Contents

Day 19

One-Sample T-Test Examples

Please see attached PDF for example procedures

Review

H_0: $\mu = \mu_0$

T-Statistic:

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

$$t \sim t(n-1)$$

Example 1 : Budwiser

ABV: Alcohol By Volume

February 2013, a lawsuit claimed actual ABV content in Budwiser was 4.7% but Budwiser advertised ABV of 5%.

Suppose 3 independent labs test Budwiser beer ABV content.

Recall: measurement error is assumed normally distributed

Assume: measurements are unbiased

read these as percents but treat as whole numbers

H_0: $\mu = 5$

H_1: $\mu = 4.7$

$\alpha = 0.05$

Find the critical region:

$$t = \frac{\bar{x} - 5}{\frac{s}{\sqrt{3}}}$$

Under $H_0 \implies t \sim t(2)$

There are two degrees of freedom.

Critical Region: $t \leq -2.92$

If $t_{\text{observed}} \leq -2.92$, accept H_1 claiming the beer is watered down.

If $t_{\text{observed}} > -2.92$, accept H_0 claiming the beer is not watered down.

Lab Measurements:

$\bar{x} = 4.927$

$s = 0.032$

$$t_{\text{observed}} = \frac{4.977 - 5}{\frac{0.032}{\sqrt{3}}} = -1.257$$

```r
qt(0.05, 2, lower.tail = TRUE)
```

Since $-1.257 > -2.92$, we accept $H_0 \implies$ beer is not watered down.

Power Analysis

Is a sample of 3 labs big enough to detect $\mu = 4.7$?

Special Case: Matched Pairs (Paired) t-Test

When it's used: matched-pairs experiment case-control observation study or any other situation where we have a numerical response variable recorded on the same subjects under two different conditions/on pairs of subjects

Commonly seen: before and after treatment studies.

We do inference on the paired differences

E.G) after - before

μ_{d} denotes the population mean of paired differences

$\bar{x_{\text{d}}}$ denotes the sample mean of paired differences

s_{d} denotes the sample standard deviation of paired differences

n denotes the number of paired differences

$$t = \frac{\bar{x_{\text{d}}} - \mu_{\text{d}}}{\frac{s_{\text{d}}}{\sqrt{n}}}$$

Very common:

H_0: $\mu_{\text{d}} = 0$

$\rightarrow$ no difference between conditions

H_a: $\mu_{\text{d}} \neq 0$

$\rightarrow$ difference between conditions

Example : Book Exercise 7.10

Comparing taste of hash browns in deep fryer (oil) vs air fryer (no oil)

Five experts rate taste of hash browns made in each fryer

Expert	Oil Rating	No Oil Rating
1	78	75
2	84	85
3	62	67
4	73	75
5	63	66

Figure 1: Hash Brown Expert Analysis

Is there a difference in taste bwtween oil and no-oil hash browns?

$\mathrm{H_0}$: $\mu_{\mathrm{d}} = 0$

$\mathrm{H_a}$: $\mu_{\mathrm{d}} \neq 0$

No oil - oil

Expert	Oil Rating	No Oil Rating	Diff
1	78	75	-3
2	84	85	1
3	62	67	5
4	73	75	2
5	63	66	3

Figure 2: Hash Brown Expert Analysis (Difference)

$\bar{x_{\mathrm{d}}} = 1.6$

$s_{\mathrm{d}} = 2.966$

$$t_{\mathrm{observed}} = \frac{1.6 - 0}{\frac{2.966}{\sqrt{5}}} = 1.96$$

$$t_{\mathrm{observed}} \sim t(4)$$

P(t $\geq$ 1.96 | $\mathrm{H_0}$ is true) = 0.061

p-value = 2(0.061) = 0.122

Compare the p-value to the significance level, which by default is 0.05

If p-value $\leq$ significance level, we reject $\mathrm{H_0}$ & accept $\mathrm{H_a}$ conclude there is a taste difference.

If p-value $>$ significance level, we fail to reject $\mathrm{H_0}$ cannot claim there is a difference.

If $\mathrm{H_a}$: $\mu_{\mathrm{d}} = 0 \rightarrow$ no-oil taste better than oil taste. This is because our difference were no_oil - oil.

Then: $t_{\mathrm{observed}} = 1.96$

p-value = P(t $\geq$ 1.96 | $\mathrm{H_0}$ is true) = 0.061

If H_a: $\mu_d < 0$, with no oil taste <u>worse</u> than oil taste.

$t = 1.96$

p-value = $P(t \leq 1.96 \mid H_[0] \text{ is true}) = 0.939$