Binomial Test

Math 122

Hypothesis Testing

Tack Flips

Claim: When a tack is flipped, the probability that it lands pointing up is 0.50.

Test the Claim: Flip a tack many times and count the number of times the tack lands pointing up.

$\# ups = 7+8+8+6+7+7+5+5+7+4+6+5+8+7+6+6+5+8+9+6+6+4+5+9 = 154$

If the claim is correct, is this result unusual?

$X = \#$ ups in 240 flips

Assume $p = 0.5$

$P(X \geq 154) = 0.0000067$ ← Unusual

Is the claim correct?

Probably not correct.

NCAA Basketball

- Computer models routinely predict 37-47 of the 64 tournament games correctly.

- Would correctly predicting at least 37 of 64 games be unusual if each prediction is simply a guess?

x = # correct guesses in 64

Assume $p = 1/2$

$P(x \geq 37) = 0.13$

Not Unusual

- What does this say about a model that is correct 37 of 64 times?

Model might be random.

In 2010, GA Tech computer science professors correctly predicted the outcomes of 51 of 64 games. What is the probability of this happening randomly?

$p = 1/2 \quad n = 64$

$P(X \geq 51) = 9.4\ldots e^{-7} = 0.00000000094$

Extremely Unusual

What does this suggest about the GA Tech model? This Model does better than guessing

These are examples of

Hypothesis Tests

Ingredients of a Hypothesis Test

- Parameters/Variables (p, n, x, O)
- Claim
- Working Assumption (Null Hypothesis) H_0
- Alternative Hypothesis H_1
- P-Value (Measure of Consistency)
  - Probability of results at least this extreme ← Assuming H_0
- Significance Level (Specifying unusualness)
- Conclusion

If P is small then H_0 is probably false

If P is big, H_0 maybe true

Parameters/Variables

- p – a proportion or probability
- n – sample size or number of trials
- x – number of successes out of n trials
- O – observed number of successes

Claim

$$p = a \quad p \neq a$$
$$p \leq a \quad p > a$$
$$p \geq a \quad p < a$$

$p = 1/2$

Null Hypothesis (Working Assumption) Denoted H_0

$$p = a$$
$$p \leq a$$
$$p \geq a$$

Alternative Hypothesis
Denoted H_1

- Opposite of H_0

The opposite of this	is this
$p = a$	$p \neq a$
$p \leq a$	$p > a$
$p \geq a$	$p < a$

H_0 H_1

- At the end of the test, we

  either believe H_1 or H_0

P-value

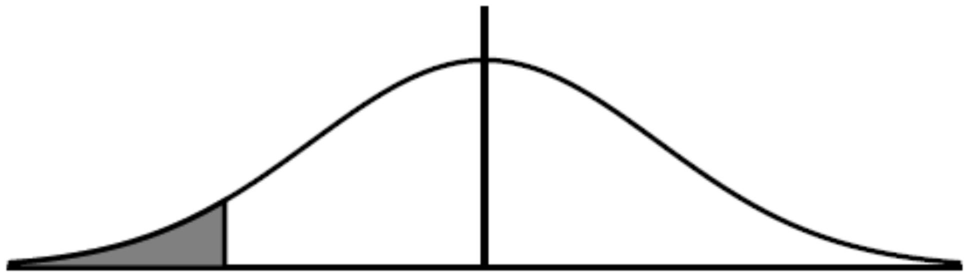
- This is the probability of getting results at least as extreme as the observed value.
- This is a measure of consistency between our observations and H_0.
- If P is **large**, then our observations are consistent with H_0.
- If P is **small**, the our observations are inconsistent with H_0, and we reject H_0.

$P \approx$ Probability that H_0 is true

How we find P-values depends on H_1

If H_1 is p<a

- We will reject H_O in favor of H_1 if x is *small*

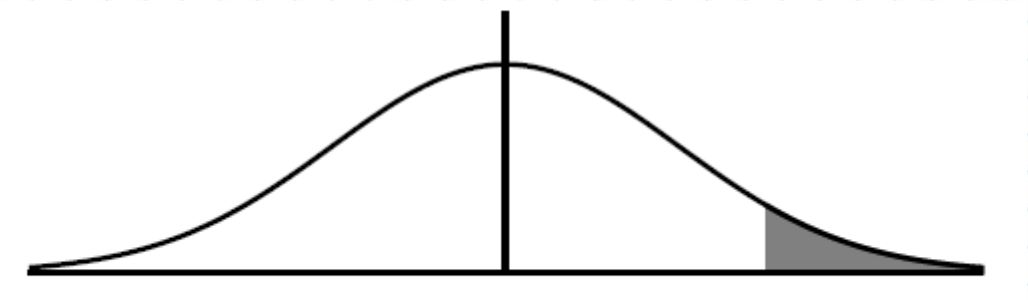


$$P = P(x \leq O) = binomcdf(n, p, O)$$

- This is a *left-tailed* test.

If H_1 is p>a

- We will reject H_O in favor of H_1 if x is *big*

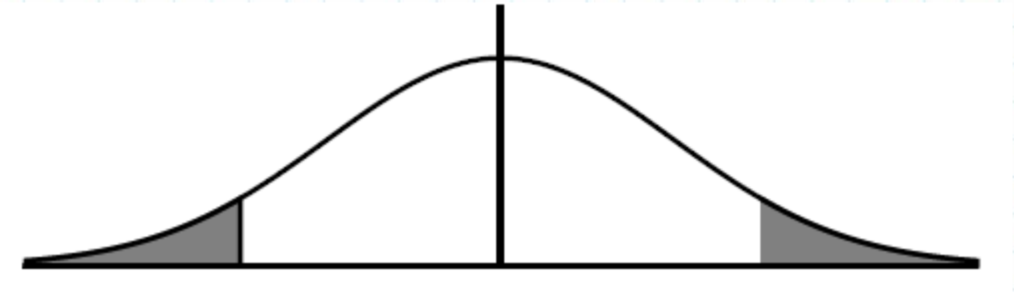


$$P = P(x \geq O) = 1 - binomcdf(n, p, O - 1)$$

- This is a *right-tailed* test.

If H_1 is p≠a

- We will reject H_O in favor of H_1 if x is *small* or *big*



$$P = P(x \leq np - d) + P(x \geq np + d)$$

- This is a *two-tailed* test.

Significance Level

- If P is **large**, the observations are consistent with H_0.
- If P is **small**, the observations are inconsistent with H_0.

- What does LARGE or SMALL mean?

Significance Level
Denoted α

alpha

- Select a significance level α.
- If $P > \alpha$, then P is large.
- If $P \leq \alpha$, then P is small
- If α is not given, use $\alpha = 5\%$ (aka the 5% Rule)

Formal Conclusion

- If $P \leq \alpha$, then our observations are inconsistent with H_0.

  We **REJECT H_0** and Support H_1.

- If $P > \alpha$, then our observations are consistent with H_0.

  We **FAIL TO REJCT H_0** and do not support H_1.

Final Conclusion

- If your **claim is H_0** select from:
  - There is enough sample evidence to **reject** the claim.
  - There is **not** enough sample evidence to **reject** the claim.
- If your **claim is H_1** select from:
  - There is enough sample evidence to **support** the claim.
  - There is **not** enough sample evidence to **support** the claim.

In a sample of 100 of a certain type of parrot, 59 were found to have red wings while 41 had green wings. Use this information to test the claim that **most of these birds have red wings**.

More Than 1/2

- p= proportion of these birds w/ red wings
- n=100
- x=# of birds w/ red wings in 100 (observed=59)
- Claim: $p > 1/2$ Opposite: $p \leq 1/2$
- $H_0: p \leq 1/2$
- $H_1: p > 1/2$ - claim
- P-value= .0443 unusual
- Formal Conclusion: Reject Ho / Support H_1
- Conclusion:

There is enough evidence to support the claim

In a clinical study of the drug Allegra, 31 patients out of 283 experienced headaches after taking Allegra. Test the claim that the rate of headaches among Allegra users is higher than 7.2% (which was the rate of headaches among those patients in the study receiving a placebo).

x

n

$\hat{p}$

p = proportion of Allega patients w/ headaches.

Claim: $p > 0.072$ Opposite: $p \leq 0.072$ — p

H_1 H_0

P = 0.0135 ← Probability of data as extreme as ours if H_0 is true.

Since P < 5% Reject H_0 / Support H_1

The sample evidence supports the claim that more than 7.2 % of Allegra users experience headaches

In a clinical trial of Xanax, 79 of 565 patients experienced depression after taking Xanax. Test the claim that the proportion of patients taking Xanax who experience depression is lower than 18% (which was the percentage of patients who experienced depression after taking a placebo).

$p =$

Claim: $p < .18$ H_1

Opposite: $p \geq .18$ H_0

$P = .006 \leftarrow$ Small

Formal Conclusion: Reject H_0 / Support H_1

Final Conclusion:

There is enough sample evidence to support the Clai x n

In a clinical trial of Xanax, 13 of 565 patients experienced weight loss after taking Xanax. Test the claim that the proportion of patients taking Xanax who experience weight loss ~~depression~~ is 3% (which was the percentage of patients who experienced weight loss after taking a placebo).

p = proportion of Xanax users who lose weight

Claim: $p = .03$ (H_0) Opposite: $p \neq 0.03$ (H_1) p

$P = 0.39$ – Probability of data as extreme as ours if H_0 is true.

Not unusual ($P > 5\%$) Do Not Reject H_0 / Do No Support H_1

There is not enough evidence to reject the claim that the prop. of Xanax users who lose weight is 3%.

The CEO of a major utility company claims that at least 80 percent of the company's 1,000,000 customers are very satisfied. To test the claim, 100 customers are surveyed using simple random sampling. Of those sampled, 76 were very satisfied.

n

x

p = prop. of customers who are very satisfied

Claim: $p \geq 0.80$

H_0

Opposite: $p < .80$

H_1

p

$P = 0.189$

Formal Conclusion: Do Not Reject H_0 / Do Not Support H_1

Final Conclusion:

Do not reject claim.