

# Stat 3202 Lab 11

Nathan Johnson.9254

2023-04-20

## Guided Example 1:

```
library(readr)
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v dplyr 1.0.10
## v tibble 3.1.8       v stringr 1.5.0
## v tidyr 1.2.1        v forcats 0.5.2
## v purrr 1.0.1
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag() masks stats::lag()

batch1 <- c(38,46,41,36,42,40,34,41,33,
46,41,39,44,45,41,37,39,36,
40,43,37,41,39,36,47,36,41,
38,38,38,33,39,41,36,38,44,39)
batch2=c(42, 37, 40, 39 ,45, 31, 45, 36, 37, 38,
41, 35, 46, 41, 44, 37, 38, 39, 38, 38,
42, 44,45, 35, 35, 42, 36, 40)
```

**Part a:** Our test statistic is:

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

**Part b:** Under the null hypothesis, the sampling distribution of this test statistic is  $Z \sim N(0, 1)$ .

**Part c:**

```
xbar1 = mean(batch1)
xbar2 = mean(batch2)

n1 = length(batch1)
n2 = length(batch2)

sigma21 = 3.5^2
sigma22 = 3.5^2

Z.obs = (xbar1 - xbar2) / sqrt(sigma21/n1 + sigma22/n2)
Z.obs
```

```
## [1] 0.04624286
```

Observed test statistic is  $Z_{obs} = 0.0462$ .

**Part d:**

```
alpha = 0.01
```

```
rr1 = qnorm(alpha/2)
```

```
rr2 = qnorm(1-alpha/2)
```

```
rr1
```

```
## [1] -2.575829
```

```
rr2
```

```
## [1] 2.575829
```

```
c(-Inf, rr1)
```

```
## [1] -Inf -2.575829
```

```
c(rr2, Inf)
```

```
## [1] 2.575829 Inf
```

Our rejection regions are  $(-\infty, -2.58], [2.58, \infty)$ .

We don't reject the null hypothesis because  $Z_{obs} = 0.0462$  is not contained in the rejection region and so we fail to reject the null hypothesis.

We conclude that there is no statistically significant difference in vanillin levels between batch 1 and batch 2.

**Part e:**

We will construct a 99% 2-sided confidence interval for  $\mu_1 - \mu_2$ .

$$(\bar{x}_1 - \bar{x}_2) \pm z_{1-\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

```
z.quant = qnorm(1-alpha/2)
```

```
CIlower = (xbar1 - xbar2) - z.quant * sqrt(sigma21/n1 + sigma22/n2)
```

```
CIupper = (xbar1 - xbar2) + z.quant * sqrt(sigma21/n1 + sigma22/n2)
```

```
c(CIlower, CIupper)
```

```
## [1] -2.217657 2.298738
```

We are 99% confident that the true difference in mean vanillin levels between batches 1 and 2 is between -2.22 and 2.30 ppt.

Since 0 is contained within our confidence interval, we fail to reject the null hypothesis.

We conclude that there is no statistically significant difference in vanillin levels between batch 1 and batch 2.

**Part f:**

$$p = P(Z > |Z_{obs}|) * 2$$

```
Z.obs
```

```
## [1] 0.04624286
```

```
p = pnorm(Z.obs, lower.tail = FALSE) * 2
p
```

```
## [1] 0.9631167
```

Since our p-value wasn't less than  $\alpha = 0.01$ , we fail to reject the null hypothesis.

We conclude that there is no statistically significant difference in vanillin levels between batch 1 and batch 2.

**Part g:**

```
library(DescTools)
```

```
ZTest(batch1, batch2, alternative = "two.sided", mu = 0, sd_pop = 3.5, conf.level = 1-alpha)
```

```
##
## Two Sample z-test
##
## data: batch1 and batch2
## z = 0.046243, Std. Dev. Population = 3.5, p-value = 0.9631
## alternative hypothesis: true difference in means is not equal to 0
## 99 percent confidence interval:
## -2.217657 2.298738
## sample estimates:
## mean of x mean of y
## 39.54054 39.50000
```

**Problem 2:**

To see if population standard deviations are equal or not, we must test if  $\frac{\sigma_1}{\sigma_2} = 1$ . So we need a confidence interval, using the F distribution, of the ratio of the variances and see if it contains 1. If it does, then we can say that the population standard deviations are equal. If it does not, then standard deviations are not equal.

```

sigma21 = var(batch1)
sigma22 = var(batch2)

n1 = length(batch1)
n2 = length(batch2)

alpha = 0.01

Flower = qf(alpha/2, n1 - 1, n2 - 2)
Fupper = qf(1-alpha/2, n1 - 1, n2 - 2)

CIlower = (sigma21/sigma22)/Fupper
CIupper = (sigma21/sigma22)/Flower

c(CIlower, CIupper)

```

```
## [1] 0.3236509 2.2221637
```

Since the confidence interval contains 1, we will have to repeat question 1 assuming the population standard deviations are unknown, but equal.

**Part a:** We will be using a t statistic.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where

$$s_p = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

**Part b:** Under the null hypothesis, the sampling distribution of this test statistic is  $t \sim t_{n_1+n_2-2}$ .

**Part c:**

```

xbar1 = mean(batch1)
xbar2 = mean(batch2)

n1 = length(batch1)
n2 = length(batch2)

sigma21 = var(batch1)
sigma22 = var(batch2)

spooled = sqrt(((n1 - 1)*sigma21 + (n2 - 1)*sigma22)/(n1+n2-2))

t.obs = (xbar1 - xbar2) / (spooled * sqrt((1/n1) + (1/n2)))
t.obs

```

```
## [1] 0.04458559
```

Observed test statistic is  $t_{obs} = 0.0446$ .

**Part d:**

```
alpha = 0.01

rr1 = qt(alpha/2, n1 + n2 - 2)
rr2 = qt(1-alpha/2, n1 + n2 - 2)

rr1
```

```
## [1] -2.656145
```

```
rr2
```

```
## [1] 2.656145
```

```
c(-Inf, rr1)
```

```
## [1] -Inf -2.656145
```

```
c(rr2, Inf)
```

```
## [1] 2.656145 Inf
```

Our rejection regions are  $(-\infty, -2.66], [2.66, \infty)$ .

We don't reject the null hypothesis because  $t_{obs} = 0.0446$  is not contained in the rejection region and so we fail to reject the null hypothesis.

We conclude that there is no statistically significant difference in vanillin levels between batch 1 and batch 2.

#### Part e:

We will construct a 99% 2-sided confidence interval for  $\mu_1 - \mu_2$ .

$$(\bar{x}_1 - \bar{x}_2) \pm t_{(1-\alpha/2, n_1+n_2-2)} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

```
t.quant = qt(1-alpha/2, n1 + n2 - 2)

spooled = sqrt(((n1 - 1)*sigma21 + (n2 - 1)*sigma22)/(n1+n2-2))

CIlower = (xbar1 - xbar2) - (t.quant * spooled * sqrt((1/n1) + (1/n2)))
CIupper = (xbar1 - xbar2) + (t.quant * spooled * sqrt((1/n1) + (1/n2)))

c(CIlower, CIupper)

## [1] -2.374625 2.455706
```

We are 99% confident that the true difference in mean vanillin levels between batches 1 and 2 is between -2.37 and 2.46 ppt.

Since 0 is contained within our confidence interval, we fail to reject the null hypothesis.

We conclude that there is no statistically significant difference in vanillin levels between batch 1 and batch 2.

#### Part f:

$$p = P(t > |t_{obs}|) * 2$$

```
Z.obs
```

```
## [1] 0.04624286
```

```
p = pnorm(t.obs, lower.tail = FALSE) * 2  
p
```

```
## [1] 0.9644376
```

Since our p-value wasn't less than  $\alpha = 0.01$ , we fail to reject the null hypothesis.

We conclude that there is no statistically significant difference in vanillin levels between batch 1 and batch 2.

**Part g:**

```
library(DescTools)
```

```
t.test(batch1, batch2, alternative = "two.sided", mu = 0, var.equal = TRUE, conf.level = 1-alpha)
```

```
##
```

```
## Two Sample t-test
```

```
##
```

```
## data: batch1 and batch2
```

```
## t = 0.044586, df = 63, p-value = 0.9646
```

```
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 99 percent confidence interval:
```

```
## -2.374625 2.455706
```

```
## sample estimates:
```

```
## mean of x mean of y
```

```
## 39.54054 39.50000
```