Stat 2470, Final Exam, Fall 2014 Name KEY

Instructions: Show all work. Use exact answers or appropriate rounding conventions. If you use your calculator, you can show work by saying which calculator commands you used.

1. A simple random sample of individuals who drive alone to work in a large metropolitan area was obtained and each individual was categorized with respect to both size of car and commuting distance. Does the accompanying data commuting distance and size of car are related in the population sampled? State the appropriate hypothesis and use an α level of 0.05 χ^2 test. (12 points)

		Commuting Distance		
		(0,10)	(10,20)	20+
Size of car	Subcompact	6	27	19
	Compact	8	36	17
	Midsize	21	45	33
	Full-size	14	18	6

into matrix A

Ho: commuting distance does not affect size of car purchased

Ha: it does

$\chi^2 = 14.1583...$

$p = .0279 < .05$

reject Ho

there does seem to be good evidence that commuting distance affects size of car purchased.

2. Sorghum is an important cereal crop whose quality and appearance could be affected by the presence of pigments in the seed shell. Genetic theory predicts a particular cross of sorghum varieties should produce seed colors in the ratio 9:3:4 in the generation studied. Does the data below contradict genetic theory? Perform a χ^2 test at the level 0.05. (14 points)

Seed Color	**Red**	**Yellow**	**White**	
Observed Frequency	195	73	100	368
	$\frac{9}{16} * 368 =$ 207	$\frac{3}{16} * 368 =$ 69	$\frac{4}{16} * 368 =$ 92	

$$\chi^2 = \sum \frac{(obs - exp)^2}{exp} = \frac{(207-195)^2}{207} + \frac{(69-73)^2}{69} + \frac{(100-92)^2}{92} = 1.623$$

$\chi^2cdf(1.623, E99, 2) = .4419... > .05$

Ho: fits the 9:3:4 ratio

Ha: fails to fit predicted ratio

fail to reject Ho

3. Let y be life of a bearing, x_1 be oil viscosity and x_2 is the load. Suppose that a multiple regression model relating life to viscosity and load is $y = 125.0 + 7.75x_1 + 0.0950x_2 - 0.0090x_1x_2 + \epsilon$.
   a. What is the mean lifetime when viscosity is 40 and load is 1100? (5 points)

$$y = 125.0 + 7.75(40) + 0.0950(1100) - 0.0090(40)(1100) =$$

$$143.50$$

   b. When viscosity is 30, what is the change in mean life associated with an increase of 1 in load? When viscosity is 40, what is the change in mean life associated with an increase of 1 in load? (5 points)

$$y = 125.0 + 7.75(30) + .0950(x_2) - .0090(30)(x_2) \Rightarrow$$
$$y = 125.0 + 7.75(40) + .0950(x_2) - .0090(40)(x_2)$$
$$y = 357.5 - .175x_2$$
$$y = 435 - .265x_2$$

decreased by −0.175 per load unit
" by −.265 per load unit

4. The accompanying data was obtained from a study of a certain method for preparing pure alcohol from refinery streams. The independent variable x is volume hourly space velocity, and the dependent variable y is the amount of conversion of iso-butylene.

x	1	1	2	3	4	4	6
y	23.0	24.5	28.0	30.9	32.0	33.6	20.0

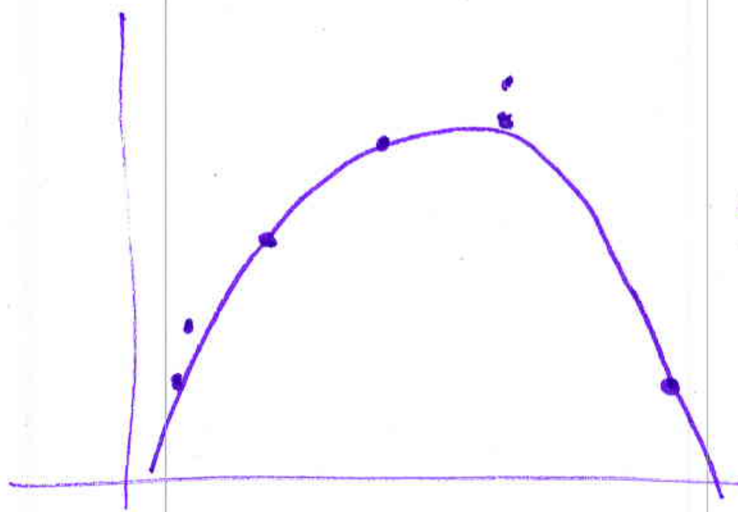
   a. Find a quadratic model to estimate the regression function. (9 points)

QuadReg → $-1.675x^2 + 11.186x + 13.733$

   b. What is the R^2 value? (7 points)

0.9389...

c. Plot the data and the regression equation on the same graph. Does the equation appear to be a good fit for the data? (9 points)



pretty well
or last value
Could be an outlier
making it appear
more quadratic than
otherwise.

5. Wear resistance of certain nuclear reactor components is partly determined by properties of the oxide layer. The following data appears in an article that proposed a new nondestructive testing method to monitor thickness of the layer. The variable x is oxide layer thickness in micrometers, and y is the eddy-current response.

x	0	7	17	114	133	142	190	218	237	285
y	20.3	19.8	19.5	15.9	15.1	14.7	11.9	11.5	8.3	6.6

a. Calculate the linear regression equation for the model. How appropriate is a simple linear regression model for this data? (8 points)

$y = -.0467x + 20.634$

it's appropriate. graph looks good and correlation is high

b. Construct an appropriate hypothesis test on the value of β_1 at the level 0.10. (7 points)

$H_0: \beta_1 = 0$

$H_a: \beta_1 \neq 0$

LinRegTTest
L1, L2
$\beta \neq 0$

$\Rightarrow$ $t = -17.875$

$p = 9.82955 \times 10^{-8}$

much less than 10% so we reject Ho.

c. What is the value of the correlation for this model? (5 points)

$r = -.9877$

d. What proportion of the data can be explained by the linear relationship between the two variables? (5 points)

97.557% (r^2)

e. Find the prediction interval for y when $x = 200$. (8 points) 95%

$y(200) = 11.29$

$S_{xx} = (101.9)^2$

$S = .792$

$11.29 \pm 2.26 * .792 \sqrt{\frac{1}{10} + \frac{(200-139.3)^2}{91809}} =$

$11.29 \pm .263 = (11.03\ 11.55)$

6. A manufacturer of nickel-hydrogen batteries randomly selects 100 nickel plates for test cells, cycles them a specified number of times, and determines that 14 of the plates have blisters. Does this provide compelling evidence for concluding that more than 10% of all plates blister under such circumstances? State and test the appropriate hypothesis test. (14 points)

1PropZTest

$H_0: p = 10\%$

$H_a: p > 10\%$

$p_0 = .10$

$x = 14$

$n = 100$

$p > p_0$

$\rightarrow$ $z = 1.33$, $p = .0912$

fail to reject Ho

7. It is thought that the front cover and nature of the first question on mail surveys influence the response rate. An article tested the theory by experimenting with different cover designs. One cover was plain, and one used the image of a skydiver. The researchers speculated that the return rate would be lower for the plain cover. (14 points)

Cover	**Number Sent**	**Number Returned**
Plain	307	156
Skydiver	388	175

Does this data support the hypothesis of the researchers? Test the relevant hypotheses using $\alpha = 0.1$. State the hypotheses, P-value, and conclusion in context of the problem.

2 propZtest

$X_1 = 156$ $n_1 = 307$

$X_2 = 175$ $n_2 = 388$

$p_1 < p_2$

$H_0: p_1 = p_2$

$H_a: p_1 < p_2$

$\Rightarrow$ $z = 1.49$

$p = .9328$

fail to reject Ho

these # actually suggest plain has a higher return rate

8. The alternating current (AC) breakdown voltage for an insulating liquid indicates its dielectric strength. An article gave the following sample observations for a particular circuit under certain conditions specified in the article.

62	50	53	57	41	53	55	61	59	64
50	53	64	62	50	68	54	55	57	50
55	50	56	55	46	55	53	54	52	47

a. Calculate a 95% confidence interval for the breakdown voltage. (7 points)

TInterval

Data, L_1

C-level: .95

$\Rightarrow$ (52.529, 56.871)

$\bar{x} = 54.7$

b. What sample size would be required for the 95% confidence interval to have a width of 4 kV (so that μ is estimated to within 2 kV)? (7 points)

$$n = \left(2 * 1.96 \cdot \frac{5.814}{4}\right)^2 = 32.46$$

$\Rightarrow$ $n = 33$ for z-interval

(a bit more for t-interval)

$n = 35$ works experimentally

9. Let X denote the proportion of allotted time that a randomly selected student spends working on a certain aptitude test, with a probability distribution function given by $f(x;\lambda) = \lambda e^{-\lambda x}, x \geq 0$. Suppose that 6 students have the following times: $\{0.9, 0.8, 0.9, 0.7, 0.9, 0.5\}$. Use this information to estimate the value of θ λ using the Maximum Likelihood Estimator function. (20 points)

$L(\lambda) =$

$\lambda e^{-.9\lambda} \cdot \lambda e^{-.8\lambda} \cdot \lambda e^{-.9\lambda} \cdot \lambda e^{-.7\lambda} \cdot \lambda e^{-.9\lambda} \cdot \lambda e^{-.5\lambda}$

$= \lambda^6 e^{-4.7\lambda}$

$L'(\lambda) = 6\lambda^5 e^{-4.7\lambda} - 4.7\lambda^6 e^{-4.7\lambda} = \lambda^5 e^{-4.7\lambda}[6 - 4.7\lambda] = 0$

$\_\_\_ = \lambda \Rightarrow$

10. Consider the discrete joint probability distribution shown below.

	$p(x,y)$	y = 0	1	2
	0	.08	.04	.02
	1	.06	.10	.06
x	2	.05	.12	.08
	3	.04	.07	.10
	4	.01	.05	.12

a. Find $P(X = 4, Y = 2)$. (3 points)

.12

b. Find p_X and p_Y. (10 points)

X	0	1	2	3	4
p(x)	.14	.22	.25	.21	.18

Y	0	1	2
P(y)	.24	.38	.38

c. Use your result in (b) to find $E(X)$ and $E(Y)$. (10 points)

$E(X) = 0(.14) + 1(.22) + 2(.25) + 3(.21) + 4(.18) = 2.07$

$E(Y) = 0(.24) + 1(.38) + 2(.38) = 1.14$

11. Suppose that $f(x,y) = \begin{cases} kx^3y^8, & 0 \le x \le 2, 0 \le y \le 2 \\ 0, & elsewhere \end{cases}$.

a. Find the value of k that makes this a valid joint probability distribution. (8 points)

$$\int_0^2 \int_0^2 kx^3y^8 \, dx\,dy = k\int_0^2 \frac{x^4}{4} y^8 \Big|_0^2 dy = \frac{4k}{4}\int_0^2 y^8 dy =$$

$$\frac{4k}{9} y^9 \Big|_0^2 = \frac{2048}{9} k = 1 \qquad k = \frac{9}{2048}$$

b. Use the result in (a) to compute $f_{Y|X}(y|x) = \frac{f(x,y)}{f(x)}$. (12 points)

$$\frac{9}{2048}\int_0^2 x^3 y^8 dy = \frac{9}{2048} \cdot \frac{y^9}{9}\Big|_0^2 x^3 = \frac{1}{4}x^3 = f(x)$$

$$f_{Y|X}(y|x) = \frac{\frac{9}{2048}x^3y^8}{\frac{1}{4}x^3} = \frac{9}{512}y^8$$

Since this depends only on y X, & Y are independent

12. The number of requests for assistance received by a towing service is a Poisson process with a rate of four per hour.

a. Compute the probability that exactly ten requests are received during a particular two-hour period. (6 points)

$\mu = 8$

poissonpdf (8, 10) = .09926

b. If the operators of the towing service take a 30-min break for lunch, with is the probability that they don't miss any calls for assistance? (6 points)

$\mu = 2$

poissonpdf (2, 0) = .13533

13. Consider the following probability table for a particular course's grade distributions where A=4, B=3, C=2, D=1, E/F=0.

x	4	3	2	1	0
$p(x)$	0.25	0.35	0.2	0.15	0.05

a. Find the expected value of the probability distribution. (7 points)

$4(.25)+3(.35)+2(.2)+1(.15)+0(.05) = 2.6$

b. What is the probability that a randomly selected student who took the course received a B or higher? (5 points)

$.25+.35 = .6$

60%

c. Suppose that 22 students take the course next semester. What is the probability that no more than four students will receive a D or less? (5 points)

binomialcdf (22, .2, 4) = .54288

54.288%

d. Given that a particular student passed the class, what is the probability that they received an A? (5 points)

$\frac{.25}{.8} = 31.25\%$

14. A department store sells shirts in three sizes and in three patterns (small, medium and large; plaid, print and stripes). The table below gives the number of shirts of each type sold on a particular, typical day.

Size	Plaid	Print	Stripes	
Small	3	2	3	8
Medium	10	5	7	22
Large	4	2	8	14
	17	9	18	44

a. What is the probability that the next shirt sold is a medium print shirt? (5 points)

$\frac{5}{44} = .1136...$

b. What is the probability that the next shirt sold will be a striped shirt? (5 points)

$\frac{18}{44} = .40909...$

c. Given that the next shirt is plaid, what is the probability that it is a medium size? (5 points)

$\frac{10}{17} = .58823...$

d. Given that the next shirt is large size, what is the probability that the shirt is print? (5 points)

$\frac{2}{14} = .142857...$

15. a. Determine the mean and standard deviation for the rainfall amount from seeded clouds shown below. (6 points)

4.1	**7.7**	**17.5**	**31.4**	**32.7**	**40.6**	**92.4**	**115.3**	**118.3**	**119.0**
129.6	198.6	200.7	242.5	255.0	274.7	302.8	334.1	430.0	489.1
703.4	978.0	1656.0	1697.8	2745.6					

1 Var Stats

$\bar{x} = 448.676$

$S_x = 663.29$

b. Construct a stemplot. Give a key for the graph. (7 points)

```
0   | 0 0 1 3 3 4 9
1   | 1 1 1 2 9
2   | 0 4 5 7
3   | 0 3
4   | 3 8
5   |
6   |
7   | 0
8   |
9   | 7
10  |
11  |
12  |
HI1 | 6 6
HI2 | 7
```

Key

```
0|0   = <10
9|1     910-919
HI1|6   1600-1699
HI2|7   2700-2799
```

note: you can use the entire data point, but values must be equally spaced to show heights of "bars" accurately.

c. Construct a boxplot for the same data. Label your primary axis, and the 5-number summary. (9 points)

min = 4.1

Q_1 = 66.5

Med = 200.7

Q_3 = 459.55

Max = 2745.6