AUA CS108, Statistics, Fall 2020

Lecture 37

Michael Poghosyan

20 Nov 2020

Contents

- Confidence Intervals

Examples

Example: Assume we are interested in the proportion of smokers in AUA. We ask 120 persons at AUA and learn that 55 of them are smokers. Construct a CI for the proportion of smokers in AUA of 95% confidence level.

Solution: OTB

Examples

Example: Assume we are interested in the proportion of smokers in AUA. We ask 120 persons at AUA and learn that 55 of them are smokers. Construct a CI for the proportion of smokers in AUA of 95% confidence level.

Solution: OTB

Example: Continuing the above Example: now assume we want to find that Proportion within the Error Margin 0.1, with the CL 95%. At least, how many persons at AUA we need to ask?

Solution: OTB

CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, σ is known, Pivotal Method

Problem: Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Assume σ^2 is known. Given $\alpha \in (0, 1)$, we want to construct a CI of CL $1 - \alpha$ for μ, using a Pivot.

CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, σ is known, Pivotal Method

Problem: Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Assume σ^2 is known. Given $\alpha \in (0, 1)$, we want to construct a CI of CL $1 - \alpha$ for μ, using a Pivot.

Answer: The interval

$$\left(\overline{X} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \;\; \overline{X} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

is a $(1 - \alpha)$-level CI for μ, when σ^2 is known (using the Pivoting).

CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, σ is known, Pivotal Method

Problem: Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Assume σ^2 is known. Given $\alpha \in (0,1)$, we want to construct a CI of CL $1-\alpha$ for μ, using a Pivot.

Answer: The interval

$$\left(\overline{X} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \;\; \overline{X} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

is a $(1-\alpha)$-level CI for μ, when σ^2 is known (using the Pivoting).

Here z_β is the β-quantile of $\mathcal{N}(0,1)$.

CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, σ is known, Pivotal Method

Problem: Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Assume σ^2 is known. Given $\alpha \in (0,1)$, we want to construct a CI of CL $1-\alpha$ for μ, using a Pivot.

Answer: The interval

$$\left(\overline{X} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \quad \overline{X} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

is a $(1-\alpha)$-level CI for μ, when σ^2 is known (using the Pivoting).

Here z_β is the β-quantile of $\mathcal{N}(0,1)$.

Note: The Margin of Error in this case is

$$z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}.$$

Example

Example: Assume we want to construct a 95% CI for μ in the $\mathcal{N}(\mu, \sigma^2)$ Model, when σ is given, known.

Example

Example: Assume we want to construct a 95% CI for μ in the $\mathcal{N}(\mu, \sigma^2)$ Model, when σ is given, known. We construct it using a Random Sample

$$X_1, ..., X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Example

Example: Assume we want to construct a 95% CI for μ in the $\mathcal{N}(\mu, \sigma^2)$ Model, when σ is given, known. We construct it using a Random Sample

$$X_1, ..., X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Since $1 - \alpha = 0.95$, then $\alpha = 0.05$.

Example

Example: Assume we want to construct a 95% CI for μ in the $\mathcal{N}(\mu, \sigma^2)$ Model, when σ is given, known. We construct it using a Random Sample

$$X_1, ..., X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Since $1 - \alpha = 0.95$, then $\alpha = 0.05$. By the above formula, we need to calculate the Standard Normal quantile $z_{1-\alpha/2} =$

Example

Example: Assume we want to construct a 95% CI for μ in the $\mathcal{N}(\mu, \sigma^2)$ Model, when σ is given, known. We construct it using a Random Sample

$$X_1, ..., X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Since $1 - \alpha = 0.95$, then $\alpha = 0.05$. By the above formula, we need to calculate the Standard Normal quantile $z_{1-\alpha/2} = z_{0.975}$.

Example

Example: Assume we want to construct a 95% CI for μ in the $\mathcal{N}(\mu, \sigma^2)$ Model, when σ is given, known. We construct it using a Random Sample

$$X_1, ..., X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Since $1 - \alpha = 0.95$, then $\alpha = 0.05$. By the above formula, we need to calculate the Standard Normal quantile $z_{1-\alpha/2} = z_{0.975}$.

R gives:

```r
qnorm(0.975)
```

```
## [1] 1.959964
```

so our 95% CI will be

$$\left(\overline{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}},\ \ \overline{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right).$$

Example

Example with R: We generate random numbers from $\mathcal{N}(2.31, 4)$ (so here we assume we know the true parameter value of μ).

```
sigma <- 2
n <- 20
smpl <- rnorm(n, mean = 2.31, sd = sigma)
smpl
```

```
##  [1]  1.0307146  3.3838580  5.6066839  3.2954158  1.581
##  [7] -2.2018976  4.4390909  6.8156058  4.0429737  1.652
## [13]  0.1586965  2.8670176  2.8652911  2.9551670  8.546
## [19] -0.4402966  2.2230537
```

Example

Example with R: We generate random numbers from $\mathcal{N}(2.31, 4)$ (so here we assume we know the true parameter value of μ).

```
sigma <- 2
n <- 20
smpl <- rnorm(n, mean = 2.31, sd = sigma)
smpl
```

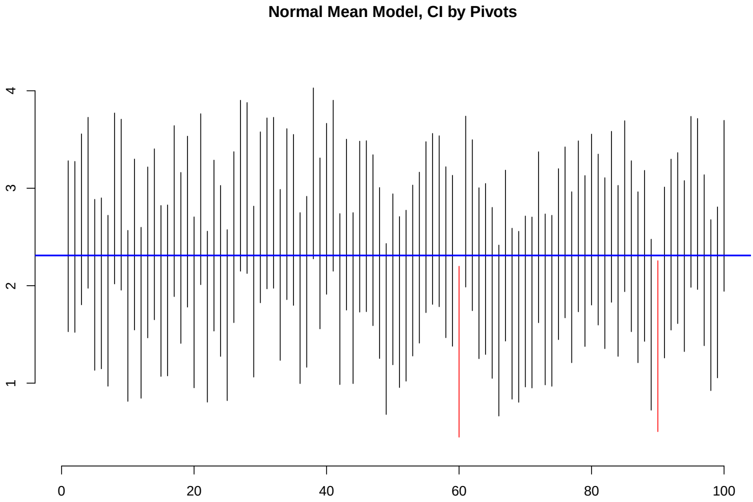
```
##  [1]  1.0307146  3.3838580  5.6066839  3.2954158  1.581
##  [7] -2.2018976  4.4390909  6.8156058  4.0429737  1.652
## [13]  0.1586965  2.8670176  2.8652911  2.9551670  8.546
## [19] -0.4402966  2.2230537
```

Now we construct the 95% CI by the above formula:

```
me <- 1.96* sigma/sqrt(n) #the Margin of Error
c(mean(smpl) - me, mean(smpl) + me)
```

```
## [1] 2.021886 3.774963
```

Example, Simulation



Example, Simulation, Code

```
mu <- 2.31; sigma <- 2
conf.level <- 0.95; a = 1 - conf.level
sample.size <- 20; no.of.intervals <- 100
z <- qnorm(1-a/2) ## our quantile
ME <- z*sigma/sqrt(sample.size) #Margin of Error

plot.new()
plot.window(xlim = c(0,no.of.intervals), ylim = c(mu-2,mu+2))
axis(1); axis(2)
title("Normal Mean Model, CI by Pivots")
for(i in 1:no.of.intervals){
  x <- rnorm(sample.size, mean = mu, sd = sigma)
  lo <- mean(x) - ME; up <- mean(x) + ME
  if(lo > mu || up < mu){
    segments(c(i), c(lo), c(i), c(up), col = "red")
  }
  else{
    segments(c(i), c(lo), c(i), c(up))
  }
}
abline(h = mu, lwd = 2, col = "blue")
```

CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, σ is **unknown**, PivMe

Problem: Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Assume σ^2 is **unknown**, which is more realistic.

CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, σ is **unknown**, PivMe

Problem: Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Assume σ^2 is **unknown**, which is more realistic. Given $\alpha \in (0, 1)$, we want to construct a CI of CL $1 - \alpha$ for μ, using a Pivot.

CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, σ is **unknown**, PivMe

Problem: Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Assume σ^2 is **unknown**, which is more realistic. Given $\alpha \in (0, 1)$, we want to construct a CI of CL $1 - \alpha$ for μ, using a Pivot.

Answer: The following interval:

$$\left(\overline{X} - t_{n-1,1-\alpha/2} \cdot \frac{S}{\sqrt{n}}, \;\; \overline{X} + t_{n-1,1-\alpha/2} \cdot \frac{S}{\sqrt{n}}\right)$$

is a $(1 - \alpha)$-level CI for Normal μ, when σ^2 is unknown.

CI for the Mean of $\mathcal{N}(\mu, \sigma^2)$, σ is **unknown**, PivMe

Problem: Assume we have a Random Sample

$$X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Assume σ^2 is **unknown**, which is more realistic. Given $\alpha \in (0, 1)$, we want to construct a CI of CL $1 - \alpha$ for μ, using a Pivot.

Answer: The following interval:

$$\left(\overline{X} - t_{n-1,1-\alpha/2} \cdot \frac{S}{\sqrt{n}}, \quad \overline{X} + t_{n-1,1-\alpha/2} \cdot \frac{S}{\sqrt{n}}\right)$$

is a $(1 - \alpha)$-level CI for Normal μ, when σ^2 is unknown.

Here $t_{n-1,\beta}$ is the β-quantile of the Student's T-Distribution with $n - 1$ degrees of freedom, which we denote by $t(n - 1)$.

CI for μ, Normal Model, Notes

Note: To compare:

- If σ is known, $(1-\alpha)$-level CI for μ is
$$\overline{X} \pm z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$
- If σ is unknown, $(1-\alpha)$-level CI for μ is
$$\overline{X} \pm t_{n-1,1-\alpha/2} \cdot \frac{S}{\sqrt{n}}$$

CI for μ, Normal Model, Notes

Note: To compare:

- If σ is known, $(1-\alpha)$-level CI for μ is
$$\overline{X} \pm z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$
- If σ is unknown, $(1-\alpha)$-level CI for μ is
$$\overline{X} \pm t_{n-1,1-\alpha/2} \cdot \frac{S}{\sqrt{n}}$$

Note: If we will compare the quantiles of the same level of $\mathcal{N}(0,1)$ with $t(n-1)$, we will see that CIs for the case when σ is unknown are wider than for the case when σ is known. This is intuitive, of course - to compensate the uncertainty in σ, we need to take a wider interval:

```
c(qnorm(0.975), qt(0.975, df = 3), qt(0.975, df = 20))
```

```
## [1] 1.959964 3.182446 2.085963
```

Example

Example: Assume we want to estimate the average time μ our Stat students are spending for a homework. We collect information from 10 students for the last hw, and get the following responses (in hrs):

```
[1] 2.17 1.42 2.13 0.56 1.21 2.22 1.35 2.37 1.47 1.70
```

Example

Example: Assume we want to estimate the average time μ our Stat students are spending for a homework. We collect information from 10 students for the last hw, and get the following responses (in hrs):

```
[1] 2.17 1.42 2.13 0.56 1.21 2.22 1.35 2.37 1.47 1.70
```

Now we assume that the time spent for solving a hw is Normally Distributed with the Mean μ and some variance σ^2.

Example

Example: Assume we want to estimate the average time μ our Stat students are spending for a homework. We collect information from 10 students for the last hw, and get the following responses (in hrs):

```
[1] 2.17 1.42 2.13 0.56 1.21 2.22 1.35 2.37 1.47 1.70
```

Now we assume that the time spent for solving a hw is Normally Distributed with the Mean μ and some variance σ^2.

So $X \sim \mathcal{N}(\mu, \sigma^2)$ shows

Example

Example: Assume we want to estimate the average time μ our Stat students are spending for a homework. We collect information from 10 students for the last hw, and get the following responses (in hrs):

```
[1] 2.17 1.42 2.13 0.56 1.21 2.22 1.35 2.37 1.47 1.70
```

Now we assume that the time spent for solving a hw is Normally Distributed with the Mean μ and some variance σ^2.

So $X \sim \mathcal{N}(\mu, \sigma^2)$ shows the time spent solving a hw for a (randomly chosen) student.

Example

Example: Assume we want to estimate the average time μ our Stat students are spending for a homework. We collect information from 10 students for the last hw, and get the following responses (in hrs):

```
[1] 2.17 1.42 2.13 0.56 1.21 2.22 1.35 2.37 1.47 1.70
```

Now we assume that the time spent for solving a hw is Normally Distributed with the Mean μ and some variance σ^2.

So $X \sim \mathcal{N}(\mu, \sigma^2)$ shows the time spent solving a hw for a (randomly chosen) student. Our unknown, μ is

Example

Example: Assume we want to estimate the average time μ our Stat students are spending for a homework. We collect information from 10 students for the last hw, and get the following responses (in hrs):

```
[1] 2.17 1.42 2.13 0.56 1.21 2.22 1.35 2.37 1.47 1.70
```

Now we assume that the time spent for solving a hw is Normally Distributed with the Mean μ and some variance σ^2.

So $X \sim \mathcal{N}(\mu, \sigma^2)$ shows the time spent solving a hw for a (randomly chosen) student. Our unknown, μ is the average time to solve a hw.

Example

Example: Assume we want to estimate the average time μ our Stat students are spending for a homework. We collect information from 10 students for the last hw, and get the following responses (in hrs):

```
[1] 2.17 1.42 2.13 0.56 1.21 2.22 1.35 2.37 1.47 1.70
```

Now we assume that the time spent for solving a hw is Normally Distributed with the Mean μ and some variance σ^2.

So $X \sim \mathcal{N}(\mu, \sigma^2)$ shows the time spent solving a hw for a (randomly chosen) student. Our unknown, μ is the average time to solve a hw. And we have an observation from a Random Sample

$$X_1, X_2, ..., X_{10} \sim \mathcal{N}(\mu, \sigma^2).$$

Example

Example: Assume we want to estimate the average time μ our Stat students are spending for a homework. We collect information from 10 students for the last hw, and get the following responses (in hrs):

```
[1] 2.17 1.42 2.13 0.56 1.21 2.22 1.35 2.37 1.47 1.70
```

Now we assume that the time spent for solving a hw is Normally Distributed with the Mean μ and some variance σ^2.

So $X \sim \mathcal{N}(\mu, \sigma^2)$ shows the time spent solving a hw for a (randomly chosen) student. Our unknown, μ is the average time to solve a hw. And we have an observation from a Random Sample

$$X_1, X_2, ..., X_{10} \sim \mathcal{N}(\mu, \sigma^2).$$

We construct a 95% CI for μ, the average time to solve the hw, by the above formula:

Example

```
n <- 10
cl <- 0.95; a = 1-cl
t <- qt(1-a/2, df = n-1)
smpl<-c(2.17,1.42,2.13,0.56,1.21,2.22,1.35,2.37,1.47,1.70)
s <- sd(smpl)
me <-  t*s/sqrt(n) #the Margin of Error
c(mean(smpl) - me, mean(smpl) + me)
```

```
## [1] 1.253748 2.066252
```