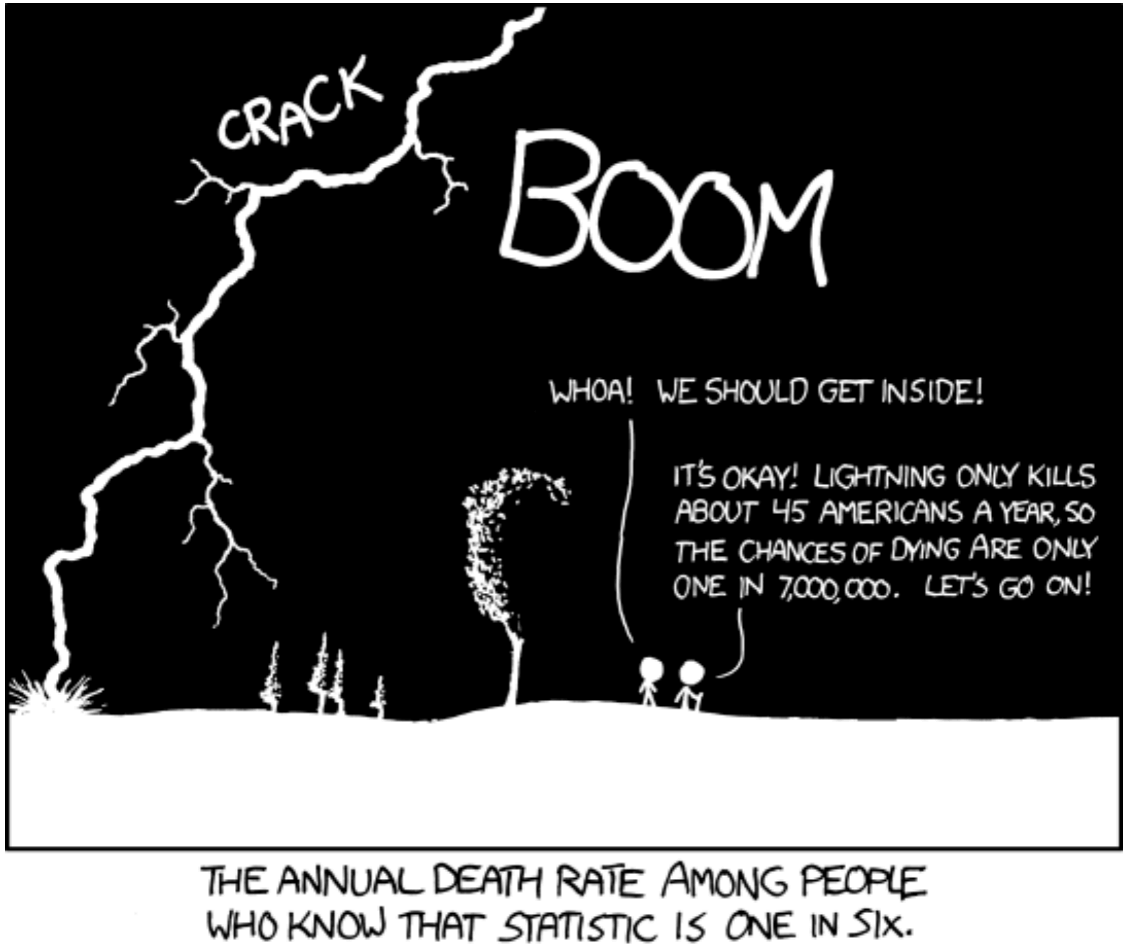
Probability and Discrete Variables

Biology 683



Heath Blackmon

Last week

1. If I say the 95% CI is 1.2-1.7, what do I mean?

2. What is the difference in a parameter and a statistic?

Today

1. Basic experimental reminders

2. Basics of Probability

3. Binomial and Chi-square tests

Some Experimental Design Considerations

Why do I need a control?

To interpret an experiment, we need to compare the experimental subjects to the correct reference group.

What about observational studies?

What is an appropriate control?

Ideal controls are identical to the experimental population, except for the one parameter being manipulated

The control population should be similar in all other respects to the experimental population

The control population should experience sham manipulations that simulate any manipulations applied to the experimental population

Sometimes you might need multiple different controls.

Avoiding Experimenter Bias

Experimenter bias is real

The results of your study can be influenced by your expectations

Some precautions

Randomize assignment of subjects to controls and treatments (**use R or random.org**).

Humans are bad at recognizing and creating randomness.

Avoiding Experimenter Bias

Use a blind or double-blind experimental design

Blind: the subject doesn't know whether it's an experimental or control subject

Double-blind: neither the researcher nor subject know which subjects are experimental versus control

How can you apply this to your research?

Confounding Variables

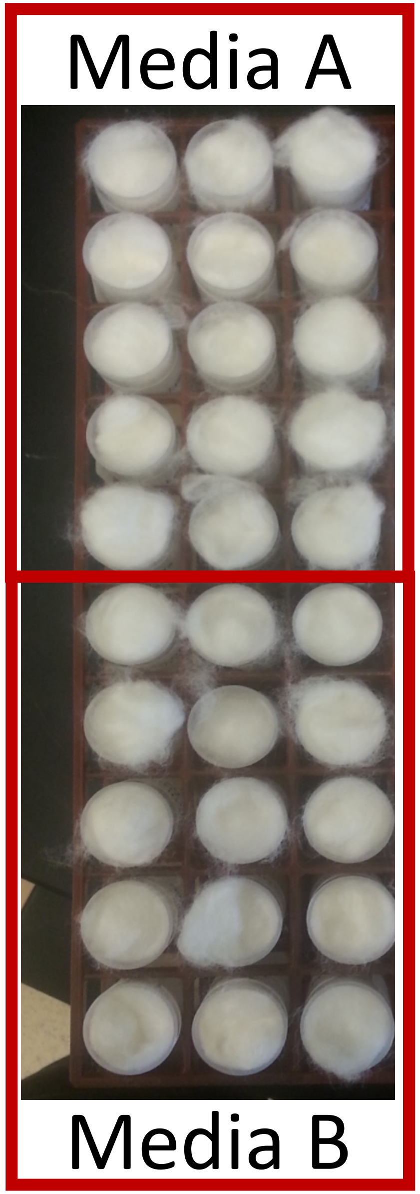
1. A difference between groups that the experimenter fails to account for
2. A hidden variable that creates an apparent causal relationship that isn’t real
3. **An experiment with confounded variables can be impossible to interpret and impossible to fix**

Confounding Example

Study type	Confounding variable
Gene expression level	recent gene duplication
Diversification	unobserved traits
Lung cancer and coffee	coffee smoking correlation
Behavior	maternal effects
Effective population size	breeding system

Redesign the procedure

- Collect 750 beetles from a population cage.
- Create 30 new vials with 25 beetles each.
- Make the first 15 of these control vials and use food media A.
- Make the next 15 of these experiment vials and use food media B.
- Place in a rack as shown and place in the incubator.
- Measure growth at day 15.



Pseudoreplication

Occurs when the apparent sample size is larger than true sample size

1. 10 rats are studied and tested on three consecutive days, resulting in 15 observations for the control group and 15 observations for the treatment groups
2. The experiment is conducted in two tanks: tank 1 has hormone added, tank 2 is the control tank. 10 fish are tested per tank.
3. We are testing for the effects of mating system on genome size. We use 5 outbreeding insects and 5 inbreeding species of beetles.
4. Beetles are segregated by sex into two vials, with 10 individuals per vial. I draw a male and female at random and test them, returning them to the vials at the end. I perform a total of 40 such tests.

Biological and Technical Replicates

- A biological replicate involves a new, independent test subject
- A technical replicate involves repeating the same procedure on a new sample from the same subject
- Technical replicates do not contribute to your estimates of population-level parameters, but they can increase the precision of measurements on individuals

Which kind of replication

- In general, biological replicates are superior to technical replicates, because biological replicates increase power.

- Technical replicates are useful when the technique in question sometimes produces extremely inaccurate results, which must be pruned from the dataset. An example is qPCR, where occasional extreme outliers are common.

Best Practices

- Ensure as much as possible that controls and experimental individuals are from identical populations (except for the factor of interest)

- Treat your controls as similarly as possible to the experimental subjects (sham injections, placebos, etc.)

- Conduct your control manipulations in parallel with your experimental manipulations

- Think about all possible confounding variables and establish a plan to eliminate or correct for them before you start!

Everything I do is an Experiment

You should approach everything you do in the lab from the perspective of an experiment

Always do the appropriate controls for PCR, transformations, etc.

Troubleshooting is experimenting

Think about how you will describe the experiment before you embark on it

You will see that simplicity is extremely valuable

Think about the analysis you will do before you get started

The Null Hypothesis

To analyze your data, you will need a statistical hypothesis to go with your scientific hypothesis

A statistical hypothesis is most easily constructed as a null hypothesis

A null hypothesis posits that the factor of interest has no effect

Frequentist test we will be looking at p-value $\Pr(statistic|null)$

Bayesian approaches usually tells us if the posterior estimate of the parameter of interest overlap in our two treatments.

Examples of Null Hypotheses

Fertilizer has no effect on the growth rate of oak trees.

Blocking olfactory cues has no effect on mate choice in swordtail fishes.

Rates of genome evolution are the same in two populations.

Mutations in the 5’ UTR of msl-2 have no effect on translation.

Rejecting the Null

- Your statistical test will attempt to reject the null hypothesis

- If you reject the null, then one of the alternative hypotheses must be true – though not necessarily the one you believe to be true!

- You cannot **<u>prove</u>** a hypothesis, but
  - As frequentist you can find support for an alternative by rejecting the null. The more convincing the null and the more well designed the experiment the more evidence you provide for your alternative.
  - As a Bayesian you can compare support for two competing hypotheses.

Type I versus Type II Error

Type I error refers to rejecting a true null hypothesis

Type II error refers to failing to reject a false null hypothesis

Power is a description of our probability of rejecting a false null hypothesis

We usually set up statistical tests to avoid Type I errors, at the expense of possibly committing Type II errors

Type I error = FALSE POSITIVE
1 – Type 2 error = POWER

Analyzing Proportions

Several chapters in the book deal with this topic

The experiment boils down to this:

- Your subjects have some alternative outcomes
- Each individual has some probability of each outcome
- You are trying to find the conditions that impact that probability

When would this type of problem come up in the biological sciences?

Terms to know for probability

Sample space: All the potential outcomes of a random trial.

Probability: The proportion of events with a given outcome if the random trial was repeated many times.

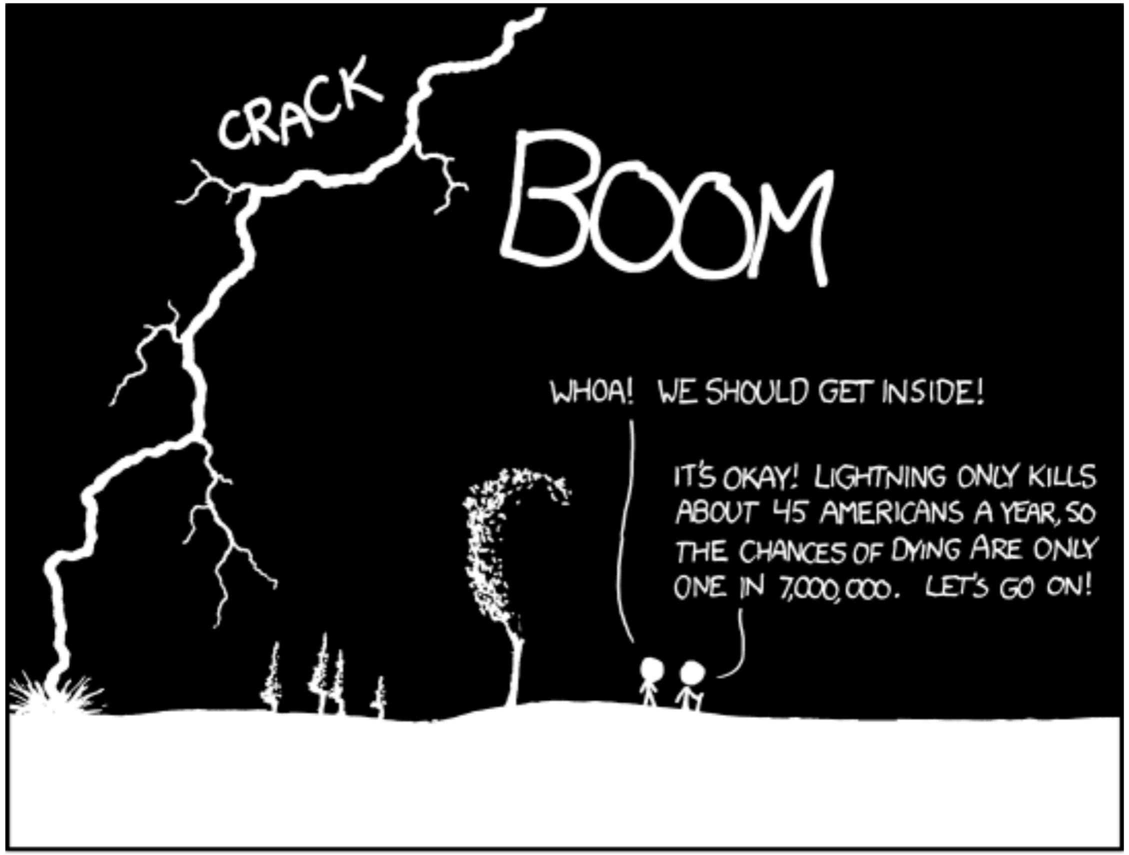
Mutually exclusive: If one outcome excludes the others, they are mutually exclusive.

Conditional probability: The probability of one outcome, if we know that some other outcome occurred.

Independent: When one outcome provides no information about another, they are independent.

Non-Independent: When knowing one outcome changes the probability of another, they are non-independent.

Terms to know for probability



THE ANNUAL DEATH RATE AMONG PEOPLE WHO KNOW THAT STATISTIC IS ONE IN SIX.

Conditional probability: The probability of being struck by lightning changes if you know the statistic about 45 American a year being killed by lightning

A simple example

What is the probability of drawing an ace then a king from a deck cards?

What is the sample space?

Are the events independent or dependent?

What is the probability of this event?

What is the conditional probability of drawing a king if we have already drawn an ace?

assume standard deck with no jokers

What is a p-value?

Is the probability of finding the observed, or more extreme, statistic when the null hypothesis is true (generating the data).

```
> x
     [,1] [,2]
[1,]  140    4
[2,]   80   13
[3,]   76   89
[4,]   20    3
> chisq.test(x)

	Pearson's Chi-squared test

data:  x
X-squared = 117.31, df = 3, p-value < 2.2e-16
```

Number of women on titanic who survived (first column) or died (second column) in first, second, third, or crew classes (rows 1:4 respectively).

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Misconceptions about p-values

The p-value is not the probability that the observed statistic is due to random chance.

A p-value is not the probability that your alternative is false.

A p-value is not the probability that the null hypothesis is true.

The magnitude of the p-value does not indicate the importance of an effect. Statistical significance does not equate to biological significance.

Studies with p-values on opposite sides of 0.05 are equally “correct”.

There are many ways of calculating p-values

Traditional statistical tests

For many questions/experiments there isn’t a ready made statistical test.
- Randomization of datasets
- Comparison to simulated datasets

Binomial Test

A test to determine whether or not the observed proportion adheres to the expected proportion under the null hypothesis

Some possible uses:

- Are frogs equally likely to be right or left handed?
- Is the sex ratio half male and half female?
- Are the offspring phenotypes a 3:1 ratio?
- Do some beetles win more fights?

Binomial Test

As in most statistical tests, a test statistic is compared to a distribution

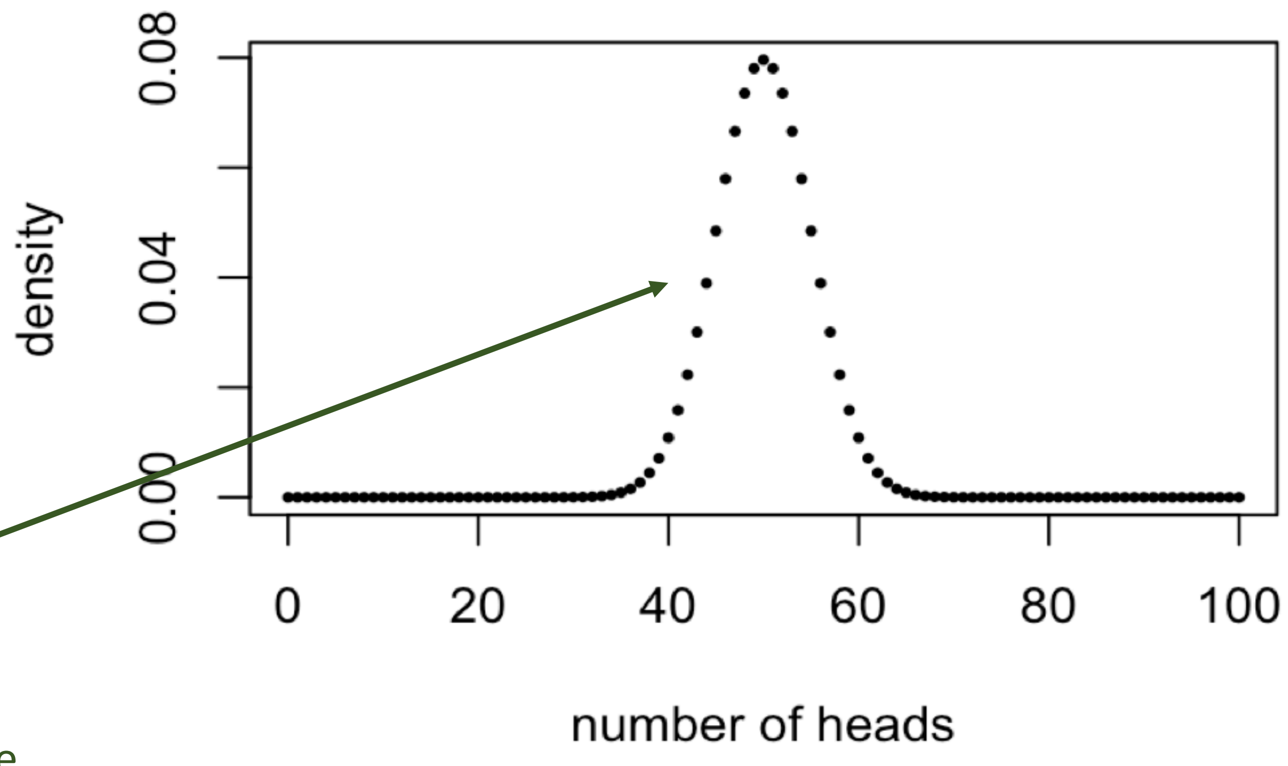
In this case, the test statistic is just the observed number (number of right-handed toads, number of females in the population, number of fights won)

Note that this test is only appropriate when there are two categories of individuals and your hypothesis allows you to provide a probability of the outcomes.

Binomial Test

With the binomial test our null hypothesis is the probability of one of the two outcomes. This probability and the number of observations defines the distribution we will compare our observation to.

```
x <- 0:100
y <- dbinom(x, size = 100, prob = .5)
plot(y~x, pch=16, cex=.5,
     xlab="number of heads")
```



Distribution when the null is 50% and we have 100 observations

use a simulation to see if you can replicate this curve

Binomial Test

Lets look at an example with sex ratio. You are hybridizing closely related species (with XY sex chromosomes) so you know Haldane's rule states that the males might be more rare. When you survey the offspring you find 23 males out of 65 offspring. Does this result support Haldane's rule occuring in your system?

```
binom.test(x = 23, n = 65, p = .5)
```

Binomial Test

Lets look at an example with sex ratio. You are hybridizing closely related species (with XY sex chromosomes) so you know Haldane's rule states that the males might be more rare. When you survey the offspring you find 23 males out of 65 offspring. Does this result support Haldane's rule occuring in your system?

```
binom.test(x = 23, n = 65, p = .5)

  data:  23 and 65
  number of successes = 23, number of trials = 65,
  p-value = 0.02481
```

Binomial Test

binom.test has an argument `alternative`

`alternative`	indicates the alternative hypothesis and must be one of `"two.sided"`, `"greater"` or `"less"`. You can specify just the initial letter.

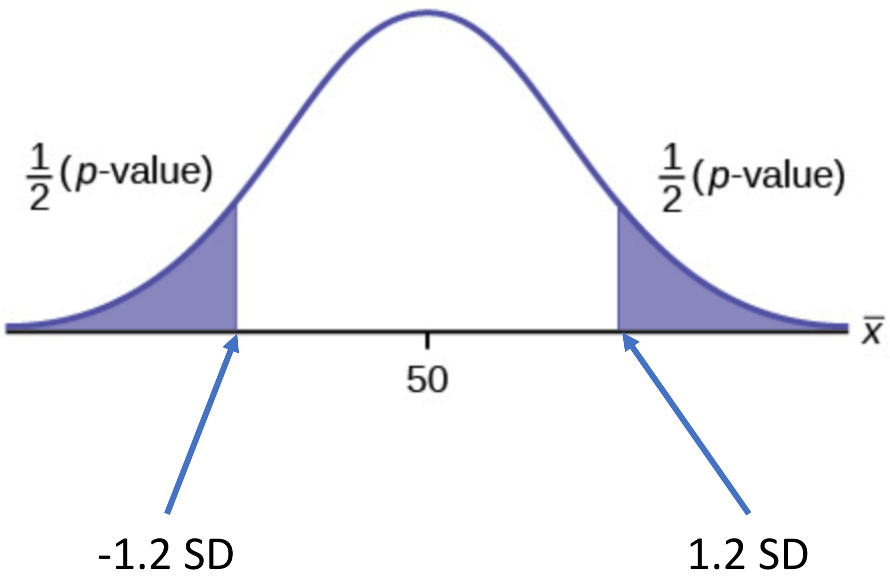
```
binom.test(x = 23, n = 65, p = .5, alternative = "t") # p-value 0.02481
binom.test(x = 23, n = 65, p = .5, alternative = "g") # p-value 0.99370
binom.test(x = 23, n = 65, p = .5, alternative = "l") # p-value 0.01241
```

Binomial Test

Alternative = two.sided

What is the probability that I would see a skew in the sex ratio this great or greater.

In this case our observed number of males was -1.2 standard deviations from the mean. So our p-value is the area under the curves above 1.2SD and below -1.2SD.

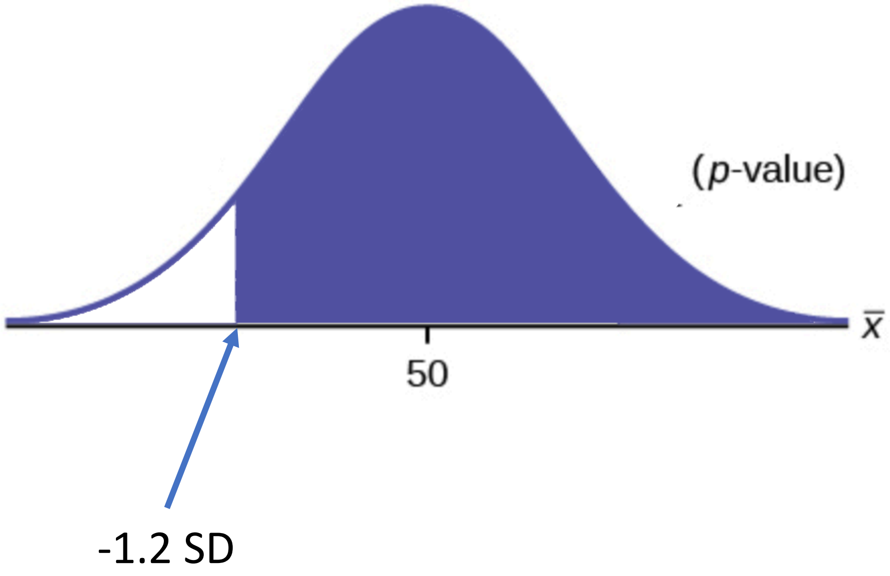


Binomial Test

Alternative = greater

What is the probability that I would see a larger number of males.

In this case our observed number of males was -1.2 standard deviations from the mean. So our p-value is the area under the curves above -1.2SD.

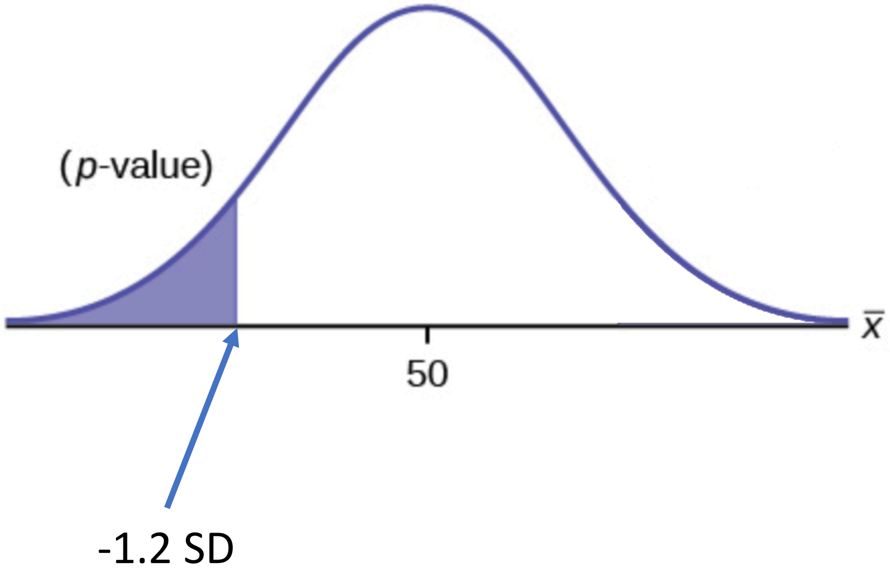


Binomial Test

Alternative = less

What is the probability that I would see this many or fewer males.

In this case our observed number of males was -1.2 standard deviations from the mean. So our p-value is the area under the curves below -1.2SD.



If before we collected this data we wanted to answer the question: “Are males more rare than females?” Then we have an apriori hypothesis that we should see fewer males than females so we could justify using this more powerful test.

Reporting the Results

```
> binom.test(x = 23, n = 65, p = .5, alternative = "l")

	Exact binomial test

data:  23 and 65
number of successes = 23, number of trials = 65,
p-value = 0.01241
alternative hypothesis: true probability of success is less than 0.5
95 percent confidence interval:
 0.0000000 0.4627116
sample estimates:
probability of success
             0.3538462
```

Our offspring ratio shows a significantly fewer males than would be expected under a 1:1 sex ratio (0.35, 95% CI: 0.24-0.48, binomial test, $n = 65$, $p < 0.025$).

Reporting the Results

This populations shows a significantly fewer males than would be expected under a 1:1 sex ration (0.35, 95% CI: 0.24-0.48, binomial test, $n = 65$, $p < 0.025$).

For very small *p*-values, we just say that *p* is very small (< 0.001 or < 0.0001).

Most journals/subdisciplines will have conventions about how certain tests are reported.

Most journals italicize mathematical variables, so *n* and *p* would be italicized. They also normally would be lower case.

χ^2 Test

This test compares the observed number in each category to expectations based on the null hypothesis (if there are only two categories, it approximates the binomial test with probability of 50%)

It can also be used to test for independence of two variables, and then it is called a contingency $\chi 2$-test.

We will use data from the Titanic and see if some females were more likely to survive than others.

Female adults on the Titanic		
	Survived	Died
1st	140	4
2nd	80	13
3rd	76	89
Crew	20	3

χ^2 Test

To calculate the statistic we just sum up the standardized deviations from the expected values in each category.

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

χ^2 Test

To calculate the statistic we just sum up the standardized deviations from the expected values in each category.

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Female adults on the Titanic			
	Survived	Died	total
1st	140	4	144
2nd	80	13	93
3rd	76	89	165
Crew	20	3	23
total	74.4%	25.6%	

χ^2 Test

To calculate the statistic we just sum up the standardized deviations from the expected values.

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Female adults on the Titanic			
	Survived	Died	
1st	140	4	144
2nd	80	13	93
3rd	76	89	165
Crew	20	3	23
total	74.4%	25.6%	

Expected		
	Survived	Died
1st	.744 x 144	.256 x 144
2nd	.744 x 93	.256 x 93
3rd	.744 x 165	.256 x 165
Crew	.744 x 23	.256 x 23

χ^2 Test

To calculate the statistic we just sum up the standardized deviations from the expected values.

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Female adults on the Titanic			
	Survived	Died	
1st	140	4	144
2nd	80	13	93
3rd	76	89	165
Crew	20	3	23
total	74.4%	25.6%	

Expected		
	Survived	Died
1st	107	37
2nd	69	24
3rd	123	42
Crew	17	6

χ^2 Test

To calculate the statistic we just sum up the standardized deviations from the expected values.

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

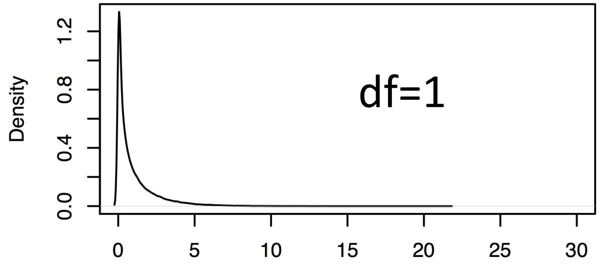
Observed		
	Survived	Died
1st	140	4
2nd	80	13
3rd	76	89
Crew	20	3

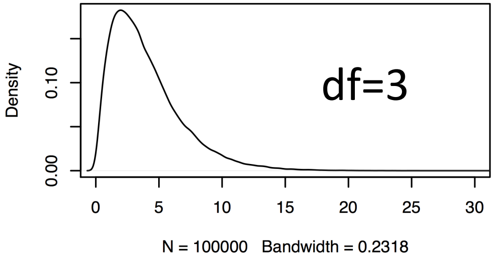
Expected		
	Survived	Died
1st	107	37
2nd	69	24
3rd	123	42
Crew	17	6

$$\chi^2 = 117$$

χ^2 Test

The shape of the chi square distribution depends on the degrees of freedom (df).





df = (no. rows – 1)(no. cols -1)

Female adults on the Titanic		
	Survived	Died
1st	140	4
2nd	80	13
3rd	76	89
Crew	20	3

df = (4– 1)(2 -1)

df = 3

```
> x
     [,1] [,2]
[1,]  140    4
[2,]   80   13
[3,]   76   89
[4,]   20    3
> chisq.test(x)

	Pearson's Chi-squared test

data:  x
X-squared = 117.31, df = 3, p-value < 2.2e-16
```

My take

Much of what was presented today forms the basis of logically thinking about your experimental results. Similarly these topics are key to your ability to evaluate the evidence presented in the papers that you read. I will definitely ask you multiple questions on your midterm and final that will require you to think clearly about these topics.

My expectations for you

Concepts		R coding	Stats
Reproducibility crisis	Probability	<- [] + - / %% : == !=	binom.test
Importance of stats training	Bayesian vs ML	%in%	chisq.test
Rules for plots	Null hypotheses	c, rep, data.frame, list, for,	
Populations and samples	P-values	if, print, rnorm, plot, hist,	
Types of data	Replicates (bio/tech)	density, names, barplot,	
Data summary		arrows, lines, points, text,	
Uncertainty		rgb, polygon, read.csv,	

Thursday

- Evaluate sex ratio of in two frog crosses. We cross species 1 and 2 and obtain 126 offspring 52 of them are male.
- 1) what is our null hypothesis going to be?
- 2) does this result support Haldane's rule?
- 3) what is the minimum number of offspring required to detect a significant deviation from our expectation under the null hypothesis?

- We cross females from one strains of fish with males from another strain. A proportion of our offspring have an unusual color pattern.

	Males	Females
Color Pattern Present	31	4
Color Pattern Absent	68	89

- 1) What null might we construct for this data?
- 2) Can we reject this null?
- 3) What might we infer from this data?