Lecture 04 - Inference with Two Categorical Variables

Data Analysis in CJ (CJUS 6103)

Outline

    I. Contingency Tables
   II. Steps in a Hypothesis Test
  III. Chi-Square Test of Independence
  IV. Computational Formula for χ^2
   V. Measures of Association
  VI. Alternative Tests of Independence in 2x2 Tables

Contingency Tables

Contingency tables display the joint distribution of two categorical variables. We characterize contingency tables by the number of rows and columns that they have. The total number of cells in a contingency table is computed as RxC.

The cells representing the totals are referred to as marginals. In a $2x2$ contingency table, there are two row marginals (RM_1, RM_2) and two column marginals (CM_1, CM_2). One important quality of a contingency table is that $RM_1 + RM_2 = n$, and $CM_1 + CM_2 = n$.

A Simple Example - Are Juvenile and Adult Arrests Independent?

	Delinquent?		
Criminal?	No	Yes	Total
No	40	20	60
Yes	10	30	40
Total	50	50	100

Here we have a $2x2$ contingency table showing the joint distribution of juvenile delinquency and adult criminality. With this type of contingency table, we would be interested in knowing if being arrested as a juvenile is associated with being arrested as an adult. In the words of probability, we want to know if juvenile crime and adult crime are statistically independent, or if they are dependent. That is, we want to know if the probability of adult crime "depends" on whether or not someone committed crime as a juvenile. If these two variables are independent, knowing whether someone was arrested as a juvenile does not help us predict whether he or she is arrested as an adult. If, on the other hand, these two variables are dependent, knowing whether someone was arrested as a juvenile does help us predict whether he or she is arrested as an adult.

One way that we might assess the relationship between juvenile crime and adult crime is to compare the probability of adult arrest conditional on juvenile arrest. When we do this, we see that $p(C \mid ND) = 15/50 = .30$ of the people in our sample that were not arrested as juveniles were subsequently arrested as adults. We also see that $p(C \mid D) = 30/50 = .60$ of the people in our sample that were arrested as juveniles were also

1

arrested as adults. The total proportion of the sample arrested as an adult, $p(C) = .40$, is clearly different from these two conditional probabilities.

This is our first indication that there may be a statistical relationship between the variables. Using the conditional probabilities, we can begin to say something substantive about the association between juvenile crime and adult crime. Our figures suggest that there is indeed an association. The problem with just comparing proportions, however, is that we do not have a reliable decision rule by which to compare them. How large should the difference be before we can claim that there is a statistical association here?

Setting Up a Hypothesis Test

Hypothesis tests are a structured means to test whether a relationship exists between two variables. A hypothesis test includes five steps:

1. Formally State Hypotheses

There are two hypotheses that you must state formally, a null or statistical hypothesis (H_0) and an alternative or research hypothesis (H_1 or H_A). The null hypothesis, with some exceptions for the specific type of test being conducted, generally indicates that no relationship exists between the independent and dependent variables. By contrast, the alternative hypothesis states that some relationship exists and, for particular types of statistical tests, you may indicate the direction of the relationship (negative/positive).

2. Obtain a Probablity Distribution

Every result of a hypothesis test (also known as a *test statistic* is associated with a probability distribution, or what we will eventually refer to as a sampling distribution in a subsequent lecture. The important part to know right now is that a sampling distribution is a **theoretical** distribution that provides us the results from an infinite number of trials or samples. It takes on a particular shape (usually normal) and is associated with a range of probabilities, which help us make informed decisions about the null hypothesis. If the null hypothesis is true, we expect our test statistic to lie relatively close to the expected mean of the sampling distribution (where probability levels are high). If the alternative hypothesis is true, we expect the sample statistic to lie in the tails of the sampling distribution (where probability values are low).

There are multiple probability distributions I will discuss this semester - for this lecture, we will use the χ^2 distribution.

3. Make Decision Rules

Once we have our theoretical probability distribution, we need to select the level of significance and determine the critical region.

First, we must select a level of significance by stating in advance how often we are willing to falsely reject the null hypothesis. By convention, researchers select a significance level of 0.05 or 0.01, which we refer to as α. This indicates that we are willing to make a Type I error 5 times out of 100 (or 1 time out of 100). In other words, we are willing to falsely reject the null hypothesis 5% of the time.

Second, we must establish the critical region of our probability distribution. This is the area of our theoretical distribution that will lead us to reject the null hypothesis. The size of this area is determined by our significance level, α. If we were to state a non-directional alternative hypothesis, we would be conducting what is called a two-tailed significance test. In other words, we would have to split the significance level into two parts, one half (0.025) of it in each tail of the distribution. We would thus have two critical regions. If we were to state a directional alternative hypothesis, we would be conducting what we call a one-tailed significance test. In other words, we would have only one critical region. In the context of the χ^2 distribution, it may only be used by one-tailed significance tests because the values of χ^2 are always positive.

There is an important reason that we select a significance level and critical region for our probability distribution. We know that since we have data from a sample, and there is sample-to-sample variation, there is bound to be some error that is due to chance and sampling. So the question is this: How far away must we be from the null hypothesis before we decide that our data are not simply the result of chance? For particular types of hypothesis tests, The critical region tells us how many standard error units our sample statistic must be from the hypothesized mean of the sampling distribution for us to decide that it is not simply the result of chance. If our statistic falls within the critical region, we are pretty certain (95% certain,

in fact) that this reflects a true difference, and so we are willing to reject the null hypothesis and accept the alternative hypothesis in its place.

4. Calculate the Test Statistic

This is the actual quantity that we obtain from our sample data and use to test the null hypothesis. We want to use the value that we obtain from our test statistic and compare it to the probability distribution to determine how likely it is that we obtained the value by chance. In most cases, there will be a formula that we must calculate by hand to derive the test statistic. For the χ^2 distribution, this test statistic is:

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_{obs} - f_{exp})^2}{f_{exp}}$$

5. Make a Decision About the Null Hypothesis

If our test statistic falls within the critical region, we will "reject H_0" and accept the alternative hypothesis in its place. If our test statistic falls outside the critical region and lies instead in the acceptance region, we will "accept H_0."

Hypothesis Testing Errors

In hypothesis testing, there are two types of error that we must be aware of. A type I error (**false positive**) is when we reject the null hypothesis, but in fact it is true. A type II error (**false negative**) is when we accept the null hypothesis, but it in fact is false.

| | Decision | |
Reality	Accept H_0	Reject H_0
H_0 is true	Correct decision	Type I Error (α)
H_0 is false	Type II Error (β)	Correct decision

We can reduce the probability of a type I error by decreasing level of significance. You should recognize, however, that there is a hydraulic relationship between type I and type II errors. If we want to ensure a small probability of falsely rejecting the null hypothesis (say 0.001), we necessarily increase the probability of falsely accepting the null hypothesis. Thus, these two errors are inversely related.

χ^2 Test of Independence

We can turn to the laws of probability to conduct hypothesis tests with purely categorical data. To determine the degree of association between juvenile crime and adult crime, we want to know what the contingency table would look like if these two variables were statistically independent of one another. We already know that, under the multiplication rule of probability, the joint probability of two events is $p(A \cap B) = p(A)p(B \mid A)$. We also know that if two events are statistically independent, $p(B \mid A) = p(B)$, and as a result the joint probability simplifies to $p(A \cap B) = p(A)p(B)$.

We can use this information to compare our observed frequencies with what we would expect to observe if juvenile crime and adult crime were indeed statistically independent. The expected frequencies can be easily computed using the formula:

$$f_{exp} = np_{exp} = \frac{RM * CM}{n}$$

Cell	f_{obs}	f_{exp}
1	40	$\frac{60*50}{100} = 30$
2	20	$\frac{60*50}{100} = 30$
3	10	$\frac{40*50}{100} = 20$
4	30	$\frac{40*50}{100} = 20$
Total	100	100

We can easily see that these two variables are not statistically independent of one another, since $f_{obs} \neq f_{exp}$ for each of the four cells. At this point, we are only one step away from actually conducting a full hypothesis test with this information. In order to do this, we rely on a probability distribution known as the chi-square distribution.

χ^2 Test of Independence for Juvenile and Adult Arrests

Let's test the hypothesis that juvenile arrest is a predictor of adult arrest, in other words, that adult arrest is not independent of juvenile arrest. We will use $\alpha = .05$ (side note - I will always provide you with this value).

Step 1: Formally state hypotheses.

The null hypothesis with a χ^2 test is always that there is no association between two variables of interest. Symbolically, we write H_0: $\chi^2 = 0$. The alternative hypothesis is always that there is an association between the two variables, so we write H_1: $\chi^2 > 0$. The alternative hypothesis is always directional, because the chi-square distribution has only one tail.

Step 2. Obtain a Probability Distribution

Our probability distribution is the chi-square distribution. The probability function for χ^2 is computed as so (don't worry, you will not have to compute this):

$$p(x) = \frac{(1/2)^{k/2}}{\Gamma(k/2)} x^{k/(2-1)} e^{-x/2}$$

Where x represents a given χ^2 that you want to know the probability of, k represents the **degrees of freedom**, and $\Gamma(\cdot)$ represents the **Gamma function** (an extension of the factorial function (e.g., $z!$) to non-integer and complex numbers).
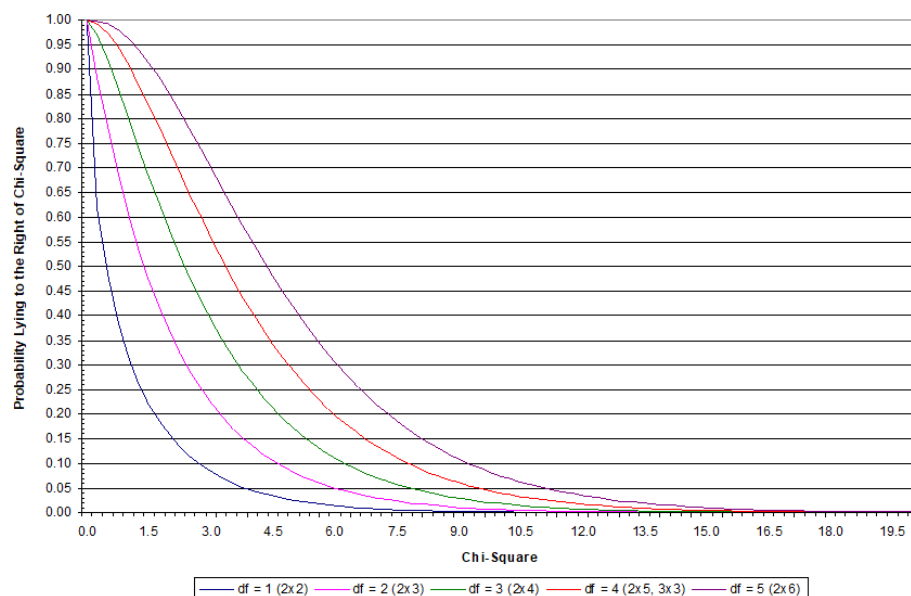
Step 3. Make Decision Rules

We have already set $\alpha = .05$. Now we need to determine the critical value - you can use the table in your textbook to identify this value or qchisq() function in R. To identify the critical value, you need to determine how many degrees of freedom you have for your hypothesis test. This is determined by the following calculation:

$$df = (\# \text{ of Rows} - 1) * (\# \text{ of Columns} - 1)$$

In this example, we have 1 degree of freedom $((2 - 1) * (2 - 1) = 1)$.

The following table displays χ^2 values for 1 to 5 degrees of freedom and for various α levels.

	Significance Level (α)			
Degrees of Freedom	.10	.05	.01	.001
1	2.706	3.841	6.635	10.828
2	4.605	5.991	9.21	13.816
3	6.251	7.815	11.345	16.266
4	7.779	9.488	13.277	18.467
5	9.236	11.07	15.086	20.515

Alternative Method to Find the Chi-Square Critical Value

An easy way to identify the χ^2 critical value is to use the qchisq() function in R.

The qchisq() function takes the following general form:

qchisq(p, df, lower.tail=FALSE)

Where p is your alpha level (.05, .01, or .001), df is your degrees of freedom, and lower.tail=FALSE tells R to return the χ^2 value necessary to reject the null hypothesis (i.e., the value that separates out .05 probability points at or above it).

Step 4. Calculate the Test Statistic

In this step, you must use the observed values in the contingency table to calculate a value for χ^2. We compute the test statistic using the following equation:

$$\chi^2 = \sum \frac{(f_{obs} - f_{exp})^2}{f_{exp}}$$

You will need to first compute the values for f_{exp} for each cell and then, similar to the calculation for a standard deviation, create a column for squared deviations $(f_{obs} - f_{exp})^2$ from those means. A notable difference is that the squared deviation is individually divided by the expected value for that cell (be careful not to sum the squared deviations together then divide by the summed expected frequencies - you will get an incorrect value that way).

Step 5. Make a Decision About the Null Hypothesis

For the χ^2 test our conclusion will either be to accept the null hypothesis or reject it. If we accept the null, we conclude that the variables are, indeed, independent of one another. If we reject the null, we may conclude that the variables are **not** independently distributed and some association exists between them. Unfortunately, we cannot say much more other than this in the absence of indicators for the strength of the association between the variables (I will talk about this later in this lecture).

Example - Juvenile and Adult Arrests

Step 1. Formally state hypotheses: H_0:$\chi^2 = 0$; H_1: $\chi^2 > 0$

Step 2. Choose a probability distribution: χ^2 distribution, df $= (2-1) * (2-1) = 1$

Step 3. Make decision rules: $\alpha = .05$; $\chi^2_{crit} = 3.841$; reject H_0 if the test statistic (TS) > 3.841.

Step 4. Calculate the test statistic:

You may do this one of two ways - using tables and manual calculations (in R or by hand) or inputting the data into R as a data frame (one row for each internal cell of the table, column and row marginals excluded). I'll show examples of both below and then I will present an automatic method that you can use to check your results from the first two methods.

Table Method - Calculations within a table

Cell	f_{obs}	f_{exp}	$f_{obs} - f_{exp}$	$(f_{obs} - f_{exp})^2$	$(f_{obs} - f_{exp})^2/f_{exp}$
11	40	30	10	100	3.33
12	20	30	-10	100	3.33
21	10	20	-10	100	5
22	30	20	10	100	5
	100	100			16.66

Though these appear as raw values, the .rmd file for these lecture notes show the actual contents of the cells before R computes each value. Below is an example where I have removed the grave accents in the final column. The quotation marks are just there to allow it to print without R yelling at me about needing more $ symbols in the expression.

Cell	f_{obs}	f_{exp}	$f_{obs} - f_{exp}$	$(f_{obs} - f_{exp})^2$	$(f_{obs} - f_{exp})^2/f_{exp}$
11	40	30	10	100	"r round$(((40 - ((6050)/100))^2)/((6050)/100), 2)$"
12	20	30	-10	100	"r round$(((20 - ((6050)/100))^2)/((6050)/100), 2)$"
21	10	20	-10	100	"r round$(((10 - ((4050)/100))^2)/((4050)/100), 2)$"
22	30	20	10	100	"r round$(((30 - ((4050)/100))^2)/((4050)/100), 2)$"
	100	100			16.66

R Method - via creating data in R

```r
cell<-c("11","12","21","22") # create character variable to indicate cell positions
obs<-c(40,20,10,30) # input observed frequencies
exp<-c((60*50)/100,(60*50)/100,(40*50)/100,(40*50)/100) # calculate expected frequencies
dev<-obs-exp # compute differences
dev_sq<-dev^2 # square the differences
chi<-dev_sq/exp # divide differences by expected frequencies
chi2<-sum(chi) # sum chi2 values for individual cells
chi2 # print out chi square value
```

```
## [1] 16.66667
```

Automatic Method

```r
juv_adult_arr<-matrix(c(40,10,20,30),ncol=2) # column values in pairs
# (top to bottom, left to right)
chisq.test(juv_adult_arr, correct=FALSE)# run chi-square test
```

```
##
##  Pearson's Chi-squared test
##
## data:  juv_adult_arr
## X-squared = 16.667, df = 1, p-value = 0.00004456
```

The *correct=FALSE* part is required to suppress a test correction that makes it harder to reject the null hypothesis. Normally, it is best to leave that as correct=TRUE in practice but for this example I wanted the χ^2 to be the same across all examples.

You cannot use this method for submitting your work, but can use it to check your math!

Final Step - Make a Decision About H_0

The obtained χ^2 value of 16.67 (or 16.667) exceeds the critical value of 3.841. Therefore, we reject H_0 and conclude that there is an association between juvenile and adult arrest.

Another Example - Military Service and Drug Use

Are military service and drug use independent?

| | **Military Service?** | | |
Use Drugs?	No ($\sim M$)	Yes (M)	Total
No ($\sim D$)	3426	407	3833
Yes (D)	629	108	737
Total	4055	515	4570

$p(D) = \frac{737}{4570} = 0.161$

$p(D \mid M) = \frac{108}{515} = 0.21$

The unconditional probability of drug use does not equal the probability of drug use conditional on military service which suggests that there is some association here.

Step 1. Formally state hypotheses: $H_0: \chi^2 = 0$; H_1: $\chi^2 > 0$

Step 2. Choose a probability distribution: χ^2 distribution, df $= (2-1) * (2-1) = 1$

Step 3. Make decision rules: $\alpha = .01$; $\chi^2_{crit} = 6.635$; reject H_0 if the test statistic (TS) > 6.635.

Step 4. Calculate the test statistic:

Table Method

Cell	f_{obs}	f_{exp}	$(f_{obs} - f_{exp})^2 / f_{exp}$
11	3426	3401.054	0.183
12	407	431.946	1.441
21	629	653.946	0.952
22	108	83.054	7.493
	4570	4570.00	10.069

Manual R Method

```r
cell<-c("11","12","21","22") # create character variable to indicate cell positions
obs<-c(3426,407,629,108) # input observed frequencies
exp<-c((3833*4055)/4570,(3833*515)/4570,(737*4055)/4570,(737*515)/4570)
# calculate expected frequencies
dev<-obs-exp # compute differences
dev_sq<-dev^2 # square the differences
chi<-dev_sq/exp # divide differences by expected frequencies
chi2<-sum(chi) # sum chi2 values for individual cells
chi2 # print out chi square value
```

```
## [1] 10.06838
```

Automatic R Method

```
drug_military<-matrix(c(3426,629,407,108),ncol=2) # column values in pairs
# (top to bottom, left to right)
chisq.test(drug_military, correct=FALSE)# run chi-square test
```

```
##
##  Pearson's Chi-squared test
##
## data:  drug_military
## X-squared = 10.068, df = 1, p-value = 0.001508
```

Final Step - Make a Decision About H_0

The obtained χ^2 value of 10.07 (or 10.068) exceeds the critical value of 6.635. Therefore, we reject H_0 and conclude that there is an association between military service and drug use.

Computational Formula for χ^2

Computational formula:

$$\chi^2 = \left(\sum_{k=1}^{k} \frac{f_{obs}^2}{f_{exp}} \right) - n$$

An advantage is that this formula does not require you to compute squared deviations

Caution Do not forget to subtract off n!!

Last Example - Employment and Delinquency

	Employment Status			
Delinquent Acts	0 Hours	1-20 Hours	21+ Hours	Total
0	3642	1605	1441	6688
1 - 4	637	374	427	1438
5+	318	201	289	808
Total	4597	2180	2157	8934

Research question: Is employment associated with delinquency?

Step 1. Formally state hypotheses: H_0:$\chi^2 = 0$; H_1: $\chi^2 > 0$

Step 2. Choose a probability distribution: χ^2 distribution, df $= (3-1)*(3-1) = 4$

Step 3. Make decision rules: $\alpha = .001$; $\chi^2_{crit} = 18.467$; reject H_0 if the test statistic (TS) > 18.467.

Step 4. Calculate the test statistic:

Cell	f_{obs}	f_{exp}	$f_{\text{obs}}^2/f_{\text{exp}}$
1	3642	3441.32	3854.38
2	1605	1631.95	1578.50
3	1714	1614.73	1285.96
4	1524	739.92	548.40
5	1179	350.89	398.63
6	671	347.19	525.16
7	221	415.76	243.23
8	109	197.16	204.91
9	8604	195.08	428.14
	8934	8934.00	9067.31

I just show the results here for the sake of demonstrating the difference in what columns your table should have - the values were all computed by hand.

Step 5: Make a decision about H_0 The TS is equal to $9067.31 - 8934 = 133.31$. Therefore, we reject H_0 and conclude that employment and delinquency are **not** independent of one another.

Measures of Association

One limitation with the χ^2 test is that it is sensitive to sample size. This means that with a large sample, we will be more likely to reject the null hypothesis. Thus, we utilize the χ^2 test only to inform us about whether there is a statistical relationship between two variables. In addition to this test, there are several measures of association that tell us about the strength of the relationship between two variables. A useful way to think about this is that chi-square tells us if there is a **statistically** significant relationship between two categorical variables, while the various measures of association tell us is if the relationship is **substantively** significant.

Phi ϕ

An important measure of association with 2×2 contingency tables is ϕ. One note of caution is that ϕ can only be used with 2×2 tables. The calculation for ϕ is simple:

$$\phi = \sqrt{\frac{\chi^2}{n}}$$

ϕ ranges from zero to one and has what we call an **explained variance** interpretation. If we multiply ϕ by 100, we can say that the independent variable explains XX% of the variance in the dependent variable.

Contingency (C) and Cramer's V (V)

Both Contingency (C) and Cramer's V (V) work for any size table. They are computed as follows:

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}}$$

$$V = \sqrt{\frac{\chi^2}{n * min(r - 1, c - 1)}}$$

Where r stands for rows and c stands for columns, and $min(r - 1, c - 1)$ returns the smaller value of the two. E.g., this value for a 2x3 contingency table would be 1 (rows-1 = 2-1 = 1; columns - 1 = 3-1 = 2). λ is a measure of proportionate reduction in error (PRE), telling us how better we can predict the DV using information about the IV.

Lambda λ

Lambda λ may also be used with any siz table, but the calculation is slightly more complicated.

$$\lambda = \sqrt{\frac{(\sum f_{IV}) - f_{DV}}{n - f_{DV}}}$$

Where f_{IV} are the largest f values in each category of the dependent variable (which should be the column variable), and f_{DV} is the largest row marginal of the dependent variable.

Note on Interpreting MOA Values

As a general rule, we interpret the strength of a relationship between two variables to be of weak (values of 0.00 to 0.29), moderate (values of 0.30 to 0.59), or strong (values of 0.60 to 1.00).

Measures of association for ordinal variables may also be positive or negative. In these cases, we may interpret negative values to mean that an **increase** in one variable is associated with a **decrease** in the other variable. The closer to -1 or 1 that the value gets, the stronger the relationship is. I do not use go over any ordinal measures of association in this lecture, but examples include Goodman-Kruskal's Gamma (γ) and Kendall's Tau-b.

Juvenile and Adult Arrest Example

	Delinquent?		
Criminal?	No	Yes	Total
No	40	20	60
Yes	10	30	40
Total	50	50	100

$$\phi = \sqrt{\frac{\chi^2}{n}} = \sqrt{\frac{16.667}{100}} = 0.408$$

$$V = \sqrt{\frac{\chi^2}{n * \min(r-1, c-1)}} = \sqrt{\frac{16.667}{100 * (2-1)}} = 0.408$$

ϕ and Cramer's V will always produce the same answer in a 2x2 table. We would say that the MOAs indicate a moderate relationship between juvenile and adult arrest. If we square the value ($0.408^2 = 0.166$) we can say that juvenile arrest explains 16.6% of the variance in adult arrest. In this specific example that applies to both ϕ and Cramer's V but that's only for a 2x2 table.

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}} = \sqrt{\frac{16.667}{100 + 16.667}} = 0.378$$

$$\lambda = \sqrt{\frac{(\sum f_{IV}) - f_{DV}}{n - f_{DV}}} = \sqrt{\frac{(40 + 30) - 60}{100 - 60}} = 0.5$$

C also indicates a moderate relationship, while λ suggests that we can reduce errors in predicting adult arrest by 50% if we were to use juvenile arrest to predict later adult arrests.

Military Service and Drug Use Example

	Military Service?		
Use Drugs?	No ($\sim M$)	Yes (M)	Total
No ($\sim D$)	3426	407	3833
Yes (D)	629	108	737
Total	4055	515	4570

$$V = \sqrt{\frac{\chi^2}{n * \min(r-1, c-1)}} = \sqrt{\frac{10.068}{4570 * (2-1)}} = 0.047$$

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}} = \sqrt{\frac{10.068}{4570 + 10.068}} = 0.047$$

$$\lambda = \sqrt{\frac{(\sum f_{IV}) - f_{DV}}{n - f_{DV}}} = \sqrt{\frac{(3426 + 407) - 3833}{4570 - 3833}} = 0$$

I skipped computing ϕ (it's the same as Cramer's V here). Based upon Cramer's V and C, we would conclude that the relationship between military service and drug use is weak. Consistent with this, λ suggests that using military service to predict drug use would not result in any reduction in prediction errors.

Employment and Delinquency Example

Delinquent Acts	Employment Status			
	0 Hours	1-20 Hours	21+ Hours	Total
0	3642	1605	1441	6688
1 - 4	637	374	427	1438
5+	318	201	289	808
Total	4597	2180	2157	8934

$$V = \sqrt{\frac{\chi^2}{n * \min(r-1, c-1)}} = \sqrt{\frac{133.31}{8934 * (3-1)}} = 0.086$$

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}} = \sqrt{\frac{133.31}{8934 + 133.31}} = 0.121$$

$$\lambda = \sqrt{\frac{(\sum f_{IV}) - f_{DV}}{n - f_{DV}}} = \sqrt{\frac{(3642 + 1605 + 1441) - 6688}{8934 - 6688}} = 0$$

The relationship between employment and delinquency is also weak and prediction errors cannot be reduced by using employment status to predict number of delinquent acts.

An Alternative Test of Independence in 2x2 Tables

A z-test for two proportions is an alternative to the χ^2 test when you want to assess if one proportion is not equal to, greater than, or lesser than some other proportion. This is the primary benefit of this test - it can be used for **directional hypotheses**.

The probability distribution for z appears in an appendix in your text (and, in general, can be easily found on the internet) but use this one instead for now.

	Significance Level (α)			
	.10	.05	.01	.001
One-tailed test	1.282	1.645	2.326	3.090
Two-tailed test	1.645	1.960	2.576	3.291

We will give a more formal treatment to the z distribution in an upcoming lecture.

z-test Example - Juvenile and Adult Arrest

Equation for z:

$$z = \frac{p_1 - p_2}{\sqrt{\pi * (1 - \pi)} * \sqrt{\frac{n_1 + n_2}{n_1 * n_2}}}$$

π is the **unconditional** probability of an event for the entire sample

p_1 and p_2 are the **conditional** probabilities for the two subsamples of interest

n_1 and n_2 are the number of cases in the two subsamples (i.e., column totals)

As a means of review, the unconditional probability of arrest as an adult in this sample is equivalent to $\pi = p(A) = 40/100 = 0.40$. The conditional probability of adult arrest given having been arrested as a juvenile is $p(A \mid J) = 30/50 = 0.60$ and the probability of an adult arrest given no arrest as a juvenile is $p(A \mid \sim J) = 10/50 = 0.20$.

Step 1: Formally state hypotheses

Hypotheses stated in terms of π, the population (unconditional) probability of arrest as an adult. Under H_1, adult arrest is more likely when a juvenile arrest (J) has occurred than when it has not ($\sim J$).

$$H_1 : \pi_J > \pi_{\sim J}$$

$$H_0 : \pi_J \leq \pi_{\sim J}$$

Notice that this is a directional test. Specifically, a *right-tailed* test. How would we have written it if we wanted a *left-tailed* test instead? In that case, we would simply flip the signs for both H_1 and H_0. By contrast, for a two-tailed test H_1 would be $p(J) \neq p(\sim J)$ and H_0 would be $p(J) = p(\sim J)$.

Step 2: Choose a probability distribution - z distribution

Step 3: Make decision rules (refer to table) - $\alpha = 0.05$ (one-tailed); $z_{crit} = 1.645$; reject H_0 if TS > 1.645

	Juvenile Arrest?		
Adult Arrest?	No ($\sim J$)	Yes (J)	Total
No ($\sim A$)	40	20	60
Yes (A)	10	30	40
Total	50	50	100

Step 4: Compute the test statistic

$$z = \frac{p_1 - p_2}{\sqrt{\pi * (1 - \pi)} * \sqrt{\frac{n_1 + n_2}{n_1 * n_2}}} = \frac{0.6 - 0.2}{\sqrt{.4 * .6} * \sqrt{\frac{50 + 50}{2500}}} = \frac{0.4}{0.0980} = 4.082$$

Note that I structure the test such that the numerator is p_1 is represented by $p(J)$ and p_2 is represented by $p(\sim J)$. I would structure the test the same way if I had the opposite hypothesis or if my hypothesis was two-tailed. The difference is that the result I expect would be negative (for a one-tailed test presuming $p(J) < p(\sim J)$)

Step 5: Make a decision about the null hypothesis - Reject H_0, conclude that the probability of adult arrest is significantly *higher* if you are arrested as a juvenile.

The fact that we obtained the test statistic we did is not a coincidence because χ^2 and z are very closely related. In fact, your conclusion from a χ^2 test will be equivalent to conducting a two-tailed z test. The primary difference is that a z test has the added flexibility of allowing us to do one-tailed tests.

$$4.082 \text{ (from the z-test)} = \sqrt{16.667} \text{ (from the chi-square test)}$$

Not **exactly** equal, but we have rounding error to blame for that.