

# Introduction to statistics

author: Alejandro Cáceres date:

autosize: true

Barcelona East School of Engineering

Universitat Politècnica de Catalunya (UPC)

# Problems on Confidence intervals

## Summary

- $(1 - \alpha) \cdot 100\%$  confidence intervals for the mean  $\mu$ :

when  $X \hookrightarrow N(\mu, \sigma^2)$

**case 1)** and we **know**  $\sigma^2$  then CI for the estimation of  $\hat{\mu}$

CI for  $\mu$ :

$$(l, u) = (\bar{x} - z_{\alpha/2}\sigma/\sqrt{n}, \bar{x} + z_{\alpha/2}\sigma/\sqrt{n})$$

**case 2)** and we **don't know**  $\sigma^2$  then

CI for  $\mu$ :

$$(l, u) = (\bar{x} - t_{\alpha/2, n-1}s/\sqrt{n}, \bar{x} + t_{\alpha/2, n-1}s/\sqrt{n})$$

- $(1 - \alpha) \cdot 100\%$  confidence intervals for the proportion  $p$ :

when  $K \rightarrow \text{Bernoulli}(p)$ , where  $k = 1$  is the event with probability  $p$ , and  $\bar{K}$  is the average of events in  $n$  trials (relative frequency of events) then the CI for the estimation of  $\hat{p}$  is

**case 3)** CI for  $p$ :

$$(l, u) = (\bar{k} - z_{\alpha/2} \left[ \frac{\bar{k}(1 - \bar{k})}{n} \right]^{1/2}, \bar{k} + z_{\alpha/2} \left[ \frac{\bar{k}(1 - \bar{k})}{n} \right]^{1/2})$$

- $(1 - \alpha) \cdot 100\%$  confidence intervals for the estimation of the variance  $\sigma^2$ :

when  $X \hookrightarrow N(\mu, \sigma^2)$

**case 4)** CI for  $\sigma$ :

$$(l, u) = \left( \frac{s^2(n-1)}{\chi_{\alpha/2, n-1}^2}, \frac{s^2(n-1)}{\chi_{1-\alpha/2, n-1}^2} \right)$$

**Remember** for the formulas:

- case 1 and 3)  $z_{\alpha/2} = \text{qnorm}(1-\alpha/2)$
- case 2)  $t_{\alpha/2, n-1} = \text{qt}(1-\alpha/2, n-1)$
- case 4)  $\chi^2_{\alpha/2, n-1} = \text{qchisq}(1-\alpha/2, n-1)$

## Problem 1

Consider:

- $n = 5$
  - $\alpha = 1 - 0.95 = 0.05$
  - $CI = (229.7, 233.5)$
- a.  $P(\mu \in (229.7, 233.5)) = 0.95?$

No.  $\mu$  is not a random variable it is a parameter of a probability function. Probabilities are defined only for random variables.

## Problem 1

- b. compute  $\bar{x}$  and  $s$

we are given 95% confidence interval

$$CI = (l, u) = (229.7, 233.5)$$

Let's remember the definition:

The 95% **random** confidence interval

$(L, U) = (\bar{X} - f_{sup}, \bar{X} - f_{inf})$  is such that

$$P(\bar{X} - f_{sup} \leq \mu \leq \bar{X} - f_{inf}) = P(f_{inf} \leq \mu - \bar{X} \leq f_{sup}) = 0.95$$

## Problem 1

if:

- $X \hookrightarrow N(\mu, \sigma^2)$  and
- and we **do not know** the variance  $\sigma^2$

Then:  $\alpha = 0.05$

$$(l, u) = (\bar{x} - t_{\alpha/2, n-1} s / \sqrt{n}, \bar{x} + t_{\alpha/2, n-1} s / \sqrt{n})$$

$$t_{0.025, n-1} * S / \sqrt{n}$$

$$(l, u) = (\bar{x} - t_{0.025, n-1} s / \sqrt{n}, \bar{x} + t_{0.025, n-1} s / \sqrt{n})$$

where

$$t_{0.025, n-1} = F_{t, n-1}^{-1}(0.975) = \text{qt}(0.975, 4) = 2.77$$

## Problem 1

Therefore we are given

$$(\bar{x} - 2.77s/\sqrt{5}, \bar{x} + 2.77s/\sqrt{5}) = (229.7, 233.5)$$

and to equations to solve for  $\bar{x}$  and 2:

$$\text{i. } \bar{x} - 1.23s = 229.7$$

$$\text{ii. } \bar{x} + 1.23s = 233.5$$

With solutions:

$$\bar{x} = (229.7 + 233.5)/2 = 231.6 \text{ and } s = (233.5 - \bar{x})/1.23 = 1.53$$

## Problem 1

c. compute 99% CI:

we have:

$$(l, u) = (\bar{x} - t_{\alpha/2, n-1}s/\sqrt{n}, \bar{x} + t_{\alpha/2, n-1}s/\sqrt{n})$$

We leave out a total of  $\alpha = 1\%$ .

or

$$(l, u) = (\bar{x} - t_{0.005, n-1}s/\sqrt{n}, \bar{x} + t_{0.005, n-1}s/\sqrt{n})$$

## Problem 1

where

$$t_{0.005, n-1} = F_{t, n-1}^{-1}(0.995) = \text{qt}(0.995, 4) = 4.60$$

$$\text{then: } (\bar{x} - 4.60s/\sqrt{5}, \bar{x} + 4.60s/\sqrt{5}) = (228.45, 234.75)$$

## Problem 2

Consider:

- $n = 1000$
- $x = 17$

Where  $X \hookrightarrow \text{Bin}(n, p)$

Remember: If we define the average  $\bar{K}$  of a Bernoulli trial ( $k = 0, 1$ ):

- $\bar{K} = \sum_{i=1}^n K_i = X/n, E(\bar{K}) = E(K) = p$
- $S^2 = \bar{K}(1 - \bar{K})$

The  $(1 - \alpha)100\%$  CI associated with a set of observations is

$$CI = (l, u) = (\bar{k} - z_{\alpha/2} \left[ \frac{\bar{k}(1-\bar{k})}{n} \right]^{1/2}, \bar{k} + z_{\alpha/2} \left[ \frac{\bar{k}(1-\bar{k})}{n} \right]^{1/2})$$

when  $np$  and  $n(p - 1) > 5$

## Problem 2

a. Then we have:  $\hat{p} = \bar{k} = 0.017$

$n\hat{p} = 17$  and  $n(1 - \hat{p}) = 983 > 5$  and we can use the approximation:

$$Z = \frac{X - \mu}{\sigma} = \frac{X - E(X)}{\sqrt{V(X)}} = \frac{X - np}{[np(1-p)]^{1/2}} = \frac{\bar{K} - p}{[p(1-p)/n]^{1/2}} \rightarrow N(0, 1)$$

## Problem 2

b. compute the 99% confidence interval then  $\alpha = 0.01$

$$CI = (l, u) = (\bar{k} - z_{0.005} \left[ \frac{\bar{k}(1-\bar{k})}{n} \right]^{1/2}, \bar{k} + z_{0.005} \left[ \frac{\bar{k}(1-\bar{k})}{n} \right]^{1/2})$$

$$\text{Since: } z_{0.005} = \Phi^{-1}(0.995) = \text{qnorm}(0.995) = 2.575829$$

Then

$$CI = (0.006474, 0.027526)$$

or

$$\hat{p} = 0.017 \pm 0.01$$

c. the estimate  $\hat{p} \leq 0.02752$  we cannot guarantee the conditions of the client with 99% confidence.

## Problem 3

Consider:

- $P(Y = 0) = a$
- $P(Y = 1) = 1 - a$
- $na(1 - a) \geq 8$

a. Compute  $n$  such that for the CI  $(l, u)$ ,  $D = u - l$  is maximum.

$Y$  is a Bernoulli variable with  $p = 1 - a$  and as  $n(1 - p)p \geq 8$  then

$$CI = (\bar{y} - z_{\alpha/2} \left[ \frac{\bar{y}(1-\bar{y})}{n} \right]^{1/2}, \bar{y} + z_{\alpha/2} \left[ \frac{\bar{y}(1-\bar{y})}{n} \right]^{1/2})$$

$$D = 2z_{\alpha/2} \left[ \frac{\bar{y}(1-\bar{y})}{n} \right]^{1/2} \text{ is maximum when } D^2 \text{ is maximum.}$$

## Problem 3

$$\frac{dD^2}{d\bar{y}} = \frac{(2z_{\alpha/2})^2}{n} (1 - 2\bar{y}) = 0 \text{ then}$$

$$D \text{ is maximum when the estimated proportion } \bar{y} = 1/2 \text{ or } D_{max} = \frac{z_{\alpha/2}}{\sqrt{n}}$$

## Problem 3

b. Consider  $D = 0.02$  and  $\alpha = 0.10$  for 90% confidence. Compute minimum  $n$ .

$$z_{\alpha/2} = z_{0.05} = \Phi^{-1}(0.95) = \text{qnorm}(0.95) = 1.644854$$

$$D_{max} = 0.02 = \frac{1.644854}{\sqrt{n_{min}}} \text{ then}$$

$$n_{min} = (1.644854/0.02)^2 = 6763.862$$

or

$$n \geq 6764 \text{ (differences with results are because from tables } z_{0.05} \sim 1.65)$$

## Problem 4

Consider:

- 682, 553, 555, 666, 657, 649, 522, 568, 700, 558
- $X \hookrightarrow N(\mu, \sigma_X^2)$

a. Compute 95% CI

- $\bar{x} = 611$
- $s = 65.51$
- $n = 10$
- As we don't know the variance of  $X$  then we use  $s$

the observed CI is:

$$(l, u) = (\bar{x} - t_{0.025, n-1} s / \sqrt{n}, \bar{x} + t_{0.025, n-1} s / \sqrt{n})$$

## Problem 4

$$t_{0.025, n-1} = F_{t, n-1}^{-1}(0.975) = \text{qt}(0.975, 9) = 2.26$$

$$\text{then: } (l, u) = (\bar{x} - 2.26s / \sqrt{n}, \bar{x} + 2.26s / \sqrt{n}) = (564.135, 657.865)$$

## Problem 5

Consider:

- $n = 9$
- the 90% CI is (118.25, 123.55)
- $X \hookrightarrow N(\mu, \sigma^2)$
- We know  $\sigma$

$$\text{the observed CI is: } (l, u) = (\bar{x} - z_{0.05} \sigma / \sqrt{n}, \bar{x} + z_{0.05} \sigma / \sqrt{n})$$

## Problem 5

where

$$z_{0.05, n-1} = \Phi^{-1}(0.95) = \text{qnorm}(0.95) = 1.644$$

then we have two equations

$$\text{i. } l = \bar{x} - 1.644 * \sigma / \sqrt{9} = 118.25$$

$$\text{ii. } u = \bar{x} + 1.644 * \sigma / \sqrt{9} = 123.55$$

a. y b. Compute  $\bar{x}$  and  $\sigma_X^2$

solving i and ii for  $\bar{x}$  and  $\sigma_X^2$

$$\bar{x} = (118.25 + 123.55) / 2 = 120.9 \text{ and } \sigma = 3 * (123.55 - \bar{x}) / 1.644 = 4.83 \text{ or } \sigma^2 = 23.36$$

## Problem 5

c. Compute 97% confidence interval, then  $\alpha = 0.03$

Remember:

$$(l, u) = (\bar{x} - z_{\alpha/2}\sigma/\sqrt{n}, \bar{x} + z_{\alpha/2}\sigma/\sqrt{n})$$

as  $z_{\alpha/2} = z_{0.015} = \Phi^{-1}(0.985) = \text{qnorm}(0.985) = 2.170$  then

$$(l, u) = (120.9 - 2.170 * 4.83/\sqrt{9}, 120.9 + 2.170 * 4.83/\sqrt{9})$$

$$= (117.4063, 124.3937)$$

## Problem 5

d. for  $d = 123.55 - 118.25 = 5.3 = (u - l)$  compute  $n$

$$d = u - l = 2 * 2.170 * 4.83/\sqrt{n} = 5.3$$

then

$$n = \left(\frac{2*2.170*4.83}{5.3}\right)^2 = 15.64 \sim 16$$

## Problem 6

consider:

- $\bar{x} = 0.5354$
- $s = 0.3479$
- $n = 51$
- $\alpha = 0.05$

a. the 95% CI for the variance

$$(l, u) = \left(\frac{s^2(n-1)}{\chi_{\alpha/2, n-1}^2}, \frac{s^2(n-1)}{\chi_{1-\alpha/2, n-1}^2}\right)$$

## Problem 6

$$l = \frac{s^2(n-1)}{\chi_{\alpha/2, n-1}^2}$$

$$\chi_{\alpha/2, n-1}^2 = \chi_{0.025, 50}^2 = \text{qchisq}(1-0.025, 50) = 71.42$$

$$\text{then } l = 0.3479^2 * 50 / 71.42 = 0.0847$$

## Problem 6

$$u = \frac{s^2(n-1)}{\chi_{1-\alpha/2, n-1}^2}$$

$$\chi_{1-\alpha/2, n-1}^2 = \chi_{0.975, 50}^2 = \text{qchisq}(1-0.975, 50) = 32.35$$

$$\text{then } u = 0.3479^2 * 50 / 32.35 = 0.1870$$

$$CI = (l, u) = (0.0847, 0.1870)$$

## Problem 6

consider:

- $X = 17$ , number of fisheries with concentrations greater than 0.700 ppm
- $\alpha = 0.01$
- then  $\hat{p} = \bar{k} = 1/3$  where  $K$  is a Bernoulli trial.

## Problem 6

since  $n\hat{p} = 17$  and  $n(1 - \hat{p}) = 34 > 5$  then CI is

$$(l, u) = (\hat{p} - z_{\alpha/2} \left[ \frac{\hat{p}(1-\hat{p})}{n} \right]^{1/2}, \hat{p} + z_{\alpha/2} \left[ \frac{\hat{p}(1-\hat{p})}{n} \right]^{1/2})$$

since:

$$z_{\alpha/2} = \text{qnorm}(1-0.005) = 2.575$$

then

$$(l, u) = (1/3 + 2.575 * \sqrt{\frac{1/3 * 2/3}{51}}, 1/3 + 2.575 * \sqrt{\frac{1/3 * 2/3}{51}}) = (0.1634, 0.5033)$$

# Introduction to statistics

author: Alejandro Cáceres date:  
autosize: true

Barcelona East School of Engineering  
Universitat Politècnica de Catalunya (UPC)

# Problems on hypothesis testing

## Summary

Hypothesis testing:

1.From the problem context, identify the parameter of interest:  $\mu$ ,  $p$  or  $\sigma$

2.State the null hypothesis, for instance for  $\mu$

- two tailed:  $H_0 : \mu = \mu_0$
- or, upper tailed:  $H_0 : \mu \leq \mu_0$
- or, lower tailed:  $H_0 : \mu \geq \mu_0$

3.Specify an appropriate alternative hypothesis:

- two tailed:  $H_1 : \mu \neq \mu_0$
- or, upper tailed:  $H_1 : \mu > \mu_0$
- or, lower tailed:  $H_1 : \mu < \mu_0$

Note: upper and lower tail refers where you expect  $\mu$  to be if  $H_1$  is true.

4.Choose a significance level:  $\alpha$  (for example  $\alpha = 0.05$  when we want 95% confidence).

5.Define the statistic

when  $X \hookrightarrow N(\mu, \sigma) = N(\mu, \sigma)$

Hypothesis for  $\mu$

**case 1)** if  $\sigma$  is known then

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

is standard ( `pnorm(z)` )



**case 2)** if  $\sigma$  is not known then

$$T = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

is t-distributed with  $n - 1$  degrees of freedom ( `pt(t,n-1)` )

when  $X \hookrightarrow N(\mu, \sigma) = N(\mu, \sigma)$  and  $n \times p_0, n \times (1 - p_0) > 5$

Hypothesis for  $p_0$

**case 3)**

$$Z = \frac{\bar{X} - p_0}{\left[ \frac{p_0(1-p_0)}{n} \right]^{1/2}}$$

is standard ( `pnorm(z)` ).

when  $X \hookrightarrow N(\mu, \sigma) = N(\mu, \sigma)$

Hypothesis for  $\sigma_0$

**case 4)**

$$W = \frac{(n-1)S^2}{\sigma_0^2}$$

is  $\chi^2$ -distributed with  $n - 1$  degrees of freedom ( `pchisq(w,n-1)` ).

6. Test the hypothesis

- define critical **the region** for the statistic under  $H_0$  with probability of  $\alpha$  at the edges
- or, compute confidence intervals at  $\alpha$  confidence limit
- or, compute  $P$ -value:
  - two tail:  $P\text{-value} = 2(1 - F(|z|))$ ,
  - upper tail:  $P\text{-value} = (1 - F(z))$ ,
  - lower tail:  $P\text{-value} = F(z)$

7. Decide

Reject the null hypothesis if:

- the observed statistic falls in the critical region
- or, the  $(1 - \alpha)100\%$  confidence interval does not contain  $\mu_0$  ( $\sigma_0, p_0$ , etc...)
- or, if  $P < \alpha$

Otherwise **do not** reject null hypothesis

*R code for testing hypothesis:*

#### case 1)

```
library(BSDA)

x <- c(...)

z.test(x, mu = 0, alternative = , sigma.x = , conf.level = )
```

#### case 2)

```
x <- c(...)

t.test(x, mu = , alternative = , conf.level = )
```

#### case 3)

```
x <- c(...)

prop.test(x, n = , p = , alternative = , conf.level = , correct = FALSE )
```

#### case 4)

```
library(EnvStats)

x <- c(...)

varTest(x, sigma.squared = , alternative = , conf.level = , )
```

## Problem 1

- consider the measurements:

204.999, 206.149, 202.150, 207.048, 203.496, 206.343, 203.496, 206.676, 205.831

- $\mu \geq 206.5$ 
  - At 90% confidence ( $\alpha = 0.1$ )
- Test for the mean where we do not know  $\sigma$

Lower tail:  $H_0 : \mu \geq 206.5$ ,  $H_1 : \mu < 206.5$

## Problem 1

- Statistic:  $T = \frac{\bar{X} - \mu_0}{S/\sqrt{9}}$ , t-distribution with  $n - 1$  degrees of freedom
- Observed value:  $t = \frac{205.132 - 206.5}{1.707/\sqrt{9}} = -2.404$

## Problem 1

$P$ -value for lower tail:

$P\text{-value} = F(T) = P(T < -2.404) = \text{pt}(-2.404, 8)$

$= 0.021 < \alpha = 0.10$

Since:  $P < \alpha$  we reject the null hypothesis

# Problem 1

b. if  $\sigma_X^2 = 4$ . Test the hypothesis  $\mu_0 = 206.5$  at 95% confidence

- Test for the mean where know  $\sigma$

Two tail:  $H_0 : \mu = 206.5$ ,  $H_1 : \mu \neq 206.5$

- Statistic:  $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{9}}$ , standard distribution
- Observed value:  $z = \frac{205.132 - 206.5}{2/\sqrt{9}} = -2.05$

# Problem 1

$P$ -value for two tail:

$$P\text{-value} = 2(1 - F(|-2.05|)) = 2*(1 - \text{pnorm}(2.05)) = 0.0403 < \alpha = 0.05$$

Since:  $P < \alpha$  we reject the null hypothesis

# Problem 2

- consider the measurements:

53700, 55500, 53000, 52400, 51000, 62000, 75000, 53800, 56600

- $\mu_0 = 62000$
- a. At 95% confidence ( $\alpha = 0.05$ )
- Test for the mean where we do not know  $\sigma$

Lower tail:  $H_0 : \mu \geq 62000$ ,  $H_1 : \mu < 62000$

# Problem 2

- Statistic:  $T = \frac{\bar{X} - \mu_0}{s/\sqrt{9}}$ , t-distribution with  $n - 1 = 8$  degrees of freedom
- Observed value:  $t = \frac{57000 - 62000}{7464.08/\sqrt{9}} = -2.01$

# Problem 2

- a. critical region

$$P(T < t_{0.95}) = 0.05$$

$$t_{0.95,8} = \text{qt}(0.05, 8) = -1.8595$$

The critical region is  $T < -1.8595$

since  $t = -2.01 < -1.8595$  we then reject the null hypothesis

# Problem 2

- b.  $P$ -value for lower tail:

$$P\text{-value} = F(T) = P(T < -2.01) =$$

$$\text{pt}(-2.01, 8) = 0.03966 < \alpha = 0.05$$

Since:  $P < \alpha$  we reject the null hypothesis

## Problem 2

c. compute the 99% CI for the if  $\sigma^2 = 54760000$

When we know the variance then the CI for an  $n$  sample of normal variables is:

$$(l, u) = (\bar{x} - z_{\alpha/2}\sigma/\sqrt{n}, \bar{x} + z_{\alpha/2}\sigma/\sqrt{n})$$

$$\bar{x} = 57000, \sigma = 7400, n = 9 \text{ and}$$

$$z_{\alpha/2} = z_{0.005} = \text{qnorm}(0.995) = 2.5758$$

Putting everything together

$$(l, u) = (50646.29, 63353.71) \text{ (since it contains 62000, we do not reject } H_0 \text{ with 99\% confidence)}$$

## Problem 3

consider:

- $\bar{x} = 7750$
- $s = 145$
- $n = 6$
- $\mu_0 = 8000$
- We don't make a claim if  $\mu \geq \mu_0$
- We make a claim if  $\mu < \mu_0$

a. if  $\alpha = 0.1$  should we make a claim?

- Test for the mean where we do not know  $\sigma$

Lower tail:  $H_0 : \mu \geq 8000, H_1 : \mu < 8000$

## Problem 3

- Statistic:  $T = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$ , t-distribution with  $n - 1$  degrees of freedom
- Observed value:  $t = \frac{7750 - 8000}{145/\sqrt{6}} = -4.22$

## Problem 3

$P$ -value for lower tail:

$$P\text{-value} = F(T) = P(T < -4.22) = \text{pt}(-4.22, 5)$$

$$= 0.004 < \alpha = 0.01$$

Since:  $P < \alpha$  we reject the null hypothesis, we make a claim.

## Problem 3

b. consider:

- $\sigma_X^2 = 136161$
- $\alpha = 0.05$

Test the same hypothesis.

- Test for the mean where we do not know  $\sigma$

Lower tail:  $H_0 : \mu \geq 8000, H_1 : \mu < 8000$

- Statistic:  $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ , is a standard variable
- Observed value:  $z = \frac{7750 - 8000}{369/\sqrt{6}} = -1.659$

## Problem 3

$P$ -value for lower tail:

$$P\text{-value} = F(Z) = P(Z < -1.659) = \text{pnorm}(-1.659) \\ = 0.004 < \alpha = 0.05$$

Since:  $P < \alpha$  we reject the null hypothesis

## Problem 3

c.

- if  $\mu_1 = 7700$  what is type II error or false negative probability?

$$\beta = P(\text{accept} | H_0 : \text{false})$$

We accept  $H_0$  when the observed value  $z$  falls in the acceptance region, at  $\alpha = 0.05$  that is

$$z > z_c = \text{qnorm}(0.05) = -1.644$$

For this critical  $z_c$ , the critical  $\bar{x}_c$  is

$$z_c = \frac{\bar{x}_c - 8000}{369/\sqrt{6}} = -1.644 \text{ then } \bar{x}_c = 7752.19$$

## Problem 3

For accepting the null hypothesis then we need to observe an average that is greater than  $\bar{x} = 7752.19$

What is the probability that if  $\mu = 7700$ , we accept  $H_0$ , that is:

$$\beta = P(\text{accept} | H_0 : \text{false}) = P(\bar{X} > 7752.19 | \mu = 7700)$$

$$P(\bar{X} > 7752.19 | \mu = 7700) = P(Z > \frac{7752.19 - 7700}{369/\sqrt{6}})$$

$$= 1 - \phi(0.35) = 1 - \text{pnorm}(0.35) = 0.36316$$

## Problem 4

- $\mu_0 = 14$
- $\sigma = 4.8$

- $n = 26$
  - $\bar{x} = 12.5$
  - $s = 2.7$
  - Test for the mean where we do not know  $\sigma$
- a.

Lower tail:  $H_0 : \mu \geq 14, H_1 : \mu < 14$

reject  $H_0$  when  $H_0$  is true is a **false positive** or type I error.

## Problem 4

**false positive** probability:

$$\alpha = P(\text{reject} | H_0 : \text{true})$$

The probability is what we leave out from the rejection zone when  $H_0$  is true.

## Problem 4

consider:

- $\alpha = 0.01$  test the hypothesis
- Test for the mean where we do not know  $\sigma$
- Statistic:  $T = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$ , t-distribution with  $n - 1$  degrees of freedom
- Observed value:  $t = \frac{12.5 - 14}{2.7/\sqrt{26}} = -2.832$

## Problem 4

$P$ -value for lower tail:

$$P\text{-value} = F(T) = P(T < -2.832) = \text{pt}(-2.832, 25)$$

$$= 0.0045 < \alpha = 0.001$$

Since:  $P < \alpha$  we reject the null hypothesis

## Problem 4

c. consider:

- $\sigma = 4.8$
- $\alpha = 0.05$

Test the same hypothesis.

- Test for the mean where we do not know  $\sigma$

Lower tail:  $H_0 : \mu \geq 14, H_1 : \mu < 14$

- Statistic:  $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ , is a standard variable

- Observed value:  $z = \frac{12.5-14}{4.8/\sqrt{26}} = -1.593$

## Problem 4

$P$ -value for lower tail:

$$P\text{-value} = F(Z) = P(Z < -1.593) = \text{pnorm}(-1.593)$$

$$= 0.0555 > \alpha = 0.05$$

Since:  $P > \alpha$  we **do not** reject the null hypothesis

## Problem 5

consider:

- $\bar{x} = 10$
- $s = 1.5$
- $n = 41$

a. What is the CI at 87.886% confidence  $\alpha = 1 - 0.87886 = 0.12114$

while we do not **know**  $\sigma^2$   $n$  is big and then we can use the standard distribution:

$$(l, u) = (\bar{x} - z_{\alpha/2}\sigma/\sqrt{n}, \bar{x} + z_{\alpha/2}\sigma/\sqrt{n})$$

## Problem 5

$$(l, u) = (\bar{x} - z_{\alpha/2}\sigma/\sqrt{n}, \bar{x} + z_{\alpha/2}\sigma/\sqrt{n})$$

where  $z_{\alpha/2} = \text{qnorm}(1-0.12114/2) = 1.55$

then

$$(l, u) = (10 - 1.55 * 1.5/\sqrt{41}, 10 + 1.55 * 1.5/\sqrt{41}) = (9.63, 10.36)$$

## Problem 5

b. consider:

- $\mu_0 = 10.5$
- $\alpha = 0.05$

Test the hypothesis.

Lower tail:  $H_0 : \mu \geq 10.5, H_1 : \mu < 10.5$

- Statistic:  $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ , is a standard variable
- Observed value:  $z = \frac{10-10.5}{1.5/\sqrt{41}} = -2.134$

## Problem 5

$P$ -value for lower tail:

$$P\text{-value} = F(Z) = P(Z < -2.134) = \text{pnorm}(-2.134)$$

$$= 0.016 < \alpha = 0.05$$

Since:  $P < \alpha$  we reject the null hypothesis

## Problem 6

consider:

- measurements: 515, 464, 558, 491
- $\sigma^2 = 10000$
- $\alpha = 0.1$

Lower tail:  $H_0 : \sigma^2 \geq 10000, H_1 : \sigma^2 < 10000$

- the sample variance is

$$s^2 = \text{sd}(c(515, 464, 558, 491))^2 = 1590$$

## Problem 6

- Statistic:  $W = \frac{(n-1)S^2}{\sigma_0^2}$ , is a random variable that follows a  $\chi^2$  with  $n - 1$  degrees of freedom.
- Observed value:  $w_{obs} = \frac{(4-1)1590^2}{10000} = 0.477$

## Problem 6

$P$ -value for lower tail:

$$P\text{-value} = F(W) = P(W < 0.477)$$

$$= \text{pchisq}(0.477, 3) = 0.076 < \alpha = 0.1$$

Since:  $P < \alpha$  we reject the null hypothesis

## Problem 6

consider

- $X \hookrightarrow N(\mu, \sigma^2)$
- $\mu = 500$
- $\sigma = 50$

b. what is  $n$  such that  $P(|\bar{X} - \mu| < 10) = 0.95$ ?

$$P(-10 < \bar{X} - \mu < 10) = 0.95$$

$$P\left(\frac{-10}{\sigma/\sqrt{n}} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{10}{\sigma/\sqrt{n}}\right) = 0.95$$

## Problem 6

$$\Phi\left(\frac{10}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{-10}{\sigma/\sqrt{n}}\right) = 0.95$$

$$\Phi\left(\frac{10}{\sigma/\sqrt{n}}\right) - (1 - \Phi\left(\frac{10}{\sigma/\sqrt{n}}\right)) = 0.95$$



$$\Phi\left(\frac{10}{\sigma/\sqrt{n}}\right) = (1 + 0.95)/2 = 0.975$$

$$\frac{10}{\sigma/\sqrt{n}} = \text{qnorm}(0.975) = 1.959$$

$$\frac{10}{\sigma/\sqrt{n}} = 1.959$$

solving for  $n$  then  $n_{min} = 96.4$ ,  $n > 97$

## Problem 7

consider

- $X \hookrightarrow N(\mu, \sigma^2)$
- $\sigma = 5$
- Upper tail:  $H_0 : \mu \leq 80$ ,  $H_1 : \mu > 80$
- $n = 100$
- $\alpha = 0.0505$

compute type II error, false negative:

$$\beta = P(\text{accept} | H_0 : \text{false})$$

We accept  $H_0$  when the observed value  $z$  falls in the acceptance region, at  $\alpha = 0.0505$

## Problem 7

Acceptance region:

$$z < z_c = \text{qnorm}(1 - 0.0505) = 1.640$$

For this critical value  $z_c$  we obtain the critical value for  $\bar{x}_c$

$$z = \frac{\bar{x}_c - 80}{5/\sqrt{100}} = 1.640 \text{ then } \bar{x}_c = (5/10) * 1.64 + 80 = 80.82$$

## Problem 7

For accepting the null hypothesis  $H_0$  we need to observe an average that is lower than  $\bar{x} = 80.82$

What is the probability that if  $\mu = 81$ , we accept  $H_0$ ?

that is:

$$\beta = P(\text{accept} | H_0 : \text{false}) = P(\bar{X} < 80.82 | \mu = 81)$$

$$P(\bar{X} < 80.82 | \mu = 81) = P(Z < \frac{80.82 - 80}{5/\sqrt{100}})$$

$$= \phi(-0.36) = \text{pnorm}(-0.36) = 0.3594236$$

## Problem 8

Consider:

- $K$  is a Bernoulli variable  $K \rightarrow \text{Bernoulli}(p)$
- The measurements give

$$\bar{k} = \hat{p} = 926/1225 = 0.7559$$

# Problem 8

a.

upper tail:  $H_0 : p \leq p_0 = 0.75, H_1 : p > p_0 = 0.75$

- Statistic:

$$Z_{obs} = \frac{\bar{X} - p_0}{\left[ \frac{p_0(1-p_0)}{n} \right]^{1/2}}$$

is a standard variable because  $np_0, n(1 - p_0) > 5$  by the CLT

- Observed value:  $z = \frac{0.7559 - 0.75}{\sqrt{0.75(1-0.75)/\sqrt{1225}}} = 0.47$

# Problem 8

$P$ -value for upper tail:

$$P\text{-value} = 1 - F(Z) = 1 - P(Z < 0.47) = 1 - \Phi(0.47)$$

$$= 1 - \text{pnorm}(0.47) = 0.31$$

only with a level of significance of 0.31 we can reject the null hypothesis, and conclude that the method produces the results within the tolerance limits.