HWCheck

1. Pick up the solutions for the "Extra Practice with Correlation....."

2. Part of your HW score should be whether you studied the packet on the Chi-Square test of Independence. Be honest ••

Extra Practice with Correlation and LSRL "by hand"

This data is a comparing Gross Domestic Product, GDP (per capita, inflation adjusted), with Infant Mortality rate, IMR (number of 0-5 year olds dying per 100).

Countries	GDP	IMR
Nigeria	1956	154
Finland	31684	3.3
Bolivia	37313	61
Congo, Dem. Rep	338	199
Mexico	11772	19
China	5450	24
Hungary	17726	7.1
Bangladesh	1326	62
South Korea	22373	5.4

a) make a scatter plot on your GDC and make a sketch of it below

b) comment on the correlation you see.

c) Calculate the correlation coefficient by hand showing both the formula and critical values. (then check against your GDC)

d) Calculate the LSRL by hand in Point-Slope form, showing all details as done in class.
If you want you can convert to slope intercept form and check against your GDC.

Today ---The full Chi-Square Test of Indep. Process

Tuesday- Special Situations + finish eval. Teenage Pregnancy Project

Wednesday- Get a list of Unit Test items

Packet P3 (Info on selecting a project and Ideas for project)

Thursday- Evaluate another project (using the scoring guide) +

Use Computer spreadsheet to calculate r and LSRL

Friday - Review Questions

Monday - Test on Statistical Applications ← Oct. 8th

Chi Square Statistic is :

$$\chi^2 = \sum \frac{(f_e - f_o)^2}{f_e}$$

← Formula packet

Calculated the Chi-Square Statistic
3 different ways

AIM:

Carry out the entire
Test of Independence

We'll step back and just observe example of the whole process

(ppt: Chi Square ppt)

Chi-Square Test of Independence.pptx

Full example of a Chi-Square Problem

Do the data provide convincing evidence of an association between age group and whether or not someone uses Facebook for all U.S. adults?

1. Do you use Facebook? Mark Z. is interested in finding out if there is an association between age group and Facebook use among U.S. adults. He takes a random sample of 250 U.S. adults aged 18 or older and records their age and whether or not they use Facebook. The two-way table summarizes the results.

	Age (years)				
Facebook user?	18–29	30–49	50–64	65+	Total
Yes	60	64	46	21	191
No	8	20	22	9	59
Total	68	84	68	30	250

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

Do the data provide convincing evidence of an association between age group and whether or not someone uses Facebook for all U.S. adults? **(Use the χ^2 statistic, Method A)**

Step 1

H_0: Age is Facebook are independent

H_1: Age and Facebook use are associated (are Not Independent)

Step 2 Calculated Expected Values

Facebook Use	18-29	30-49	50-64	65+
Yes	51.952	64.176	51.952	22.92
No	16.048	19.824	16.048	7.08

$\frac{(59)(68)}{250}$

FROM matrix B

$$\chi^2 = \sum \frac{(o-fe)^2}{fe}$$

Step 3
All expected are greater than 5
so test result will be valid

Step 4 : $\alpha = .05$ (5%)

$\alpha = .1$
$\alpha = .01$

Step 5 degrees of freedom $df = (4-1)(3-1) = 3$

$\chi^2_{critical} = 7.815$

TABLE

Step 6
Reject H_0 if $\chi^2 > 7.815$

Step 7 $\chi^2 = 8.86$ ↑ GDC

Step 8
- Since $8.86 > 7.815$, then we reject H_0.

There is evidence to support the claim that age and Facebook usage are associated

Critical values of the χ^2 distribution

$\alpha = .1$ $\alpha = .05$

significance level

degree of freedom → p

ν	10%	5%		1%	
1	2.706	3.841	5.024	6.635	7.879
2	4.605	5.991	7.378	9.210	10.597
3	6.251	7.815	9.348	11.345	12.838
4	7.779	9.488	11.143	13.277	14.860
5	9.236	11.070	12.833	15.086	16.750
6	10.645	12.592	14.449	16.812	18.548
7	12.017	14.067	16.013	18.475	20.278
8	13.362	15.507	17.535	20.090	21.955
9	14.684	16.919	19.023	21.666	23.589
10	15.987	18.307	20.483	23.209	25.188
11	17.275	19.675	21.920	24.725	26.757
12	18.549	21.026	23.337	26.217	28.300
13	19.812	22.362	24.736	27.688	29.819
14	21.064	23.685	26.119	29.141	31.319
15	22.307	24.996	27.488	30.578	32.801
16	23.542	26.296	28.845	32.000	34.267

2. Are you headed for college ?

A curious student wondered if students in her high school would respond differently to a teacher than to a student when asked about their intention to go to college after high school. One hundred students from her high school volunteered to participate in an experiment, and they were randomly assigned to have either a student or a teacher interview them. This two-way table summarizes the data.

		Interviewer		
		Student	Teacher	Total
Going to college?	Yes	14	27	41
	No	36	23	59
	Total	50	50	100

Expected Values

	Student	Teacher
Yes	20.5	20.5
No	29.5	29.5

$\frac{(50)(59)}{100}$

Does the data provide convincing evidence that there is an association between how a student responds when asked whether they will attend college and who asks the question? **(Use the χ^2 statistic, Method A)**

Step 1

H_0: Whether the interview is a Teacher or student is independent of whether the student indicates College intention.

H_1: Interviewer and indicating college intention is associated.

Step 3 The test will be valid because all four expected values are greater than 5

Step 4 $\alpha = .05$

Step 5 $df = (2-1)(2-1) = 1$

$\chi^2_{critical} = \cancel{6.67}\ \cancel{6.64}\ 3.841$

Step 6 Reject H_0 if $\chi^2 > 3.841$

Step 7 $\chi^2 = 6.986$

Step 7. We reject H_0 because $6.986 > 3.841$

There is evidence to support the claim that the interviewer and college intention are assoc.

significance level

degree of freedom

p	10%	5%		1%	
$\nu = 1$	2.706	3.841	5.024	6.635	7.879
2	4.605	5.991	7.378	9.210	10.597
3	6.251	7.815	9.348	11.345	12.838
4	7.779	9.488	11.143	13.277	14.860
5	9.236	11.070	12.833	15.086	16.750
6	10.645	12.592	14.449	16.812	18.548
7	12.017	14.067	16.013	18.475	20.278
8	13.362	15.507	17.535	20.090	21.955
9	14.684	16.919	19.023	21.666	23.589
10	15.987	18.307	20.483	23.209	25.188

B.B.

Now using Method B

B. The Chi Squared Test of Indep.
(using Probability)

You may be asked to perform the test using **probability** instead of the Chi-Square statistic. In that case, you will follow steps 1 through 4. But continue with the following steps instead:

5. State the ***rejection inequality***

"If the p-value is **less than** the significance level " (5% or 0.05 for many cases), then I will reject H_0.

6. **Calculate the *p-value*.**
This step must be done on a calculator. Follow the same calculator steps as if you are calculating the Chi-square statistic.

3. **Chi-Square by finding the *p-value*, Method B.**
Are people who are prone to sudden anger more likely to develop heart disease? An observational study followed a random sample of 8474 people with normal blood pressure for about four years.

Each person took the Spielberger Trait Anger Scale test. Researchers also recorded whether each individual developed coronary heart disease (CHD).

		Anger level			
		Low	Moderate	High	Total
CHD status	Yes	53	110	27	190
	No	3057	4621	606	8284
	Total	3110	4731	633	8474

Do these data provide convincing evidence of an association between the variables in the larger population? (Make the determination by finding the p-value, Method B)

H_0: CHD status and anger level are independent.

H_1: CHD status and anger level are associated

Step 2 and 3: Check if expected values are greater than 5

Test will be valid because . . . .

Step 4: $\alpha = .05$

Step 5: Reject H_0 if the p-value is less than 5% (or .05)

↑ significance level

χ^2 critical table (known) is needed

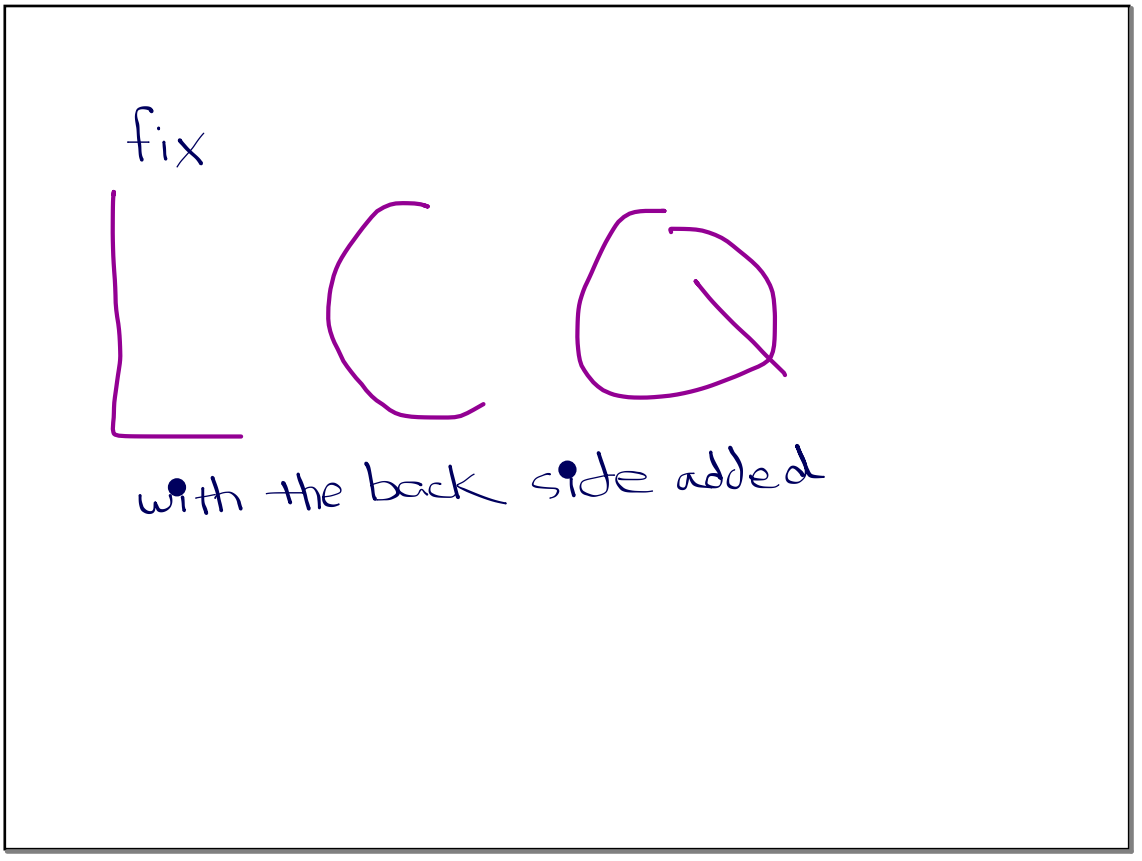
Step 6: p-value from GDC = .0313

Since $p = .0313$ is less than .05, reject H_0.

there is evidence to support the claim that CHD and anger level are associated

How do you tell one bathroom full of statisticians from another?

Check the p-value



(b) Using the same data calculate the Pearson's product moment correlation coefficient quickly using your calculator.

(c) Now calculate the Pearson's correlation coefficient again, this time by "hand" using the formula. Show all three critical values and then the final answer

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2 \cdot \Sigma(y-\bar{y})^2}} = \frac{\qquad\qquad}{\sqrt{\qquad\qquad\cdot\qquad\qquad}} =$$

(d) Using this correlation coefficient, comment on the correlation.

Assignment: Ch.11 Packet

1. p. 337.... 2abc
2. p. 341.... #1 (use X^2 statistic)
3. p. 341....#3 (Use P-value, method B)