Lecture 12

Inference for One Mean

agenda

Revisions	Discuss the process for submitting midterm 1 revisions
lecture part 1	the t-distribution and the sampling distribution of the sample mean (`ht_one_mean.Rmd`)
R activity	simulating the t-distribution
lecture part 2	inference for one mean
R activity	hypothesis testing for one mean (`ht_one_mean.Rmd`)

Review: parameters vs. statistics

$\hat{p}$

sample proportion

$\hat{p}$ = 10/50
10 students out of 50 own a cat

$\bar{x}$

sample mean

$\bar{x}$ = \$117.25
avg vet visit cost of 159 cat owners

s

sample standard deviation

s = \$23.71
std dev of vet visit cost of 159 cat owners

p

population proportion

p = ?
true proportion of students who own cats

μ

population mean

μ = ?
true average vet visit cost for cat owners

σ

population standard deviation

σ = ?
true std dev vet visit cost for cat owners

The t-distribution

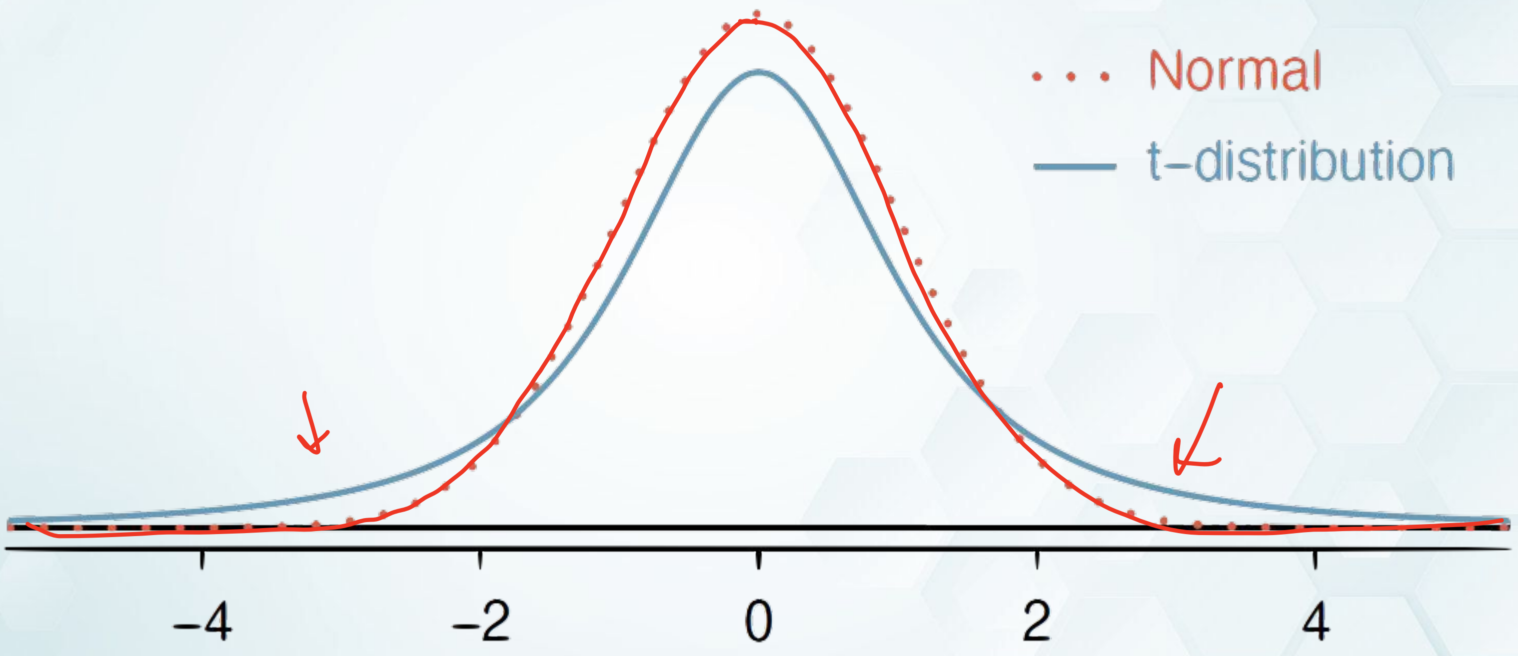


Figure 7.1: Comparison of a t-distribution and a normal distribution.

Properties of the t-distribution

$x_1, x_2, \ldots, x_{n-1}, x_n$ $\quad \bar{x}$

Parameter

df = degrees of freedom, the number of independent observations

Shape

like a standard normal but with fatter tails

Why we use it: it is more *conservative* than the standard normal distribution

The t-distribution

Let $X_1, X_2, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma)$. Define $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

CLT review:

$X_1, X_2, \ldots, X_n \sim f(\ )$, $E[X_i] = \mu = \mu_X$

$SD[X_i] = \sigma = \sigma_X$

- If X_i's are independent
- $n \geq 30$ or $n < 30$, data not strongly skewed

then $\bar{X} \sim N\left(\mu_X = \mu, \sigma_X = \frac{\sigma}{\sqrt{n}}\right)$

* σ almost always unknown
→ we estimate with s

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{df=n-1}$$

$n = 50 \rightarrow t_{49}$

As $df \rightarrow \infty$, $t \xrightarrow{d} Z$

mean of t-dist'n $= 0$

Let's simulate it in R!

Please download and open

`ht_one_mean.Rmd` in the Week 8 module

Hypothesis Testing

Purpose: to test a claim about the population using information collected from a sample

How: finding sufficient evidence against the null

Hypothesis testing for one mean

1) Hypotheses $H_0 : \mu = \mu_0$

$H_A : \mu < / \neq / > \mu_0$

2) Test conditions:
   - Independence
   - Normality

3) Calculate test statistic

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \sim t_{n-1}$$

4) Compute p-value

```
<: pt(t, df)
>: pt(t, df, lower.tail = FALSE)
≠: 2*pt(-abs(t), df)
```

5) Decision and conclusion in context

$$\text{test stats} = \frac{\text{statistic} - \text{null value}}{\text{se}}$$

Is the average guess of Wendy's age older ~~different~~ from her true age?

What is the parameter of interest? Define in symbols and interpret the parameter in context.

μ = average guess of Wendy's age

Hypotheses

null

$$H_0 :\ \mu = \mu_0$$

$H_0 : \mu = 29$

alternative

$<$

$$H_A : \mu \neq \mu_0$$

$>$

$H_A : \mu > 29$

Check conditions

Independence

observational units are randomly selected

✓

Normality

- $n \geq 30$ and no outliers ✓

  OR
- $n < 30$ and data are approximately normal

$$t = \frac{\text{statistic - null value}}{\text{standard error of statistic}}$$

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

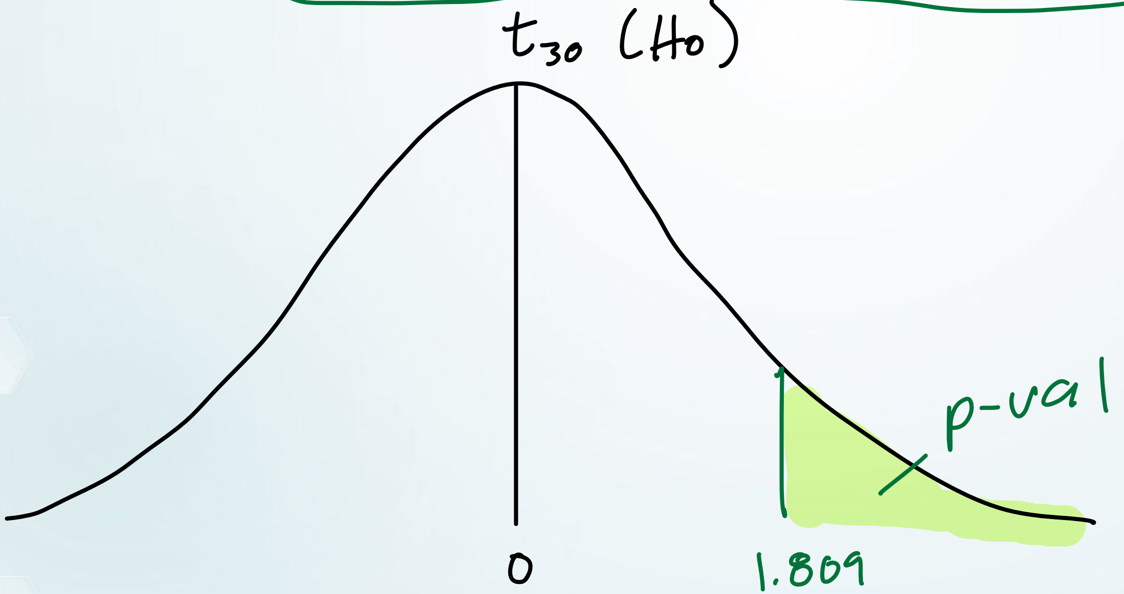
$$= \frac{29.85 - 29}{2.63/\sqrt{31}}$$

$$= \boxed{1.809}$$

Calculate test statistic

Compute p-value

```
<: pt(t, df)  | >: pt(t, df, lower.tail = FALSE) |  ≠: 2*pt(-abs(t), df)
```



$\boxed{\text{p-val} = 0.04}$

Conclusion in context

Decision :

p-val = 0.04 < α = 0.05

Reject the null: We have enough evidence that the true mean _______ is less than/different from/ more than μ_0.

Fail to reject the null: We do not have enough evidence that the true mean _______ is less than/different from/ more than μ_0.

Conclusion in context

We have enough evidence that the avg guess of wendy's age is older than 29.

Confidence intervals

$$\bar{x} \pm \boxed{t^* \frac{s}{\sqrt{n}}}$$

statistic ± crit. value x se

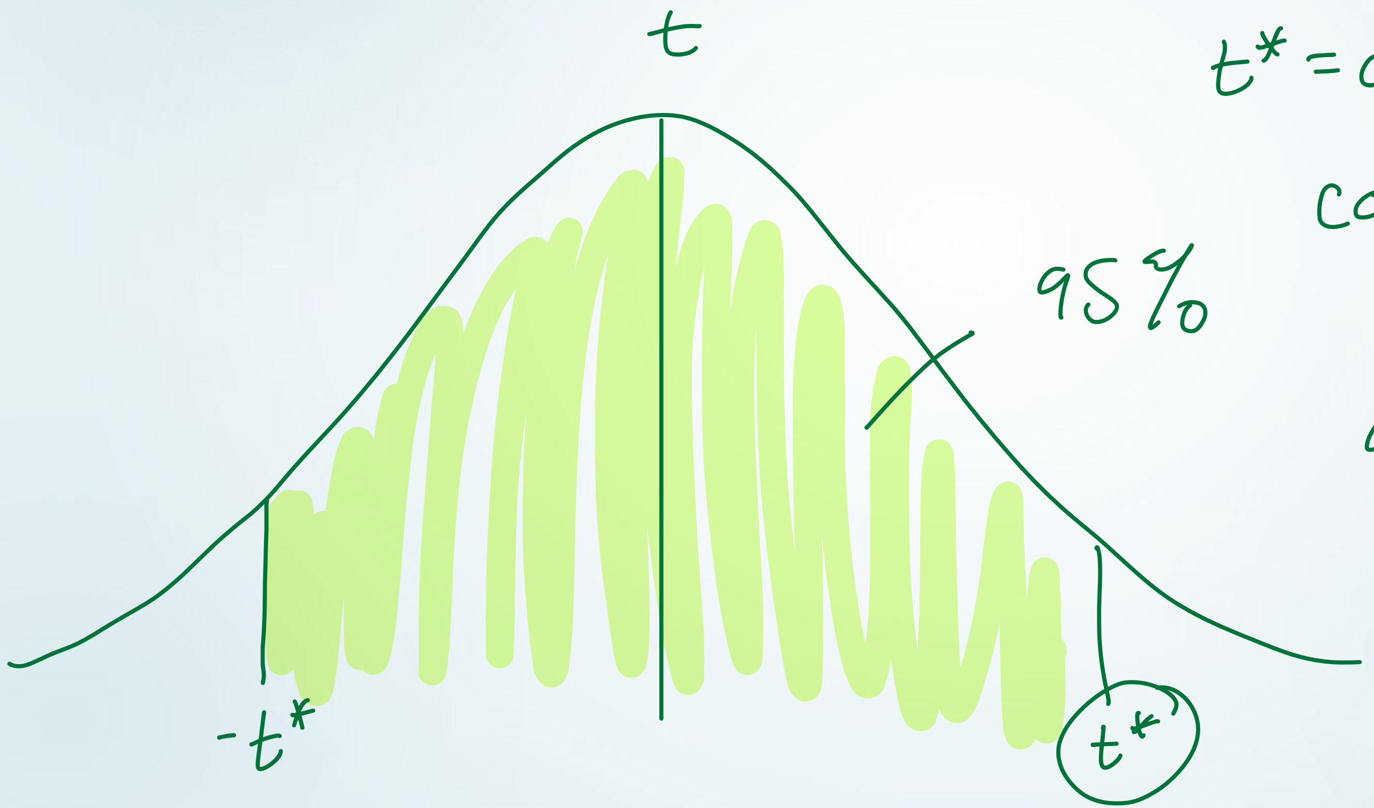
CI: (28.89, 30.82)

$H_0: \mu = 29$

$H_A: \mu \neq 29$

Decision: Fail to reject H_0

Confidence intervals



t^* = crit. value

conf. level = CL = 95%
= .95

qt(0.975, n-1)

$CL = 80\%$

$\downarrow$

$\alpha = 0.2$

$qt(1 - \frac{\alpha}{2}, n-1)$

to R!

back to `ht_one_mean.Rmd`