Intermediate Econometrics

Comparing Two Population Means

Vincenzo Gioia

16/10/2026

Comparing Two Population Means

Setting

- We have two populations and we collect data from a sample of each
- y_{i1}, $i = 1, \ldots, n_1$: observed characteristic in the first sample
- y_{i2}, $i = 1, \ldots, n_2$: measurements on the same characteristic in the second sample

Goal: compare the population means μ_1 and μ_2

- Parameter: $\mu_1 - \mu_2$
- Parameter estimate: $\bar{y}_1 - \bar{y}_2$
- Estimator: $\bar{Y}_1 - \bar{Y}_2$

Comparing Two Population Means: Assumptions

Assumptions

- $Y_{i1} \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $i = 1, \ldots, n_1$
- $Y_{i2} \sim \mathcal{N}(\mu_2, \sigma_2^2)$, $i = 1, \ldots, n_2$
- Y_{i1} and Y_{i2} are independent
- We also assume $\sigma_1^2 = \sigma_2^2 = \sigma^2$

Pooled variance estimator

We only know that $\sigma_1^2 = \sigma_2^2 = \sigma^2$, but σ^2 is an unknown parameter: we need to estimate it.

$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Two-Sided Confidence Interval for $\mu_1 - \mu_2$

Pivotal quantity

Let $\overline{Y}_1$ and $\overline{Y}_2$ be the sample means of the two groups

$$T = \frac{\overline{Y}_1 - \overline{Y}_2 - (\mu_1 - \mu_2)}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

Confidence interval at level $1 - \alpha$

$$IC^{1-\alpha}_{\mu_1-\mu_2} = \left(\bar{y}_1 - \bar{y}_2 \mp t_{n_1+n_2-2;1-\alpha/2} \times s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

that is, lower bound with $-$ and upper bound with $+$:

$$\bar{y}_1 - \bar{y}_2 - t_{n_1+n_2-2;1-\alpha/2}\, s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \quad \bar{y}_1 + \bar{y}_2 + t_{n_1+n_2-2;1-\alpha/2}\, s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Example: Anorexia Study

Cognitive behavioral therapy

- A study uses cognitive behavioral therapy to treat a sample of teenage girls who suffered from anorexia
- There is also a control group that received no treatment
- Researchers analyzed how the mean weight change compared between treatment and control groups
- The girls were randomly assigned to the therapy (Group 1) or to the control group (Group 2)

Let μ_1 and μ_2 denote the mean weight gains (in pounds) for these groups.

Example: Confidence Interval with `t.test()`

```r
Anor <- read.table("http://stat4ds.rwth-aachen.de/data/Anorexia.dat",
                   header=TRUE)
# Get difference post-pre treatment for the group cb and c
cogbehav <- Anor$after[Anor$therapy == "cb"] -
  Anor$before[Anor$therapy == "cb"]
control <- Anor$after[Anor$therapy == "c"] -
  Anor$before[Anor$therapy == "c"]

# Get the 95% CI via t.test function
res <- t.test(cogbehav, control, var.equal = TRUE, conf.level = 0.95)
res$conf.int
```

```
## [1] -0.680137  7.593930
## attr(,"conf.level")
## [1] 0.95
```

Example: Interpretation

Reading the interval

- The mean weight change for the cognitive behavioral therapy could be as much as 0.68 pounds lower or as much as 7.59 pounds higher than the mean weight change for the control group
- The interval includes 0: it is plausible that the population means are identical (we can also see the results underlying the hypothesis test)
- The confidence interval is relatively wide: sample sizes are not large

Example: Confidence Interval by Hand

```
n1 <- length(cogbehav)
n2 <- length(control)
v1 <- var(cogbehav)
v2 <- var(control)

s2 <- ((n1 - 1) * v1 + (n2 - 1) * v2)/(n1 + n2 - 2)

CI <- mean(cogbehav) - mean(control) +
  c(-1,1) * qt(0.975, df = n1 + n2 - 2) *  sqrt(s2 * (1/n1 + 1/n2))
CI
```

```
## [1] -0.680137  7.593930
```

Example: Same result using the linear model

```r
Anor$diff <- Anor$after - Anor$before
lm1 <- lm(diff~therapy, data=Anor[Anor$therapy!="f",])
summary(lm1)
```

```
##
## Call:
## lm(formula = diff ~ therapy, data = Anor[Anor$therapy != "f",
##     ])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -12.107  -4.678  -1.307   3.500  17.893
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -0.450      1.498  -0.300   0.7650
## therapycb      3.457      2.063   1.676   0.0996 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.637 on 53 degrees of freedom
## Multiple R-squared:  0.05033,    Adjusted R-squared:  0.03241
## F-statistic: 2.809 on 1 and 53 DF,  p-value: 0.09963
```

Example: Same result using the linear model

- As we can see, the CI for the mean difference can be obtained using the quantities of a simple linear model regressing the outcome on the binary variable representing the membership to the two groups

```
n1 <- length(cogbehav)
n2 <- length(control)
summary(lm1)$coefficients[2,1] +
  qt(0.975, df = n1 + n2 -2) *
  summary(lm1)$coefficients[2,2]
```

```
## [1] 7.59393
```

```
summary(lm1)$coefficients[2,1] -
  qt(0.975, df = n1 + n2 -2) *
  summary(lm1)$coefficients[2,2]
```

```
## [1] -0.680137
```

Test for the Mean Difference

Anorexia example, again

We have two groups and we want to ask whether the mean weight change between the two groups can be considered equal.

- Does the cognitive behavioral therapy group have mean weight change equal to that of the control group?
- Or does it have a greater mean weight change than the control group?

Under the same assumptions detailed above, we compare μ_1 and μ_2 through the following hypothesis tests (consider $\alpha = 0.05$)

Test for the Mean Difference: Hypotheses

Two-sided two-sample test

$$\begin{cases} H_0 : \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 - \mu_2 \neq 0 \end{cases}$$

One-sided two-sample test

$$\begin{cases} H_0 : \mu_1 - \mu_2 = 0 \quad \text{(equivalently} \quad \mu_1 - \mu_2 \leq 0) \\ H_1 : \mu_1 - \mu_2 > 0 \end{cases}$$

Test statistic

Assuming that $\sigma_1^2 = \sigma_2^2 = \sigma^2$, the test statistic under H_0 has the form

$$T = \frac{\overline{Y}_1 - \overline{Y}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \underset{H_0}{\sim} t_{n_1+n_2-2}$$

Two-Sided Test with `t.test()`

```
res.two <- t.test(cogbehav, control, var.equal = TRUE)
res.two
```

```
##
##  Two Sample t-test
##
## data:  cogbehav and control
## t = 1.676, df = 53, p-value = 0.09963
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.680137  7.593930
## sample estimates:
## mean of x mean of y
##  3.006897 -0.450000
```

One-Sided Test with `t.test()`

```
res.one <- t.test(cogbehav, control, var.equal = TRUE, alternative = "g
res.one
```

```
##
##  Two Sample t-test
##
## data:  cogbehav and control
## t = 1.676, df = 53, p-value = 0.04981
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.003879504         Inf
## sample estimates:
## mean of x mean of y
##  3.006897 -0.450000
```

Test by Hand and using the linear model

```
s <- sqrt(s2 * (1/n1 +1/n2))
testStat <- (mean(cogbehav) - mean(control))/s
## two-sided
2 * pt(testStat, df = n1+n2-2, lower = FALSE )
```

```
## [1] 0.09962901
```

```
summary(lm1)$coefficients[2,4]
```

```
## [1] 0.09962901
```

```
## one-sided
pt(testStat, df = n1+n2-2, lower = FALSE )
```

```
## [1] 0.04981451
```

What are the conclusions?

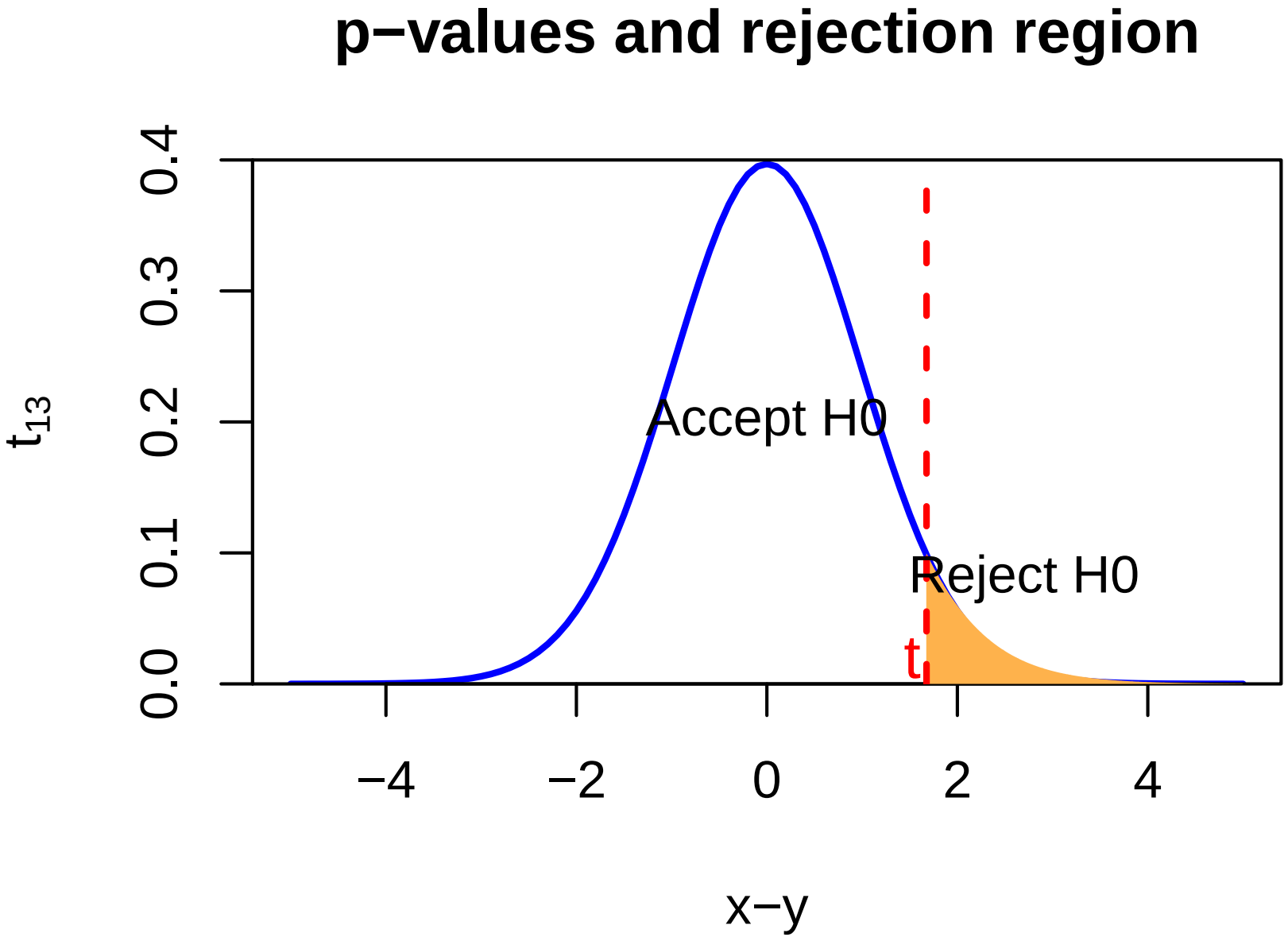
p-Values and Rejection Region

```
library(RColorBrewer)
plotclr <- brewer.pal(6, "YlOrRd")

curve(dt(x, n1 + n2 - 2), xlim = c(-5, 5), ylim = c(0, 0.4),
      main = "p-values and rejection region", col = "blue",
      lwd = 2, xlab = "x-y",  ylab = expression(t[13]),  yaxs="i")
cord.x <- c(qt(0.95, n1 + n2 - 2),seq(qt(0.95, n1 + n2 - 2), 5, 0.01),
cord.y <- c(0, dt(seq(qt(0.95, n1 + n2 - 2), 5, 0.01), 13), 0)
polygon(cord.x, cord.y, col = plotclr[3], border = NA )

abline(v = res.one$statistic, lty = 2, lwd = 2, col = "red")
text(0, 0.2, paste("Accept", expression(H0)))
text(2.7, 0.08, paste("Reject", expression(H0)))
text(as.double(res.one$statistic) - 0.15, 0.02, "t",
     col = "red", cex = 1.2)
```

p-Values and Rejection Region



p-Values and Rejection Region: Remark

Note on the plot

- The observed test statistic varies as a function of the sample
- In this particular case it is so close to the critical value (for α) that you cannot detect a graphical difference
- Different data (with the same sample size) will produce a completely different observed test statistic
- Considering the same α (and the same distribution of the test statistic), the area in orange stays the same