Intro to Hypothesis Testing

Let's play a game: I flip a coin

Heads: Pay you \$1.50

Tails: You pay me \$1

$E[X] = .5(\$1.50) + .5(-\$1) = \$0.25$

If coin was fair, $\hat{p} \approx 50\%$ distribution of $\hat{p}$

What p was less than 50%

When we sample $\hat{p}$, it comes from the true distribution.

Hypothesis Testing allows us to answer a specific question about a parameter.

A report found that 1/4 Americans who had recent (≤12 mo) police interactions "not a positive interaction" over "positive experience"

Proportion of gen population reporting negative interactions

$p_G = 25\%$

Question: Is the proportion of Black Americans reporting neg interactions higher than gen population?

We start with two competing hypotheses

Null hypothesis

Alternative hypothesis

Null hypothesis

The thing we will assume and attempt to disprove

The prop of Black Americans reporting neg interactions is the same as gen pop^n.

$H_0: p_{BA} = p_G = \underline{25\%}$

Alternative hypothesis

The thing that we want to show.

The prop of Black American reporting neg interactions is higher.

$H_a: p_{BA} > p_G$

If we take a sample of 72 Black Americans, we find that $\frac{26}{72} = 36.\bar{1}\%$ report negative interactions,

How likely is this result or something more extreme if H_0 is true ($p_{BA} = .25$)?

This is the p-value

Expect $\hat{p} \sim N\left(.25, \sqrt{\frac{.25(1-.25)}{72}}\right)$

How unlikely is this result?

Normal Approx

Use continuity correction

$\hat{p} = \frac{26}{72}$, we will use $\hat{p} = \frac{26-.5}{72}$

Exact (Binomial)

$P(X \geq 26 \mid n = 72, p = .25)$

$= \underline{2.364\%}$

$\hat{p} = \frac{26-.5}{72}$ $= \underline{2.364}$

$$Z = \frac{\hat{p} - p}{\underbrace{\sqrt{\frac{p(1-p)}{n}}}_{SE}}$$

$$Z = \frac{\frac{25.5}{72} - .25}{\sqrt{\frac{.25 \times .75}{72}}} = 2.0412$$

$P(Z \geq 2.0412) = 2.061\%$

p-value $= 2.36\%$ If H_0 is true, the results we see should only happen in about 1/40 samples.

Conclusion

If the results we see are likely:

Then we fail to reject H_0.

If results are unlikely:

Then we reject H_0 and conclude the alternative.

What's "unlikely"?

If H_0 is true, we don't want to reject it.

If H_0 is false, we want to reject it.

Typically we consider results "significant" if the p-value is below some predefined value (α, $\alpha = 5\%$)

" is below some predefined value (α),

If p-value $< \alpha$, we reject H_0 and conclude H_α

If p-value $> \alpha$, we fail to reject H_0.

Based on this study, we reject the hypothesis that Black Americans report negative interactions at rate equal to gen pop^n and conclude that the proportion of neg interactions is $p_{BA} > 25\%$

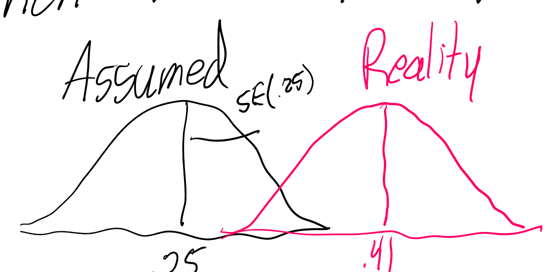
How do we know these results aren't due to chance.

We don't.

So how does a hypothesis test work?

Truth is, approx 41% of Black Americans reported negative interactions

When we sampled $\hat{p}$.



If we repeat this process, how often do we reject H_0.

To find this we need to know when we reject

Normal

$P(z \ldots) \leq .05$

Binomial

$P(X \geq x \mid n = 72, p = .25) \leq .05$

$P(Z \geq z) \leq .05$

$P(X \geq x \mid n = 72, p = .25) \leq .05$

We reject if $X \geq 25$

$p + Z \cdot SE$

$SE = \sqrt{\frac{p(1-p)}{n}}$ — p assumed

We reject if $\hat{p} \geq 33.4\%$

If $p = .41$, what is the prob we reject?

Normal

$SE = \sqrt{\frac{.41(1-.41)}{72}}$

$Z = \frac{.334 - .41}{SE} = -1.312$

$P(Z > -1.312 \mid p = .41) = \underline{90.8\%}$

Binomial

$P(X \geq 25 \mid n = 72, p = .41)$

$= \underline{88.6\%}$

If $p_{BA} = .25$

95% of samples will fail to reject H_0. ✓

5% of samples will reject H_0. X Type I error

False positive

If $p_{BA} = .41$

88.6% of samples will reject H_0 ✓

11.4% of samples will fail to reject X Type II error

11.4% of samples will fail to reject X Type II error

False negative