Exam 3 Study Guide Key

Exam 3 Study Guide

The only 2 formulas that will be given to you on Exam 3 are:

$$SD_{errors} = \sqrt{1-r^2} * SD_y = RMSE$$

$$SE_{slope} = \frac{SD_{errors}}{\sqrt{n} * SD_x} = \frac{\sqrt{1-r^2}}{\sqrt{n}} * \frac{SD_y}{SD_x}$$

Formulas you need to know:

- Slope of the regression line $= r\frac{SD_y}{SD}$
- **Correlation Coefficient,** $r = \frac{\sum_{i=1}^{n} Z_x Z_y}{n}$
- **Z and t test stats for testing H_0 : slope=0 in simple regression (1 slope):**

  $$Z = \frac{r}{\sqrt{1-r^2}} * \sqrt{n} \quad \text{and} \quad t_{(n-2)} = \frac{r}{\sqrt{1-r^2}} * \sqrt{n-2}$$

- **Chi square and F stats for testing H_0 : All slopes = 0 in multiple regression**

  $$\chi^2_{(p-1)} = \frac{R^2}{1-R^2} * n \quad \text{and} \quad F_{(p-1, n-p)} = \frac{R^2}{1-R^2} * \frac{n-p}{p-1}$$

  Hint: For simple regression $Z^2 = \chi^2$ and $t^2 = F$ so just take the square root of the Chi Square and F stats to find the Z and t stats for simple regression. Remember for simple regression p=2 and R=|r|.

- **ANOVA for regression and means: SST =SSM + SSSE, degrees of freedom: n-1 = p-1 + n- p**

p = # parameters in your model For **regression**: p = # of β's in regression equation, for **means: p = g** (# of groups)

Source	SS (Sum of Squares)	df
Model	$R^2 SST$ SSM (reg) SSB (means)	 p-1 g-1
Error	$(1-R^2)SST$ SSE (reg) SSW (means)	n-p n-g
Total	SST	n-1

Exam 3 Study Guide

Part VIII Chapter 21 (pages 107-111)

Question 1: 717 Stat 100 students rated their belief in the existence of ghosts on a scale of 1-10 (1 is certain ghosts don't exist and 10 is certain they do exist). They also classified their hometowns into 4 types: Small Town, Medium City, Suburb, and Big City.
Here are the results:

	Level of hometown	Average	SD	n
Ghosts	small_town	4.769	3.096	121
Ghosts	medium_city	4.115	2.833	104
Ghosts	suburb	5.140	3.106	356
Ghosts	big_city	5.309	3.064	136

more than 2 means ⇒ can't use z or t

a) What's the appropriate significance test the null that all 4 group means are the SAME in the "population". We just happen to see small differences in our sample due to chance?
   i) Z test only ii) t test only iii) F-test only iv) either z, t or F v) either t or F

   χ^2 would be okay too

b) What's the alternative hypothesis at least one group mean is different
   i) All 4 group means are different than each other in the population.
   ii) Some group means are different than each other in the population.
   iii) One of the group means is different than the others in the population.
   iv) That either i, ii, or iii is true.

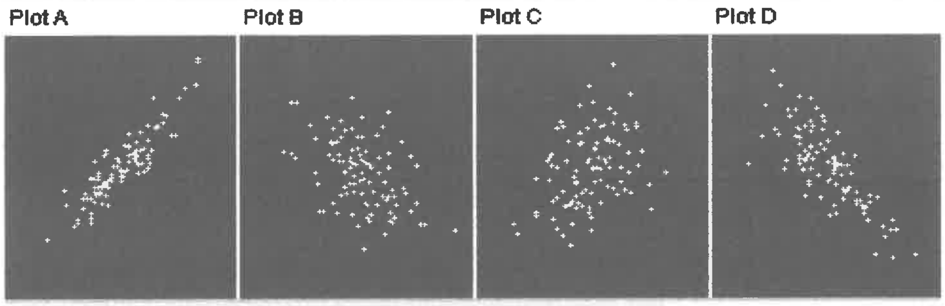
Fill out the ANOVA table below to test the null hypothesis that all the group means are the same in the population we just happen to see differences in the sample due to sampling variation. ***Show work inside each box (except for the df column). Write your answers in the blanks provided.***

Source	SS (Sum of Squares)	df	Mean Square	F Statistic	P-value
Model	SSB=107	df= 3 $g-1$	MSB= 35.7 (round to 1 decimal place) $\frac{107}{3}$	F= 3.8 $\frac{MSB}{MSW} = \frac{35.7}{9.4}$	You would have to look at the F curve with $df_{between}$ = 3, df_{within} = 713 F^* at a=0.05 is = 2.6 Reject null at at a=0.05? Yes or No
Error	SSW= 6706 6813-107 =	df= 713 $n-g$	MSW= 9.4 (round to 2 decimal places) $\frac{6706}{713}$	SD^+_{errors} = 3.07 (Round to 2 decimal places) $\sqrt{MSW} = \sqrt{9.4} = 3.07$	
Total	SST=6813	df= 716 $n-1$		R^2= 0.016 (Round to 3 decimal places) $\frac{SSB}{SST} = \frac{107}{6813} = 0.016$	

Exam 3 Study Guide

Part IX Simple Regression: Chapters 23-25

Question 2 pertains to the 4 scatter plots below:



Write the letter of the plot next to the correlation coefficient that is closest to it.

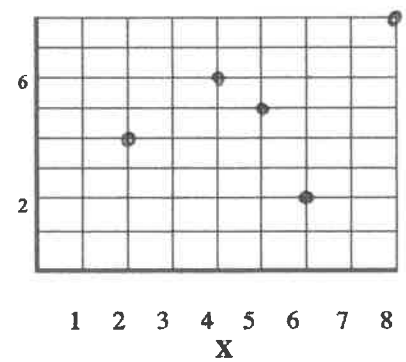
a) r = 0.36 C **b)** r = 0.9 A **c)** r = -0.79 D **d)** r = -0.46 B

Question 3

Compute the correlation coefficient (r) between X and Y by filling in the table below. Plot the points on the graph and check that the plot and r agree.

$Z = \frac{\text{val} - \text{avg}}{SD}$

X	Y	X in Standard Units	Y in Standard Units	Products
2	4	-1.5	-0.5	0.75
4	6	-0.5	0.5	- 0.25
5	5	0	0	0
6	2	0.5	-1.5	-0.75
8	8	1.5	1.5	2.25



avg of x's: 5
SD of x's: 2

avg of y's: 5
SD of y's: 2

a) The correlation coefficient, r = 0.4

$r = \text{avg of products} = \frac{0.75 + (-0.25) + 0 + (-0.75) + 2.25}{5}$

b) Using the result of part (a), determine the correlation coefficient for each of the following data sets. No computation is necessary. *Write your answers in the blanks provided.* ***Your answer should be a number.***

x	*y*
2	-8
4	-12
5	-10
6	-4
8	-16
r = -0.4	

multiply y's by -2

x	*y*
8	8
5	5
2	4
4	6
6	2
r = 0.4	

switch order

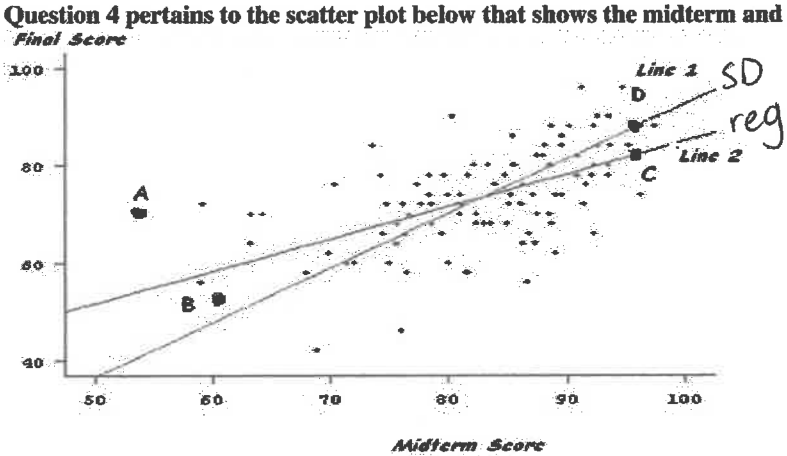
x	*y*
4	4
6	6
7	5
8	2
10	8
r = 0.4	

add 2 to x's

x	*y*
4	2
6	4
5	5
2	6
8	8
r = 0.4	

switch x's + y's

Question 4 pertains to the scatter plot below that shows the midterm and final exam scores for 107 students.



	Average	*SD*
Midterm	*83*	*9*
Final	*74*	*10*

Correlation: r = 0.6

a) Which is the regression line? ***Choose one:*** i) Line 1 ii) Line 2 flatter line

b) Look at students A, B, C and D on the graph. How did their actual scores on the final compare to their predicted scores? For each student circle whether their actual final exam scores were better than, worse than, or the same as the regression line predicted from their midterm scores.

	Actual Final Scores Compared to Predicted Ones			
Student A	*Choose One:*	*Better*	*Worse*	*The Same*
Student B	*Choose One:*	*Better*	*Worse*	*The Same*
Student C	*Choose One:*	*Better*	*Worse*	*The Same*
Student D	*Choose One:*	*Better*	*Worse*	*The Same*

c) Without any information about a particular student's midterm score, what would you expect him to score on the Final? 74 (average)

d) About 68% of the time, your prediction in part (c) will be correct to within 10 points. 1 RMSE when r = 0

e) Suppose you are told that the student has a midterm score of 74. Now what would you predict for his score on the final exam? Use the 3 step process (not the regression equation) ***Show your work!*** ***Circle answer.***

midterm 74 $\quad z_m = \frac{74-83}{9} = -1 \times 0.6 = -0.6 \; (z_F)$ $\quad$ Final $74 + (-0.6)(10) = 68$

f) About 68% of the time, your prediction in part (e) will be correct to within 8 points. ***Show your work!***

1 RMSE

$RMSE = SD_{errors} = \sqrt{1-r^2} \times SD_y = \sqrt{1-0.6^2} \times 10 = 8$

g) If a student was exactly average on both the midterm and the final which line ~~would he~~ fall on? point of averages
Choose one: Only the SD Line Only the Regression Line Both Neither

h) If a student was exactly 1 SD above average on both the midterm and the final which line would he fall on? $z_m = 1$ $\; z_f = 1$
Choose one: Only the SD Line Only the Regression Line Both Neither

assumes $z_x = z_y$

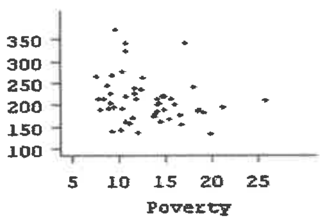
i) If a new scatter plot was drawn with 10 pts. added to everyone's final score then the correlation between midterm and final scores would.... ***Choose one:*** i) increase ii) decrease iii) stay the same
(For (i) and (j) assume that final scores are allowed to exceed 100)

j) If a new scatter plot was drawn with 10 % added to everyone's final score then the correlation between midterm and final scores would__________ ***Choose one:*** i) stay the same ii) decrease iii) increase

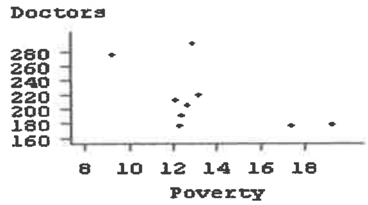
h) If point A was removed the, r would ... i) Decrease ii) Increase iii) Stay the Same

Question 5 The following scatter plots show the relation between poverty level (percentage of people living below the poverty line) and number of doctors (per 100,000 people) by state and by geographical region. The graph on the left has 50 points, one for each **individual** state's poverty and doctor level. The graph on the right has the same information condensed into 9 points, one for each of the 9 geographical regions. (In other words, the 50 states were divided into 9 geographical regions. The average poverty and doctor level was computed for each region.)

By State



By Division



see p. 119

a) The correlation coefficient for the graph on the left is -0.2. The correlation for the graph on the right is closest to
   i) -0.2   ii) -0.6   iii) 0   iv) 0.2   v) 0.6

b) The scatter plots above are an illustration of
   i) The Regression Effect   ii) Simpson's Paradox   iii) Ecological Correlation   iv) Negative Correlation

Question 6 For each of the following pairs of variables, check the box under the column heading that best describes its correlation among typical STAT 100 students:

	Correlation		Exactly -1	Between -1 and 0	About 0	Between 0 and 1	Exactly +1
a)	Weight in lbs.	Weight in kilograms (There are 2.2 lbs./kg)	☐	☐	☐	☐	☑
b)	Weight in lbs.	GPA	☐	☐	☑	☐	☐
c)	Freshman GPA	Sophomore GPA	☐	☐	☐	☑	☐
d)	How much you fall asleep in class	How much sleep you got the night before	☐	☑	☐	☐	☐
e)	Number of Points scored on Exam 1	Number of points missed on Exam 1	☑	☐	☐	☐	☐

Is there a relationship?

Yes ↓   No ⇒ r=0

Is it positive or negative?

positive ↓ yes: is it perfect
- Yes: r=+1
- No: r is between 0 + 1

negative ↓ yes: is it perfect
- Yes: r=-1
- No: r is between -1 and 0

Exam 3 Study Guide

Question 7
Here are the (rounded) summary statistics for height and weight of the 325 men in our class who completed Survey 1.

	Average	*SD*
Height	*71"*	*3"*
Weight	*175 lbs.*	*30 lbs.*

Correlation: ***r = 0.5***

$z = \frac{val - avg}{SD}$ $val = avg + (z)(SD)$

a) One student is exactly one SD above average in height and falls on the regression line. How many lbs. does he weigh?

$z_{height} = 1 \times \overset{r}{0.5} = 0.5 = z_{weight}$ $Weight = 175 + (0.5)(30) = 190$

b) Another student is 65" tall, predict how many lbs he weighs. ***Show work. Circle answer.***

Height 65" z_{height}: $z = \frac{65-71}{3} = -2 \times 0.5 =$ (r) z_{weight}: -1 Weight: $weight = 175 + (-1)(30) = 145$

c) What is the RMSE when predicting weight from height? ***Show work. Circle answer. Round your answer to the nearest lb.***

$RMSE = SD_{errors} = \sqrt{1-r^2} \times SD_y = \sqrt{1-0.5^2} \times 30 = 25.98$ or 26

d) If a student is 71" and weighs 175 lbs. he would fall on the ….. point of averages
Choose one:
i) SD line only   ii) regression line only   iii) Neither   iv) Both

e) What is the slope for ***predicting*** $\overset{y}{\textbf{\textit{weight}}}$ from height?
Show work, circle answer.

$slope = r \times \frac{SD_y}{SD_x} = 0.5 \times \frac{30}{3} = 5$

f) The men in our class who are 68" weigh 160 lbs. on the average. Can you conclude that the men in our class who weigh 160 lbs. are 68" tall on the average?
Choose one:
i) Yes
ii) No, they'd be taller than 68" on the average.
iii) No, they'd be shorter than 68" on the average.

$\frac{160-175}{30} = -0.5 \times 0.5 = -0.25$ $71 + (-.25)(3) = 70.25$

g) The regression equation for predicting height from weight is : **Height = .05 inch/lb * (Weight) +** 62.25
Find the y-intercept. ***Show work, write answer in blank below. Give your answer to 2 decimal places.***

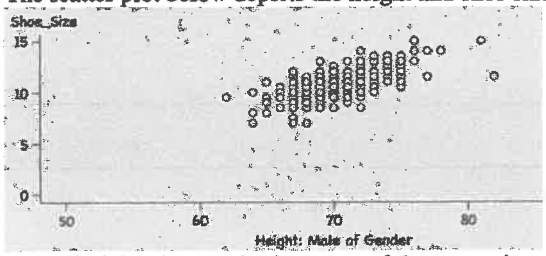
$71 = 0.05(175) + b$
$b = 62.25$

h) If all the heights of the men were converted to centimeters (by multiplying each height by 2.54 cm/inch) the correlation coefficient would …
Choose one: i) increase   ii) decrease   iii) stay the same   iv) not enough information given

Exam 3 Study Guide

Part X: Inference for Regression: Chapters 26-29

Question 8 Part I
The scatter plot below depicts the height and shoe size of 100 UI male undergrads



	Avg	*SD*
Height	71"	3"
Shoe Size	11	1.5

r = 0.7

a) Find the slope and y-intercept of the regression equation for predicting shoe size from height.

Shoe Size = 0.35 Height + -13.85 (Round to 2 decimal places.)

$$\text{Slope} = r \times \frac{SD_y}{SD_x} = 0.7 \times \frac{1.5}{3} = 0.35 \qquad 11 = 0.35(71) + b$$
$$b = -13.85$$

b) What is the SD_{errors} for predicting shoe size from height? $= \sqrt{1-0.7^2} \times 1.5 = 1.07$

i) 3 ii) 1.5 iii) 0.51 iv) 0.71 v) 1.07 vi) 2.14

Question 8 part II deals with *inference*—using the sample slope to make inferences about the population slope.

Now suppose the 100 students from Question 7 were randomly chosen from all male UI undergrads.

a) This corresponds to drawing 100 points, at random without replacement from a scatter plot depicting (write a number in the first blank and "with" or "without" in the second blank)

i) the heights and shoe sizes of all male UI undergrads population
ii) the heights and shoe sizes of the 100 randomly drawn students
iii) the heights and shoe sizes of all UI undergrads

b) Our best estimate of the **slope** for the whole population = 0.35 with a SE = 0.036
Show work for SE. Round to 3 decimal places. You don't need to re-calculate the sample slope.

$$SE_{slope} = \frac{\sqrt{1-r^2} \times SD_y}{\sqrt{n} \times SD_x} = \frac{1.07}{\sqrt{100} \times 3} = 0.036$$

c) Find the following confidence intervals for the slope of ***all*** UI undergrads when predicting shoe size from height. (Round answers to 3 decimal places.) Use the Normal Curve.

90% Confidence Interval = 0.35 +/- 1.65 SE_{slope} = (0.2906 to 0.4094)

95% Confidence Interval = 0.35 +/- 2 SE_{slope} = (0.278 to 0.422)

sample slope ± $z^* SE_{slope}$

d) In part (c) above we saw that a 90% confidence interval for slope did not include 0.
Based only on that information, you could conclude that a Z test for slope would reject the null hypothesis that $slope_{pop} = 0$ against the alternative that slope ≠ 0 at α= .10.

Fill in the 1st blank with "reject" or "not reject" and the 2nd with ">" or "≠".
(Hint: 90% CI interval has 5% area in each tail.)

Question 8 Part III: Z and t tests for Slope in Simple Regression
Formulas you'll need to know. (Or derive them from the 2 formulas you're given.)

$$Z_{slope} = \frac{\text{obs slope - exp slope}}{SE_{slope}} = \sqrt{n}\frac{r}{\sqrt{1-r^2}} \qquad t_{slope} = \frac{\text{obs slope - exp slope}}{SE^{+}_{slope}} = \sqrt{n-2}\frac{r}{\sqrt{1-r^2}}$$

a) Compute the Z statistic to test **H_0: $slope_{pop}$=0** **H_a: $slope_{pop}$>0**

$$Z = \sqrt{100} \times \frac{0.7}{\sqrt{1-0.7^2}} = 9.8$$

b) To change the Z-stat above to a t-statistic you would multiply by ________.

i) $\sqrt{\frac{98}{100}}$ ii) $\sqrt{\frac{100}{98}}$ iii) $\frac{100}{98}$ iv) $\frac{98}{100}$ v) $\sqrt{\frac{99}{100}}$ vi) $\sqrt{\frac{100}{99}}$

t should be tinier

c) How many degrees of freedom does the t-test have? 98

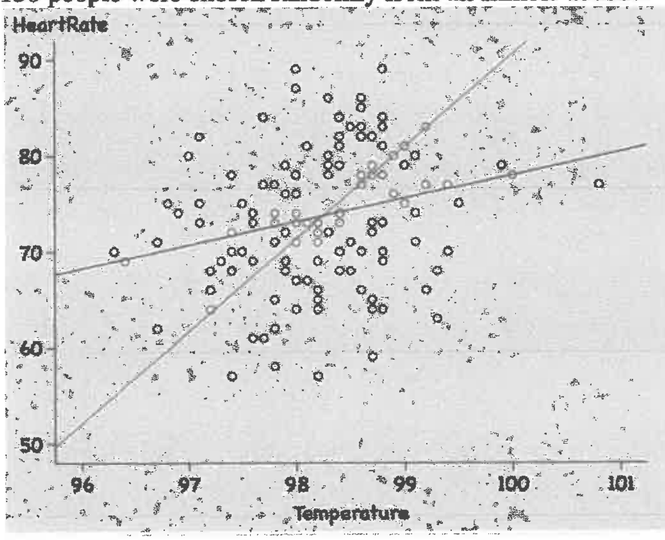
$n-p = n-2 = 100-2 = 98$

d) How do p-values for Z and t tests compare when performed on the same data sets with the same null and alternative hypotheses?

i) Z tests will always yield smaller p-values
~~ii)~~ Z tests will always yield larger p-values
~~iii)~~ Both tests will yield exactly the same p-values
~~iv)~~ Depending on the sample size the p-values from the z test could be larger, smaller or the same as the corresponding p-values from the t-test.

Question 9

The scatter plot below depicts the body temperatures and heart rates (beats per minute) of 130 adults. Pretend the 130 people were chosen randomly from all Illinois adults.



	Avg	***SD***
Temp	98	0.7
HR	74	7

r = 0.25

Sample Regression Equation
Heart Rate= -171 +2.5(Temperature)

a)What is the SE of the sample slope? Show work and round your answer to 2 decimal places.

$$SE_{slope} = \frac{\sqrt{1-0.25^2} \times 7}{\sqrt{130} \times 0.7} = 0.85$$

b) A 95% confidence interval for the population slope using the Normal Curve is (0.8 **to** 4.2 **). Round your answers to 2 decimal places.**

95% CI = sample slope $\pm 2SE_{slope}$ = $2.5 \pm 2(0.85)$

c) The confidence interval above didn't include 0, so if we did a 2 sided Z test, testing the null hypothesis that the slope = 0 for the whole population we should ______________the null. Reject? or Not Reject? Circle one.

d) Do the hypothesis test by calculating Z and the p-value. The null and alternative are:

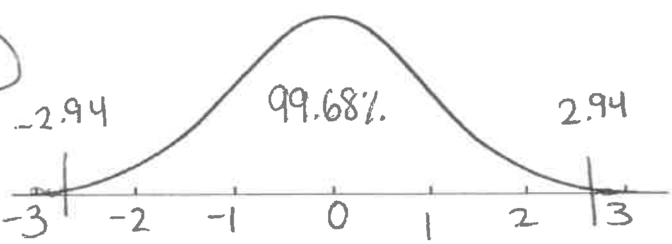
H_0: Slope of the regression equation for the *whole* population is 0. We just happened to get a small slope of 2.5 in our sample of n=130 due to the luck of the draw.

H_a: Slope of the regression equation for the whole population ≠ 0. Our sample slope of 2.5 is too big to be due to chance variation.

i) Calculate the test statistic Z for the slope.

$$Z = \sqrt{n} \times \frac{r}{\sqrt{1-r^2}} = \frac{\sqrt{130} \times 0.25}{\sqrt{1-0.25^2}} = 2.94$$

ii) Mark Z on the Normal Curve and find p-value.



iii) Conclusion? Reject null?

$$\text{p-value} = \frac{100-99.68}{2} = 0.16\%$$

2-sided p-value = 2 × 0.16% = 0.32%

less than 5% ⇒ reject Ho.

There is evidence that the slope is not equal to 0.

Exam 3 Study Guide

Question 10

We're trying to fit a simple linear regression model for the whole population: $Y=\beta_0 + \beta_1 X + \varepsilon$. (Assume ε are independent and normally distributed with constant variance). We draw a random sample of **n=7** from the population and get a sample correlation **r = 0.6**. Compute the 4 test statistics for testing the **null H_0: β_1=0.** (same as testing **H_0: $r_{population}$=0.**) ***(Round your final answers to 4 decimal places, but don't round during intermediate steps.)***

$p = 2 = \#$ of parameters

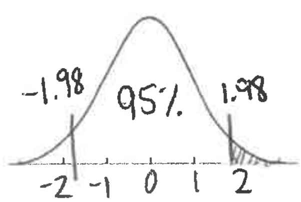
$n = 7$

a) R^2 = 0.36 **$1-R^2$ =** 0.64

b) Now compute the 4 statistics below.

	Z	χ^2	t	F
Compute the values of the 4 test statistics. ***Show work below your answers.***	Z= 1.9843 *(1 pt.)* $Z=\sqrt{\chi^2}$ $=\sqrt{3.9375}$ $=1.9843$	χ^2 = 3.9375 *(1 pt.)* $\chi^2=\frac{R^2}{1-R^2}\times n$ $=\frac{0.36}{0.64}\times 7$ $=3.9375$	t = 1.6771 *(1 pt.)* $t=\sqrt{F}$ $=\sqrt{2.8125}$ $=1.6771$	F= 2.8125 *(1 pt.)* $F=\frac{R^2}{1-R^2}\cdot\frac{n-p}{p-1}$ $=\frac{0.36}{0.64}\times\frac{5}{1}$ $=2.8125$

c) Compute the p-values for each statistic. Assume the alternative for the Z and t test is 1-sided: H_A: $\beta_1 > 0$, and assume the alternative for the χ^2 and F is 2-sided: H_A: $\beta_1 \neq 0$.

Z	χ^2	t	F
p-value= 2.5 % **Label Z on the normal curve below and shade the area representing the p-value.**  *	p-value= 5 % How many degrees of freedom? 1 $p-1=1$	Choose one: i) 1% ii) 2% iii) 7.7% How many degrees of freedom? 5 $n-p=7-2=5$	p-value=______% Choose one: i) 2% ii) 4% iii) 15.4% How many degrees of freedom in numerator? 1 in denominator? 5

***If Z is between 2 lines on the Normal Table you may approximate middle area.**

d) Suppose our sample y values are: 1, 2, 3, 4, 5, 6, 7. Compute the SST. (Show work). avg of y's = 4

$SST=(1-4)^2+(2-4)^2+(3-4)^2+(4-4)^2+(5-4)^2+(6-4)^2+(7-4)^2$

$=9+4+1+0+1+4+9=28$

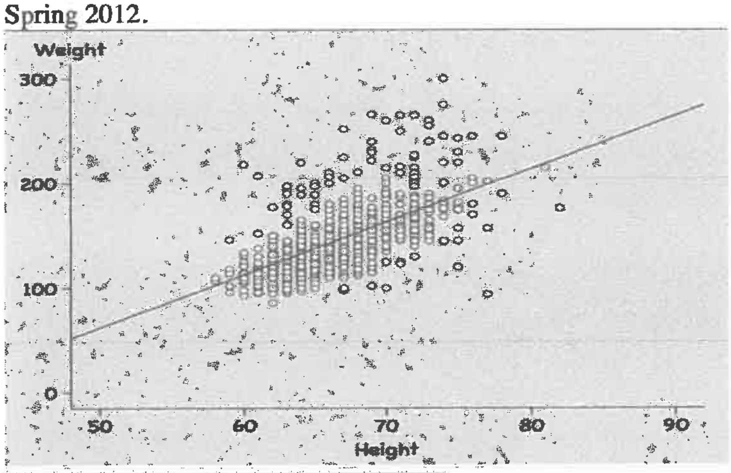
e) Compute SSM. _______ **Hint: Use part (a)**

$SSM=R^2 SST$

$=0.36(28)=10.08$

Exam 3 Study Guide

Part X1: Binary Variables in a Regression Model (Chapter 30)
Question 11
Here's a scatter plot of the heights and weights of the 227 males and 485 females who responded to Survey 1 in Spring 2012.



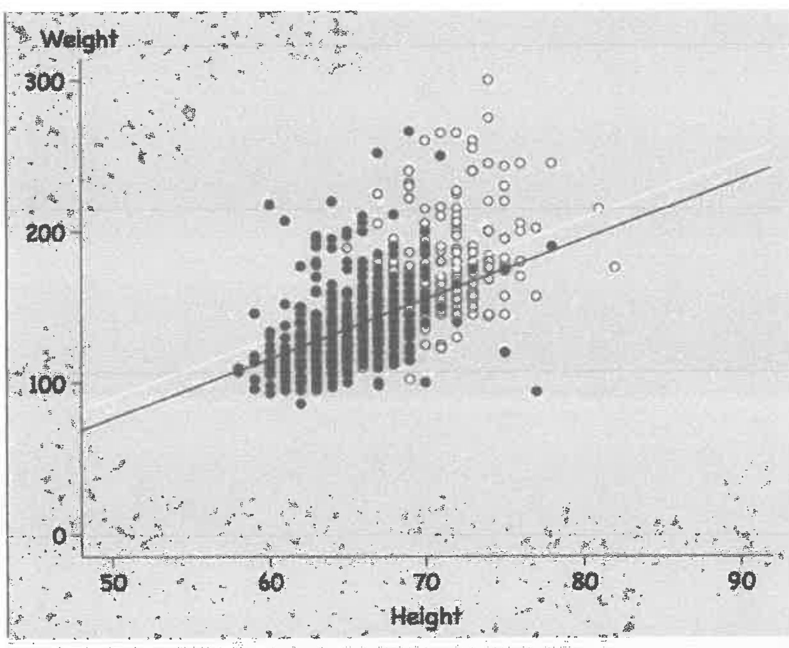
Equation	r	SD_{errors}
Weight = -190.8 + 5.061 × Height	0.6149	25.84

a) Why is r (0.6149) larger in the combined male female plot to the left than in the 2 separate regression below?
Circle all that are true.

- ~~I)~~ Because there are more points in the combined plot
- II) Because males are both taller and heavier on the average so combining males and females in one plot increases r.
- III) Whenever two positively correlated groups are combined the correlation increases.

Think of sumo wrestlers example from HW!

b) Here's the scatter plot broken into groups fitted with the same slope



Gender	Equation	r	SD_{errors}
Male	Weight = -106.2 + 3.931 × Height	0.4382	29.53
Female	Weight = -119.9 + 3.931 × Height	0.3795	23.16

Fill in the blanks in the equations below to match the equations for males and females above.
(Hint: Males are coded 0, so fit the equation to males first by subbing in 0 for Gender)

Weight = -106.2 **+** 3.931 **Hgt +** -13.7 **Gender**

plug in 0 for gender + look at equation for males

same slope

plug in 1 for gender + look @ equation for females

Exam 3 Study Guide

Chapter 37

Question 12 There are 3 sections to the MCAT: Physical Science (PS); Biological Science (BS); and Verbal Reasoning (VR). Each is scored on a scale of 1-15. Suppose we randomly selected 55 UI pre-meds from all UI pre-meds who took the MCAT last year and got the following sample multiple regression equation for predicting PS from both VR and BS: $\hat{PS} = 1.6 + 0.2 \times VR + 0.6 \times BS$

$p=3$

$n=55$

$SD_y = \sqrt{\frac{SST}{n}}$

$SST = SD_y^2 \cdot n$

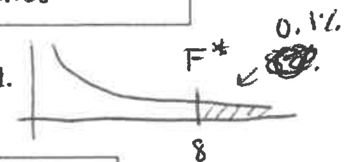
Summary Stats

	Average	Median	SD	Min	Max	n
VR	9.764	10.00	1.768	6.000	13.00	55.00
BS	9.782	10.00	1.522	6.000	14.00	55.00
y PS	9.709	10.00	1.659	5.000	14.00	55.00

a) Here's the ANOVA table to test the overall regression effect. Fill in them missing values. You'll need to use some info from the summary stats above to calculate SST.

Source	SS (Round to nearest whole number)	df	MS (Round to 1 decimal place)	(Round to 2 decimal places)
Model	SSM= 63	$p-1 = 2$	$\frac{63}{2} = 31.5$	F= $\frac{31.5}{1.7} = 18.53$
Error	SSE= 88.38	$n-p=52$	$\frac{88.38}{52} = 1.7$	$SD^{+}_{errors} = \sqrt{1.7} = 1.3$
Total	SST= $SD_y^2 \cdot n$ $= 1.659^2 \times 55 = 151.38$	54		$R^2 = \frac{SSM}{SST} = \frac{63}{151.38} = 0.42$

b) Our F is > or < F* = __>__ so our p-value > or < __<.1__% so we can or cannot reject the null.
Circle the correct ">" or "<" signs, fill in the 2 blanks and circle "can" or "cannot"



F Distribution critical values for P=0.001

Denominator ↓ / Numerator DF

DF	1	2	3	4	5	7	10	15	20	30	60	120	500	1000
1	405284	499999	540379	562500	576405	592873	605621	615764	620908	626099	631337	633972	635983	636301
2	998.50	999.00	999.17	999.25	999.30	999.36	999.40	999.43	999.45	999.47	999.48	999.49	999.50	999.50
3	167.03	148.50	141.11	137.10	134.58	131.58	129.25	127.37	126.42	125.45	124.47	123.97	123.59	123.53
4	74.137	61.245	56.177	53.436	51.712	49.658	48.053	46.761	46.100	45.429	44.746	44.400	44.135	44.093
5	47.181	37.122	33.202	31.085	29.752	28.163	26.917	25.911	25.395	24.869	24.333	24.061	23.852	23.819
7	29.245	21.689	18.772	17.198	16.206	15.019	14.083	13.324	12.932	12.530	12.119	11.909	11.747	11.722
10	21.040	14.905	12.553	11.283	10.481	9.5174	8.7539	8.1288	7.8038	7.4688	7.1224	6.9443	6.8085	6.7846
15	16.587	11.339	9.3352	8.2526	7.5673	6.7408	6.0808	5.5351	5.2484	4.9502	4.6378	4.4749	4.3478	4.3275
20	14.819	9.9526	8.0984	7.0960	6.4606	5.6920	5.0753	4.5618	4.2900	4.0051	3.7030	3.5439	3.4184	3.3981
30	13.293	8.7734	7.0544	6.1245	5.5339	4.8173	4.2389	3.7528	3.4928	3.2171	2.9197	2.7595	2.6310	2.6100
60	11.973	7.7678	6.1712	5.3067	4.7565	4.0864	3.5415	3.0781	2.8265	2.5549	2.2522	2.0821	1.9390	1.9150
120	11.380	7.3212	5.7814	4.9471	4.4157	3.7669	3.2372	2.7833	2.5345	2.2621	1.9502	1.7668	1.6027	1.5736

c) When the null is true we'd expect our F to be about __1__. Given how your F compares to that you'd expect the p-value to be about __0__%. see p-value calculator

d) Suppose you decided to reject the null, you'd conclude that

- i. Both slopes must be significant
- ii. The VR slope must be significant
- iii. The BS slope must be significant
- ~~iv.~~ The intercept must be significant
- v. Either the VR or the BS slope or both must be significant

Exam 3 Study Guide

$p=3 \quad n=764$

Question 13

On a Stat 100 Survey, 764 students reported how many drinks they typically consumed per week, how many hours they typically exercised per week and their GPA. The multiple regression equation predicting GPA from drinks and exercise yielded R=0.04. Assume these students were randomly sampled from a larger population of possible Stat 100 students.

a) Do a χ^2 test for the overall regression effect. How many degrees of freedom? $p-1 = 2$

$$\chi^2 = \frac{R^2}{1-R^2} \cdot n = \frac{0.04^2}{1-0.04^2} \times 764 = 1.224$$

b) Compute the F stat. How many df in the numerator 2 ? The denominator 761 ?

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-p}{p-1} = \frac{0.04^2}{1-0.04^2} \times \frac{761}{2} = 0.61$$

c) Here are the p-values for the 2 tests. Which one is for the χ^2 and which is for the F?

55.74% and 55.58% Label each as either χ^2 or F. $F_{p\text{-value}} > \chi^2_{p\text{-value}}$

F χ^2

Question 13

In the overall regression test, the null hypothesis is that the population slopes all = 0.

That's equivalent to the null hypothesis that in the population ....

i) $R = 0$

ii) $R^2 = 0$

iii) $Y = \bar{Y}$

iv) all of the above

v) none of the above

Question 14

If the χ^2 test doesn't yield significant results, is it possible the F test still would?

i) Yes, since the F test yields slightly more precise tests.

ii) Yes, if the sample size is relatively small, the F test results could yield significantly different results.

iii) No, the p-value for the F test will always be greater so it could never yield more significant results.

iv) It's impossible to know since F is centered at 1 when the null is true and the χ^2 is centered at its degrees of freedom making comparisons of results statistically meaningless.