y_i - \hat{y}_i$$

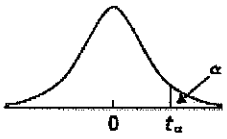
$$SS(resid) = \sum_{i=1}^{n}(e_i)^2 = \sum_{i=1}^{n}(y_i - \hat{y})^2 \qquad s_e = \sqrt{\frac{SS(resid)}{n-2}}$$

Inferrence

$$SEb_1 = \frac{s_e}{s_x\sqrt{n-1}} \qquad b_1 \pm t_{0.025}SEb_1$$

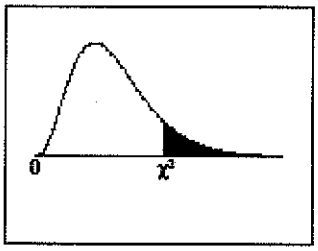
$$t_s = \frac{b_1}{SEb_1} = r\sqrt{\frac{n-2}{1-r^2}} \text{ with } df = n\text{-}2$$

TABLE IV
Values of t_α



df	$t_{0.10}$	$t_{0.05}$	$t_{0.025}$	$t_{0.01}$	$t_{0.005}$	df
1	3.078	6.314	12.706	31.821	63.657	1
2	1.886	2.920	4.303	6.965	9.925	2
3	1.638	2.353	3.182	4.541	5.841	3
4	1.533	2.132	2.776	3.747	4.604	4
5	1.476	2.015	2.571	3.365	4.032	5
6	1.440	1.943	2.447	3.143	3.707	6
7	1.415	1.895	2.365	2.998	3.499	7
8	1.397	1.860	2.306	2.896	3.355	8
9	1.383	1.833	2.262	2.821	3.250	9
10	1.372	1.812	2.228	2.764	3.169	10
11	1.363	1.796	2.201	2.718	3.106	11
12	1.356	1.782	2.179	2.681	3.055	12
13	1.350	1.771	2.160	2.650	3.012	13
14	1.345	1.761	2.145	2.624	2.977	14
15	1.341	1.753	2.131	2.602	2.947	15
16	1.337	1.746	2.120	2.583	2.921	16
17	1.333	1.740	2.110	2.567	2.898	17
18	1.330	1.734	2.101	2.552	2.878	18
19	1.328	1.729	2.093	2.539	2.861	19
20	1.325	1.725	2.086	2.528	2.845	20
21	1.323	1.721	2.080	2.518	2.831	21
22	1.321	1.717	2.074	2.508	2.819	22
23	1.319	1.714	2.069	2.500	2.807	23
24	1.318	1.711	2.064	2.492	2.797	24
25	1.316	1.708	2.060	2.485	2.787	25
26	1.315	1.706	2.056	2.479	2.779	26
27	1.314	1.703	2.052	2.473	2.771	27
28	1.313	1.701	2.048	2.467	2.763	28
29	1.311	1.699	2.045	2.462	2.756	29
30	1.310	1.697	2.042	2.457	2.750	30
31	1.309	1.696	2.040	2.453	2.744	31
32	1.309	1.694	2.037	2.449	2.738	32
33	1.308	1.692	2.035	2.445	2.733	33
34	1.307	1.691	2.032	2.441	2.728	34
35	1.306	1.690	2.030	2.438	2.724	35
36	1.306	1.688	2.028	2.434	2.719	36
37	1.305	1.687	2.026	2.431	2.715	37
38	1.304	1.686	2.024	2.429	2.712	38
39	1.304	1.685	2.023	2.426	2.708	39
40	1.303	1.684	2.021	2.423	2.704	40
41	1.303	1.683	2.020	2.421	2.701	41
42	1.302	1.682	2.018	2.418	2.698	42
43	1.302	1.681	2.017	2.416	2.695	43
44	1.301	1.680	2.015	2.414	2.692	44
45	1.301	1.679	2.014	2.412	2.690	45
46	1.300	1.679	2.013	2.410	2.687	46
47	1.300	1.678	2.012	2.408	2.685	47
48	1.299	1.677	2.011	2.407	2.682	48
49	1.299	1.677	2.010	2.405	2.680	49

Chi-Square Distribution Table



The shaded area is equal to α for $\chi^2 = \chi^2_\alpha$.

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169