Introduction to statistics

author: Alejandro Cáceres date:
autosize: true

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

Problems on Confidence intervals

Summary

- $(1-\alpha)$*100% confidence intervals for the mean μ:

case 1:

when

1. $X \to N(\mu, \sigma^2)$
2. we **know** σ^2 then CI for the estimation of $\hat{\mu}$

CI for μ:

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

case 2:

when

1. $X \to N(\mu, \sigma^2)$
2. and we **don't know** σ^2 then

CI for μ:

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

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

case 3:

when

1. $X \to Bernoulli(p)$ 2 and np and $n(1-p) > 5$

CI for p:

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

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

case 4:

when

1. $X \to N(\mu, \sigma^2)$

CI for σ:

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

Remember for the formulas:

- In cases 1 and 3: $z_{\alpha/2} =$ `qnorm(1-alpha/2)`
- In case 2: $t_{\alpha/2,n-1} =$ `qt(1-alpha/2, n-1)`
- In case 4: $\chi^2_{\alpha/2,n-1} =$ `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. μ is not a random variable it is a parameter of a probability function. Probabilities are defined only for random variables.

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

we are given 95% confidence interval

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

Let's remember the definition for **case 2**

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

for $\alpha = 0.05$: $t_{0.025,n-1}$

then

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

where

$t_{0.025,n-1} = F^{-1}_{t,n-1}(0.975) =$ `qt(0.975, 4)` $= 2.77$

Therefore we are given the confidence interval:

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

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

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

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

With solutions:

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

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 $\alpha = 0.01$.

or

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

where

$t_{0.005,n-1} = F^{-1}_{t,n-1}(0.995) =$ `qt(0.995, 4)` $= 4.60$

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

Problem 2

Consider:

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

Where $X \to Bernoulli(p)$

Remember: the sample mean of n Bernoulli trials $\bar{X}$ is an estimator of p:

- $\bar{X} = \sum_{i=1}^{n} X_i$ where $E(\bar{x}) = E(X) = p$
- $S^2 = \bar{x}(1 - \bar{x})$ where $E(S^2) = \sigma^2 = p(1-p)$

The $(1-\alpha)100\%$ CI of sample is (case 3)

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

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

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

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

$Z = \frac{\bar{x}-p}{\big[p(1-p)/n\big]^{1/2}} \to N(0,1)$

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

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

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

Then

$CI = (0.006474, 0.027526)$

or

$\bar{x} = 0.017 \pm 0.01$

c. the estimate $\bar{x} \leq 0.02752$ 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

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

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

$\frac{d(\bar{y}(1-\bar{y}))}{d\bar{y}}=(1-2\bar{y})=0$ then$D$ is maximum when $\bar{y}=1/2$. That is $D_{max}=\frac{z_{\alpha/2}}{\sqrt{n}}$

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

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

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

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

or

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

Problem 4

Consider:

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

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:

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

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

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 $(118.25,123.55)$

- $X \to N(\mu, \sigma^2)$
- We know σ

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

where

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

then we have two equations

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

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

a. y b. Compute $\bar{x}$ and σ_X^2

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

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

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 $z_{\alpha/2} = z_{0.015} = \phi^{-1}(0.985) =$ `qnorm(0.985)` $= 2.170$ then

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

$= (117.4063, 124.3937)$

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

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

Problem 6

consider:

- $\bar{x} = 0.5354$
- $s = 0.3479 * \sqrt(51/50) = 0.351395$
- $n = 51$
- $\alpha = 0.05$

a. the 95% CI for the variance

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

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

$\chi^2_{\alpha/2,n-1} = \chi^2_{0.025,50} =$ `qchisq(1-0.025, 50)` $= 71.42$

then $l = 0.351395^2 * 50/71.42 = 0.08644529$

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

$\chi^2_{1-\alpha/2,n-1} = \chi^2_{0.975,50} =$ `qchisq(1-0.975, 50)` $= 32.35$

then $u = 0.351395^2 * 50/32.35 = 0.1908477$

$CI = (l, u) = (0.08644529, 0.1908477)$

```
library(Ecfun); confint.var(0.351395^2, 50)
```

consider:

- $X = 17$, number of fisheries with concentrations greater than 0.700 ppm
- $\alpha = 0.01$
- then $\bar{x} = \bar{x} = 1/3$ where $X \rightarrow Bernoulli(p)$.

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

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

since:

$z_{\alpha/2} =$ `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

For hypothesis testing, follow steps from 1 to 7:

1.From the problem context, identify the parameter of interest: μ, p or σ^2

2.State the hypothesis contrast , for instance for μ

- two tailed: a. $H_0 : \mu = \mu_0$, b. $H_1 : \mu \neq \mu_0$
- or, upper tailed: a. $H_0 : \mu \leq \mu_0$, b. $H_1 : \mu > \mu_0$
- or, lower tailed: a. $H_0 : \mu \geq \mu_0$, b. $H_1 : \mu < \mu_0$

Note: the type of tail refers to H_1.

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

4.Define the statistic

case 1

Hypothesis for μ

i. $X \to N(\mu, \sigma^2)$

ii. we **know** σ^2

then

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

is standard. In R: `pnorm(z)`

case 2

Hypothesis for μ

i. $X \to N(\mu, \sigma^2)$
ii. and we **don't know** σ^2 then

then

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

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

case 3

Hypothesis for p

i. $X \rightarrow Bernoulli(p)$
ii. and np and $n(1-p) > 5$

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

is standard. In R: `pnorm(z)` .

case 4

Hypothesis for σ^2

i. $X \rightarrow N(\mu, \sigma^2)$

then

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

is χ^2-distributed with $n-1$ degrees of freedom. In R: `pchisq(w,n-1)` .

5.Test the hypothesis

a. compute the confidence interval at $(1-\alpha)100\%$ confidence assuming H_0 is true

b. or, define the acceptance **region** for the statistic assuming H_0 is true with probability of α at the edges

c. or, compute the P-value:

- two tail: P-value$= 2(1-F(|z|))$,
- upper tail: P-value$= (1-F(z))$,
- lower tail: P-value$= F(z)$

7.Decide

Reject H_0 (accept H_1) if:

a. the $(1-\alpha)100\%$ confidence interval does not contain μ_0 (σ_0, p_0, etc…)

b. or, the observed statistic falls outside the acceptance region

c. or, if the P-value $< \alpha$

Otherwise **do not** reject H_0 and accept the null.

R code for testing hypotheses:

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

- $\mu \geq 206.5$

a. At 90% confidence ($\alpha = 0.1$)

- Test for the mean where we do not know σ

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

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

P-value for lower tail:

P-value$= F(T) = P(T < -2.404) =$ `pt(-2.404, 8)`

$= 0.021 < \alpha = 0.10$

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

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

- Test for the mean where know σ

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: $z = \frac{205.132-206.5}{2/\sqrt{9}} = -2.05$

P-value for two tail:

P-value$= 2(1-F(|-2.05|)) =$ `2*(1-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 ($\alpha = 0.5$)

- Test for the mean where we do not know σ

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

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

a. critical region

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

$t_{0.95,8} =$ `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

b. P-value for lower tail:

P-value$= F(T) = P(T < -2.01) =$

`pt(-2.01, 8)` $= 0.03966 < \alpha = 0.05$

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

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

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

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

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

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

Putting everything together

$(l, u) = (50646.29, 63353.71)$ (since it contains 62000, we do not reject H_0 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 σ

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

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

P-value for lower tail:

P-value$= F(T) = P(T < -4.22) =$ `pt(-4.22, 5)`

$= 0.004 < \alpha = 0.01$

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

b. consider:

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

Test the same hypothesis.

- Test for the mean where we do not know σ

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: $z = \frac{7750-8000}{369/\sqrt{6}} = -1.659$

P-value for lower tail:

P-value$= F(Z) = P(Z < -1.659) =$ `pnorm(-1.659)`

$= 0.004 < \alpha = 0.05$

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

c.

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

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

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

$z > z_c =$ `qnorm(0.05)` $= -1.644$

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

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

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:

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

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

$= 1 - \phi(0.35) =$ `1-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 σ

   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.

false positive probability:

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

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

consider:

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

P-value for lower tail:

P-value$= F(T) = P(T < -2.832) =$ `pt(-2.832, 25)`

$= 0.0045 < \alpha = 0.001$

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

   c. consider:

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

Test the same hypothesis.

- Test for the mean where we do not know σ

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

P-value for lower tail:

P-value$= F(Z) = P(Z < -1.593) =$ `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** σ^2 n is big and then we can use the standard distribution:

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

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

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

then

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

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: $z = \frac{10-10.5}{1.5/\sqrt{41}} = -2.134$

P-value for lower tail:

P-value$= F(Z) = P(Z < -2.134) =$ `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

$s^2 =$ `sd(c(515, 464, 558, 491))^2` $= 1590$

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

P-value for lower tail:

P-value$= F(W) = P(W < 0.477)$

$=$ `pchisq(0.477, 3)` $= 0.076 < \alpha = 0.1$

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

consider

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

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

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

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

$\phi(\frac{10}{\sigma/\sqrt{n}}) - \phi(\frac{-10}{\sigma/\sqrt{n}}) = 0.95$

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

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

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

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

compute type II error, false negative:

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

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

Acceptance region:

$z < z_c =$ `qnorm(1-0.0505)` $= 1.640$

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

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

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:

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

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

$= \phi(-0.36) =$ `pnorm(-0.36)` $= 0.3594236$

Problem 8

Consider:

- X is a Bernoulli variable $X \rightarrow Bernoully(p)$
- The measurements give

$\bar{x} = \bar{x} = 926/1225 = 0.7559$

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

P-value for upper tail:

P-value$= 1 - F(Z) = 1 - P(Z < 0.47) = 1 - \phi(0.47)$

$=$ `1-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.