Introductory Statistics for Economics

ECON1013: LAB 3

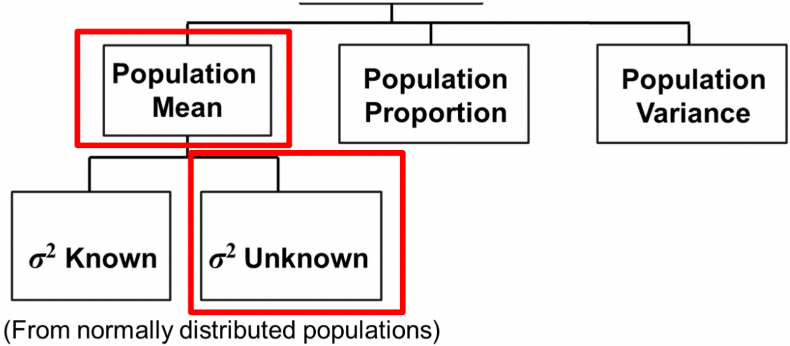
Duong Trinh

University of Glasgow

March 2022

PRELIMINARIES

Estimation and Hypothesis Testing



Exercise 1

Picture the Scenario

- **Objective**: The manager at a plant that bottles drinking water wants to be sure that the process to fill one-gallon bottles is operating properly (1 gallon ≈ 3.785 liters). Currently, the company is testing the volumes of one-gallon bottles. A random sample of 75 one-gallon bottles is tested.

- **Dataset**: `Water.csv`

Questions

(a) Find the 95% confidence interval estimate of the population mean volume.

(b) Without doing calculations, state whether an 80% confidence interval for the population mean would be wider than, narrower than or the same as the answer to part (a).

(c) Without dong calculations, test the null hypothesis $H_0 : \mu = 3.785$ against the $H_1 : \mu \neq 3.785$ at 1% significance level.

Solution

```r
# Load Dataset
Water <- read.csv('Water.csv')
str(Water)
```

```
## 'data.frame':    75 obs. of  1 variable:
##  $ Weights: num  3.93 3.78 3.98 3.82 3.77 3.94 3.76 4.11 3.78
```

```r
# Volumes of bottles in the sample
x <- Water$Weights
```

(a) Find the 95% confidence interval estimate of the population mean volume.

The $100(1-\alpha)\%$ confidence interval for the population mean (when population variance is unknown) is given by:

$$LB = \bar{x} - t_{n-1,\frac{\alpha}{2}} \times \frac{s}{\sqrt{n}} \quad \text{and} \quad UB = \bar{x} + t_{n-1,\frac{\alpha}{2}} \times \frac{s}{\sqrt{n}}$$

where $t_{n-1,\frac{\alpha}{2}}$ is the critical value of the t distribution with $n-1$ degrees of freedom satisfying:

$$P\left(\frac{\bar{X}-\mu}{s/\sqrt{n}} > t_{n-1,\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$$

(a) Find the 95% confidence interval estimate of the population mean volume.

```
# Compute the level of significance
alpha <- 1 - 0.95
# Compute the sample size
n <- length(x)
# Compute the critical value t_c from the t-distribution with n-
t_c <- qt(alpha/2, n-1, lower.tail = FALSE)
# Compute the lower bound (lb) and the upper bound (ub)
lb <-  mean(x) - t_c*sd(x)/sqrt(n)
ub <-  mean(x) + t_c*sd(x)/sqrt(n)
# Compute the confidence interval manually
c(lb,ub)
```

```
## [1] 3.784305 3.831428
```

(a) Find the 95% confidence interval estimate of the population mean volume.

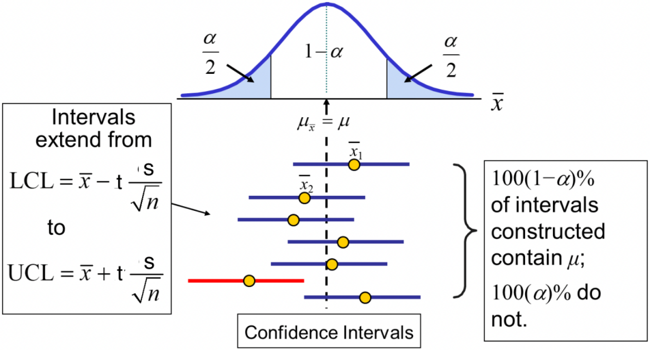
```r
# Alternative: Use the command in R
t.test(x)
```

```
##
##  One Sample t-test
##
## data:  x
## t = 322.02, df = 74, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  3.784305 3.831428
## sample estimates:
## mean of x
##  3.807867
```

(a) Find the 95% confidence interval estimate of the population mean volume.

(!) Interpretation

- We are 95% confident that the true mean volume is between 3.7843 and 3.8314 kg
- Although the true mean may or may not be in this interval, 95% of intervals formed in this manner will contain the true mean



(b) Whether an 80% confidence interval for the population mean would be wider than, narrower than or the same as the answer to part (1).

Length of the $100(1-\alpha)\%$ confidence interval is:

$$UB - LB = 2 \times t_{n-1,\frac{\alpha}{2}} \times \frac{s}{\sqrt{n}}$$

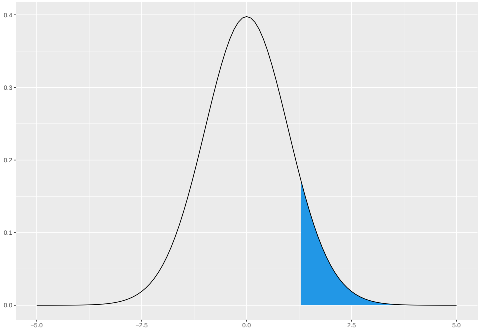


Figure 2: Student t distribution (df = n-1)

(b) Whether an 80% confidence interval for the population mean would be wider than, narrower than or the same as the answer to part (1).

Conclusion: The interval will be narrower, because the critical values will be smaller in absolute value.

(c) Test the null hypothesis $H_0 : \mu = 3.785$ against the $H_1 : \mu \neq 3.785$ at 1% significance level.

(c) Test the null hypothesis $H_0 : \mu = 3.785$ against the $H_1 : \mu \neq 3.785$ at 1% significance level.

In (a), $\mu = 3.785$ lies inside the 95% confidence interval $[3.7843, 3.8314]$, so the null hypothesis is already not rejected at 5%.

(c) Test the null hypothesis $H_0 : \mu = 3.785$ against the $H_1 : \mu \neq 3.785$ at 1% significance level.

In (1), $\mu = 3.785$ lies inside the 95% confidence interval $[3.7843, 3.8314]$, so the null hypothesis is already not rejected at 5%.

The 99% confidence interval contains the 95% confidence interval (when being contructed from the same sample), so $\mu = 3.785$ also lies inside this interval. Hence, the null hypothesis is not rejected at 1%.

Exercise 2

Picture the Scenario

- **Objective**: You have accepted a job in Taiwan and want to optimize your commuting costs and rental costs. You are in the process of buying an apartment. You have been informed that a house in Taiwan is considered overpriced if it is sold at a price higher than 3,500 USD/m^2. You wonder if the population mean price of the houses sold in 2012-2013 was considered overpriced.
- **Dataset**: `TaiwanRealEstate.csv`

Questions

(a) Use a classical hypothesis test to determine if there is enough evidence to conclude that the population mean price of the houses sold in 2012-2013 was considered overpriced. Use a probability of Type I error equal to 0.05.

(b) Prepare a power curve for the test (*Hint*: Find the population mean values for $\beta = 0.50$, $\beta = 0.25$, $\beta = 0.10$ and $\beta = 0.05$, and plot those means versus the power of the test).

Solution

```r
# Load Dataset
Taiwan <- read.csv('TaiwanRealEstate.csv')
str(Taiwan)
# Price of houses in the sample
price <- Taiwan$House.price..usd.m.2.
# Check if there is any "Not Available" data point
anyNA(price)
# is.na(price)
# sum(is.na(price))
# Omit the "Not Available" elements
x <- na.omit(price)
```

(a) Hypothesis test using $\alpha = 0.05$

Procedure includes 4 steps:

- Null hypothesis H_0
- Alternative hypothesis H_1
- Decision rule
- Conclusion

(a) Hypothesis test using $\alpha = 0.05$

- Null hypothesis H_0
  - H_0: the population mean price of the houses sold in 2012-2013 was not considered overpriced
    $H_0 : \mu \leq 3500$
- Alternative hypothesis H_1
  - H_1: the population mean price of the houses sold in 2012-2013 was considered overpriced
    $H_1 : \mu > 3500$
- Decision rule
- Conclusion

(a) Hypothesis test using $\alpha = 0.05$

- Null hypothesis H_0
  - H_0: the population mean price of the houses sold in 2012-2013 was not considered overpriced
    $H_0 : \mu \leq 3500$
- Alternative hypothesis H_1
  - H_1: the population mean price of the houses sold in 2012-2013 was considered overpriced
    $H_1 : \mu > 3500$

  $\Longrightarrow$ This is an *upper-tail test* since the alternative hypothesis is focused on the upper tail above the mean of 3500
- Decision rule
- Conclusion

Decision Rule

Assume the population is normal, and the population variance is unknown:

$$\frac{\bar{X} - \mu_0}{s\sqrt{n}} \sim t_{n-1}$$

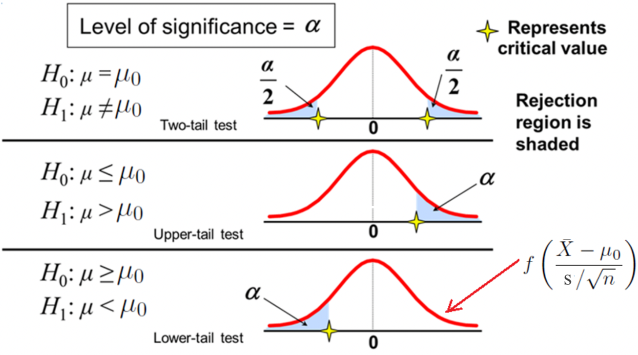


Figure 3: Level of Significance and the Rejection Region: one-sided vs two-sided alternatives

Decision Rule

For upper-tail test, reject H_0 if:

$$t = \frac{\bar{x} - \mu_0}{s\sqrt{n}} > t_{n-1,\alpha}$$

$$\Leftrightarrow \bar{x} > \bar{x}_c = \mu_0 + t_{n-1,\alpha} \times \frac{s}{\sqrt{n}}$$

Decision Rule

```
# Compute the level of significance
alpha <- 0.05
# Compute the sample size
n <- length(x)
# Compute the critical value t_c
t_c <- qt(alpha, n-1, lower.tail = FALSE)
# Compute the critical value \bar{x}_c
x_c <- 3500 + t_c*sd(x)/sqrt(n)
x_c
```

```
## [1] 3608.572
```

```
# Compute the sample mean \bar{x}
mean(x)
```

```
## [1] 3740.474
```

```
# Test manually
mean(x) > x_c
```

```
## [1] TRUE
```

Decision Rule

```r
# Alternative: Use the command in R
t.test(x, mu = 3500, conf.level = 0.95, alternative = "greater")
```

```
##
##  One Sample t-test
##
## data:  x
## t = 3.6513, df = 413, p-value = 0.0001472
## alternative hypothesis: true mean is greater than 3500
## 95 percent confidence interval:
##  3631.902      Inf
## sample estimates:
## mean of x
##  3740.474
```

(a) Hypothesis test using $\alpha = 0.05$

- Null hypothesis:
  - H_0: $\mu \leq 3500$
- Alternative hypothesis:
  - H_1: $\mu > 3500$
- Decision rule:
  - Since $\bar{x} = 3740.474 > 3608.572 = \bar{x}_c$, we reject H_0 at $\alpha = 0.05$.
- Conclusion:
  - There is sufficient evidence to conclude that the population mean price of the houses sold in 2012-2013 was considered overpriced.

(b) Prepare a power curve for the test

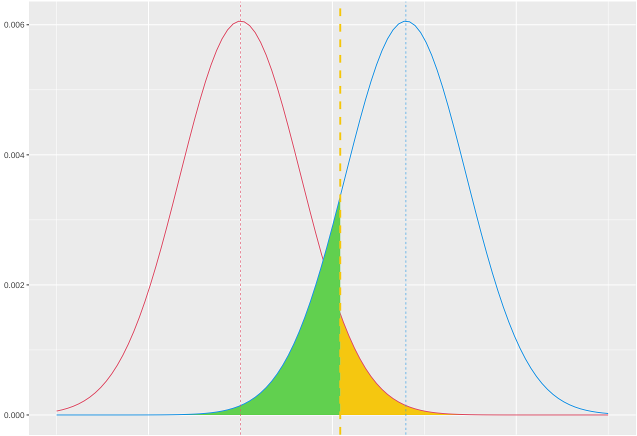
Key:
Outcome
(Probability)

	Actual Situation	
Decision	H_0 True	H_0 False
Do Not Reject H_0	**Correct Decision** **($1 - \alpha$)**	**Type II Error** **(β)**
Reject H_0	**Type I Error** **(α)**	**Correct Decision** **($1 - \beta$)**

- $1 - \beta$ is defined as the power of the test, the probability that a false null hypothesis is rejected.
- The value of β and the power will be different for each value of true mean μ^*.

(b) Prepare a power curve for the test

Suppose we do not reject $H_0 : \mu \leq 3500$ when in fact, the true mean $\mu = \mu^* > 3500$:



(b) Prepare a power curve for the test

$$\beta = P(\text{ Do not reject } H_0 | H_0 \text{ is false})$$

$$\Leftrightarrow \beta = P(\bar{X} \leq \bar{x}_c | \mu = \mu^*)$$

$$\Leftrightarrow \beta = P\left(\frac{\bar{X} - \mu^*}{s/\sqrt{n}} \leq \frac{\bar{x}_c - \mu^*}{s/\sqrt{n}} \mid \mu = \mu^*\right)$$

$$\Leftrightarrow \beta = P\left(t \leq \frac{\bar{x}_c - \mu^*}{s/\sqrt{n}}\right) \text{ where } t = \frac{\bar{X} - \mu^*}{s/\sqrt{n}} \sim t_{n-1}$$

$$\Leftrightarrow 1 - \beta = P\left(t > \frac{\bar{x}_c - \mu^*}{s/\sqrt{n}}\right)$$

$$\Leftrightarrow \frac{\bar{x}_c - \mu^*}{s/\sqrt{n}} = t_{n-1,1-\beta}$$

$$\Leftrightarrow \mu^* = \bar{x}_c - t_{n-1,1-\beta} \times \frac{s}{\sqrt{n}}$$

(b) Prepare a power curve for the test

```
# Assign values for probability of Type II error
beta = c(.5,.25,.1,.05)
# Compute corresponding power of the test
power = 1 - beta
# Compute true value of population mean
mu_star = x_c - qt(power, df = n-1, lower.tail = FALSE)*sd(x)/sq
# Power curve
plot(mu_star, power, col = 'red')
```