

HW8

Manny Fomundam

2023-12-04

Given a normal population with known variance σ^2 by what factor must the sample size be increased to reduce the length of a confidence interval for the mean by a factor of k ?

```
# Define the known variance
sigma_sq <- 1

# Define the factor by which you want to reduce the length of the confidence interval
k <- 2

# Calculate the original sample size needed for a certain length of confidence interval
# (This is just an example, replace 0.05 with the half length of your original confidence interval)
n_original <- (2*1.96*sqrt(sigma_sq)/0.05)^2

# Calculate the new sample size needed to reduce the length of the confidence interval by a factor of k
n_new <- k^2 * n_original

# Print the new sample size
print(n_new)
```

```
## [1] 24586.24
```

In a study conducted at Appalachian State University, students used digital oral thermometers to record their temperatures each day they came to class. A randomly selected day of student temperatures is provided in the following table and in the data frame STATTEMPS. Information is also provided with regard to subject gender and the hour of the day when the students' temperatures were measured.

- Construct a 95% confidence interval for the true average temperature difference between females and males. Does the interval contain the value zero? What does this suggest about gender temperature differences?
- Construct a 95% confidence interval for the true average temperature difference between students taking their temperatures at 8 a.m. and students taking their temperatures at 9 a.m. Give a reason why one group appears to have a higher temperature reading.

```
library(PASWR2)
```

```
## Warning: package 'PASWR2' was built under R version 4.3.2
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```
data("STATTEMPS")
```

```
# For gender difference
```

```
male_temps <- STATTEMPS$temperature[STATTEMPS$gender == "Male"]
```

```
female_temps <- STATTEMPS$temperature[STATTEMPS$gender == "Female"]
```

```

mean_diff_gender <- mean(female_temps) - mean(male_temps)
print(mean_diff_gender)

## [1] 0.3027668

std_error_gender <- sqrt(var(female_temps)/length(female_temps) + var(male_temps)/length(male_temps))
print(std_error_gender)

## [1] 0.6483239

# 95% confidence interval
ci_gender <- mean_diff_gender + c(-1, 1) * qt(0.975, df=length(female_temps)+length(male_temps)-2) * std_error_gender
print(ci_gender)

## [1] -1.017826 1.623359

# For time difference
temps_8am <- STATTEMPS$temperature[STATTEMPS$hour == 8]
print(temps_8am)

## numeric(0)

temps_9am <- STATTEMPS$temperature[STATTEMPS$hour == 9]
print(temps_9am)

## numeric(0)

mean_diff_time <- mean(temps_8am) - mean(temps_9am)
print(mean_diff_time)

## [1] NaN

std_error_time <- sqrt(var(temps_8am)/length(temps_8am) + var(temps_9am)/length(temps_9am))
print(std_error_time)

## [1] NA

# 95% confidence interval
ci_time <- mean_diff_time + c(-1, 1) * qt(0.975, df=length(temps_8am)+length(temps_9am)-2) * std_error_time
print(ci_time)

## Warning in qt(0.975, df = length(temps_8am) + length(temps_9am) - 2): NaNs
## produced

## [1] NaN NaN

# Print the confidence intervals
print(paste("95% confidence interval for gender difference: [", ci_gender[1], ",", ci_gender[2], "]"))

## [1] "95% confidence interval for gender difference: [ -1.01782574162748 , 1.62335933846543 ]"

print(paste("95% confidence interval for time difference: [", ci_time[1], ",", ci_time[2], "]"))

## [1] "95% confidence interval for time difference: [ NaN , NaN ]"

```

The following data were taken to measure the unknown pH values μ of a solution in a chemical experiment: 8.01, 8.05, 7.96, 8.04, 8.03, 8.03, 8.02, 7.98, 8.05, 8.03. If the pH meter has a systematic error, Δ , and a normally distributed random error, $\epsilon \sim N(0, \sigma)$, then it can be assumed that the observations come from a normal random variable, $X \sim N(\mu + \Delta, \sigma)$ (a) Compute a 95% confidence interval for μ when $\Delta = 0$ and $\sigma = 0.05$. Compute the interval assuming that the variance is unknown. (b) Repeat part (a) with $\Delta = 0.2$.

```

# Assuming your data is stored in a vector called 'data'
data <- c(8.01, 8.05, 7.96, 8.04, 8.03, 8.03, 8.02, 7.98, 8.05, 8.03)

# Compute sample mean and standard deviation
mean_data <- mean(data)
sd_data <- sd(data)

# Compute 95% confidence interval for mu when systematic error = 0 and random error = 0.05
ci_1 <- mean_data + c(-1, 1) * qt(0.975, df=length(data)-1) * sd_data/sqrt(length(data))

# Compute 95% confidence interval for mu when systematic error = 0.2
ci_2 <- mean_data + 0.2 + c(-1, 1) * qt(0.975, df=length(data)-1) * sd_data/sqrt(length(data))

# Print the confidence intervals
print(paste("95% confidence interval for mu when systematic error = 0 and random error = 0.05: [", ci_1

## [1] "95% confidence interval for mu when systematic error = 0 and random error = 0.05: [ 7.998940462

print(paste("95% confidence interval for mu when systematic error = 0.2: [", ci_2[1], ",", ci_2[2], "]".

## [1] "95% confidence interval for mu when systematic error = 0.2: [ 8.19894046290797 , 8.241059537092

```