Week 10 10.2 Note

Lecture 30: 1 Sample Z and T Tests

---

1. The Z-Test

- The Z-Test:
- 1 sided:
  - H: Hypothesis
    - H0, population mean = c vs. H1, population mean < c (or > c).
  - A: Assumptions
    - Sample is random.
    - Population is Normal. / Sample size is large enough (for the CLT).
    - Population SD is known.
  - T: Test Statistic
    - $$\text{Test statistic} = \frac{\text{observed mean (OV) - expected mean (EV)}}{\text{SD}/\sqrt{n}}$$
  - P: p-value
    - Use *normal curve* to find *tail area* for observed test statistic.
  - C: Conclusion
    - Retain or reject H0.
- 2 sided:
  - H: Hypothesis
    - H0, population mean = c vs. H1, population mean =/ c.
  - A: Assumptions
    - Same as 1 sided one.
  - T: Test Statistic
    - Same as 1 sided one.
  - P: p-value
    - Use *normal curve* to find *2 tail areas* for observed test statistic.
  - C: Conclusion
    - Same as 1 sided one.
- Comparing Z-Test with the Proportion Test:
  - Both tests use a Z-test based on the normal model. However, **this Z-test is used for continuous data** (*time*), whereas **the one last week was binary** (*success or failure*).
  - The Z-test can only apply to **binary data** when there is **a large sample size** *allowing CLT*. If the sample proportion is not normal, we may have to *bootstrap the distribution* based on the same proportion.
  - The SE of this one is for the sample **mean**, whereas the one last week was for the sample **sum**.

2. The T-Test

- When we don't know the population SD:
  - We need to know the *population SD* to use the Z-Test, and this is often *not known*!
  - Solution 1: **Estimate the population SD from the sample SD** and use the Z-Test.
    - **This estimation will add extra variability** to the test statistic, as the sample SD varies from sample to sample. (Large sample? OK. Small sample? No!)
    - For *larger samples*, **the difference between the population SD and sample SD should be small**, so this may be appropriate.
    - For *smaller samples*, the difference is noticeable, so we should use the T-Test.
  - Solution 2: Use the **T-Test**
- The T-Test:
  - 1 sided:
    - H: Hypothesis
      - H0, population mean = c vs. H1, population mean < c (or > c).
    - A: Assumptions
      - Sample is random.
      - Population is Normal. / Sample size is large enough (for the CLT).
    - T: Test Statistic
      - $\text{Test statistic} = \dfrac{\text{observed mean - population mean}}{\text{sample SD}/\sqrt{n}}$
    - P: p-value
      - Use *t(n-1) curve* to find *tail area* for observed test statistic.
      - Instead of using `1 - pnorm(T)` for 1 sided alternative, and `2 * (1 - pnorm(T))` for 2 sided alternative, **we use `1 - pt(T, n - 1)` and `2 * (1 - pt(T, n - 1))` to calculate P-value.** (`pt()`: T-distribution; T: Test Statistics; n - 1: sample size - 1 (also known as `df`, "degree of freedom"))
    - C: Conclusion
      - Retain or reject H0.
  - 2 sided:
    - In the next lecture.
- Speedy way in R:
  - The `t.test` command calculates the test statistic and p-value.
  - Example:

```
t.test(caf0, mu = 45)
```

```
##
##  One Sample t-test
##
## data:  caf0
## t = 0.34592, df = 8, p-value = 0.7383
## alternative hypothesis: true mean is not equal to 45
## 95 percent confidence interval:
##  36.84067 56.03933
## sample estimates:
## mean of x
##     46.44
```

36.8 ~ 56.0
(In this range)
CI tells all the Hos that will be valid.
Our Ho: 45
We retain Ho (valid)

Note:

- The 95% CI is equivalent to a 2 sided Test with $\alpha = 0.05$.
- As 45 falls inside the CI hence we retain Ho.

3. The Paired T-Test

- Difference between 0 and 13 caffeine levels:
  - We now **focus on the difference** in time (minutes) to exhaustion for 0 and 13 mg caffeine per kg body weight.
  - This sample of differences is now our focus:

```
caf0 = c(36.05,52.47,56.55,45.2,35.25,66.38,40.57,57.15,28.34)
caf13 = c(37.55,59.3,79.12,58.33,70.54,69.47,46.48,66.35,36.2)
cafdiff = caf13-caf0  # This 1 sample is our focus: the "differences".
mean(cafdiff)
```

```
## [1] 11.70889
```

```
sd(cafdiff)
```

```
## [1] 10.79987
```

```
t.test(cafdiff, mu = 0)
```

Ho in CI? No!

We reject Ho

```
##
##  One Sample t-test
##
## data:  cafdiff
## t = 3.2525, df = 8, p-value = 0.01166
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   3.407372 20.010405
## sample estimates:
## mean of x
##  11.70889
```

< 0.05

- Summary of T-Test:

H

$$H_0 \ : \text{mean exhaustion time for the differences} = 0 \text{ (mins)}$$

$$H_1 \ : \text{mean exhaustion time for the differences} \neq 0$$

A

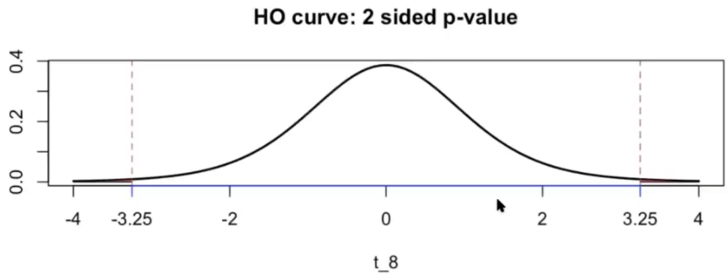
- We assume the sample of cyclists is random - they are all independent (eg none related to each other).
- We assume the population of differences is Normal.

T

```
teststat2= (mean(cafdiff)-0)/(sd(cafdiff)/sqrt(length(cafdiff)))
teststat2
```

```
## [1] 3.252508
```

P




```
2*pt(teststat2,8,lower.tail=F) # 2 tailed
```

```
## [1] 0.01165724
```

C

- As the p-value is so small, we reject Ho and conclude that "there is a difference" - ie caffeine consumption affects endurance.

- Note on SD vs. SE:
  - The SD tells us **how far each individual varies from the sample mean** of 9 cyclists.
  - The SE tells us **how far the sample means vary from the true population mean**, for all elite cyclists.
  - The sample size of 9 may not indicate normal distribution of sample means under the CLT. If we use the normal model to construct a confident interval, we need to assume that the distribution of the box (population for each measurement) is normal.
    - *Extension* - Calculating the CI by hand:

```
# T value for 95% CI (ie upper tail is 5%/2 = 0.025)
tval = qt(0.025,8,lower.tail=F)
tval
```

```
## [1] 2.306004
```

```
mean(cafdiff)
```

```
## [1] 11.70889
```

```
sd(cafdiff)
```

```
## [1] 10.79987
```

```
c(mean(cafdiff) - tval * sd(cafdiff)/sqrt(9),mean(cafdiff) + tval * sd(cafdiff)/sqrt(9))
```

```
## [1]  3.407372 20.010405
```

Lecture 31: 2 Sample T Tests

1. Inference: The 2 Sample T-Test

> Inference: Making a *decision* about population parameter(s) based on a *sample*. How would we go about this?

- HATPC:
    - H: Hypothesis
        - Let u1 = mean heart rate of our control (no Red Bull)
        - Let u2 = mean heart rate of our treatment (Red Bull)
        - H0, there is no difference: u1 = u2, or u1 - u2 = 0.
        - H1, there is a difference: u1 =/ u2, or u1 - u2 =/ 0.
    - A: Assumptions
        - The 2 samples are **independent**.
            - The 2 samples contain different people.
        - The 2 populations have **equal spread** (SD or variance).
            - We assume that the 2 populations have the same variation in heart rate.
        - The 2 populations are **normal**.
    - T: Test Statistic (for 2 sample T test)
        - > We compare 2 populations, so the observed test statistic is
          >
          > $$\text{test statistic} = \frac{\text{OV - EV}}{\text{SE}} = \frac{\underbrace{\boxed{\bar{x}_1 - \bar{x}_2}}_{\text{difference of 2 samples}} - \boxed{0}}{\text{SE}}$$
          >
          > We expect "no difference"
          >
          > where
          >
          > $$SE = \sqrt{\mathrm{SD}_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \qquad df = n_1 + n_2 - 2$$
          >
          > based on the pooled SD, where $\mathrm{SD}_p^2 = \frac{(n_1-1)\mathrm{SD}_1^2 + (n_2-1)\mathrm{SD}_2^2}{n_1+n_2-2}$
    - P: p-value
        - Using `t.test`:
          The test statistic and P-value can be obtained using `t.test` in one go.

```
t.test(No_RB, RB, var.equal = T)
```

```
##
##  Two Sample t-test
##
## data:  No_RB and RB
## t = -1.5418, df = 24, p-value = 0.1362
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -13.892538   2.011586
## sample estimates:
## mean of x mean of y
##  73.64286  79.58333
```

          Note the argument `var.equal = T` indicates that we are doing a two sample T test assuming equal population variances.
    - C: Conclusion
        - Statistical Conclusion:
            - As the p-value > 0.05, we retain the H0.
            - The data is consistent with hypothesis that the mean heart rates are equal.
        - Scientific Conclusion:
            - The data suggests that the consumption of Redbull by university students does not have an effect on heart rate.

2. Checking Assumptions for 2 Sample T-Test

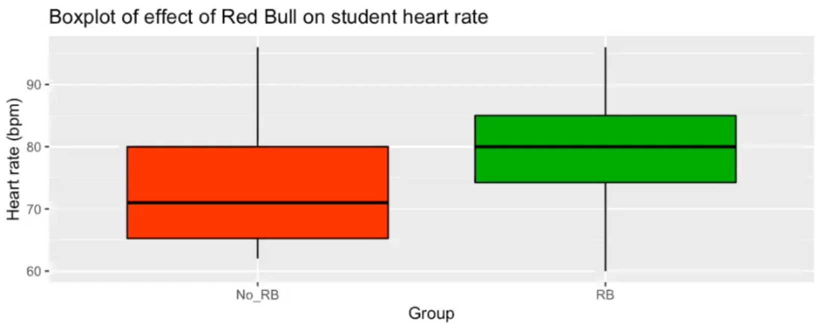
- A: Assumptions

- A1: The 2 samples are independent.
  - Check: Context.
- A2: The 2 populations have equal spread (SD or variance).
  - Check: Boxplots, Histograms, Variance Test
- A3: The 2 populations are Normal.
  - Check: Boxplots (expect no or few outliers), Histograms, QQ-Plots, Normality Test

- Various Tests:

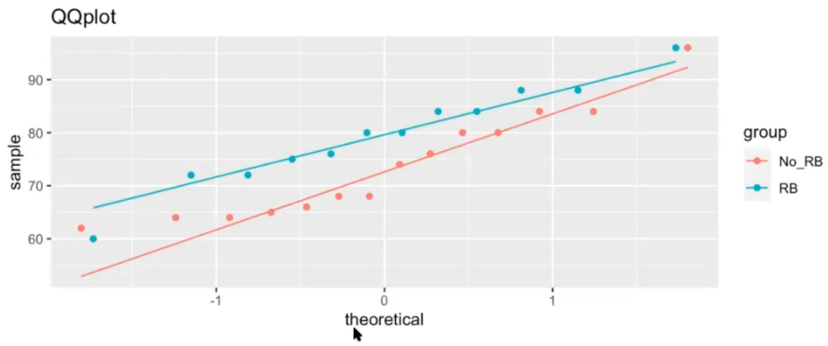
Comparative Boxplots - for normality and equality of variance assumptions

- Normality: the 2 samples look symmetrical and so are consistent with Normality, although the control one looks a bit asymmetric.
- Equality of variance: the 2 samples have similar spread and so are consistent with the equality of variance assumption.



Q-Q Plot - for Normality

- A quantile-quantile plot (Q-Q Plot) graphs the theoretical quantiles based on the normal curve against the actual quantiles. If the line formed by the points is reasonably straight, then we can safely assume that the data is normally distributed. Both plots (red = no Redbull, green = Redbull) show a reasonably straight line.




```
require(ggplot2)
p3 = ggplot(RB_data, aes(sample = rate, colour = group)) +
  stat_qq() + stat_qq_line() + ggtitle("QQplot")
p3
```

Shapiro-Wilk Test - for Normality

- The Shapiro-Wilk test tests the null hypothesis that the data is Normal.
  - H_0 : The data is consistent with Normal
  - H_1 : The data is not consistent with Normal
- While useful, this test is very sensitive to sample size (for example, small samples will almost always be retained as Normal). Therefore, it is recommended to use this test in conjunction with graphical methods.

```
shapiro.test(No_RB)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  No_RB
## W = 0.90604, p-value = 0.138
```

```
shapiro.test(RB)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  RB
## W = 0.97459, p-value = 0.9524
```

- Here, both tests give large p-values (p-value>0.05), so considering the previous plots, we conclude that the populations could be Normally distributed.

Levene's Test (F-Test) - for equal spread

- The F-Test tests the null hypothesis that the spread is equal between the two populations.
- Let σ_1^2 = variance of the Control group (No Redbull)
- Let σ_2^2 = variance of the Treament group (Redbull)
  - H_0 : There is no difference: $\sigma_1^2 = \sigma_2^2$
  - H_1 : There is a difference: $\sigma_1^2 \neq \sigma_2^2$
- If the variation is very different between the 2 samples, we may increase the chance of falsely rejecting the H_0 (called type I error).

```
var.test(No_RB,RB)
```

```
##
##  F test to compare two variances
##
## data:  No_RB and RB
## F = 1.1357, num df = 13, denom df = 11, p-value = 0.8428
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.334832 3.631266
## sample estimates:
## ratio of variances
##           1.135659
```

The F-test gives p-value > 0.05 suggesting that data is consistent with equal variance.

- Overall Conclusion about Assumptions:
  - A2: The 2 populations have equal spread (SD or variance).
    - The boxplots suggest that the variation is similar between 2 populations, as is confirmed by the F-test.
  - A3: The 2 populations are normal.
    - The 2 samples suggest tht 2 populations could be normal, from both the boxplots, the Q-Q Plot and the Shapiro-Wilk Test.
  - Hence, we proceed with a 2 Sample T Test (assuming equal variance).

3. Other Tests

- Assumptions not met?

- In this case we use the argument `var.equal = FALSE`. This performs the Welch 2 Sample T Test. The p-value is > 0.05 so our conclusion concurs with the previous 2 Sample T Test for equal variances.

```
t.test(No_RB, RB, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  No_RB and RB
## t = -1.6111, df = 19.747, p-value = 0.123
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -21.180686   2.728887
## sample estimates:
## mean of x mean of y
##  70.35743  79.58333
```

A lot of more tests with different assumptions to deal with the following situations:

Data	Appropriate Test(s)
2 samples have unequal variance	Welch 2 Sample T Test
2 samples suggest Non-Normality	Transformations or Non Parametric Tests
2 samples are matched pairs	Paired T-Test (using sample of differences)

- Unequal Variance - Welch 2 Sample T Test:

  Suppose the variance was much larger in the group who have not had caffeine (Redbull).

```
set.seed(10)
No_RB <- No_RB + rnorm(14, 0, 20)  # New simulated version of No-RB with larger variance
var.test(No_RB, RB)
```

```
##
##  F test to compare two variances
##
## data:  No_RB and RB
## F = 3.9709, num df = 13, denom df = 11, p-value = 0.02813
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##   1.170767 12.697012
## sample estimates:
## ratio of variances
##           3.970923
```

  Now we see that the p-value from the variance test is < 0.05. Therefore we reject the null hypothesis and question whether the variances of the 2 populations are equal.

- Non-Normality - Mann-Whitney-Wilcoxon Test:
  - Dealing with non-Normal data is an area hotly debated in the stats world.
  - One option is to transform non-Normal data (eg log or square root) before performing a T Test.
  - An alternative, particularly for smaller datasets, is to use non-parametric tests such as the Mann-Whitney-Wilcoxon test: `wilcox.test`.

- Non-Independent Data - Paired T Test:
  - Sometimes it is desirable to analyse **dependent** data. We often design an experiment to take advantage of this dependency in order to control variation between experimental groups.
  - Suppose we measure the heart rate of 12 students **before** and **after** they consume a Redbull can, like the elite cyclists example. We call this **paired data** because we have one experimental unit (a student) with two treatments (heart rate before and after RedBull), as seen in previous lecture.
  - We can record the data as follows:

StudentdID	1	2	3	4	5	6	7	8	9	10	11	12
Before_Redbull	84	76	68	80	64	62	74	84	68	96	80	64
After_Redbull	72	88	72	88	76	75	84	80	60	96	80	84

  - Calculate the **differences** between `Before_Redbull - After_Redbull`.

```
diff = RB  - No_RB
diff
```

```
## [1] -12  12   4   8  12  13  10  -4  -8   0   0  20
```

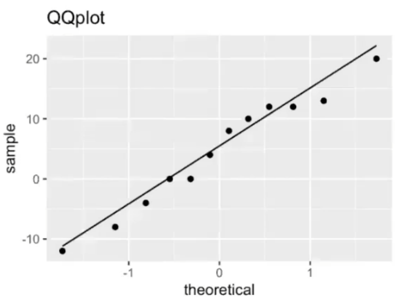
  - The Paired T Test is just a 1 Sample T Test on the differences!

H

- H_0 : The difference in means is zero : $\mu_d = 0$
- H_1 : The difference in means is not zero : $\mu_d \neq 0$

A

We assume that the population of differences is Normally distributed.



- The QQ-plot looks reasonably linear suggesting the data is reasonably Normally distributed.

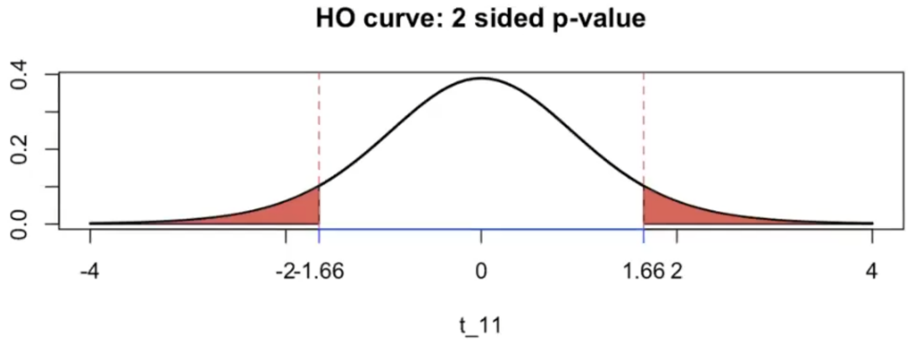
T

Use the `t.test` function with the argument `paired = T`.

```
No_RB <- c(84, 76, 68, 80, 64, 62, 74, 84, 68, 96, 80, 64)
RB <- c(72, 88, 72, 88, 76, 75, 84, 80, 60, 96, 80, 84)
t.test(No_RB, RB, paired = T)
```

P

```
##
##  Paired t-test
##
## data:  values by ind
## t = -1.6578, df = 11, p-value = 0.1256
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -10.668294   1.501628
## sample estimates:
## mean of the differences
##               -4.583333
```



C

Statistical conclusion:

- As the p-value > 0.05, we retain the null hypothesis.
- Note also that that the 95% Confidence Interval for the mean difference is (-10.668294,1.501628) which contains the null hypothesis value 0.)
- Thus the data is consistent with the hypothesis that the difference in mean heart rates is 0.

Scientific conclusion:

- The data suggests that the consumption of Redbull by university students does not have an effect on heart rate.