Stat 2470, Exam #3, Spring 2015 Name KEY

Instructions: Show all work. State any formulas used. If you use the calculator, you should say which function you used, and what you entered into it, as well as any output. I can only give partial correct for incorrect answers if I have something to grade.

1. Conduct an appropriate hypothesis test comparing two different age groups and their stance duration (ms) to determine if older people have a shorter stance duration than younger people. Clearly state the appropriate hypotheses and compare to a 5% significance level. (10 points)

Age Group	Sample Average	Sample Standard Deviation	Sample Size
Older	780	117	28
Younger	801	73	16

2-SampleTTest Stats

$\bar{x}_1 = 780$

$Sx_1 = 117$

$n_1 = 28$

$\bar{x}_2 = 801$

$Sx_2 = 73$

$n_2 = 16$

not pooled

$\Rightarrow t = -.732479$

$p = .23399 > .05$

$H_0: \mu_1 = \mu_2$

$H_a: \mu_1 < \mu_2$

fail to reject Ho
not enough evidence to think stance duration declines w/age

2. Calculate the 90% confidence interval for the difference between the two stance duration means. (6 points)

2-SampTInt Stats
Same as above
C-level: .90

(-69.23, 27.232)

3. Data shown in the table below represent individual student test scores after reading the book only, or after seeing a 30-minute video lecture on the same material. Determine if watching the video helped students improve their test scores by more than 5 points. Conduct an appropriate hypothesis test at the 10% significance level. (13 points)

SUBJECT	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
READING	23	13	15	28	22	23	17	27	25	14	21	14	15	7	15	27
VIDEO	27	24	24	36	30	29	26	30	26	28	22	24	16	23	22	38

Paired TTest

$L_2 - L_1 \Rightarrow L_3$

TTest on L_3 Data $\Rightarrow$

$\mu_0 = 5$

$\mu > \mu_0$

$H_0: \mu_2 - \mu_1 = 5$

$H_a: \mu_2 - \mu_1 > 5$

$t = 2.138$

$p = .02469 < .10$

reject Ho
there is good reason to think watching the video improves scores by at least 5 points

4. According to a recent poll, a majority of Americans now support legal recognition for same-sex marriage. The numbers break down by religious affiliation and can be summarized in the table below. Conduct a hypothesis test to determine if there is strong evidence to believe that those Americans who do not attend church regularly support same-sex marriage at a higher rate than those who do. Clearly state your hypotheses, and your conclusion in the context of the problem. (10 points)

Group	Sample Size	Percent that Support Legal Recognition for Same-Sex Marriage
Does not attend church regularly/unaffiliated	2305	68.5%
Attends church regularly	4375	49.7%

2PropZtest

$H_0: p_1 - p_2 = 0$

$H_a: p_1 - p_2 > 0$

$x_1 = .685 * 2305 = 1578.925 \rightarrow 1579$

$n_1 = 2305$

$x_2 = .497 * 4375 = 2174.375 \rightarrow 2174$

$n_2 = 4375$

$z = 14.73$

$p = 2.123 \times 10^{-49} < .05$ reject Ho.

There is good evidence that the less religious people do have a higher support for same-sex marriage than the more religious.

5. Use the data in #4 to construction a confidence interval for the difference between the two proportions. (7 points)

2PropZInt

Same as above

C-level: .95

(.16405, .21218)

or (16.4%, 21.2%)

6. Conduct an ANOVA test on the following datasets. Clearly state the hypotheses and the result on the test in the context of the problem. The data represents samples of the heights (in) of women in various sports at a particular high school. (10 points)

Gymnastics	**57**	**59**	**62**	**61**	**62**	**58**	**62**	**63**	**60**
Soccer	69	69	64	67	70	65	67	58	64
Volleyball	68	70	69	67	66	73	67	68	69
Baseball	59	69	76	51	70	71	63	72	59

enter in L_1, L_2, L_3, L_4

ANOVA(L_1, L_2, L_3, L_4)

$F = 4.7677$

$p = .0073837 < .05$

Factor

$df = 3$

$SS = 310$

$MS = 103.33$

Error

$df = 32$

$SS = 693.556$ $\quad S_{xp} = 4.655$

$MS = 21.6736$

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

H_a: at least one $\mu_i \neq \mu_j$ for some $i \neq j$

reject H_0

at least one mean is different (heights of different sports players are different)

7. Conduct Tukey's procedure to confirm the conclusion of the test in #6 and if you rejected the null hypothesis, use it to group the sports with similar heights. Use the attached Q table and a significance level of 5%. (10 points)

1VarStats on each list

$\bar{X}_1 = 60.44$

$\bar{X}_2 = 65.89$

$\bar{X}_3 = 68.56$

$\bar{X}_4 = 65.56$

$\underline{X_1 \quad X_4} \quad X_2 \quad X_3$

$Q_{0.05, 3, 32} = 3.475$

$w = 3.475\sqrt{\frac{21.6736}{9}} = 5.3926$

$MSE = 21.6736$

$J = 9$

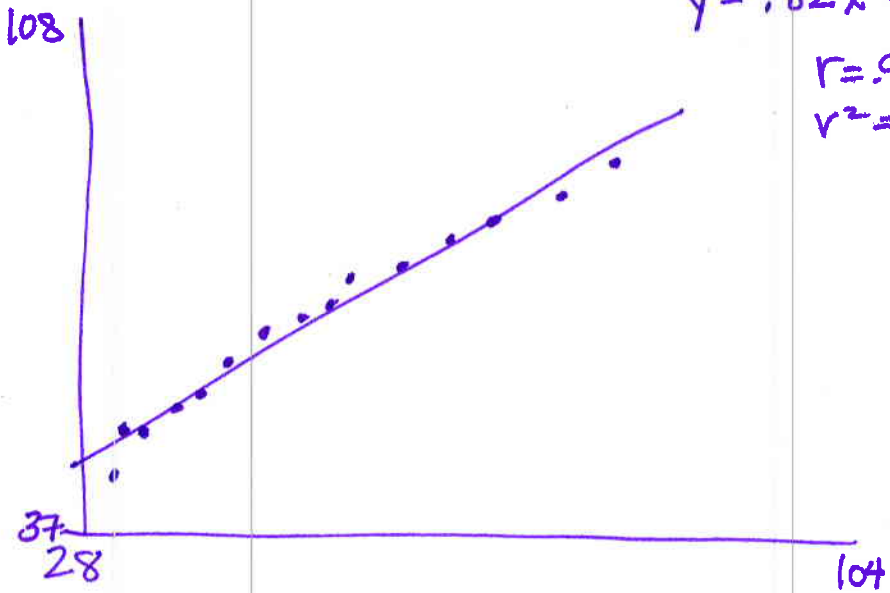
8. Use the data to find a linear regression model. Plot the data on a scatterplot along with the regression equation. State the correlation for the model. (10 points)

x	35	37	41	46	50	53	59	64	67	68	73	78	81	89	97
y	46	56	57	61	64	72	78	81	83	85	86	91	92	93	99

$y = .82x + 24.90$

$r = .9712$

$r^2 = .9433$



9. What proportion of the change in y (in #8) is explained by the relationship with x? (4 points)

94% (r^2)

10. Test the hypothesis that the correlation is non-zero for the model in question #8. Clearly state your hypotheses and conclusions. (8 points)

LinRegTTest

XList: L_1

YList: L_2

$\rho \neq 0$

$H_0: \rho = 0$

$H_a: \rho \neq 0$

$t = 14.7098$

$p = 1.756 \times 10^{-9} < .05$ reject H_0

there is good reason to believe the correlation is non-zero

11. Find the 95% confidence interval for β_1 for problem #8. (7 points)

LinRegTInt
XList: L_1
YList: L_2
Clevel: .95

(.70074, .94201)

12. Create a prediction interval for $\hat{y}$ at $x^* = 70$ using the regression equation derived in #8. (9 points)

$S = 3.96$

$\hat{Y} = 82.40$

1VarStats L_1
$\bar{x} = 62.53$

$82.40 \pm 2.16(3.96)\sqrt{1 + \frac{1}{15} + \frac{(70-62.53)^2}{5037.5}}$

$t_{\alpha/2, n-2} = 2.16$

82.40 ± 8.88

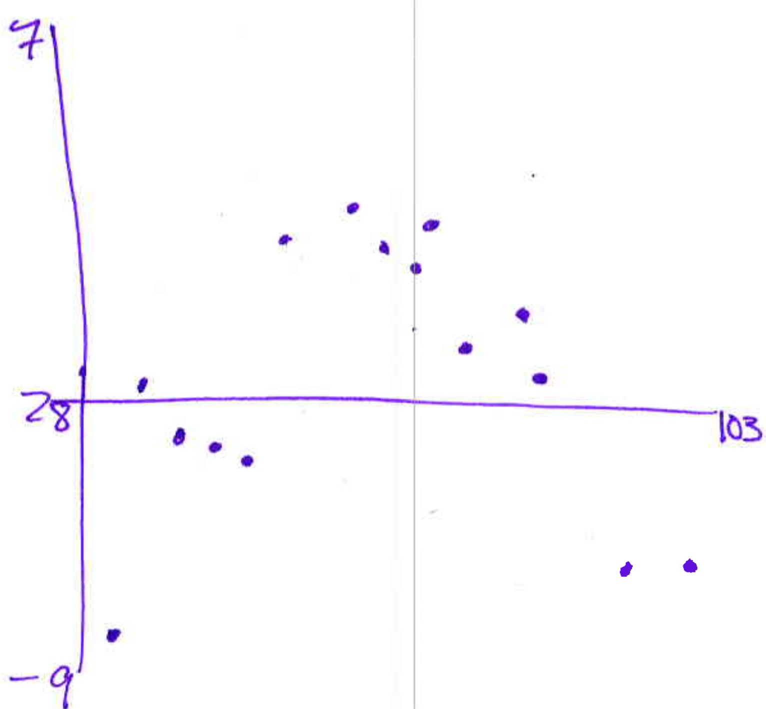
$S_x = 18.969$

$(73.52, 91.28)$

$S_{xx} = (18.969)^2(14) = 5037.5$

13. Use the linear regression equation to plot a residual plot and describe any potential problems with the model. Sketch the graph of the residuals vs. x here. (8 points)

$Y_1(L_1) \rightarrow L_3$
$L_2 - L_3 \rightarrow L_4$
scatterplot of
L_1 w/ L_4



They don't seem as random as they could be. Could mean a nonlinear model is better, or that the errors are not normally distributed.

14. Use the data in #8 to create a logarithmic model on $y = \alpha + \beta\ln(x)$. Give the equation here. Does the r^2 $(or\ R^2)$ value go up, stay the same, or go down? (8 points)

LnReg Y_2

$y = -128.82 + 50.14 \ln x$

$r = .98975$

$r^2 = .9796$

this went up from 94% to 98%.

15. Use the data in #8 to create a quadratic model of the data. Give the equation here, and say whether the r^2 $(or\ R^2)$ value went up, stayed the same, or went down. (8 points)

QuadReg Y_3

$y = -.009x^2 + 2.0199x - 10.36$

$R^2 = .98378$

this is the highest R^2 value so far, but a hypothesis test on the coefficient of x^2 (β_2) being nonzero may fail. This should be checked before accepting it.

16. Compare the three models you created from #8, #13, and #14. Which one has the best fit? If you were modeling this data for a project, which model would you choose and why? Explain your reasoning. (5 points)

While the quadratic model has the best fit
I would use the log model as the best fit for
the fewest parameters.
we should check the adjusted R^2 value

17. Let y be the error percentage for subjects reading a 4-digit liquid crystal display, and let x_1 be the level of backlight, x_2 be the character subtense, x_3 be the viewing angle and x_4 be the level of ambient light. The model to fit the data was $y = 1.52 + 0.02x_1 - 1.5x_2 + 0.02x_3 - 0.0006x_4 + \epsilon$.

   a. Estimate the mean value of the error percentage when $x_1 = 10, x_2 = 0.5, x_3 = 50, x_4 = 100$. (5 points)

   $y = 1.52 + .02(10) - 1.5(0.5) + .02(50) - .0006(100) = 1.91$

   b. Interpret the meaning of β_3 when all other variables are held constant. (5 points)

   as the viewing angle increases, the error rate
   goes up by 0.02 % per degree.