

CA05-STAT100

Fábio Rodrigues Pereira

Oppgave 0

Se på løsningsforsalget til modul 7, diskuter tre ting dere ikke fikk til helt eller forstod fra forrige uke, som dere nå skjønner litt bedre. Skriv ned hvilke tema som ble diskutert og gjerne en setning av hva dere kom frem til.

1d) $P(\text{type 1 error}) = P(\text{reject } H_0 \mid H_0 \text{ is true}) \leq 0,05 \quad \square$

1e) H_0 and H_1 are claims not stochastic variables and there have observations which we are interested in. \square

1f) \rightarrow Since $\bar{x} = 17,5$, then we don't reject H_0 . \square

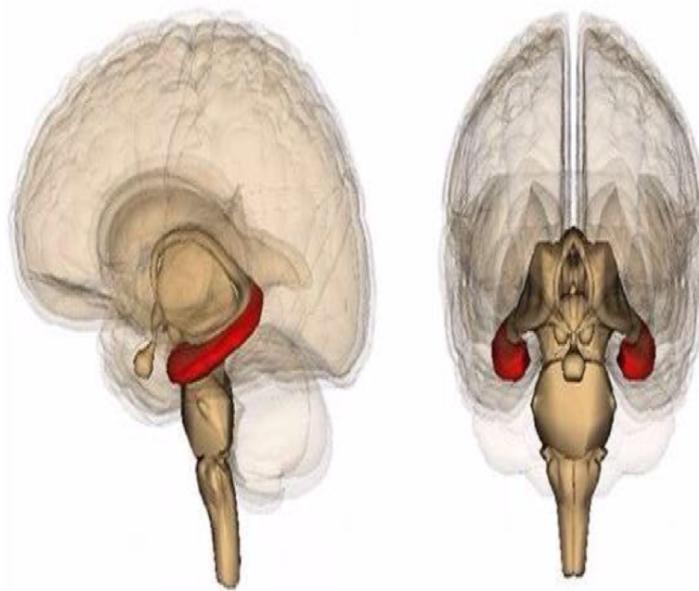
1m) Note $KI\ 95\% = \bar{x} \pm T_{24, 0,025} \cdot \frac{\sqrt{m}}{SD}$, but $KI\ 95\% = \bar{x} \pm t_{m-1, \frac{\alpha}{2}} \cdot \frac{SD}{\sqrt{m}}$ \square

1n) \rightarrow Since H_0 value for μ_0 is covered by $KI\ 95\%$, then we can not reject H_0 . This is the same for two-sided test at $\alpha = 0,05$. \square

Oppgave 1

Are there physiological indicators associated with schizophrenia? In a 1990 article, researchers reported the results of a study that controlled for genetic and socioeconomic differences by examining 15 pairs of identical twins, where one of the twins was schizophrenic and the other not. The researchers used magnetic resonance imaging to measure the volumes (in cubic centimeters) of several regions and subregions inside the twins' brains. The following data came from one of the subregions, the left hippocampus:

Pair	Unaffectionate	Affectionate
1	1.94	1.27
2	1.44	1.63
3	1.56	1.47
4	1.58	1.39
5	2.06	1.93
6	1.66	1.26
7	1.75	1.71
8	1.77	1.67
9	1.78	1.28
10	1.92	1.85
11	1.25	1.02
12	1.93	1.34
13	2.04	2.02
14	1.62	1.59
15	2.08	1.97



- i) What is the magnitude of the difference in the volumes of the left hippocampus between (all) unaffected and affected individuals?

R-Notebook:

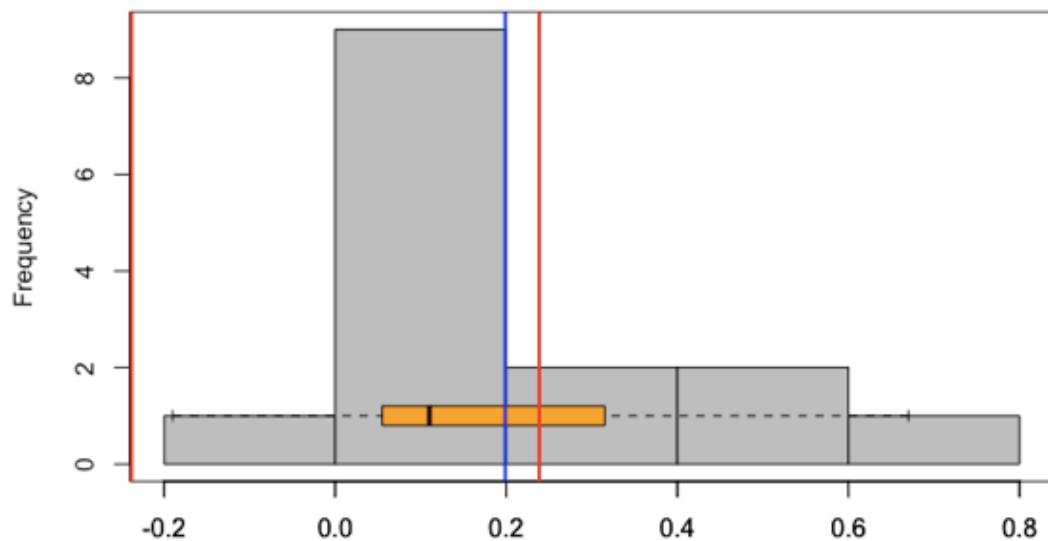
```
1- ---
2 title: "STAT100 - Compulsory Assignment 05 - Fabio Rodrigues Pereira"
3 output: html_notebook
4 ---
5 ````{r}
6 # Create Variables:
7 Unaffected <- c(1.94, 1.44, 1.56, 1.58, 2.06, 1.66, 1.75, 1.77, 1.78, 1.92, 1.25, 1.93, 2.04, 1.62, 2.08)
8 Affected <- c(1.27, 1.63, 1.47, 1.39, 1.93, 1.26, 1.71, 1.67, 1.28, 1.85, 1.02, 1.34, 2.02, 1.59, 1.97)
9 D <- Unaffected - Affected
10
11 # Create data frame with the variables:
12 schizophrenia <- data.frame(Unaffected, Affected, D)
13
14 # Show dimension and data frame:
15 cat("Dimension (size, variables): ", dim(schizophrenia))
16 cat("\n\n")
17 schizophrenia
18
19 ````
```

Dimension (size, variables): 15 3

	Unaffected	Affected	D
1	1.94	1.27	0.67
2	1.44	1.63	-0.19
3	1.56	1.47	0.09
4	1.58	1.39	0.19
5	2.06	1.93	0.13
6	1.66	1.26	0.40
7	1.75	1.71	0.04
8	1.77	1.67	0.10
9	1.78	1.28	0.50
10	1.92	1.85	0.07
11	1.25	1.02	0.23
12	1.93	1.34	0.59
13	2.04	2.02	0.02
14	1.62	1.59	0.03
15	2.08	1.97	0.11

```
20 ````{r}
21 # Histogram and boxplot together:
22 hist(schizophrenia$D, main="Distribution of the variable D and its mean (blue) and SD (red) lines:",
23       xlab="Difference in the volumes between Unaffected and Affected individuals", col="grey")
24 boxplot(schizophrenia$D, horizontal = TRUE, col="orange", add=TRUE)
25
26 # Mean line:
27 abline(v=mean(schizophrenia$D), col="blue", lw=2)
28 # 2*SD lines:
29 abline(v=sd(schizophrenia$D), col="red", lw=2)
30 abline(v=-1*sd(schizophrenia$D), col="red", lw=2)
31 ````
```

Distribution of the variable D and its mean (blue) and SD (red) lines:



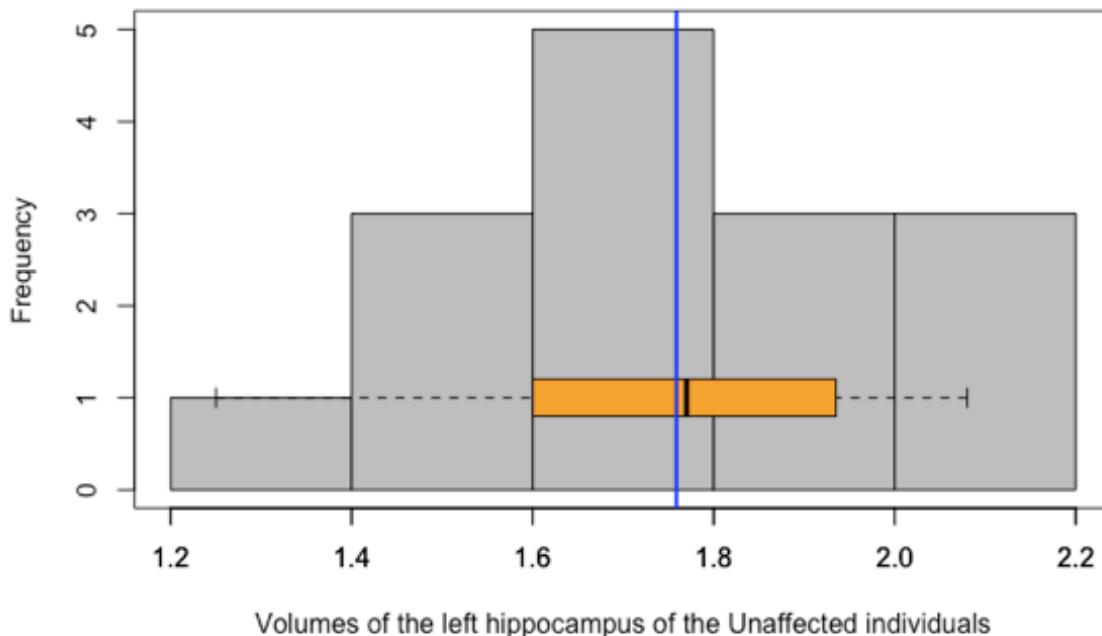
Difference in the volumes between Unaffected and Affected individuals

```

32 # Histogram and boxplot together:
33 hist(schizophrenia$Unaffected, main="Distribution of the variable Unaffected and its mean (blue) and SD (red) lines:", xlab="Volumes of the left hippocampus of the Unaffected individuals", col="grey")
34 boxplot(schizophrenia$Unaffected, horizontal = TRUE, col="orange",add=TRUE)
35
36 # Mean line:
37 abline(v=mean(schizophrenia$Unaffected), col="blue",lw=2)
38 # 1*SD lines:
39 abline(v=sd(schizophrenia$Unaffected), col="red",lw=2)
40 abline(v=-1*sd(schizophrenia$Unaffected),col="red",lw=2)
41 ``

```

Distribution of the variable Unaffected and its mean (blue) and SD (red) line

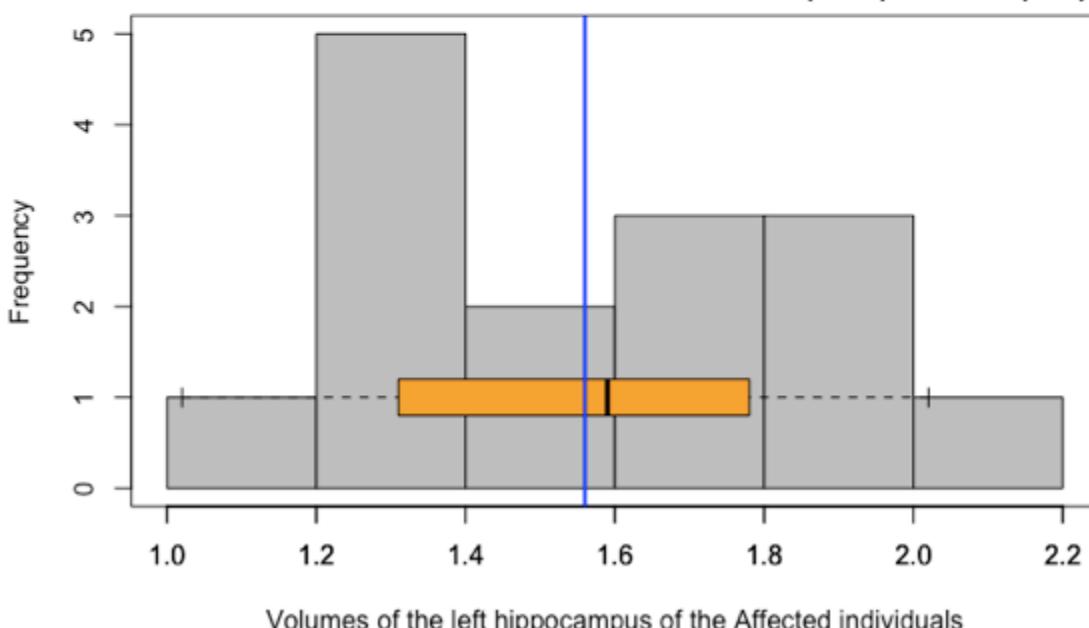


```

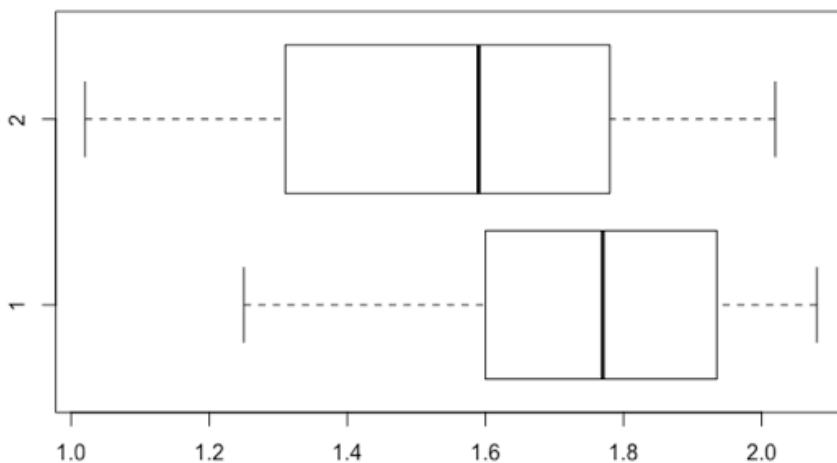
42 ``
43 # Histogram and boxplot together:
44 hist(schizophrenia$Affected, main="Distribution of the variable Affected and its mean (blue) and SD (red) lines:", xlab="Volumes of the left hippocampus of the Affected individuals", col="grey")
45 boxplot(schizophrenia$Affected, horizontal = TRUE, col="orange",add=TRUE)
46
47 # Mean line:
48 abline(v=mean(schizophrenia$Affected), col="blue",lw=2)
49 # 1*SD lines:
50 abline(v=sd(schizophrenia$Affected), col="red",lw=2)
51 abline(v=-1*sd(schizophrenia$Affected),col="red",lw=2)
52 ``

```

Distribution of the variable Affected and its mean (blue) and SD (red) lines



```
53 - ````{r}
54 # Plot combined boxplots between Unaffected and Affected:
55 boxplot(schizophrenia$Unaffected, schizophrenia$Affected, horizontal=TRUE)
56 ````
```



```
57 - ````{r}
58 # Descriptive Statistics:
59 summary(schizophrenia)
60 ````
```

Unaffected	Affected	D
Min. :1.250	Min. :1.02	Min. :-0.1900
1st Qu.:1.600	1st Qu.:1.31	1st Qu.: 0.0550
Median :1.770	Median :1.59	Median : 0.1100
Mean :1.759	Mean :1.56	Mean : 0.1987
3rd Qu.:1.935	3rd Qu.:1.78	3rd Qu.: 0.3150
Max. :2.080	Max. :2.02	Max. : 0.6700

Dataene i eksemplet over er hentet fra en tvillingstudie og effekten av schizophreni på størrelsen på en hjernedel som heter Hippocampus. Denne delen av hjernen er sentral for å huske koblinger mellom steder man har vært og hendelser man har opplevd. Den er nyttig for å holde orden på ting, rett og slett. 15 tvillingpar hvor én tvilling har schizophreni, den andre ikke, ble scannet med såkalt fMRI og størrelsen på deres hippocampus ble målt. Man var ute etter å se på om det er noen forskjell i forventet hippocampusstørrelse mellom syke og friske.

Merk at Det etterhvert kan bli tidkrevende (og dønn kjedelig) å bruke lommeregner. Se oppgave H hvis du vil unngå noe av dette. Dataene ligger også vedlagt nederst i filen: Tabell 1.

- A) La X_i være størrelsen av Hippocampus hos frisk (unaffected) og Y_i størrelsen av Hippocampus hos syk (affected) i et tilfeldig tvillingpar nr i . Det er rimelig å anta at både X og Y er normalfordelte.

(i) Sett opp sannsynlighetsmodeller for X og Y separat.

(ii) Prøv å forklar hvorfor X og Y er avhengige av hverandre innen tvillingpar, men at ulike tvillingpar kan betraktes som uavhengige.

$$\hookrightarrow X_i := \text{Unaffected} ; Y_i := \text{Affected} ; i = 1, \dots, 15 = m ;$$

$$\hookrightarrow X_i \sim N(\mu_X, \sigma_X) \text{ and } Y_i \sim N(\mu_Y, \sigma_Y) \text{ and } D_i \sim N(\mu_D, \sigma_D) ;$$

(iii) $\rightarrow X$ and Y are dependent between twin siblings because they are supposed to have quite similar genetics. On the other hand, X and Y that are not twin siblings are supposed to have different genetics, then independents. \square

- B) For å håndtere denne avhengigheten kan vi betrakte dette som parvis data og se på differansen. Som er regnet ut i Tabell 1.

\hookrightarrow pairwise

$$\text{La } D_i = X_i - Y_i,$$

- (i) Hva uttrykker D_i ? \rightarrow Difference in the volumes between Unaffected and Affected twin siblings. //
- (ii) Sett opp en modell for D_i .
- (iii) Hvordan tolker du de to parameterne som inngår i denne modellen?
- (iv) Estimer begge disse parameterne (bruk dataene i tabellen nedenfor i den grad du finner det nødvendig.)
- (v) Hva blir standardfeilen til estimatet for forventningen?

(vi) \rightarrow Since the variables are paired, then $D_i = X_i - Y_i$, where $D_i = \mu_D + E_i$ where $E_i \sim N(0, \sigma_D)$ for $i = 1, \dots, 15$. //

(vii) μ_D := overgangsmean of D_{15} ; σ_D := standard deviation or dispersion of the outcomes from the mean. //

(viii) \rightarrow We estimate $\hat{\mu}_D$ using the unbiased estimator \bar{D} , such that:

$$\hookrightarrow \hat{\mu}_D = \frac{1}{m} \cdot \sum_{i=1}^m D_i = 0,1986667 \text{ for } m=15;$$

We estimate $\hat{\sigma}_D$ using the unbiased estimator $\frac{SD_D}{\sqrt{m}}$, such that:

$$\hookrightarrow \hat{\sigma}_D = SD[\bar{D}] = \frac{1}{m-1} \cdot \sum_{i=1}^m (D_i - \bar{D})^2 = 0,23 \text{ and } SE[\hat{\mu}_D] = \frac{SD[\bar{D}]}{\sqrt{m}} = \frac{0,23}{\sqrt{15}} \approx 0,06 \quad \square$$

C) Her må vi bruke t-fordeling i stedet for standard normalfordeling når vi skal lage konfidensintervaller og teste hypoteser.

i) Hvorfor?

ii) Hva skiller t-fordeling fra en standard normalfordeling?

iii) Prøv å forklar på en muntlig måte hvorfor t-fordelingen har tyngre haler enn normalfordelingen.

i) We must use T-student because the population standard deviation is unknown. //

ii) T-student has degrees of freedom to reduce the deviation of the sample parameter. //

iii) T-student has wider curve than normal distribution, but as $n \rightarrow \infty$ as T-student \rightarrow normal distribution. //

D) Finn et 95% KI for forventet forskjell mellom gruppene.

i) Gi en tolkning av intervallet.

ii) Hvorfor er det interessant å se om intervallet dekker 0?

$$KI 95\% = \bar{D} \pm t_{14, 0,025} \cdot SE[\bar{D}] = 0,1986667 \pm 2,145 \cdot 0,06152713 \stackrel{\text{approx}}{=} [0,0667041; 0,3306424], //$$

iii) We are 95% sure that the average of the difference (\bar{D}) is not inside this interval. //

iv) It is interesting to see that the interval covers 0 because the hypothesis of the claim $H_0: \mu_D = 0$ is plausible. Then we could infer that we were 95% sure that there was no difference between the volumes. //

Ekstra: Hvilke faktorer avgjør om intervallet vil dekke 0?

↳ I did not understand the question. Is not the same as iii)?

E) Hvordan vil du teste om det er forskjell mellom gruppene?

Sett opp hypoteser for denne testen.

$$H_0: \mu_x = \mu_y \text{ or } \mu_D = 0 \quad \text{vs} \quad H_1: \mu_x \neq \mu_y \text{ or } \mu_D \neq 0 //$$

Forklar hvordan du kan bruke konfidensintervallet til å konkludere.

Hvilken signifikansnivå tester du i så fall på?

↳ Since we formulate a two-sided test, then the confidence interval can be used to determine if the claim H_0 should be rejected. The significance level is $\alpha/2 = 0,025$. Since the KI 95% does not cover H_0 , then H_0 should be rejected. //

F) Utfør også en ensidig hypotesetest på 5% nivå for å teste om forventet størrelse på hippocampus er mindre for syke enn for friske. Hva blir hypotesene i dette tilfellet?

$$H_0: \mu_D = 0 \quad \text{vs} \quad H_1: \mu_D > 0, \quad \bar{D} \sim T(\hat{\mu}_D = 0,2; \hat{\sigma}_{\bar{D}} = 0,23) \text{ for } i=1, \dots, 15 \text{ and } \alpha = 0,05; \quad T = \frac{\bar{D} - \mu_D}{SE} = \frac{0,2 - 0}{0,23/\sqrt{15}} = \frac{0,2}{0,06} = 3,33 //$$

↳ Since $T = 3,33 > 1,761$, then reject H_0 //

G) Hvorfor kan du nå ikke regne ut p-verdien til testen i E (uten datamaskin).

$$\text{p-value} = P\left(T \geq \frac{\bar{D} - \mu_D}{SE} \mid H_0 \text{ is true}\right) = P\left(T \geq \frac{0,1986667}{0,06152713} \mid H_0 \text{ is true}\right) = P\left(T \geq 3,2289 \mid H_0 \text{ is true}\right)$$

↳ because we don't have a table of two-sided cumulative distribution values of T-student. //

H) Nå vil det etter hvert bli grusomt tidkrevende og dønn kjedelig å bruke lommeregner. Analysene kan enkel gjøres i RStudio. Last ned schizophrenia.Rdata på egen maskin (du finner den på Canvas under modulen Data), og åpne den i Rstudio:

Finn fila i Canvas, trykk på den, deretter på **Download schizophrenia.Rdata**. Start Rstudio og åpne fila du nettopp lastet ned, ved å bruke menyen øverst til venstre: File – Open File – (browse) ‘schizophrenia.Rdata’ **Open**. Du vil få opp meldingen «Do you want to load the R data file into your global environment?», svar **Yes**.

I Console-vinduet kan du først skrive

```
View(schizophrenia)  
for å få se på datafila, og deretter  
?t.test
```

for å få se dokumentasjonen om t-testen som du kan bruke/gjøre i Rstudio. Noen synes denne dokumentasjonen er nyttig, andre finner mindre hjelp i den. Hvis du er en av de sistnevnte, er det fint om du spør en på kollokviegruppa di eller en hjelpearer om å forklare litt av dette for deg.

Kjør en parvis analyse:

```
t.test(schizophrenia$Unaffected, schizophrenia$Affected, paired=TRUE)
```

Lim inn utskriften i svaret deres og forklar hva de forskjellige tallene er.

(Dersom man ikke rekker dette på gruppa, bør man prøve selv etterpå.)

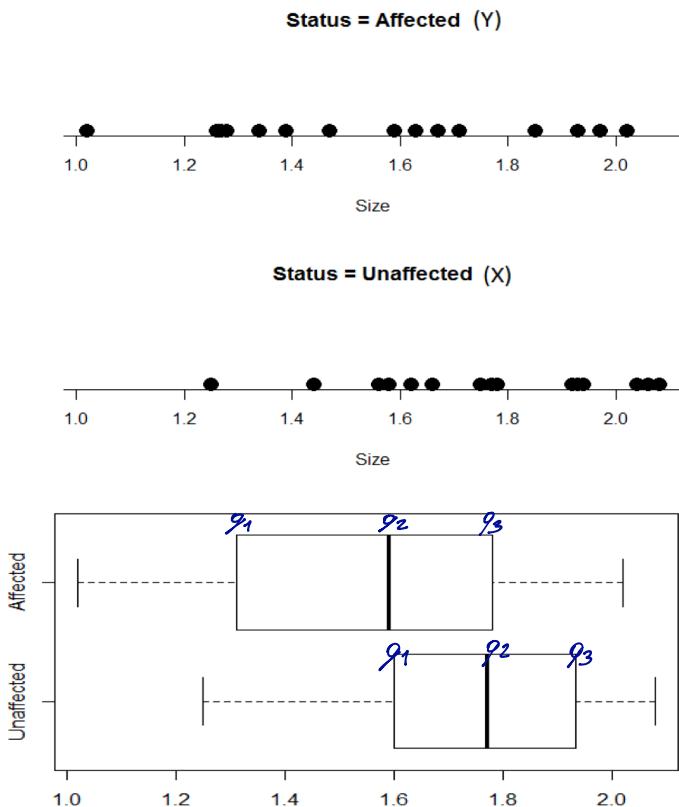
↳ Both, my calculations and R computations are the same.

```
61 - ````{r}  
62 # Variance and size of X:  
63 var_x <- var(schizophrenia$Unaffected)  
64  
65 # Variance and size of Y:  
66 var_y <- var(schizophrenia$Affected)  
67  
68 # Spooled variance of D:  
69 var_d <- var(schizophrenia$D)  
70  
71 # Sample Variance:  
72 cat("Sample Variance for Unaffected: ", var_x,  
    "\nSample Variance for Affected: ", var_y,  
    "\nSample Variance for pairwised D: ", var_d)  
73  
74 ````  
75  
    Sample Variance for Unaffected: 0.05876952  
    Sample Variance for Affected: 0.09075714  
    Sample Variance for pairwised D: 0.05678381  
76 - ````{r}  
77 # Point estimated for Standard Deviation of X:  
78 nx <- length(schizophrenia$Unaffected)  
79 std_x <- sqrt(var_x/nx)  
80  
81 # Point estimated for Standard Deviation of Y:  
82 ny <- length(schizophrenia$Affected)  
83 std_y <- sqrt(var_y/ny)  
84  
85 # Point estimated for Standard Deviation of pairwised D:  
86 nd <- length(schizophrenia$D)  
87 std_d <- sqrt(var_d/nd)  
88  
89 # Sample Standard Deviation:  
90 cat("Sample Standard Deviation for Unaffected: ", std_x,  
91     "\nSample Standard Deviation for Affected: ", std_y,  
92     "\nSample Standard Deviation/Error for pairwised D: ", std_d)  
93  
    Sample Standard Deviation for Unaffected: 0.06259368  
    Sample Standard Deviation for Affected: 0.07778481  
    Sample Standard Deviation/Error for pairwised D: 0.06152713  
94 - ````{r}  
95 t.test(schizophrenia$Unaffected, schizophrenia$Affected, mu=0, paired=TRUE, var.equal=FALSE,  
96 conf.level=0.95)  
97  
98  
Paired t-test  
data: schizophrenia$Unaffected and schizophrenia$Affected  
t = 3.2289, df = 14, p-value = 0.006062  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
 0.0667041 0.3306292  
sample estimates:  
mean of the differences  
 0.1986667
```



Vi antar nå at vi IKKE har noe opplysninger om at personene er i slekt. Totalt er det nå $n_1 = 15$ friske og $n_2 = 15$ syke helt tilfeldige personer som ble målt.

- I) Nedenfor er et såkalt dotplot laget i R Commander som viser de observerte størrelsene for de syke (affected) og for de friske (unaffected), og et boksplot av de samme dataene, laget i RStudio.



④ We see here that the volumes have wider dispersion when we compare with the plot below.

④ The dispersion of the dots are smaller.

④ we see that the first 50% of affected (up to g_2) covers only 25% of unaffected (up to g_1) and the final 50% of affected covers approx. 75% of unaffected, then not similar but with clearly differences. Therefore, their expected volumes might be different at a certain significant level. □

Ut fra kun å betrakte figurene, diskuter om det er rimelig å tro at forventet hippocampusstørrelse er forskjellig hos friske og syke (vi skal teste dette nedenfor).

- J) La X være hippocampusstørrelsen hos frisk (unaffected) og Y hippocampusstørrelsen hos syk (affected). Det er rimelig å anta at både X og Y er normalfordelte.

i) Sett opp sannsynlighetsmodeller for X og Y.

ii) Gi en tolkning av de 4 parametrerne

iii) Bruk dataene nederst i oppgaven eller Rstudio og fila schizophrenia.Rdata til å estimere de ukjente parametrerne.

iv) $X_i \sim T(\mu_x, \sigma_x)$ and $Y_i \sim T(\mu_y, \sigma_y)$;

v) μ_x and μ_y are population average of the volume of the hippocampus of each groups.

σ_x and σ_y are population standard deviation or dispersion from the mean of each groups.

$$X_i = \mu_x + \epsilon_i \text{ where } \epsilon_i \sim N(0, \sigma_x^2) \text{ for } i=1, \dots, 15; \hat{\mu}_x = \bar{X} = \frac{1}{m_x} \sum_{i=1}^{m_x} x_i \approx 1,75 \text{ and } \hat{\sigma}_x = \sqrt{\frac{1}{m_x-1} \sum_{i=1}^{m_x} (x_i - \bar{X})^2} \approx 0,24$$

$$Y_i = \mu_y + \epsilon_i \text{ where } \epsilon_i \sim N(0, \sigma_y^2) \text{ for } i=1, \dots, 15; \hat{\mu}_y = \bar{Y} = \frac{1}{m_y} \sum_{i=1}^{m_y} y_i \approx 1,56 \text{ and } \hat{\sigma}_y = \sqrt{\frac{1}{m_y-1} \sum_{i=1}^{m_y} (y_i - \bar{Y})^2} \approx 0,30$$

K) Hvordan kan vi lage et felles estimat for σ^2 basert på begge utvalgene?

- (i) Finn estimatet.
- (ii) Finn estimatet for σ .
- (iv) Kan du tenke deg hvorfor det er lurt å lage et slikt felles estimat (dersom antagelsen om samme standardavvik er rimelig)?
- (v) Gjør en to-utvalgs t-test for å teste om de to gruppene har samme forventning mht størrelsen på hippocampus. Sett opp hypoteser og utfør testen. (Evt kontrollér svaret vha RStudio.)
Bruk dataene nederst i oppgaven eller Rstudio og fila schizophrenia.Rdata til å utføre testen.

(i) A common variance can be done when we calculate its pooled value, such that:

$$(ii) S_p^2 = \frac{(n_x-1) \cdot S_x^2 + (n_y-1) \cdot S_y^2}{n_x+n_y-2} = \frac{(n-1)}{(n-1)} \cdot \frac{S_x^2 + S_y^2}{2} = \frac{0,05876952 + 0,09075714}{2} = 0,07476333 //$$

$$(iii) Spooled variance := \frac{(n_x-1) S_x^2 + (n_y-1) S_y^2}{n_x+n_y-2} = \hat{\sigma}_D^2 = 0,074 //$$

$$\hat{\sigma}_D = \sqrt{\hat{\sigma}_D^2} \approx \sqrt{0,074} \approx 0,272 // \text{ and } SE[\hat{D}] = \hat{\sigma}_D \cdot \sqrt{\frac{1}{n_x} + \frac{1}{n_y}} \approx 0,099 //$$

(iv) It is reasonable to calculate a new pooled parameter, because we can get more precise estimations for our models. //

(v) $H_0: \mu_{\bar{x}} = \mu_{\bar{y}} \text{ or } \mu_b = 0 \quad \forall b$ $H_1: \mu_{\bar{x}} \neq \mu_{\bar{y}} \text{ or } \mu_b \neq 0$

$$\hookrightarrow \bar{D} \sim T(\mu_D; S_p \cdot \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}) \sim T(0,198667; 0,0998421)$$

$$\hookrightarrow T = \frac{\bar{d} - 0}{SE[\bar{D}]} = \frac{0,198667}{0,0998421} = 1,989812 // \quad |T| > t_{28, 0,0025} \rightarrow \text{reject } H_0$$

\hookrightarrow Since $T = 1,989812 < t_{n_x+n_y-2, 0,0025} = t_{28, 0,0025} = 2,048$, then we don't reject H_0 .

\hookrightarrow KI 95% = $\bar{d} \pm t_{28, 0,0025} \cdot SE[\bar{D}] = [-0,006; 0,403]$. Since KI 95% covers the H_0 , then we don't reject H_0 . //

```

97 - ````{r}
98 # Spooled Variance for not pairwised D:
99 var_p = ((nx-1)*var_x + (ny-1)*var_y) / (nx + ny - 2)
100
101 # Spooled Standard Deviation for not pairwised D:
102 std_d = sqrt(var_p) * sqrt((1/nx) + (1/ny))
103
104 # Sample Standard Deviation:
105 cat("Sample Spooled Variance for not pairwised D: ", var_p,
106     "\nSample Standard Deviation/Error for not pairwised D: ", std_d)
107 ```

    Sample Spooled Variance for not pairwised D: 0.07476333
    Sample Standard Deviation/Error for not pairwised D: 0.0998421

108 - ````{r}
109 t.test(schizophrenia$Unaffected, schizophrenia$Affected, mu=0, paired=FALSE, var.equal=FALSE)
110 ```

    Welch Two Sample t-test

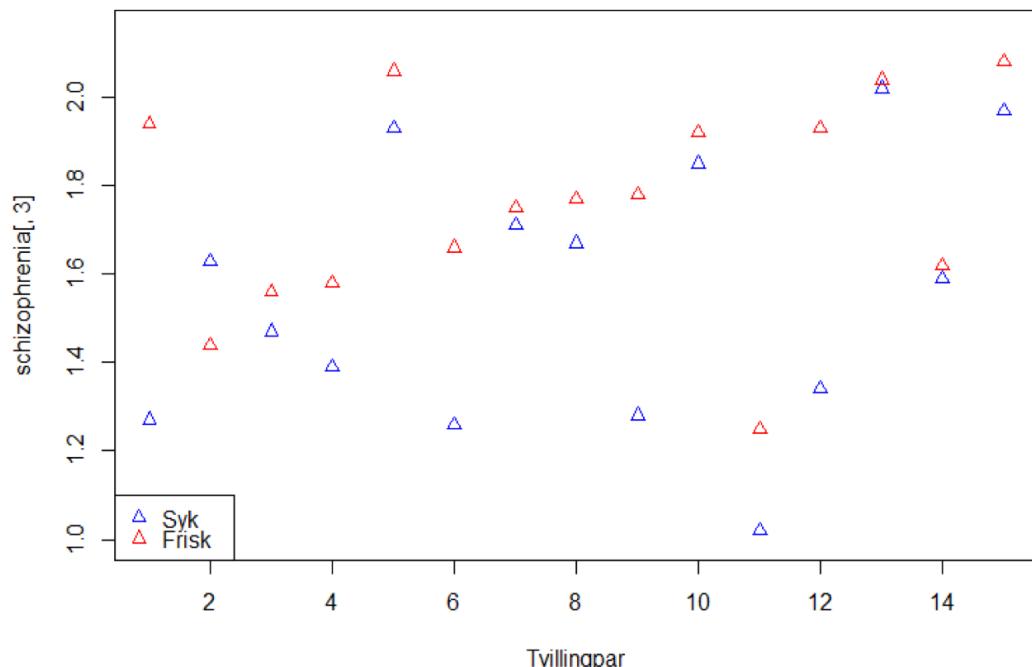
    data: schizophrenia$Unaffected and schizophrenia$Affected
    t = 1.9898, df = 26.775, p-value = 0.05691
    alternative hypothesis: true difference in means is not equal to 0
    95 percent confidence interval:
    -0.006273094 0.403606427
    sample estimates:
    mean of x mean of y
    1.758667 1.560000
  
```

Ekstra:

Som noen kanskje har skjønt så har vi brukt samme data i denne oppgaven både for paret og uparet test, men vi har byttet om på rekkefølgen av observasjonene, både for X og Y. Vi har dermed tapt informasjonen om at én observasjon av X og én av Y tilhører samme tvillingpar og kan tenke oss at de dermed representerer to tilfeldige utvalg fra populasjonen av friske og syke.

I Figur 2 har vi imidlertid tatt inn igjen informasjonen om tvillingpar og plottet verdiene av X (rød) og Y (blå) for hvert tvillingpar. I oppgave L fant vi at vi ikke kunne forkaste nullhypotesen på 5% nivå ved å teste som om det var to uavhengige utvalg, mens vi i oppgaven med parvis test og tvillinginformasjon kunne forkaste og konkludere med at syke hadde mindre forventet volum på hippocampus enn de friske.

Bruk Figur 2 til å forklare hvorfor det var enklere å finne en forskjell mellom syke og friske ved en parvis test enn ved testen som antok to uavhengige utvalg. Hvilken uønsket kilde til variasjon i data har vi blitt «kvitt» ved å bruke parvis test?



Figur 2: Plott av X og Y for hvert tvillingpar nr 1 til 15.

* when we use the pairwise technique we eliminate the great variation that one of the variable has, then that is the reason we didn't reject H_0 . ☒