Regression HW5 Solutions

Jun Lu, Mingfan Meng

2023

Contents

Total: 70 pts

1. Finish Lab 3 (Separate submission.)

2. Read midterm exam instructions.

Grocery retailer data (problem 6.9)

A large, national grocery retailer tracks productivity and costs of its facilities closely. Data below were obtained from a single distribution center for a one-year period. Each data point for each variable represents one week of activity. The variables included are the number of cases shipped (X_1) the indirect costs of the total labor hours as a percentage (X_2), a qualitative predictor called holiday that is coded 1 if the week has a holiday and 0 otherwise (X_3), and the total labor hours (Y).

```
gr <- read.table("./CH06PR09.txt", header = F)
colnames(gr) <- c("Y_labor", "X_case", "X_costs", "X_holiday")
gr.mlr <- lm(Y_labor ~ X_case + X_costs+ X_holiday, data = gr)
gr.mlr
```

```
##
## Call:
## lm(formula = Y_labor ~ X_case + X_costs + X_holiday, data = gr)
##
## Coefficients:
## (Intercept)       X_case      X_costs    X_holiday
##   4.150e+03    7.871e-04   -1.317e+01    6.236e+02
```

3. Problem 6.11. Grocery retailer (p. 250, 15 pts)

Assume that regression model (6.5) for three predictor variables with independent normal error terms is appropriate.

a. Test whether there is a regression relation, using level of significance 0.05. State the alternatives, decision rule, and conclusion. What does your test result imply about β_1, β_2, β_3? What is the p-value of the test?

```
gr.mlr <- lm(Y_labor ~ X_case + X_costs+ X_holiday, data = gr)
summary(gr.mlr)
```

```
##
## Call:
## lm(formula = Y_labor ~ X_case + X_costs + X_holiday, data = gr)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -264.05 -110.73  -22.52   79.29  295.75
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.150e+03  1.956e+02  21.220  < 2e-16 ***
## X_case       7.871e-04  3.646e-04   2.159   0.0359 *
## X_costs     -1.317e+01  2.309e+01  -0.570   0.5712
## X_holiday    6.236e+02  6.264e+01   9.954 2.94e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 143.3 on 48 degrees of freedom
## Multiple R-squared:  0.6883, Adjusted R-squared:  0.6689
## F-statistic: 35.34 on 3 and 48 DF,  p-value: 3.316e-12
```

```
gr_null <- lm(Y_labor ~ 1, data=gr)
anova(gr_null, gr.mlr)
```

```
## Analysis of Variance Table
##
## Model 1: Y_labor ~ 1
## Model 2: Y_labor ~ X_case + X_costs + X_holiday
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1     51 3162136
## 2     48  985530  3   2176606 35.337 3.316e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Estimated regression function:
  - $\hat{Y} = 4150 + 0.000787 * X_1 - 13.166 * X_2 + 623.554 * X_3$
- Hypothesis:
  - $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ (None of predictor variables is associated with response variable)
  - H_a: At least one $\beta_k \neq 0$ (At least one predictor variable is associated with response variable)
- Calculation:
  - $SST = SSR + SSE$, $df_T = df_R + df_E$.
  - $SST = 3162136$
  - $MSR = (SSR)/df_R = (2176606)/3 = 725535.3$. $(df_R = p - 1)$
  - $MSE = SSE/df_E = 985530/48 = 20532$. $(df_E = n - p)$
  - $F_{obs} = \frac{MSR}{MSE} = 35.34$.
  - p-value $= P(F_{(df_R, df_E)} > F_{obs}) = P(F_{(3, 48)} > 35.34) = 3.316e - 12$

    ```
    p <- 1- pf(35.34, 3, 48)
    ```

  - Note that F_{obs} and the p-value are in the `summary()` output and the `anova()` output.

- Decision rules:
  - Reject H_0 if p-value $< \alpha$.
  - Alternatively, one can use F-critical value, and reject H_0 if $F_{obs} > F_{(1-\alpha,\, df_1=df_R,\, df_2=df_E)}$
- Conclusion: According to the p-value=$3.316 \times 10^{-12} < 0.05$, we should reject H_0. We have significant evidence showing that at least one predictor variables is associated with total labor hours. The model is significant overall at 0.05 level.

b. Estimate β_1 and β_3 jointly by the Bonferroni procedure, using a 95 percent family confidence coefficient. Interpret your results.

- There are 2 intervals to be evaluated jointly as a "family". By Bonferroni procedure, compute CI for β_1 and β_3 at 97.5% confidence level. ($g = 2$, family confidence level $= 0.95 = 1 - 0.05$, each interval will use $1 - 0.05/2 = 0.975$ level.)

```
family_alpha <- 0.05  # 0.95 family confidence level
g <- 2                # 2 intervals in the family
p <- 4                # Model has 4 parameters (1 intercept, 3 slopes)
n <- 52   # Sample size is 52 in this data set

B <- qt(1- (family_alpha/g)/2, df=(n-p))
B
```

```
## [1] 2.313899
```

- From the `summary()` output: $\hat{\beta}_1 = 0.0007871$, $se(\hat{\beta}_1) = 0.000365$
- CI for β_1 : $\hat{\beta}_1 \pm B * se(\hat{\beta}_1) = 0.000787 \pm 2.3139(0.000365)$
- From the `summary()` output: $\hat{\beta}_3 = 623.554$, $se(\hat{\beta}_3) = 62.64$
- CI for β_3 : $\hat{\beta}_3 \pm B * se(\hat{\beta}_3) = 623.554 \pm 2.3139(62.6409)$
- At 95% family confidence level, the Bonferroni joint confidence interval for β_1 and β_3 are (-0.000058, 0.00163) and (478.6092, 768.4988), respectively.
- The above formula-based computation can be confirmed in R.

```
confint(gr.mlr, level=0.975) # Each CI use 97.5% level after Bonferroni.
```

```
##                    1.25 %      98.75 %
## (Intercept)  3.697369e+03 4.602406e+03
## X_case      -5.646080e-05 1.630622e-03
## X_costs     -6.659796e+01 4.026592e+01
## X_holiday    4.786096e+02 7.684993e+02
```

- We are 95% confident that: the total labor hours is expected to increase between -0.000058 and 0.00163 hours for each additional unit increase in cases after adjusting for indirect-labor-cost and holiday; and the total labor hours is expected to increase between 478.6092 and 768.4988 hours for holiday weeks, after adjusting for cases-shipped and indirect-labor-cost. This is evaluated using Bonferroni adjustment.

c. Calculate the coefficient of multiple determination R_2. How is this measure interpreted here?

- According to the ANOVA table in part a:
  - $SSR = 2176606$.
  - $SS_{TO} = SSR + SSE = 3162136$.
  - $R^2 = \frac{SSR}{SS_{TO}} = 0.6883$.
- R^2 is the proportion of the variation in response variable that can be determined by the current multiple linear regression model using predictor variables.

- About 69% the total variation in variables the total labor hours can be determined using the current regression model with predictor variables case numbers, cost and holiday.
- Assuming other model assumptions are valid, the model fits the data moderately well.

4. Problem 6.12. Grocery retailer (15 pts)

```
GR_new <- read.table("./CH06PR12.txt", header = F)
colnames(GR_new) <- c("X1", "X2", "X3")
```

a. Management desires simultaneous interval estimates of the total labor hours for the following five typical weekly shipments: Obtain the family of estimates using a 95 percent family confidence coefficient. Employ the Working-Hotelling and the Bonferroni procedure, whichever is more efficient.

- There are 5 intervals to be estiamted jointly as a "family."
- $W = \sqrt{p * F_{(1-\alpha,\, df_1=p,\, df_2=n-p)}} = \sqrt{4 * F_{(0.95,\, df_1=4,\, df_2=48)}} = 3.2033$
- $B = t_{(1-\alpha/(2g), n-p)} = t_{(1-(0.05/5)/2,\, n-p)} = 2.6822$

```
f.alp <- 0.05  # 95% family level
g <- 5    # 5 intervals
p <- 4    # 4 parameters (1 intercept, 3 slopes)
n <- 52   # Sample size is 52 in this data set

# Wordking-Hotelling procedure
W <- sqrt(p * qf((1-f.alp), p, (n-p)))
paste("Working-Hotelling critical value: W =",  W)
```

```
## [1] "Working-Hotelling critical value: W = 3.2032736433087"
```

```
# Bonferroni procedure
B <- qt(1 - (f.alp/g)/2, (n-p))
paste("Bonferrroni Critical value: B =", B)
```

```
## [1] "Bonferrroni Critical value: B = 2.68220402695022"
```

- Because of the $B < W$, we will choose the Bonferroni adjustment because it will produce narrower intervals.
- Bonferroni procedure CI: using software directly. We need to set `level=` according the confidence level that applies to each interval $(1 - \alpha/g)$.

```
f.alp <- 0.05
g <- 5

predict(gr.mlr, newdata=data.frame(X_case= GR_new$X1,
                                   X_costs=GR_new$X2, X_holiday=GR_new$X3),
        se=F, level= (1 - (0.05/5)), interval="confidence")
```

```
##        fit      lwr      upr
## 1 4292.790 4235.507 4350.073
## 2 4245.293 4165.626 4324.960
## 3 4279.424 4213.859 4344.989
## 4 4333.203 4255.609 4410.798
## 5 4917.418 4749.781 5085.055
```

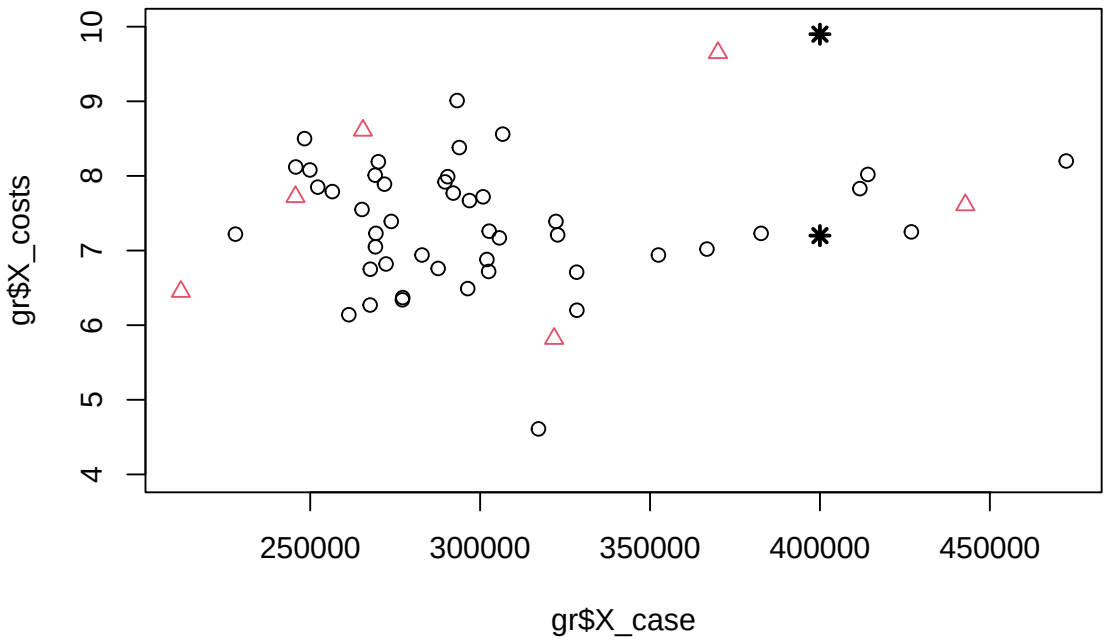
- Conclusion: at a family confidence level of 95%, the joint confidence intervals of the mean labor hours of the 5 weekly shipments are

- $\hat{Y}_1 = 4292.790$, CI is (4235.507, 4350.073)
- $\hat{Y}_2 = 4245.293$, CI is (4165.626, 4324.960)
- $\hat{Y}_3 = 4279.424$, CI is (4213.859, 4344.989)
- $\hat{Y}_4 = 4333.203$, CI is (4255.609, 4410.798)
- $\hat{Y}_5 = 4917.418$, CI is (4749.781, 5085.055)

- *Remark*: Please refer to class examples and previous homework to review how to construct the CI for mean response (E(Y)) by using the formula: $\hat{y} \pm (critical\ value) * se(\hat{y}_{mean})$, where the **critical value** can be t_{crit} (the regular t-critical value), B (t_critical value with Bonferroni adjustment), or W (Working-Hotelling critical value).

b. For the data in Problem 6.9 on which the regression fit is based, would you consider a shipment of 400,000 cases with an indirect percentage of 7.20 on a nonholiday week to be within the scope of the model? What about a shipment of 400,000 cases with an indirect percentage of 9.9 on a nonholiday week? Support your answers by preparing a relevant plot.

```
plot(gr$X_case, gr$X_costs, col=gr$X_holiday + 1, pch=1+gr$X_holiday, ylim=c(4, 10))
points(c(400000, 400000), c(7.2, 9.9), pch=8, lwd=2)
```



- The 2 sets of new X-values are marked as $*$ in the plot. The first shipment (400000, 7.2) is within the scope of the model (within the "cluster" of observed X-values), no extrapolation. The second one (400000, 9.9) is not within the scope of the model (not in the "cluster" of the observed X-values for non-holidays), and has extrapolation problem.

5. Problem 6.13. Grocery retailer (10 pts)

Management desires predictions of the handling times for these shipments so that the actual handling times can be compared with the predicted times to determine whether any are out of line. Develop the needed predictions, using the most efficient approach and a family confidence coefficient of 95 percent.

```
GR_new2 <- read.table("./CH06PR13.txt", header = F)
colnames(GR_new2) <- c("X1", "X2", "X3")
```

- There are 4 intervals to be estiamted jointly as a "family."
- $S = \sqrt{g \cdot F_{(1-\alpha;\, g,\, n-p)}} = \sqrt{4 \cdot F_{(1-\alpha;\, 4,\, 48)}} = 3.2033$
- $B = t_{(1-\alpha/(2g),n-p)} = t_{(1-0.05/4,48)} = 2.5953$

```
f.alp <- 0.05  # 95% family conf. level.
g <- 4  # 4 intervals
p <- 4  # Model has 4 parameters (1 intercept, 3 slopes)
n <- 52   # Sample size is 52 in this data set

# Scheffe's method
S <- sqrt(g*qf(1-f.alp, df1=g, df2=n-p))
paste("Scheffe's critical value S=", S)
```

```
## [1] "Scheffe's critical value S= 3.2032736433087"
```

```
#Bonferroni method
B <- qt(1- (f.alp/g)/2, df=n-p)
paste("Bonferroni critical value B=", B)
```

```
## [1] "Bonferroni critical value B= 2.59532272638378"
```

- Since $S > B$, Bonferroni methods will produce narrower prediction interval in this case.
- Bonferroni method. (set `level=` as the level that applies to each interval.)

```
f.alp <- 0.05  # 95% family conf. level.
g <- 4
predict(gr.mlr, newdata=data.frame(X_case=GR_new2$X1,
                                   X_costs = GR_new2$X2,
                                   X_holiday=GR_new2$X3), se=F,
    level = 1 - f.alp/g, interval="prediction")
```

```
##        fit      lwr      upr
## 1 4232.171 3849.911 4614.430
## 2 4250.545 3871.478 4629.613
## 3 4276.791 3900.122 4653.460
## 4 4326.649 3947.913 4705.385
```

- Conclusion: at a family confidence level of 95%, the joint prediction intervals of the labor hours of the 4 shipments are
  - $\hat{Y}_1 = 4232.171$, PI is (3849.911, 4614.430)
  - $\hat{Y}_2 = 4250.545$, PI is (3871.478, 4629.613)
  - $\hat{Y}_3 = 4276.791$, PI is (3900.122,4653.460)
  - $\hat{Y}_4 = 4326.649$, PI is (3947.913,4705.3859)
- *Remark*: Please refer to class examples and previous homework to review how to construct the PI for individual response (new Y) by using the formula: $\hat{y} \pm (critical\ value) * se(\hat{y}_{new})$, where $se(\hat{y}_{new}) = \sqrt{(se(\hat{y}_{mean}))^2 + MSE}$ and the **critical value** can be t_{crit} (the regular t-critical value), B (t_critical value with Bonferroni adjustment), or S (Scheffe critical value).

Patient satisfaction data (Problem 6.15)

A hospital administrator wished to study the l-elation between patient satisfaction (Y) and patient's age (X_1, in years), severity of illness (X_2, an index), and anxiety level (X_3 , an index). The administrator randomly

selected 46 patients and collected the data presented below, where larger values of Y, X_2 , and X_3 are, respectively, associated with more satisfaction, increased severity of illness, and more anxiety.

```
ps <- read.table("./CH06PR15.txt", header = F)
colnames(ps) <- c("Y_sa", "X1_age", "X2_severity", "X3_anxiety")
```

6. Problem 7.5 Patient satisfaction (b, c. 10 pts)

```
ps.mlr <- lm(Y_sa ~ X2_severity +X1_age +  X3_anxiety, data = ps)
```

a. (Optional) Obtain the analysis of variance table that decomposes the regression sum of squares into extra sums of squares associated with X_2; with X_1 given X_2; and with X_3 , given X_2 and X_1.

```
anova(ps.mlr)
```

```
## Analysis of Variance Table
##
## Response: Y_sa
##             Df Sum Sq Mean Sq F value    Pr(>F)
## X2_severity  1 4860.3  4860.3 48.0439 1.822e-08 ***
## X1_age       1 3896.0  3896.0 38.5126 2.008e-07 ***
## X3_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
```

- $SSR(X_2) = 4860.3$
- $SSR(X_1|X_2) = 3896.0$ $(= SSR(X_1, X_2) - SSR(X_2))$
- $SSR(X_3|X_1, X_2) = 364.2$ $(= SSR(X_1, X_2, X_3) - SSR(X_1, X_2))$
- $SSE(X_1, X_2, X_3) = 4248.8$
- df are 1, 1, 1, 42, respectively.

b. Test whether X_3 can be dropped from the regression model given that X_1 and X_2 are retained. Use the F^* test statistic and level of significance 0.025. State the alternatives, decision rule, and conclusion. What is the p-value of the test?

- Full model:$Y_{sa} = \beta_0 + \beta_1 X_{age} + \beta_2 X_{serverity} + \beta_3 X_{anxiety} + \epsilon$
- Hypothesis:
  - $H_0 : \beta_3 = 0$, I.e. The anxiety level can be dropped from the model while retaining age and serverity.
  - $H_a : \beta_3 \neq 0$, I.e. The anxiety level is needed in the model.
- Reduced model: $Y_{sa} = \beta_0 + \beta_1 X_{age} + \beta_2 X_{illness} + \epsilon$
- Partial F-test using software directly (doesn't apply to SPSS)

```
ps.mlrb<- lm(Y_sa ~ X1_age + X2_severity, data=ps)
anova(ps.mlrb, ps.mlr)
```

```
## Analysis of Variance Table
##
## Model 1: Y_sa ~ X1_age + X2_severity
## Model 2: Y_sa ~ X2_severity + X1_age + X3_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
```

- Partial F test using formulas. (SSE are obtained from the software output. You can use either `anova(full.model)` or `anova(reduced.model, full.model)` to obtain the SSE.)
  - $SSE(F) = SSE(X_1, X_2, X_3) = 4248.8$, $df_E(F) = 42$.
  - $SSE(R) = SSE(X_1, X_2) = 4613$, $df_E(R) = 43$.
  - $F_{partial} = \frac{[SSE(R) - SSE(F)]/(df_E(R) - df_E(F))}{SSE(F)/df_E(F)} = \frac{(4613 - 4248.8)(43-42)}{4248.8/42} = 3.60$
  - p-value $= P(F_{(df_1 = df_E(R) - df_E(F),\ df_2 = df_E(F))} > F_{partial}) = P(F_{(df_1=1,\ df_2=42)} > 3.6) = 0.0647$

```
p <- 1- pf(3.60, 1, 42)
p
```

```
## [1] 0.06466867
```

- Decision rules:
  - Reject H_0 if p-value $< \alpha$.
  - Alternatively, using F-critictal value . Reject H_0 when $F_{partial} < F_{(1-\alpha,\ df1,\ df2)} = F(0.975; 1, 42) = 5.40$

```
qf(0.975, 1, 42)
```

```
## [1] 5.403859
```

- Since p-value $= 0.0647 > \alpha = 0.025$ (and $F_{partial} = 3.60 < F(0.975; 1, 42) = 5.40$), we do not reject H_0. At 0.025 significance level, We do not have significant evidence to support that the anxiety level is needed in the model when age and severity of illness are in the model. Therefore, the anxiety level can be dropped from the model when the other 2 predictors are in the model.

c. Repeat part (b), but use a t-test. Report the t-test statistic and p-value from the output. Compare the t-test results to the Partial F-test result from part (b), and comment.

```
summary(ps.mlr)
```

```
##
## Call:
## lm(formula = Y_sa ~ X2_severity + X1_age + X3_anxiety, data = ps)
##
## 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 ***
## X2_severity  -0.4420     0.4920  -0.898   0.3741
## X1_age       -1.1416     0.2148  -5.315 3.81e-06 ***
## X3_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
```

- Hypothesis:

  - $H_0 : \beta_3 = 0$, I.e. The anxiety level can be dropped from the model while retaining age and serverity.
  - $H_a : \beta_3 \neq 0$, I.e. The anxiety level is needed in the model.
- t-test statistic and p-value (these results are also in the `summary()` output).
  - $t_{obs} = (b_3 - 0)/se(b3) = (-13.4702 - 0)/7.0997 = -1.8973$
  - p-value $= 2 \cdot P(t_{(df=df_E)} > |t_{obs}) = 2 \cdot P(t_{df=42}) > | - 1.8973| = 0.0647$.

```
tobs <- (-13.4702 - 0)/ 7.0997
data.frame("t_statistic" = tobs, "p_value"= 2* pt(tobs, df=42))
```

```
##   t_statistic    p_value
## 1   -1.897291 0.06467884
```

- Since the p-value is $0.0647 > \alpha = 0.025$ (and $F_{partial} = 3.60 < F(0.975; 1, 42) = 5.40$), we do not reject H_0. At 0.025 significance level, We do not have significant evidence to support that the anxiety level is needed in the model when age and severity of illness are in the model. Therefore, the anxiety level can be dropped from the model when the other 2 predictors are in the model.
- The resulting p-values from the t-test and the partial F-test are the same. They both lead to the same conclusion.

7. Problem 7.6 Patient satisfaction (10pts. 3 for formula-based calculation, 4 for models and software, 3 for Ho/Ha and conclusion.)

Test whether both X_2 and X_3 can be dropped from the regression model given that X_1 is retained. Use $\alpha = 0.025$. State the alternatives, decision rule, and conclusion. What is the p-value of the test?

- Full model:$Y_{sa} = \beta_0 + \beta_1 X_{age} + \beta_2 X_{illness} + \beta_3 X_{al} + \epsilon$
- Hypothesis:
  - $H_0 : \beta_2 = \beta_3 = 0$, I.e. Both severity of illness and the anxiety level can be dropped while age is in the model.
  - H_a: At least one $\beta \neq 0$, I.e. At least one of severity of illness and the anxiety level should remain in the model, while age is in the model.
- Reduced model: $Y_{sa} = \beta_0 + \beta_1 X_{age} + \epsilon$.

```
ps.mlr4<- lm(Y_sa~X1_age, data=ps)
anova(ps.mlr4, ps.mlr)
```

```
## Analysis of Variance Table
##
## Model 1: Y_sa ~ X1_age
## Model 2: Y_sa ~ X2_severity + X1_age + X3_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
```

- Formula-based calculation: (the resulting F-statistic and p-value match the output above.)
  - $SSE(F) = SSE(X_1, X_2, X_3) = 4248.84$
  - $SSE(R) = SSE(X_1) = 5093.9$
  - $SSE(R) - SSE(F) = 845.07$ (This is also $SSR(X_2, X_3|X_4)$, extra sum of squares.)
  - $F_{partial} = \frac{(845.07/2)}{(4248.84/42)} = 4.1768$
  - p-value $= P(F_{(2,42)} > 4.1768) = 0.022$

- Decision rule:
  - Reject H_0 when p-value $< \alpha$.
- Since the p-value is 0.02216 is smaller than $\alpha = 0.025$, We should reject H_0. At 0.025 significance level, we have significant evidence to support that at least one of X_2 and X_3 should remain in the model given that X_1 is retained.

8. Problem 7.9 Patient satisfaction (10pts. 3 for formula-based calculation, 4 for running models, 3 for Ho/Ha and conclusion.)

Test whetherβ_1 = -1.0 and$\beta_2 = 0$; use $\alpha = 0.025$ State the alternatives, full and reduced models, decision rule, and conclusion.

- Full model:$Y_{sa} = \beta_0 + \beta_1 X_{age} + \beta_2 X_{severity} + \beta_3 X_{anxiety} + \epsilon_i$
- Hypothesis:
  - $H_0 : \beta_1 = -1, \beta_2 = 0$.
  - H_a: At least one is inequality.
- Reduced model: $Y_{sa} + X_{age} = \beta_0 + \beta_3 X_{anxiety} + \epsilon_i$
- Calculation: (Use the MSE from the above output)

```
ps$NewY <- ps$Y_sa + ps$X1_age
ps.mlr5 <- lm(NewY~X3_anxiety,data = ps)
anova(ps.mlr)  # Full
```

```
## Analysis of Variance Table
##
## Response: Y_sa
##              Df Sum Sq Mean Sq F value    Pr(>F)
## X2_severity   1 4860.3  4860.3 48.0439 1.822e-08 ***
## X1_age        1 3896.0  3896.0 38.5126 2.008e-07 ***
## X3_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
```

```
anova(ps.mlr5)  # Reduced
```

```
## Analysis of Variance Table
##
## Response: NewY
##             Df Sum Sq Mean Sq F value    Pr(>F)
## X3_anxiety   1 1636.3 1636.26   16.26 0.0002162 ***
## Residuals   44 4427.7  100.63
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

  - $SSE(F) = 4248.84$, $df_E F = 42$
  - $SSE(R) = 4427.7$, $df_E R = 44$
  - $F_{partial} = \frac{[SSE(R) - SSE(F)]/(df_E(R) - df_E(F))}{SSE(F)/df_E(F)} = \frac{[(4427.7 - 4248.84)/2]}{(4248.84/42)} = 0.8840$
  - p-value $= P(F_{(2,42)} > 0.8840) = 0.42$

```
f <- (4427.7-4248.8)/(44-42)/(4248.8/42)
f
```

```
## [1] 0.8842261
```

```
1-pf(f, 2, 42)
```

```
## [1] 0.4205815
```

- Decision rule:
  - Reject H_0 when p-value $< \alpha$.
- Conclusion: Since the resulting p-value $= 0.42 > \alpha = 0.025$, We do not reject H_0. There is not significant evidence to doubt that $\beta_1 = -1$,and $\beta_2 = 0$ at 0.025 significance level.

—— ***This is the end of HW 5.*** ——