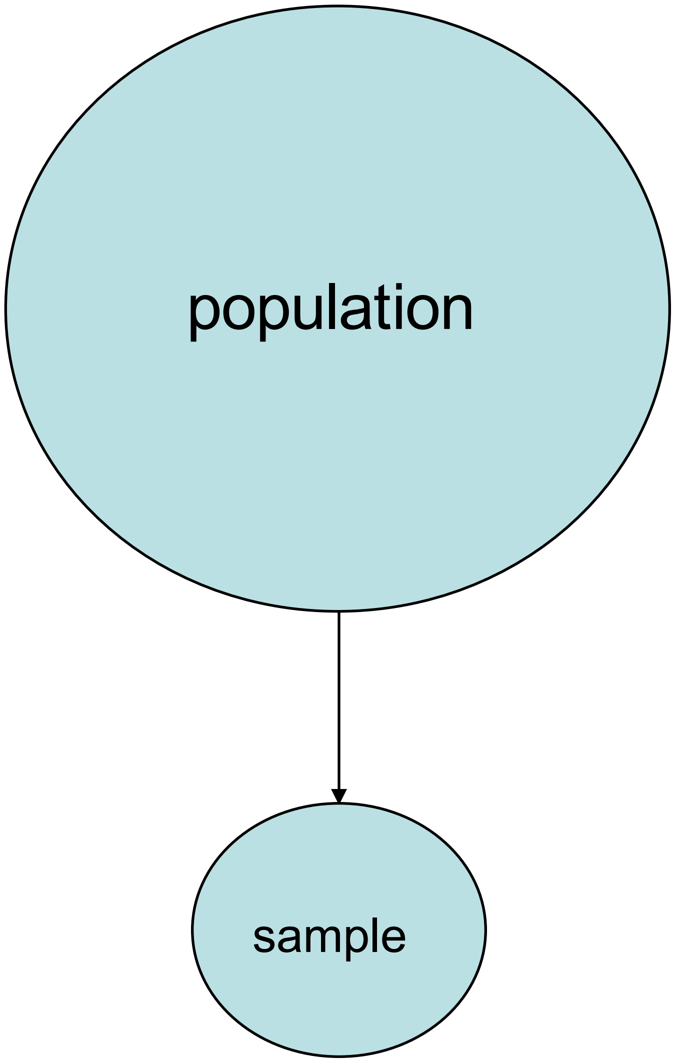
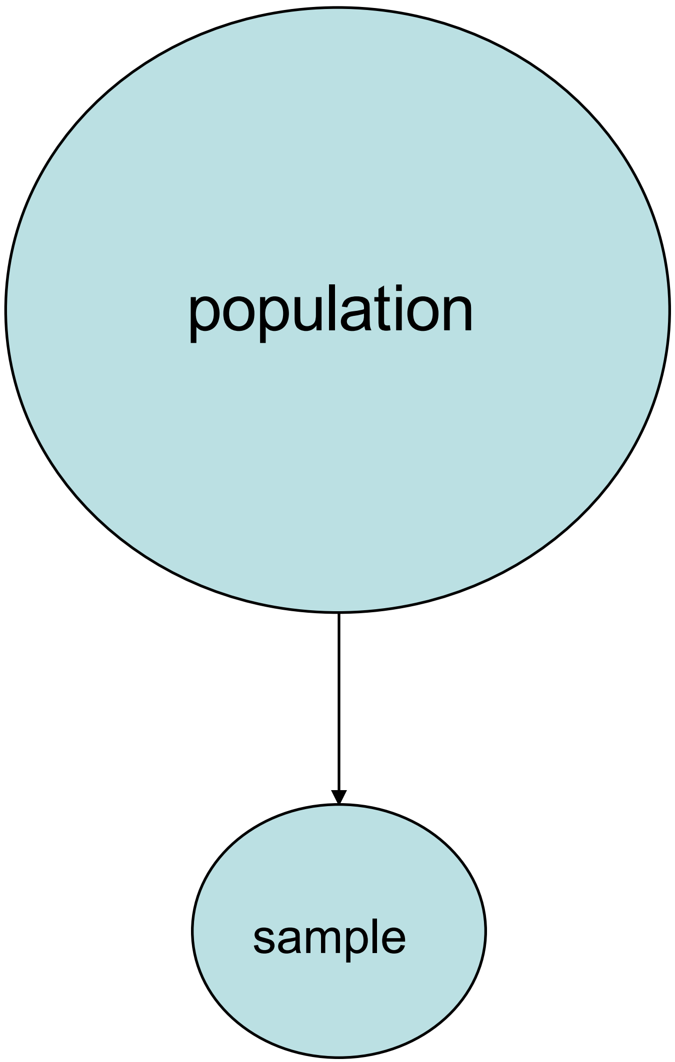
Hypothesis testing



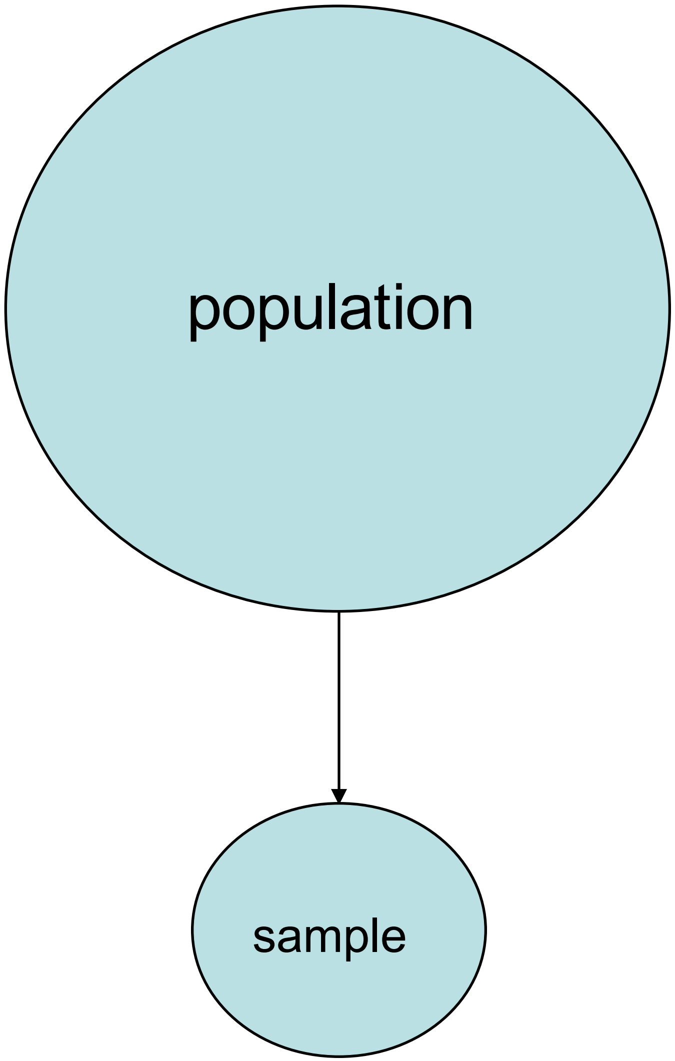
μ Mean
σ Standard deviation

$\bar{Y}$ Mean
s Standard deviation



Last topic:

probability



Last topic:

probability

Why did we take a sample?
What are we interested in?
What are our expectation

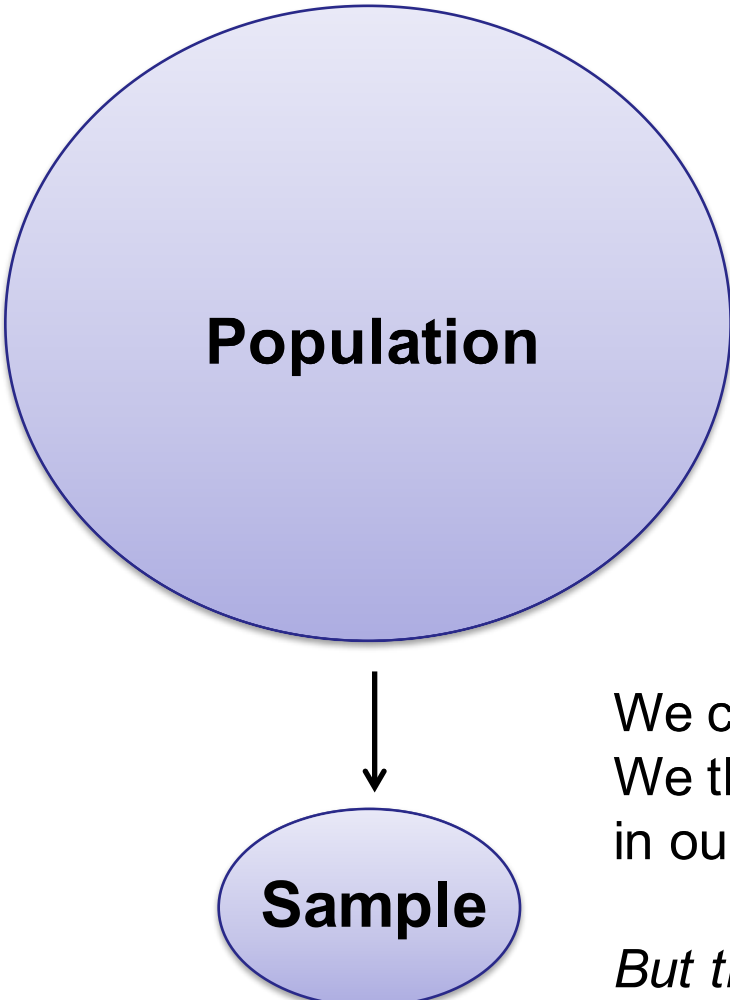
…..

What is our hypothesis?

Hypothesis testing asks how unusual it is to get data that differ from the null hypothesis.

If the data would be quite unlikely under H_0, we reject H_0.

Hypothesis testing: a quick introduction



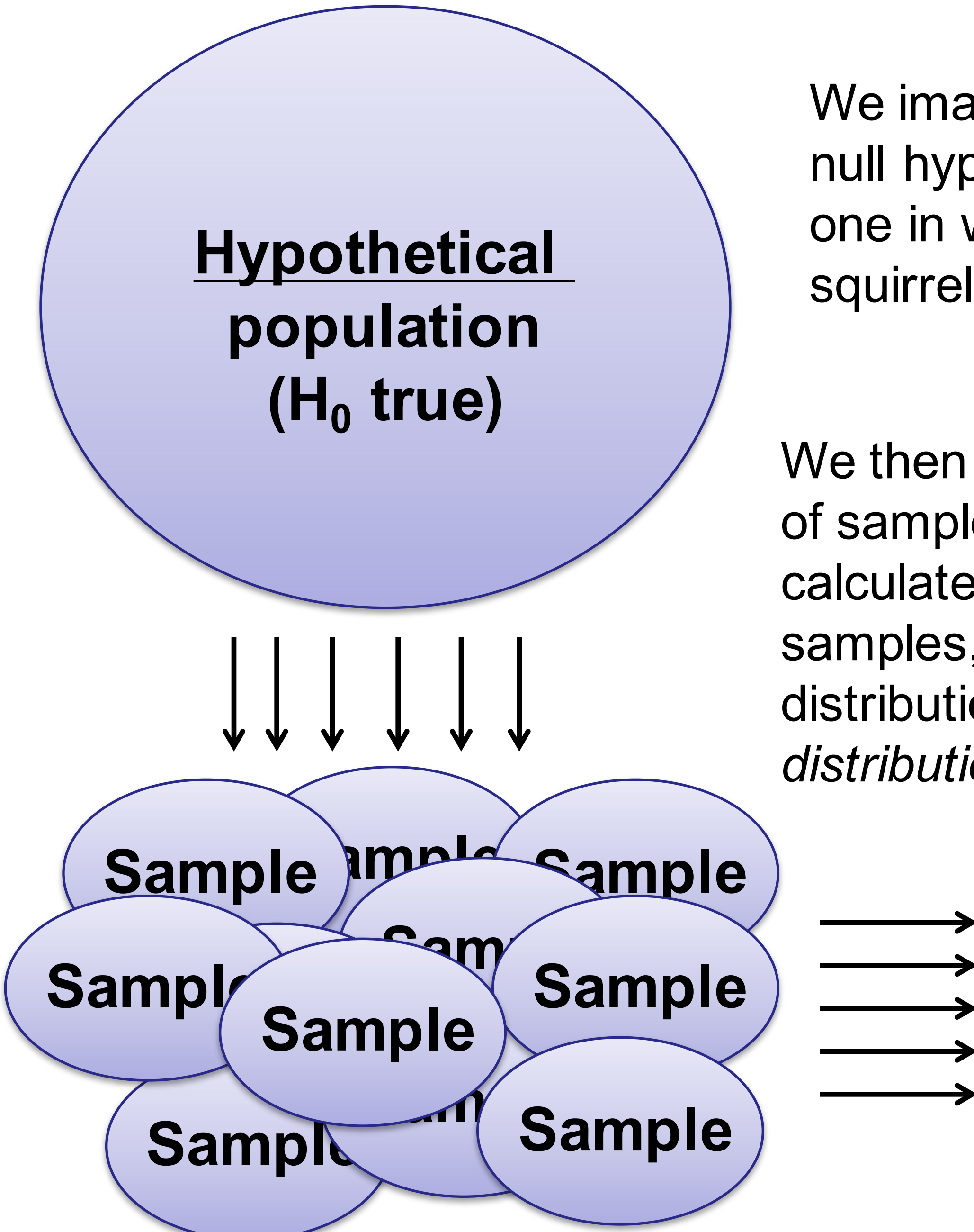
We want to know something about this population. For example, are male and female squirrels the same length, on average, or do they differ?

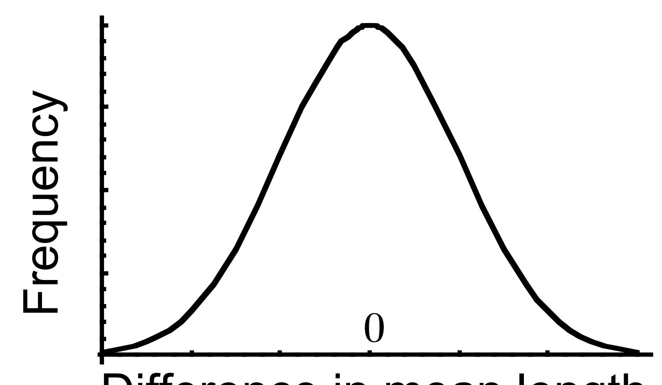
We can't study everyone, so we take a sample. We then determine whether males and females in our sample have different sizes, on average.

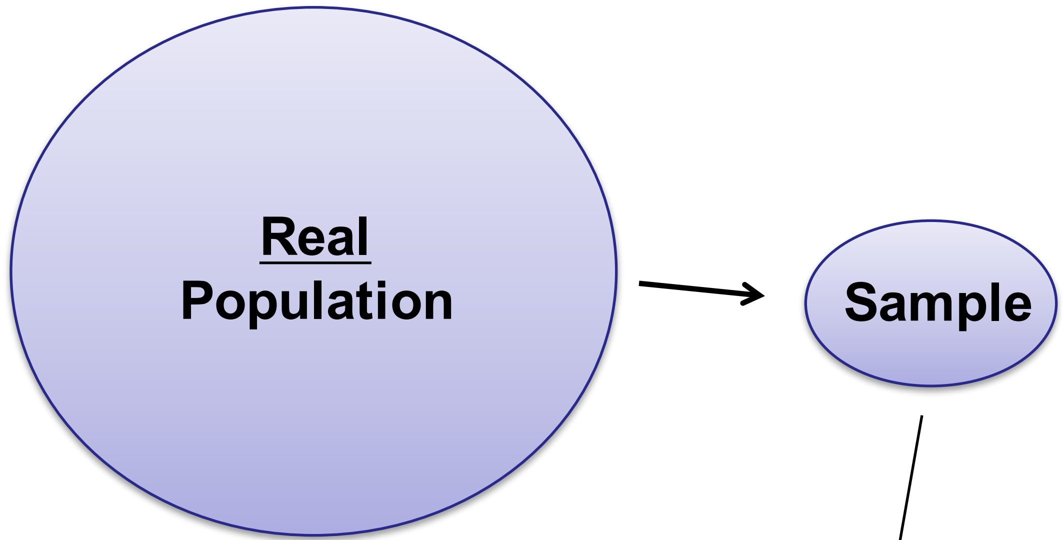
But there is a problem: our sample differs from the population because of sampling error.

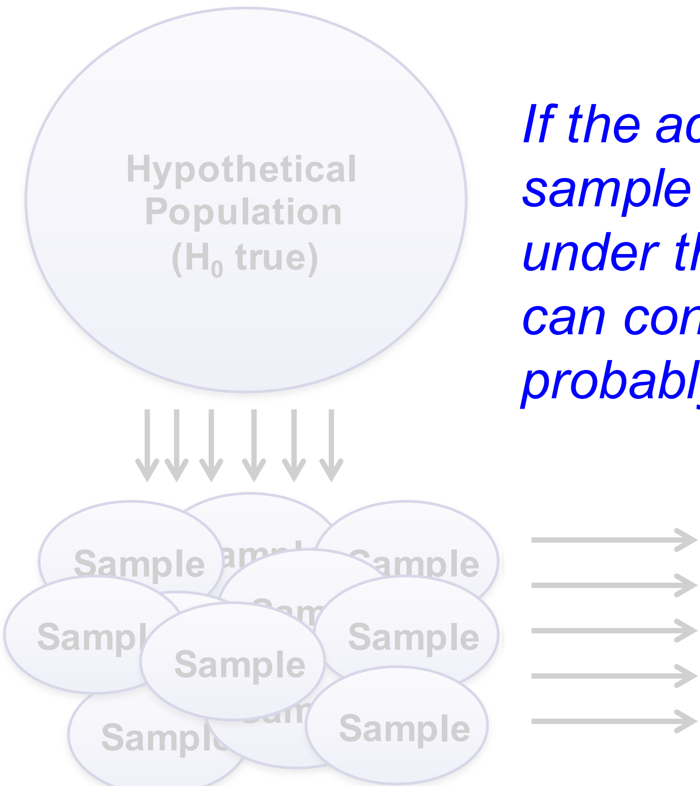
We imagine a population in which the null hypothesis is true (in this case, one in which male and female squirrels are the same length).

We then imagine taking an infinite number of samples from this population. We calculate an estimate from each of these samples, and generate a sampling distribution of that estimate (the *null distribution*).

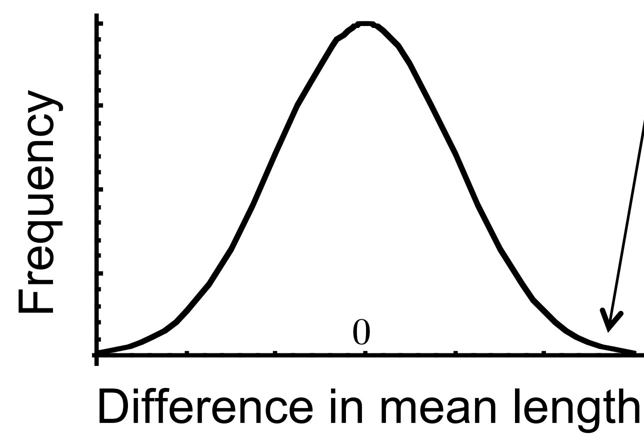








If the actual estimate from the real sample would be highly unusual under the null hypothesis, then we can conclude that the null hypothesis probably isn't true.



Hypotheses are about populations, but are tested with data from samples

Hypothesis testing usually assumes that sampling is random.

Null hypothesis: a specific statement about a population parameter made for the purposes of argument.

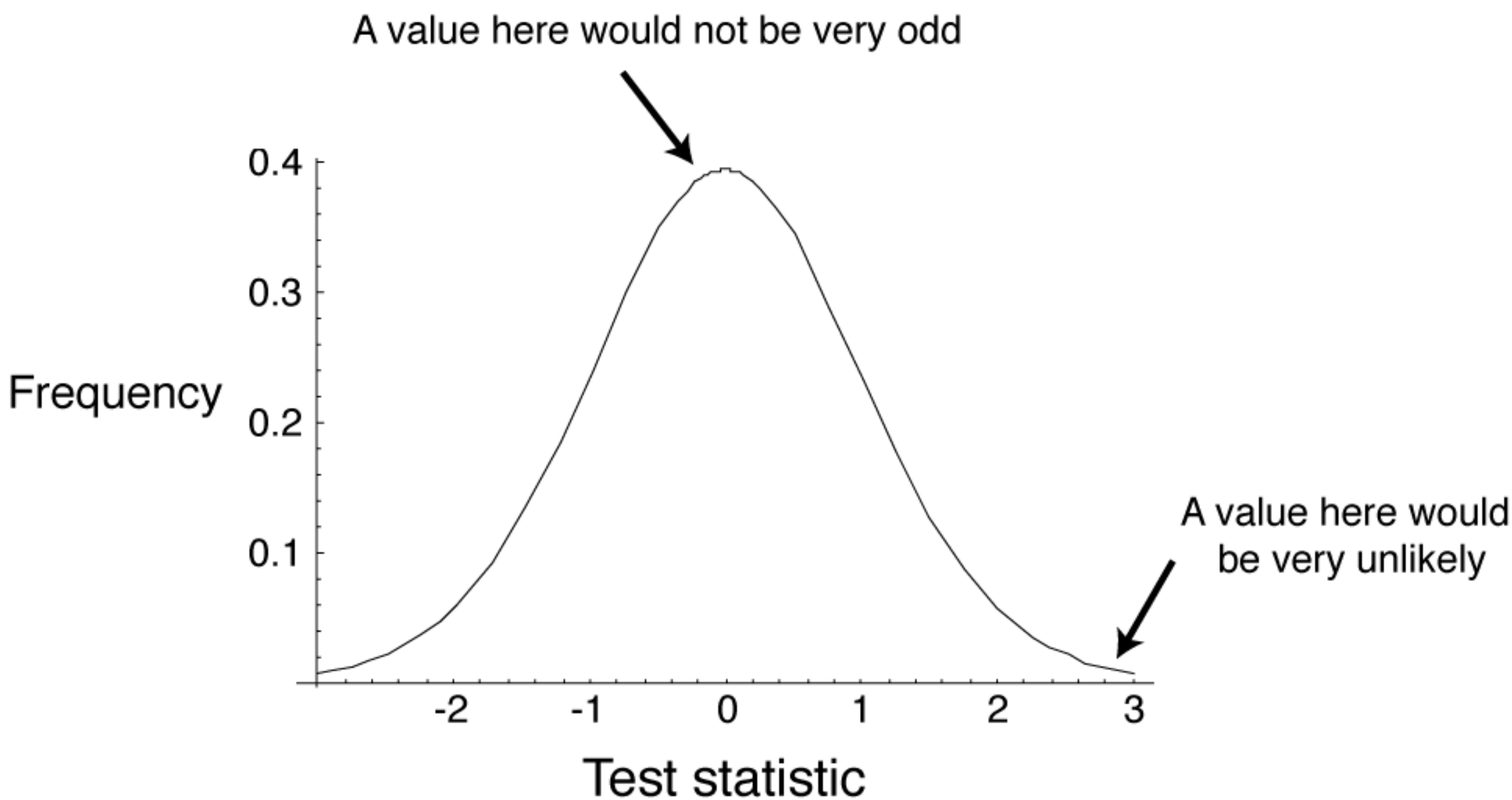
Null hypothesis: a specific statement about a population parameter made for the purposes of argument.

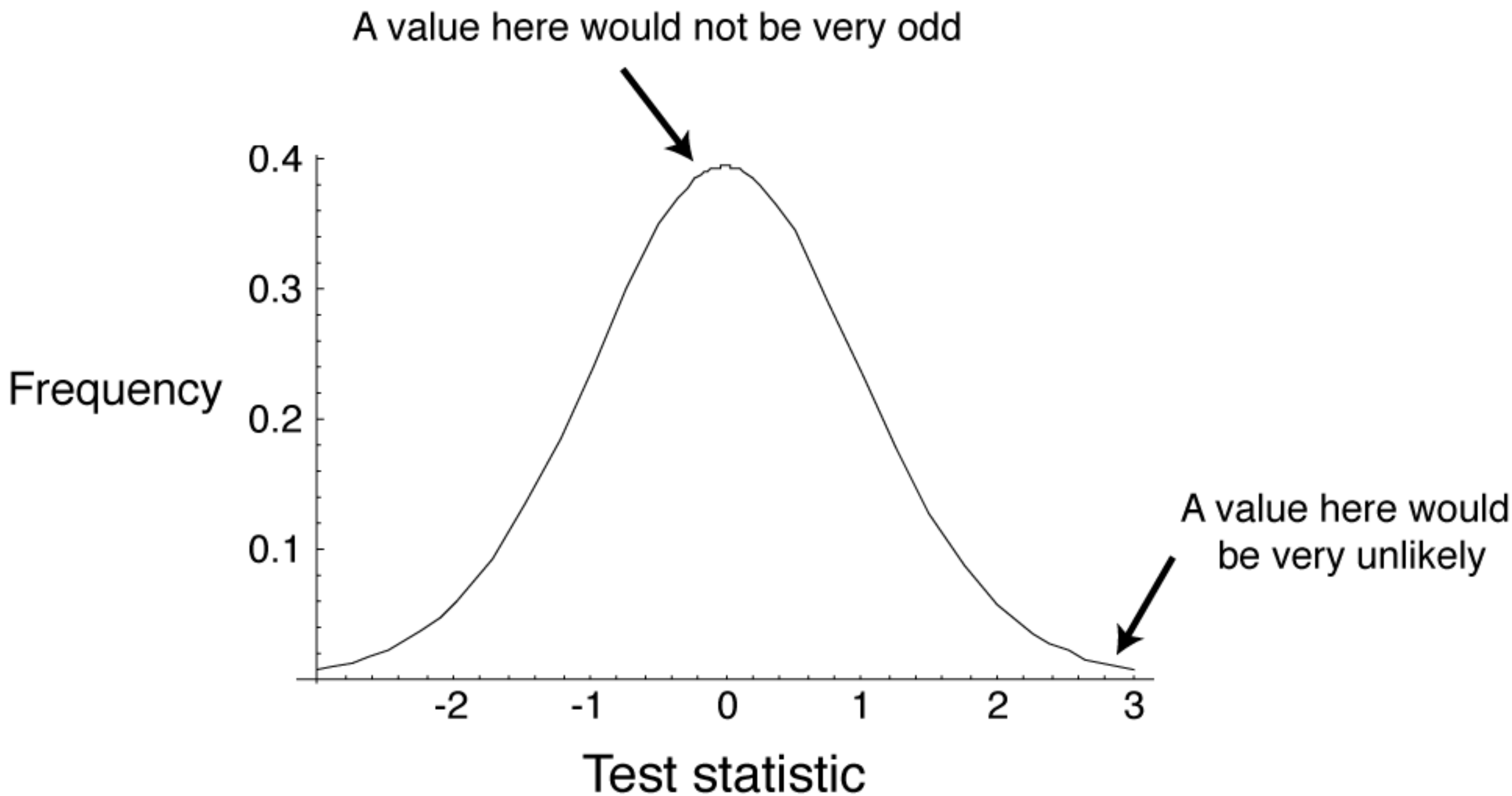
Alternate hypothesis: represents all other possible parameter values except that stated in the null hypothesis.

The *null hypothesis* is usually the simplest statement, whereas the *alternative hypothesis* is usually the statement of greatest interest.

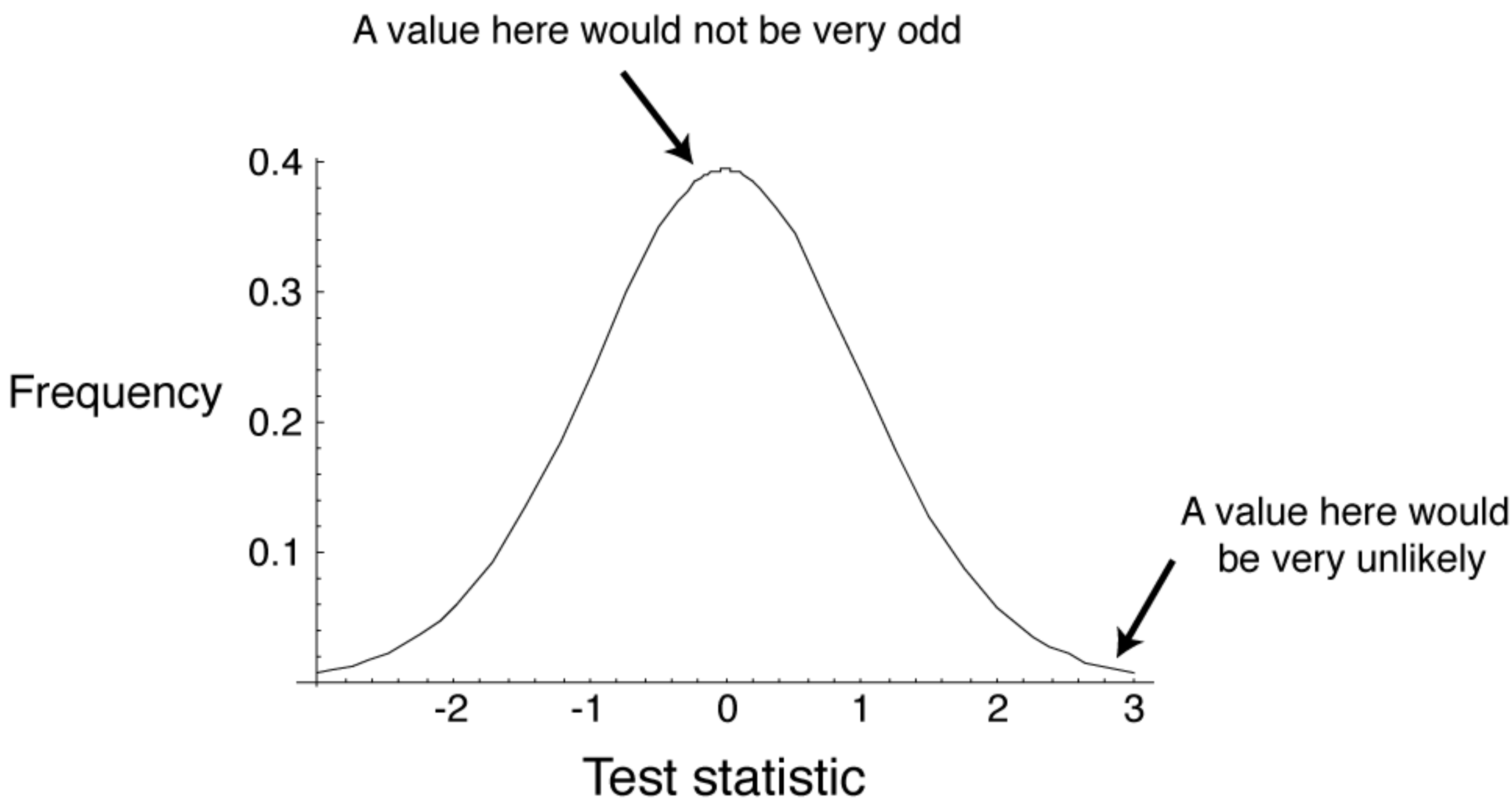
A good null hypothesis is one that would generate interest if proven wrong.

A null hypothesis is specific;
an alternate hypothesis is not.


A value here would not be very odd
Frequency
0.4
0.3
0.2
0.1
-2
-1
0
1
2
3
Test statistic
A value here would be very unlikely

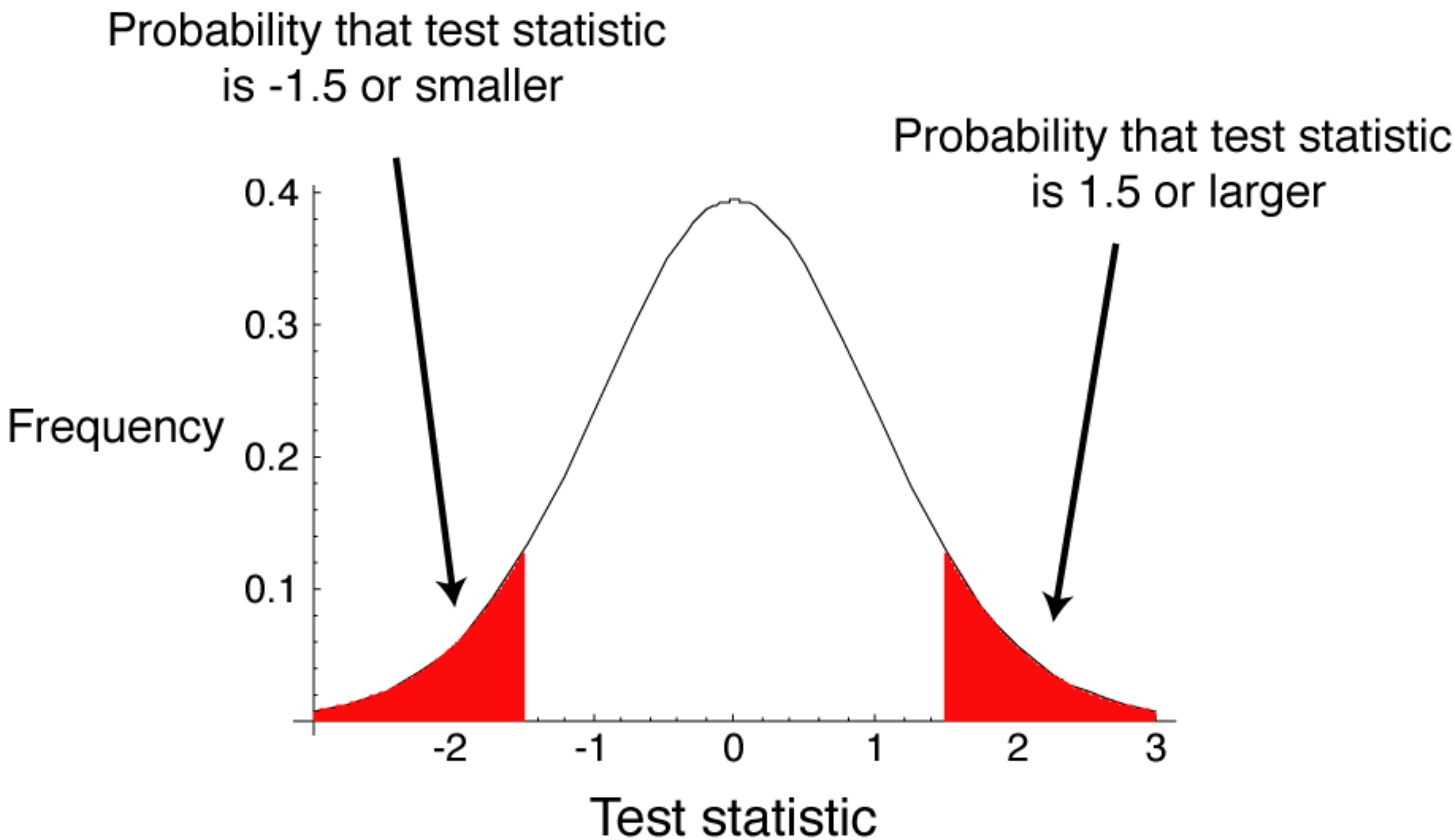


A test statistic summarizes the match between the data and the null hypothesis



Null distribution: the sampling distribution of outcomes for a test statistic under the assumption that the null hypothesis is true

P-value


Probability that test statistic
is -1.5 or smaller
Probability that test statistic
is 1.5 or larger
0.4
0.3
0.2
0.1
Frequency
-2
-1
0
1
2
3
Test statistic

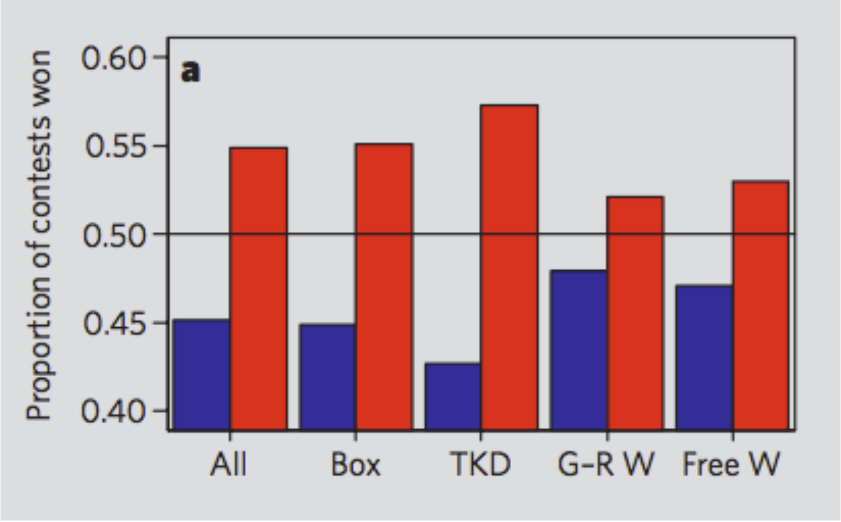
A *P*-value is the probability of getting the data, or something as or more unusual, if the null hypothesis were true.

Hypothesis testing: an example

Does a red shirt help win wrestling?

The experiment and the results

- Animals use red as a sign of aggression
- Does red influence the outcome of wrestling, taekwondo, and boxing?



- 16 of 20 rounds > red-shirted winners
- Shirt color was randomly assigned

Hill, RA, and RA Burton 2005. Red enhances human performance in contests Nature 435:293.

Stating the hypotheses

H_0: Red- and blue-shirted athletes are <u>equally likely</u> to win (*proportion* = 0.5).

H_A: Red- and blue-shirted athletes are <u>not equally likely</u> to win (*proportion* ≠ 0.5).

Estimating the value

- 16 of 20 is a proportion, *proportion* = 0.8

- This is a discrepancy of 0.3 from the proportion proposed by the null hypothesis, *proportion* = 0.5

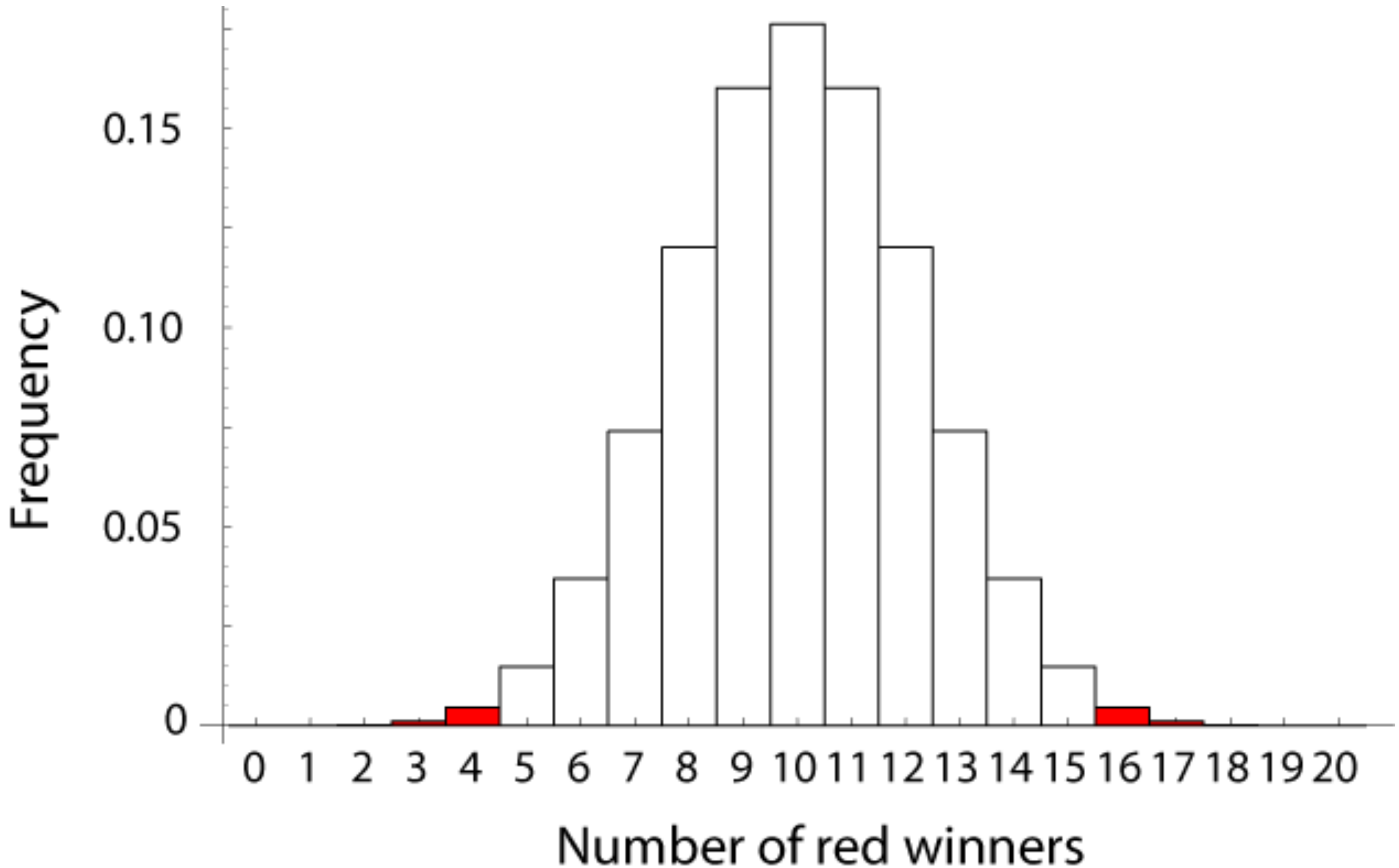
Is this discrepancy by chance alone?

We can calculate the probability of such an extreme result, assuming the null hypothesis is true

Null distribution

We use the *null distribution* to estimate the probability of getting the data, or something as or more unusual, if the null hypothesis were true (i.e. to get the P-value for our test statistic)

The null distribution of the *sample proportion*



Calculating the P-value from the null distribution

The P-value is calculated as

$P = 2 \times [\Pr(16) + \Pr(17) + \Pr(18) + \Pr(19) + \Pr(20)] = 0.012.$

Statistical significance

The *significance level*, α, is a probability used as a criterion for rejecting the null hypothesis.

If the *P*-value for a test is less than or equal to α, then the null hypothesis is rejected.

α is often 0.05

Significance for the red shirt example

- $P = 0.012$

- $P < \alpha$, so we can reject the null hypothesis

- Athletes in red shirts were more likely to win.

Hypothesis testing: another example

Do dogs resemble their owners?

Common wisdom holds that dogs resemble their owners. Is this true?

- 41 dog owners approached in parks; photos taken of dog and owner separately
- Photos of owner and dog, along with a photo of another dog, shown to students to match

Roy, M.M., & Christenfeld, N.J.S. (2004). Do dogs resemble their owners? *Psychological Science*, **15**, 361–363

Hypotheses

H_0: The proportion of correct matches is *proportion* = 0.5.

H_A: The proportion of correct matches is different from *proportion* = 0.5.

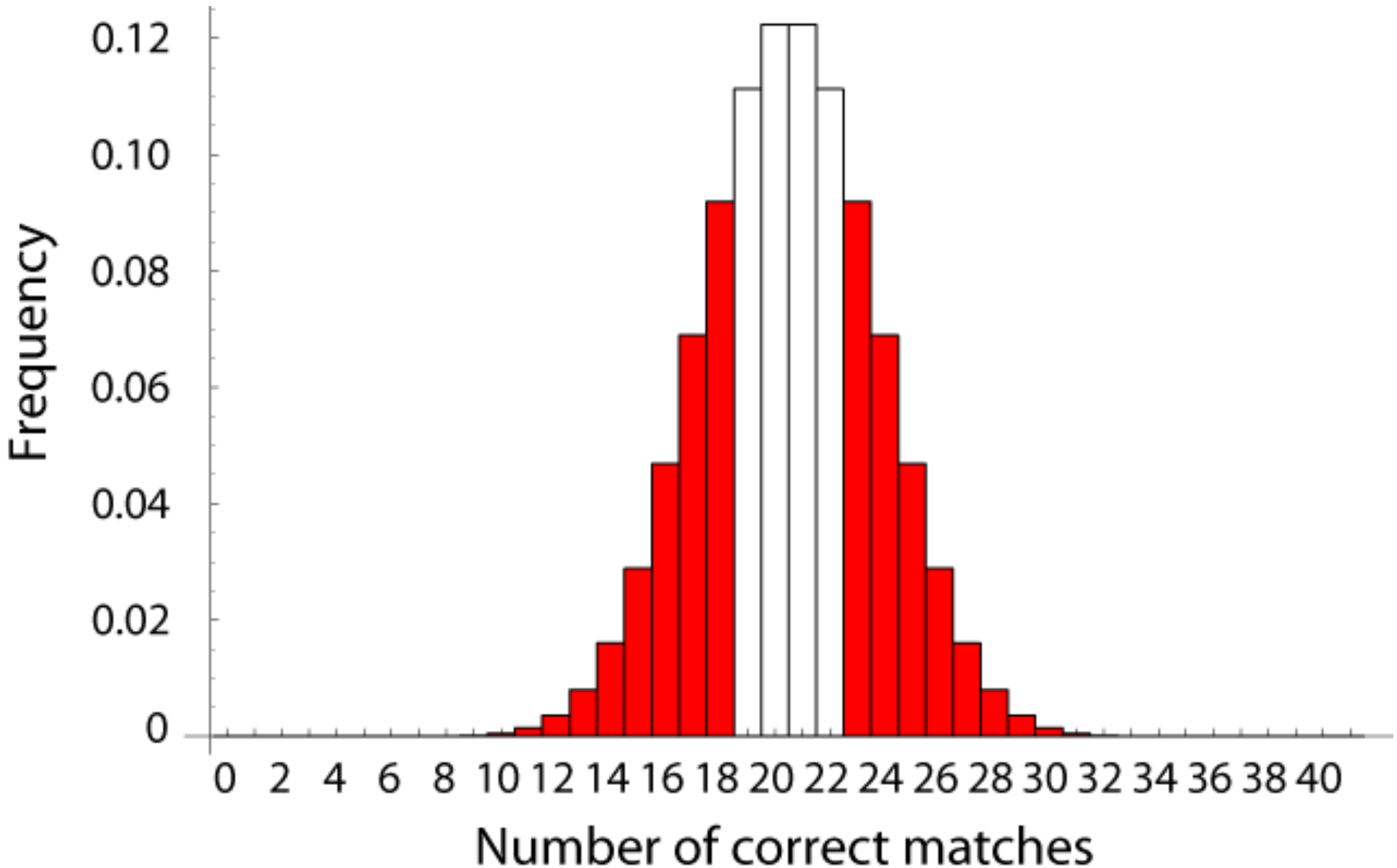
Data

Of 41 matches, 23 were correct and 18 were incorrect.

Estimating the proportion

$$sample\ proportion = \frac{23}{41} = 0.56$$

Null distribution for dog/owner resemblance



The *P*-value:

$$P = 0.53$$

We do not reject the null hypothesis that dogs do not resemble their owners.

How to construct null distributions and find *P*-values

- Simulation
- Infer from theory
- Re-sample from data (an advanced topic)

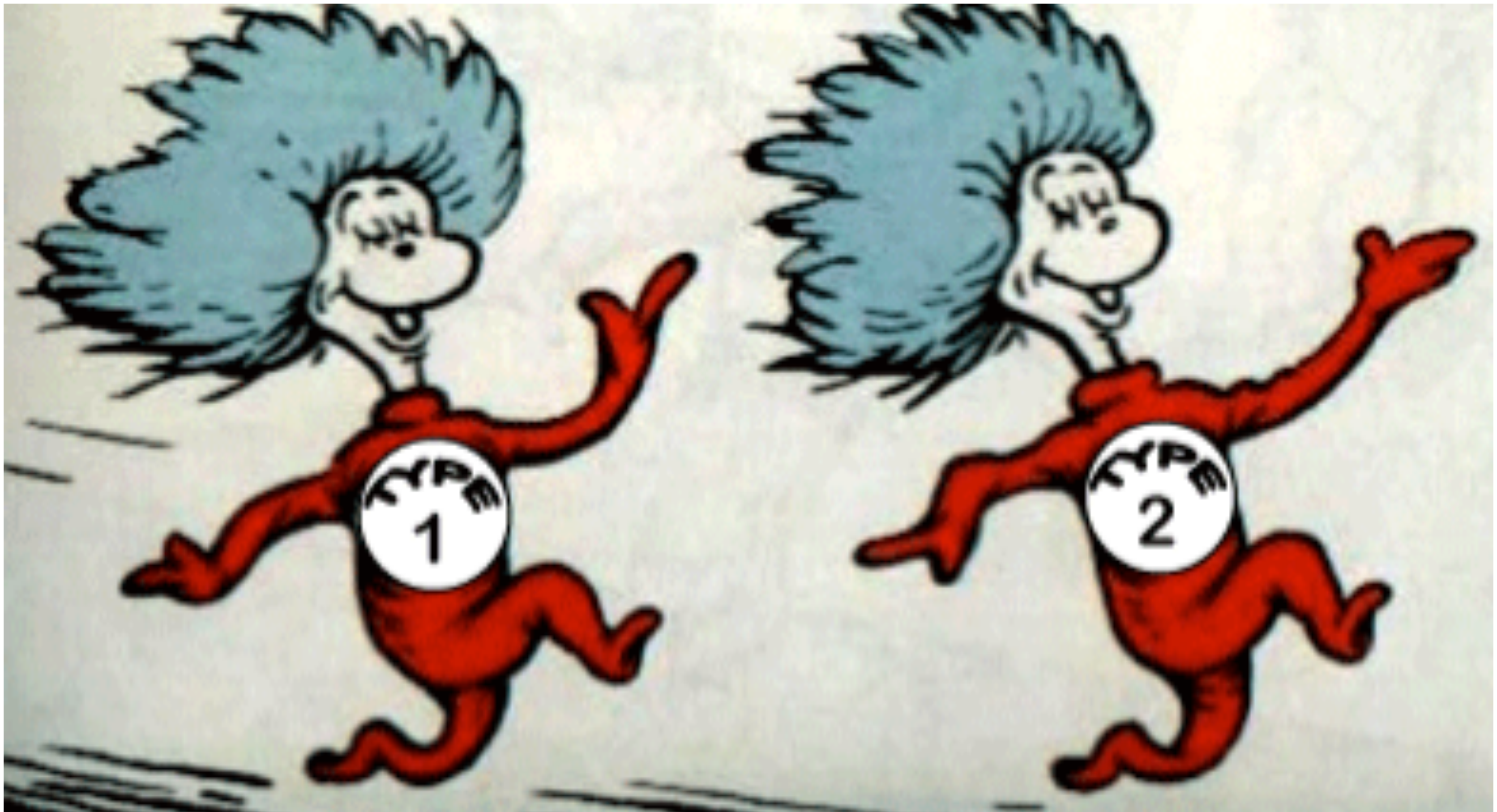
Significance level summary

- The acceptable probability of rejecting a true null hypothesis
- Called α
- For many purposes, $\alpha = 0.05$ is acceptable

“Statistically significant”

- $P < \alpha$
- We can “reject the null hypothesis”
- We **never** “accept the null hypothesis”

Two types of error in hypothesis testing



Type I error

- Rejecting a *true* null hypothesis

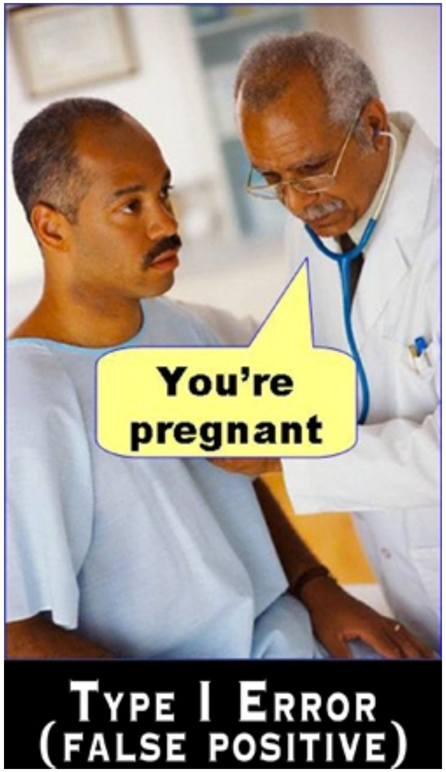
- Probability of Type I error is α (the significance level)

Type II error

- Not rejecting a *false* null hypothesis
- The probability of a Type II error is β.
- The smaller β, the more *power* a test has.

Decision	Reality H_0 true	H_0 false
Reject H_0	**type I error**	no error
Do not reject H_0	no error	**type II error**

A pregnancy example



http://illuminutti.com/tag/false-negative/

Type I and type II errors are linked

Power

- The ability of a test to reject a false null hypothesis

- Power $= 1 - \beta$

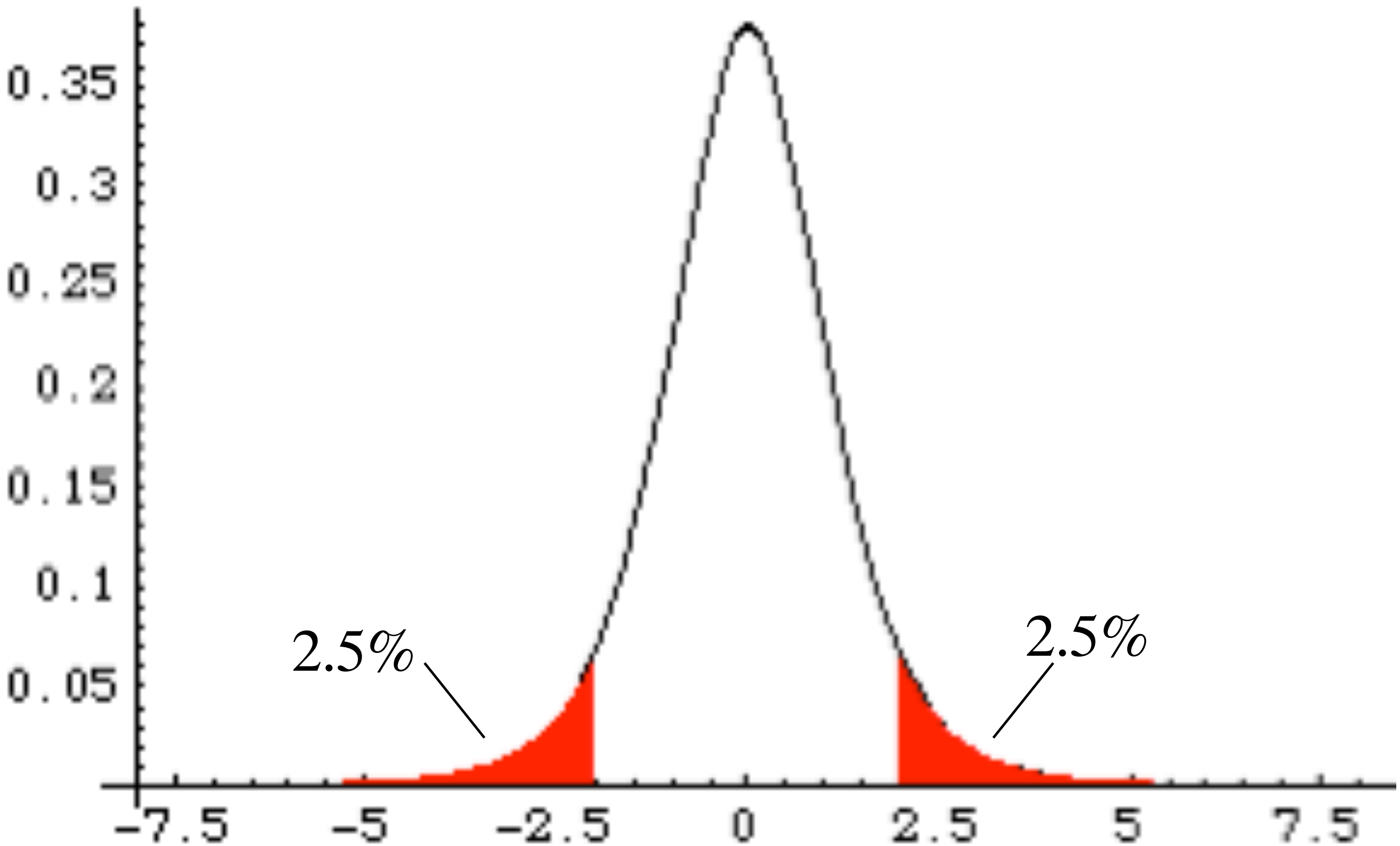
Larger samples give more information

- A larger sample will tend to give and estimate with a smaller confidence interval

- A larger sample will give more power to reject a false null hypothesis

One- and two-tailed tests

- Most tests are *two-tailed tests.*
- This means that a deviation in either direction would reject the null hypothesis.
- Normally α is divided into $\alpha/2$ on one side and $\alpha/2$ on the other.



Test statistic

One-tailed tests

- Only used when the other tail is nonsensical
- For example, comparing grades on a multiple choice test to that expected by random guessing

Test Statistic

- A number calculated to represent the match between a set of data and the null hypothesis

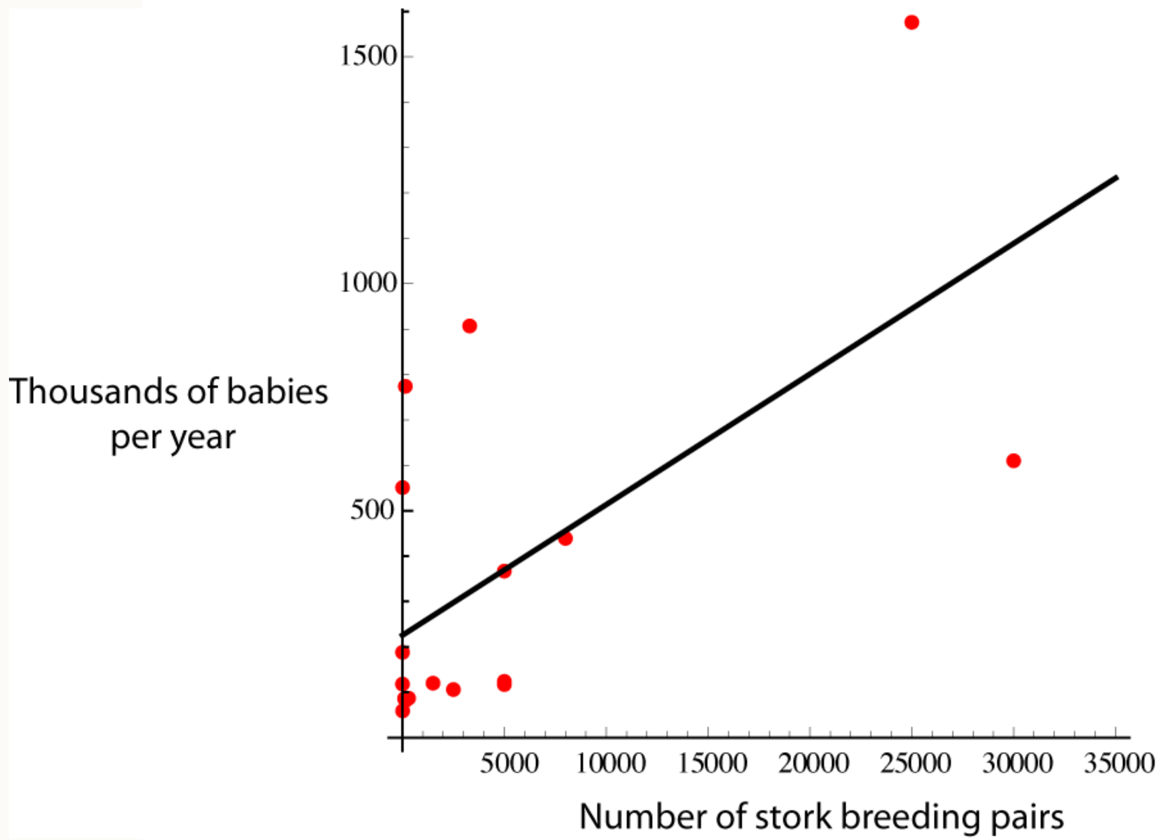
- Can be compared to a general distribution to infer probability

Critical value

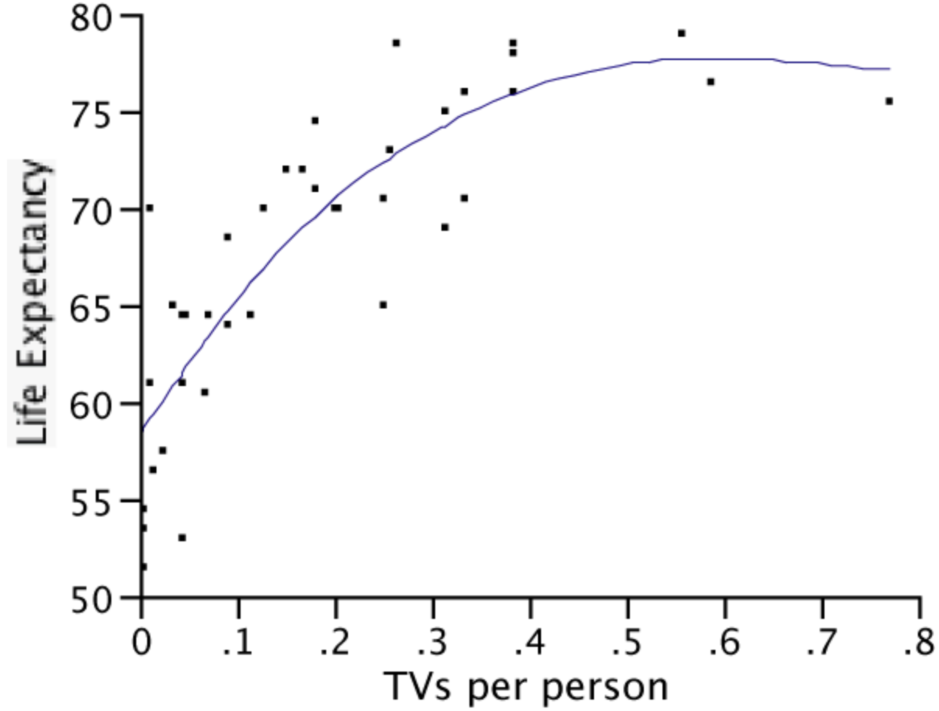
- The value of a test statistic beyond which the null hypothesis can be rejected

Correlation does not automatically imply causation

Correlation does not automatically imply causation



Life expectancy by country:



Confounding variable

An unmeasured variable that may cause both X and Y

Observations vs. Experiments

Statistical significance ≠ Biological importance

<table>
<tr><th></th><th>Important</th><th>Unimportant</th></tr>
<tr><td>Significant</td><td>Polio vaccine reduces incidence of polio</td><td>Things you don’t care about, or already well known things

BRIEFS
Study Shows Frequent Sex Enhances Pregnancy Chances
By The Associated Press
BOSTON — A study that researchers say gives the best estimate ever of nature's window of female fertility comes to a not-so-startling conclusion: The best way to make babies is to have lots of sex.
Every other day is better than once a week. Daily is best.
The discovery, however, is not
ples don't want to use other forms of birth control.
Researchers say there are six days in every menstrual month when a woman can get pregnant.
Until now, estimates of women's fertility ranged from two days in a menstrual cycle to 10 or more. The new study found that conception is possible if a woman has intercourse on the five days before ovulation as well as on the day</td></tr>
<tr><td>Insignificant</td><td>Small study shows a possible effect, leading to larger study which finds significance
or
Large study showing no effect of drug that was thought to be beneficial</td><td>Studies with small sample size and high P-value
or
Things you don’t care about</td></tr>
</table>

Proportions

A *proportion* is the fraction of individuals having a particular attribute.

Can range from 0 to 1

Example:

2092 adult passengers on the Titanic;
654 survived

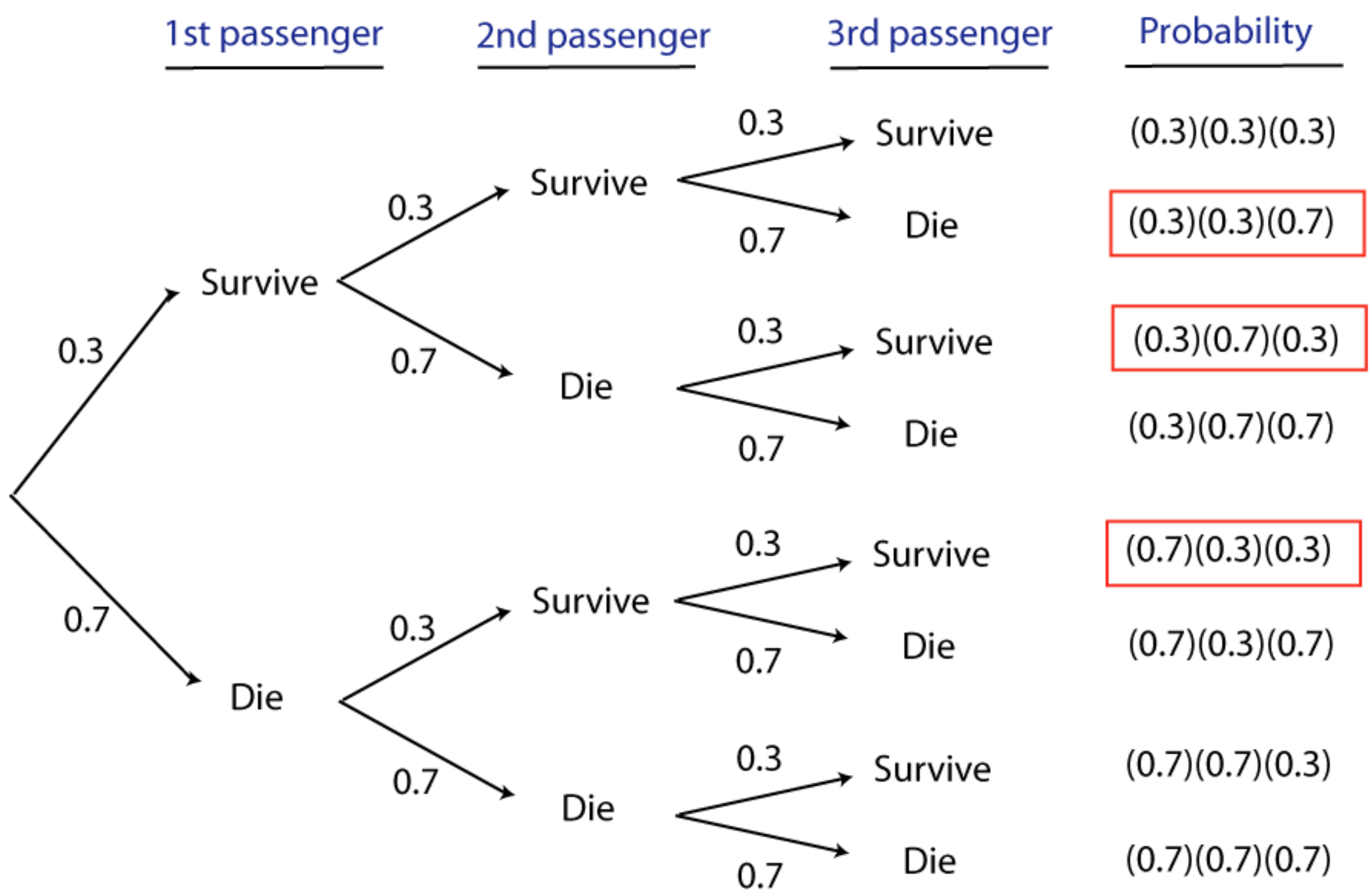
www.history.com

Example:
2092 adult passengers on the Titanic;

654 survived

Proportion of survivors = $654/2092 \cong 0.3$

Probability that two out of three randomly chosen passengers survived the Titanic



Binomial distribution

The *binomial distribution* describes the probability of a given number of "successes" from a fixed number of independent trials, when the probability of success is the same in each trial.

Binomial distribution

Used when individuals can be divided into two (*bi*-) mutually exclusive named groups (-*nomial*).

For example:

- Left handed or right handed
- Alive or dead
- University student or not university student

We call the two groups “successes” vs. “failures”

n trials; *p* probability of success

$$\Pr[X] = \binom{n}{X} p^{X} (1 - p)^{n-X}$$

n trials; *p* probability of success

$$\Pr[X] = \binom{n}{X} \underbrace{p^{X}(1-p)^{n-X}}$$

Probability of *X* successes in *n* trials

n trials; *p* probability of success

$$\Pr[X] = \binom{n}{X} p^X (1-p)^{n-X}$$

Probability of *X* successes in *n* trials

"*n* choose *X*"
The # of unique ordered sequences of successes and failures that yield *X* successes in *n* trials

n trials; *p* probability of success

$$\Pr[X] = \binom{n}{X}\underbrace{p^{X}(1-p)^{n-X}}$$

Probability of *X* successes in *n* trials

"*n* choose *X*"
The # of unique ordered sequences of successes and failures that yield *X* successes in *n* trials

Probability of a given ordered sequence of successes and failures that yield *X* successes in *n* trials

n trials; p probability of success

$$\Pr[X] = \binom{n}{X} p^X (1-p)^{n-X}$$

$$\binom{n}{X} = \frac{n!}{X!(n-X)!}$$

$$n! = n \times n\text{-}1 \times n\text{-}2 \times \ldots 3 \times 2 \times 1$$

$$6! = 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$$

$$0! = 1$$

$$1! = 1$$

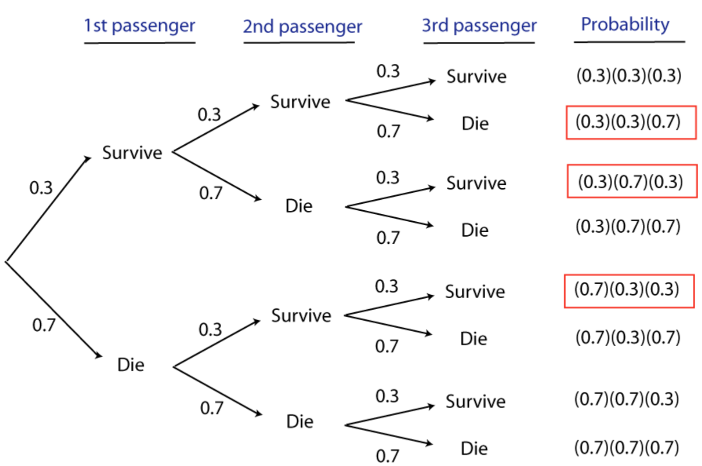
Binomial distribution

Assumptions:

- The number of trials (n) is fixed
- Separate trials are independent
- The probability of successes (p) is the same in every trial

Probability that two out of three randomly chosen passengers survived the Titanic

$$\Pr[2] = \binom{3}{2}(0.3)^2(1-0.3)^{3-2}$$



$$= \frac{3!}{2!\times 1!}(0.3)^2(0.7)^1$$

$$= 3(0.3)^2(0.7) = 0.189$$

Probability that two out of three randomly chosen passengers survived the Titanic

$$\Pr[2] = \binom{3}{2}(0.3)^2(1-0.3)^{3-2}$$

Number of ways to get 2 survivors out of 3 passengers

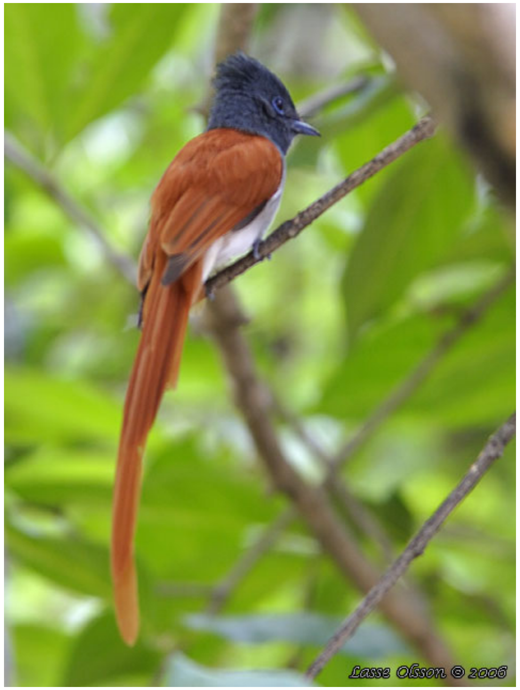
Probability of 2 survivors

Probability of 1 death

Example: Paradise flycatchers

A population of paradise flycatchers has 80% brown males and 20% white. Your field assistant captures 5 male flycatchers at random.

What is the chance that 3 of those are brown and 2 are white?



Call brown "success"

$p = 0.8$ $\qquad n = 5$ $\qquad X = 3$

$$\Pr[3] = \binom{5}{3} 0.8^3 (1 - 0.8)^{5-3} = \frac{120}{6 \times 2} 0.8^3 (0.2)^2 = 0.205$$

In-class Exercise:

What is the probability that 3 or more are brown?

In-class Exercise:

What is the probability that 3 or more are brown?

$$\Pr\left[3 \; or \; more \; are \; brown\right] = \Pr[3] + \Pr\left[4\right] + \Pr\left[5\right]$$

In-class Exercise:

What is the probability that 3 or more are brown?

$$\Pr\left[3 \textit{ or more are brown}\right] = \Pr[3] + \Pr\left[4\right] + \Pr\left[5\right]$$

$$\Pr[3] = \binom{5}{3} 0.8^3 (1-0.8)^{5-3} = 0.205$$

$$\Pr[4] = \binom{5}{4} 0.8^4 (1-0.8)^{5-4} = 0.410$$

$$\Pr[5] = \binom{5}{5} 0.8^5 = 0.328$$

In-class Exercise:

What is the probability that 3 or more are brown?

$$\Pr\left[3 \ \textit{or more are brown}\right] = \Pr[3] + \Pr\left[4\right] + \Pr\left[5\right]$$
$$= 0.205 + 0.410 + 0.328$$
$$= 0.943$$

Hypothesis testing on proportions

The binomial test

Binomial test

The *binomial test* uses data to test whether a population proportion p matches a null expectation for the proportion.

H_0: The relative frequency of successes in the population is p_0 .

H_A: The relative frequency of successes in the population is not p_0 .

Binomial distribution

Represents the sampling distribution for the number of successes (X) in a random sample of n trials, when the probability of success is the same in each trial

Rather than using a computer to simulate a vast number of random samples, we can use this to calculate the null distribution!

Example

An example: Imagine a student takes a multiple choice test before starting a statistics class. Each of the 10 questions on the test have 5 possible answers, only one of which is correct. This student gets 4 answers right. Can we deduce from this that this student knows anything at all about statistics?

Hypotheses

H_0: Student got correct answers randomly.

H_A: Student got more answers correct than random.

This is properly a one-tailed test.

Hypotheses

H_0: Student got correct answers randomly.

H_0: $p = 0.2$

H_A: Student got more answers correct than random.

H_A: $p > 0.2$

$N = 10,\ p = 0.2$

$$
\begin{aligned}
P &= \Pr[4] + \Pr[5] + \Pr[6] + \ldots + \Pr[10] \\
&= \binom{10}{4}(0.2)^4(0.8)^6 + \binom{10}{5}(0.2)^5(0.8)^5 + \binom{10}{6}(0.2)^6(0.8)^4 + \ldots \\
&= 0.12
\end{aligned}
$$

Note: The capital *P* here is used for the *P*-value, in contrast to the population proportion with a small *p*.

$P = 0.12$

This is greater than the α value of 0.05, so we would *not* reject the null hypothesis.

It is plausible that the student had four answers correct just by guessing randomly.

Estimating Proportions: Proportion of successes in a sample

p is the true population proportion

$$\hat{p} = \frac{X}{n}$$

The hat (^) shows that this is an estimate of *p.*

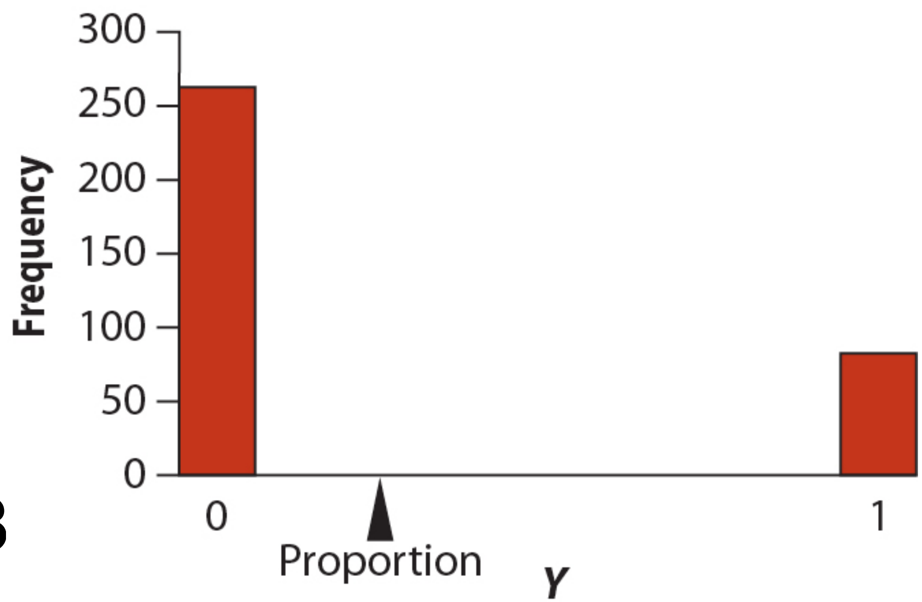
A proportion is like a mean

Yes = 1

No = 0

82/344 = 0.238

(82*1 + 262*0)/344 = 0.238



Variance of the estimate of a proportion is p(1-p)

Case	Worth It?	Score (X)	Mean	(X-mean)	$(X\text{-mean})^2$
1	yes	1	0.6	0.4	0.16
2	no	0	0.6	-0.6	0.36
3	no	0	0.6	-0.6	0.36
4	yes	1	0.6	0.4	0.16
5	yes	1	0.6	0.4	0.16
6	yes	1	0.6	0.4	0.16
7	yes	1	0.6	0.4	0.16
8	no	0	0.6	-0.6	0.36
9	yes	1	0.6	0.4	0.16
10	no	0	0.6	-0.6	0.36
		6/10 =.6 (mean of proportion)			· = 2.4 (sum of squares)

Variance = 2.4/10 = 0.6 * 0.4 = 0.24

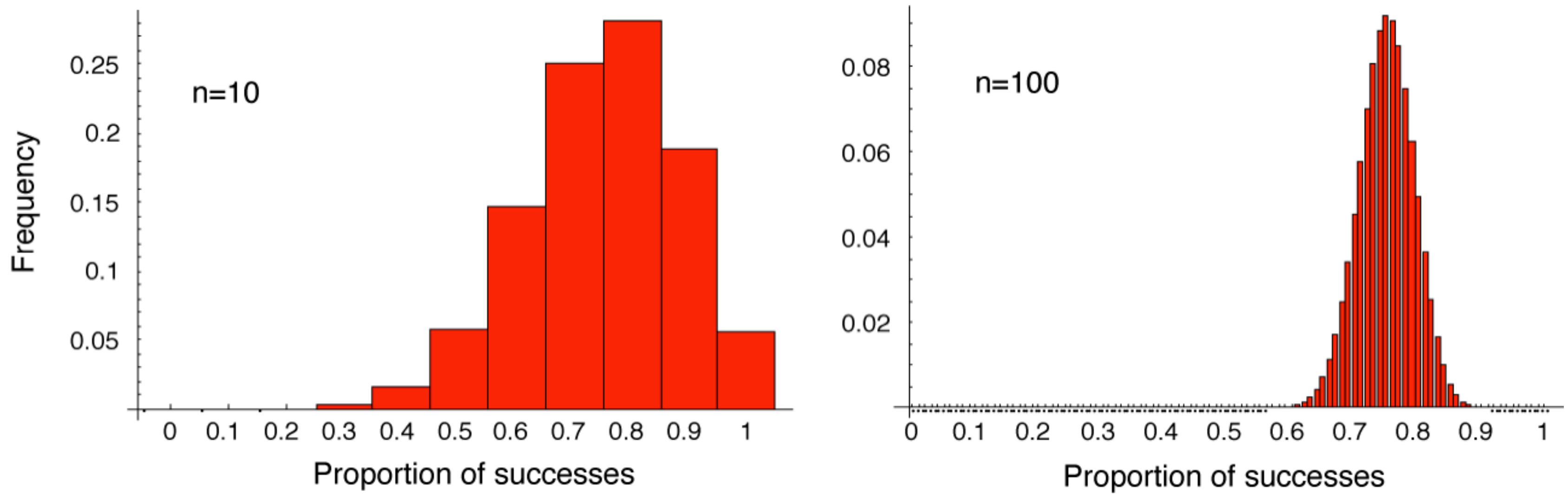
Standard error of the estimate of a proportion is the standard deviation of the sampling distribution

$$\sigma_{\hat{\rho}} = \sqrt{\frac{p(1-p)}{n}}$$

We usually don't know *p* so we estimate the standard error with $\hat{p}$

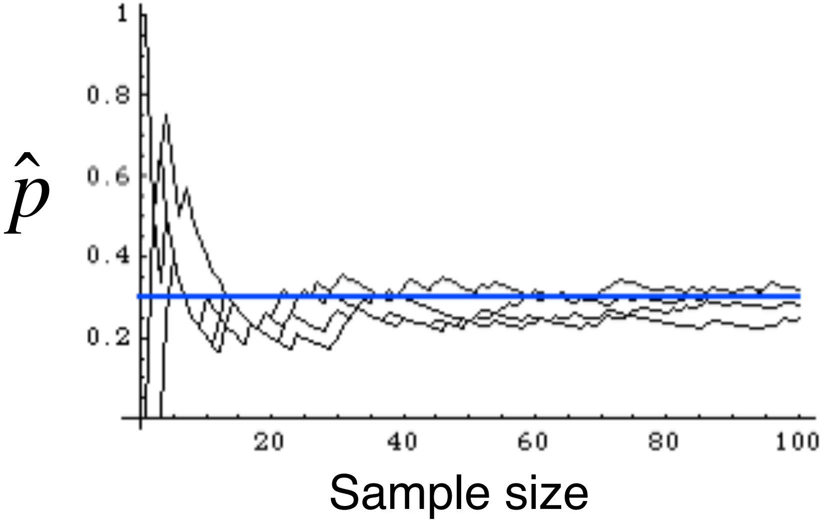
$$SE_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

A larger sample has a lower standard error



The law of large numbers

The greater the sample size, the closer an estimate of a proportion is likely to be to its true value.



95% confidence interval for a proportion

$$p' = \frac{X+2}{n+4}$$

$$\left(p' - 1.96\sqrt{\frac{p'(1-p')}{n+4}} \right) \le p \le \left(p' + 1.96\sqrt{\frac{p'(1-p')}{n+4}} \right)$$

This is the Agresti-Coull confidence interval

Example: The daughters of radiologists

30 out of 87 offspring of male radiologists are males, and the rest female. What is the best estimate of the proportion of sons among radiologists?

Example: The daughters of radiologists

30 out of 87 offspring of ,male radiologists are males, and the rest female. What is the best estimate of the proportion of sons among radiologists?

$$\hat{p} = 30/87, \text{ or } 0.345$$

Example: The daughters of radiologists

30 out of 87 offspring of male radiologists are males, and the rest female. What is the best estimate of the proportion of sons among radiologists? What is the 95% confidence interval for this estimate?

$$\hat{p} = 30/87, \text{ or } 0.345$$

Example: The daughters of radiologists

30 out of 87 offspring of male radiologists are males, and the rest female. What is the best estimate of the proportion of sons among radiologists? What is the 95% confidence interval for this estimate?

$$p' = \frac{X+2}{n+4} = \frac{30+2}{87+4} = 0.352$$

Example: The daughters of radiologists

30 out of 87 offspring of male radiologists are males, and the rest female. What is the best estimate of the proportion of sons among radiologists? What is the 95% confidence interval for this estimate?

$$p' = \frac{X+2}{n+4} = \frac{30+2}{87+4} = 0.352$$

$$p' \pm 1.96\sqrt{\frac{p'(1-p')}{n+4}} = 0.352 \pm 1.96\sqrt{\frac{0.352(1-0.352)}{87+4}}$$

$$= 0.352 \pm 0.098$$

Example: The daughters of radiologists

30 out of 87 offspring of male radiologists are males, and the rest female. What is the best estimate of the proportion of sons among radiologists? What is the 95% confidence interval for this estimate?

$$p' = \frac{X+2}{n+4} = \frac{30+2}{87+4} = 0.352$$

$$p' \pm Z\sqrt{\frac{p'(1-p')}{n+4}} = 0.352 \pm 1.96\sqrt{\frac{0.352(1-0.352)}{87+4}}$$

$$= 0.352 \pm 0.098$$

$$0.254 < p < 0.450$$

Fitting probability models to frequency data

Probability Model

A probability distribution that represents how we think a natural process works

Discrete distribution

A probability distribution describing a discrete numerical random variable

For example,

- Number of heads from 10 flips of a coin
- Number of flowers in a square meter
- Number of disease outbreaks in a year

χ^2 Goodness-of-fit test

Compares counts to a discrete probability distribution

Hypotheses for χ^2 test

H_0: The data come from a particular discrete probability distribution.

H_A: The data do <u>not</u> come from that distribution.

Test statistic for χ^2 test

$$\chi^2 = \sum_{all\ classes} \frac{\left(Observed_i - Expected_i\right)^2}{Expected_i}$$

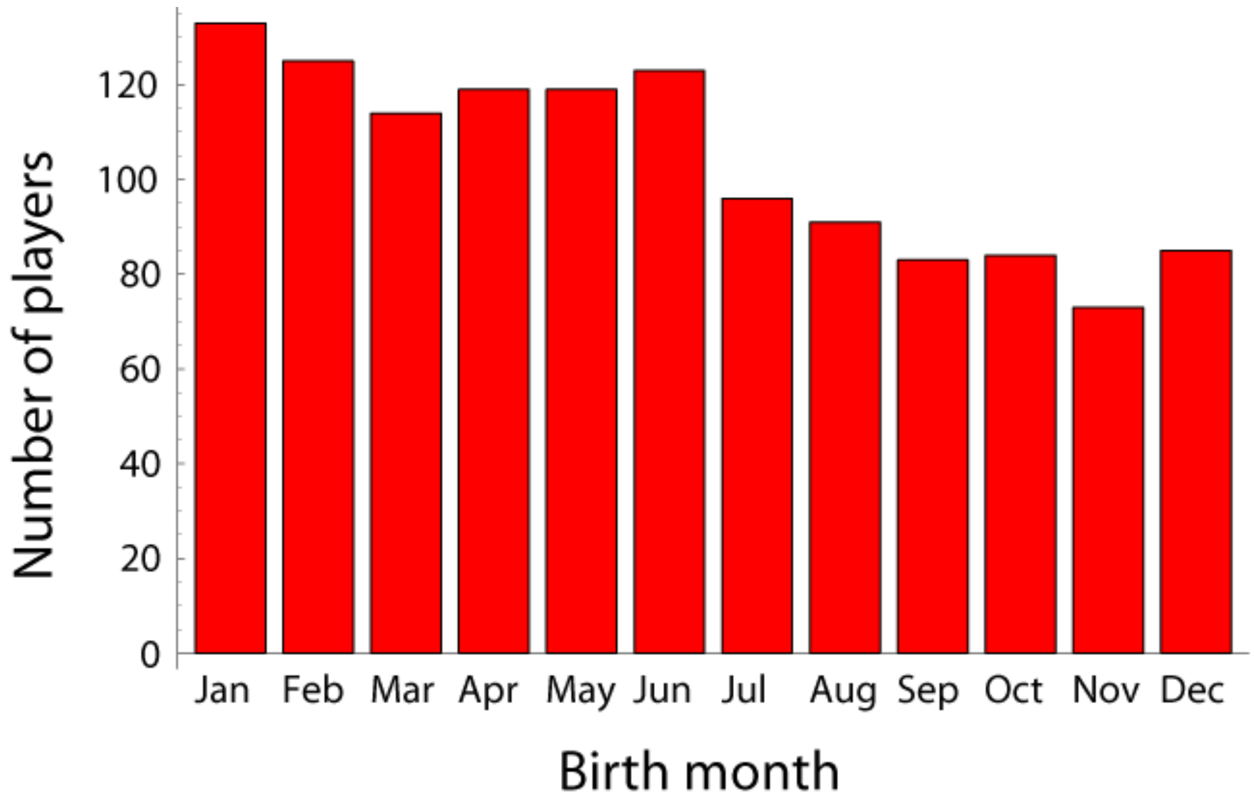
Example: Hockey player size

http://viralinvestor.com/hockey-players-are-big-bullies/

The month of birth for 1245 NHL players

Month	**Number of players**
January	133
February	125
March	114
April	119
May	119
June	123
July	96
August	91
September	83
October	84
November	73
December	85

Data from http://www.nhl.com/players/search/all.html in 2006


Number of players
0
20
40
60
80
100
120
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Birth month

A *Goodness-of-Fit test* compares count data to a model of the expected frequencies of a set of categories.

Hypotheses for birth month example

H_0: The probability of a NHL birth occurring on any given month is equal to national proportions.

H_A: The probability of a NHL birth occurring on any given month is *not* equal to national proportions.

NHL compared to all Canadians

Month	Number of players	Expected (%)
January	133	7.94
February	125	7.63
March	114	8.72
April	119	8.63
May	119	8.95
June	123	8.57
July	96	8.76
August	91	8.5
September	83	8.54
October	84	8.19
November	73	7.70
December	85	7.86
Total	1245	100

Computing Expected values

Month	Number of players	Expected (%)	Expected (of 1245)
January	133	7.94	99
February	125	7.63	95
March	114	8.72	109
April	119	8.63	107
May	119	8.95	111
June	123	8.57	107
July	96	8.76	109
August	91	8.5	106
September	83	8.54	106
October	84	8.19	102
November	73	7.70	96
December	85	7.86	98
Total	1245	100%	1245

Note: For simplicity, we have rounded the expected column to integers. In any real calculation, we would keep a couple decimal places.

The calculation for January

$$\frac{\left(Observed - Expected\right)^2}{Expected} = \frac{\left(133 - 99\right)^2}{99} = \frac{1156}{99}$$

Calculating χ^2

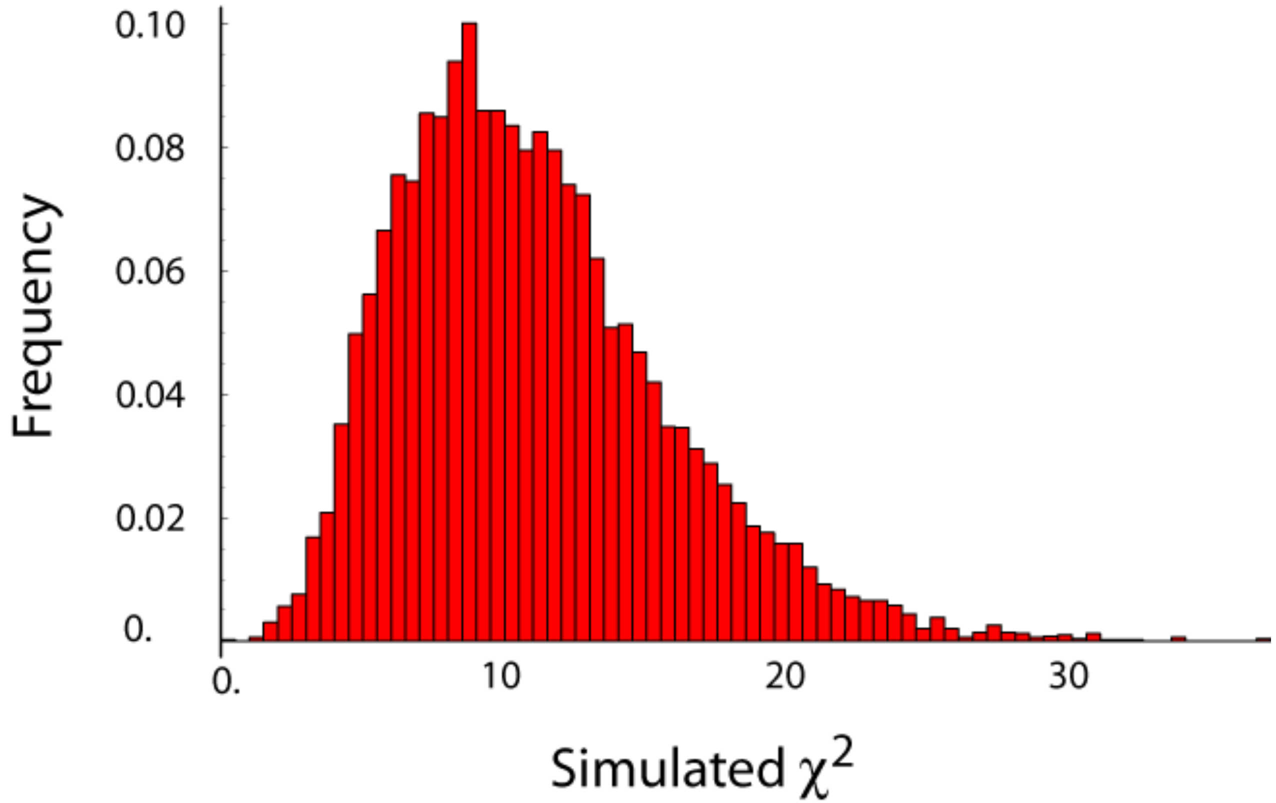
$$\chi^2 = \sum_{all\ classes} \frac{\left(Observed_i - Expected_i\right)^2}{Expected_i}$$

$$= \frac{1156}{99} + \frac{900}{95} + \frac{25}{109} + \frac{144}{107} + \frac{64}{111} + \frac{256}{107} +$$

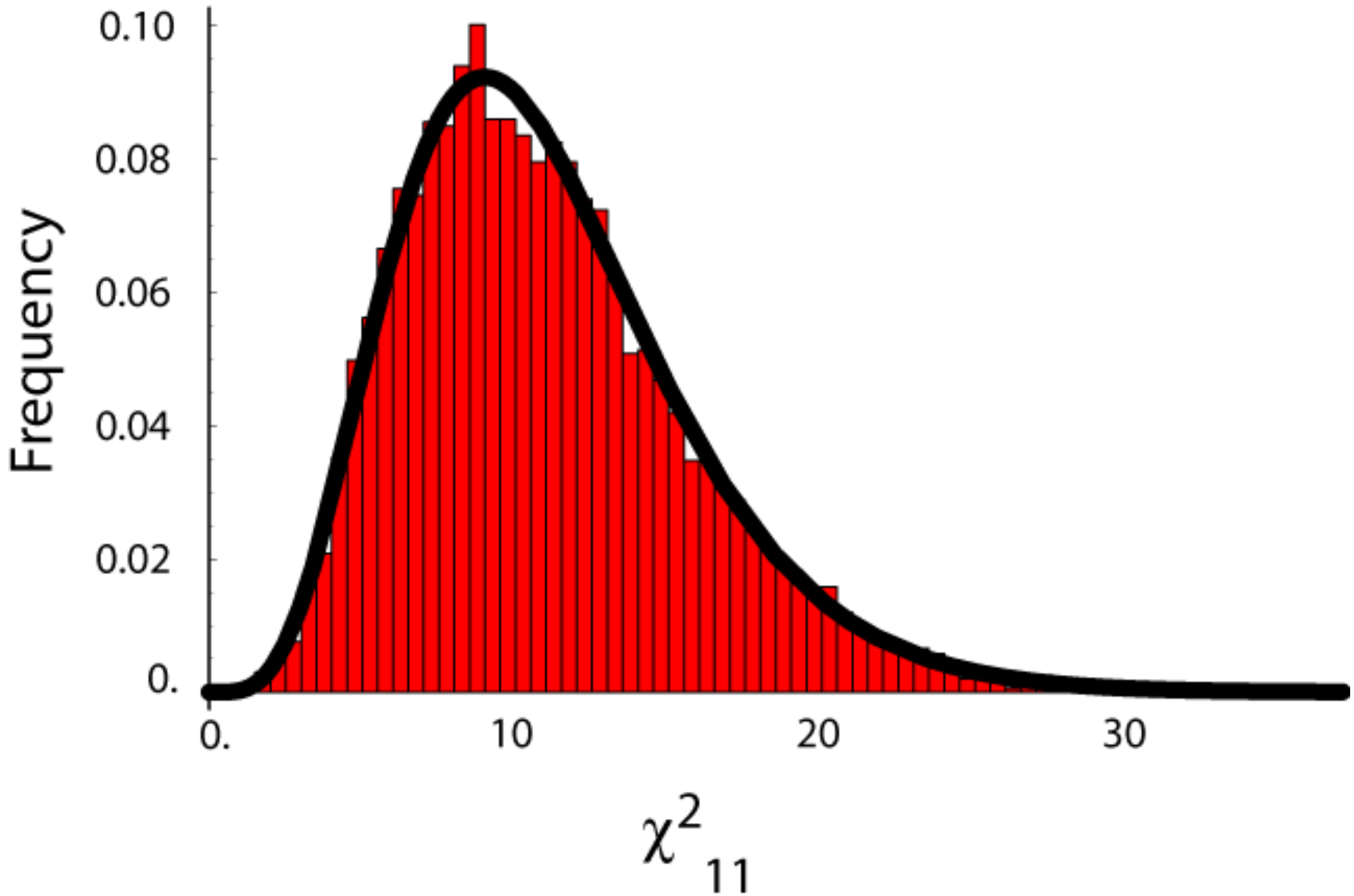
$$\frac{169}{109} + \frac{225}{106} + \frac{529}{106} + \frac{324}{102} + \frac{529}{96} + \frac{169}{98}$$

$$= 44.77$$

The sampling distribution of χ^2 by simulation



Sampling distribution of χ^2 by the χ^2 distribution



Degrees of freedom

The number of *degrees of freedom* of a test specifies which of a family of distributions to use.

Degrees of freedom for χ^2 test

df = (Number of categories)

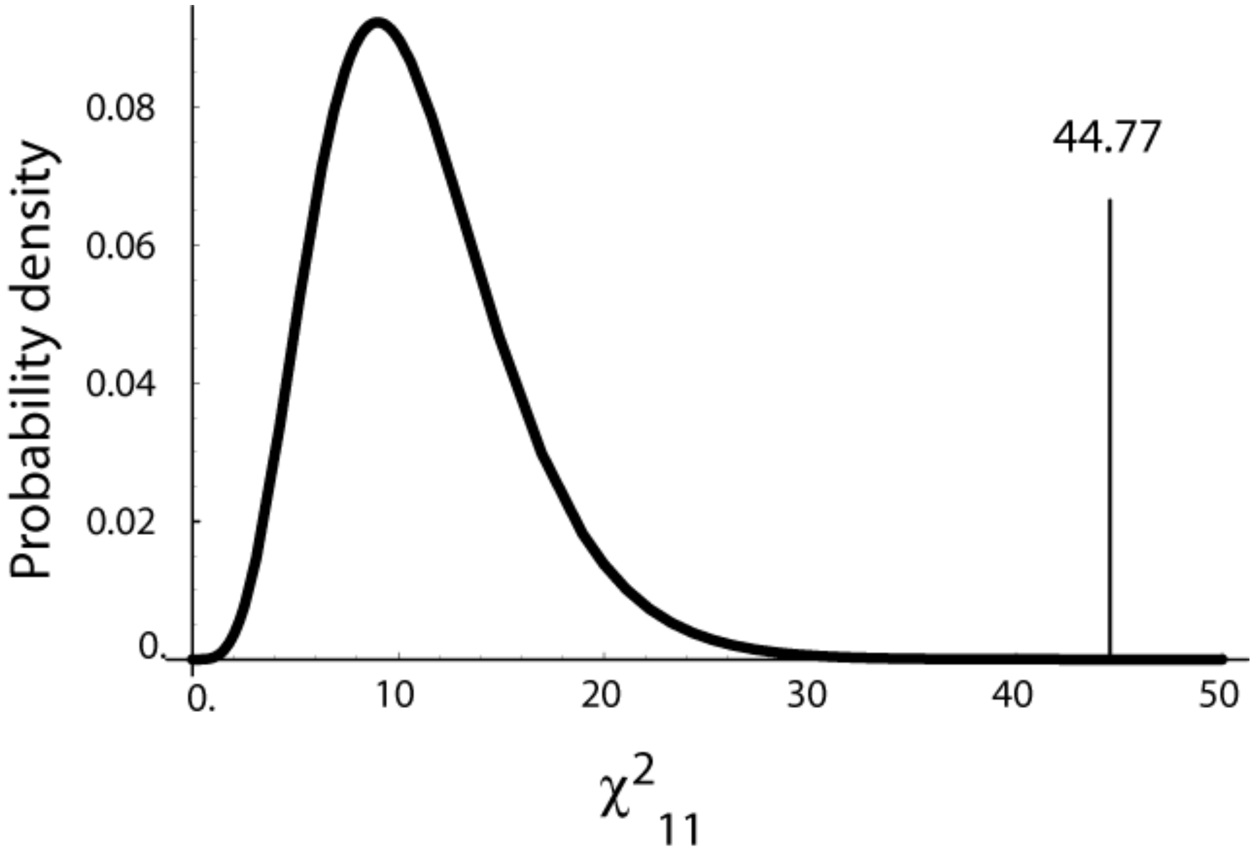
– (Number of parameters estimated from the data)

– 1

Degrees of freedom for NHL month of birth

$$df = 12 - 0 - 1 = 11$$

Finding the *P*-value



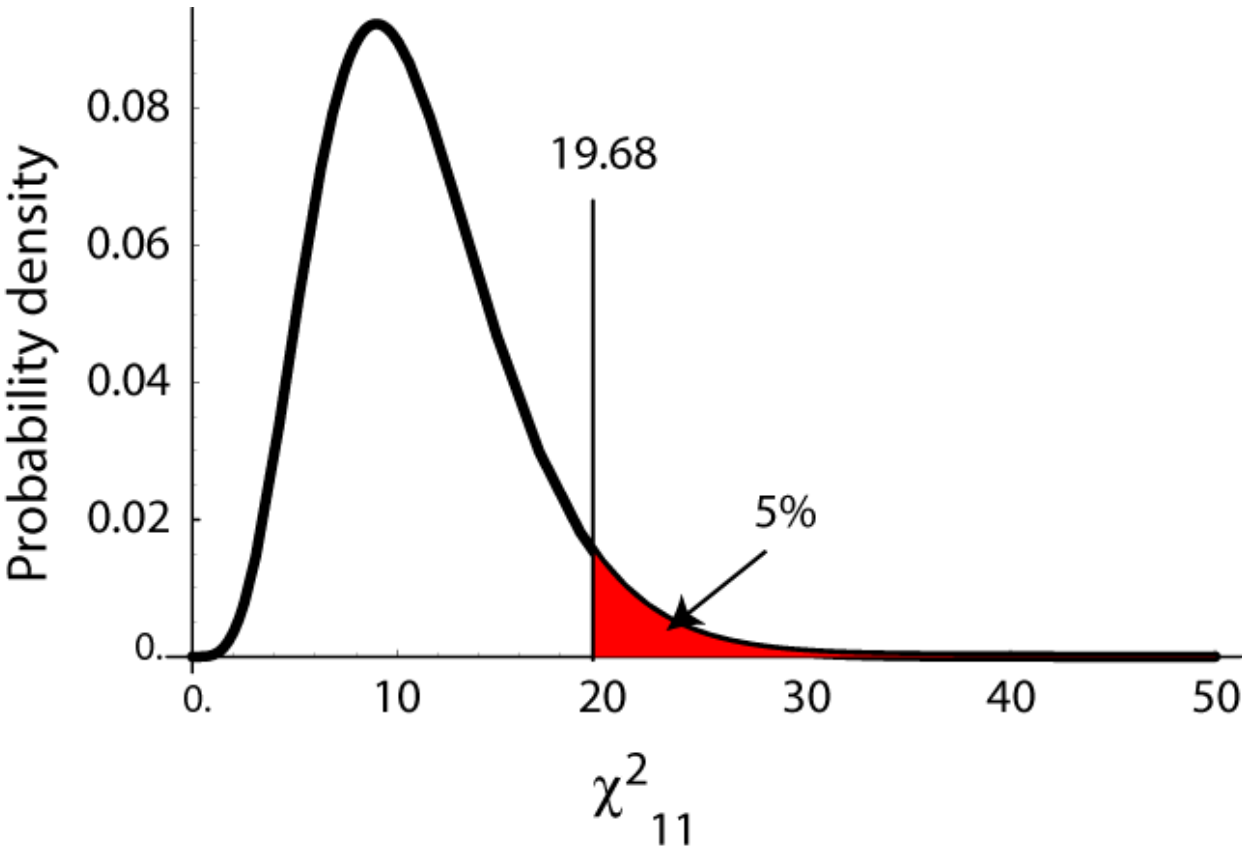
Critical value

The value of the test statistic where $P = \alpha$

Table A - χ^2 distribution

	α									
df	**0.999**	**0.995**	**0.99**	**0.975**	**0.95**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**
1	0.0000016	0.000039	0.00016	0.00098	0.00393	3.84	5.02	6.63	7.88	10.83
2	0.002	0.01	0.02	0.05	0.10	5.99	7.38	9.21	10.60	13.82
3	0.02	0.07	0.11	0.22	0.35	7.81	9.35	11.34	12.84	16.27
4	0.09	0.21	0.30	0.48	0.71	9.49	11.14	13.28	14.86	18.47
5	0.21	0.41	0.55	0.83	1.15	11.07	12.83	15.09	16.75	20.52
6	0.38	0.68	0.87	1.24	1.64	12.59	14.45	16.81	18.55	22.46
7	0.60	0.99	1.24	1.69	2.17	14.07	16.01	18.48	20.28	24.32
8	0.86	1.34	1.65	2.18	2.73	15.51	17.53	20.09	21.95	26.12
9	1.15	1.73	2.09	2.70	3.33	16.92	19.02	21.67	23.59	27.88
10	1.48	2.16	2.56	3.25	3.94	18.31	20.48	23.21	25.19	29.59
11	1.83	2.60	3.05	3.82	4.57	19.68	21.92	24.72	26.76	31.26
12	2.21	3.07	3.57	4.40	5.23	21.03	23.34	26.22	28.30	32.91

The 5% critical value



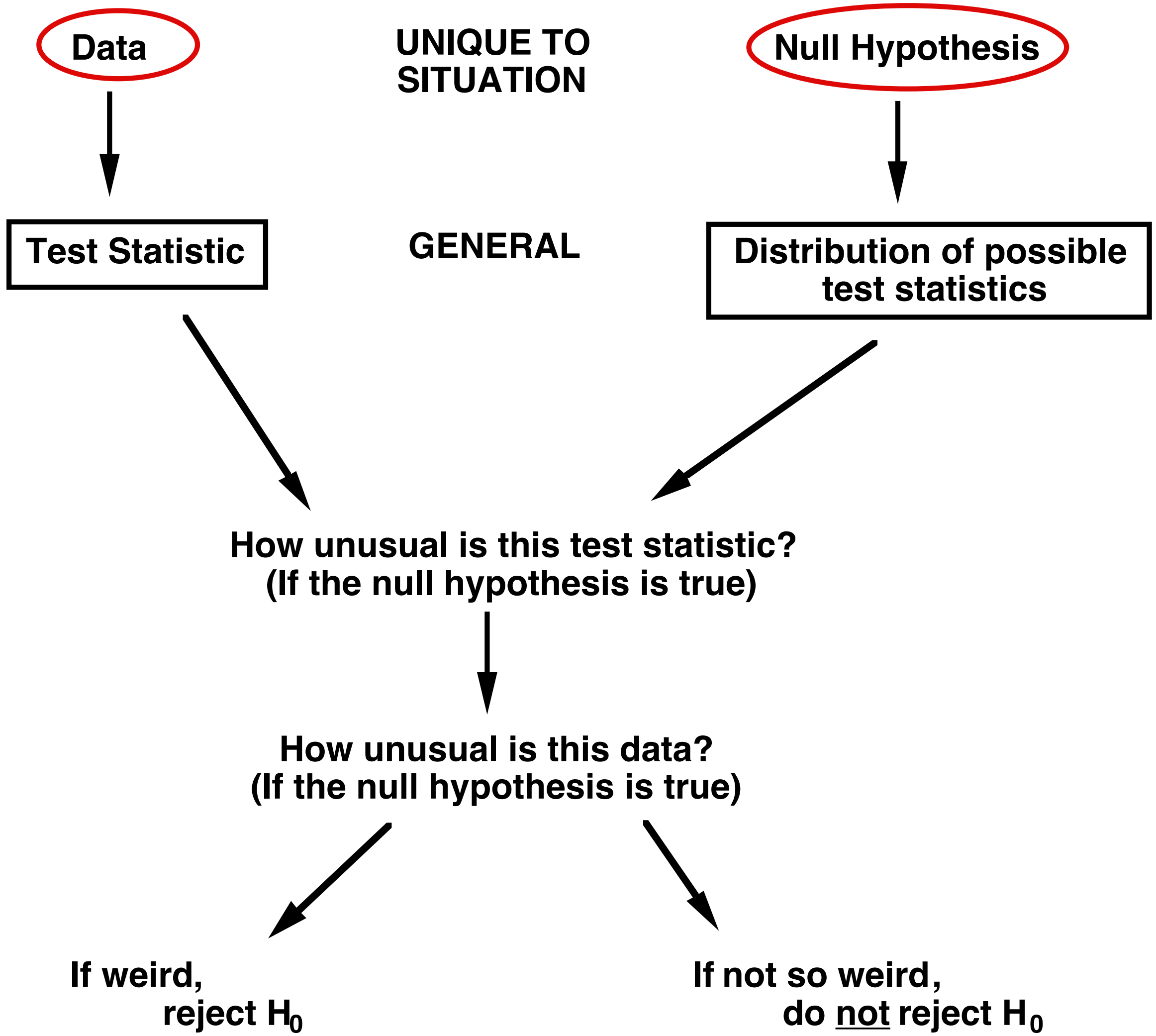
$P < 0.05$,
so we can reject the null hypothesis

NHL players are not born in the same proportions per month as the population at large.

Test statistics

A *test statistic* is a number calculated from the data and the null hypothesis that can be compared to a standard distribution to find the *P*-value of the test.

Test Statistics and Hypothesis Testing



χ^2 test as approximation of binomial test

- χ^2 goodness-of-fit test works even when there are only two categories, so it can be used as a substitute for the binomial test.
- Very useful if the number of data points is large.
  - Imagine if, in our red/blue wrestler example, rather than 16/20 wins by red, we had 1600/2000 wins by red. Imagine calculating: $$\Pr[1600] = \frac{2000!}{1600!\,400!} 0.5^{1600} 0.5^{400}$$
- See text for an example.

Assumptions of χ^2 test

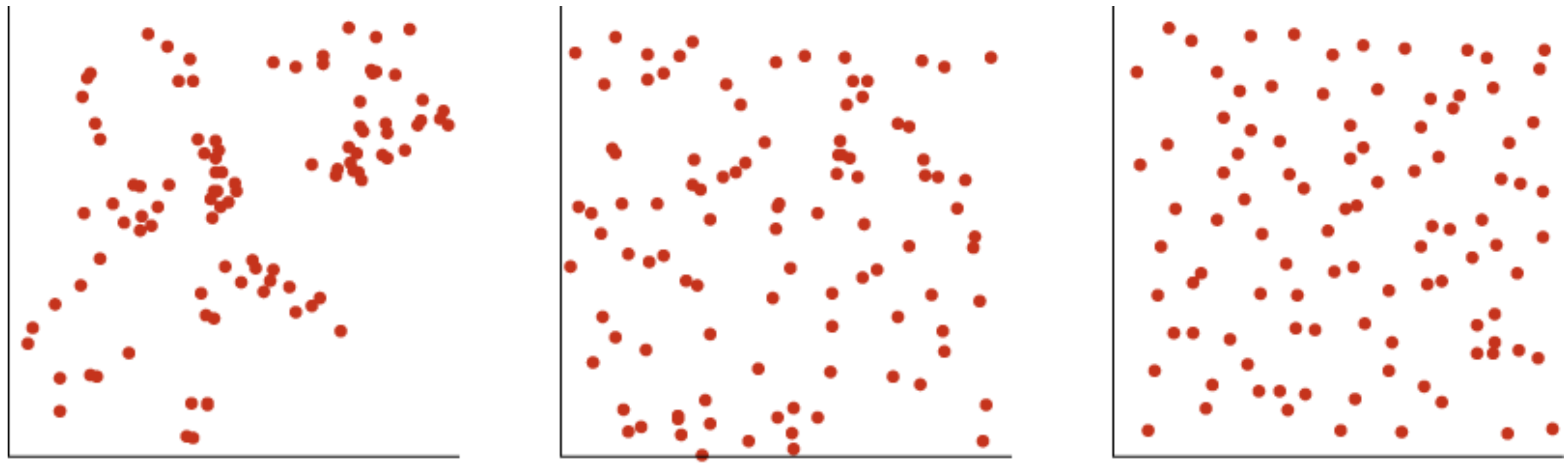
- No more than 20% of categories have *Expected* < 5

- No category with *Expected* ≤ 1

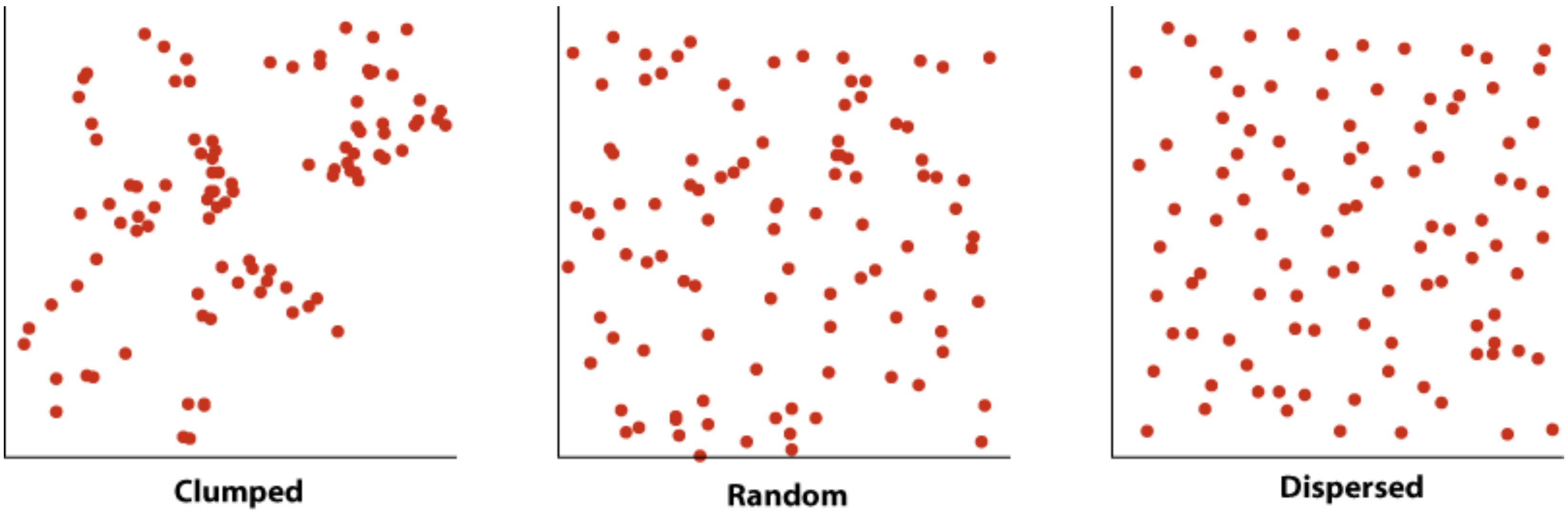
Fitting other distributions: the Poisson distribution

The **Poisson distribution** describes the probability that a certain number of events occur in a block of time or space, when those events happen *independently* of each other and occur *with equal probability* at every point in time or space.

Which one is random?



Which one is random?



Poisson distribution

$$\Pr[X] = \frac{e^{-\mu} \mu^{X}}{X!}$$

Poisson distribution

$$\Pr[X] = \frac{e^{-\mu}\mu^{X}}{X!}$$

Probability to get X success occurring in any block of time or space

Poisson distribution

$$\Pr[X] = \frac{e^{-\mu}\mu^{X}}{X!}$$

probability to get X success occurring in any block of time or space

mean number independent successes in time or space

Poisson distribution

base of natural logarithm (~2.718)

$$\Pr[X] = \frac{e^{-\mu}\mu^{X}}{X!}$$

probability to get X success occurring in any block of time or space

mean number independent successes in time or space

2002 World Cup Soccer example

Example: Number of goals per side in World Cup Soccer

Q: Is the outcome of a soccer game (at this level) random?

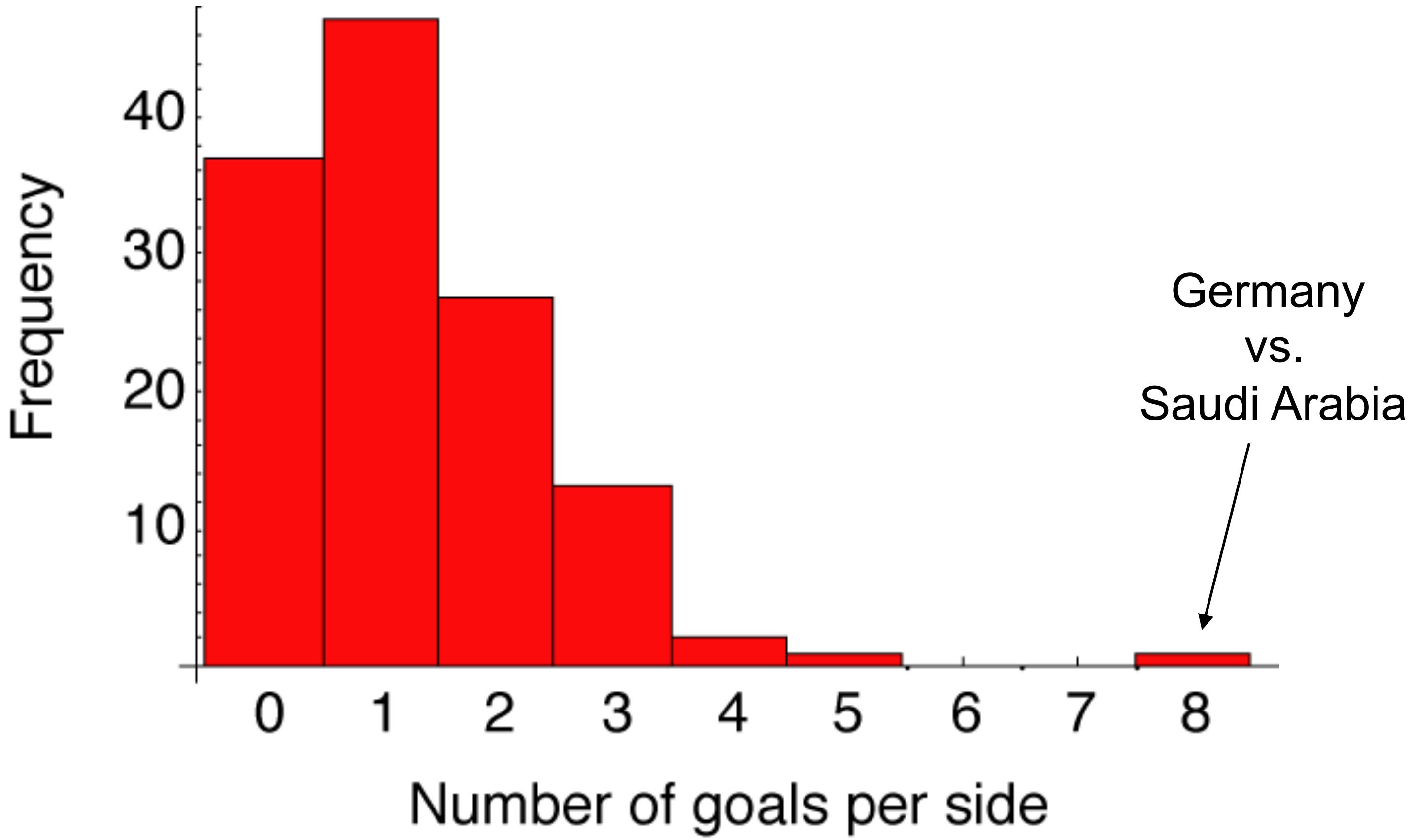
In other words, is the number of goals per team distributed as expected by pure chance?

Hypotheses

- H_0: Number of goals per side follows a Poisson distribution.

- H_A: Number of goals per side does not follow a Poisson distribution.

World Cup 2002 scores



Number of goals for a team (World Cup 2002)

Number of goals	Frequency
0	37
1	47
2	27
3	13
4	2
5	1
6	0
7	0
8	1
Total	128

What's the mean, μ?

$$\bar{x} = \frac{37(0) + 47(1) + 27(2) + 13(3) + 2(4) + 1(5) + 1(8)}{128}$$

$$= \frac{161}{128}$$

$$= 1.26$$

What is what?

Mean number of goals = $\bar{x}$ of previous slide

$$\Pr[X] = \frac{e^{-\mu}\mu^{X}}{X!}$$

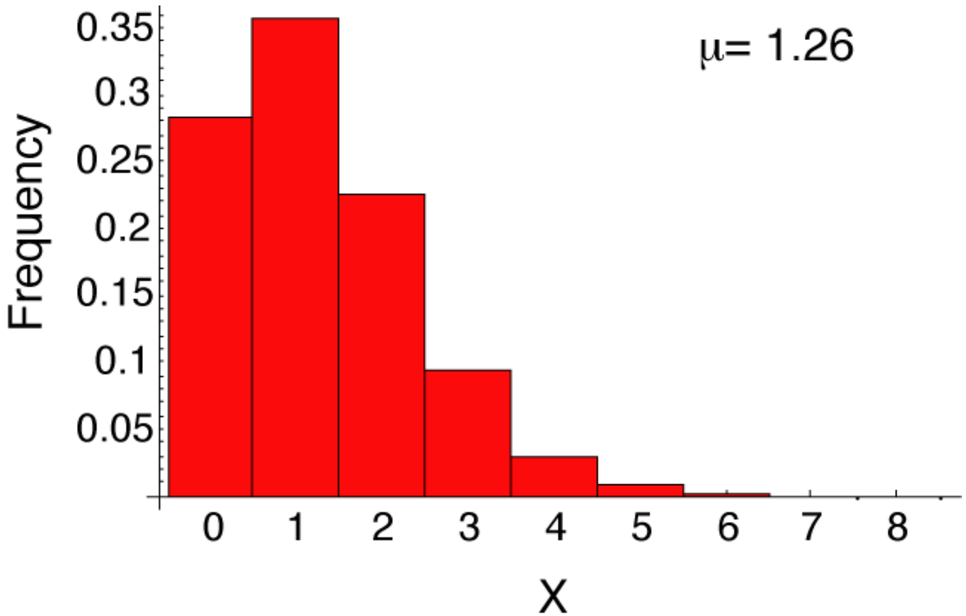
Varies from 0 to 8

Poisson with μ = 1.26

Example:

$$\Pr[2] = \frac{e^{-\mu}\mu^{X}}{X!} = \frac{e^{-1.26}(1.26)^{2}}{2!} = \frac{(0.284)1.59}{2} = 0.225$$

Poisson with $\mu = 1.26$



X	Pr[X]
0	0.284
1	0.357
2	0.225
3	0.095
4	0.030
5	0.008
6	0.002
7	0
≥8	0

Finding the *Expected*

X	Pr[X]	*Expected*
0	0.284	36.3
1	0.357	45.7
2	0.225	28.8
3	0.095	12.1
4	0.030	3.8
5	0.008	1.0
6	0.002	0.2
7	0	0.04
≥8	0	0.007

Finding the *Expected*

X	Pr[X]	*Expected*
0	0.284	36.3
1	0.357	45.7
2	0.225	28.8
3	0.095	12.1
4	0.030	3.8
5	0.008	1.0
6	0.002	0.2
7	0	0.04
≥8	0	0.007

Any problems?

Finding the *Expected*

X	Pr[X]	*Expected*
0	0.284	36.3
1	0.357	45.7
2	0.225	28.8
3	0.095	12.1
4	0.030	3.8
5	0.008	1.0
6	0.002	0.2
7	0	0.04
≥8	0	0.007

too small!

Calculating χ^2

X	*Expected*	*Observed*	$\frac{(Observed_i - Expected_i)^2}{Expected_i}$
0	36.3	37	0.013
1	45.7	47	0.037
2	28.8	27	0.113
3	12.1	13	0.067
≥ 4	5.0	4	0.200

$$\chi^2 = \sum_{all\ classes} \frac{(Observed_i - Expected_i)^2}{Expected_i} = 0.429$$

Degrees of freedom

df = (Number of categories)

– (Number of parameters estimated from the data)

– 1

= 5 – **?** – 1 = ..

Degrees of freedom

df = (Number of categories)

– (Number of parameters estimated from the data)

– 1

$= 5 - \underbrace{1}_{\text{for estimating } \mu} - 1 = 3$

Critical value

	α									
df	**0.999**	**0.995**	**0.99**	**0.975**	**0.95**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**
1	0.0000016	0.000039	0.00016	0.00098	0.00393	3.84	5.02	6.63	7.88	10.83
2	0.002	0.01	0.02	0.05	0.10	5.99	7.38	9.21	10.60	13.82
3	0.02	0.07	0.11	0.22	0.35	7.81	9.35	11.34	12.84	16.27
4	0.09	0.21	0.30	0.48	0.71	9.49	11.14	13.28	14.86	18.47
5	0.21	0.41	0.55	0.83	1.15	11.07	12.83	15.09	16.75	20.52
6	0.38	0.68	0.87	1.24	1.64	12.59	14.45	16.81	18.55	22.46
7	0.60	0.99	1.24	1.69	2.17	14.07	16.01	18.48	20.28	24.32
8	0.86	1.34	1.65	2.18	2.73	15.51	17.53	20.09	21.95	26.12
9	1.15	1.73	2.09	2.70	3.33	16.92	19.02	21.67	23.59	27.88
10	1.48	2.16	2.56	3.25	3.94	18.31	20.48	23.21	25.19	29.59
11	1.83	2.60	3.05	3.82	4.57	19.68	21.92	24.72	26.76	31.26
12	2.21	3.07	3.57	4.40	5.23	21.03	23.34	26.22	28.30	32.91

Comparing χ^2 to the critical value for $\alpha = 0.05$

$$\chi^2 = 0.429$$

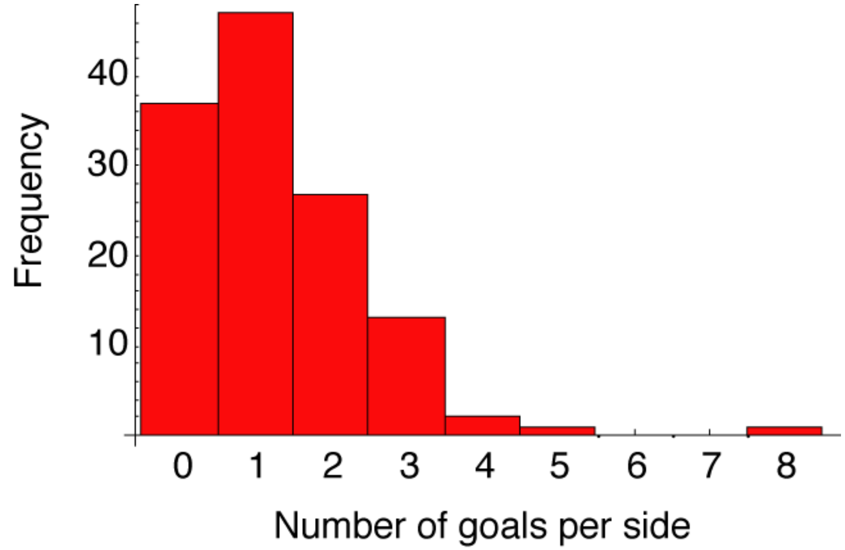
$$\chi_3^2 = 7.81$$

$$0.429 < 7.81$$

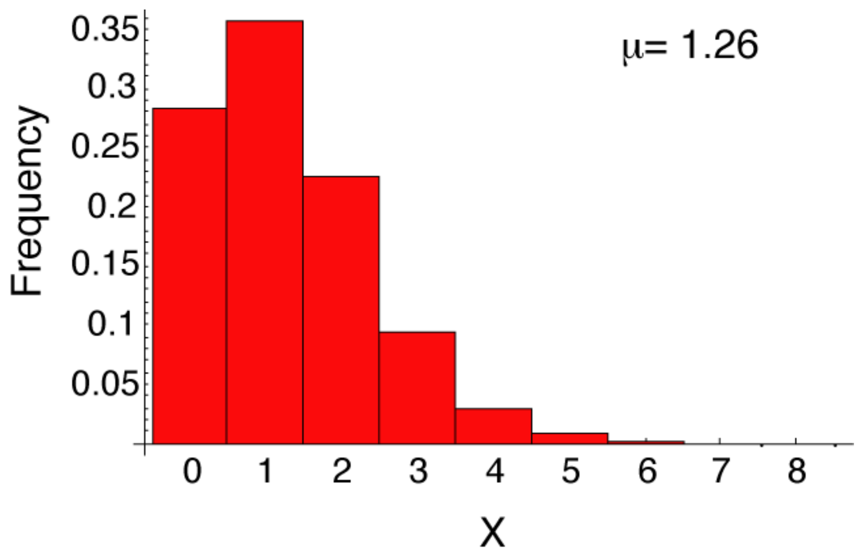
So we cannot reject the null hypothesis.

There is no evidence that the score of a World Cup Soccer game is not Poisson distributed.

World Cup 2002 scores



Poisson distribution



In-class Exercise

The parasitic nematode *Camallanus oxycephalus* infects many freshwater fish, including shad. Do they infect fish at random?

Number of nematodes	Number of fish
0	103
1	72
2	44
3	14
4	3
5	1
6	1

In-class Exercise

The parasitic nematode *Camallanus oxycephalus* infects many freshwater fish, including shad. Do they infect fish at random?

H_0: The number of nematodes per fish has a Poisson distribution

H_A: The number of nematodes per fish does **not** have a Poisson distribution

In-class Exercise

The parasitic nematode *Camallanus oxycephalus* infects many freshwater fish, including shad. Do they infect fish at random?

Number of nematodes	Number of fish	Poisson expectation
0	103	
1	72	
2	44	
3	14	
4	3	
5	1	
6	1	

$$\Pr[X] = \frac{e^{-\mu}\mu^{X}}{X!}$$

$$\bar{Y} = \frac{103(0)+72(1)+44(2)+14(3)+3(4)+1(5)+1(6)}{238}$$

$$= 0.945$$

In-class Exercise

The parasitic nematode *Camallanus oxycephalus* infects many freshwater fish, including shad. Do they infect fish at random?

Number of nematodes	Number of fish	Poisson expectation
0	103	0.389
1	72	0.367
2	44	0.174
3	14	0.055
4	3	0.013
5	1	0.002
6	1	0.000

In-class Exercise

The parasitic nematode *Camallanus oxycephalus* infects many freshwater fish, including shad. Do they infect fish at random?

Number of nematodes	Number of fish	Poisson expectation	Expected
0	103	0.389	92.58
1	72	0.367	87.35
2	44	0.174	41.41
3	14	0.055	13.09
4	3	0.013	3.09
5	1	0.002	0.48
6	1	0.000	0.00

In-class Exercise

The parasitic nematode *Camallanus oxycephalus* infects many freshwater fish, including shad. Do they infect fish at random?

Number of nematodes	Number of fish	Poisson expectation	Expected
0	103	0.389	92.58
1	72	0.367	87.35
2	44	0.174	41.41
3	14	0.055	13.09
4	3	0.013	3.09
5	1	0.002	0.48
6	1	0.000	0.00

Problems?

In-class Exercise

The parasitic nematode *Camallanus oxycephalus* infects many freshwater fish, including shad. Do they infect fish at random?

Number of nematodes	Number of fish	Expected	$(O-E)^2 / E$
0	103	92.58	1.17
1	72	87.35	2.70
2	44	41.41	0.16
3	14	13.09	0.06
≥4	5	3.57	0.57
			$\chi^2 = 4.66$

In-class Exercise

The parasitic nematode *Camallanus oxycephalus* infects many freshwater fish, including shad. Do they infect fish at random?

$\chi^2 = 4.66$

df = 5 – 1 – 1 = 3

Example: Fitting the Binomial Distribution

One thousand coins were flipped eight times, and the number of heads was recorded for each coin. Were they fair coins?

	α									
df	0.999	0.995	0.99	0.975	0.95	0.05	0.025	0.01	0.005	0.001
1	0.0000016	0.000039	0.00016	0.00098	0.00393	3.84	5.02	6.63	7.88	10.83
2	0.002	0.01	0.02	0.05	0.10	5.99	7.38	9.21	10.60	13.82
3	0.02	0.07	0.11	0.22	0.35	7.81	9.35	11.34	12.84	16.27
4	0.09	0.21	0.30	0.48	0.71	9.49	11.14	13.28	14.86	18.47
5	0.21	0.41	0.55	0.83	1.15	11.07	12.83	15.09	16.75	20.52
6	0.38	0.68	0.87	1.24	1.64	12.59	14.45	16.81	18.55	22.46
7	0.60	0.99	1.24	1.69	2.17	14.07	16.01	18.48	20.28	24.32
8	0.86	1.34	1.65	2.18	2.73	15.51	17.53	20.09	21.95	26.12
9	1.15	1.73	2.09	2.70	3.33	16.92	19.02	21.67	23.59	27.88
10	1.48	2.16	2.56	3.25	3.94	18.31	20.48	23.21	25.19	29.59
11	1.83	2.60	3.05	3.82	4.57	19.68	21.92	24.72	26.76	31.26
12	2.21	3.07	3.57	4.40	5.23	21.03	23.34	26.22	28.30	32.91

In-class Exercise

The parasitic nematode *Camallanus oxycephalus* infects many freshwater fish, including shad. Do they infect fish at random?

$\chi^2 = 4.66$

df = 5 – 1 – 1 = 3

$\chi^2_{0.05,3} = 7.81$

P>0.05

We do not reject the null hypothesis.

There is no evidence that nematodes do not infect fish randomly.