Sample MATH 105-Test #3a

1. Which of the following is **NOT** true of the χ^2 (chi-square) test of independence?
   a) If the observed and expected frequencies are not close, the χ^2 test statistic will be large and the p-value will be small.
   b) A large p-value leads us to conclude that the two variables have an association.
   c) The χ^2 test statistic is based on differences between the observed and expected values.

2. Which of the following statements is **false**:
a) The correlation coefficient, r, is always between -1 and 1.
b) The correlation coefficient is unitless.
c) A correlation coefficient of .62 suggests a stronger correlation than a correlation coefficient of -.82.
d) The correlation coefficient should only be used when the association is linear.

3. The regression equation is $\hat{y} = -2.7 - 13x$. Identify the slope and the y-intercept.
a) y-intercept=-2.7, slope=-13 b) y-intercept=-13, slope=-2.7
c) y-intercept=-2.7, slope=13 d) y-intercept=2.7, slope=13

4. A study is done to predict the age of a certain type of tree based on the tree's diameter. The response variable in this study is
a) the diameter b) the tree c) the type of tree d) the age

5. If faculty who are good researchers tend to be poor teachers and faculty who are poor researchers tend to be good teachers, then the correlation between the quality of research and the quality of teaching is:
   a) negative b) positive c) zero

6. The correlation coefficient measures
   a) whether or not a scatter diagram shows an interesting pattern
   b) whether a cause and effect relationship exists between two variables
   c) the strength of a straight line relationship between two variables
   d) whether there is a relationship between two variables.

7. An article in the *Archives of Internal Medicine* reported that in a sample of 244 men, 73 had elevated total cholesterol levels (>200 milligrams per deciliter). In a sample of 232 women, 44 had elevated cholesterol levels. Compute a 95% confidence interval for the difference in proportion of men and women with high cholesterol, p_m-p_w. Interpret your confidence interval in context of the problem.

$\hat{p}_m = \frac{73}{244} = .299$

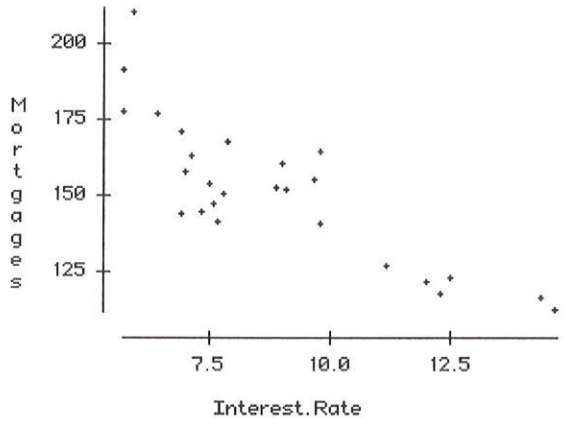
$\hat{p}_w = \frac{44}{232} = .190$

$(.299 - .190) \pm 1.96\sqrt{\frac{.299 \times .701}{244} + \frac{.190 \times .81}{232}} = (.033, .185)$

We can be 95% confident that between 3.3% to 18.5% more men will have elevated cholesterol than women.

Calc. Function: B: 2-PropZInt

8. Since 1980, average mortgage interest rates have fluctuated from a low of under 6% to a high of over 14%. Is there a relationship between the amount of money people borrow and the interest rate that's offered. Here is a scatterplot of *Total Mortgages* in the US (in millions) versus *Interest Rate* at various times over the past 26 years.



a) Which is the most likely value for the correlation coefficient?

a) .84 b) -.84 c) 0 d) .20 e) -.20

b) The regression equation to predict mortgages in millions (y) from interest rate (x) is

$\hat{y}$ =222.9-7.68 x

What is the predicted total mortgages borrowed if the interest rate is at 8.0%?

$\hat{y} = 222.9 - 7.68 \times 8.0 = \$161.46$ million

c) Suppose the actual amount people borrowed the last time the interest rate was 8.0% was 155 million dollars. What is the residual for this observation?

Resid = $y - \hat{y} = 155 - 161.46 = \$-6.46$ million

obs-pred.

d) Interpret the slope in context of the problem in one sentence. For every additional 1% in interest rates the amount borrowed increases by $7.68 million.

e) If the interest rate increases by 3%, how much would the predicted total mortgage borrow decrease?

For every 1% it ↑ by $7.68 million so for 3% it would decrease by

$7.68 \times 3 = \$23.04$ million

9. The table below shows the years in school and salary (in 1000s) for a sample of 18 men 40 years old.

Years School	Salary ($1000s)
10	28
16	38
12	16
6	13
12	25
12	30
16	35
16	27
18	28
12	28
10	26
12	21
16	34
14	30
11	21
12	27
19	29
16	24

a) Which variable is the explanatory variable and which variable is the response variable?

→ Years School

→ Salary

b) The correlation coefficient is .623. Calculate R^2.

$.623^2 = .388$

c) Interpret R^2 in context of the problem.

38.8% of the variation in salary can be explained by the regression equation.

d) If you were included in the sample, would you prefer to have a negative or positive residual for your salary? Explain.

Positive since this would mean you would be making more than predicted by the line.

10. Below is a two-way table of the variables & values are given in the parentheses.

→ 44 − 48.54 = −4.54

	Math-Science	Social Science	Humanities	Total
Male	45 (33.77) 11.23	41 (47.69) −6.69	44 (48.54) −4.54	130
Female	35 (46.23) −11.23	72 (65.31) 6.69	71 (66.46) 4.54	178
Total	80	113	115	308

Identify the cells with the largest residuals. Explain what this signifies in context of the data.

The math-science group had the largest residuals. There are more male math/science major than expected & less female math/science majors than expected. Both the social science & humanities majors had more females than expected & less males.

11. Do females have more facebook friends than males? Recall that we collected data in the beginning of the semester. One of the questions was "how many Facebook friends do you have?. The StatCrunch results appear below. Use the output to answer the questions below.

Hypothesis test results:

μ_1 : mean of population 1
μ_2 : mean of population 2
$\mu_1 - \mu_2$: mean difference
$H_0 : \mu_1 - \mu_2 = 0$
$H_A : \mu_1 - \mu_2 > 0$
(with pooled variances)

Difference	Sample Mean	Std. Err.	DF	T-Stat	P-value
$\mu_1 - \mu_2$	40.43	17.11614	3175	2.3620977	0.0091

95% confidence interval results:

μ_1 : mean of population 1
μ_2 : mean of population 2
$\mu_1 - \mu_2$: mean difference
(with pooled variances)

Difference	Sample Mean	Std. Err.	DF	L. Limit	U. Limit
$\mu_1 - \mu_2$	40.43	17.11614	3175	6.87019	73.98981

a) Using a significance level of 0.05, give a conclusion for the hypothesis test in context of the problem.

$H_0: \mu_f = \mu_m$
$H_1: \mu_f > \mu_m$
p-value = .0091

We can reject H_0 & conclude that there is evidence that, on average, females have more facebook friends than males.

b) Interpret a 95% confidence interval for the difference in the average number of Facebook friends for males and females.

We can be 95% confident that, on average, females have between 6.9 & 74.0 more facebook friends than males.

$z = \frac{.899 - .850}{\sqrt{\frac{.8752 \times .1248}{386} + \frac{.8752 \times .1248}{359}}} = 2.022$ p-value $= 2 \times .0216 = .0432$

12. A pew research Center poll asked randomly selected subjects if they agreed with the statement that "It is morally wrong for married people to have an affair." Among the 386 women surveyed, 347 agreed with the statement. Among the 359 men surveyed, 305 agreed with the statement. Use a .05 significance level to test the claim that the percentage of women who agree is different from the percentage of men who agree.

a) State the null and alternative hypotheses.

$H_0: p_w = p_m$ $H_1: p_w \neq p_m$

b) Calculate the test statistic and p-value. $\bar{p} = \frac{347+305}{386+359} = .8752$

$\hat{p}_w = \frac{347}{386} = .899$ $\hat{p}_m = \frac{305}{359} = .850$ Calc function: 6: 2PropZTest

c) State your conclusion in context of the problem.

We can reject H_0 & conclude that the percentage of women who agree w. the statement is different than the percentage of males

d) Calculate and interpret a 95% confidence interval for the difference between men and women with regards to how they feel about this question.

$(.899 - .850) \pm 1.96\sqrt{\frac{.899 \times .101}{386} + \frac{.85 \times .15}{359}}$

$= (.0014, .0966)$ Calc function: B: 2PropZInt

We can be 95% confident that women are between .14% & 9.66% more likely to agree with the statement.

e) Does your confidence interval and hypothesis test agree? Explain.

Yes because we concluded in the hypothesis test that they were different & the confidence interval does not include 0.

13. Winning team data were collected for teams in different sports, with the results given in the table below.

	Basketball	Baseball	Hockey	Football	Total
Home Team Wins	131	56	72	73	332
Visiting Team Wins	53	53	38	40	184
Total	184	109	110	113	516

a) (8) State the null and alternative for conducting a chi-square test of independence on the two variables.

H_0: Sport & homefield advantage are independent

H_1: " " " " dependent

b) The p-value for the above test was .008. Write a conclusion in context of the problem.

We can reject H_0 & conclude that homefield advantage is dependent on type of sport.