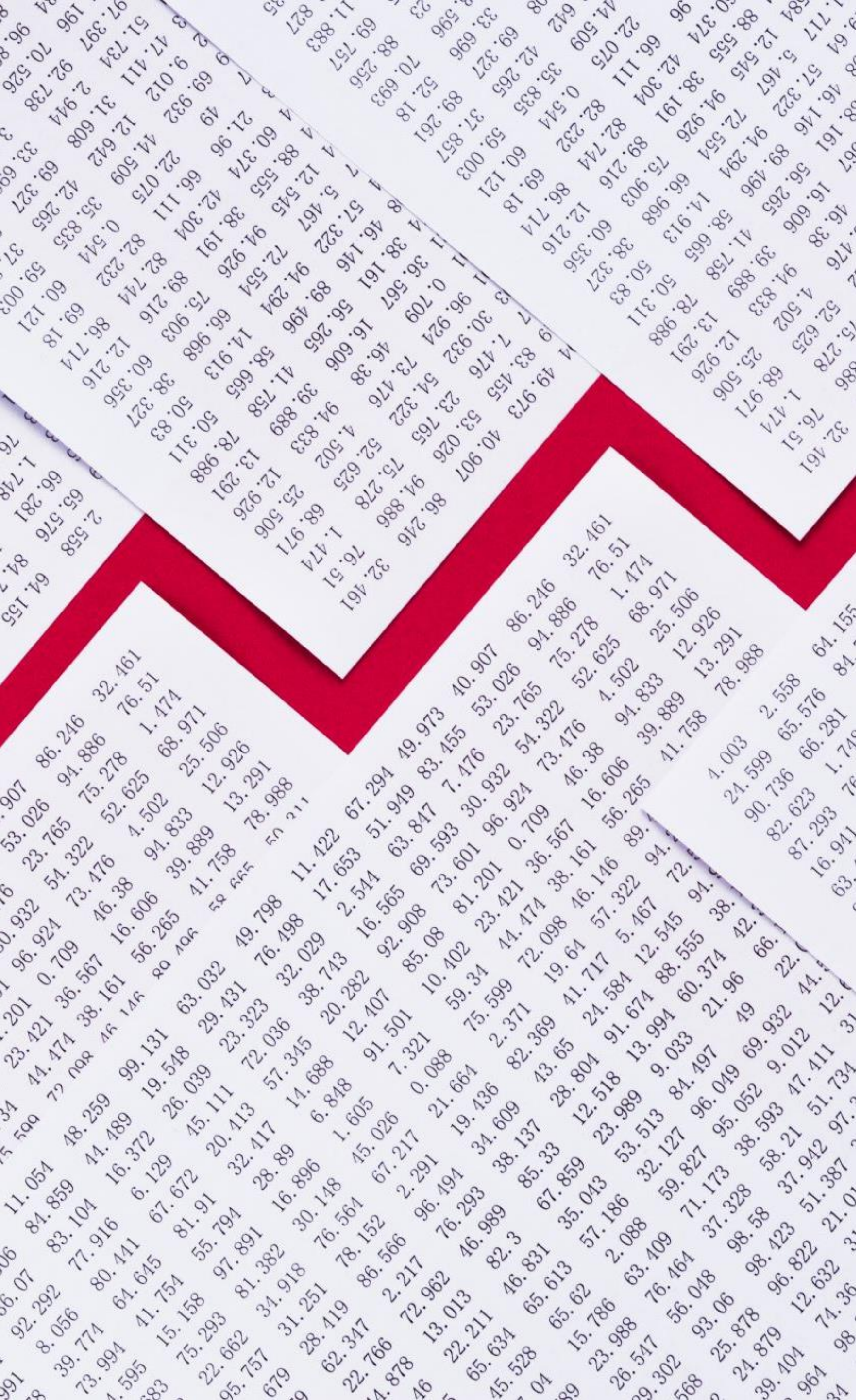
Risk, Odds, Odds Ratios and Probabilities



Four Primary Risk Measures

- Prevalence
- Incidence
- Relative Risk
- Odds Ratios

Measures of Risk

- Concepts: Uncertainty, Probability and Odds
- Common Measures of Frequency
  - Prevalence
  - Incidence
- Relationship between Incidence, Duration & Prevalence
- Risk Estimates (and their uses)

Measures of Risk

- Risk estimates for many different measures
  - Drug use in a population; by type and frequency
  - Developing a mental health problem or some other outcome following a violent victimization
  - The drawbacks/benefits of a proposed treatment
- These can all be discussed as:
  - Frequencies
  - Proportions, probabilities & odds/odds ratio
  - Prevalence and incidence rates
  - Risk (absolute and relative)   (e.g., the risk of something happening (or not) v the risk of one thing in relation to another)

Relationship between prevalence, probability and odds

Let p = the probability of some outcome, then

Odds for that outcome is defined as

$$Odds = \frac{p}{(1 - p)}$$

Example

Example: In North Carolina the probability of a pre-term birth is .1136. Compute and interpret:

(1) The odds that a woman will have a pre-term birth; and

(2) The odds that a woman will not have a pre-term birth.

Solution

- Odds for pre-term birth = $(.\mathbf{1136})/(\mathbf{1}-.\mathbf{1136})=(.\mathbf{1136})/(.\mathbf{8864})=.\mathbf{1282}$
- Odds for full-term birth = $(.\mathbf{8864})/(\mathbf{1}-.\mathbf{8864})=(.\mathbf{8864})/(.\mathbf{1136})=\mathbf{7}.\mathbf{803}$

Interpretation

- Three possibilities
  - If Odds are < 1 then the odds are “lower”
  - If Odds are > 1 then odds are higher
  - If Odds = 1 then the odds are the same
- Here .1282 < 1, the odds of pre-term birth are lower than full-term births
- **Recall:** the Null Hypothesis is one of no difference, what is the null hypothesis for a test involving odds?

Relationship between probability and odds

- Let *odds =* the ODDS that some outcome will occur, then
- The probability for that outcome is defined as

$$p = \frac{Odds\ of\ the\ outcome}{(1{+}dds\ of\ the\ outcome)} = \frac{Odds}{(1{+}Odds)}$$

Intuition: The odds of a child being substantiated for neglect are 3:4 after being reported for maltreatment. What is the probability of being substantiated?

The odds are 3 to 4 substantiated versus not substantiated therefore the probability of a substantiation is 3 **out of 4** or .75 (following our rules of probability)

Example

- **Example:** At OSU, the ODDS that a student will graduate are twice as high compared to those who do not graduate. Compute the probability that a randomly selected person will graduate from OSU:

$$p = \frac{ODDS\ of\ the\ outcome}{(1+ODDS\ of\ the\ outcome)} = \frac{2}{(1+2)} = \frac{2}{3}$$

The odds are 2 to 1 (2:1) graduating versus not graduating
therefore the probability of graduating is
2 **out of** 3 or .667 (following our rules of probability)

Probability	Odds
0.80	4
0.67	**2**
0.60	1.5
0.50	1.0
0.40	0.67
0.33	0.5
0.25	0.33
0.20	***0.25***
0.10	***0.11***
0.05	***0.053***
0.01	***0.0101***

Intuition

- Probability and odds are more alike the lower the probability of the outcome, $p = P$ (outcome)
- Check these for yourself, by hand

Prevalence

- Def’n: the proportion of a defined group or population that has a condition or outcome at a ***given point in time***
- The prevalence ranges from 0 to 1 (i.e., it’s a proportion), but usually referred to as a rate and is often expressed as a %.

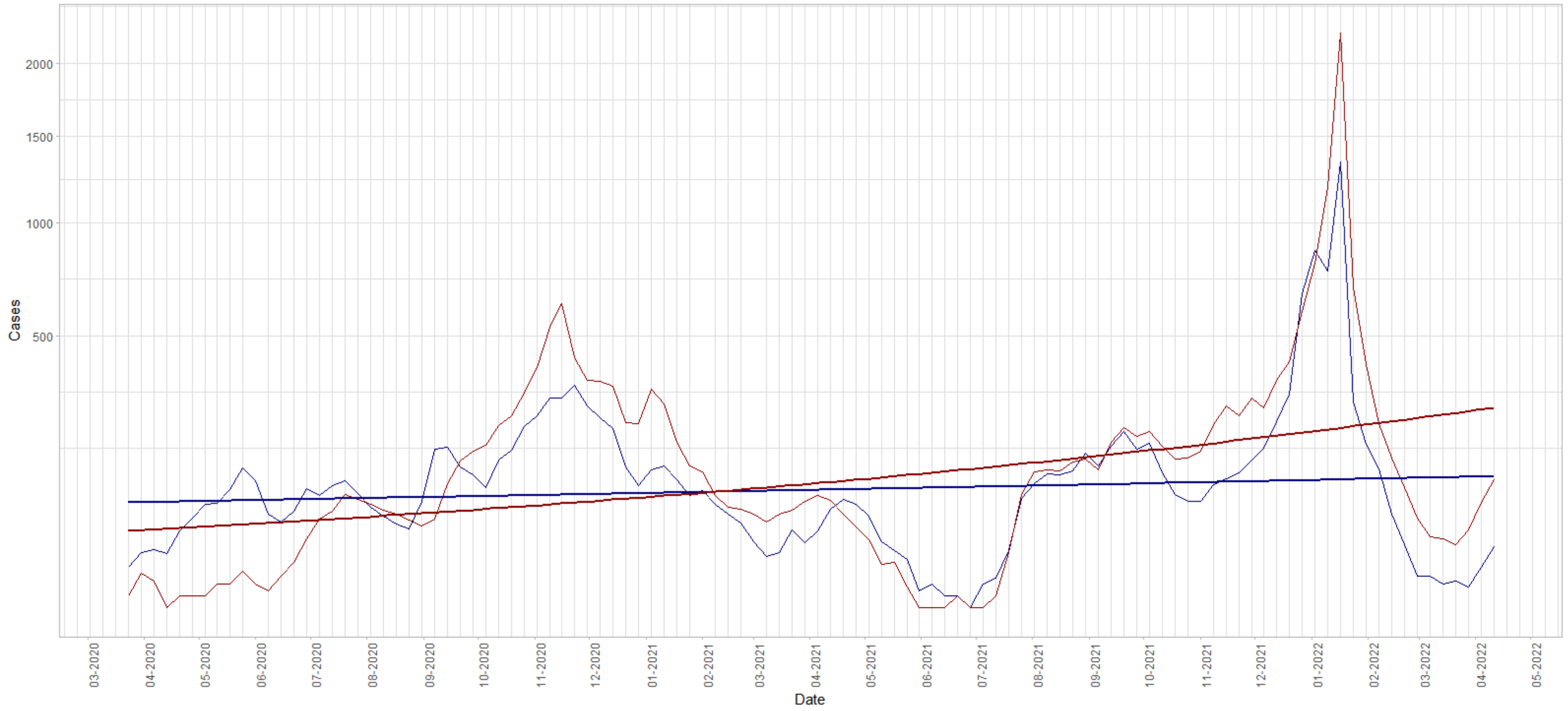
$$100\%\ X\ \frac{N}{T},$$

- where N= number of cases observed at a particular time, *t* and *T* = total number of individuals at time *t*
  - **Example:** Of 100 incarcerated persons, 35 reported child abuse histories. Therefore, our estimate of the prevalence of child abuse among persons who are incarcerated = 35%
- The prevalence rate answers the question: “what fraction of the group is affected at this moment in time?”

Incidence

- A special type of proportion that includes a specific *<u>time period</u>* and *<u>population-at-risk</u>*
- Numerator = the number of *newly affected* individuals occurring over a specified time period
- Denominator = the *population-at-risk* over the same time period

Daily COVID-19 Cases (7 Day Moving Average)



Blue areas represent most deprived 20% of neighborhoods
Red areas represent least deprived 20% of neighborhoods

Quantifying Risk and Interpreting Risk

- Information on the effect of a potential risk factor or beneficial treatment can be presented in several different ways:
  - Relative Risk (RR)
  - Odds Ratio (OR)
  - Percent decrease or increase in Odds
- The way risk information is ***presented and interpreted*** can have a profound effect on decision-making processes with implications for policy and treatment outcomes

Categorical Association Estimates

- We might also think of risk and odds as measures of association for categorical variables
- Measures of association quantify the potential relationship between “exposure” and “outcome” among two groups.
- Recall there are two types of association (+) and (-)
- The two main measures are Relative Risk (RR) and Odds Ratio (OR)

Relative Risk (AKA Risk Ratio, or Rate Ratio)

The general definition of RR is

$$\frac{Risk\ of\ an\ outcome\ in\ the\ group\ of\ primary\ interest}{Risk\ of\ that\ same\ outcome\ in\ a\ comparison\ group}$$

- Here we could be taking the ratio of two prevalence measures (i.e. proportions) or two rate measures (e.g., murder rates).
- A RR = 1.0 the risk is the same for both groups, if RR > 1.0 the risk is greater for group in numerator and if RR < 1.0 it indicates decreased risk for group in numerator.

Example: Incidence of Tuberculosis Infection Among HIV+ Patients By Hospital Ward

	Developed Tuberculosis?		
Wing	**Yes**	**No**	
East Wing	28	129	**157**
West Wing	4	133	**137**
	32	**262**	**Total = 294**

Column marginals

The column marginal answers the question: *Among patients who developed tuberculosis what percentage reside in the east wing?* **28/32**

Row marginals

The row marginal answers the question: Among patients *residing in the East Wing, what percentage developed tuberculosis?* **28/157**

Note: this is NOT the same as the probability of having tuberculosis AND being in the East Wing (28/294)

	Developed Tuberculosis?		
Wing	**Yes**	**No**	
East Wing	28	129	**157**
West Wing	4	133	**137**
	32	**262**	**Total = 294**

What is the risk ratio (RR) of developing tuberculosis among patients in the east wing compared to the west wing?

What percentage of East Wing inmates developed tuberculosis? This is also the 'risk'

28/157 = 17.8%

What percentage of West Wing inmates developed tuberculosis? This is also the 'risk'

4/137 = 2.9%

$$RR = \frac{17.8}{2.9} = 6.1$$

Inmates who resided in the East Wing were 6.1 times more likely to develop tuberculosis compared to inmates in the West Wing

Your Turn: Relative Risk of COVID-19 in a sample of individuals who received the Pfizer Vaccine

	Developed Coronavirus?		
Vaccinated?	**Yes**	**No**	
Yes	18	134	**152**
No	3	4	**7**
			Total = 159

What is the relative risk of developing COVID-19 among individuals who have been vaccinated (compared to those who have not?

What is the outcome we are looking for?

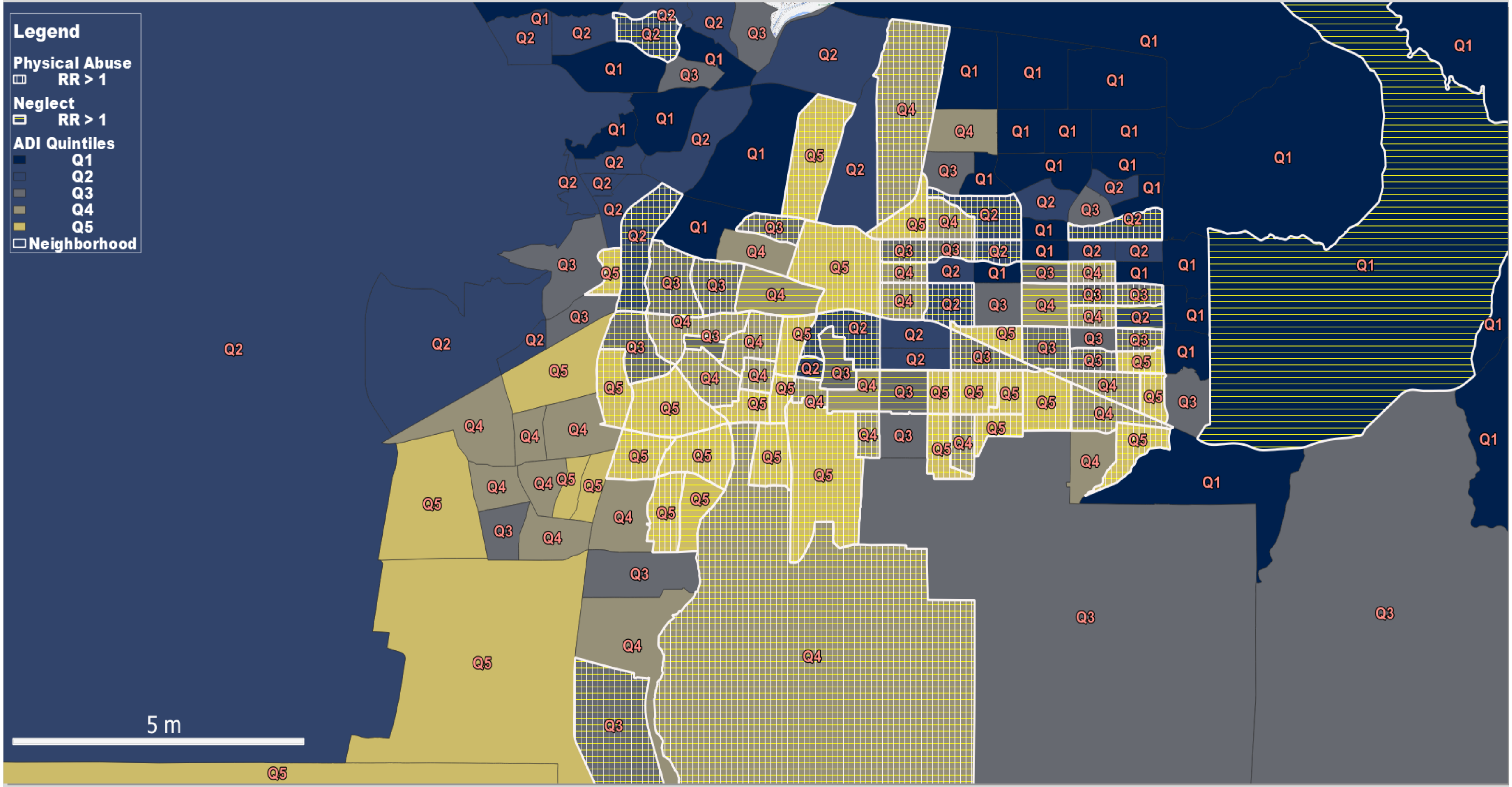
What percent of vaccinated persons developed corona?

What percent of unvaccinated persons developed corona?

Compute and interpret the relative risk…

The risk of developing COVID-19 among those who are vaccinated by the Pfizer vaccine is .28 times that of those who are not

$$RR = \frac{11.8}{42.9} = .28 < 1$$


Legend
Physical Abuse
RR > 1
Neglect
RR > 1
ADI Quintiles
Q1
Q2
Q3
Q4
Q5
Neighborhood
5 m

Computing Odds

- Recall the odds of an event or outcome, A, is defined as

$$ODDS = \frac{probability\ of\ the\ outcome}{(1 - the\ probability\ of\ the\ outcome)} = \frac{p(A)}{1 - p(A)}$$

For example, suppose we roll a single (fair) die, compute the odds of rolling a 3:

(1) Let A – we observe a "3" after 1 roll of a fair die
(2) What is the P(A)? 1/6
(3) What is 1 - P(A)? 5/6
(4) Compute

$$ODDS = \frac{p(A)}{1 - p(A)} = \frac{1/6}{5/6} = \frac{1}{5}$$

(5) Interpret: The odds of getting a "3" are 1 in 5 (intuitive) and the odds of not getting a "3" are 5 in 1

Example

- To compute the odds we need to introduce the idea of conditional probability

	Developed Tuberculosis?		
Wing	**Yes**	**No**	
East Wing	28	129	**157**
West Wing	4	133	**137**
	32	**262**	**Total = 294**

Outcome

Column marginals

"Risk factor"

Different questions can be asked....

Probability of getting Tuberculosis?

$$\frac{\text{P(Tuberculosis)}}{\text{Total}} = 32/294$$

Probability of getting Tuberculosis conditional on being in the East Wing?

$$\frac{\text{P(Tuberculosis | East Wing)}}{\text{East Wing Total}} = 28/157$$

The Odds *Ratio*

- The Odds Ratio (OR) for an outcome associated with a risk factor is ratio of the odds for the outcome for those with risk factor and the odds for the outcome for those without the risk factor

Odds for outcome among those with risk factor present

$$\mathrm{OR} = \frac{\dfrac{P(\text{Outcome} \mid \text{Risk Factor})}{1 - P(\text{Outcome} \mid \text{Risk Factor})}}{\dfrac{P(\text{Outcome} \mid \text{No Risk Factor})}{1 - P(\text{Outcome} \mid \text{No Risk Factor})}}$$

Odds for outcome among those without the risk factor.

Example: For the following data, compute **the odds ratio** of 1st preg. before (or at) age 25 among women who have cervical cancer and those who do not – **we are only looking at individuals aged 25 or under**

Age at 1st Pregnancy	Cervical Cancer	No Cervical Cancer	Row Totals
Age $\leq$ 25	**42**	**203**	245
Age > 25	**7**	**114**	121
Column Totals	49	317	n = 366

a) What is P(preg age $\leq$ 25|cervical cancer) & P(preg age $\leq$ 25|No cervical cancer) ?

P(preg age $\leq$ 25|Cancer) = 42/49 = .857 or 85.7% prev

P(preg age $\leq$ 25|No Cancer) = 203/317 = .640 or 64.0%

Age at 1st Pregnancy	Cervical Cancer	No Cervical Cancer	Row Totals
Age $\leq$ 25	42	203	245
Age > 25	7	114	121
Column Totals	49	317	n = 366

b) What are the odds for preg $\leq$ 25 among women with cervical cancer? What are the odds among women without cervical cancer?

Odds of 1st preg. $\leq$ 25 among those with cervical cancer = .857/(1-.857) = 5.99

Odds of 1st preg. $\leq$ age 25 among those with no cervical cancer = .64/(1- .64) = 1.78

Age at 1st Pregnancy	Cervical Cancer	No Cervical Cancer	Row Totals
Age $\leq$ 25	**42**	**203**	245
Age > 25	**7**	**114**	121
Column Totals	49	317	n = 366

c) What is the odds ratio for having a pregnancy $\leq$ age 25 among women with cervical cancer versus those without cervical cancer?

Odds Ratio (OR) = 5.99/1.78 = 3.37, the odds for having 1st child on or before age 25 are 3.37 times higher for women with cervical cancer versus those without cervical cancer.

Easy Formula to use

- Again, the question is: compute the odds ratio of having a 1st preg. before (or at) age 25 among women who have cervical cancer and those who do not

Risk Factor Status	**Case**	**Control**
Risk Factor Present	a	b
Risk Factor Absent	c	d

$$OR = \frac{a \times d}{b \times c}$$

$$\mathbf{OR} = \frac{\mathbf{a \times d}}{\mathbf{b \times c}}$$

Easy Formula to use

- Again, the question is: compute the odds ratio of having a child before (or at) age 25 among women who have cervical cancer and those who do not

Age at 1st Pregnancy	Cancer	No Cancer
Age <= 25	a = 42	b = 7
Age > 25	c = 203	d = 114

$$OR = \frac{42 \times 114}{7 \times 203}$$

4902 x 1421 = 3.4!

Other interpretations

- Compute the odds as a percent increase or decrease (this is often the better interpretation because people understand percentages better than they understand odds)
- The percent change in the odds is defined as $(OR - 1)\ X\ 100$
- **Example**: Odds ratio = 1.89 calculate the percent change in the odds
- |(1 – 1.89)| X 100 = 89% increase
- **Example**: the odds of a first pregnancy among women aged 25 or younger is [(3.4 - 1) * 100 =] 240% higher among women who had cervical cancer compared to those who have not
- If the odds are .69 what is the percent change in the odds?

Example from my work

	Adjusted Estimates	
Domains of structural disadvantage	RR (95% Credible Interval)	Percent Change
Time trend	1.068 (1.029, 1.105)	6.78% (2.94, 10.51)
Food & Housing Insecurity		
Eviction rate	1.060 (1.037, 1.084)	5.99% (3.70, 8.38)
Rent burden	1.013 (1.007, 1.019)	1.30% (0.73, 1.89)
Labor Market Characteristics		
Origin-Destination job flows	0.937 (0.806, .986)	-6.35% (-19.39, -1.41)
Commute times > 60 min.	1.137 (1.015, 1.270)	13.71% (1.53, 26.97)
Percent travel to work (auto)	0.830 (0.706, 0.974)	-17.03% (-29.43, -2.57)
Percent travel to work (public transportation)	0.645 (0.416, 0.998)	-35.47% (-58.38, -0.24)
Origin-Destination Job Flows	0.937 (0.806, 0.986)	-6.35% (-19.39, -1.41)

For every *one-unit increase* in the neighborhood eviction rate and rent burden rate, the relative risk of substantiated CAN increased by 6% and 1.3%; the risk was 13.71% higher in areas where commuting times were > 60 minutes on average, and 6.35% lower in areas with more job flows *controlling for financial strength and economic inequality.*

Data to practice: Less focus on socioeconomic *status, more focus on* socioeconomic context and economic inequality

- Targeted interventions to most socioeconomically vulnerable areas
  - Socioeconomic gradients & cumulative disadvantage
- Sustainable food and housing policy
  - Safe and affordable housing
  - Programs to address food insecurity
    - 11.7% of the tracts in Bernalillo County were defined as food insecure, most advantaged areas not food desert;
- Local governments and non-profits
  - SNAP, EITC, community land trusts, strengthening local food systems
- Job market versus labor force characteristics
  - Commuting burden – work locations have increasingly become spread out over time and associated with higher maltreatment risk
  - Job proximity and accessibility in socioeconomically vulnerable tracts

Measuring Disparity

Many issues involve a comparison between two proportions (or probabilities). Consider the example of a test taken by males and females on which males are observed to have a higher pass rate.

How would you describe the disparity in pass (or fail) rates (and what are the problems of interpretation)?

- 1. The difference in pass (fail) rates
- 2. The ratio of the pass (fail) rates
- 3. The odds ratio

The difference in pass rates

- Assume that the passage rate for males is 97% and the passage rate for females is 90%
  - What is the difference in pass rates?
  - What is the difference in fail rates?

The ratio of the pass/fail rates (relative risk)

- Assume that the passage rate for males is 97% and the passage rate for females is 90%
  - What is the ratio of the pass rates? $\frac{.97}{.90} = 1.07$
  - Interpretation?

  - What is the ratio of the fail rates? $\frac{.10}{.03} = 3.33$
  - Interpretation?

Odds of Passing

- What are the odds of passing for males and females?
- Calculate the odds ratio in pass rates
  - First calculate the odds of passing for males

$$Odds = \frac{p_M}{1 - p_M} = \frac{.97}{1 - .97} = \frac{.97}{.03} = 32.3 : 1$$

  - Second calculate the odds of passing for females

$$Odds = \frac{p_W}{1 - p_W} = \frac{.90}{1 - .90} = \frac{.90}{.10} = 9 : 1$$

Odds Ratio of the Fail Rates

- The odds ratio of the fail rates is defined as the odds of failing for women divided by the odds of failing for men

$$Odds = \frac{p_W}{1 - p_W} = \frac{.10}{1 - .10} = \frac{.10}{.90} = \frac{1}{9} = 1:9$$

$$Odds = \frac{p_M}{1 - p_M} = \frac{.03}{1 - .03} = \frac{.03}{.97} = \frac{3}{97} = 1:32.3$$

- Calculate the odds ratio in fail rates
  - First calculate the odds of failing for women
    - For every woman that fails, 9 will pass
  - Second calculate the odds of failing for men
  - Divide the two
    - Odds ratio = $\dfrac{\frac{1}{9}}{\frac{3}{97}} = \frac{97}{27} = 3.6$

Contingency Tables and Chi-Square

STEPS IN STATISTICAL DECISION-MAKING AND THE CHI-SQUARE GOODNESS OF FIT

Review

Last week, we used probability theory to identify the observed significance level in a test of statistical significance (whether you realized it or not)

Here, we review assumptions behind our test statistic and take a step-by-step approach to hypothesis testing

The steps are illustrated with a specific research problem that is addressed using the **binomial distribution (for which I've made up)**

The binomial distribution is only used in one situation (?: hint go back to the definition of the binomial distribution)

Don't lose sight of the fact that we are still establishing a general model for presenting tests of statistical significance

Review: What is Hypothesis Testing?

Hypothesis testing: A procedure, based on sample evidence and probability theory, used to determine whether the hypothesis is a reasonable statement and should not be rejected, or is unreasonable and should be rejected.

Null Hypothesis H_0: A statement about the value of a population parameter that is ***assumed to be true*** for the purpose of testing

- Example: Twenty percent of all juvenile offenders are caught and sentenced to prison.

Alternative Hypothesis H_a: A statement about the value of a population parameter that is assumed to be true if the Null Hypothesis is rejected

- Example: The percentage of all juveniles offenders are caught and sentenced to prison is not equal to 20%

Review: Parametric vs. Non-parametric

Parametric Tests – Statistical tests that involve assumptions about population parameters

- E.g., normal distribution, interval/ratio level measurement

Nonparametric Tests

- Also known as distribution-free tests; do not rely on assumptions about distributions
- E.g., does not assume interval/ratio, no normality assumption

Depending on the type of data and hypotheses, the shape of the distribution of the test statistic must be chosen

- The second half of the course is devoted to the best decision about which test statistic to choose
- We already saw the binomial distribution, today we cover Chi-Square, later Z, *t,* and F distributions

Hypothesis Generation

These two hypotheses, designated H_O and H_A (or H_1), are mutually exclusive

Properties of the Null

- Specifies no difference or no change from a standard or theoretical value
- Always specifies something about a particular population parameter
- **Used in constructing a sampling distribution**
  - **For the subsequent quantitative work, the null hypothesis is assumed to be true**

Properties of the Alternative

- Usually stated in general terms
- Mutually exclusive - no overlap with H_O
- Can be directional or non-directional

Directional vs. Non-directional H_As

Non-directional H_A - usually stated as "does not equal" or "is different than"

Directional H_A - stated as "greater than" or "less than"

- note that a non-directional hypothesis is equal to the two directional hypotheses "greater than" or "less than"

Model Assumptions

Step 1:

- **What is the level of measurement in your data?**: state at the outset the type of measurement required by a test
  - For the binomial test, the measures must be nominal, with two outcomes (?)
- **Shape of the Population Distribution**
  - Two types of tests
    - Parametric
    - Non-parametric

Statistical Assumptions

- **Sampling Method**
  - Random sampling → external validity
    - Recall the definition of validity from week 1
  - Convenience sampling (not always a bad thing, but must acknowledge threats to validity)
    - **Example**. Studying the behavior of HIV+ individuals
- **Sampling Frame**
  - The sampling frame is the list of things in your study that you sample from
  - Example: Compile a list of all hot spots in LA and sample those during covid-19
    - The sampling frame is: all hot spots of child abuse and neglect in the city of Los Angeles

Statistical Assumptions

- The research hypothesis is the **alternative hypothesis**

1. **Hypothesis 1**. Child abuse and neglect in COVID-19 hot spots have changed during the COVID-19 pandemic.
2. **Hypothesis 2**. Child abuse and neglect in COVID-19 hot spots have different levels of unemployment, chronic absenteeism and poor child health compared to cold spots.
3. **Hypothesis 3. Child abuse and neglect will change if appropriate interventions are placed.**

- The null hypothesis is that there is no difference …
- Assume we were interested in (3). We saw 10 of 11 total hot spots change after the intervention. How likely is it to observe 10 hot spots change relative to control locations if the null hypothesis of no change is true? What is the sampling distribution?

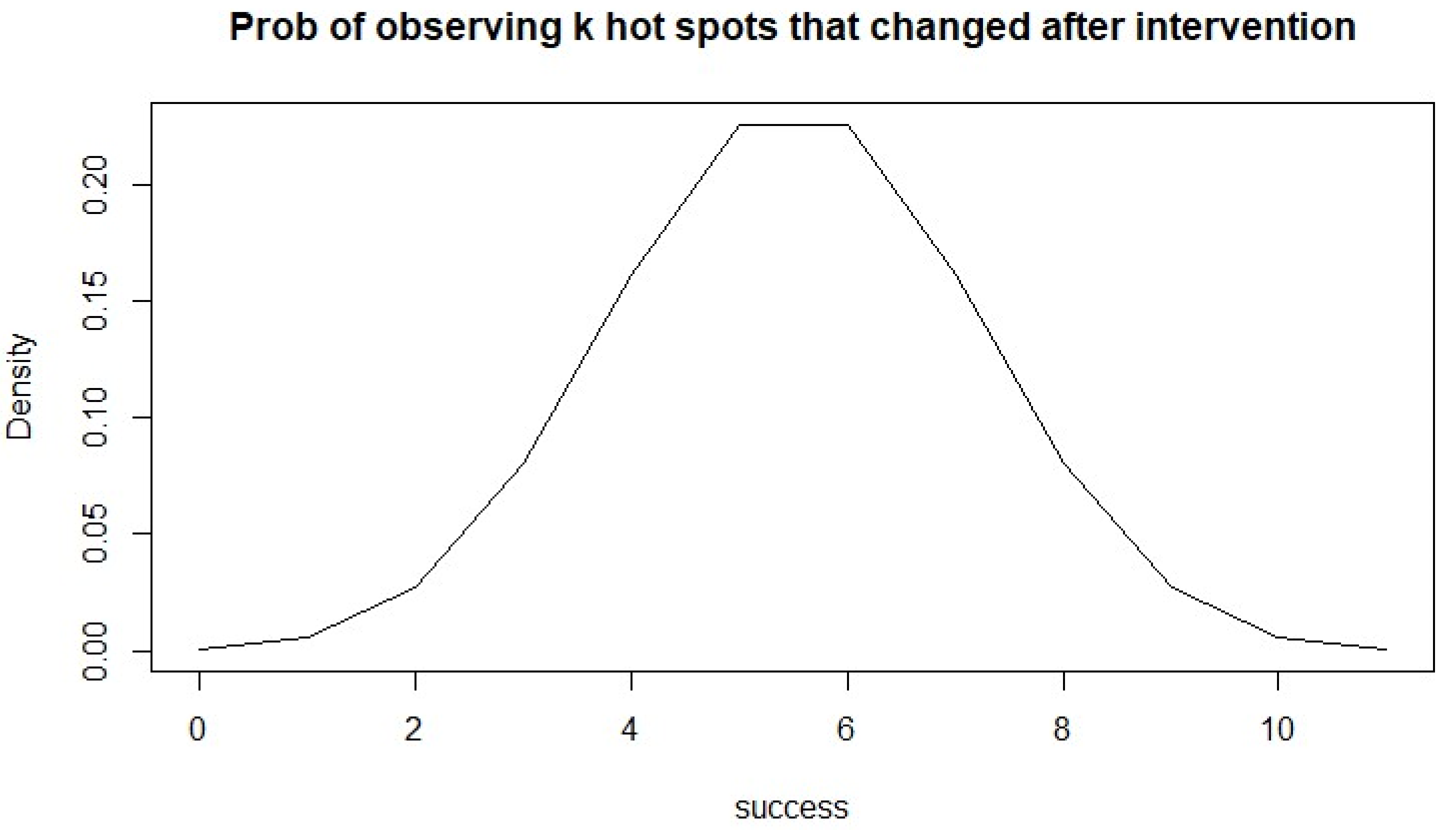
Mathematical Hypos

- The null and alternative can be stated in mathematical terms:
  - H_o: The level of CAN in treatment hot spots relative to CAN in control hot spots is no different after my intervention or
    - H_o: Prob = .50 (the treatment and control hotspots are equally likely to improve)
  - H_A: The level of CAN in treatment hot spots relative to CAN in control hot spots changes after my intervention, or H_A: P ≠.50

Stating All Assumptions

- Level of Measurement: Nominal binary (change or no change).
- Population Distribution: Binomial.
- Sampling Method: Independent random sampling (no replacement).
- Sampling Frame: Hot spots of child abuse and neglect during Covid-19
- Hypotheses:
  - H0: The level of CAN in treatment hot spots does not change relative to CAN in control hot spots after my intervention, $P = 0.50$.
  - H1: The level of CAN in treatment hot spots relative to incivilities in control hot spots changes after my intervention, $P \neq 0.50$.
- **Notice this is a two-tailed test of significance**

Run hot spot example.R code



Sampling Distribution of Success or Failure in 11 Trials

OUTCOME OF TRIALS	OVERALL PROBABILITY
0 successes	0.00049
1 success	0.00537
2 successes	0.02686
3 successes	0.08057
4 successes	0.16113
5 successes	0.22559
6 successes	0.22559
7 successes	0.16113
8 successes	0.08057
9 successes	0.02686
10 successes	0.00537
11 successes	0.00049

The Sampling Distribution

```
dbinom(0, size=11, prob=.5)
dbinom(1, size=11, prob=.5)
dbinom(2, size=11, prob=.5)
dbinom(3, size=11, prob=.5)
dbinom(4, size=11, prob=.5)
dbinom(5, size=11, prob=.5)
dbinom(6, size=11, prob=.5)
dbinom(7, size=11, prob=.5)
dbinom(8, size=11, prob=.5)
dbinom(9, size=11, prob=.5)
dbinom(10, size=11, prob=.5)
dbinom(11, size=11, prob=.5)
```

```
n <- 11
success <- 0:11
dbinom(success, n, .5)

plot(success,
   dbinom(success, n, .5),
   type = "l",
   ylab = "Density",
   main = "Prob of observing k hot spots
   that changed after intervention")
```

```
[1] 0.0004882812 0.0053710938 0.0268554688 0.0805664062
[5] 0.1611328125 0.2255859375 0.2255859375 0.1611328125
[9] 0.0805664062 0.0268554687 0.0053710938 0.0004882812
```

Significance Level & Rejection Region

We selected the distribution that will be used to assess Type I error (theoretically we can select any level for our type I error, in practice we select .05)

Our first step is to choose the significance level of our test

The significance level defines the Type I error but does not tell us what outcomes in our sample lead us to reject the null hypothesis (ex. How many successes do we need before the probability is less than .05?)

Now we define outcomes that will lead us to reject the null hypothesis

- Define rejection region – the area in the rejection region that includes the outcomes that lead us to reject the null hypothesis
- **The area covered by the rejection region is equivalent to the significance level of a test**

Choosing a One-Tailed or a Two-Tailed Rejection Region

We are trying to decide whether the outcomes observed in our sample are very different from the outcomes that would be expected under the null hypothesis

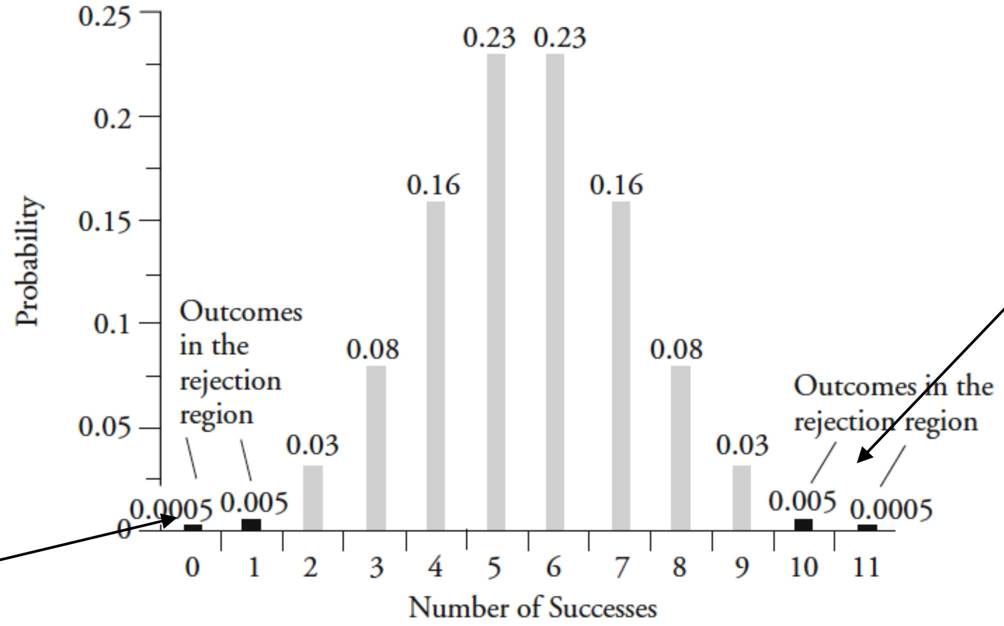
What is your intuition as to where to place the rejection region – i.e. where are the events unlikely

IF we choose a non-directional research hypothesis we believe we might observe values that are much larger than we expect or much smaller → two tailed test of significance

IF we chose a one-sided research hypothesis, we are concerned only with outcomes on one side of the distribution --> one tailed hypothesis

Rejecting the Null for a Two-Tailed Hypothesis

My intervention is a huge success …
we saw many differences in control and treatment areas

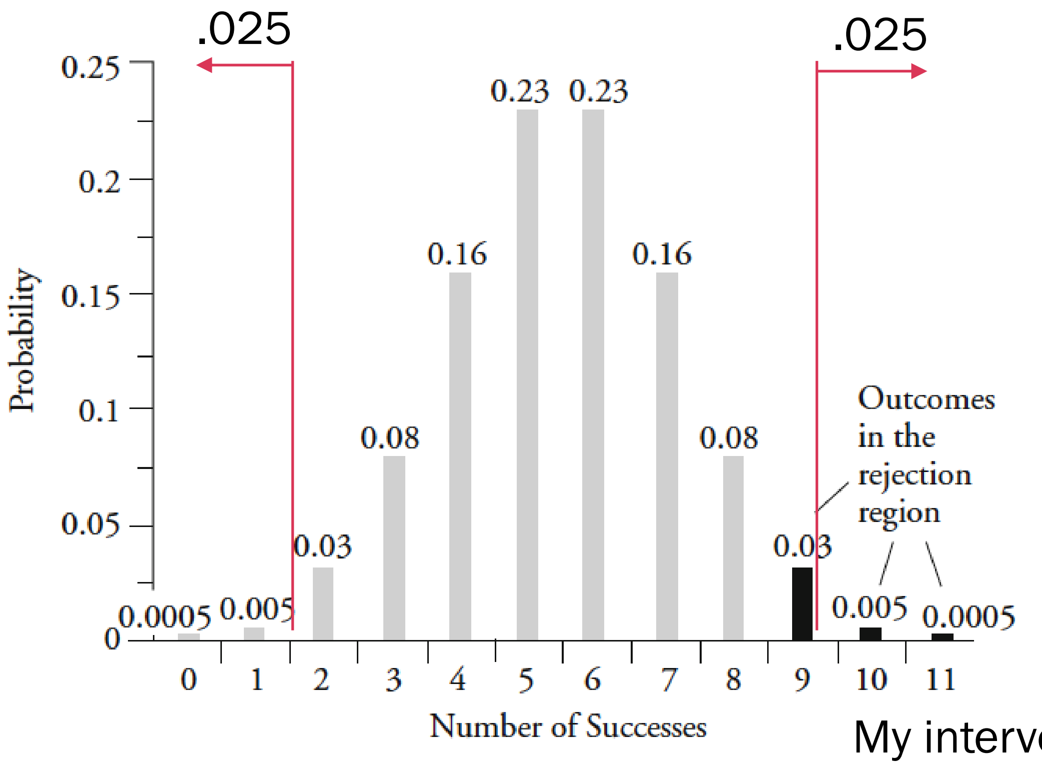


My intervention is a huge failure… we saw few differences in control and treatment areas

When we add 0 and 1 successes or 10 and 11 successes, we gain a probability value of 0.00586 (in each tail of the distribution, 0.00049 + 0.00537). This is less than the 0.025 value that we have defined
as the rejection region for each tail of our test.

If we included outcomes 2 and 9 the probability would be greater than .05 (.066 > .05)

Therefore we only reject the null that incivility is equally likely in control and treatment hot spots if EITHER 0 or 1 or 10 or 11 comparisons that changed – we observed 10 changes so we reject the null because our test statistic (10) is within our rejection region…



My intervention works

Sampling Distribution of Success or Failure in 11 Trials

OUTCOME OF TRIALS	OVERALL PROBABILITY
0 successes	0.00049
1 success	0.00537
2 successes	0.02686
3 successes	0.08057
4 successes	0.16113
5 successes	0.22559
6 successes	0.22559
7 successes	0.16113
8 successes	0.08057
9 successes	0.02686
10 successes	0.00537
11 successes	0.00049

Review: Definitions

Level of Significance: The probability of rejecting the null hypothesis when it is actually true. (signified by a)

Type I Error: Rejecting the null hypothesis when it is actually true.

Type II Error: Failing to reject the null hypothesis when it is actually false.

Review: Definitions

Significance level: the probability value associated with the decision rule represented by alpha → α (you choose alpha, typical value is .05)

- Results are considered significant when they are extreme (they can happen less than 5% of the time)

Critical value(s): The dividing point(s) between the region where the null hypothesis is rejected and the region where it is not rejected. The critical value determines the decision rule.

Rejection Region: Region(s) of the Statistical Model which contain the values of the Test Statistic where the Null Hypothesis will be rejected. The area of the Rejection Region = a

p-value: the probability of observing a test statistic as large (or larger) than we did if the null hypothesis is true, this is computed by you.

If $p < \alpha$, reject null

p-Value in Hypothesis Testing

p-Value: the probability, assuming that the null hypothesis is true, of getting a value of the test statistic at least as extreme as the computed value for the test.

If the p-value is smaller than the significance level, H_0 is rejected.

If the p-value is larger than the significance level, H_0 is not rejected.

Comparing p-value to a

Both **p-value** and **a** are probabilities.

The **p-value** is determined by the **data**, and is the probability of getting results as extreme as the data assuming H_0 is true. Small values make one more likely to reject H_0.

a is determined by **design**, and is the maximum probability the experimenter is willing to accept of rejecting a true H_0.

Reject H_0 if p-value < a for ALL MODELS.

The Test Statistic

A test statistic expresses the value of your outcome in units of the sampling distribution employed in your test. For the binomial distribution, the units are simply the number of successes in the total number of trials.

The test statistic for my intervention example is 10.

Making a Decision

If your test statistic falls within the rejection region, then you reject the null hypothesis.

If the test statistic does not fall in the rejection region, you cannot reject the null hypothesis

In our example, the test statistic falls within/outside the rejection region, so you reject/do not reject null?

What is statistical significance?

Remember always distinguish legal or practical significance

We also are not stating that we are certain that the null hypothesis is untrue for the population (how can we know?)

A statistically significant result is one that is unlikely if the null hypothesis is true for the ***population***

Whenever we make a statement that a result is statistically significant, we do it with the recognition that we are risking a certain level of Type I error

Chances of success & failure

8.6 Use the following binomial distribution showing the chances of success and failure for 12 trials.

Number of Successes	Probability
0 successes	0.00118
1 success	0.01065
2 successes	0.04418
3 successes	0.11110
4 successes	0.18857
5 successes	0.22761
6 successes	0.20032
7 successes	0.12953
8 successes	0.06107
9 successes	0.02048
10 successes	0.00463
11 successes	0.00064
12 successes	0.00004

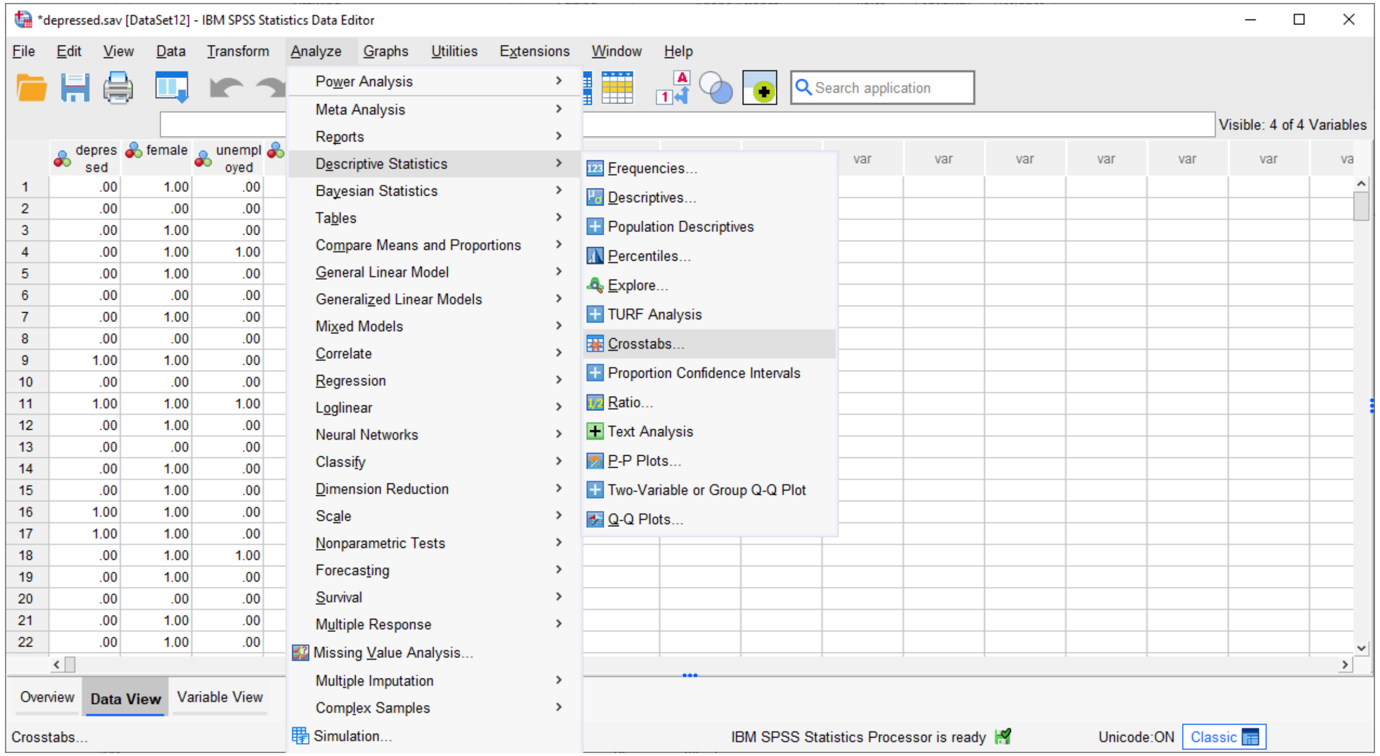
Using a significance level of 0.05, what outcomes would lead you to reject the null hypothesis for each of the following pairs of hypotheses?

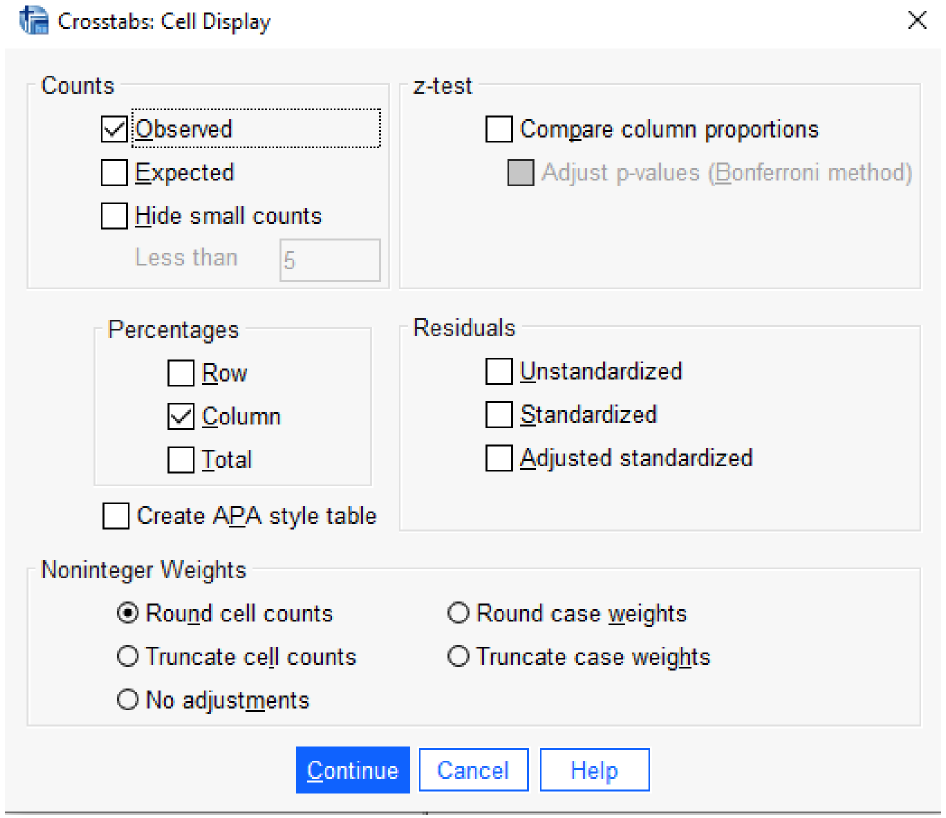
a. H_0: $P = 0.50$
   H_1: $P \neq 0.50$

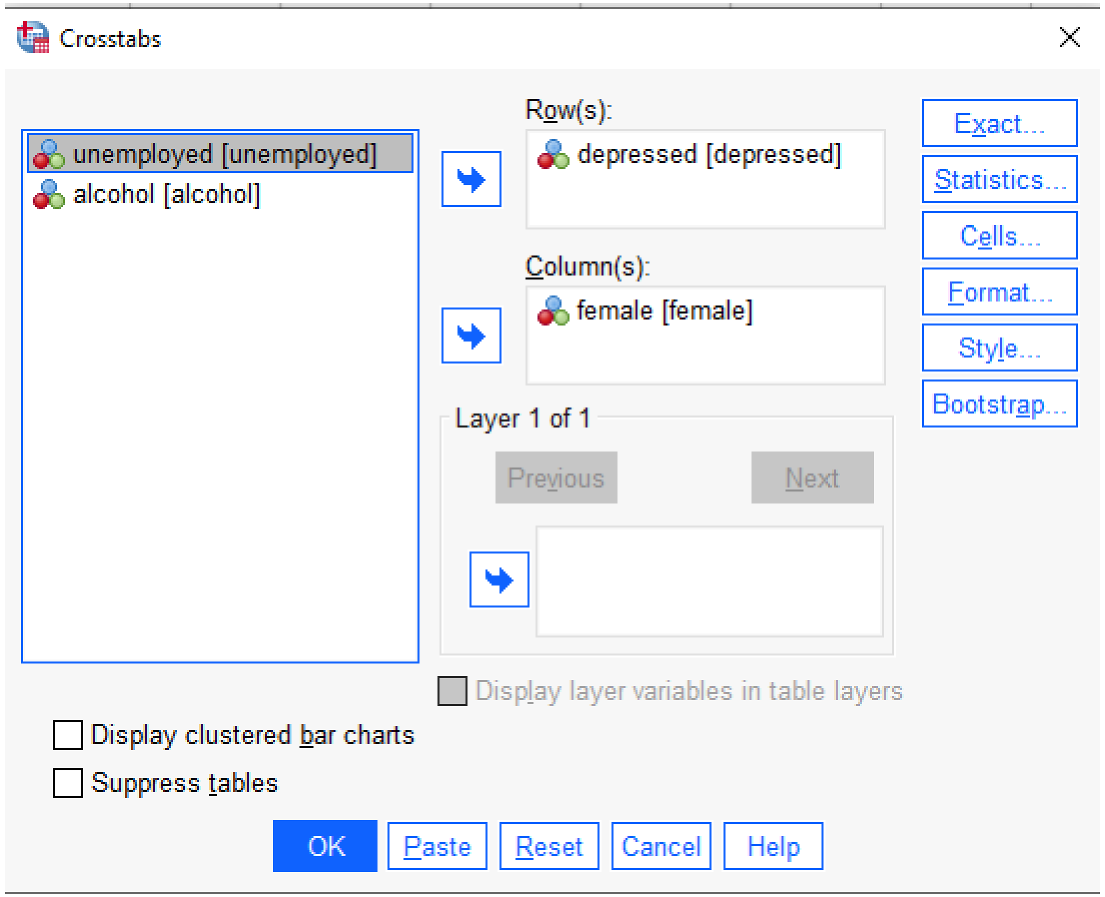
b. H_0: $P = 0.50$
   H_1: $P < 0.50$

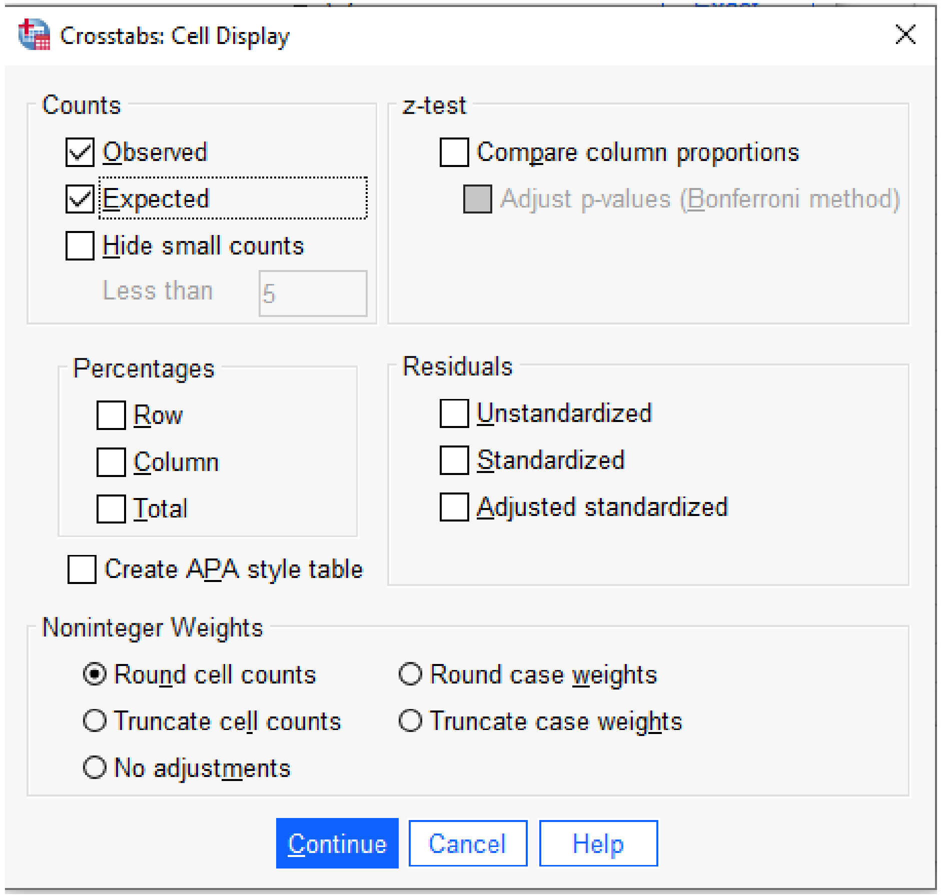
Analyze → Descriptives → Crosstabs

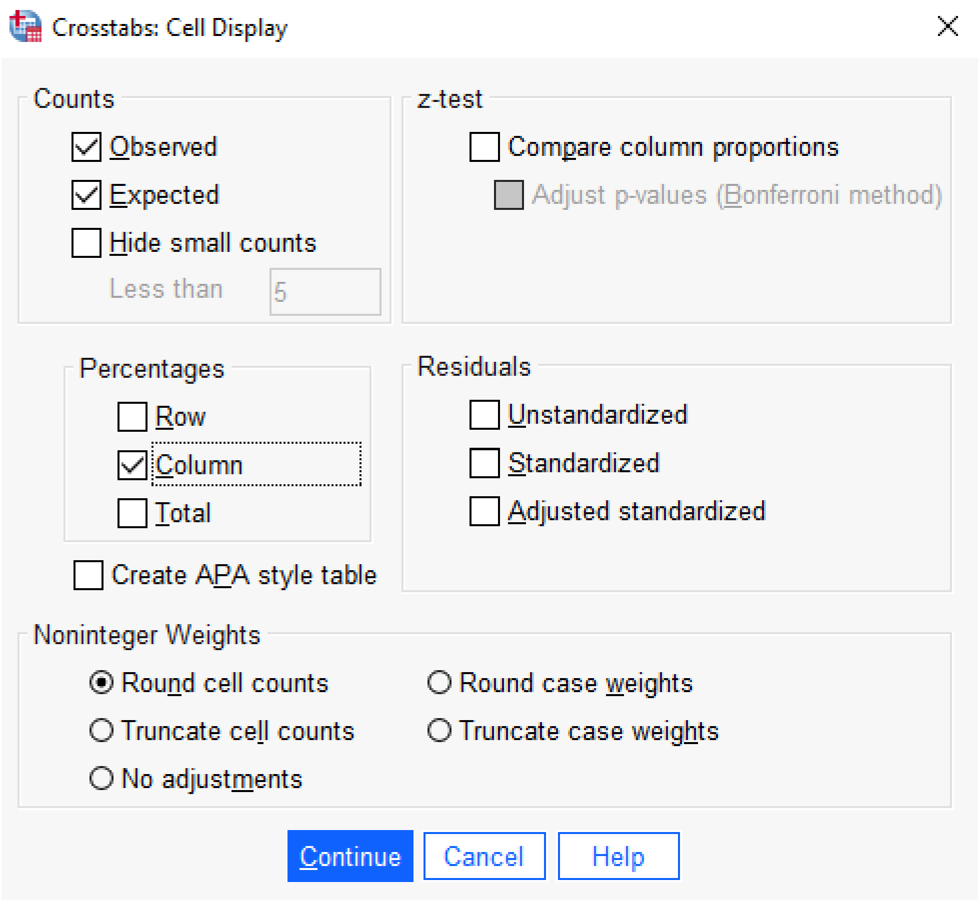
- Let's do this in SPSS

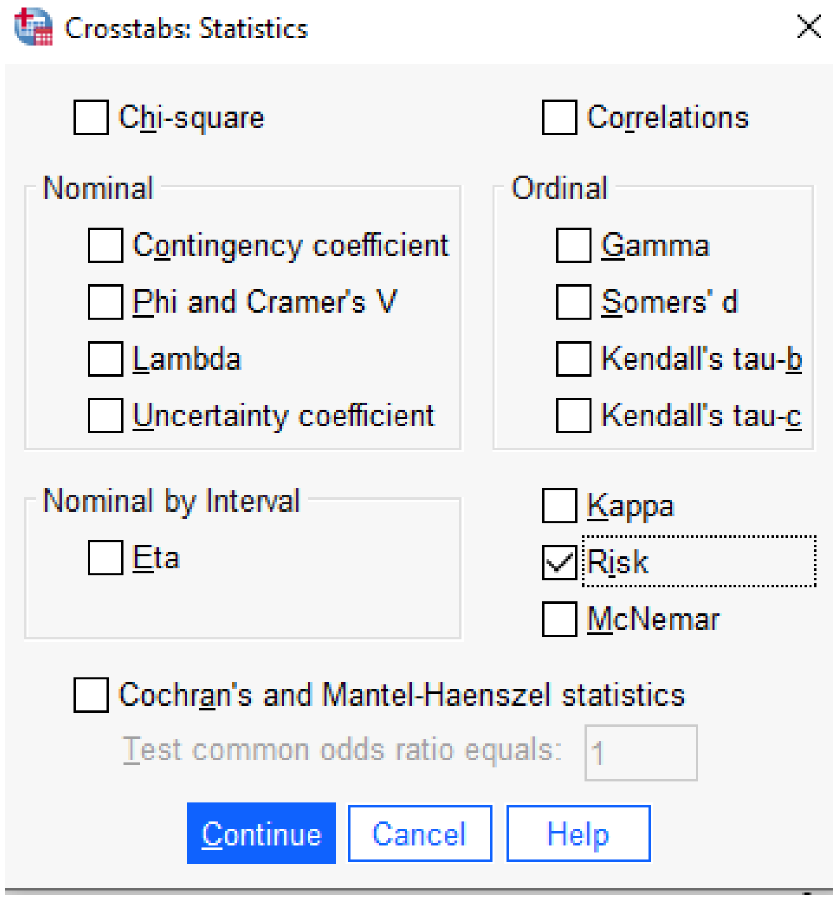





Crosstabs
unemployed [unemployed]
alcohol [alcohol]
Row(s):
depressed [depressed]
Column(s):
female [female]
Exact...
Statistics...
Cells...
Format...
Style...
Bootstrap...
Layer 1 of 1
Previous
Next
Display layer variables in table layers
Display clustered bar charts
Suppress tables
OK
Paste
Reset
Cancel
Help


Crosstabs: Cell Display
Counts
Observed
Expected
Hide small counts
Less than
5
z-test
Compare column proportions
Adjust p-values (Bonferroni method)
Percentages
Row
Column
Total
Residuals
Unstandardized
Standardized
Adjusted standardized
Create APA style table
Noninteger Weights
Round cell counts
Round case weights
Truncate cell counts
Truncate case weights
No adjustments
Continue
Cancel
Help


Crosstabs: Cell Display
Counts
Observed
Expected
Hide small counts
Less than
5
z-test
Compare column proportions
Adjust p-values (Bonferroni method)
Percentages
Row
Column
Total
Residuals
Unstandardized
Standardized
Adjusted standardized
Create APA style table
Noninteger Weights
Round cell counts
Round case weights
Truncate cell counts
Truncate case weights
No adjustments
Continue
Cancel
Help


Crosstabs: Statistics
Chi-square
Correlations
Nominal
Contingency coefficient
Phi and Cramer's V
Lambda
Uncertainty coefficient
Ordinal
Gamma
Somers' d
Kendall's tau-b
Kendall's tau-c
Nominal by Interval
Eta
Kappa
Risk
McNemar
Cochran's and Mantel-Haenszel statistics
Test common odds ratio equals:
1
Continue
Cancel
Help

depressed * female Crosstabulation

			female .00	female 1.00	Total
depressed	.00	Count	101	143	244
		% within depressed	41.4%	58.6%	100.0%
	1.00	Count	10	40	50
		% within depressed	20.0%	80.0%	100.0%
Total		Count	111	183	294
		% within depressed	37.8%	62.2%	100.0%

Risk Estimate

	Value	95% Confidence Interval Lower	95% Confidence Interval Upper
Odds Ratio for depressed (.00 / 1.00)	2.825	1.350	5.911
For cohort female = .00	2.070	1.166	3.675
For cohort female = 1.00	.733	.615	.872
N of Valid Cases	294		

Odds

Risk

First, verify that for cohort female = 0 (i.e., males) the risk of not being depressed compared to being depressed is twice as high.

Then, look only within the column female = 0 (i.e., males)

- The proportion of NOT being depressed = 101/244
- The proportion of being depressed = 10/50
- This means the relative risk (which is the ratio of the two proportions) is (101/244)/(10/50) which is 2.07 as the table indicates

YOU DO THE SAME FOR COHORT FEMALE = 1 and get .733

depressed * female Crosstabulation

			female		
			.00	1.00	Total
depressed	.00	Count	101	143	244
		% within depressed	41.4%	58.6%	100.0%
	1.00	Count	10	40	50
		% within depressed	20.0%	80.0%	100.0%
Total		Count	111	183	294
		% within depressed	37.8%	62.2%	100.0%

NOW, let's compute the odds of NOT being depressed for males and females.

- For males the odds of not being depressed versus being depressed is:

(101/111)/(10/111) = (this is p/1-p) = 0.90991/0.09009 = 10.1

- For females the odds of not being depressed versus being depressed are much lower:

(143/183)/(40/180) = 0.781421/0.218579=3.575

- To double check this is correct, recall the odds of not being depressed versus being depressed for males divided by the odds of not being depressed versus being depressed for females (i.e., the odds ratio) is simply the ratio of the odds

Double check that 10.1/3.575 = 2.825 from the SPSS output

Trickery: to get the odds of BEING depressed (versus not) for males (versus females) just take the inverse 1/2.825 = .354

Compute the odds ratio of being unemployed for males and females; interpret

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for unemployed (.00 / 1.00)	1.546	.473	5.054
For cohort female = .00	1.338	.577	3.103
For cohort female = 1.00	.865	.613	1.220
N of Valid Cases	294		

The Chi-square Test

GOODNESS OF FIT AND INDEPENDENCE

Chi-Square

The χ^2 Goodness-of-Fit test

- Used when we have distributions of frequencies across two or more categories on ***<u>one</u>*** variable.
- Test determines how well a hypothesized distribution fits an obtained distribution.

The χ^2 test of independence.

- Used when we compare the distribution of frequencies across categories in two or more independent samples
- Used in a single sample when we want to know whether two categorical variables are related.

The Chi-Square Goodness of Fit Test

Our first real statistical test

Used for nominal level measures

The binomial distribution assumes two outcomes

The chi-square test can be used when there is *more than one outcome*; the result depends on how many degrees of freedom are associated with the test (which are associated with how many categories we have)

To define the degrees of freedom of a chi-square test, we ask how many categories would have to be known for us to predict the remaining categories with certainty?

Intuition: Chi-square Goodness of Fit test

Quarter Tossing

- Probability of Head?
- Probability of Tails?
- How can you tell if a Quarter is unfair when tossed?

- Imagine a flipped a quarter 50 times, what would we expect?

Heads	Tails
25	25

Intuition: What if we get the following on any toss?

Which of these scenarios seems probable with a “fair” coin?

Heads	Tails
20	30

Heads	Tails
15	35

Heads	Tails
10	40

Heads	Tails
5	45

Chi-square Goodness of fit test helps us know which distribution is most like what we observed in any toss

Ask: What would we expect given the coin is “fair” and what was actually observed in our toss?

	Heads	Tails
Observed	17	33
Expected	25	25
O-E	-8	8

We observed 8 less heads and 8 more tails than we would expect if the coin is fair. Is this statistically significant?

The idea

If your hypothesis is supported by data

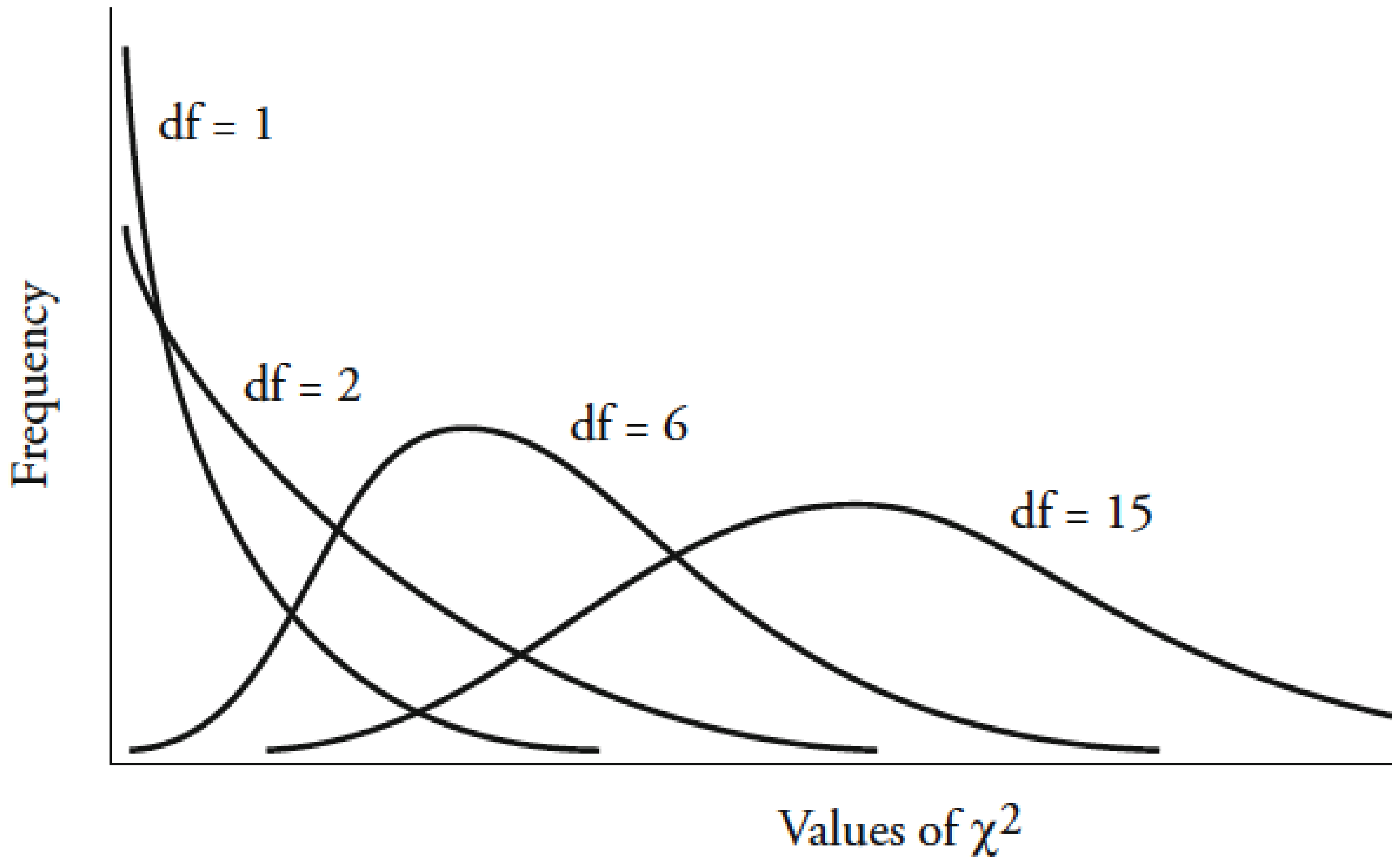
- you are claiming that the test is **random and independent**

If your hypothesis is not supported by data

- you are seeing that the deviation between observed and expected is very far apart…something non-random must be occurring….

The Chi-Square Distribution

chi-square distributions vary according to the number of degrees of freedom



Chi-square Goodness of fit test

The test statistic is compared to a theoretical probability distribution.

In order to use this distribution properly you need to determine the degrees of freedom.

We can test to see if our observed frequencies “Fit” our expectations

This is the χ^2 Goodness-of-Fit test

$$\chi^2 = \sum \frac{(O - E)^2}{E}; df = \#categories - 1$$

This converts the difference between the frequencies we observe and the frequencies we expect to a distribution with known probabilities

Calculating the Chi-Square Statistic

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_o - f_e)^2}{f_e} = \sum_{i=1}^{k} \frac{(Obs - Exp)^2}{Exp}$$

where:
f_o = O*bs* = observed frequency in a particular cell

f_e = *Exp* = expected frequency in a particular cell if H_0 is true

The expected frequency in cell *i* is np_i

Example: frequencies

	Age <=25	Age >25	
Male	10 (cell 1)	5	15
Not Male	3	17	20
	13	22	35

Prob(Male) = 10/35 → $np_1 = 10$

Example: degrees of freedom → the number of parameters that can vary 'freely'

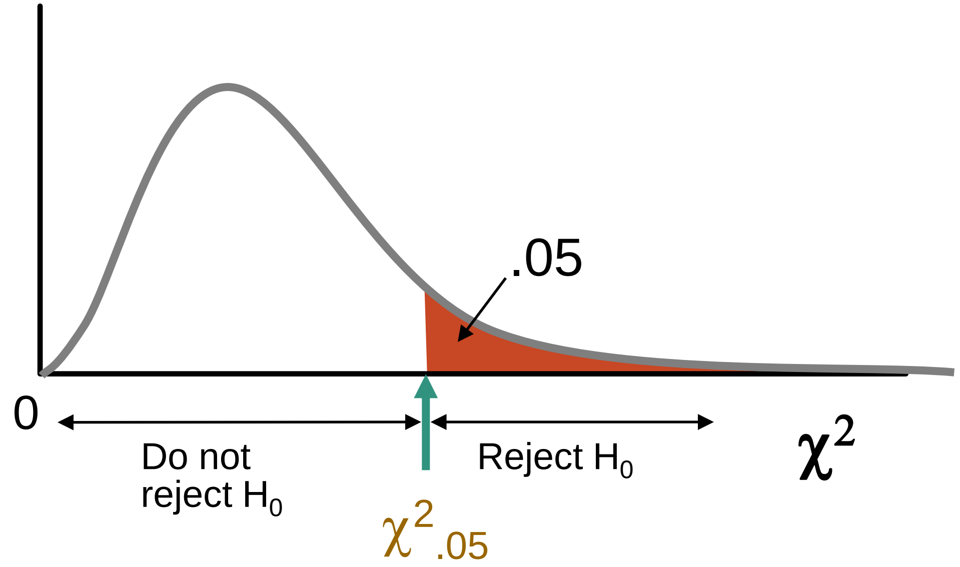
	Age <=25	Age >25	
Male	10	5	15
Not Male	3	17	20
	13	22	35

df = (c-1)(r-1) where c = # of columns; r = # of rows

The Chi-Square Test Statistic

1. The χ^2 test statistic approximately follows a chi-squared distribution with k-1 degrees of freedom, where k is the number of categories.
2. If the χ^2 test statistic is large, this is evidence against the null hypothesis. $\boxed{\chi^2 = \sum_{\text{all cells}} \frac{(Obs - Exp)^2}{Exp}}$

Decision Rule:
If $\chi^2 > \chi^2_{.05}$, or $p < \alpha$, reject H_0, otherwise, do not reject H_0.



Linking the Chi-Square Statistic to Probabilities: The Chi-Square Table

The “back of the book” approach presents a table of probabilities associated with chi-square distributions with varying ranges of degrees of freedom. The chi-square table does not give us the probability associated with every possible outcome, but rather provides probabilities and then lists the chi-square statistics associated with them.

Critical Values of χ^2 Distribution

	α					
df	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.001**
1	1.642	2.706	3.841	5.412	6.635	10.827
2	3.219	4.605	5.991	7.824	9.210	13.815
3	4.642	6.251	7.815	9.837	11.341	16.268
4	5.989	7.779	9.488	11.668	13.277	18.465
5	7.289	9.236	11.070	13.388	15.086	20.517
6	8.558	10.645	12.592	15.033	16.812	22.457
7	9.803	12.017	14.067	16.622	18.475	24.322
8	11.030	13.362	15.507	18.168	20.090	26.125
9	12.242	14.684	16.919	19.679	21.666	27.877
10	13.442	15.987	18.307	21.161	23.209	29.588
11	14.631	17.275	19.675	22.618	24.725	31.264
12	15.812	18.549	21.026	24.054	26.217	32.909
13	16.985	19.812	22.362	25.472	27.688	34.528
14	18.151	21.064	23.685	26.873	29.141	36.123
15	19.311	22.307	24.996	28.259	30.578	37.697
16	20.465	23.542	26.296	29.633	32.000	39.252
17	21.615	24.769	27.587	30.995	33.409	40.790
18	22.760	25.989	28.869	32.346	34.805	42.312
19	23.900	27.204	30.144	33.687	36.191	43.820
20	25.038	28.412	31.410	35.020	37.566	45.315
21	26.171	29.615	32.671	36.343	38.932	46.797
22	27.301	30.813	33.924	37.659	40.289	48.268
23	28.429	32.007	35.172	38.968	41.638	49.728
24	29.553	33.196	36.415	40.270	42.980	51.179
25	30.675	34.382	37.652	41.566	44.314	52.620
26	31.795	35.563	38.885	42.856	45.642	54.052
27	32.912	36.741	40.113	44.140	46.963	55.476
28	34.027	37.916	41.337	45.419	48.278	56.893
29	35.139	39.087	42.557	46.693	49.588	58.302
30	36.250	40.256	43.773	47.962	50.892	59.703

Source: From Table IV of R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research* (London: Longman Group Ltd., 1974). (Previously published by Oliver & Boyd, Edinburgh.) Reprinted by permission of Pearson Education Ltd.

We compute the chi-square test statistic and then look up the value for our significance level in the book.

For example, under one degree of freedom, a statistic of 2.706 is associated with a significance level of 0.10, a statistic of 3.841 with an value of 0.05, and a statistic of 10.827 with an value of 0.001.

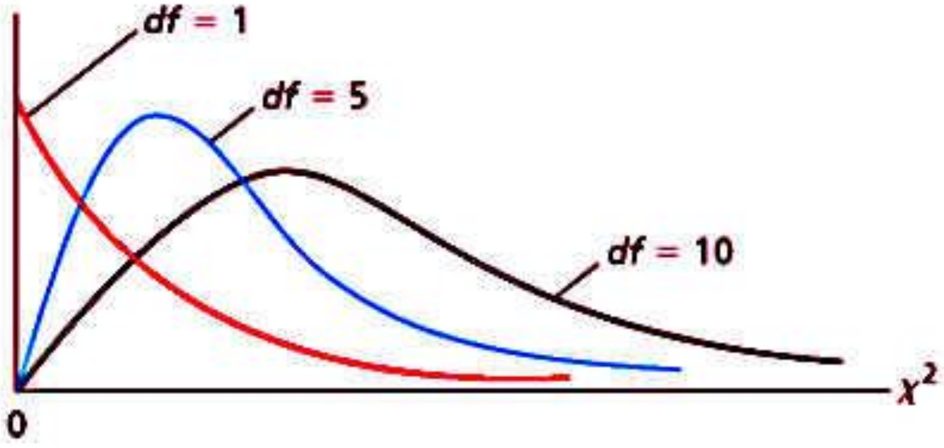
Chi-square Goodness of fit test

Hypothesis Test

1. H_0: P(heads) = .5
2. H_1: P(heads) ≠ .5
3. $\alpha = .05$
4. Type of test = χ^2 goodness-of-fit

Chi Square Distribution

5. DF = 2 – 1 = 1;



$\chi^2(1) = 3.841$; If χ^2 observed is larger than 3.841, reject the null hypothesis

Chi-square Goodness of fit test

6. Do the test: For our coin example

$$\chi^2 = \sum \frac{(O - E)^2}{E} = \frac{(17 - 25)^2}{25} + \frac{(33 - 25)^2}{25} =$$

$$\chi^2 = \frac{(-8)^2}{25} + \frac{(8)^2}{25} = \frac{64}{25} + \frac{64}{25} = 2.56 + 2.56 = 6.55$$

	Heads	Tails
Observed	17	33
Expected	25	25

7. Look up value: Since 6.55 > 3.84, reject the null.

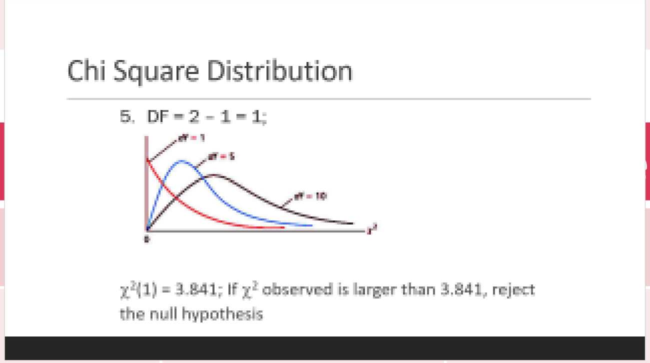
There is evidence that the coin is not fair.

df	0.20	0.10	0.05	0.02	0.01	0.001
1	1.642	2.706	3.841	5.412	6.635	10.827
2	3.219	4.605	5.991	7.824	9.210	13.815
3	4.642	6.251	7.815	9.837	11.341	16.268

Example: Car Accidents and day of the week

Are the accidents equally likely to occur on any day of the working week?

Number of Collisions by day of week					
Mon	Tue	Wed	Thu	Fri	Total
133	[illegible]	[illegible]	136	113	667



Num[illegible]ek						
Mon	[illegible]	[illegible]	Thu	Fri	Total	
133	[illegible]	[illegible]	136	113	667	Observed
133.4	133.4	133.4	133.4	133.4	667	Expected

	f_o	f_e	$(f_o - f_e)$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
Mon	**133**	133.4	-0.4	0.16	0.0011994
Tue	**126**	133.4	-7.4	54.76	0.41049475
Wed	**159**	133.4	25.6	655.36	4.91274363
Thu	**136**	133.4	2.6	6.76	0.05067466
FrI	**113**	133.4	-20.4	416.16	3.11964018
				$\sum_{i=1}^{k} \frac{(f_o - f_e)^2}{f_e} =$	8.49

	α					
df	0.20	0.10	0.05	0.02	0.01	0.001
1	1.642	2.706	3.841	5.412	6.635	10.827
2	3.219	4.605	5.991	7.824	9.210	13.815
3	4.642	6.251	7.815	9.837	11.341	16.268
4	5.989	7.779	9.488	11.668	13.277	18.465
5	7.289	9.236	11.070	13.388	15.086	20.517
6	8.558	10.645	12.592	15.033	16.812	22.457
7	9.803	12.017	14.067	16.622	18.475	24.322
8	11.030	13.362	15.507	18.168	20.090	26.125
9	12.242	14.684	16.919	19.679	21.666	27.877
10	13.442	15.987	18.307	21.161	23.209	29.588
11	14.631	17.275	19.675	22.618	24.725	31.264
12	15.812	18.549	21.026	24.054	26.217	32.909
13	16.985	19.812	22.362	25.472	27.688	34.528
14	18.151	21.064	23.685	26.873	29.141	36.123
15	19.311	22.307	24.996	28.259	30.578	37.697
16	20.465	23.542	26.296	29.633	32.000	39.252
17	21.615	24.769	27.587	30.995	33.409	40.790
18	22.760	25.989	28.869	32.346	34.805	42.312
19	23.900	27.204	30.144	33.687	36.191	43.820
20	25.038	28.412	31.410	35.020	37.566	45.315
21	26.171	29.615	32.671	36.343	38.932	46.797
22	27.301	30.813	33.924	37.659	40.289	48.268
23	28.429	32.007	35.172	38.968	41.638	49.728
24	29.553	33.196	36.415	40.270	42.980	51.179
25	30.675	34.382	37.652	41.566	44.314	52.620
26	31.795	35.563	38.885	42.856	45.642	54.052
27	32.912	36.741	40.113	44.140	46.963	55.476
28	34.027	37.916	41.337	45.419	48.278	56.893
29	35.139	39.087	42.557	46.693	49.588	58.302
30	36.250	40.256	43.773	47.962	50.892	59.703

Source: From Table IV of R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research* (London: Longman Group Ltd., 1974). (Previously published by Oliver & Boyd, Edinburgh.) Reprinted by permission of Pearson Education Ltd.

$7.78 < X^2 = 8.49 < 9.49$

The p-value of the test statistic is
.05 < p-value < .10
➔ DNR Null

We don't know the exact P-value but we DO know that P-value > 0.05, thus we conclude that … there is no significant evidence of different car accident rates for different weekdays when the driver was using a cell phone.

It turns out we can get the exact p-value from excel as follows

=CHISQ.TEST(array of observed values, array of expected values)

Compare this to .05 and make decision…

Example 3: Are successful people more likely to be born under some astrological signs than others? ☺

256 executives of Fortune 400 companies have birthday signs shown at the right.

There is some variation in the number of births per sign, and there are more Pisces.

Can we claim that successful people are more likely to be born under some signs than others?

Births (Observed)	Births (Expected)	Sign
23	21.3	Aries
20	21.3	Taurus
18	21.3	Gemini
23	21.3	Cancer
20	21.3	Leo
19	21.3	Virgo
18	21.3	Libra
21	21.3	Scorpio
19	21.3	Sagittarius
22	21.3	Capricorn
24	21.3	Aquarius
29	21.3	Pisces
256	256	

A Substantive Example: The Relationship Between Assault Victims and Offenders

Relationship Between Assault Victim and Offender	
CATEGORY	Frequency (N)
Stranger	166
Acquaintance	61
Friend	35
Boyfriend/girlfriend	38
Spouse	66
Other relative	44
Total	410

What is a simple research question?

State Assumptions & Perform Test

Level of measurement

Shape of the population distribution

Fully independent sampling with replacement

Null and Alternative Hypothesis

Sampling Distribution (and degrees of freedom)

Significance Level & Rejection Region

Test Statistic

Decision

Computation of Chi-Square for Type of Victim-Offender Relationship

	f_o	f_e	$(f_o - f_e)$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
Stranger	166	68.333	97.667	9,538.843	139.593
Acquaintance	61	68.333	7.333	53.773	0.787
Friend	35	68.333	33.333	1,111.089	16.260
Boyfriend/girlfriend	38	68.333	30.333	920.091	13.465
Spouse	66	68.333	2.333	5.443	0.080
Other relative	44	68.333	24.333	592.095	8.665
					178.849

Find critical value and compare calculated value --> make a decision

Revisit data

Calculate Chi-square test on depression by gender; unemployment by gender