Week 6: Hypothesis Testing

2. p-value

Stat 140 - 04

Mount Holyoke College

Dr. Shan Shan

Slides posted at *http://sshanshans.github.io/stat140*

http://www.youtube.com/watch?v=3ESGpRUMj9E

In 2008, Paul the Octopus predicted 8 World Cup games, and predicted them all correctly.

Is this evidence that Paul's chance of guessing correctly, p, is really greater than 50%?

Poll question

What are the null and alternative hypotheses?

a. $H_0 : p \neq 0.5, H_a : p = 0.5$
b. $H_0 : p = 0.5, H_a : p \neq 0.5$
c. $H_0 : p = 0.5, H_a : p > 0.5$
d. $H_0 : p > 0.5, H_a : p = 0.5$

How unusual is it to see a sample statistic as extreme as that observed, if H_0 is true?

- If it is very unusual, we have statistically significant evidence against the null hypothesis
- **Today:** How do we measure how unusual a sample statistic is, if H_0 is true?

SIMULATE what would happen if H_0 were true!

To see if a statistic provides evidence against H_0, we need to see what kind of sample statistics we would observe, just by random chance, if H_0 were true!

- We need to know what kinds of statistics we would observe just by random chance, if the null hypothesis were true
- How could we figure this out???

Simulate many samples of size $n = 8$ *with* $p = 0.5$

Flip a fair coin

Each coin flip = a guess between two teams

- Head: correct
- Tail: incorrect

Let's flip together

Flip a coin 8 times, count the number of heads, and calculate the sample proportion of heads

- Did you get all 8 heads (correct)?
- How extreme is Paul's sample proportion of 1?

Poll question

Based on your simulation results, for a sample size of n $= 8$, do you think $\hat{p} = 1$ is statistically significant?

a. Yes
b. No

Poll question

Based on your simulation results, for a sample size of n $= 8$, do you think $\hat{p} = 1$ is statistically significant?

a. Yes
b. No

For a better randomization distribution, we need many more simulations!

A randomization distribution is a collection of statistics from samples simulated assuming the null hypothesis is true

The randomization distribution shows what types of statistics would be observed, just by random chance, if the null hypothesis were true

Center

A randomization distribution simulates samples assuming the null hypothesis is true, so it is centered at the value of the parameter given in the null hypothesis.

Spread

The spread of the distribution measures how much the statistic varies from sample to sample under the null hypothesis

Shape

For most of the statistics we consider, if the sample size is large enough the randomization distribution will be symmetric and bell-shaped.

In a hypothesis test for

$$H_0 : \mu = 12 \text{ vs } H_a : \mu < 12$$

we have a sample with $n = 45$ and $\bar{x} = 10.2$.

Poll question

What do we require about the method to produce randomization samples?

a. $\mu = 12$
b. $\mu < 12$
c. $\hat{x} = 12$

In a hypothesis test for

$$H_0 : \mu = 12 \text{ vs } H_a : \mu < 12$$

we have a sample with $n = 45$ and $\bar{x} = 10.2$.

Poll question

What do we require about the method to produce randomization samples?

a. $\mu = 12$
b. $\mu < 12$
c. $\hat{x} = 12$

We need to generate randomization samples assuming the null hypothesis is true.

In a hypothesis test for

$$H_0 : \mu = 12 \text{ vs } H_a : \mu < 12$$

we have a sample with $n = 45$ and $\bar{x} = 10.2$.

Poll question

Where will the randomization distribution be centered?

a. 10.2
b. 12
c. 45

In a hypothesis test for

$$H_0 : \mu = 12 \text{ vs } H_a : \mu < 12$$

we have a sample with $n = 45$ and $\bar{x} = 10.2$.

Poll question

Where will the randomization distribution be centered?

a. 10.2
b. 12
c. 45

Randomization distributions are always centered around the null hypothesized value.

In a hypothesis test for

$$H_0 : \mu = 12 \text{ vs } H_a : \mu < 12$$

we have a sample with $n = 45$ and $\bar{x} = 10.2$.

Poll question

What will we look for on the randomization distribution?

a. How extreme 10.2 is
b. How extreme 12 is
c. How extreme 45 is
d. How many randomization samples we collected
e. The center point

In a hypothesis test for

$$H_0 : \mu = 12 \text{ vs } H_a : \mu < 12$$

we have a sample with $n = 45$ and $\bar{x} = 10.2$.

Poll question

What will we look for on the randomization distribution?

a. How extreme 10.2 is
b. How extreme 12 is
c. How extreme 45 is
d. How many randomization samples we collected
e. The center point

We want to see how extreme the observed statistic is.

In a hypothesis test for

$$H_0 : \mu_1 = \mu_2 \text{ vs } H_a : \mu_1 > \mu_2$$

we have a sample with $\bar{x}_1 = 24$ and $\bar{x}_2 = 21$.

Answer the previous questions

- What do we require about the method to produce randomization samples?
- Where will the randomization distribution be centered?
- What will we look for on the randomization distribution?

We need a way to quantify evidence against the null...

The p-value is the chance of obtaining a sample statistic as extreme (or more extreme) than the observed sample statistic, if the null hypothesis is true

How do you compute the p-value?

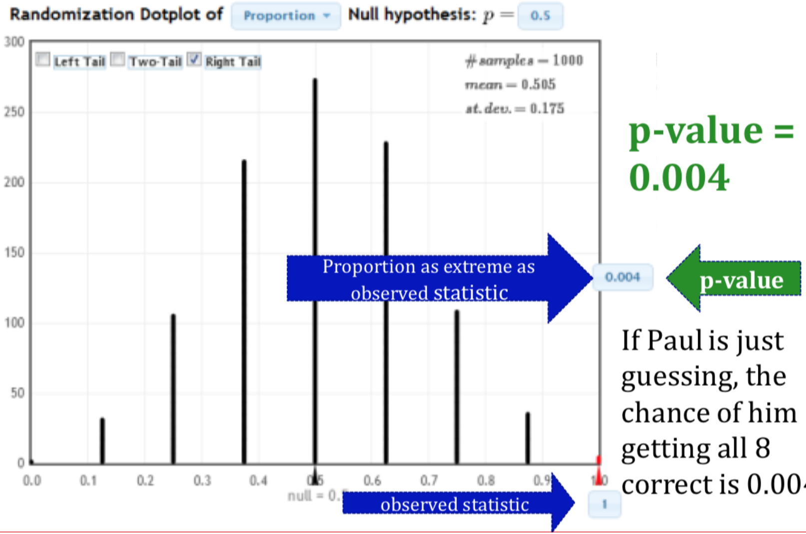
The p-value can be calculated as the proportion of statistics in a randomization distribution that are as extreme (or more extreme) than the observed sample statistic

Paul the Octopus

- The p-value is the chance of getting all 8 out of 8 guesses correct, if $p = 0.5$
- What proportion of statistics in the randomization distribution are as extreme as $\hat{p} = 1$

Let's find it together

`http://www.lock5stat.com/StatKey/randomization_1_quant_1_cat/randomization_1_quant_1_cat.html`



p-value = 0.004

If Paul is just guessing, the chance of him getting all 8 correct is 0.004.

For a sufficiently large sample size, the distribution of sample proportion or sample mean is normal.

We model the randomization distribution with

$$\mathcal{N}(\text{Parameter in } H_0, \text{Standard Error})$$

Parameter	**Distribution**	**Standard Error (CI)**	**Standard Error (Test)**[1]
Proportion	Normal	$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$\sqrt{\frac{p_0(1-p_0)}{n}}$, for $H_0 : p = p_0$
Mean	$t,\ df = n-1$	$\frac{s}{\sqrt{n}}$	$\frac{s}{\sqrt{n}}$
Difference in Proportions	Normal	$\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$	$\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1}+\frac{\hat{p}(1-\hat{p})}{n_2}}$, $\hat{p} = \frac{\hat{p}_1 n_1 + \hat{p}_2 n_2}{n_1+n_2}$
Difference in Means	$t,\ df = \min(n_1, n_2) - 1$	$\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$	$\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$
Correlation	$t,\ df = n-2$	$\sqrt{\frac{1-r^2}{n-2}}$	$\sqrt{\frac{1}{n-2}}$

A large city's DMV claimed that 80% of candidates pass driving tests, but a survey of 90 randomly selected local teens who had taken the test found only 61 who passed.

Does this finding suggest that the passing rate for teenagers is lower than the DMV reported?

Set up the hypotheses Let p denote the passing rate of the driving test for teenagers

$$H_0 : p = 0.8$$
$$H_a : p < 0.8$$

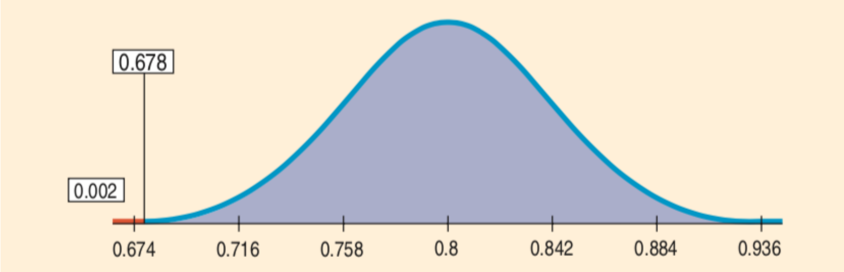
If in fact the null hypothesis is true, $\hat{p}$ is distributed nearly normally with mean $p = 0.8$ and standard error given by $\frac{\sqrt{p_0(1-p_0)}}{n}$.

We would like to find out how likely it is to observe a sample proportion at least as far from the data as our current sample proportion ($61/90 = 0.678$), if in fact the null hypothesis is true.

The standard error is given by

$$\frac{\sqrt{p_0(1-p_0)}}{n} = \frac{\sqrt{0.8(0.2)}}{90} = 0.0402$$

We are interested in finding the area to the left of 0.678 under the curve of the normal model with mean 0.8 and standard deviation 0.0402.


0.678
0.002
0.674
0.716
0.758
0.8
0.842
0.884
0.936

Compute the z-score

$$z = \frac{\hat{p} - p_0}{\mathsf{SE}} = \frac{0.678 - 0.800}{0.042} = -2.90$$

The p-value is then given by

$$P(z < -2.90) = 0.002$$

Are the conditions for inference satisfied?

- Randomization Condition: The 90 teens surveyed were a random sample of local teenage driving candidates.
- 10% Condition (for independence): 90 is fewer than 10% of the teenagers who take driving tests in a large city.
- Success/Failure Condition:
  We expect $n \times p_0 = 90(0.80) = 72$ successes and $n \times (1 - p_0) = 90(0.20) = 18$ failures. Both are at least 10.

ANSWER: The conditions are satisfied, so it's okay to use a Normal model to compute the p-value

- If the p-value is small, then a statistic as extreme as that observed would be unlikely if the null hypothesis were true, providing significant evidence against H_0
- The smaller the p-value, the stronger the evidence against the H_0 and in favor of the H_a

How small?

Randomization distribution

What kinds of statistics would we get, just by random chance, if the null hypothesis were true?

p-value

The chance of obtaining a sample statistic asextreme (or more extreme) than the observed sample statistic, ifthe null hypothesis is true

- What proportion of these statistics are as extreme as our original sample statistic?
- CLT based method