HW9

```
library(alr4)
```

```
## Loading required package: car
```

```
## Loading required package: carData
```

```
## Loading required package: effects
```

```
## lattice theme set by effectsTheme()
## See ?effectsTheme for details.
```

Problem1

These two models are not nested model. So we can't compare these model using F-test.

```
head(UN11)
```

```
##                 region  group fertility   ppgdp lifeExpF pctUrban
## Afghanistan       Asia  other     5.968   499.0    49.49       23
## Albania         Europe  other     1.525  3677.2    80.40       53
## Algeria         Africa africa     2.142  4473.0    75.00       67
## Angola          Africa africa     5.135  4321.9    53.17       59
## Anguilla     Caribbean  other     2.000 13750.1    81.10      100
## Argentina   Latin Amer  other     2.172  9162.1    79.89       93
```

```
attach(UN11)
model1 = lm(lifeExpF~group)
model2 = lm(lifeExpF~log(ppgdp))
```

Problem2

From summary P-value is lower than significane level. So we can reject null hypothesis which has just intercept.

```
head(fuel2001)
```

```
##     Drivers    FuelC Income  Miles      MPC      Pop  Tax
## AL  3559897  2382507  23471  94440 12737.00  3451586 18.0
## AK   472211   235400  30064  13628  7639.16   457728  8.0
## AZ  3550367  2428430  25578  55245  9411.55  3907526 18.0
## AR  1961883  1358174  22257  98132 11268.40  2072622 21.7
## CA 21623793 14691753  32275 168771  8923.89 25599275 18.0
## CO  3287922  2048664  32949  85854  9722.73  3322455 22.0
```

```
attach(fuel2001)
alt_model= lm(FuelC~Tax+Drivers+Income+log(Miles))
summary(alt_model)
```

```
##
## Call:
## lm(formula = FuelC ~ Tax + Drivers + Income + log(Miles))
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1676904  -126002   -21638   146118  1849371
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.274e+05  1.219e+06   0.186    0.853
## Tax         -2.270e+04  1.436e+04  -1.581    0.121
## Drivers      6.566e-01  2.198e-02  29.868   <2e-16 ***
## Income      -1.820e+01  1.745e+01  -1.043    0.302
## log(Miles)   7.579e+04  8.503e+04   0.891    0.377
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 453600 on 46 degrees of freedom
## Multiple R-squared:  0.974,  Adjusted R-squared:  0.9717
## F-statistic: 430.6 on 4 and 46 DF,  p-value: < 2.2e-16
```

Problem3

Alternative model is significant compare to each null model.

```
head(cakes)
```

```
##   block X1  X2    Y
## 1     0 33 340 3.89
## 2     0 37 340 6.36
## 3     0 33 360 7.65
## 4     0 37 360 6.79
## 5     0 35 350 8.36
## 6     0 35 350 7.63
```

```
attach(cakes)
model_a = lm(Y~X1+I(X1^2)+X2+I(X2^2)+I(X1*X2))
summary(model_a)
```

```
##
## Call:
## lm(formula = Y ~ X1 + I(X1^2) + X2 + I(X2^2) + I(X1 * X2))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.4912 -0.3080  0.0200  0.2658  0.5454
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.204e+03  2.416e+02  -9.125 1.67e-05 ***
## X1           2.592e+01  4.659e+00   5.563 0.000533 ***
## I(X1^2)     -1.569e-01  3.945e-02  -3.977 0.004079 **
## X2           9.918e+00  1.167e+00   8.502 2.81e-05 ***
## I(X2^2)     -1.195e-02  1.578e-03  -7.574 6.46e-05 ***
## I(X1 * X2)  -4.163e-02  1.072e-02  -3.883 0.004654 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.4288 on 8 degrees of freedom
## Multiple R-squared:  0.9487, Adjusted R-squared:  0.9167
## F-statistic:  29.6 on 5 and 8 DF,  p-value: 5.864e-05
```

```
model_n1 = lm(Y~X1+I(X1^2)+X2+I(X2^2))
anova(model_n1, model_a)
```

```
## Analysis of Variance Table
##
## Model 1: Y ~ X1 + I(X1^2) + X2 + I(X2^2)
## Model 2: Y ~ X1 + I(X1^2) + X2 + I(X2^2) + I(X1 * X2)
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1      9 4.2430
## 2      8 1.4707  1    2.7722 15.079 0.004654 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
model_n2 = lm(Y~X1+X2+I(X2^2)+I(X1*X2))
anova(model_n2, model_a)
```

```
## Analysis of Variance Table
##
## Model 1: Y ~ X1 + X2 + I(X2^2) + I(X1 * X2)
## Model 2: Y ~ X1 + I(X1^2) + X2 + I(X2^2) + I(X1 * X2)
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1      9 4.3785
## 2      8 1.4707  1    2.9077 15.816 0.004079 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
model_n3 = lm(Y~X2+I(X2^2))
anova(model_n3, model_a)
```

```
## Analysis of Variance Table
##
## Model 1: Y ~ X2 + I(X2^2)
## Model 2: Y ~ X1 + I(X1^2) + X2 + I(X2^2) + I(X1 * X2)
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1     11 11.4739
## 2      8  1.4707  3    10.003 18.137 0.0006293 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Manual

```
RSS_n1 = sum(resid(model_n1)^2)
RSS_a = sum(resid(model_a)^2)
df_n1 = 9
df_a = 8
f1 = ((RSS_n1-RSS_a)/(df_n1-df_a))/(RSS_a/df_a)
p_val1 = pf(f1,1,8, lower.tail = FALSE)
p_val1
```

```
## [1] 0.004653652
```

Problem4

$NH : b_2 = 0, AH : b_2 \neq 0$

P_value is 0.08085 so that we can't reject null hypothesis. Which means $b_2 = 0$ has probability 0.0404 to be above or below 1.751 each.

```
head(Rateprof)
```

```
##   gender numYears numRaters numCourses pepper discipline             dept
## 1   male        7        11          5     no        Hum          English
## 2   male        6        11          5     no        Hum Religious Studies
## 3   male       10        43          2     no        Hum              Art
## 4   male       11        24          5     no        Hum          English
## 5   male       11        19          7     no        Hum          Spanish
## 6   male       10        15          9     no        Hum          Spanish
##    quality helpfulness  clarity easiness raterInterest sdQuality sdHelpfulness
## 1 4.636364    4.636364 4.636364 4.818182      3.545455 0.5518564     0.6741999
## 2 4.318182    4.545455 4.090909 4.363636      4.000000 0.9020179     0.9341987
## 3 4.790698    4.720930 4.860465 4.604651      3.432432 0.4529343     0.6663898
## 4 4.250000    4.458333 4.041667 2.791667      3.181818 0.9325048     0.9315329
## 5 4.684211    4.684211 4.684211 4.473684      4.214286 0.6500112     0.8200699
## 6 4.233333    4.266667 4.200000 4.533333      3.916667 0.8632717     1.0327956
##   sdClarity sdEasiness sdRaterInterest
## 1 0.5045250  0.4045199       1.1281521
## 2 0.9438798  0.5045250       1.0744356
## 3 0.4129681  0.5407021       1.2369438
## 4 0.9990938  0.5882300       1.3322506
## 5 0.5823927  0.6117753       0.9749613
## 6 0.7745967  0.6399405       0.6685579
```

```
attach(Rateprof)
gender = as.factor(gender)
pepper = as.factor(pepper)
discipline = as.factor(discipline)
prof_model = lm(quality~gender+numYears+pepper+discipline+easiness+raterInterest)
summary(prof_model)
```

```
##
## Call:
## lm(formula = quality ~ gender + numYears + pepper + discipline +
##     easiness + raterInterest)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.63978 -0.42534  0.03105  0.41535  1.26088
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -0.18066    0.24240  -0.745  0.45658
## gendermale          0.04678    0.06492   0.721  0.47162
## numYears            0.01760    0.01005   1.751  0.08085 .
## pepperyes           0.56166    0.09934   5.654 3.22e-08 ***
## disciplineSocSci    0.01865    0.08889   0.210  0.83393
## disciplineSTEM      0.29475    0.08148   3.618  0.00034 ***
## disciplinePre-prof  0.09656    0.09139   1.057  0.29144
```

```
## easiness            0.51288    0.04245  12.082  < 2e-16 ***
## raterInterest       0.54413    0.05937   9.165  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5892 on 357 degrees of freedom
## Multiple R-squared:  0.5158, Adjusted R-squared:  0.505
## F-statistic: 47.54 on 8 and 357 DF,  p-value: < 2.2e-16
```

$NH : b_2 = 0, AH : b_2 < 0$

P_value is 0.9595852 so that we can't reject null hypothesis again.

$NH : b_2 = 0, AH : b_2 > 0$

P_value is 0.04041476 so that we can reject null hypothesis.

```
pt(1.751, 347, lower.tail = TRUE)
```

```
## [1] 0.9595852
```

```
pt(1.751, 347, lower.tail = FALSE)
```

```
## [1] 0.04041476
```

```
Anova(prof_model, type = "II")
```

```
## Anova Table (Type II tests)
##
## Response: quality
##                Sum Sq  Df  F value    Pr(>F)
## gender          0.180   1   0.5193  0.471621
## numYears        1.064   1   3.0651  0.080848 .
## pepper         11.098   1  31.9647 3.218e-08 ***
## discipline      5.007   3   4.8072  0.002698 **
## easiness       50.680   1 145.9684 < 2.2e-16 ***
## raterInterest  29.161   1  83.9908 < 2.2e-16 ***
## Residuals     123.949 357
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
prof_model2 = lm(quality~numYears+pepper+discipline+easiness+raterInterest)
Anova(prof_model, type = "II")
```

```
## Anova Table (Type II tests)
##
## Response: quality
##                Sum Sq  Df  F value    Pr(>F)
## gender          0.180   1   0.5193  0.471621
## numYears        1.064   1   3.0651  0.080848 .
## pepper         11.098   1  31.9647 3.218e-08 ***
## discipline      5.007   3   4.8072  0.002698 **
## easiness       50.680   1 145.9684 < 2.2e-16 ***
## raterInterest  29.161   1  83.9908 < 2.2e-16 ***
## Residuals     123.949 357
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
prof_model3 = lm(quality~pepper+discipline+easiness+raterInterest)
Anova(prof_model, type = "II")
```

```
## Anova Table (Type II tests)
##
## Response: quality
##                Sum Sq  Df  F value    Pr(>F)
## gender          0.180   1   0.5193  0.471621
## numYears        1.064   1   3.0651  0.080848 .
## pepper         11.098   1  31.9647 3.218e-08 ***
## discipline      5.007   3   4.8072  0.002698 **
## easiness       50.680   1 145.9684 < 2.2e-16 ***
## raterInterest  29.161   1  83.9908 < 2.2e-16 ***
## Residuals     123.949 357
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(prof_model3)
```

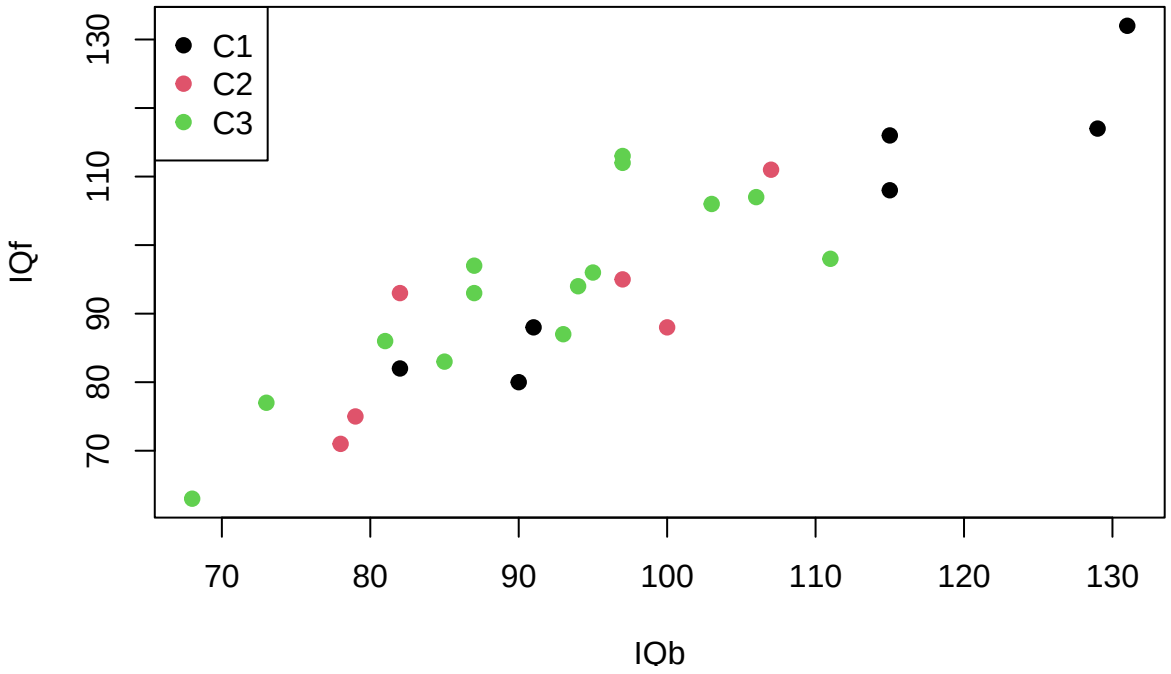
```
##
## Call:
## lm(formula = quality ~ pepper + discipline + easiness + raterInterest)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.71206 -0.41720  0.04295  0.42812  1.29817
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -0.01980    0.22697  -0.087    0.931
## pepperyes           0.52477    0.09763   5.375 1.38e-07 ***
## disciplineSocSci    0.01287    0.08906   0.144    0.885
## disciplineSTEM      0.31389    0.07973   3.937 9.91e-05 ***
## disciplinePre-prof  0.09181    0.09116   1.007    0.315
## easiness            0.50514    0.04237  11.922  < 2e-16 ***
## raterInterest       0.55556    0.05889   9.435  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5908 on 359 degrees of freedom
## Multiple R-squared:  0.5105, Adjusted R-squared:  0.5023
## F-statistic:  62.4 on 6 and 359 DF,  p-value: < 2.2e-16
```

Problem5

```r
head(twins)
```

```
##    C IQb IQf
## 1 C1  82  82
## 2 C1  90  80
## 3 C1  91  88
## 4 C1 115 108
## 5 C1 115 116
## 6 C1 129 117
```

```r
attach(twins)
C = as.factor(C)
plot(IQb, IQf,pch = 19,col=C)
legend("topleft",
       legend = levels(C),
       pch = 19,
       col = factor(levels(C)))
```




```r
twins_model = lm(IQf~IQb+C+IQb:C)
summary(twins_model)
```

```
##
## Call:
## lm(formula = IQf ~ IQb + C + IQb:C)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -14.479  -5.248  -0.155   4.582  13.798
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.872044  17.808264  -0.105    0.917
## IQb          0.977562   0.163192   5.990 6.04e-06 ***
## CC2          2.688068  31.604178   0.085    0.933
## CC3          9.076654  24.448704   0.371    0.714
## IQb:CC2     -0.004995   0.329525  -0.015    0.988
## IQb:CC3     -0.029140   0.244580  -0.119    0.906
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.921 on 21 degrees of freedom
## Multiple R-squared:  0.8041, Adjusted R-squared:  0.7574
## F-statistic: 17.24 on 5 and 21 DF,  p-value: 8.31e-07
```

```r
Anova(twins_model, type="II")
```

```
## Anova Table (Type II tests)
##
## Response: IQf
##           Sum Sq Df F value    Pr(>F)
## IQb       4674.7  1 74.5132 2.382e-08 ***
## C          175.1  2  1.3958    0.2697
## IQb:C        0.9  2  0.0074    0.9926
## Residuals 1317.5 21
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
twins_model2 = lm(IQf~IQb+C)
Anova(twins_model2, type="II")
```

```
## Anova Table (Type II tests)
##
## Response: IQf
##           Sum Sq Df F value    Pr(>F)
## IQb       4674.7  1 81.5521 5.047e-09 ***
## C          175.1  2  1.5276    0.2383
## Residuals 1318.4 23
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
twins_model3 = lm(IQf~IQb)
Anova(twins_model3, type="II")
```

```
## Anova Table (Type II tests)
##
## Response: IQf
##           Sum Sq Df F value    Pr(>F)
## IQb       5231.1  1  87.563 1.204e-09 ***
## Residuals 1493.5 25
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(twins_model3)
```

```
##
## Call:
## lm(formula = IQf ~ IQb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.3512  -5.7311   0.0574   4.3244  16.3531
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9.20760    9.29990   0.990    0.332
## IQb          0.90144    0.09633   9.358  1.2e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.729 on 25 degrees of freedom
```

```
## Multiple R-squared:  0.7779, Adjusted R-squared:  0.769
## F-statistic: 87.56 on 1 and 25 DF,  p-value: 1.204e-09
```