HW6

2023-09-30

42*7.5. Refer to **Patient satisfaction Problem 6.15.**

7.5a Obtain the **analysis of variance table** that decomposes the regression sum of squares into **extra sums of squares** associated with **X2**; with **X1**, given X2; and with **X3**, given X2 and X1.

Note: For 7.5 (a) ANOVA table, follow the example Tables 7.3 & 7.4 in the textbook (pp.261-262), but pay attention that in Problem 7.5, X2 is entered first instead of X1.

	SS	MS	DF	
Regression (R)	**9120.5**	**3040.167**	**3**	
(X2)	4860.3	4860.3	1	n=46
(X1|X2)	3896	3896	1	p=4
(X3|X2,X1)	364.2	364.2	1	
Error	**4248.8**	**101.1619**	**42**	
Total	**13369.3**	**297.0956**	**45**	

```r
# -----------------------------------------------------------------------
# 7.5 Fit the 2nd model X2, X1, X3.
m2 <- lm(Satisfaction ~ Illness + Age + Anxiety, data = Satisfaction)
summary(m2)
anova(m2)

n <- 46
p <- 4
SSR <- 4860.3 + 3896.0 + 364.2
SSE <- 4248.8
SSTO <- SSR + SSE
MSE <- 101.2
MSR <- SSR / (p-1)
MSTO <- SSTO / (n-1)
SSR;SSE;SSTO
MSR;MSE;MSTO
```

```
> anova(m2)
Analysis of Variance Table

Response: Satisfaction
          Df Sum Sq Mean Sq F value    Pr(>F)    
Illness    1 4860.3  4860.3 48.0439 1.822e-08 ***
Age        1 3896.0  3896.0 38.5126 2.008e-07 ***
Anxiety    1  364.2   364.2  3.5997   0.06468 .  
Residuals 42 4248.8   101.2                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> SSR;SSE;SSTO
[1] 9120.5
[1] 4248.8
[1] 13369.3
> MSR;MSE;MSTO
[1] 3040.167
[1] 101.2
[1] 297.0956
```

7.5b. Test **whether X3 can be dropped from the regression model given that X1, and X2 are retained.** Use the F* test statistic and level of significance 0.025. State the alternatives, decision rule, and conclusion. What is the P -yalue of the test?

Note: For 7.5 (b), an example is given in the textbook Section 7.2 (the first example), which shows you can use quantities available in the ANOVA table you created in part (a) to compute the test statistic. Alternatively, you can also use the "general linear test approach".

```
################ The F-test of extra sum of squares ################
# b. whether X3 can be dropped from the reg. model given X1 & X2 being retained.
# H0: β3=0, X3 can be dropped.
# H1: β1≠0, X3 should be included in the model.
MSR.X3 <- 364.2
MSE <- 101.2
F.obs.X3 <- MSR.X3/MSE
F.obs.X3
1-pf(q=F.obs.X3, df1=1, df2=42)
pf(q = F.obs.X3, df1 = 1, df2 = 42, lower.tail=F)
anova(m2)
# Conclusion: With p=0.06471098 >α, we failed to reject H0: β3=0.
# There is not enough evidence suggesting that X3 should be included.
# Thus, X3 can be dropped from the model.
```

```
> anova(m2)
Analysis of Variance Table

Response: Satisfaction
          Df Sum Sq Mean Sq F value    Pr(>F)    
Illness    1 4860.3  4860.3 48.0439 1.822e-08 ***
Age        1 3896.0  3896.0 38.5126 2.008e-07 ***
Anxiety    1  364.2   364.2  3.5997   0.06468 .  
Residuals 42 4248.8   101.2
```

```
> MSR.X3 <- 364.2
> MSE <- 101.2
> F.obs.X3 <- MSR.X3/MSE
> F.obs.X3
[1] 3.598814
> 1-pf(q=F.obs.X3, df1=1, df2=42)
[1] 0.06471098
```

```
# Effect size by X3, partial R-squared
# R2.Y3_12 <- SSR(X3|X1,2)/SSE(X1,2)
# R2.Y3_12 <- SSR3_12/SSE.m3
m3 <- lm(Satisfaction ~ Illness + Age, data = Satisfaction )
anova(m3)
SSE.m3 <- 4613.0
SSR3_12 <- 364.2
R2.Y3_12 <- SSR3_12/SSE.m3

# R2.Y3_12_1 <- (SSE.m3-SSE.m2)/SSE.m3
SSE.m2 <- 4248.8
R2.Y3_12_1 <- (SSE.m3-SSE.m2)/SSE.m3
R2.Y3_12; R2.Y3_12_1

# Verify with what an add-on function gives:
install.packages("rsq")
library(rsq)
rsq::rsq.partial(objF = m2, objR = m3)

# The effect size by addubg X3 is 7.89%.
# By addubg X3 to the model where X1 and X2 already exist, the variation in Y
# unexplained by the model reduces by R2y_X3|X1,2 is 7.89%.
```

```
> R2.Y3_12; R2.Y3_12_1
[1] 0.07895079
[1] 0.07895079
> rsq::rsq.partial(objF = m2, objR = m3)
$adjustment
[1] FALSE

$variables.full
[1] "Illness" "Age"     "Anxiety"

$variables.reduced
[1] "Illness" "Age"

$partial.rsq
[1] 0.07894201
```

```
# general linear test approach
m.F <- lm(Satisfaction ~ Illness + Age + Anxiety, data = Satisfaction)
m.R <- lm(Satisfaction ~ Illness + Age, data = Satisfaction )
# F-test comparing the two models
anova(m.R, m.F)
# Effect size: partial R-squared
rsq::rsq.partial(objF = m.F, objR = m.R)
# Impression: The anova comparison shows that the difference between two models
# are not significant with the p=0.0648.The effect size by the X3 is 7.89%.
```

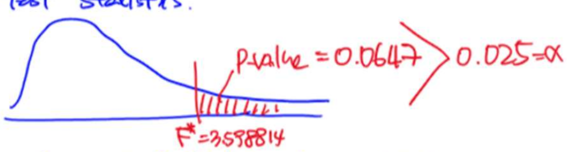
```
> anova(m.R, m.F)
Analysis of Variance Table

Model 1: Satisfaction ~ Illness + Age
Model 2: Satisfaction ~ Illness + Age + Anxiety
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1     43 4613.0
2     42 4248.8  1    364.16 3.5997 0.06468 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> # Effect size: partial R-squared
> rsq::rsq.partial(objF = m.F, objR = m.R)
$adjustment
[1] FALSE

$variables.full
[1] "Illness" "Age"     "Anxiety"

$variables.reduced
[1] "Illness" "Age"

$partial.rsq
[1] 0.07894201
```

7.5b. Test whether X_3 can be dropped given X_1, X_2 are retained.

Step 1. Assumption: $\varepsilon \overset{iid}{\approx} N(0, \sigma^2)$.
Diagnostics are already completed

Step 2. Hypothesis
$H_0: \beta_3 = 0$, X_3 can be dropped
$H_1: \beta_3 \neq 0$, X_3 should be included in the model.

Step 3. Test Statistics.



$$F^*_{obs} = \frac{MSR(X_3|X_1,X_2)}{MSE} = \frac{364.2}{101.2} \approx 3.598814 \simeq 3.599$$

Step 4. p-value.
$1 - pf(q = 3.598814,\ df1 = 1,\ df2 = 46 - 4 = 42) = 0.06471098$

Step 5. With $p = 0.0647$, greater than pre-determined significance level (0.025),
we fail to reject Ho.
Thus, we conclude that X_3(Anxiety) can be dropped from the regression model.
the effect size by adding X_3, where X_1 & X_2 are already exist, is 7.89%.
$(R^2_{Y3|1,2})$

***7.6**. Refer to **Patient satisfaction Problem 6.15**. Test whether both X2 and X3 can be dropped from the regression model given that X1 is retained. Use **α = 0.025**. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

Note: Solve 7.6 **using the "general linear test approach".**

For work done by applying the "general linear test approach", **clearly state the full model and the reduced model** in Step 2 (hypotheses). Then **give H0 where you specify which parameters are constrained at what value,** and then clearly state H1 as well.

```
# whether both X2 and X3 can be dropped from the regression model given that
# X1 is retained.
# H0: β2 = β3 = 0. Both X2 & X3 can be dropped. Reduced model is better.
# H1: β2≠0 or β3≠0. X2, X3 or both should be included. Full model is better.

m.F <- lm(Satisfaction ~ Age + Illness + Anxiety, data=Satisfaction)
m.R <- lm(Satisfaction ~ Age, data=Satisfaction)
anova(m.R, m.F)
F.obs <- 4.1768 # From anova result
SSR12_1 <- 845.07 # From anova result
SSE.m.R <- 5093.9 # From anova result
SSE.m.F <- 4248.8 # From anova result
MSE.m.F <- SSE.m.F /42 # From anova result
F_obs <- ((SSE.m.R-SSE.m.F)/2) / MSE.m.F # From anova result
F..obs <- ((SSR12_1)/2) / MSE.m.F # From anova result
F.obs; F_obs; F..obs
1-pf(q=F.obs, df1=2, df2=42) # From anova result
# Impression: With p=0.02216 < α=0.025, we have enough evidence to reject H0.
# We conclude that the full model fits the data significantly better than
# the reduced model. At least one or more of X2 and X3 should be included in the model.

# Effect size
rsq::rsq.partial(objF = m.F, objR = m.R)
```

```
> anova(m.R, m.F)
Analysis of Variance Table

Model 1: Satisfaction ~ Age
Model 2: Satisfaction ~ Age + Illness + Anxiety
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1     44 5093.9
2     42 4248.8  2    845.07 4.1768 0.02216 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> F.obs <- 4.1768 # From anova result
> SSR12_1 <- 845.07 # From anova result
> SSE.m.R <- 5093.9 # From anova result
> SSE.m.F <- 4248.8 # From anova result
> MSE.m.F <- SSE.m.F /42 # From anova result
> F_obs <- ((SSE.m.R-SSE.m.F)/2) / MSE.m.F # From anova result
> F..obs <- ((SSR12_1)/2) / MSE.m.F # From anova result
> F.obs; F_obs; F..obs
[1] 4.1768
[1] 4.176968
[1] 4.176819
> 1-pf(q=F.obs, df1=2, df2=42) # From anova result
[1] 0.02216124
> rsq::rsq.partial(objF = m.F, objR = m.R)
$adjustment
[1] FALSE

$variables.full
[1] "Age"     "Illness" "Anxiety"

$variables.reduced
[1] "Age"

$partial.rsq
[1] 0.1658989
```

Step1. Assumption : $\varepsilon \overset{iid}{\sim} N(0, \sigma^2)$

(1) Linear relationship between each X and Y.
(2) No outlier (stud. resid $\sim \hat{Y}$)
(3) Not heteroskedastic. (rstudent $\sim$ predict)
(4) Normal Error distribution $N(0, \sigma^2)$, by QQ plot
(5) Independent observation : randomly selected patients.

Step2. Hypothesis

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$

$H_0 : \beta_2 = \beta_3 = 0$ → m.R : Both X_2, X_3 can be dropped.
$H_1 : \beta_2 \neq 0$ or $\beta_3 \neq 0$ → m.F : Atleast one of X_2 or X_3 should be included in the model.

Full model (m.F) ; Satisfaction = $\beta_0 + \beta_1$ Age + β_2 illness + β_3 Anxiety.
Reduced model (m.R); Satisfaction = $\beta_0 + \beta_1$ Age

Step3. Test statistic

$$F^* = \frac{\frac{SSR(F) - SSR(R)}{df_R(F) - df_R(R)}}{MSE.F} = \frac{\frac{SSE(R) - SSE(F)}{df.e(R) - df.e(F)}}{MSE.F}$$

$$= \frac{\frac{842.07}{2}}{\frac{4248.8}{42}} = \frac{\frac{5093.9 - 4248.8}{44 - 42}}{\frac{4248.8}{48}}$$

$\simeq 4.1768$

Step4. P-Value.

1-pf (q = 4.1768, df1 = 44-42 , df2 = 42)

df.sseR - df.sse.F   df.sse(F)

= 0.0221612Y

Step5. Conclusion.

With p = 0.02216 , less than α = 0.025 , we reject H_0.
thus, we conclude that full model fits the data significantly better, indicating at least one or more of X_2 and X_3 should be included.

*7.9. Refer to **Patient satisfaction Problem 6.15**. Test whether β1 = -1.0 and, β2 = 0; use α = 0.025. State the alternatives, full and reduced models, decision rule, and conclusion.

Step 1. Assumption : $\varepsilon \overset{iid}{\sim} N(0,\sigma^2)$

(1) linear relationship
(2) outlier (no)
(3) Not heteroskedastic
(4) Normal error distribution (Q-Q plot)
(5) Independent observation (random sampling)

Step 2. Hypothesis

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$
Test $\beta_1 = -1,\ \beta_2 = 0$

Full model M.F : $Y = \beta_0 + \overset{-1}{\beta_1} X_1 + \overset{0}{\beta_2} X_2 + \beta_3 X_3$
Reduced model M.R : $Y + X_1 = \beta_3 X_3$

H_0 : $\beta_1 = -1$ and $\beta_2 = 0$
H_1 : $\beta_1 \neq -1$ or $\beta_2 \neq 0$, The full model fits data better than the restricted model. $(\alpha = 0.025)$

Step 3. Test Statistics.

$$F^*_{obs} = \frac{\dfrac{SSE(R) - SSE(F)}{Df_{sse}(R) - Df_{sse}(F)}}{\dfrac{SSE(F)}{Df_{sse}(F)} = MSE.F\ (=101.2)} = \frac{\dfrac{4427.7 - 4248.8}{44-42}}{\dfrac{4248.8}{42}} = 0.8842$$

$$= \frac{\dfrac{SSR(F) - SSR(R) = \Delta SSR}{Df_{ssr}(F) - Df_{ssr}(R) = 2}}{MSE.F} = \frac{\dfrac{178.81}{2}}{101.2} = 0.8837813$$

≈ 0.884

Step 4. P-value

$1 - pf(q = 0.884,\ df_1 = 2,\ df_2 = 42) = 0.4207535$

Step 5. Conclusion.

With $P \approx 0.4208 > \alpha = 0.025$, we fail to reject H_0.

There is Not enough evidence supporting the restricted model is significantly different to the full model.
The constrained model with $\beta_1 = -1$ and $\beta_2 = 0$ fits the data about as well as the full model.
thus, we may choose the smaller model (restricted model).

```
# ------------------------------------------------------------------------
# Test whether β1 = -1.0 and, β2 = 0; use α = 0.025.
# Y = β0 -1*X1 + 0*X2 + β3*X3
# Y + X1 = β0 + β3X3
# Full model
data.full <- Satisfaction
data.full$Y <- data.full$Satisfaction
m.F <- lm(Y ~ Age + Illness + Anxiety, data=data.full)
# Reduced model
data.reduced <- Satisfaction
data.reduced$Y <- with(Satisfaction, Satisfaction + Age) # Y + X1
m.R <- lm(Y ~ Anxiety, data=data.reduced) # β0 + β3X3
# F-test comparing the two models
anova(m.R, m.F)
F_obs <- 0.8838
# F = {SSE(R-F)/DFe(R-F)}/ (SSE.F/DFe.F)
F.obs <- ((4427.7 - 4248.8)/(44-42))/(4248.8/42)
# F = {SSR.delta/Df}/ MSE
SSR.delta <- 178.81
F..obs <- (SSR.delta/2)/(4248.8/42)
F_obs; F.obs; F..obs

1-pf(q=0.8838, df1=44-42, df2=42)
# With p=0.4207535 > α= 0.025, we fail to reject H0.
# There isn't enough evidence suggesting that the constraints β1=-1 and β2=0
# are significantly different to the full model.
# The constrained model fits the data about as well as the full model.
```

```
> anova(m.R, m.F)
Analysis of Variance Table

Model 1: Y ~ Anxiety
Model 2: Y ~ Age + Illness + Anxiety
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     44 4427.7
2     42 4248.8  2    178.81 0.8838 0.4208
> F_obs <- 0.8838
> # F = {SSE(R-F)/DFe(R-F)}/ (SSE.F/DFe.F)
> F.obs <- ((4427.7 - 4248.8)/(44-42))/(4248.8/42)
> # F = {SSR.delta/Df}/ MSE
> SSR.delta <- 178.81
> F..obs <- (SSR.delta/2)/(4248.8/42)
> F_obs; F.obs; F..obs
[1] 0.8838
[1] 0.8842261
[1] 0.8837813
> 1-pf(q=0.8838, df1=44-42, df2=42)
[1] 0.4207535
```

R code and the results

Import data

```
rm(list=ls()) # Clean up the workspace for the new analysis
# Load packages: Every time opining a new R session, need to load packages.

library(Hmisc) # describe functions of Hmisc and Psych are same, thus rename it.

Hmisc.describe <- describe
library(psych) # describe functions of Hmisc and Psych are same, thus rename it.

Psych.describe <- describe
library(rgl)
```

```
library(rsq)

# Set the following to your own folder
setwd("C:/Users/jyang/OneDrive - Arizona State University/10
Classes_OneDrive/2023_STP530_Regression/HW6")

############### Read data, check data ###############
Satisfaction <- read.table("CH06PR15.txt")
head(Satisfaction)

##   V1 V2 V3  V4
## 1 48 50 51 2.3
## 2 57 36 46 2.3
## 3 66 40 48 2.2
## 4 70 41 44 1.8
## 5 89 28 43 1.8
## 6 36 49 54 2.9

colnames(Satisfaction) <- c("Satisfaction","Age","Illness", "Anxiety")
str(Satisfaction) #Displays a concise structure of an R object

## 'data.frame':    46 obs. of  4 variables:
##  $ Satisfaction: int  48 57 66 70 89 36 46 54 26 77 ...
##  $ Age         : int  50 36 40 41 28 49 42 45 52 29 ...
##  $ Illness     : int  51 46 48 44 43 54 50 48 62 50 ...
##  $ Anxiety     : num  2.3 2.3 2.2 1.8 1.8 2.9 2.2 2.4 2.9 2.1 ...

attach(Satisfaction) #Adds a database to R's search path,

#allowing direct variable referencing. Use with caution.
```

7.5a.Obtain the analysis of variance table that decomposes the regression sum of squares into extra sums of squares associated with X2; with X1, given X2; and with X3, given X2 and X1.

```
################ Fit the 2nd model ###############
# Residual diagnostics are done in the previous HW5

# 7.5 Fit the 2nd model X2, X1, X3.
m1 <- lm(Satisfaction ~ Age + Illness + Anxiety, data = Satisfaction)
m2 <- lm(Satisfaction ~ Illness + Age + Anxiety, data = Satisfaction)
summary(m2)

##
## Call:
## lm(formula = Satisfaction ~ Illness + Age + Anxiety, data = Satisfaction)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.3524  -6.4230   0.5196   8.3715  17.1601
##
```

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 158.4913    18.1259   8.744 5.26e-11 ***
## Illness      -0.4420     0.4920  -0.898   0.3741
## Age          -1.1416     0.2148  -5.315 3.81e-06 ***
## Anxiety     -13.4702     7.0997  -1.897   0.0647 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.06 on 42 degrees of freedom
## Multiple R-squared:  0.6822, Adjusted R-squared:  0.6595
## F-statistic: 30.05 on 3 and 42 DF,  p-value: 1.542e-10
```

```
anova(m2)
```

```
## Analysis of Variance Table
##
## Response: Satisfaction
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Illness    1 4860.3  4860.3 48.0439 1.822e-08 ***
## Age        1 3896.0  3896.0 38.5126 2.008e-07 ***
## Anxiety    1  364.2   364.2  3.5997   0.06468 .
## Residuals 42 4248.8   101.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
n <- 46
p <- 4
SSR <- 4860.3 + 3896.0 + 364.2
SSE <- 4248.8
SSTO <- SSR + SSE
MSE <- 101.2
MSR <- SSR / (p-1)
MSTO <- SSTO / (n-1)
SSR;SSE;SSTO
```

```
## [1] 9120.5
```

```
## [1] 4248.8
```

```
## [1] 13369.3
```

```
MSR;MSE;MSTO
```

```
## [1] 3040.167
```

```
## [1] 101.2
```

```
## [1] 297.0956
```

7.5b. whether X3 can be dropped from the reg. model given X1 & X2 being retained.

```
################ The F-test of extra sum of squares ################
# H0: β3=0, X3 can be dropped.
# H1: β1≠0, X3 should be included in the model.
MSR.X3 <- 364.2
MSE <- 101.2
F.obs.X3 <- MSR.X3/MSE
F.obs.X3

## [1] 3.598814

1-pf(q=F.obs.X3, df1=1, df2=42)

## [1] 0.06471098

pf(q = F.obs.X3, df1 = 1, df2 = 42, lower.tail=F)

## [1] 0.06471098

anova(m2)

## Analysis of Variance Table
##
## Response: Satisfaction
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Illness    1 4860.3  4860.3 48.0439 1.822e-08 ***
## Age        1 3896.0  3896.0 38.5126 2.008e-07 ***
## Anxiety    1  364.2   364.2  3.5997   0.06468 .
## Residuals 42 4248.8   101.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Conclusion: With p=0.06471098 >α, we failed to reject H0: β3=0.
# There is not enough evidence suggesting that X3 should be included.
# Thus, X3 can be dropped from the model.

# ---------------------------------------------------------------------------
# Effect size by X3, partial R-squared
# R2.Y3_12 <- SSR(X3|X1,2)/SSE(X1,2)
# R2.Y3_12 <- SSR3_12/SSE.m3
m3 <- lm(Satisfaction ~ Illness + Age, data = Satisfaction )
anova(m3)

## Analysis of Variance Table
##
## Response: Satisfaction
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Illness    1 4860.3  4860.3  45.305 3.161e-08 ***
## Age        1 3896.0  3896.0  36.317 3.348e-07 ***
## Residuals 43 4613.0   107.3
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
SSE.m3 <- 4613.0
SSR3_12 <- 364.2
R2.Y3_12 <- SSR3_12/SSE.m3

# R2.Y3_12_1 <- (SSE.m3-SSE.m2)/SSE.m3
SSE.m2 <- 4248.8
R2.Y3_12_1 <- (SSE.m3-SSE.m2)/SSE.m3
R2.Y3_12; R2.Y3_12_1
```

```
## [1] 0.07895079
```

```
## [1] 0.07895079
```

```
# Verify with what an add-on function gives:
rsq::rsq.partial(objF = m2, objR = m3)
```

```
## $adjustment
## [1] FALSE
##
## $variables.full
## [1] "Illness" "Age"     "Anxiety"
##
## $variables.reduced
## [1] "Illness" "Age"
##
## $partial.rsq
## [1] 0.07894201
```

```
# The effect size by addubg X3 is 7.89%.
# By addubg X3 to the model where X1 and X2 already exist, the variation in Y
# unexplained by the model reduces by R2y_X3|X1,2 is 7.89%.

# ---------------------------------------------------------------------------
# general linear test approach
m.F <- lm(Satisfaction ~ Illness + Age + Anxiety, data = Satisfaction)
m.R <- lm(Satisfaction ~ Illness + Age, data = Satisfaction )
# F-test comparing the two models
anova(m.R, m.F)
```

```
## Analysis of Variance Table
##
## Model 1: Satisfaction ~ Illness + Age
## Model 2: Satisfaction ~ Illness + Age + Anxiety
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     43 4613.0
## 2     42 4248.8  1    364.16 3.5997 0.06468 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Effect size: partial R-squared
rsq::rsq.partial(objF = m.F, objR = m.R)

## $adjustment
## [1] FALSE
##
## $variables.full
## [1] "Illness" "Age"     "Anxiety"
##
## $variables.reduced
## [1] "Illness" "Age"
##
## $partial.rsq
## [1] 0.07894201

# Impression: The anova comparison shows that the difference between two
models
# are not significant with the p=0.0648.The effect size by the X3 is 7.89%.
```

7.6. whether both X2 and X3 can be dropped from the regression model given that X1 is retained.

```
# Use "general linear test approach"
# H0: β2 = β3 = 0. Both X2 & X3 can be dropped. Reduced model is better.
# H1: β2≠0 or β3≠0. X2, X3 or both should be included. Full model is better.

m.F <- lm(Satisfaction ~ Age + Illness + Anxiety, data=Satisfaction)
m.R <- lm(Satisfaction ~ Age, data=Satisfaction)
anova(m.R, m.F)

## Analysis of Variance Table
##
## Model 1: Satisfaction ~ Age
## Model 2: Satisfaction ~ Age + Illness + Anxiety
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     44 5093.9
## 2     42 4248.8  2    845.07 4.1768 0.02216 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

F.obs <- 4.1768 # From anova result
SSR12_1 <- 845.07 # From anova result
SSE.m.R <- 5093.9 # From anova result
SSE.m.F <- 4248.8 # From anova result
MSE.m.F <- SSE.m.F /42 # From anova result
F_obs <- ((SSE.m.R-SSE.m.F)/2) / MSE.m.F # From anova result
F..obs <- ((SSR12_1)/2) / MSE.m.F # From anova result
F.obs; F_obs; F..obs

## [1] 4.1768

## [1] 4.176968
```

```
## [1] 4.176819

1-pf(q=F.obs, df1=2, df2=42) # From anova result

## [1] 0.02216124

# Impression: With p=0.02216 < α=0.025, we have enough evidence to reject H0.
# We conclude that the full model fits the data significantly better than
# the reduced model. At least one or more of X2 and X3 should be included in
the model.

# Effect size
rsq::rsq.partial(objF = m.F, objR = m.R)

## $adjustment
## [1] FALSE
##
## $variables.full
## [1] "Age"     "Illness" "Anxiety"
##
## $variables.reduced
## [1] "Age"
##
## $partial.rsq
## [1] 0.1658989

# The effect size by both X2 & X3 is 16.59%. By adding both X2 and X3 to the
# model given X1, the Y unexplained reduces by 16.59%.
```

7.9. Test whether β1 = -1.0 and, β2 = 0; use α = 0.025.

```
# -----------------------------------------------------------------------
# Test whether β1 = -1.0 and, β2 = 0; use α = 0.025.
# Y = β0 -1*X1 + 0*X2 + β3*X3
# Y + X1 = β0 + β3X3
# Full model
data.full <- Satisfaction
data.full$Y <- data.full$Satisfaction
m.F <- lm(Y ~ Age + Illness + Anxiety, data=data.full)
# Reduced model
data.reduced <- Satisfaction
data.reduced$Y <- with(Satisfaction, Satisfaction + Age) # Y + X1
m.R <- lm(Y ~ Anxiety, data=data.reduced) # β0 + β3X3
# F-test comparing the two models
anova(m.R, m.F)

## Analysis of Variance Table
##
## Model 1: Y ~ Anxiety
## Model 2: Y ~ Age + Illness + Anxiety
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
```

```
## 1     44 4427.7
## 2     42 4248.8  2    178.81 0.8838 0.4208

F_obs <- 0.8838
# F = {SSE(R-F)/DFe(R-F)}/ (SSE.F/DFe.F)
F.obs <- ((4427.7 - 4248.8)/(44-42))/(4248.8/42)
# F = {SSR.delta/Df}/ MSE
SSR.delta <- 178.81
F..obs <- (SSR.delta/2)/(4248.8/42)
F_obs; F.obs; F..obs

## [1] 0.8838

## [1] 0.8842261

## [1] 0.8837813

1-pf(q=0.8838, df1=44-42, df2=42)

## [1] 0.4207535

# With p=0.4207535 > α= 0.025, we fail to reject H0.
# There isn't enough evidence suggesting that the constraints β1=-1 and β2=0
# are significantly different to the full model.
# The constrained model fits the data about as well as the full model.

# Effect size: partial R-squared
rsq::rsq.partial(objF = m.F, objR = m.R)

## $adjustment
## [1] FALSE
##
## $variables.full
## [1] "Age"     "Illness" "Anxiety"
##
## $variables.reduced
## [1] "Anxiety"
##
## $partial.rsq
## [1] 0.5647479

# The effect size is 56.47%
```