

140.615.HW.03.Jin.Vincent

Vincent Jin

2023-02-24

140.615.HW.03.Jin.Vincent

Vincent Jin

1. Suppose I measure some treatment response on a set of 10 mice, and receive the following data:

84 96 89 103 100 97 125 105 111 108

Calculate a 95% confidence interval for the population mean treatment response. Derive the answers analytically by calculating the sample mean, the sample standard deviation, etc, as well as a call to the R function `t.test()`.

Answer

```
mice <- c(84, 96, 89, 103, 100, 97, 125, 105, 111, 108)
mean_mice <- mean(mice)
cat(paste("the mean of the set of 10 mice is:", round(mean_mice, 2), sep = "\n"))

## the mean of the set of 10 mice is:
## 101.8

sd_mice <- sd(mice)
cat(paste(" ", "the standard deviation of the set of 10 mice is:", round(sd_mice, 2), sep = "\n"))

##
## the standard deviation of the set of 10 mice is:
## 11.61

se_mice <- sd_mice / sqrt(length(mice))
cat(paste(" ", "the standard error of the set of 10 mice is:", round(se_mice, 2), sep = "\n"))

##
## the standard error of the set of 10 mice is:
## 3.67
```

```
# get t
t_stats <- qt(0.975, length(mice) - 1)
mice_low <- mean_mice - t_stats * se_mice
mice_up <- mean_mice + t_stats * se_mice

string_helper <- paste("the 95% confidence interval for the sample mean is: \n(", round(mice_low, 2), "
cat(paste(" ", string_helper, sep = "\n"))
```

```
##
## the 95% confidence interval for the sample mean is:
## (93.49, 110.11)
```

```
cat("\n \nthe results from t-test is:")
```

```
##
##
## the results from t-test is:
```

```
print(t.test(mice))
```

```
##
## One Sample t-test
##
## data: mice
## t = 27.722, df = 9, p-value = 5.02e-10
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## 93.49309 110.10691
## sample estimates:
## mean of x
## 101.8
```

2. We obtain the log10 gene expression for 5 tissue samples as follows:

8.52 8.89 8.13 8.66 8.28

Provide a 99% confidence interval for the population mean by using the formula based on the sample mean and the estimated standard error and by a call to the R function `t.test()`.

Answer

```
gene <- c(8.52, 8.89, 8.13, 8.66, 8.28)
mean_gene <- mean(gene)
sd_gene <- sd(gene)
se_gene <- sd_gene / sqrt(length(gene))
t_stats <- qt(1 - ((1 - 0.99) / 2), length(gene) - 1)
gene_low <- mean_gene - t_stats * se_gene
gene_up <- mean_gene + t_stats * se_gene

cat(paste("the mean of the set of 5 tissue samples is:", round(mean_gene, 2), sep = "\n"))
```

```

## the mean of the set of 5 tissue samples is:
## 8.5

cat(paste(" ", "the standard deviation of the set of 5 tissue samples is:", round(sd_gene, 2), sep = "\n"))

##
## the standard deviation of the set of 5 tissue samples is:
## 0.3

cat(paste(" ", "the standard error of the set of 5 tissue samples is:", round(se_gene, 2), sep = "\n"))

##
## the standard error of the set of 5 tissue samples is:
## 0.13

string_helper <- paste("the 99% confidence interval for the sample mean is: \n(", round(gene_low, 2), "\n",
cat(paste(" ", string_helper, sep = "\n"))

##
## the 99% confidence interval for the sample mean is:
## (7.88, 9.12)

cat("\n \nthe results from t-test is:")

##
##
## the results from t-test is:

print(t.test(gene, conf.level = 0.99))

##
## One Sample t-test
##
## data: gene
## t = 63.035, df = 4, p-value = 3.794e-07
## alternative hypothesis: true mean is not equal to 0
## 99 percent confidence interval:
## 7.875454 9.116546
## sample estimates:
## mean of x
## 8.496

```

3. We are interested in estimating the concentration of substance X in the Baltimore water supply on the basis of measurements of a number of samples. Suppose measurements of such samples will be

approximately normally distributed with unknown mean (the true concentration) and known SD = 1.5 ppb. How many samples should we measure if we wish our 95% confidence interval for the true concentration to have width < 1 ppb?

Answer

```
sd <- 1.5
width <- 1

n <- (1.96 * sd * 2 / width) ^ 2
print(ceiling(n))
```

```
## [1] 35
```

To have our 95% confidence interval for the true concentration to have width <1 ppb, we need to have at least 35 samples.

4. I measure some treatment response on a set of 10 mice from strain A, and get the following data:

84 106 99 101 100 99 127 105 101 108

I measure the treatment response on a set of 5 mice from strain B, and get the following data:

56 62 67 81 69

Calculate a 95% and a 99% confidence interval for the difference in the mean treatment responses of strains A and B, assuming that the standard deviations within the two groups are the same. Derive the answers analytically as well as a call to the R function `t.test()`.

Answer

```
mice_A <- c(84, 106, 99, 101, 100, 99, 127, 105, 101, 108)
mice_B <- c(56, 62, 67, 81, 69)
mean_A <- mean(mice_A)
mean_B <- mean(mice_B)
sd_A <- sd(mice_A)
sd_B <- sd(mice_B)
se_A <- sd_A / sqrt(length(mice_A))
se_B <- sd_B / sqrt(length(mice_B))
mean_dif <- abs(mean_B - mean_A)
cat("since we assume the standard deviations within the two groups are the same and we do not know the p
```

```
## since we assume the standard deviations within the two groups are the same and we do not know the p
```

```
sd_pooled <- sqrt((sd_A ^ 2 * (length(mice_A) - 1) + sd_B ^ 2 * (length(mice_B) - 1)) / (length(mice_A) + length(mice_B) - 2))
sd_dif <- sd_pooled * sqrt(1 / length(mice_A) + 1 / length(mice_B))
t_stats <- qt(0.975, length(mice_A) + length(mice_B) - 2)

dif_95_low <- mean_dif - t_stats * sd_dif
dif_95_up <- mean_dif + t_stats * sd_dif

t_stats <- qt(0.995, length(mice_A) + length(mice_B) - 2)
dif_99_low <- mean_dif - t_stats * sd_dif
dif_99_up <- mean_dif + t_stats * sd_dif

string_helper <- paste("the 95% confidence interval for the differences among sample means is: \n(", round(dif_95_low, 2), " to ", round(dif_95_up, 2), ")", sep = "\n")
cat(paste(" ", string_helper, sep = "\n"))
```

```
##
## the 95% confidence interval for the differences among sample means is:
## (23.85, 48.15)

string_helper <- paste("the 99% confidence interval for the differences among sample means is: \n(", round(mice_A - mice_B, 2),
cat(paste(" ", string_helper, sep = "\n"))

##
## the 99% confidence interval for the differences among sample means is:
## (19.06, 52.94)

cat("the results from t.test function are:\n")

## the results from t.test function are:

t.test(mice_A, mice_B, var.equal = TRUE)

##
## Two Sample t-test
##
## data: mice_A and mice_B
## t = 6.4026, df = 13, p-value = 2.334e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 23.85277 48.14723
## sample estimates:
## mean of x mean of y
## 103 67

t.test(mice_A, mice_B, var.equal = TRUE, conf.level = 0.99)

##
## Two Sample t-test
##
## data: mice_A and mice_B
## t = 6.4026, df = 13, p-value = 2.334e-05
## alternative hypothesis: true difference in means is not equal to 0
## 99 percent confidence interval:
## 19.06271 52.93729
## sample estimates:
## mean of x mean of y
## 103 67
```

5. Suppose I measure some treatment response on a set of 10 mice, and receive the following data:

101.4 102.5 99.2 97.3 97.5 100.3 100.5 99.5 99.2 95.3

Imagine that the data are independent draws from some normal distribution.

- (a) Calculate a 95% confidence interval for the population mean.

Answer

```
draws <- c(101.4, 102.5, 99.2, 97.3, 97.5, 100.3, 100.5, 99.5, 99.2, 95.3)
draw_mean <- t.test(draws)
print(draw_mean[4])
```

```
## $conf.int
## [1] 97.75473 100.78527
## attr(,"conf.level")
## [1] 0.95
```

The 95% confidence interval for the population mean is (97.75, 100.79)

(b) Calculate a 95% confidence interval for the population standard deviation.

Answer

```
sd_draw <- sd(draws)
df <- length(draws) - 1
L <- qchisq(0.025, df)
U <- qchisq(0.975, df)
sd_low <- sd_draw * sqrt(df / U)
sd_up <- sd_draw * sqrt(df / L)

string_helper <- paste("the 95% confidence interval for the population standard deviation is: \n(", round(
cat(paste(" ", string_helper, sep = "\n"))
```

```
##
## the 95% confidence interval for the population standard deviation is:
## (1.46, 3.87)
```

(c) Calculate a 95% confidence interval for the population variance.

Answer

```
var_up <- sd_draw ^ 2 * df / L
var_low <- sd_draw ^ 2 * df / U

string_helper <- paste("the 95% confidence interval for the population variance is: \n(", round(var_low)
cat(paste(" ", string_helper, sep = "\n"))
```

```
##
## the 95% confidence interval for the population variance is:
## (2.12, 14.95)
```