Lecture 20b: Practice Problems for Lecture 20 "Chi-Square"

Everything that appears in these lecture notes is fair game for the test. They are the best "study guide" I can provide. It is impossible to provide a "list" that is more comprehensive than the lecture notes above. However, here are a few additional practice exercises or practice concepts.

Goodness of Fit Chi-Square

Recall that a goodness of fit test requires one nominal (or ordinal) level variable. We can test whether or not the distribution is uniform (the expected counts of each variable are equal) or we can specify a customized distribution. In the following example we will test the easiest: whether or not the distribution is uniform:

H_o: The population distribution is uniform – an equal percentage of the population falls in each category of the variable
H_1: The population distribution is not uniform – **an unequal** percentage of the population falls in each category of the variable

So for the following problems n=10 and we will expect an equal number of people to fall in each category.

Perform the 7 steps to a Goodness of Fit Chi-Square for the problems below

1. type of computer system? α =.05

An administrator needs to upgrade the computers for her division. She wants to know what sort of computer system her workers prefer. She gives three choices: windows, mac, or no preference. Test the hypothesis or theory that an equal percentage of the population prefers each type of computer system . The data and SPSS output is below.

NOTE: An assumption of Chi-Square is violated in this test. See if you can spot it. Also, if you violate this assumption on a take home test w/ your data you will have to note this!

Type of computer	observed
windows	2
mac	6
linux	2

ANSWER BELOW

1. State null and alternative hypothesis.

H_o: The population distribution is uniform – an equal percentage of the population prefers each type of computer system
H_1: The population distribution is not uniform – **an unequal** percentage of the population prefers each type of computer system

2. State level of significance or a "alpha."

For this example we'll use alpha =.05

3. Determine the test distribution to use – Chi Square tests use X^2 distribution.

4. Define the rejection regions. And draw a picture!

df= k-1 (where k= # of categories in variable). In this case df = 3-1 = 2. Using Appendix 6 critical value df=1 area in tail - .05 = 5.99

5. State the decision rule.

Reject the null if the TR >5.99, otherwise FTR.

6. Perform necessary calculations on data and compute TR value.

TR= $\sum \frac{(O-E)^2}{E}$

O = # of observed cases E = # of expected cases

computer	observed	expected	(O-E)	(O-E)2	(O - E)2/E
mac	6	3.3	2.7	7.29	2.21
linux	2	3.3	-1.3	1.69	0.51
windows	2	3.3	-1.3	1.69	0.51
				sum=	**3.23**

SPSS output

system pref

	Observed N	Expected N	Residual
windows	2	3.3	-1.3
mac	6	3.3	2.7
linux	2	3.3	-1.3
Total	10		

Test Statistics

	system pref
Chi-Square	3.200[a]
df	2
Asymp. Sig.	.202

a. 3 cells (100.0%) have expected frequencies less than 5. The minimum expected cell frequency is 3.3.

7. Compare TR value with the decision rule and make a statistical decision. (Write out decision in English! -- my addition)

In this case TR falls in FTR region and conclude the population distribution is uniform – there is insufficient evidence to reject the theory that the an equal percentage of the population prefers each type of computer system. NOTE: we do not state a level of confidence in this statement because we failed to reject the null. Only when you REJECT the null can you state a level of confidence.
Note the p-value =.202 or 20.2%

Note that the assumption violated was that the expected counts in each cell are NOT 5. They are supposed to be at least 5 for any chi-square.

2. <u>Prior DUI Convictions?</u>α =.05

An administrator working for the police department wants to know whether or not the people arrested for DUI are likely to have prior DUI convictions on their record as do not. The data and SPSS output is below.

data points	observed
prior	3
no priors	7

ANSWER BELOW

1. State null and alternative hypothesis.

H_o: The population distribution is uniform – an equal percentage of the population have DUI prior convictions as do not

H_1: The population distribution is not uniform – **<u>an unequal</u>** percentage of the population have DUI prior convictions as do not

2. State level of significance or a "alpha."

For this example we'll use alpha =.05

3. Determine the test distribution to use – Chi Square tests use X^2 distribution.

4. Define the rejection regions. And draw a picture!

df= k-1 (where k= # of categories in variable). In this case df = 2-1 = 1. Using Appendix 6 critical value df=1 area in tail - .05 = 3.84

5. State the decision rule.

Reject the null if the TR >3.84, otherwise FTR.

6. Perform necessary calculations on data and compute TR value.

TR= $\sum \frac{(O-E)^2}{E}$

O = # of observed cases E = # of expected cases

data points	observed	expected	(O-E)	(O-E)2	(O - E)2/E
prior	3	5	-2	4	0.80
no priors	7	5	2	4	0.80
				sum=	**1.60**

SPSS output for goodness of fit

prior DUI? 1=y 0-no

	Observed N	Expected N	Residual
no	7	5.0	2.0
yes	3	5.0	-2.0
Total	10		

Test Statistics

	prior DUI? 1=y 0-no
Chi-Square	1.600[a]
df	1
Asymp. Sig.	.206

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 5.0.

7. Compare TR value with the decision rule and make a statistical decision. (Write out decision in English! -- my addition)

In this case TR falls in FTR region and conclude the population distribution is uniform – there is insufficient evidence to reject the theory that the percentage of the population has an equal number of people with DUI prior convictions as do not. NOTE: we do not state a level of confidence in this statement because we failed to reject the null. Only when you REJECT the null can you state a level of confidence.

Note the p-value =.206 or 20.6%

3. Prior Drug Convictions? α =.05

An administrator working for the police department wants to know whether or not the people arrested for DUI are likely to have prior drug convictions on their record as do not. The data and SPSS output is below.

data points	observed
prior	2
no priors	8

ANSWER BELOW

1. State null and alternative hypothesis.

H_o: The population distribution is uniform – an equal percentage of the population have prior drug convictions as do not

H_1: The population distribution is not uniform – **an unequal** percentage of the population have prior drug convictions as do not

2. State level of significance or a "alpha."

For this example we'll use alpha =.05

3. Determine the test distribution to use – Chi Square tests use X^2 distribution.

4. Define the rejection regions. And draw a picture!

df= k-1 (where k= # of categories in variable). In this case df = 2-1 = 1. Using Appendix 6 critical value df=1 area in tail - .05 = 3.84

5. State the decision rule.

Reject the null if the TR >3.84, otherwise FTR.

6. Perform necessary calculations on data and compute TR value.

TR= $\sum \frac{(O-E)^2}{E}$

O = # of observed cases   E = # of expected cases

data points	observed	expected	(O-E)	(O-E)2	(O - E)2/E
prior	2	5	-3	9	1.80
no priors	8	5	3	9	1.80
				sum=	**3.60**

SPSS output

prior Drug offense 1=y 0=n

	Observed N	Expected N	Residual
no	8	5.0	3.0
yes	2	5.0	-3.0
Total	10		

Test Statistics

	prior Drug offense 1=y 0=n
Chi-Square	3.600[a]
df	1
Asymp. Sig.	.058

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 5.0.

7. Compare TR value with the decision rule and make a statistical decision. (Write out decision in English! -- my addition)

In this case TR falls in FTR region and conclude the population distribution is uniform – there is insufficient evidence to reject the theory that the percentage of the population has an equal number of people with prior drug convictions as do not. NOTE: we do not state a level of confidence in this statement because we failed to reject the null. Only when you REJECT the null can you state a level of confidence.

Note the p-value =.058 or 5.8%

4. **prefer 4 day work week?** α **=.05**

An administrator working for the prisons wants to know if prison guards show a preference for a 4 day work week (4 x 10 hr days) vs. a traditional five day work week (5 x 8 hour days). Test the hypothesis or theory that an equal percentage of the populations prefers a 4 day work week as do not. The data and SPSS output is below.

Like 4 days	observed
yes	9
no	1

ANSWER BELOW

1. State null and alternative hypothesis.

H_o: The population distribution is uniform – an equal percentage of the population prefers a 4 day work week as do not

H_1: The population distribution is not uniform – **an unequal** percentage of the population prefers a 4 day work week as do not

2. State level of significance or a "alpha."

For this example we'll use alpha =.05

3. Determine the test distribution to use – Chi Square tests use X^2 distribution.

4. Define the rejection regions. And draw a picture!

df= k-1 (where k= # of categories in variable). In this case df = 2-1 = 1. Using Appendix 6 critical value df=1 area in tail - .05 = 3.84

5. State the decision rule.

Reject the null if the TR >3.84, otherwise FTR.

6. Perform necessary calculations on data and compute TR value.

TR= $\sum \frac{(O-E)^2}{E}$

O = # of observed cases E = # of expected cases

like 4 day	observed	expected	(O-E)	(O-E)2	(O - E)2/E
yes	9	5	4	16	3.20
no	1	5	-4	16	3.20
				sum=	**6.40**

SPSS output

prefer 4 day week?

	Observed N	Expected N	Residual
no	1	5.0	-4.0
yes	9	5.0	4.0
Total	10		

Test Statistics

	prefer 4 day week?
Chi-Square	6.400[a]
df	1
Asymp. Sig.	.011

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 5.0.

7. Compare TR value with the decision rule and make a statistical decision. (Write out decision in English! -- my addition)

In this case TR falls in rejection region and conclude the population distribution is NOT uniform – there is sufficient evidence to reject the theory at the 95% confidence level that the percentage of the population has an equal number of people who prefer a 4 day work wee as do not. Another way to say this is you could be 95% confident that the percentage of the population who prefers a 4 day work is not equal to the percentage of the population that do not. Or a third way to say it (there are many many ways to say it) is you could be 95% confident that the percentage of the population who prefers a 4 day work week and the percentage of the population who does not prefer a 4 day work week are NOT equal.

Note the p-value =.011 or 1.1%

5. <u>Political Party: Democrat &Republican</u> α =.05

A public administrator wants to figure out if there are an equal number of democrats and republicans appointed to key positions in the state. She takes a random sample of 10 people (n=10). The data and SPSS output is below.

data points	observed
dem	4
rep	6

ANSWER BELOW

1. **State null and alternative hypothesis**.

H_o: The population distribution is uniform – an equal percentage of the population are democrats and republicans

H_1: The population distribution is not uniform – **<u>an unequal</u>** percentage of the population are democrats and republicans

2. **State level of significance or a "alpha."**

For this example we'll use alpha =.05

3. **Determine the test distribution to use –** Chi Square tests use X^2 distribution.

4. **Define the rejection regions. And draw a picture!**

df= k-1 (where k= # of categories in variable). In this case df = 2-1 = 1. Using Appendix 6 critical value df=1 area in tail - .05 = 3.84

5. **State the decision rule.**

Reject the null if the TR >3.84, otherwise FTR.

6. **Perform necessary calculations on data and compute TR value.**

TR= $\sum \frac{(O-E)^2}{E}$

O = # of observed cases E = # of expected cases

data points	observed	expected	(O-E)	(O-E)2	(O - E)2/E
dem	4	5	-1	1	0.20
rep	6	5	1	1	0.20
				sum=	**0.40**

SPSS output for goodness of fit

dem=1 rep=2

	Observed N	Expected N	Residual
democrat	4	5.0	-1.0
republican	6	5.0	1.0
Total	10		

Test Statistics

	dem=1 rep=2
Chi-Square	.400[a]
df	1
Asymp. Sig.	.527

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 5.0.

7. **Compare TR value with the decision rule and make a statistical decision. (Write out decision in English! -- my addition)**

In this case TR falls in FTR region and conclude the population distribution is uniform – there is insufficient evidence to reject the theory that the percentage of the population has an equal number of democrats as republicans. NOTE: we do not state a level of confidence in this statement because we failed to reject the null. Only when you REJECT the null can you state a level of confidence.
Note the p-value =.527 or 52.7%

5. type of drug arrested for

We could look at "market share" of felons in prison for drug dealing. Pretend want to measure policing of drug crimes and we want to know what sorts of drug dealers are being caught? Pretend the mayor has asked the cops to concentrate on arresting methamphetamine dealers. So we could test the theory that the police arresting an equal number of Ice (or methamphetamine), marijuana, and crack dealers. If so then about 33% of the arrests should be for crack, 33% for Ice, and 33% for marijuana. If we can prove this theory wrong, the police are NOT arresting all types of drug dealers equally.

Say, that our population is felony drug arrests on Oahu, Hawaii. We'll just see if the market share is equal or not: that means we would expect the distribution to be approx 33.3% for each type of drug (33.3 x 3= approx 100).

We can call the variable *DRUG* and take a random sample of n=75 and find out the following:
DRUG – type of drug person was caught selling
1 = Ice
2 = Crack
3 = Marijuana

Drug	# of persons preferring it
Ice	28
Crack	25
Marijuana	22

In plain English we will test the theory that the population distribution is uniform – an equal percentage of the population arrested for drug dealing was caught selling each type of drug. Or we will try to prove the population distribution is not uniform – an unequal percentage of the population arrested for drug dealing was caught selling each type of drug.

1State null and alternative hypothesis.
H_o: The population distribution is uniform – an equal percentage of the population arrested for drug dealing was caught selling each type of drug

H_1: The population distribution is not uniform – **an unequal** percentage of the population arrested for drug dealing was caught selling each type of drug.

2State level of significance or a "alpha."

For this example we'll use alpha =.05

3Determine the test distribution to use – Chi Square tests use X^2 distribution.

4Define the rejection regions. And draw a picture!

df= k-1 (where k= # of categories in variable). In this case df = 3-1 = 2. Using Appendix 6 critical value df=2 area in tail - .05 = 5.99

5State the decision rule.

Reject the null if the TR >5.99, otherwise FTR.

6Perform necessary calculations on data and compute TR value.

$$TR=\sum \frac{(O-E)^2}{E}$$

O = # of observed cases E = # of expected cases

note how the TR formula above looks *a bit* like a variance formula

$\sigma^2 = \frac{\sum (x-\mu)^2}{N} =$ This (population) variance formulas sort of asks "on average how much to all of these x's differ from the mean?"

Well the TR formula above essentially asks "on average how much do our observations differ from what our theory expected?" If they differ a lot from the theory, the we will reject the theory. If the do not differ a lot from the theory, then the theory is probably correct, or at least we do not have sufficient evidence to reject the theory.

Large differences between expected and observed values will tend to have larger numerators in the fraction and thus larger TR values and thus "tend towards significance."

What happens when the number on the top of the fraction gets larger relative the the number on the bottom of the fraction? The number represented by the fraction "gets bigger:"

1/10 2/10 3/10 5/10 10/10 20/10 note how the numbers get bigger as the top number of the fraction grows larger.

Test Ratio Computation					
DRUG	**Observed**	**Expected**	**O - E**	**$(O - E)^2$**	**$(O - E)^2/E$**
Ice	28	25	3	9	.36
Crack	25	25	0	0	0

Marijuana	22	25	-3	9	.36
Σ (sum)	**80**	**75**			**.72**

TR = .72

7Compare TR value with the decision rule and make a statistical decision. (Write out decision in English! -- my addition)
In this case TR is less than 5.99 and therefore we FTR and conclude the population distribution is uniform – there is insufficient evidence to reject the theory that the percentage of the population arrested for selling each type of drug is equal. NOTE: we do not state a level of confidence in this statement because we failed to reject the null. Only when you REJECT the null can you state a level of confidence.

Using SPSS

Below is the output for these data using SPSS. See lecture 20c_SPSS.pdf for how to have SPSS create this output.

type of drug

	Observed N	Expected N	Residual
Ice	28	25.0	3.0
Crack	25	25.0	.0
marijuana	22	25.0	-3.0
Total	75		

Test Statistics

	type of drug
Chi-Square	.720[a]
df	2
Asymp. Sig.	.698

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 25.0.

Test of Independence (or Contingency Table) Chi-Square

Recall that a test of independence chi-square requires two minal (or ordinal) level variables. We can test whether or not the two variables are independent. In the following example we will test the easiest: whether or not the distribution is uniform:

H_0: The two variables are ***independent*** of each other
H_1: The two variables are ***dependent on or related to*** each other

So for the following problems n=10 and we will expect an equal number of people to fall in each category.

Perform the 7 steps to a Test of Independence (or Contingency Table) Chi-Square for the problems below

1. gender and political party where <u>alpha =.01</u>

Pretend a person working for a politician wants to see if gender and political party are related to each other. She has a random sample of 10 people with the following data. ***NOTE: An assumption of Chi-Square is violated in this test. See if you can spot it. Also, if you violate this assumption on a take home test w/ your data you will have to note this!***

gender m=1 f=2 * dem=1 rep=2 Crosstabulation

			dem=1 rep=2		
			democrat	republican	Total
gender m=1 f=2	male	Count	2	1	3
		Expected Count	1.8	1.2	3.0
	female	Count	4	3	7
		Expected Count	4.2	2.8	7.0
	Total	Count	6	4	10
		Expected Count	6.0	4.0	10.0

1. State null and alternative hypothesis.

H_o: Gender is ***independent*** of one's political party.

H_1: Gender is ***dependent on or related to*** one's political party.

2. State level of significance or a "alpha."

For this example we'll use <u>**alpha =.01**</u>

3. Determine the test distribution to use – Chi Square tests use X^2 distribution.

4. Define the rejection regions. And draw a picture!

df= (r-1) (c -1) (where r= # of rows in table and c= # of columns in the table). In this case df = (2-1) (2-1) = 1. Using Appendix 6 critical value df=1 area in tail - .01 = 6.63

5. State the decision rule.

Reject the null if the TR > 6.63, otherwise FTR.

6. Perform necessary calculations on data and compute TR value.

gender m=1 f=2 * dem=1 rep=2 Crosstabulation

			dem=1 rep=2		
			democrat	republican	Total
gender m=1 f=2	male	Count	2	1	3
		Expected Count	1.8	1.2	3.0
	female	Count	4	3	7
		Expected Count	4.2	2.8	7.0
	Total	Count	6	4	10
		Expected Count	6.0	4.0	10.0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.079[a]	1	.778		
Continuity Correction[b]	.000	1	1.000		
Likelihood Ratio	.080	1	.777		
Fisher's Exact Test				1.000	.667
Linear-by-Linear ...	.071	1	.789		
N of Valid Cases	10				

a. 4 cells (100.0%) have expected count less than 5. The minimum expected count is 1.20.

TR= .079

7. Compare TR value with the decision rule and make a statistical decision. (Write out decision in English! -- my addition)

Since TR is falls in the FTR region. In English: There is insufficient evidence to reject the theory that gender and political party are independent.

p-value =.778 indicating that if you were to conclude that the two variables were related to (or dependent upon each other) you'd have to accept a 77.8% chance of error.

Note that the assumption violated was that the expected counts in each cell are NOT 5. They are supposed to be at least 5 for any chi-square.

2. gender and computer system where alpha =.05

There are 2 categories in gender and 3 in computer system

Pretend a person working for a politician wants to see if gender and type of computer system preferred are related to each other. (I'm sorry but I can't think of a good reason why the would be, but I just wanted to have the df in step 3 and 4 be a bit different for practice.) She has a random sample of 10 people with the following data. ***NOTE: An assumption of Chi-Square is violated in this test. See if you can spot it. Also, if you violate this assumption on a take home test w/ your data you will have to note this!***

gender m=1 f=2 * system pref Crosstabulation

			system pref			
			windows	mac	linux	Total
gender m=1 f=2	male	Count	1	1	1	3
		Expected Count	.6	1.8	.6	3.0
	female	Count	1	5	1	7
		Expected Count	1.4	4.2	1.4	7.0
	Total	Count	2	6	2	10
		Expected Count	2.0	6.0	2.0	10.0

1. State null and alternative hypothesis.

H_o: Gender is ***independent*** of type of computer system preferred.

H_1: Gender is ***dependent on or related to*** type of computer system preferred.

2. State level of significance or a "alpha."

For this example we'll use **alpha =.05**

3. Determine the test distribution to use – Chi Square tests use X^2 distribution.

4. Define the rejection regions. And draw a picture!

df= (r-1) (c -1) (where r= # of rows in table and c= # of columns in the table). In this case df = (2-1) (3-1) = 2. Using Appendix 6 critical value df=2 area in tail - .05 = 5.99

5. State the decision rule.

Reject the null if the TR > 5.99, otherwise FTR.

6. Perform necessary calculations on data and compute TR value.

gender m=1 f=2 * system pref Crosstabulation

			system pref			
			windows	mac	linux	Total
gender m=1 f=2	male	Count	1	1	1	3
		Expected Count	.6	1.8	.6	3.0
	female	Count	1	5	1	7
		Expected Count	1.4	4.2	1.4	7.0
	Total	Count	2	6	2	10
		Expected Count	2.0	6.0	2.0	10.0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	1.270[a]	2	.530
Likelihood Ratio	1.265	2	.531
Linear-by-Linear ...	.000	1	1.000
N of Valid Cases	10		

a. 6 cells (100.0%) have expected count less than 5. The minimum expected count is .60.

TR= 1.270

7. Compare TR value with the decision rule and make a statistical decision. (Write out decision in English! -- my addition)

Since TR is falls in the FTR region. In English: There is insufficient evidence to reject the theory that gender and type of computer system are independent.

p-value =.530 indicating that if you were to conclude that the two variables were related to (or dependent upon each other) you'd have to accept a 53% chance of error.

Note that the assumption violated was that the expected counts in each cell are NOT 5. They are supposed to be at least 5 for any chi-square.

3. gender and opinion on abortion where <u>alpha =.01</u>

Pretend a person working for a politician wants to see if gender and opinion on abortion are related to each other. She has a random sample of 10 people with the following data. ***NOTE: An assumption of Chi-Square is violated in this test. See if you can spot it. Also, if you violate this assumption on a take home test w/ your data you will have to note this!***

gender m=1 f=2 * opinion on abortion Crosstabulation

			opinion on abortion		
			legal	illegal	Total
gender m=1 f=2	male	Count	1	2	3
		Expected Count	1.2	1.8	3.0
	female	Count	3	4	7
		Expected Count	2.8	4.2	7.0
	Total	Count	4	6	10
		Expected Count	4.0	6.0	10.0

1. State null and alternative hypothesis.

H_o: Gender is ***independent*** of one's opinion about abortion.
H_1: Gender is ***dependent on or related to*** one's political party.

2. State level of significance or a "alpha."

For this example we'll use **<u>alpha =.01</u>**

3. Determine the test distribution to use – Chi Square tests use X^2 distribution.

4. Define the rejection regions. And draw a picture!

df= (r-1) (c -1) (where r= # of rows in table and c= # of columns in the table). In this case df = (2-1) (2-1) = 1. Using Appendix 6 critical value df=1 area in tail - .01 = 6.63

5. State the decision rule.

Reject the null if the TR > 6.63, otherwise FTR.

6. Perform necessary calculations on data and compute TR value.

gender m=1 f=2 * opinion on abortion Crosstabulation

			opinion on abortion		
			legal	illegal	Total
gender m=1 f=2	male	Count	1	2	3
		Expected Count	1.2	1.8	3.0
	female	Count	3	4	7
		Expected Count	2.8	4.2	7.0
	Total	Count	4	6	10
		Expected Count	4.0	6.0	10.0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.079[a]	1	.778		
Continuity Correction[b]	.000	1	1.000		
Likelihood Ratio	.080	1	.777		
Fisher's Exact Test				1.000	.667
Linear-by-Linear ...	.071	1	.789		
N of Valid Cases	10				

a. 4 cells (100.0%) have expected count less than 5. The minimum expected count is 1.20.

b. Computed only for a 2x2 table

TR= .079

7. Compare TR value with the decision rule and make a statistical decision. (Write out decision in English! -- my addition)

Since TR is falls in the FTR region. In English: There is insufficient evidence to reject the theory that gender and opinion about abortion are independent.

p-value =.778 indicating that if you were to conclude that the two variables were related to (or dependent upon each other) you'd have to accept a 77.8% chance of error.

Note that the assumption violated was that the expected counts in each cell are NOT 5. They are supposed to be at least 5 for any chi-square.

4. Political party and opinion on abortion where <u>alpha =.01</u>

Pretend a person working for a politician wants to see if political party and opinion on abortion are related to each other. She has a random sample of 10 people with the following data. ***NOTE: An assumption of Chi-Square is violated in this test. See if you can spot it. Also, if you violate this assumption on a take home test w/ your data you will have to note this!***

dem=1 rep=2 * opinion on abortion Crosstabulation

			opinion on abortion		
			legal	illegal	Total
dem=1 rep=2	democrat	Count	3	3	6
		Expected Count	2.4	3.6	6.0
	republican	Count	1	3	4
		Expected Count	1.6	2.4	4.0
	Total	Count	4	6	10
		Expected Count	4.0	6.0	10.0

1. State null and alternative hypothesis.

H_o: Political party is ***independent*** of one's opinion about abortion.

H_1: Political party is ***dependent on or related to*** one's opinion about abortion.

2. State level of significance or a "alpha."

For this example we'll use **<u>alpha =.01</u>**

3. Determine the test distribution to use – Chi Square tests use X^2 distribution.

4. Define the rejection regions. And draw a picture!

df= (r-1) (c -1) (where r= # of rows in table and c= # of columns in the table). In this case df = (2-1) (2-1) = 1. Using Appendix 6 critical value df=1 area in tail - .01 = 6.63

5. State the decision rule.

Reject the null if the TR > 6.63, otherwise FTR.

6. Perform necessary calculations on data and compute TR value.

dem=1 rep=2 * opinion on abortion Crosstabulation

			opinion on abortion		
			legal	illegal	Total
dem=1 rep=2	democrat	Count	3	3	6
		Expected Count	2.4	3.6	6.0
	republican	Count	1	3	4
		Expected Count	1.6	2.4	4.0
	Total	Count	4	6	10
		Expected Count	4.0	6.0	10.0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.625[a]	1	.429		
Continuity Correction[b]	.017	1	.895		
Likelihood Ratio	.644	1	.422		
Fisher's Exact Test				.571	.452
Linear-by-Linear ...	.563	1	.453		
N of Valid Cases	10				

a. 4 cells (100.0%) have expected count less than 5. The minimum expected count is 1.60.

b. Computed only for a 2x2 table

TR= .625

7. Compare TR value with the decision rule and make a statistical decision. (Write out decision in English! -- my addition)

Since TR is falls in the FTR region. In English: There is insufficient evidence to reject the theory that political party and opinion about abortion are independent.

p-value =.429 indicating that if you were to conclude that the two variables were related to (or dependent upon each other) you'd have to accept a 42.9% chance of error.

Note that the assumption violated was that the expected counts in each cell are NOT 5. They are supposed to be at least 5 for any chi-square.

5. Political party and opinion on whether DUI laws are too weak or soft on criminals where <u>alpha =.05</u>

Pretend a person working for a politician wants to see if political party and opinion whether or not DUI laws are too weak or soft of DUI criminals are related to each other. She has a random sample of 10 people with the following data. ***NOTE: An assumption of Chi-Square is violated in this test. See if you can spot it. Also, if you violate this assumption on a take home test w/ your data you will have to note this!***

dem=1 rep=2 * opinion: DUI laws weak? Crosstabulation

			opinion: DUI laws weak?			
			Disagree	nuetral	agree	Total
dem=1 rep=2	democrat	Count	2	3	1	6
		Expected Count	1.2	1.8	3.0	6.0
	republican	Count	0	0	4	4
		Expected Count	.8	1.2	2.0	4.0
	Total	Count	2	3	5	10
		Expected Count	2.0	3.0	5.0	10.0

1. State null and alternative hypothesis.

H_o: Political party is ***independent*** of one's opinion about DUI laws.
H_1: Political party is ***dependent on or related to*** one's opinion about DUI laws.

2. State level of significance or a "alpha."

For this example we'll use **<u>alpha =.05</u>**

3. Determine the test distribution to use – Chi Square tests use X^2 distribution.

4. Define the rejection regions. And draw a picture!

df= (r-1) (c -1) (where r= # of rows in table and c= # of columns in the table). In this case df = (2-1) (3-1) = 2. Using Appendix 6 critical value df=2 area in tail - .05 = 5.99

5. State the decision rule.

Reject the null if the TR > 5.99 otherwise FTR.

6. Perform necessary calculations on data and compute TR value.

dem=1 rep=2 * opinion: DUI laws weak? Crosstabulation

			opinion: DUI laws weak?			
			Disagree	nuetral	agree	Total
dem=1 rep=2	democrat	Count	2	3	1	6
		Expected Count	1.2	1.8	3.0	6.0
	republican	Count	0	0	4	4
		Expected Count	.8	1.2	2.0	4.0
	Total	Count	2	3	5	10
		Expected Count	2.0	3.0	5.0	10.0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	6.667[a]	2	.036
Likelihood Ratio	8.456	2	.015
Linear-by-Linear ...	4.820	1	.028
N of Valid Cases	10		

a. 6 cells (100.0%) have expected count less than 5. The minimum expected count is .80.

TR= 6.667

7. Compare TR value with the decision rule and make a statistical decision. (Write out decision in English! -- my addition)

Since TR is falls in the rejection region. In English: There is sufficient evidence to reject the theory that political party and opinion about DUI laws are independent. So we conclude, with 95% confidence that political party and opinion about DUI laws are dependent or related to each other.

p-value =.036 indicating that if you were to conclude that the two variables were related to (or dependent upon each other) you'd have to accept a 3.6% chance of error.

Note that the assumption violated was that the expected counts in each cell are NOT 5. They are supposed to be at least 5 for any chi-square.

6. where you live and where you surf alpha =.05

Pretend the C&C of Honolulu wants to make their lifeguards more effective at preventing ocean rescues. They have lifeguards patrol the beaches on days with dangerous surf and counsel some people not to enter the water. It is a preventative measure. For every person they persuade not to enter the water, that is potential ocean rescues avoided. So they think that bodyboarders will tend to listen best to lifeguards who are bodyborders, and that shortboardes will tend to listen best to lifeguards who also ride shortboards. Finally, they think longboardes will tend to listen best to lifeguards who also ride longboards.

There are four lifeguarding districts on Oahu that correspond to each side of the island: North Shore, West Shore, South Shore and East Shore.
So if a certain side of the island has more shortboarders, then they want to stations more shortboarding lifeguards in that part of the island. So they are going to do a study of surfers to see if side of the island where people surf (the neighborhood if you will) is related to or dependent upon their favorite way to surf (longboard, shortboard, or boogieboard).

So, in plain English, the administrator working for the lifeguards will do a Contingency Table Chi Square (or a Chi-Square Test of Independence) to test the theory that side of the island where people surf (the neighborhood if you will) is NOT related to or independent upon their favorite way to surf (longboard, shortboard, or boogieboard). Or conversely, the administrator will attempt do the chi-square test to prove that side of the island where people surf (the neighborhood if you will) is related to or dependent upon their favorite way to surf (longboard, shortboard, or boogieboard).

So we have two variables:
Hood – where do you like to surf on the island?
1=south shore 2= North Shore 3=West Side
Favorite -- Favorite way to surf
1 = shortboard 2=bodyboard/body surf 3= longboard

Say the two variables “cross-tabulate” like this:

Favorite	South shore	North Shore	West Side	row totals
shortboard	33	75	20	128
bodyboard	67	25	45	137
longboard	20	20	55	95
column totals	120	120	120	360

1. **State null and alternative hypothesis**.

H_o: The favorite way to surf is ***independent*** of where one lives on the island.
H_1: The favorite way to surf is ***dependent on or related to*** where one lives on the island.

2. **State level of significance or a “alpha.”**

For this example we’ll use alpha =.05

3. **Determine the test distribution to use –** Chi Square tests use X^2 distribution.

4. **Define the rejection regions. And draw a picture!**

df= (r-1) (c -1) (where r= # of rows in table and c= # of columns in the table). In this case df = (3-1) (3-1) = 4. Using Appendix 6 critical value df=4 area in tail - .05 = 9.49

5. **State the decision rule.**

Reject the null if the TR > 9.49, otherwise FTR.

6. **Perform necessary calculations on data and compute TR value.**

TR= $\sum \frac{(O-E)^2}{E}$

O = # of observed cases   E = # of expected cases

Again, note how the TR formula above looks *a bit* like a variance formula

$\sigma^2 = \frac{\sum (x-\mu)^2}{N}$ = This (population) variance formulas sort of asks "on average how much to all of these x's differ from the mean?"

Well the TR formula above essentially asks "on average how much do our observations differ from what our theory expected?" If they differ a lot from the theory, the we will reject the theory. If the do not differ a lot from the theory, then the theory is probably correct, or at least we do not have sufficient evidence to reject the theory.

Large differences between expected and observed values will tend to have larger numerators in the fraction and thus larger TR values and thus "tend towards significance."

What happens when the number on the top of the fraction gets larger relative the the number on the bottom of the fraction? The number represented by the fraction "gets bigger:"

1/10 2/10 3/10 5/10 10/10 20/10 note how the numbers get bigger as the top number of the fraction grows larger.

E = # of expected cases [E=(row total) (column total)/grand total]

Favorite	South shore	Country	West Side	row totals
short (O)	33	75	20	128
(E)	43	43	43	
body(O)	67	25	45	137
(E)	46	46	46	
long (O)	20	20	55	95
(E)	32	32	32	
column totals	120	120	120	360

row-col (cell)	**Observed**	**Expected**	**O -E**	**(O - E)2**	**(O - E)2/E**
1-1	33	43	-10	100	2.3
1-2	75	43	32	1024	23.8

1-3	20	43	-23	529	12.3
2-1	67	46	21	441	9.6
2-2	25	46	-21	441	9.6
2-3	45	46	-1	1	0.0
3-1	20	32	-12	144	4.5
3-2	20	32	-12	144	4.5
3-3	55	32	23	529	16.5
				$\sum=$	**83.2**

TR= 83.2

7. **Compare TR value with the decision rule and make a statistical decision. (Write out decision in English! -- my addition)**

Since TR is greater than 9.49 we reject null and conclude alternative. In English: We say that favorite way to surf is dependent on or related to where one lives on the island.

Again, TR value is essentially a ratio that computes a variance or “an average of the sum of the squared deviations from the expected value” where large differences in the numerator tend to make the fraction (or ratio) large. Large differences thus tend towards significance.