Hypothesis tests

with binomial example

STAT 587 (Engineering)
Iowa State University

September 30, 2020

Statistical hypothesis testing

A hypothesis test consists of two hypotheses,

- null hypothesis (H_0) and
- an alternative hypothesis (H_A),

Statistical hypothesis testing

A hypothesis test consists of two hypotheses,

- null hypothesis (H_0) and
- an alternative hypothesis (H_A),

which make claims about parameter(s) in a model,

Statistical hypothesis testing

A hypothesis test consists of two hypotheses,

- null hypothesis (H_0) and
- an alternative hypothesis (H_A),

which make claims about parameter(s) in a model, and a decision to either

- reject the null hypothesis or
- fail to reject the null hypothesis.

Binomial model

If $Y \sim Bin(n, \theta)$, then some hypothesis tests are

$$H_0 : \theta = \theta_0$$

Binomial model

If $Y \sim Bin(n, \theta)$, then some hypothesis tests are

$$H_0 : \theta = \theta_0 \qquad \text{versus} \qquad H_A : \theta \neq \theta_0$$

Binomial model

If $Y \sim Bin(n, \theta)$, then some hypothesis tests are

$$H_0 : \theta = \theta_0 \qquad \text{versus} \qquad H_A : \theta \neq \theta_0$$

or

$$H_0 : \theta \leq \theta_0 \qquad \text{versus} \qquad H_A : \theta > \theta_0$$

Binomial model

If $Y \sim Bin(n, \theta)$, then some hypothesis tests are

$$H_0 : \theta = \theta_0 \qquad \text{versus} \qquad H_A : \theta \neq \theta_0$$

or

$$H_0 : \theta \leq \theta_0 \qquad \text{versus} \qquad H_A : \theta > \theta_0$$

or

$$H_0 : \theta \geq \theta_0 \qquad \text{versus} \qquad H_A : \theta < \theta_0$$

Binomial model

If $Y \sim Bin(n, \theta)$, then some hypothesis tests are

$$H_0 : \theta = \theta_0 \qquad \text{versus} \qquad H_A : \theta \neq \theta_0$$

or

$$H_0 : \theta = \theta_0 \qquad \text{versus} \qquad H_A : \theta > \theta_0$$

or

$$H_0 : \theta = \theta_0 \qquad \text{versus} \qquad H_A : \theta < \theta_0$$

Small data

Let $Y \sim Bin(n, \theta)$ with

$$H_0 : \theta = 0.5 \qquad \text{versus} \qquad H_A : \theta \neq 0.5.$$

Small data

Let $Y \sim Bin(n, \theta)$ with

$$H_0 : \theta = 0.5 \qquad \text{versus} \qquad H_A : \theta \neq 0.5.$$

You collect data and observe $y = 6$ out of $n = 13$ attempts.

Small data

Let $Y \sim Bin(n, \theta)$ with

$$H_0 : \theta = 0.5 \qquad \text{versus} \qquad H_A : \theta \neq 0.5.$$

You collect data and observe $y = 6$ out of $n = 13$ attempts. Should you reject H_0?

Small data

Let $Y \sim Bin(n, \theta)$ with

$$H_0 : \theta = 0.5 \qquad \text{versus} \qquad H_A : \theta \neq 0.5.$$

You collect data and observe $y = 6$ out of $n = 13$ attempts. Should you reject H_0? Probably not since $6 \approx E[Y] = 6.5$ if H_0 is true.

Small data

Let $Y \sim Bin(n, \theta)$ with

$$H_0 : \theta = 0.5 \qquad \text{versus} \qquad H_A : \theta \neq 0.5.$$

You collect data and observe $y = 6$ out of $n = 13$ attempts. Should you reject H_0? Probably not since $6 \approx E[Y] = 6.5$ if H_0 is true.

What if you observed $y = 2$?

Small data

Let $Y \sim Bin(n, \theta)$ with

$$H_0 : \theta = 0.5 \qquad \text{versus} \qquad H_A : \theta \neq 0.5.$$

You collect data and observe $y = 6$ out of $n = 13$ attempts. Should you reject H_0? Probably not since $6 \approx E[Y] = 6.5$ if H_0 is true.

What if you observed $y = 2$? Well, $P(Y = 2) \approx 0.01$.

Large data

Let $Y \sim Bin(n, \theta)$ with

$$H_0 : \theta = 0.5 \qquad \text{versus} \qquad H_A : \theta \neq 0.5.$$

You collect data and observe $y = 6500$ out of $n = 13000$ attempts.

Large data

Let $Y \sim Bin(n, \theta)$ with

$$H_0 : \theta = 0.5 \qquad \text{versus} \qquad H_A : \theta \neq 0.5.$$

You collect data and observe $y = 6500$ out of $n = 13000$ attempts. Should you reject H_0?

Large data

Let $Y \sim Bin(n, \theta)$ with

$$H_0 : \theta = 0.5 \qquad \text{versus} \qquad H_A : \theta \neq 0.5.$$

You collect data and observe $y = 6500$ out of $n = 13000$ attempts. Should you reject H_0? Probably not since $6500 = E[Y]$ if H_0 is true.

Large data

Let $Y \sim Bin(n, \theta)$ with

$$H_0 : \theta = 0.5 \qquad \text{versus} \qquad H_A : \theta \neq 0.5.$$

You collect data and observe $y = 6500$ out of $n = 13000$ attempts. Should you reject H_0? Probably not since $6500 = E[Y]$ if H_0 is true. But $P(Y = 6500) \approx 0.007$.

p-values

p-value: the probability of observing a test statistic as or more extreme than observed if the null hypothesis is true

p-values

p-value: the probability of observing a test statistic as or more extreme than observed if the null hypothesis is true

The as or more extreme region is determined by the alternative hypothesis.

p-values

p-value: the probability of observing a test statistic as or more extreme than observed if the null hypothesis is true

The as or more extreme region is determined by the alternative hypothesis.

For example, if $Y \sim Bin(n, \theta)$ and $H_0 : \theta = \theta_0$ then

$$H_A : \theta < \theta_0$$

p-values

p-value: the probability of observing a test statistic as or more extreme than observed if the null hypothesis is true

The as or more extreme region is determined by the alternative hypothesis.

For example, if $Y \sim Bin(n, \theta)$ and $H_0 : \theta = \theta_0$ then

$$H_A : \theta < \theta_0 \implies Y \leq y$$

p-values

p-value: the probability of observing a test statistic as or more extreme than observed if the null hypothesis is true

The as or more extreme region is determined by the alternative hypothesis.

For example, if $Y \sim Bin(n, \theta)$ and $H_0 : \theta = \theta_0$ then

$$H_A : \theta < \theta_0 \implies Y \leq y$$

or

$$H_A : \theta > \theta_0$$

p-values

p-value: the probability of observing a test statistic as or more extreme than observed if the null hypothesis is true

The as or more extreme region is determined by the alternative hypothesis.

For example, if $Y \sim Bin(n, \theta)$ and $H_0 : \theta = \theta_0$ then

$$H_A : \theta < \theta_0 \implies Y \leq y$$

or

$$H_A : \theta > \theta_0 \implies Y \geq y$$

p-values

p-value: the probability of observing a test statistic as or more extreme than observed if the null hypothesis is true

The as or more extreme region is determined by the alternative hypothesis.

For example, if $Y \sim Bin(n, \theta)$ and $H_0 : \theta = \theta_0$ then

$$H_A : \theta < \theta_0 \implies Y \leq y$$

or

$$H_A : \theta > \theta_0 \implies Y \geq y$$

or

$$H_A : \theta \neq \theta_0$$

p-values

p-value: the probability of observing a test statistic as or more extreme than observed if the null hypothesis is true

The as or more extreme region is determined by the alternative hypothesis.

For example, if $Y \sim Bin(n, \theta)$ and $H_0 : \theta = \theta_0$ then

$$H_A : \theta < \theta_0 \implies Y \leq y$$

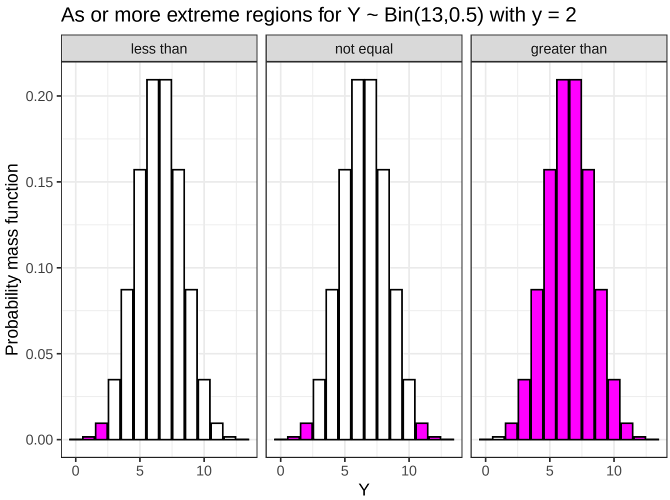
or

$$H_A : \theta > \theta_0 \implies Y \geq y$$

or

$$H_A : \theta \neq \theta_0 \implies |Y - n\theta_0| \geq |y - n\theta_0|.$$

as or more extreme regions



R "hand" calculation

$$H_A : \theta < 0.5 \implies p\text{-value} = P(Y \leq y)$$

R "hand" calculation

$$H_A : \theta < 0.5 \implies p\text{-value} = P(Y \leq y)$$

```
pbinom(y, size = n, prob = theta0)
```

```
[1] 0.01123047
```

R "hand" calculation

$$H_A : \theta < 0.5 \implies p\text{-value} = P(Y \leq y)$$

```
pbinom(y, size = n, prob = theta0)
```

```
[1] 0.01123047
```

$$H_A : \theta > 0.5 \implies p\text{-value} = P(Y \geq y)$$

R "hand" calculation

$$H_A : \theta < 0.5 \implies p\text{-value} = P(Y \leq y)$$

```
pbinom(y, size = n, prob = theta0)

[1] 0.01123047
```

$$H_A : \theta > 0.5 \implies p\text{-value} = P(Y \geq y) = 1 - P(Y \leq y - 1)$$

R "hand" calculation

$$H_A : \theta < 0.5 \implies p\text{-value} = P(Y \leq y)$$

```r
pbinom(y, size = n, prob = theta0)
```

```
[1] 0.01123047
```

$$H_A : \theta > 0.5 \implies p\text{-value} = P(Y \geq y) = 1 - P(Y \leq y - 1)$$

```r
1-pbinom(y-1, size = n, prob = theta0)
```

```
[1] 0.998291
```

R "hand" calculation

$$H_A : \theta < 0.5 \implies p\text{-value} = P(Y \leq y)$$

```
pbinom(y, size = n, prob = theta0)

[1] 0.01123047
```

$$H_A : \theta > 0.5 \implies p\text{-value} = P(Y \geq y) = 1 - P(Y \leq y - 1)$$

```
1-pbinom(y-1, size = n, prob = theta0)

[1] 0.998291
```

$$H_A : \theta \neq 0.5 \implies p\text{-value} = P(|Y - n\theta_0| \leq |y - n\theta_0|)$$

R "hand" calculation

$$H_A : \theta < 0.5 \implies p\text{-value} = P(Y \leq y)$$

```
pbinom(y, size = n, prob = theta0)

[1] 0.01123047
```

$$H_A : \theta > 0.5 \implies p\text{-value} = P(Y \geq y) = 1 - P(Y \leq y - 1)$$

```
1-pbinom(y-1, size = n, prob = theta0)

[1] 0.998291
```

$$H_A : \theta \neq 0.5 \implies p\text{-value} = P(|Y - n\theta_0| \leq |y - n\theta_0|)$$

```
2*pbinom(y, size = n, prob = theta0)

[1] 0.02246094
```

R Calculation

$$H_A : \theta < 0.5$$

R Calculation

$$H_A : \theta < 0.5$$

```
binom.test(y, n, p = theta0, alternative = "less")$p.value
```

```
[1] 0.01123047
```

R Calculation

$$H_A : \theta < 0.5$$

```
binom.test(y, n, p = theta0, alternative = "less")$p.value

[1] 0.01123047
```

$$H_A : \theta > 0.5$$

R Calculation

$$H_A : \theta < 0.5$$

```
binom.test(y, n, p = theta0, alternative = "less")$p.value

[1] 0.01123047
```

$$H_A : \theta > 0.5$$

```
binom.test(y, n, p = theta0, alternative = "greater")$p.value

[1] 0.998291
```

R Calculation

$$H_A : \theta < 0.5$$

```
binom.test(y, n, p = theta0, alternative = "less")$p.value

[1] 0.01123047
```

$$H_A : \theta > 0.5$$

```
binom.test(y, n, p = theta0, alternative = "greater")$p.value

[1] 0.998291
```

$$H_A : \theta \neq 0.5$$

R Calculation

$$H_A : \theta < 0.5$$

```
binom.test(y, n, p = theta0, alternative = "less")$p.value

[1] 0.01123047
```

$$H_A : \theta > 0.5$$

```
binom.test(y, n, p = theta0, alternative = "greater")$p.value

[1] 0.998291
```

$$H_A : \theta \neq 0.5$$

```
binom.test(y, n, p = theta0, alternative = "two.sided")$p.value

[1] 0.02246094
```

Significance level

Make a decision to either

- reject the null hypothesis or
- fail to reject the null hypothesis.

Significance level

Make a decision to either

- reject the null hypothesis or
- fail to reject the null hypothesis.

Select a significance level a

Significance level

Make a decision to either

- reject the null hypothesis or
- fail to reject the null hypothesis.

Select a significance level a and

- reject if p-value $< a$

Significance level

Make a decision to either

- reject the null hypothesis or
- fail to reject the null hypothesis.

Select a significance level a and

- reject if p-value $< a$ otherwise
- fail to reject.

Decisions

Decision	Truth	
	H_0 true	H_0 not true
reject H_0	type I error	correct
fail to reject H_0	correct	type II error

Decisions

Decision	Truth	
	H_0 true	H_0 not true
reject H_0	type I error	correct
fail to reject H_0	correct	type II error

Then

significance level a is $P(\text{reject } H_0 | H_0 \text{ true})$

Decisions

Decision	Truth	
	H_0 true	H_0 not true
reject H_0	type I error	correct
fail to reject H_0	correct	type II error

Then

significance level a is $P(\text{reject } H_0 | H_0 \text{ true})$

and

power is $P(\text{reject } H_0 | H_0 \text{ not true})$.

Interpretation

The null hypothesis is a model.

Interpretation

The null hypothesis is a model. For example,

$$H_0 : Y \sim Bin(n, \theta_0)$$

if we reject H_0, then we are saying the data are incompatible with this model.

Interpretation

The null hypothesis is a model. For example,

$$H_0 : Y \sim Bin(n, \theta_0)$$

if we reject H_0, then we are saying the data are incompatible with this model.

Recall that $Y = \sum_{i=1}^{n} X_i$ for $X_i \stackrel{ind}{\sim} Ber(\theta)$.

Interpretation

The null hypothesis is a model. For example,

$$H_0 : Y \sim Bin(n, \theta_0)$$

if we reject H_0, then we are saying the data are incompatible with this model.

Recall that $Y = \sum_{i=1}^{n} X_i$ for $X_i \stackrel{ind}{\sim} Ber(\theta)$.

So, possibly

- the X_i are not independent

Interpretation

The null hypothesis is a model. For example,

$$H_0 : Y \sim Bin(n, \theta_0)$$

if we reject H_0, then we are saying the data are incompatible with this model.

Recall that $Y = \sum_{i=1}^{n} X_i$ for $X_i \overset{ind}{\sim} Ber(\theta)$.

So, possibly

- the X_i are not independent or
- they don't have a common θ

Interpretation

The null hypothesis is a model. For example,

$$H_0 : Y \sim Bin(n, \theta_0)$$

if we reject H_0, then we are saying the data are incompatible with this model.

Recall that $Y = \sum_{i=1}^{n} X_i$ for $X_i \stackrel{ind}{\sim} Ber(\theta)$.

So, possibly

- the X_i are not independent or
- they don't have a common θ or
- $\theta \neq \theta_0$

Interpretation

The null hypothesis is a model. For example,

$$H_0 : Y \sim Bin(n, \theta_0)$$

if we reject H_0, then we are saying the data are incompatible with this model.

Recall that $Y = \sum_{i=1}^n X_i$ for $X_i \overset{ind}{\sim} Ber(\theta)$.

So, possibly

- the X_i are not independent or
- they don't have a common θ or
- $\theta \neq \theta_0$ or
- you just got unlucky.

Interpretation

The null hypothesis is a model. For example,

$$H_0 : Y \sim Bin(n, \theta_0)$$

if we reject H_0, then we are saying the data are incompatible with this model.

Recall that $Y = \sum_{i=1}^n X_i$ for $X_i \stackrel{ind}{\sim} Ber(\theta)$.

So, possibly

- the X_i are not independent or
- they don't have a common θ or
- $\theta \neq \theta_0$ or
- you just got unlucky.

If we fail to reject H_0,

Interpretation

The null hypothesis is a model. For example,

$$H_0 : Y \sim Bin(n, \theta_0)$$

if we reject H_0, then we are saying the data are incompatible with this model.

Recall that $Y = \sum_{i=1}^{n} X_i$ for $X_i \stackrel{ind}{\sim} Ber(\theta)$.

So, possibly

- the X_i are not independent or
- they don't have a common θ or
- $\theta \neq \theta_0$ or
- you just got unlucky.

If we fail to reject H_0, insufficient evidence to say that the data are incompatible with this model.

Die tossing example

You are playing a game of Dragonwood and a friend rolled a four 3 times in 6 attempts.

Die tossing example

You are playing a game of Dragonwood and a friend rolled a four 3 times in 6 attempts. Did your friend (somehow) increase the probability of rolling a 4?

Die tossing example

You are playing a game of Dragonwood and a friend rolled a four 3 times in 6 attempts. Did your friend (somehow) increase the probability of rolling a 4?

Let Y be the number of fours rolled and assume $Y \sim Bin(6, \theta)$.

Die tossing example

You are playing a game of Dragonwood and a friend rolled a four 3 times in 6 attempts. Did your friend (somehow) increase the probability of rolling a 4?

Let Y be the number of fours rolled and assume $Y \sim Bin(6, \theta)$. You observed $y = 3$

Die tossing example

You are playing a game of Dragonwood and a friend rolled a four 3 times in 6 attempts. Did your friend (somehow) increase the probability of rolling a 4?

Let Y be the number of fours rolled and assume $Y \sim Bin(6, \theta)$. You observed $y = 3$ and are testing

$$H_0 : \theta = \frac{1}{6}$$

Die tossing example

You are playing a game of Dragonwood and a friend rolled a four 3 times in 6 attempts. Did your friend (somehow) increase the probability of rolling a 4?

Let Y be the number of fours rolled and assume $Y \sim Bin(6, \theta)$. You observed $y = 3$ and are testing

$$H_0 : \theta = \frac{1}{6} \qquad \text{versus} \qquad H_A : \theta > \frac{1}{6}.$$

Die tossing example

You are playing a game of Dragonwood and a friend rolled a four 3 times in 6 attempts. Did your friend (somehow) increase the probability of rolling a 4?

Let Y be the number of fours rolled and assume $Y \sim Bin(6, \theta)$. You observed $y = 3$ and are testing

$$H_0 : \theta = \frac{1}{6} \qquad \text{versus} \qquad H_A : \theta > \frac{1}{6}.$$

```
binom.test(3, 6, p = 1/6, alternative = "greater")$p.value

[1] 0.06228567
```

Die tossing example

You are playing a game of Dragonwood and a friend rolled a four 3 times in 6 attempts. Did your friend (somehow) increase the probability of rolling a 4?

Let Y be the number of fours rolled and assume $Y \sim Bin(6, \theta)$. You observed $y = 3$ and are testing

$$H_0 : \theta = \frac{1}{6} \qquad \text{versus} \qquad H_A : \theta > \frac{1}{6}.$$

```
binom.test(3, 6, p = 1/6, alternative = "greater")$p.value
```

```
[1] 0.06228567
```

With a signficance level of $a = 0.05$, you fail to reject the null hypothesis.

Summary

- Hypothesis tests:

$$H_0 : \theta = \theta_0 \qquad \text{versus} \qquad H_A : \theta \neq \theta_0$$

Summary

- Hypothesis tests:

$$H_0 : \theta = \theta_0 \qquad \text{versus} \qquad H_A : \theta \neq \theta_0$$

- Use p-values to determine whether to
  - reject the null hypothesis or
  - fail to reject the null hypothesis.

Summary

- Hypothesis tests:

$$H_0 : \theta = \theta_0 \qquad \text{versus} \qquad H_A : \theta \neq \theta_0$$

- Use p-values to determine whether to
  - reject the null hypothesis or
  - fail to reject the null hypothesis.
- More assessment is required to determine if other model assumptions hold.