

Example

[Carroll, *J Med Entomol* **38**:114–117, 2001]

Place tick on clay island surrounded by water, with two capillary tubes: one treated with deer-gland-substance; one untreated.

Does the tick go to the treated or the untreated tube?

Tick sex	Leg	Deer sex	treated	untreated
male	fore	female	24	5
female	fore	female	18	5
male	fore	male	23	4
female	fore	male	20	4
male	hind	female	17	8
female	hind	female	25	3
male	hind	male	21	6
female	hind	male	25	2

Is the tick more likely to go to the treated tube?

1

Test for a proportion

Suppose $X \sim \text{binomial}(n, p)$.

Test $H_0 : p = \frac{1}{2}$ vs $H_a : p \neq \frac{1}{2}$

Reject H_0 if $X \geq H$ or $X \leq L$

Choose H and L such that

$$\Pr(X \geq H \mid p = \frac{1}{2}) \leq \alpha/2 \text{ and } \Pr(X \leq L \mid p = \frac{1}{2}) \leq \alpha/2$$

Thus $\Pr(\text{Reject } H_0 \mid H_0 \text{ is true}) \leq \alpha$.

The difficulty: The binomial distribution is hard to work with. Because of its discrete nature, you can't get **exactly** your desired significance level (α).

2

Rejection region

Consider $X \sim \text{binomial}(n=29, p)$

Test of $H_0 : p = \frac{1}{2}$ vs $H_a : p \neq \frac{1}{2}$ at significance level $\alpha = 0.05$

Lower critical value:

$$\text{qbinom}(0.025, 29, 0.5) = 9$$

$$\Pr(X \leq 9) = \text{pbinom}(9, 29, 0.5) = 0.031 \rightarrow L = 8$$

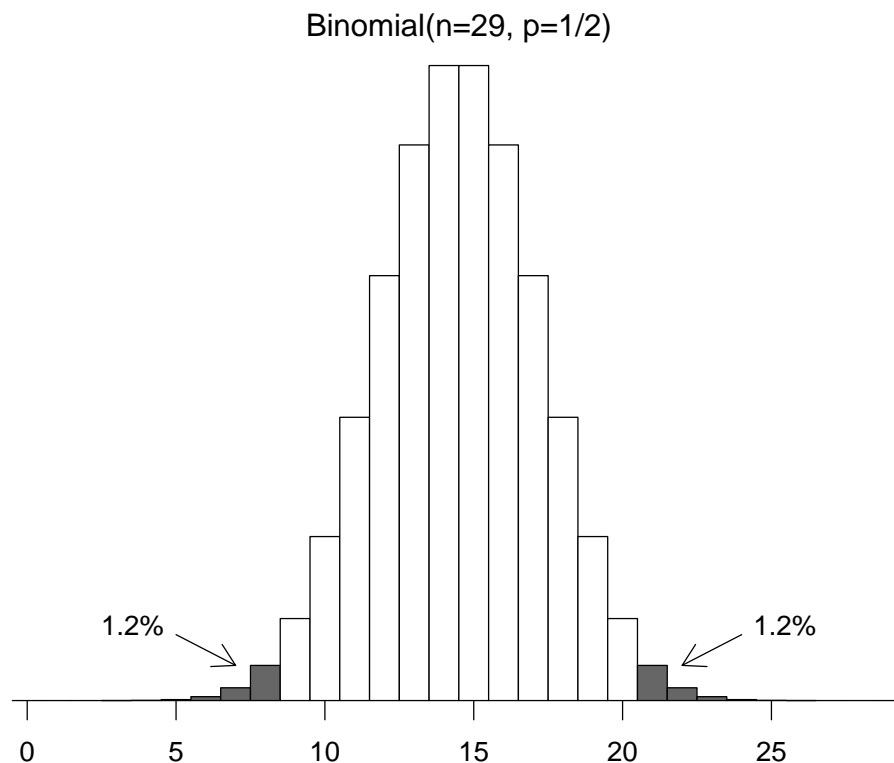
Upper critical value:

$$\text{qbinom}(0.975, 29, 0.5) = 20$$

$$\Pr(X \geq 20) = 1 - \text{pbinom}(20, 29, 0.5) = 0.031 \rightarrow H = 21$$

Reject H_0 if $X \leq 8$ or $X \geq 21$. (For testing $H_0 : p = \frac{1}{2}$, $H = n - L$.)

3



4

Significance level

Consider $X \sim \text{binomial}(n=29, p)$

Test of $H_0 : p = \frac{1}{2}$ vs $H_a : p \neq \frac{1}{2}$ at significance level $\alpha = 0.05$

Reject H_0 if $X \leq 8$ or $X \geq 21$.

Actual significance level:

$$\begin{aligned}\alpha &= \Pr(X \leq 8 \text{ or } X \geq 21 \mid p = \frac{1}{2}) \\ &= \Pr(X \leq 8 \mid p = \frac{1}{2}) + [1 - \Pr(X \leq 20 \mid p = \frac{1}{2})] \\ &= \text{pbinom}(8, 29, 0.5) + 1 - \text{pbinom}(20, 29, 0.5) \\ &= 0.024\end{aligned}$$

If we used, instead, “Reject H_0 if $X \leq 9$ or $X \geq 20$,” the significance level would be:

$$\text{pbinom}(9, 29, 0.5) + 1 - \text{pbinom}(19, 29, 0.5) = 0.061$$

5

Example

Observe $X = 24$ (for $n = 29$)

Reject $H_0 : p = \frac{1}{2}$ if $X \leq 8$ or $X \geq 21$.

Thus we reject H_0 and conclude that the ticks were more likely to go to the deer-gland-substance-treated tube.

$$\begin{aligned}\text{P-value} &= 2 \times \Pr(X \geq 24 \mid p = \frac{1}{2}) \\ &= 2 * (1 - \text{pbinom}(23, 29, 0.5)) \\ &= 5/10,000.\end{aligned}$$

Alternatively: `binom.test(24, 29)`

6

Example 2

Observe $X = 17$ (for $n = 25$); assume $X \sim \text{binomial}(n=25, p)$

Test $H_0 : p = \frac{1}{2}$ vs $H_a : p \neq \frac{1}{2}$

Rejection rule: Reject H_0 if $X \leq 7$ or $X \geq 18$

`qbinom(0.025, 25, 0.5) = 8`

`pbinom(8, 25, 0.5) = 0.054`

`pbinom(7, 25, 0.5) = 0.022`

Significance level:

`pbinom(7, 25, 0.5) + 1 - pbinom(17, 25, 0.5) = 0.043`

Since we observed $X = 17$, we fail to reject H_0

P-value = $2 * (1 - \text{pbinom}(16, 25, 0.5)) = 0.11$

7

Confidence interval for a proportion

Suppose $X \sim \text{binomial}(n=29, p)$ and we observe $X = 24$.

Consider the test of $H_0 : p = p_0$ vs $H_a : p \neq p_0$

We reject H_0 if

$$\Pr(X \leq 24 \mid p = p_0) \leq \alpha/2 \quad \text{or} \quad \Pr(X \geq 24 \mid p = p_0) \leq \alpha/2$$

95% confidence interval for p :

The set of p_0 for which a two-tailed test of $H_0 : p = p_0$ would **not** be rejected, for the observed data, with $\alpha = 0.05$.

The “plausible” values of p .

8

Example

$X \sim \text{binomial}(n=29, p)$; observe $X = 24$

Lower bound of 95% confidence interval:

Largest p_0 such that $\Pr(X \geq 24 \mid p = p_0) \leq 0.025$

Upper bound of 95% confidence interval:

Smallest p_0 such that $\Pr(X \leq 24 \mid p = p_0) \leq 0.025$

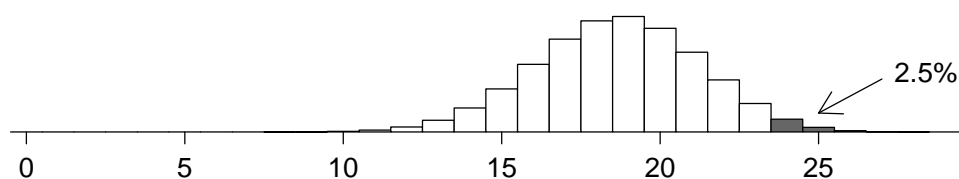
In R: `binom.test(24, 29)`

95% CI for p : (0.642, 0.942)

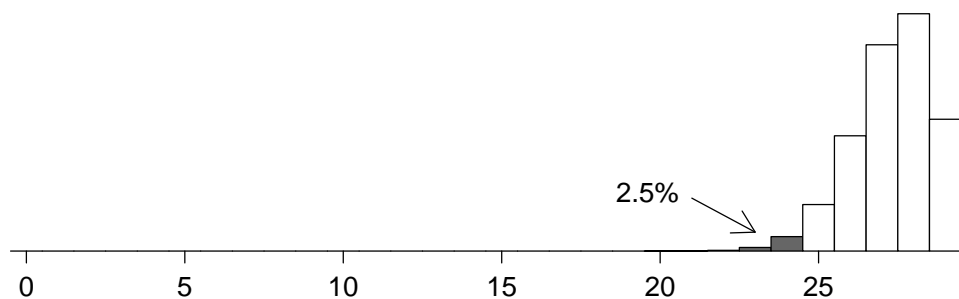
Note: $\hat{p} = 24/29 = 0.83$ is not the midpoint of the CI

9

Binomial($n=29$, $p=0.64$)



Binomial($n=29$, $p=0.94$)



Example 2

$X \sim \text{binomial}(n=25, p)$; observe $X = 17$

Lower bound of 95% confidence interval:

p_L such that 17 is the 97.5 percentile of $\text{binomial}(n=25, p_L)$

Upper bound of 95% confidence interval:

p_H such that 17 is the 2.5 percentile of $\text{binomial}(n=25, p_H)$

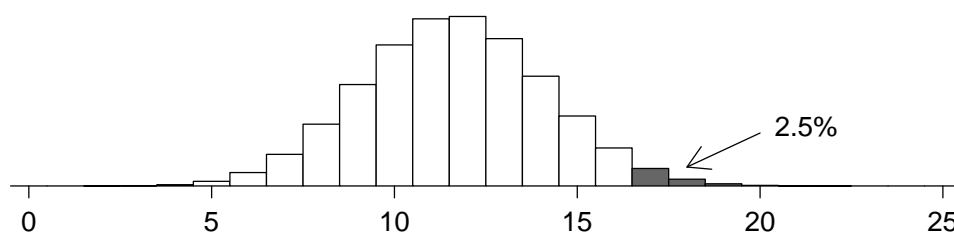
In R: `binom.test(17,25)`

95% CI for p : (0.465, 0.851)

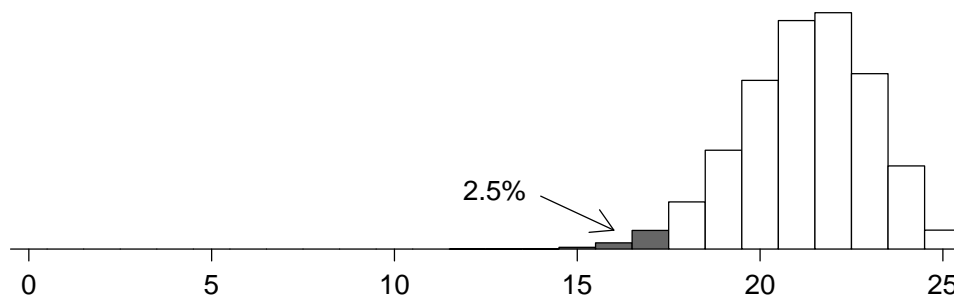
Again, $\hat{p} = 17/25 = 0.68$ is not the midpoint of the CI

11

Binomial($n=25, p=0.46$)



Binomial($n=25, p=0.85$)



12

The case $X = 0$

Suppose $X \sim \text{binomial}(n, p)$ and we observe $X = 0$.

Lower limit of 95% confidence interval for p : 0

Upper limit of 95% confidence interval for p :

p_H such that

$$\begin{aligned}\Pr(X \leq 0 \mid p = p_H) &= 0.025 \\ \implies \Pr(X = 0 \mid p = p_H) &= 0.025 \\ \implies (1 - p_H)^n &= 0.025 \\ \implies 1 - p_H &= \sqrt[n]{0.025} \\ \implies p_H &= 1 - \sqrt[n]{0.025}\end{aligned}$$

In the case $n = 10$ and $X = 0$, the 95% CI for p is (0, 0.31)

13

A mad cow example

New York Times, Feb 3, 2004:

The department [of Agriculture] has not changed last year's plans to test 40,000 cows nationwide this year, out of 30 million slaughtered. Janet Riley, a spokeswoman for the American Meat Institute, which represents slaughterhouses, called that "plenty sufficient from a statistical standpoint."

Suppose that the 40,000 cows tested are chosen at random from the population of 30 million cows, and suppose that 0 (or 1, or 2) are found to be infected.

How many of the 30 million total cows would we estimate to be infected?

What is the 95% confidence interval for the total number of infected cows?

No. infected		
Obs'd	Est'd	95% CI
0	0	0 – 2763
1	750	19 – 4173
2	1500	181 – 5411

14

The case $X = n$

Suppose $X \sim \text{binomial}(n, p)$ and we observe $X = n$.

Upper limit of 95% confidence interval for p : 1

Lower limit of 95% confidence interval for p :

p_L such that

$$\Pr(X \geq n \mid p = p_L) = 0.025$$

$$\implies \Pr(X = n \mid p = p_L) = 0.025$$

$$\implies (p_L)^n = 0.025$$

$$\implies p_L = \sqrt[n]{0.025}$$

In the case $n = 25$ and $X = 25$, the 95% CI for p is (0.86, 1.00)

15

Large n and medium p

Suppose $X \sim \text{binomial}(n, p)$.

$$E(X) = np \quad SD(X) = \sqrt{np(1-p)}$$

$$\hat{p} = X/n \quad E(\hat{p}) = p \quad SD(\hat{p}) = \sqrt{\frac{p(1-p)}{n}}$$

For large n and medium p , $\hat{p} \sim \text{normal}\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$

Use 95% confidence interval $\hat{p} \pm 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Unfortunately, this usually behaves poorly.

Fortunately, you can just use `binom.test()`

16