- Read
- Hypotheses
- Q's
- Solve

Inference using the t distribution

McCaig Statistics Group
Jordan Bannister

Example 0.1. You are a brewer at Guinness breweries. You've invented a new brewing technique that you hypothesize will reduce the bitterness of Guinness beer. You know the mean bitterness of the current brewing technique. You've also brewed 10 batches with your new technique and measured their bitterness.

Techniques:
Confidence interval for a sample mean,
1 sided 1-sample t-test

$H_0: \mu_X = 0$

$H_A: \mu_X > 0$

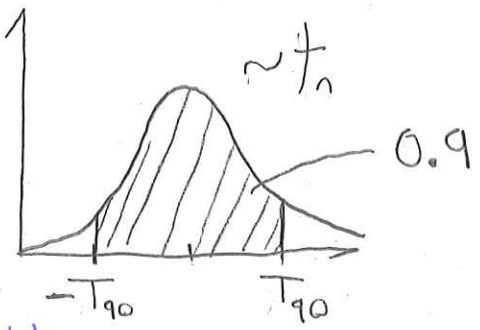
$X = \beta_{old} - \beta_{new}$

What is being compared? ~~sample~~ mean w. fixed value

Assumptions: Parametric stats, random sampling

questions?

$$\frac{\bar{X} - \mu_X}{SE(\bar{X})} \sim t_n$$



- interpretation

$\bar{X} = 5 \quad SE(x) = 3$

$n = 10 \quad T_{90} = 1.83$

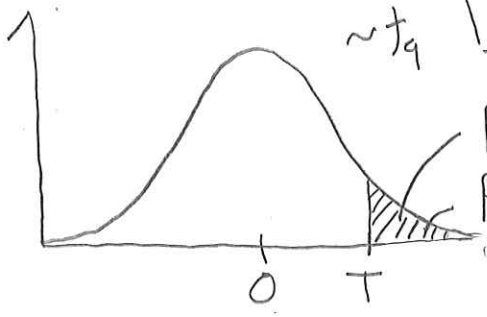
$\mu_X = \bar{X} \pm T_{90} \cdot SE(\bar{X})$

$\mu_X = 5 \pm 5.5$

- 90% of 90% CI's contain the population mean
- CI bounds are random variables!

questions? infer

$$\frac{\bar{X} - \mu_X}{SE(\bar{X})} \sim t_n$$



- assume H_0
- calculate: $P(data \mid H_0)$

$P(t_9 > T \mid H_0)$

$P(\bar{X} > 5 \mid H_0)$

- Directional t-test for directional H_A

$T = \frac{5}{3} \rightarrow p = 0.13$

questions

- p-values are random variables
- $p \neq P(\neg H_0)$ • $p \neq P(H_A)$
- How much does $P(data \mid H_0)$ tell you about $P(H_A)$?

Example 0.2. In 1990, researchers reported the results of a study that examined pairs of monozygotic twins, where one of the twins was schizophrenic and the other was not. They hypothesized that there were structural differences between the brains of the schizophrenic and the normal populations. They took a head MRI of each subject and measured the differences in volume of the left hippocampus between twins.

Technique:
2 sided paired 2 sample t-test

H_0: $\mu_X = 0$

H_A: $\mu_X \neq 0$

$X = V_{norm} - V_{schiz}$

What is being compared? ~~the~~ mean with a fixed value

Assumptions: Parametric Stats, random sampling

$$\frac{\bar{X} - \mu_X}{SE(\bar{X})} \sim t_{14}$$

- Because samples are paired we can treat the difference between pairs as a single ~~parameter~~ population

$\bar{X} = 0.2$ $s = 0.24$ $SE(\bar{x}) = s/\sqrt{n}$

$n = 15$ $T = 3.23$

questions?

$P(t_{14} > T \vee t_{14} < -T) = 0.006$

- non-directional H_A, 2 sided test

Example 0.3. In the late 1900's, biologists Peter and Rosemary Grant studied finches on one of the Galapagos islands in search of evidence supporting Darwin's theory of natural selection. In 1977 there was a severe drought on the island that caused vegetation to whither. The only remaining food source for the finches was a large tough seed that the finches usually didn't eat. The Grants hypothesized that the drought had selected for large beak depth in the finch population. They measured the beak depth of the entire finch population before the drought in 1976, and after the drought in 1978.

Technique:
1 sided unpaired 2-sample t-test

H_0: $D = 0$ • Full population? $D = \mu_{pre} - \mu_{post}$

H_A: $D < 0$ Doesn't matter

• parameters not variables
• different than before

What is being compared? ~~2 sample~~ means

Assumptions: Parametric stats, random, independent

Questions?

$$\frac{(\bar{X}_{pre} - \bar{X}_{post}) - D}{SE(\bar{X}_{pre} - \bar{X}_{post})} \sim t_{[n_{pre} + n_{post} - 2]}$$

$$SE(\bar{X}_{pre} - \bar{X}_{post}) = s_p \sqrt{\frac{1}{n_{pre}} + \frac{1}{n_{post}}}$$

$$s_p = \sqrt{\frac{(n_{pre}-1)s_{pre}^2 + (n_{post}-1)s_{post}^2}{(n_{pre} + n_{post} - 2)}}$$

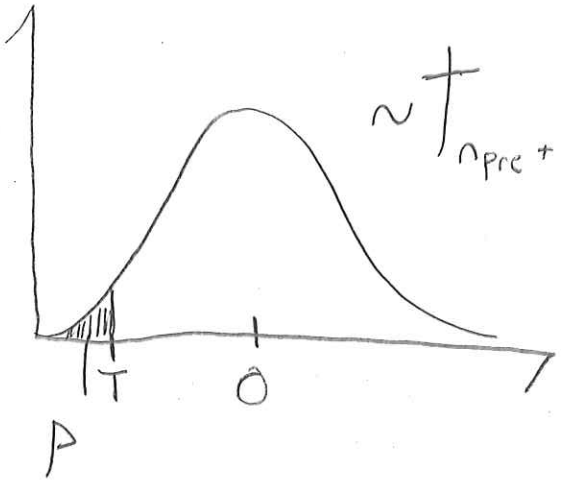
• Because we sampled 2 populations, independently.
• Need to add the variances associated with both sample sets.

Questions?

$T = 4.6$

$P = 0.0000004$

$\sim t_{n_{pre} + n_{post} - 2}$



Q_1: What is a Confidence interval? eg. 95%. 1° for a pop mean

A_1: interval s.t. 95% of the time it will contain μ_x

Q_2: What is a p-value?

A_2: $P(\text{data} \mid H_0)$

Q_3: How do you decide between a directional t-test (1-sided) and a 2 sided test

A_3: Look at H_A

Q_4: What is effect size?

A_4: $\frac{\bar{x}_1 - \bar{x}_2}{s_p}$ or $\frac{\mu_1 - \mu_2}{\sigma}$