stat models B EX1

Roi hezkiyahu - 205884018 Dov tuch - 207049719

24 2 2022

Q.1

תחת הנחת הנורמליות מתקיים

$Y_{ij} \sim N(\mu+\alpha_i, \sigma^2)$

ולכן פונקציית הנראות הינה:

$\Pi_i \Pi_j \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y_{ij}-\mu-\alpha_i)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\sum_i \sum_j \frac{-(y_{ij}-\mu-\alpha_i)^2}{2\sigma^2}}$

מיקסום הנראות שקול למיקסום הלוג נראות וכן הלוג נראות הינה:

$ln(\frac{1}{\sqrt{2\pi\sigma^2}}) + \frac{\sum_i \sum_j -(y_{ij}-\mu-\alpha_i)^2}{2\sigma^2}$

כעת נשים לב שבעת גזירה לפי אלפא ומיו נקבל כי המחובר השמאלי יעלם ובעת השוואה לאפס יעלם המכנה לכן מיקסום של הלוג נראות לפי פרמטרים אלו שקול למיקסום של:

$\sum_i \sum_j -(y_{ij}-\mu-\alpha_i)^2$

וכן ביטוי זה זהה למינוס הביטוי בכיתה כלומר האומדים הנ"ל הינם אומדי נראות מריבית שכן האומדים בכיתה הביאו את הפונקציה בכיתה למינימום

Q.2

נחשב תחילה תוחלת של SSA

$E(SSA) = E[\sum_{i=1}^I \sum_{j=1}^{n_i} (\bar{Y_{i.}} - \bar{Y})^2] = E\sum_{i=1}^I \sum_{j=1}^{n_i} [(\mu+\alpha_i + \frac{1}{n_i}\sum_j^{n_i} \varepsilon_{ij} - \mu - \sum_i^I \frac{n_i\alpha_i}{N} + \sum_{i=1}^I \sum_{j=1}^{n_i} \varepsilon_{ij})^2] =$
$\sum_{i=1}^I \sum_{j=1}^{n_i} E[(\alpha_i + \frac{1}{n_i}\sum_j^{n_i} \varepsilon_{ij} - \sum_i^I \frac{n_i\alpha_i}{N} + \sum_{i=1}^I \sum_{j=1}^{n_i} \varepsilon_{ij})^2] = \sum_{i=1}^I \sum_{j=1}^{n_i} E[(\alpha_i + \bar{\varepsilon}_{i.} - \sum_i^I \frac{n_i\alpha_i}{N} + \bar{\varepsilon})^2] =$
$\sum_{i=1}^I \sum_{j=1}^{n_i} E[(\alpha_i - \sum_i^I \frac{n_i\alpha_i}{N}) + ((\bar{\varepsilon}_{i.} - \bar{\varepsilon}))^2] =$
$\sum_{i=1}^I \sum_{j=1}^{n_i} E[(\alpha_i - \sum_i^I \frac{n_i\alpha_i}{N})^2 + (\bar{\varepsilon}_{i.} - \bar{\varepsilon})^2 + 2(\alpha_i - \sum_i^I \frac{n_i\alpha_i}{N})(\bar{\varepsilon}_{i.} - \bar{\varepsilon})] := \sum_{i=1}^I \sum_{j=1}^{n_i} [E(A^2) + E(B^2) + E(AB)]$

נפרק כל אחד מהמחוברים

קבוע ולכן A:

$E(AB) = AE(B) = 0$

השוויון האחרון נובע מכך שתוחלת האפסילונים היא 0

$\sum_{i=1}^I \sum_{j=1}^{n_i} E(B^2) = V(B) + E(B)^2 = \sum_{i=1}^I \sum_{j=1}^{n_i} V(B) = \sum_{i=1}^I \sum_{j=1}^{n_i} [\frac{1}{n_i^2} V(\sum_{j=1}^{n_i} \varepsilon_{ij}) - \frac{1}{N^2} V(\sum_{i=1}^I \sum_{j=1}^{n_i} \varepsilon_{ij})]$
$= \sum_{i=1}^I \sum_{j=1}^{n_i} [\frac{\sigma^2}{n_i} - \frac{\sigma^2}{N}] = I\sigma^2 - \sigma^2 = (I-1)\sigma^2$

$\sum_{i=1}^I \sum_{j=1}^{n_i} E(A^2) = \sum_{i=1}^I \sum_{j=1}^{n_i} E(A)^2 = \sum_{i=1}^I \sum_{j=1}^{n_i} (\alpha_i - \sum_i^I \frac{n_i\alpha_i}{N})^2 = \sum_{i=1}^I \sum_{j=1}^{n_i} (\alpha_i^2) = \sum_{i=1}^I n_i\alpha_i^2$

השוויון השלישי נובע מהאילוץ

סה"כ קיבלנו כי:

$E[MSA] = E[\frac{SSA}{I-1}] = \sigma^2 + \frac{\sum_{i=1}^I n_i\alpha_i^2}{I-1}$

Q.3

```
library(tidyverse)
library(Stat2Data)
library(glue)
library(lawstat)
data(Ricci)
Ricci_W <- Ricci%>%
  filter(Race == "W")%>%
  select(Position,Written,Race)
W_C <- select(filter(Ricci_W,Position == "Captain"),Written)
W_L <- select(filter(Ricci_W,Position != "Captain"),Written)
```

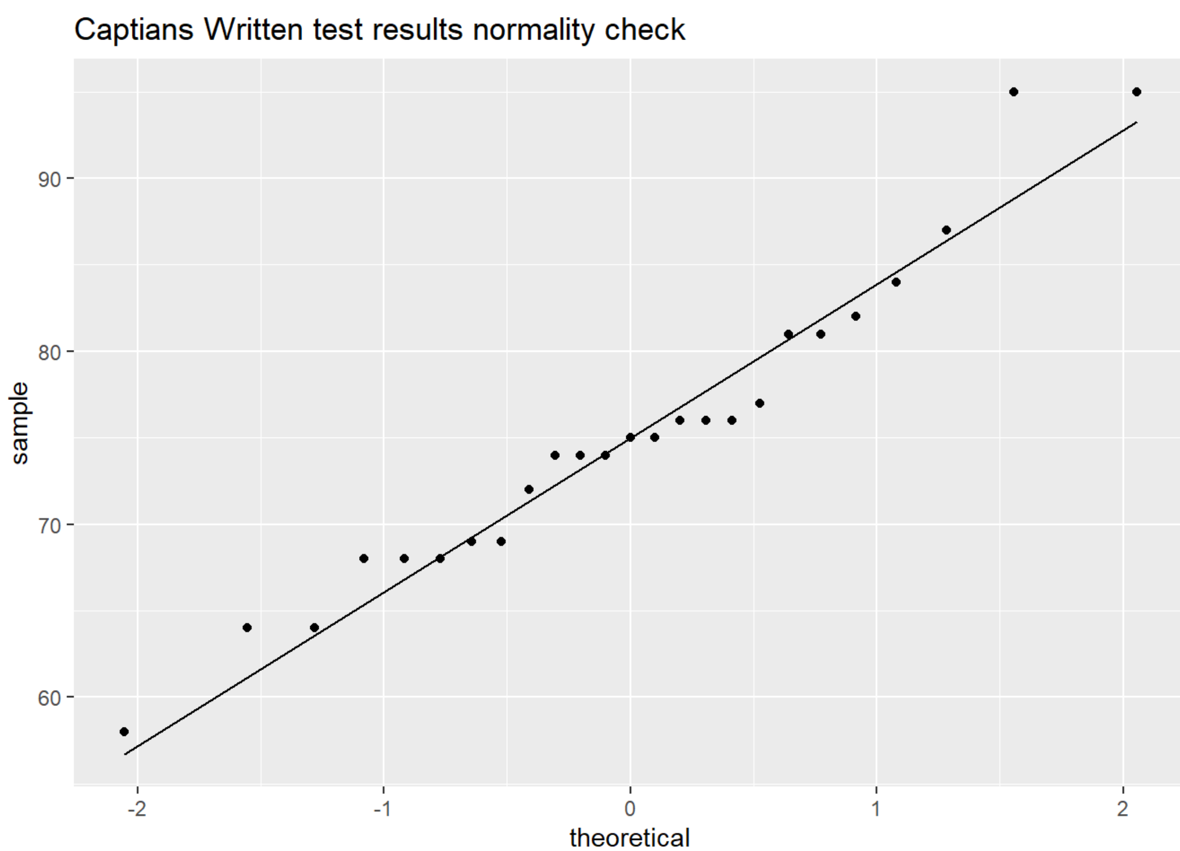
א

ההנחות הינן:נורמליות,שוויון שונויות,אי תלות בין התצפיות

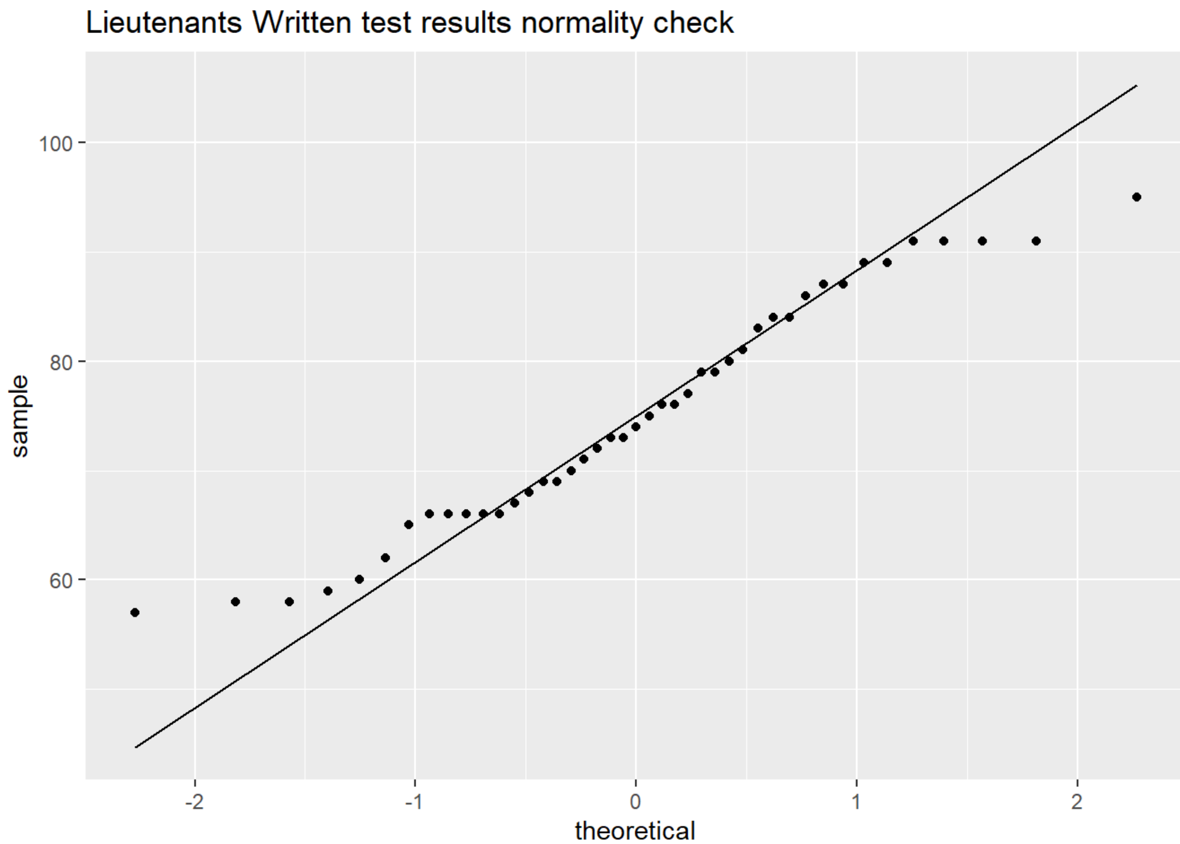
נבדוק את הנחות הנורמליות ושוויון שונויות

```
## normality test via QQplot
```

```
ggplot(W_C,aes(sample = Written))+
  stat_qq()+
  stat_qq_line()+
  ggtitle("Captians Written test results normality check")
```




```
ggplot(W_L,aes(sample = Written))+
  stat_qq()+
  stat_qq_line()+
  ggtitle("Lieutenants Written test results normality check")
```




```
f_test <- var.test(Written~Position,data = Ricci_W)
```

```
## both qqlpots suggets noramlity assupmtion is ok
```

```
## equal variance test via F-test
```

```
## F test pvalue is 0.334786258197595 therefore assumption for equal variance is ok
```

ב

```
t_test <- t.test(W_C,W_L,var.equal = TRUE)
t_test_pv <- t_test$p.value
aov_pos <- aov(Written~Position,data = Ricci_W)
pv_aov <- summary(aov_pos)[[1]][["Pr(>F)"]][1]
```

```
## t test pvalue is 0.919985536261316
```

```
## anova pvalue is 0.919985536261319
```

```
## [1] "pvalue for both test > 0.05 therefore we will not reject the null and conclude that there is no diff
rence between groups"
```

ג

```
glue("we can see that the pvalues are the same")
```

```
## we can see that the pvalues are the same
```

```
t_statistic <- t_test$statistic
F_statistic <- summary(aov_pos)[[1]][["F value"]][1]
glue("we can also see that t_statistic^2:{round(t_statistic^2,4)} is the same as F_statistic:{round(F_statis
tic,4)}")
```

```
## we can also see that t_statistic^2:0.0102 is the same as F_statistic:0.0102
```

מבחן אנובה עבור 2 קבוצות שקול למבחן טי עבור 2 קבוצות וזאת מכיוון ובמקרה זה יש שקילות בין ערך טי לערך אף

ראינו שקילות זאת במודלים א/תיאוריה סטטיסטית

Q.4

א

```
Ricci_L <- Ricci%>%
  filter(Position == "Lieutenant")%>%
  select(Combine,Race)
levene_test <-levene.test(Ricci_L$Combine,Ricci_L$Race,location = "mean")
glue("pvalue for levene test is {round(levene_test$p.value,4)} therefore we will not reject the null and con
clude equal varicane")
```

```
## pvalue for levene test is 0.1795 therefore we will not reject the null and conclude equal varicane
```

ב

```
my_lev <- function(vals,groups){
  df = tibble(values = vals,groups_names = groups)
z_df <- df %>%
  group_by(groups_names)%>%
  mutate(mean_val = mean(values))%>%
  mutate(z = abs(values-mean_val))%>%
  select(groups_names,z)
  aov_res <- aov(z~groups_names,data = z_df)
  sum_aov <- summary(aov_res)
  return(list("Test Statistic"=sum_aov[[1]][["F value"]][1],"p-value" = sum_aov[[1]][["Pr(>F)"]][1]))
}
my_lev_res <- my_lev(Ricci_L$Combine,Ricci_L$Race)

glue("pvalue for my levene test is {round(my_lev_res$`p-value`,4)} therefore we will not reject the null and
conclude equal varicane")
```

```
## pvalue for my levene test is 0.1795 therefore we will not reject the null and conclude equal varicane
```