Testing proportions

BIO5312 FALL2017

STEPHANIE J. SPIELMAN, PHD

Estimation

An **estimator** is a statistic (~formula) for estimating a parameter

A good estimator is **unbiased**
- The expected value (expectation) of the estimator should equal the parameter being estimated
- Mean of the sampling distribution of the statistic should equal the parameter being estimated

A good estimator is **consistent**
- Increasing the sample size produces an estimate with smaller SE

A good estimator is **efficient**
- Has the smallest SE among any estimator you could have chosen

We are usually interested in point estimate, SE, and CI

Normally-distributed variable

- $\widehat{\mu} = \bar{x}$

- $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$

- Known σ
  - SE = $\frac{\sigma}{\sqrt{n}}$
  - 95% CI = $\bar{x} \pm Z_{0.025} SE$

- Unknown σ
  - SE = $\frac{s}{\sqrt{n}}$
  - 95% CI = $\bar{x} \pm t_{0.025} SE$

Hypothesis testing frameworks

t-tests compare means for continuous quantitative data

Today we will learn to analyze discrete count data ("proportions"):

- ◦ Binomial test
- ◦ χ^2 goodness-of-fit
- ◦ Contingency table analysis
  - ◦ χ^2 association/homogeneity and Fisher exact test

Binomial test

$$P(k\ successes) = \binom{n}{k}p^k(1-p)^{(n-k)} = \binom{n}{k}p^k q^{(n-k)}$$

◦ Binomial coefficient: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

Null proportion of successes to test against

Hypothesis test:

◦ H_0 : The relative frequency of success in the underlying population is p_0

◦ H_A : The relative frequency of success in the underlying population is not p_0

◦ H_A : The relative frequency of success in the underlying population is $> / < p_0$

Binomial test assumption: BInS conditions are satisfied

Binary outcomes

Independent trials (outcomes do not influence each other)

n is fixed before the trials begin

Same probability of success, p, for all trials

Binomial test: Example

In a certain species of wasp, each wasp has a 30% chance of being male. I collect 12 wasps, of which 5 are male. Does my sample show evidence that 30% of wasps are male? Use α=0.05.

In other words, is the observed success proportion 5/12 (41.67%) consistent with a population whose probability of success is 0.3?

Verifying assumptions

Binary outcomes: Male or female

Independent trials: Wasp sex does not influence sex of other wasps

n is fixed before the trials begin: I collect 12 wasps

Same probability of success, p, for all trials: P(male) = 0.3 for every wasp
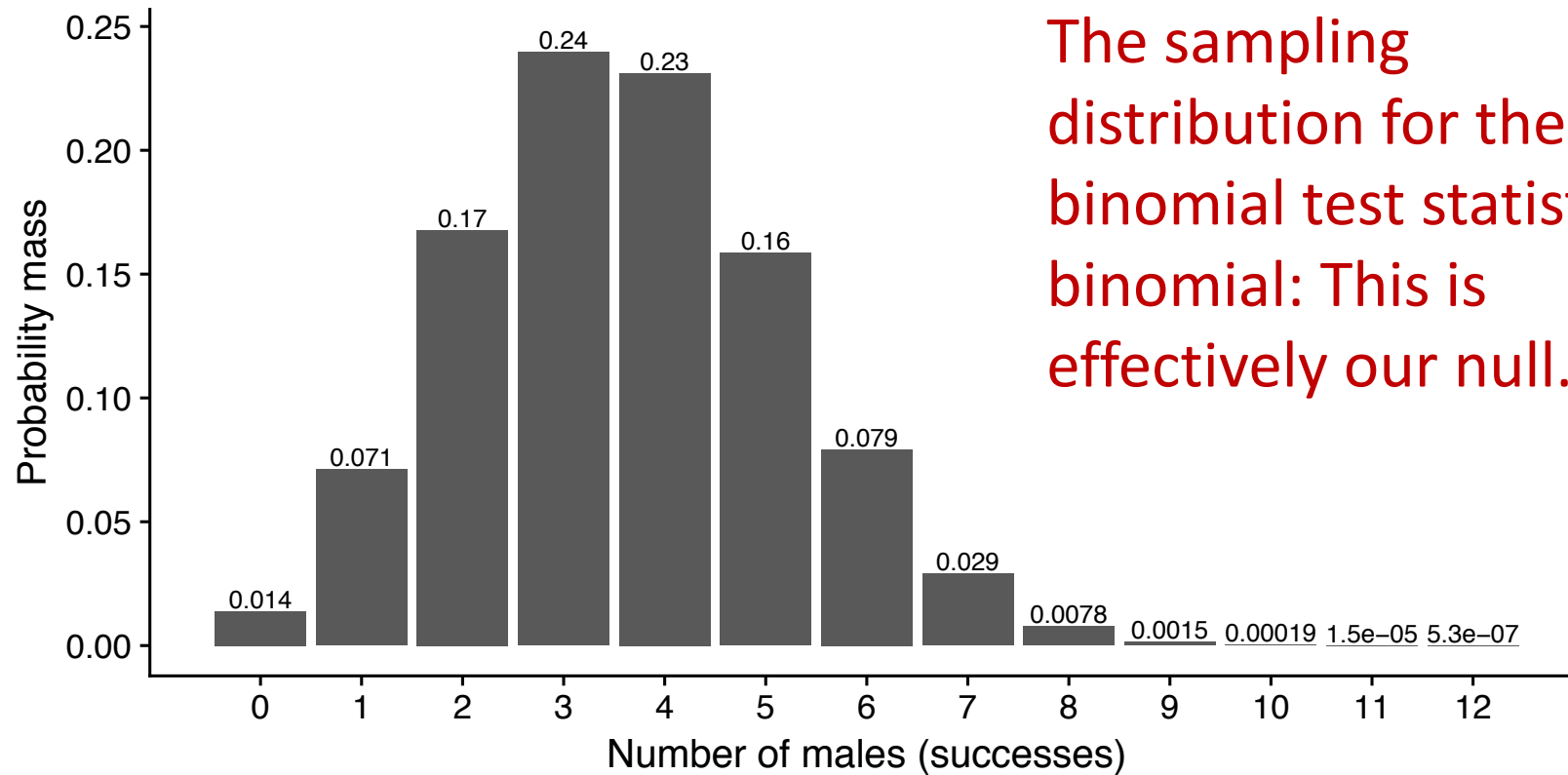
Performing the binomial test

My sample:
- p = 5/12 = 0.417
- n = 12
- X = 5 ⟵ We generally say X instead of k when performing hypothesis tests, by convention

H_0 : The probability of being a male wasp is p_0 = 0.3

H_A: The probability of being a male wasp differs from p_0 = 0.3
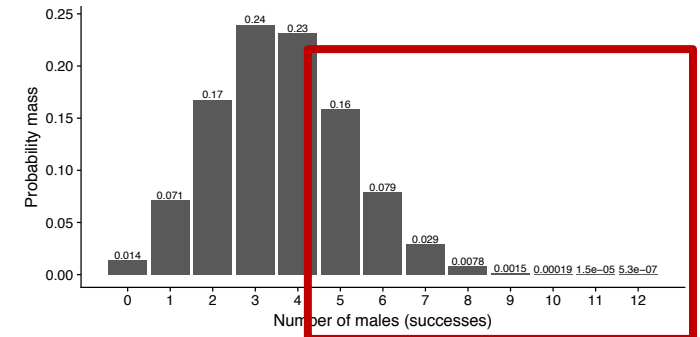
The PMF for wasp sex

The sampling distribution for the binomial test statistic is binomial: This is effectively our null.

Performing the test

Recall, the P-value is the probability of obtaining a result *as extreme or more*

◦ Therefore, P-value is P(number of successes >=5)

$$P(X \geq 5) = \binom{12}{5}0.3^5 0.7^{(12-5)} + \binom{12}{6}0.3^6 0.7^{(12-6)} + \cdots + \binom{12}{12}0.3^{12} 0.7^{(12-12)}$$

```
> 1 - pbinom(4, 12, 0.3)
[1] 0.2673445
```

Conclusions, round 1

Our P-value of 0.276 is much greater than α. Therefore we fail to reject the null hypothesis and we have no evidence that the population proportion of males corresponding to our sample differs from 0.3.

Notes on binomial tests

Computing two-sided P-values is non-trivial

◦ Binomial distribution symmetric **only when p=0.5**

```
> binom.test(5, 12, 0.3)
Exact binomial test
```

This is not 0.276*2!

```
data:  5 and 12
number of successes = 5, number of trials = 12, p-value = 0.3614
alternative hypothesis: true probability of success is not equal to 0.3
95 percent confidence interval:
 0.1516522 0.7233303
sample estimates:
probability of success
          0.4166667
```

Computing the binomial standard error

$$SE_{\hat{p}} = \sqrt{\hat{p}(1 - \hat{p})/n}$$

$$= \sqrt{\frac{0.417(1 - 0.417)}{12}} = 0.142$$

What is this value?

1. The standard deviation of the sampling distribution of the probability of success
2. Quantifies the precision of $\hat{p}$, our estimate of the population prob. of success
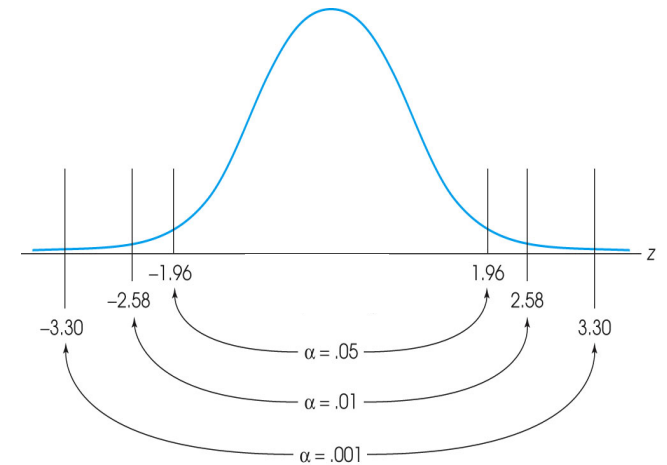
Computing the binomial confidence interval

Classically, we use the **Wald method**
- Note: Only "precise" when n is not very large (>0.8) or small (<0.2)

$$\hat{p} - \left(Z_{0.025} * SE_{\hat{p}}\right) < p < \hat{p} + \left(Z_{0.025} * SE_{\hat{p}}\right)$$

- $\hat{p}$ is the estimated proportion of success, X/n = 0.417
- $Z_{0.025}$ is 1.96
- $SE_{\hat{p}} = \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\dfrac{0.417(1-0.417)}{12}} = 0.142$

Calculating the binomial CI

$$\hat{p} - (Z_{0.025} * SE_{\hat{p}}) < p < \hat{p} + (Z_{0.025} * SE_{\hat{p}})$$

0.417 − 0.278 < p < 0.417 + 0.278 →   0.417 ± 0.278

```
> binom.test(5, 12, 0.3)
Exact binomial test


data:  5 and 12
number of successes = 5, number of trials = 12, p-value = 0.3614
alternative hypothesis: true probability of success is not equal to 0.3
95 percent confidence interval:
 0.1516522 0.7233303
sample estimates:
probability of success
             0.4166667
```

**R uses a more exact method, the Clopper-Pearson interval**

Final conclusions

Our P-value of 0.276 is much greater than α. Therefore we fail to reject the null hypothesis and we have no evidence that the population proportion of males corresponding to our sample differs from 0.3.

Our estimated proportion of success is 0.417 with SE =0.142 and a 95% CI of 0.417 ± 0.278.

Pause: Binomial exercise

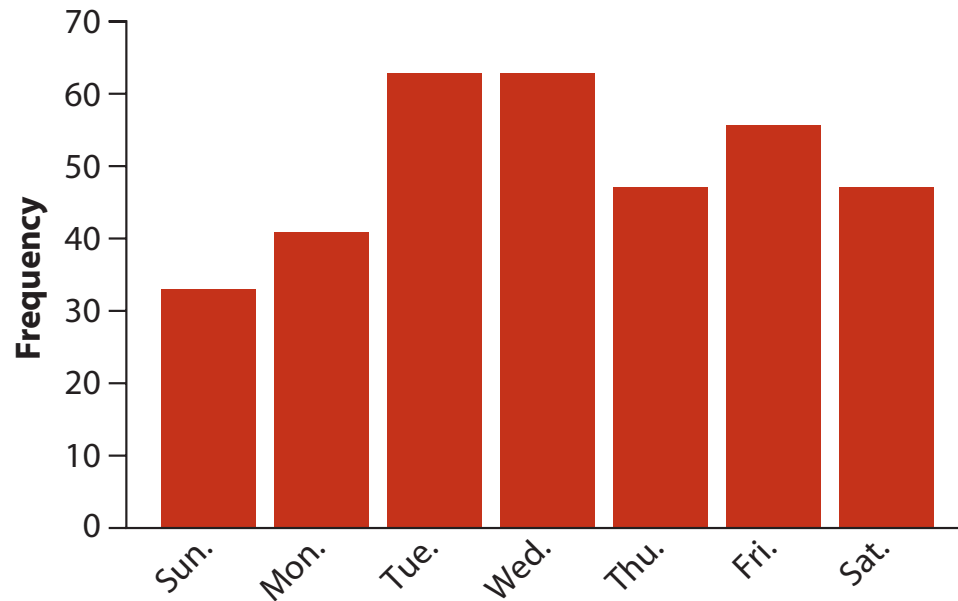# Use χ^2 Goodness-of-fit test if we do not have binary outcomes

Goodness-of-fit test asks if observed proportions are equal to a null proportion
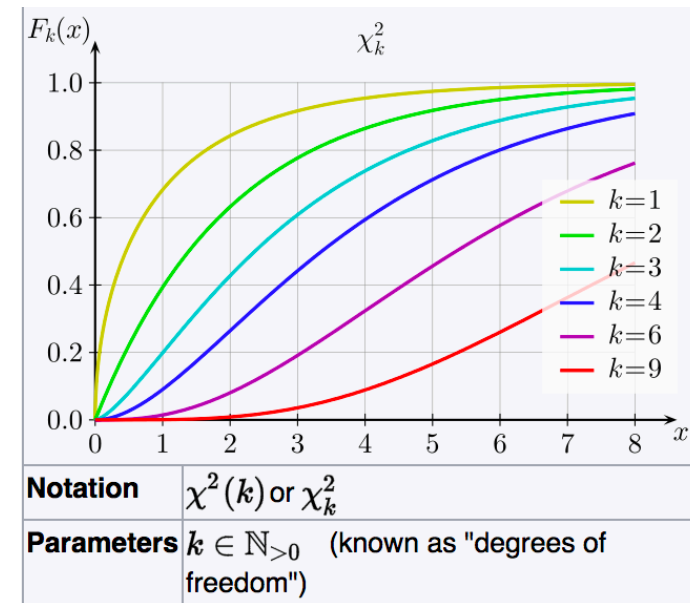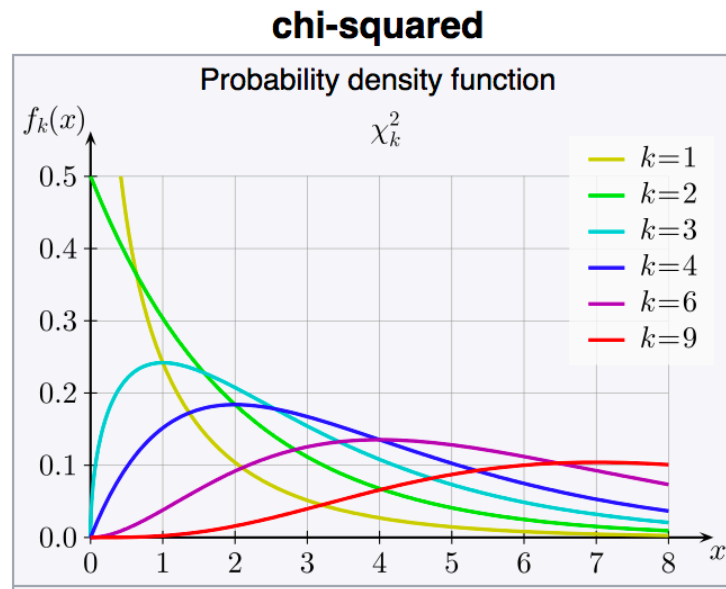
df = (number of categories) − 1 − (number of parameters estimated from data)

0 for goodness-of-fit test

Example: Are babies born with the same frequency every day of the week?

Day in 1999	# births
Sunday	33
Monday	41
Tuesday	63
Wednesday	63
Thursday	47
Friday	56
Saturday	47

H_0 : The probability of birth was the same every day of the week in 1999.

H_A: The probability of birth was not the same every day of the week in 1999.

Test statistic

$$\chi^2 = \sum_i \frac{(\# \, observed_i - \# \, expected_i)^2}{\# \, expected_i}$$

| Day	# Observed births	# days in 1999	Expected prop	# Expected births
Sunday	33	52	52/365 = 0.142	0.142*52 = 49.863
Monday	41	52	0.142	49.863
Tuesday	63	52	0.142	49.863
Wednesday	63	52	0.142	49.863
Thursday	47	52	0.142	49.863
Friday	56	53	0.145	50.822
Saturday	47	52	0.142	49.863
Total	**350**	**365**	**1**	**1**

Calculating the test statistic and df

$$\chi^2 = \sum_i \frac{(\# \ observed_i - \# \ expected_i)^2}{\# \ expected_i}$$

$$= \frac{(33-49.863)^2}{49.863} + \frac{(41-49.863)^2}{49.863} + \frac{(63-49.863)^2}{49.863} + \frac{(63-49.863)^2}{49.863} + \frac{(47-49.863)^2}{49.863} + \frac{(56-50.822)^2}{50.822} + \frac{(47-49.863)^2}{49.863}$$

= 15.05

df = #categories − 1 = 7 − 1 = 6

Our categorical variable is **Days of week**
It has seven **categories**

Day	# Observed births	# Expected births
Sunday	33	0.142*52 = 49.863
Monday	41	49.863
Tuesday	63	49.863
Wednesday	63	49.863
Thursday	47	49.863
Friday	56	50.822
Saturday	47	49.863
Total	**350**	**1**

Reports and conclusions
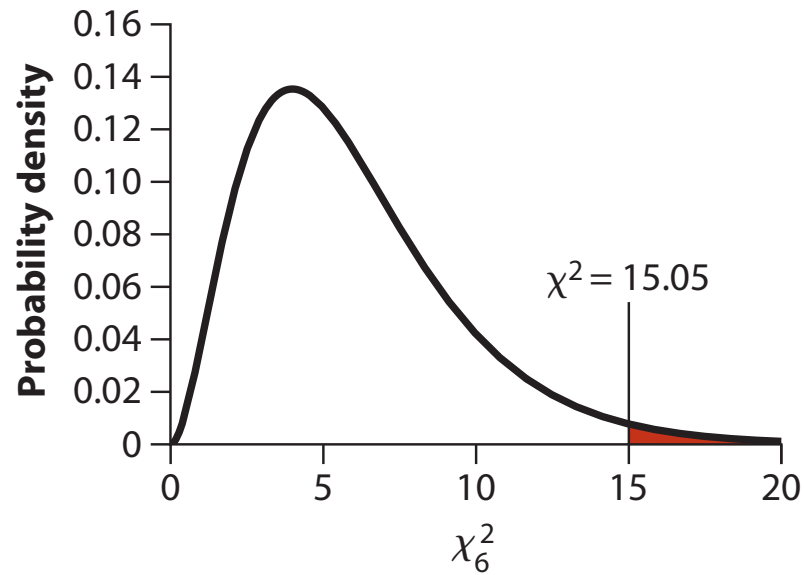

```
> 1 - pchisq(15.05, 6)
[1] 0.01987137
```

At 0.0199, we reject the null hypothesis that are births are equally distributed across days in 1999. We have evidence that frequency of births differs across days.

Notes on χ^2 Goodness-of-fit test

Assumptions for **all** χ^2 tests

- Randomly sampled data from population
- Two or more categories of a categorical variable (data is **counts**)
- Expected frequencies must be >=1
- No more than 20% of expected frequencies are < 5

We take only >= test statistic for P-value

- General to all χ^2 tests

χ^2 goodness-of-fit in R

```
#### Prepare data: Observed counts and expected proportions ####
> births <- c(33,41,63,63,47,56,47)
> expected <- c(52,52,52,52,52,53,52)
> expected <- expected/sum(expected)
> expected
[1] 0.1424658 0.1424658 0.1424658 0.1424658 0.1424658 0.1452055 0.1424658

> chisq.test(births, p = expected)


        Chi-squared test for given probabilities


data:  births
X-squared = 15.057, df = 6, p-value = 0.01982
```

Binomial is preferred for two groups

Temple University students are 52% female, 48% male. Does this class reflect the Temple student population?

We have 19 students: 7 females and 12 males.

Binomial P-values are more precise

```
> binom.test(7, 19, 0.52)


        Exact binomial test


data:  7 and 19
number of successes = 7, number of trials = 19, p-value = 0.251
alternative hypothesis: true probability of success is not equal to 0.52
95 percent confidence interval:
 0.1628859 0.6164221
sample estimates:
probability of success
             0.3684211
```

```
> chisq.test(c(7,12), p = c(0.52, 0.48))


        Chi-squared test for given probabilities


data:  c(7, 12)
X-squared = 1.749, df = 1, p-value = 0.186
```

Pause: Goodness of fit exercise

Contingency table analysis

Test for an association between two (or more) categorical variables

◦ Are heart attacks more likely for people who take aspirin daily?

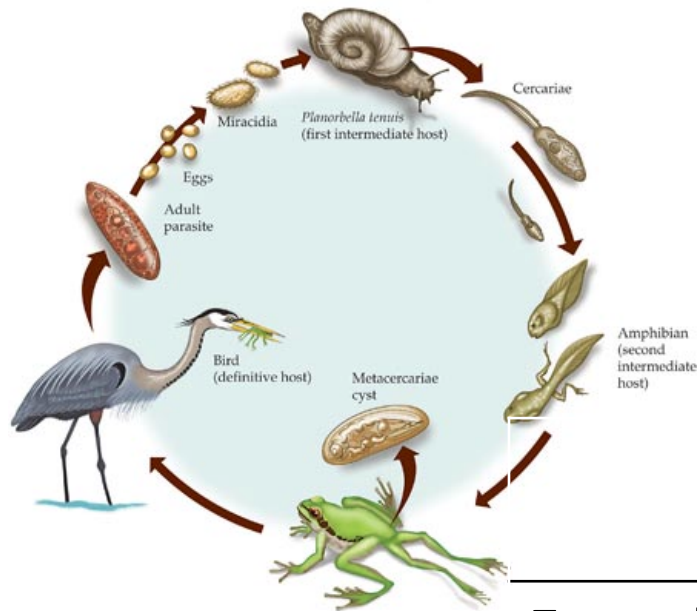◦ Are smokers more likely to drink than non-smokers?

Two flavors:

◦ χ^2 test for independence (or homogeneity)

◦ Fisher's Exact test

Contingency tables show associated counts for two+ categorical variables

	Takes daily aspirin	No daily aspirin
Heart attack	75	62
No heart attack	108	71

Example: χ^2 test for independence/association

Life cycle of *R. ondatrae*

2 variables:
Eaten (2 categories yes/no)
Infected (2 categories yes/no)

	Uninfected frog	Infected frog
Eaten by bird	1	47
Not eaten by bird	49	44

Example: χ^2 test for independence

H_0 : Infection and being eaten are independent

H_A : Infection and being eaten are not independent

	Uninfected frog	Infected frog	TOTAL
Eaten by bird	1	47	**48**
Not eaten	49	44	**93**
TOTAL	**50**	**91**	**141**

Computing the test statistic

$$\chi^2 = \sum_c \sum_r \frac{(\# \ observed_{r,c} - \# \ expected_{r,c})^2}{\# \ expected_{r,c}}$$

Under the null hypothesis, the variables are independent.
Expected calculations employ **P[A and B] = P[A] x P[B]**

	Uninfected	Infected	TOTAL
Eaten	1	47	**48**
Not eaten	49	44	**93**
TOTAL	**50**	**91**	**141**

P[eaten and uninfected] = P[eaten]   x   P[uninfected]
                                  = 48/141   x   50/141   =   0.1207
Expected count = P[eaten and uninfected] x total   = **17.02**
… = (row/total) x (column/total) x (total)

Performing the test

	Uninfected	Infected
Eaten	1 17.02	44 30.9
Not eaten	49 33.3	47 60.2

$$\chi^2 = \sum_c \sum_r \frac{\left(\# \, observed_{r,c} - \# \, expected_{r,c}\right)^2}{\# \, expected_{r,c}}$$

$$= \frac{(1-17.02)^2}{17.02} + \frac{(44-30.9)^2}{30.9} + \frac{(49-33.3)^2}{33.3} + \frac{(47-60.2)^2}{60.2} = 31.9$$

df = (#r − 1)(#c − 1) = (2 − 1)(2 − 1) = 1

```
1 - pchisq(31.9, 1)
[1] 1.623172e-08
```

We reject the null hypothesis (P << α) that infection and being eaten are independent. We have evidence that being infected with this trematode is associated with being eaten by a bird.

Performing the test in R

```
> data.table <- rbind(c(1,49), c(44,47))
> data.table
     [,1] [,2]
[1,]    1   49
[2,]   44   47

> chisq.test(data.table)
Pearson's Chi-squared test with Yates' continuity correction
data:  data.table
X-squared = 29.809, df = 1, p-value = 4.768e-08

> chisq.test(data.table, correct=FALSE)
Pearson's Chi-squared test
data:  data.table
X-squared = 31.906, df = 1, p-value = 1.618e-08
```

This is what we calculated on the last slide ("R as calculator"). Differences are from using rounded expected counts.

Yates continuity correction

$$\chi^2 = \sum_c \sum_r \frac{\left(\# \, observed_{r,c} - \# \, expected_{r,c}\right)^2}{\# \, expected_{r,c}}$$ **Without correction**

$$\chi^2 = \sum_c \sum_r \frac{\left(\left|\# \, observed_{r,c} - \# \, expected_{r,c}\right| - 0.5\right)^2}{\# \, expected_{r,c}}$$ **Yates continuity correction**

Decreases the test statistic and increases the P-value

Odds

The **odds** of success are the probability of success divided by failure

$$O = \frac{p}{1-p}$$

	Uninfected	Infected	TOTAL
Eaten	1	47	48
Not eaten	49	44	93
TOTAL	50	91	141

The **odds** of being eaten while infected

$$O = \frac{P[eaten\ and\ infected]}{1-P[eaten\ and\ infected]} = \frac{47/91}{1-47/91} = 1.07$$

$$O = \frac{P[eaten\ and\ infected]}{P[not\ eaten\ and\ infected]} = \frac{47}{44} = 1.07$$

Odds ratio, for 2x2 tables

The **odds ratio** is the odds of success in one group divided by odds of success in a second group

$$OR = \frac{p_1/(1-p_1)}{p_2/(1-p_2)}$$

ORs quantify the deviation from null in 2x2 contingency table tests.

Interpretation

◦ **OR = 1:** Odds of success is the same for either group

◦ **OR < 1:** Odds of success in group 2 are higher than group 1

◦ **OR > 1:** Odds of success in group 1 are higher than group 2

Odds ratio calculations: Are the odds higher that you are eaten while infected?

	Uninfected	Infected	TOTAL
Eaten	1	47	**48**
Not eaten	49	44	**93**
TOTAL	**50**	**91**	**141**

$$O_1 = \frac{P[eaten\ and\ infected]}{1 - P[eaten\ and\ infected]} = \frac{P[eaten\ and\ infected]}{P[not\ eaten\ and\ infected]} = \frac{47}{44} = 1.07$$

$$O_2 = \frac{P[eaten\ and\ uninfected]}{1 - P[eaten\ and\ uninfected]} = \frac{P[eaten\ and\ uninfected]}{P[not\ eaten\ and\ uninfected]} = \frac{1}{49} = 0.02$$

$$OR = \frac{1.07}{0.02} = 52.3$$

Infected frogs have 52.3 the odds of being eaten compared to uninfected frogs.

Odds ratio calculations: Are the odds higher that you are eaten while infected?

	Uninfected	Infected	TOTAL
Eaten	1	47	**48**
Not eaten	49	44	**93**
TOTAL	**50**	**91**	**141**

$$O_1 = \frac{P[eaten\ and\ infected]}{1 - P[eaten\ and\ infected]} = \frac{P[eaten\ and\ infected]}{P[not\ eaten\ and\ infected]} = \frac{47}{44} = 1.07$$

$$O_2 = \frac{P[eaten\ and\ uninfected]}{1 - P[eaten\ and\ uninfected]} = \frac{P[eaten\ and\ uninfected]}{P[not\ eaten\ and\ uninfected]} = \frac{1}{49} = 0.02$$

$$OR = \frac{1.07}{0.02} = 52.3$$

Infected frogs have 52.3 the odds of being eaten compared to uninfected frogs.

Odds ratio calculations: Are the odds higher that you are eaten while infected?

$$O_1 = \frac{P[eaten\ and\ infected]}{P[not\ eaten\ and\ infected]} = 1.07$$

$$O_2 = \frac{P[eaten\ and\ uninfected]}{P[not\ eaten\ and\ uninfected]} = 0.02$$

Eaten frogs have 52.3 the odds of being infected compared to uneaten frogs.

	Uninfected	Infected	TOTAL
Eaten	1	47	48
Not eaten	49	44	93
TOTAL	50	91	141

$$O_1 = \frac{P[infected\ and\ eaten]}{P[uninfected\ and\ eaten]} = \frac{47}{1} = 47$$

$$O_2 = \frac{P[infected\ and\ uneaten]}{P[uninfected\ and\ uneaten]} = \frac{44}{49} = 0.899$$

$$OR = \frac{47}{0.899} = 52.3$$

Infected frogs have 52.3 the odds of being eaten compared to uninfected frogs.

	Uninfected	Infected	TOTAL
Eaten	1	47	48
Not eaten	49	44	93
TOTAL	50	91	141

There are two ways to calculate OR

One will be > 1 (52.3) and one will be < 1 (1/52.3 = 0.019)

◦ We generally use the >1 option

◦ **Convince yourself that this is true.**

	Uninfected	Infected	TOTAL
Eaten	a 1	c 47	48
Not eaten	b 49	d 44	93
TOTAL	50	91	141

Fun fact: $OR = \frac{a*d}{b*c} = \frac{1*44}{49*47} = 0.019$

Often we report *log odds* = ln(OR)

```
> log(52.3)
[1] 3.956996
```

Calculating the OR standard error

$$SE[\ln(OR)] = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

	blah	blah2
blob	a	c
blob1	b	d

$$SE[\ln(OR)] = \sqrt{\frac{1}{1} + \frac{1}{49} + \frac{1}{47} + \frac{1}{44}} = 1.03$$

	Uninfected	Infected
Eaten	1	47
Not eaten	49	44

Calculating the log odds CI

$$ln(\widehat{OR}) - (Z_{0.025} * SE_{\widehat{OR}}) < ln(OR) < \widehat{ln(OR)} + (Z_{0.025} * SE_{\widehat{OR}})$$

3.96 − (1.96*1.03)< $ln(OR)$< 3.96 + (1.96*1.03) ➔   3.96 ± 2.02

Conclusions, with log odds

We reject the null hypothesis (P << α) that infection and being eaten are independent. We have evidence that being infected with this trematode is associated with being eaten by a bird.

Furthermore, frogs that are eaten are more likely to be infected compared to uneaten frogs, with a log odds ratio of 3.96 and log odds CI of 1.94 – 5.98 .

χ^2 test for homogeneity

Independence: measure **two properties** from **one set of subjects**

◦ We measured **eaten** and **infection** for **frogs**

Homogeneity: measure **one property** on **two sets of subjects from different populations**

◦ Measure **effect of medicine** in **sample of cancer individuals** and **sample of healthy individuals**

Example: test of homogeneity

	Drug	Placebo
Cancer	75	62
Healthy	108	71

H_0 : The probability that symptoms improve is the same for both cancer and healthy groups.

H_A: The probability that symptoms improve differs between cancer and healthy groups.

In practical terms, this uses the exact same procedure as a test for independence.

Fisher's Exact test

More exact than χ^2 and used for low-count tables

Compute the **exact** probability of observing table with counts:

$$P(a, b, c, d) = \frac{(a + b)!\,(c + d)!\,(a + c)!\,(b + d)!}{n!\,a!\,b!\,c!\,d!}$$

	blah	blah2
blob	a	c
blob1	b	d

Fisher's test computes this value for *all possible tables* with the same row/column totals (margins)

Computes P-value by summing probabilities for tables with as extreme or more count distributions

Fisher's exact test

	Uninfected	Infected	TOTAL
Eaten	1	47	48
Not eaten	49	44	93
TOTAL	50	91	141

```
> chisq.test(data.table, correct=FALSE)
Pearson's Chi-squared test

data:  data.table
X-squared = 31.906, df = 1, p-value = 1.618e-08
```
Approximate P-value

```
> fisher.test(data.table)
Fisher's Exact Test for Count Data

data:  data.table
p-value = 8.37e-10
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.0005344122 0.1417331275
sample estimates:
odds ratio
0.02222648
```
Exact P-value

Our OR = 52.3, or 0.019. Slight differences are expected because fisher.test() uses ML

Relative risk: It's not the OR

Commonly measured in epidemiological studies

Relative risk is the probability of an event (ie disease) in an exposed group, relative to unexposed group
- ◦ RR = P(event when exposed) / P(event when not exposed)

Relative risk example

RR = P(event when exposed) / P(event when not exposed)

RR of cancer due to smoking exposure:
 = P(cancer | smoker)/P(cancer | not smoker)
= [525/(525 + 450)] / [32/(32+621)]

	Lung cancer	No lung cancer
Smoker	525	450
Non-smoker	32	621

= **10.99**

→ **Smokers have a 10.99 times higher risk than do non-smokers to develop lung cancer.**

Live exercise: Calculate the odds ratio for a smoker developing cancer relative to a non-smoker.

The Odds Ratio

	Lung cancer	No lung cancer
Smoker	525	450
Non-smoker	32	621

$$O_1 = \frac{P[smoker\ and\ cancer]}{P[non-smoker\ and\ cancer]} = \frac{525}{32}$$

$$O_2 = \frac{P[smoker\ and\ no\ cancer]}{P[non-smoker\ no\ cancer]} = \frac{450}{621}$$

$$OR = \frac{525/32}{450/621} = \mathbf{22.64}$$

→ Smokers have 22.64 times the odds of getting lung cancer than non-smokers.

What's the practical difference?

Odds ratios measure the extent of association between variables.

- ◦ It is the ratio of two **odds** (ratio of prob event : prob non-event)

Relative risk is the more intuitive quantity that we "understand"

- ◦ It is the ratio of two **probabilities** (prob event)

Recap on estimation

Normally-distributed variable

- $\hat{\mu} = \bar{x}$

- $\hat{\sigma}^2 = s^2$

- Known σ

  - $SE_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$

  - 95% CI = $\bar{x} \pm Z_{0.025}SE$

- Unknown σ

  - $SE_{\bar{x}} = \dfrac{s}{\sqrt{n}}$

  - 95% CI = $\bar{x} \pm t_{0.025}SE$

Binomially-distributed variable

- $\hat{p} = \dfrac{k}{n}$

- $SE_{\hat{p}} = \sqrt{\hat{p}(1-\hat{p})/n}$

- 95% CI = $\hat{p} \pm Z_{0.025}SE_{\hat{p}}$

Log-Odds ratio

- $\log \widehat{OR} = \ln\left[\dfrac{\hat{p}_1/(1-\hat{p}_1)}{\hat{p}_2/(1-\hat{p}_2)}\right]$

- $SE_{\log \widehat{OR}} = \sqrt{\dfrac{1}{a} + \dfrac{1}{b} + \dfrac{1}{c} + \dfrac{1}{d}}$

- 95% CI = $\log \widehat{OR} \pm Z_{0.025}SE_{\hat{p}}$

Choose your own adventure, so far