Sample Answer Lab02: Multiple Regression Analysis, Factors and Interaction Effects

Handed out: Monday, February 10, 2020

Return date: Monday, February 24, 2020, at the beginning of the class.

Grading: This lab counts 12 % towards your final grade

Task 1. Partial Regression Coefficient [2 points]

Use the **CONCORD1.SAV** file for this task. You will demonstrate that in multiple regression the partial effect of an independent variable is free from any linear effects of the remaining independent variables.

Task 1.1: Run the multiple model **water81~income+water80+educat** and *interpret* its regression coefficients. [0.5 points]

```
lm1<- lm(water81 ~ income + water80 + educat, data = concord)
summary(lm1)

Call:
lm(formula = water81 ~ income + water80 + educat, data = concord)

Residuals:
    Min      1Q  Median      3Q     Max 
-4635.2  -473.4   -65.0   405.6  4831.1 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 811.18243  196.29879   4.132 4.22e-05 ***
income       24.71652    3.54917   6.964 1.06e-11 ***
water80       0.59130    0.02477  23.872  < 2e-16 ***
educat      -49.88881   14.18671  -3.517 0.000478 ***
---
Signif. codes:  
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 914.9 on 492 degrees of freedom
Multiple R-squared:  0.6233,   Adjusted R-squared:  0.621 
F-statistic: 271.3 on 3 and 492 DF,  p-value: < 2.2e-16
```

Comment: The intercept and all regression coefficients are significant in this model. **Income** and **water80** have positive effects on **water81**, however **educat** has a negative influence on **water81**. When **income** increases one-thousand dollars, the water81 consumption increases 24.7 ft^3 units because affluent people have more money to support high water consumption. When **water80** increases one ft^3, **water81** increases by **0.59** ft^3. In other words, water consumption in 1981 is positively correlated the water consumption in the previous year. The water consumption in the previous year can be considered as the baseline demand not captured by the other variable in the model. Higher educated people tend to consume less water because they are concerned with saving water either because of environmental considerations or because they are better informed about saving water and, therefore, reduce their water bills. Thus, if the household head has one more year of education, the water consumption in 1981 decreases 49.9 ft^3. Overall, 62% of the variation in **water81** is explained by the independent variables.

Task 1.2: Calculate the residuals of the two models [a] **water81~income+water80** and [b] **educat~income+water80**. *What are these residuals specifically measuring*? [0.5 points]

[a] **water81~income+water80**

```
lm2 <- lm(water81~income+water80, data=concord)
summary(lm2)

Call:
lm(formula = water81 ~ income + water80, data = concord)

Residuals:
    Min      1Q  Median      3Q     Max 
-4861.1  -439.5   -67.5   382.5  4984.0 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 203.82169   94.36129   2.160   0.0313 *  
income       20.54504    3.38341   6.072 2.52e-09 ***
water80       0.59313    0.02505  23.679  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 925.4 on 493 degrees of freedom
Multiple R-squared:  0.6138,	Adjusted R-squared:  0.6122 
F-statistic: 391.8 on 2 and 493 DF,  p-value: < 2.2e-16
```

[b] **educat~income+water80**

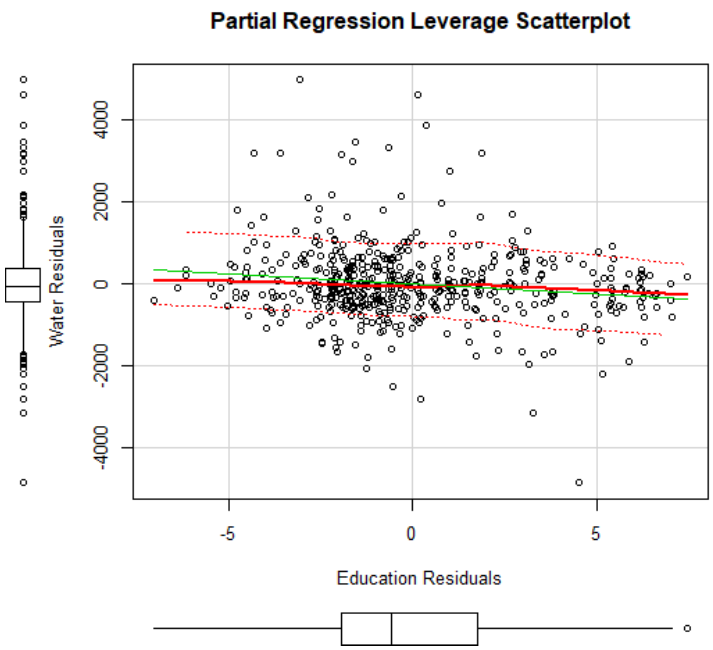
```
lm3 <- lm(educat~income+water80, data=concord)
summary(lm3)

Call:
lm(formula = educat ~ income + water80, data = concord)

Residuals:
    Min      1Q  Median      3Q     Max 
-7.0278 -1.9509 -0.5896  1.7503  7.4588 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.217e+01  2.962e-01  41.106  < 2e-16 ***
income       8.362e-02  1.062e-02   7.874 2.21e-14 ***
water80     -3.654e-05  7.862e-05  -0.465    0.642    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.905 on 493 degrees of freedom
Multiple R-squared:  0.1203,	Adjusted R-squared:  0.1167 
F-statistic:  33.7 on 2 and 493 DF,  p-value: 1.911e-14
```

Comment: Residuals of model [a] and model [b] measure the unexplained variations of **water81** and **educat**, respectively, after controlling for the influence from **income** and **water80**. We can observe the regression coefficients of **income** have significant positive effects in both the model [a] and the model [b]. This positive

confounding correlation between **educat** and **income** causes **educat** in a bivariate model having a positive effect on **water81** compared to the multiple model. To solve this problem, we need to control for the effect of the confounding variable **income** to get the pure negative effect of **educat**.

Task 1.3: Generate the partial regression leverage scatterplot of the water residuals against the education residuals. Make sure to use properly labeled axes. *Briefly interpret the scatterplot*. [0.5 points]

```
library(car)
scatterplot(resid(lm2)~resid(lm3), xlab="Education Residuals",
            ylab="Water Residuals", main="Partial Regression Leverage
Scatterplot")
```



Comment: the water residuals and education residuals have a negative relationship. When education residuals increase, the water residuals decrease.

Task 1.4: Estimate a regression model of the water residuals on the education residuals and *compare* its estimate slope coefficient against the slope coefficient for **educat** of the multiple model from task 1.1. *Why are you allowed to suppress the intercept in this model*? [0.5 points]

```
> lm4 <- lm(resid(lm2)~resid(lm3))
> summary(lm4)

Call:
lm(formula = resid(lm2) ~ resid(lm3))

Residuals:
    Min      1Q  Median      3Q     Max
-4635.2  -473.4   -65.0   405.6  4831.1

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.757e-14  4.100e+01   0.000 1.000000
resid(lm3)  -49.89      1.416e+01  -3.524 0.000465 ***
```

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

Residual standard error: 913.1 on 494 degrees of freedom
Multiple R-squared:  0.02452, Adjusted R-squared:  0.02254
F-statistic: 12.42 on 1 and 494 DF,  p-value: 0.0004651
```

Comment: The estimate slope coefficient in this model and the slope coefficient for **educat** of the multiple model from task 1.1 are identical because both models control the confounding effect of **income.** The intercept can be suppressed because the mean of both residual vectors is zero, i.e., they are centered around zero. Therefore, the origin point (0,0) is on the regression line.

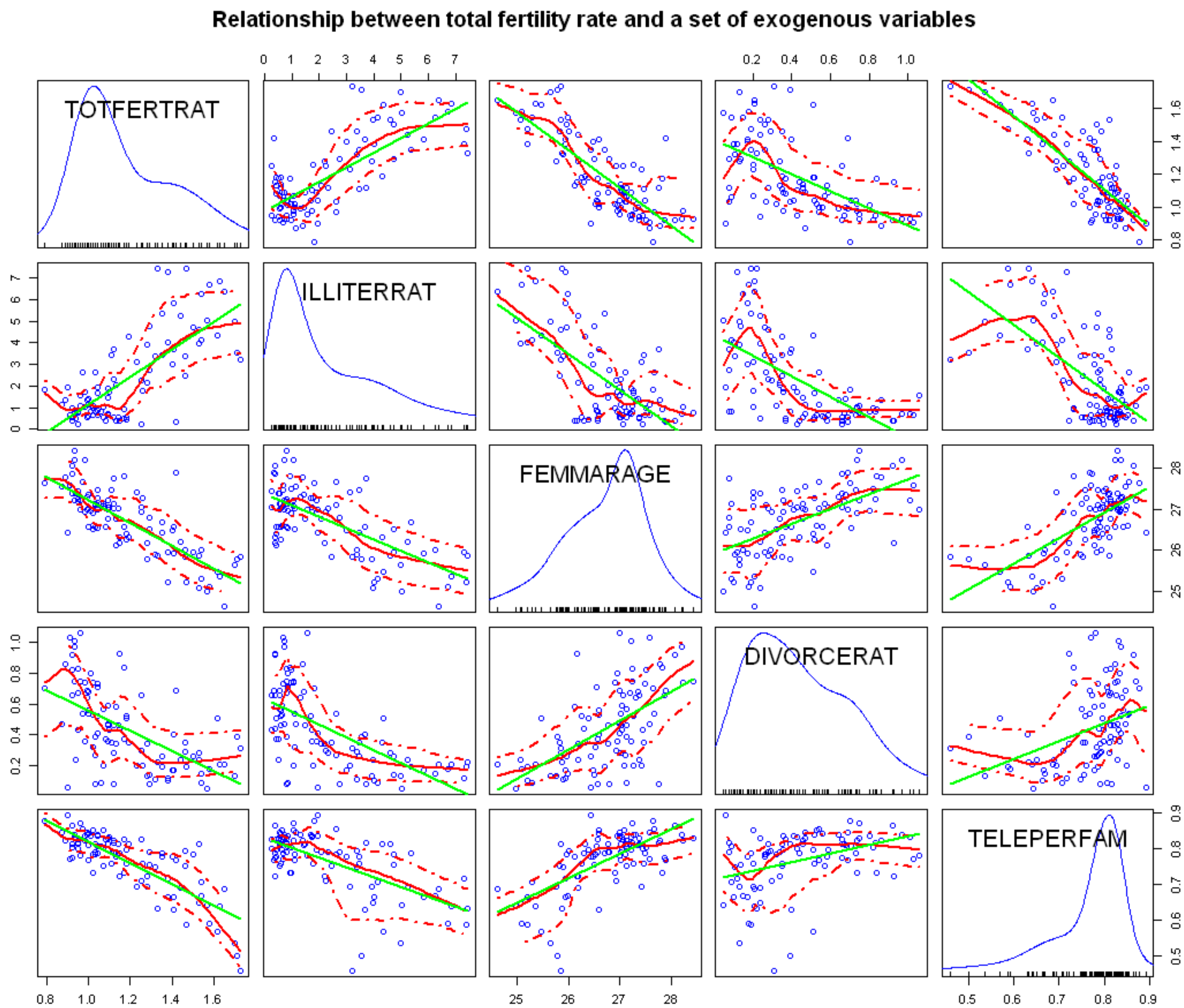
Task 2: A Multiple Regression Model with Factors and Partial *F*-test [4 points]

Task 2.1: Use common sense arguments ***how*** these four metric variables will influence the provincial fertility rates. Use one or two sentences per explanation, and formulate one or two-sided null and alternative hypotheses based on your explanation. Format everything in at table. [0.5 points]

Variable	Common Sense Arguments	Statistical Hypotheses
ILLITERRAT	A higher illiteracy rate leads to higher fertility rate due to lack of education.	$H_{0}: \beta \leq 0$ $H_{1}: \beta > 0$
FEMMARAGE	The latter a woman marries the lower will be her likelihood to have many children.	$H_{0}: \beta \geq 0$ $H_{1}: \beta < 0$
DIVORCERAT	A higher divorce rate leads to lower chance of having many children.	$H_{0}: \beta \geq 0$ $H_{1}: \beta < 0$
TELEPERFAM	An increased number of televisions will lead to more distractions and decreased fertility rate.	$H_{0}: \beta \geq 0$ $H_{1}: \beta < 0$

Task 2.2: Generate a scatterplot matrix showing the dependent variable and the four metric independent variables. *Briefly interpret the scatterplot matrix.* [0.5 points]

```
setwd("E:\\Lectures2020\\GISC7310\\Labs\\Lab02")
provItaly <- foreign::read.dbf("provinces.dbf")
car::scatterplotMatrix(~TOTFERTRAT+ILLITERRAT+FEMMARAGE+DIVORCERAT+TELEPERFAM, data=provItaly,
              main="Relationship between total fertility rate and a set of exogenous variables",
              pch=1, smooth=list(span = 0.35,lty.smooth=1, col.smooth="red", col.var="red"),
              regLine=list(col="green"))
```



Comments:

[a] Distributional characteristics: The distributions of the dependent variable and the four independent variables are unimodal. **TOTFERTRAT**, **DIVORCERAT**, and **ILLITERRAT** are positively skewed, and **FEMMARAGE** and **TELEPERFAM** are negatively skewed.
[b] Y-X relationships: **FEMMARAGE**, **DIVORCERAT**, and **TELEPERFAM** have strong negative effects on **TOTFERTRAT**. However, **ILLITERRAT** has a positive relationship with **TOTFERTRAT**.
[c] Positive X-X relationships: **FEMMARAGE- DIVORCERAT**, **FEMMARAGE- TELEPERFAM**, and **DIVORCERAT-TELEPERFAM** have positive relationships.

[d] Negative X-X relationships: **FEMMARAGE- ILLITERRAT**, **DIVORCERAT- ILLITERRAT**, and **ILLITERRAT-TELEPERFAM** have negative relationships.

Task 2.3: Run a base model multiple regression with the four metric variables to explain the variation of the fertility rates. Interpret this model [a] in the light of your earlier stated hypotheses in task 2.1, [b] the significances of the estimate regression coefficients and [c] the goodness of fit. [0.5 points]

```
lm1<- lm(TOTFERTRAT~ ILLITERRAT+ FEMMARAGE+ DIVORCERAT+ TELEPERFAM,
data=Province)
summary(lm1)
```

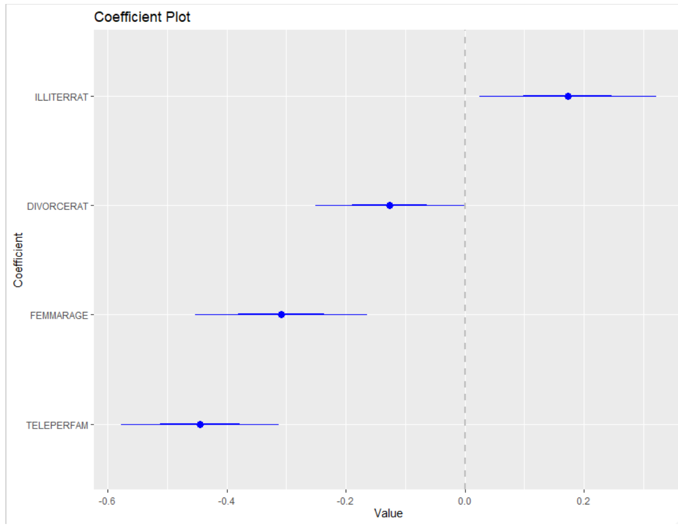
```
Call:
lm(formula = TOTFERTRAT ~ ILLITERRAT + FEMMARAGE + DIVORCERAT +
    TELEPERFAM, data = Province)

Residuals:
     Min       1Q   Median       3Q      Max
-0.21906 -0.06267 -0.00966  0.05425  0.41272

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.496337   0.513726   8.752 1.13e-13 ***
ILLITERRAT   0.020377   0.008735   2.333   0.0219 *
FEMMARAGE   -0.088837   0.020771  -4.277 4.71e-05 ***
DIVORCERAT  -0.112265   0.055648  -2.017   0.0466 *
TELEPERFAM  -1.226364   0.183037  -6.700 1.76e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1035 on 90 degrees of freedom
Multiple R-squared:  0.8096,  Adjusted R-squared:  0.8012
F-statistic: 95.69 on 4 and 90 DF,  p-value: < 2.2e-16
```

Comment: All independent variables exhibit a relationship with the dependent variable as stated by the one-sided alternative hypotheses in task 2.1. All regression coefficients are significantly different from zero at an error probability of $\alpha = 0.05$. Since the reported error probabilities are associated with two-sided tests; for one-sided tests they need to be divided by 2. The overall goodness of fit of this model is high ($R^2_{adj} = 0.8012$).

Task 2.4: Calculate the standardized *beta-coefficients* for the multiple model in task 2.3. Rank the independent variables *according to the absolute strength of their effects* on the fertility rates and plot the beta coefficients with the **`coefplot( )`** function. Use proper options for the **`coefplot( )`** function. [1 point]

```
prov <- foreign::read.dbf("provinces.dbf")
prov <- prov[,13:17]
prov <- as.data.frame(scale(prov))
beta.lm <- lm(TOTFERTRAT~., data=prov)
summary(beta.lm)
coefplot::coefplot(beta.lm, sort="magnitude", intercept=F)
```

Variables	Coefficients (absolute value)	Rank
TELEPERFAM	0.44537	1
FEMMARAGE	0.30905	2
ILLITERRATE	0.17303	3
DIVORCERAT	0.12638	4



Comment: The influence strengths of the independent variables on the variation of the dependent variable are: **DIVORCERAT < ILLITERRAT < FEMMARAGE < TELEPERFAM.**

Task 2.5: Run five separate regressions on the [a] independent variables and [b] the fertility rates using the factor **REGION** as independent variable.

Does the **REGION** factor *explain the variation* of the four independent variables as well as the fertility rates, i.e., is this factor highly correlated with other variables? [0.5 points]

```
lm2 <-lm(cbind(TOTFERTRAT,ILLITERRAT,FEMMARAGE,DIVORCERAT,TELEPERFAM)~REGION,
           data=Province)
summary(lm2)
Response TOTFERTRAT :

Call:
lm(formula = TOTFERTRAT ~ REGION, data = Province)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.23300 -0.09275 -0.01300  0.07167  0.36333 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)     1.02300    0.02205  46.405   <2e-16 ***
REGIONNorth     0.03367    0.03118   1.080    0.283    
REGIONSardinia  0.09950    0.06427   1.548    0.125    
REGIONSicily    0.53700    0.04589  11.702   <2e-16 ***
REGIONSouth     0.39473    0.03389  11.647   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1207 on 90 degrees of freedom
Multiple R-squared:  0.7409,	Adjusted R-squared:  0.7294 
F-statistic: 64.34 on 4 and 90 DF,  p-value: < 2.2e-16


Response ILLITERRAT :

Call:
lm(formula = ILLITERRAT ~ REGION, data = Province)
```

```
Residuals:
     Min       1Q   Median       3Q      Max 
-2.82455 -0.50394 -0.07267  0.34756  2.83545 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)      1.4157     0.1844   7.676 1.89e-11 ***
REGIONNorth     -0.7630     0.2608  -2.925  0.00435 ** 
REGIONSardinia   1.8068     0.5377   3.360  0.00114 ** 
REGIONSicily     3.2466     0.3839   8.457 4.64e-13 ***
REGIONSouth      3.1689     0.2835  11.176  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.01 on 90 degrees of freedom
Multiple R-squared:  0.7485,  Adjusted R-squared:  0.7373 
F-statistic: 66.97 on 4 and 90 DF,  p-value: < 2.2e-16


Response FEMMARAGE :

Call:
lm(formula = FEMMARAGE ~ REGION, data = Province)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.96636 -0.31017 -0.04033  0.29057  1.21000 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)    27.25033    0.09179 296.888  < 2e-16 ***
REGIONNorth    -0.27333    0.12981  -2.106    0.038 *  
REGIONSardinia  0.18217    0.26760   0.681    0.498    
REGIONSicily   -1.88033    0.19107  -9.841 6.11e-16 ***
REGIONSouth    -1.19397    0.14111  -8.461 4.54e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5027 on 90 degrees of freedom
Multiple R-squared:  0.6288,  Adjusted R-squared:  0.6124 
F-statistic: 38.12 on 4 and 90 DF,  p-value: < 2.2e-16


Response DIVORCERAT :

Call:
lm(formula = DIVORCERAT ~ REGION, data = Province)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.51233 -0.10767 -0.01267  0.13591  0.46767 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)     0.59233    0.03703  15.994  < 2e-16 ***
REGIONNorth    -0.04967    0.05237  -0.948    0.346    
```

```
REGIONSardinia -0.36733    0.10797  -3.402    0.001 ***
REGIONSicily   -0.34789    0.07709  -4.513 1.93e-05 ***
REGIONSouth    -0.38324    0.05694  -6.731 1.53e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2028 on 90 degrees of freedom
Multiple R-squared:  0.423,   Adjusted R-squared:  0.3973
F-statistic: 16.49 on 4 and 90 DF,  p-value: 3.596e-10


Response TELEPERFAM :

Call:
lm(formula = TELEPERFAM ~ REGION, data = Province)

Residuals:
      Min        1Q    Median        3Q       Max
-0.247360 -0.025351  0.008611  0.033786  0.106420

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     0.812704   0.010775  75.422  < 2e-16 ***
REGIONNorth    -0.006805   0.015239  -0.447    0.656
REGIONSardinia -0.049512   0.031416  -1.576    0.119
REGIONSicily   -0.180715   0.022431  -8.057 3.12e-12 ***
REGIONSouth    -0.106844   0.016566  -6.449 5.48e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05902 on 90 degrees of freedom
Multiple R-squared:  0.5306,  Adjusted R-squared:  0.5098
F-statistic: 25.43 on 4 and 90 DF,  p-value: 4.118e-14
```

Comment: **Region** is a significant variable in all five models. It can explain the 42.3% variation of **DIVORCERAT**, 74.85% variation of **ILLITERRAT**, 62.88% variation of **FEMMARAGE**, 53.06% variation of **TELEPERFAM**, and 74.09% variation of **TOTFERTRAT**. In other words, **Region** is highly correlated with the dependent variable and all four independent variables.

Task 2.6: Run the multiple regression model with the four metric variables plus the **REGION** factor to explain the variation of the fertility rates.
Speculate in an informed way why some independent metric variables are no longer significant? [0.5 points]

```
lm3 <- lm(TOTFERTRAT~FEMMARAGE+DIVORCERAT+ILLITERRAT+TELEPERFAM+REGION,
           data=Province)
summary(lm3)

Call:
lm(formula = TOTFERTRAT ~ FEMMARAGE + DIVORCERAT + ILLITERRAT +
    TELEPERFAM + REGION, data = Province)

Residuals:
     Min       1Q   Median       3Q      Max
-0.17487 -0.06724  0.00231  0.04516  0.39168
```

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      3.538560   0.609124   5.809 1.03e-07 ***
FEMMARAGE       -0.060186   0.023278  -2.585 0.011409 *
DIVORCERAT      -0.086453   0.055473  -1.558 0.122795
ILLITERRAT      -0.001634   0.010915  -0.150 0.881362
TELEPERFAM      -1.011377   0.182312  -5.547 3.14e-07 ***
REGIONNorth      0.004793   0.027447   0.175 0.861794
REGIONSardinia   0.031584   0.059267   0.533 0.595473
REGIONSicily     0.216287   0.060134   3.597 0.000537 ***
REGIONSouth      0.186853   0.046051   4.057 0.000109 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09589 on 86 degrees of freedom
Multiple R-squared:  0.8438,  Adjusted R-squared:  0.8293
F-statistic: 58.09 on 8 and 86 DF,  p-value: < 2.2e-16
```

Comment: **DIVORCERAT**, **ILLITERRAT**, and **TELEPERFAM** are no longer significant in this model. And the significance of **FEMMARAGE** also decreases dramatically. This drop in the significance is induced by the high correlation of these variables with the factor **REGION** which now also captures most of the variability in the dependent variable **TOTFERTRAT**. A high degree of multicollineary is present in the independent variables.

Task 2.7: Use a partial *F*-test to check whether the model in task 2.6 has improved the model fit of the base model in task 2.3 significantly. [0.5 points]
That is, test the null hypothesis: $H_0: \beta_{Region\,1} = \beta_{Region\,2} = \cdots = \beta_{Region\,J} = 0$ against the alternative hypothesis is $H_0: \beta_{Region\,j} \neq 0 \; \textit{for at least one}\; j \in \{1,2,\ldots,J\}$.

```
anova(lm3, lm1)
Analysis of Variance Table

Model 1: TOTFERTRAT ~ FEMMARAGE + DIVORCERAT + ILLITERRAT + TELEPERFAM +
REGION
Model 2: TOTFERTRAT ~ ILLITERRAT + FEMMARAGE + DIVORCERAT + TELEPERFAM
  Res.Df     RSS Df Sum of Sq      F   Pr(>F)
1     86 0.79079
2     90 0.96406 -4  -0.17327 4.7107 0.001743 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Comment: The *p*-value (0.001743) is substantially smaller than 0.05, thus the null hypothesis can be rejected. We can conclude that the effect of the factor **REGION** is a significantly different from zero and the model in task 2.6 has improved the model fit of the base model in task 2.3 significantly.

Task 3. Identification of the Underlying Model Structure [6 points]

Use the workspace **`ModelSpecs.RData`** for this task. It contains the six data-frames **`mod1`** to **`mod6`**. Each data-frame is comprised of three variables: **`y`** for the dependent variables, **`g`** for a binary ***factor***, and **`x`** for a ***metric*** variable. Each of these data-frames is best ***statistically*** described by one of these competing models:

Name	Models Structure

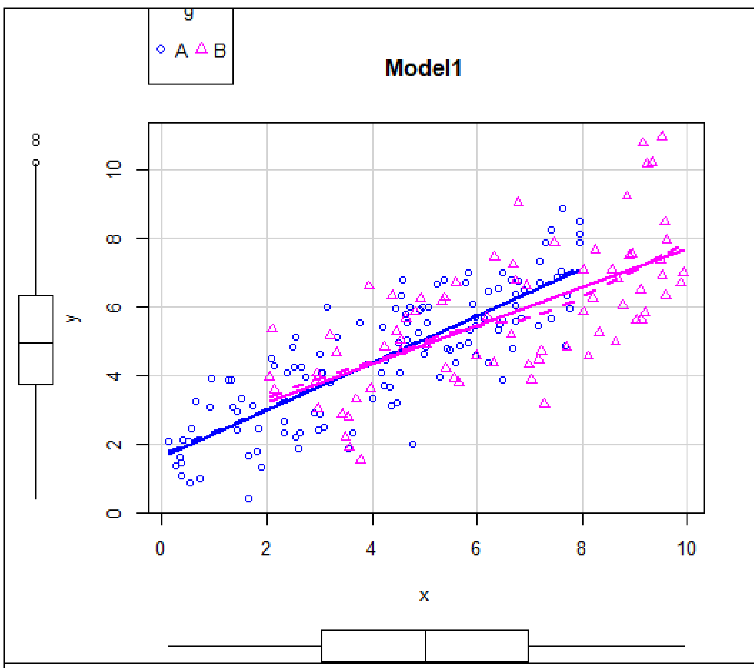
Full interaction model	`lm(y~g+x+g:x, data=mod?)` ⇔ `lm(y~g*x, data=mod?)`
Intercept model	`lm(y~g+x, data=mod?)`
Slope model	`lm(y~g:x, data=mod?)`
Means model	`lm(y~g, data=mod?)`
Plain regression model	`lm(y~x, data=mod?)`

For each of the data-frame generate an informative scatterplot showing the regression regimes for both groups of observations. You can employ the syntax:

```
car::scatterplot(y~x|g,smoother=F,boxplots="xy",data=mod?,main="Model?")
```

Then identify which of the competing model structures best describes the given data-frame. If several competing model structures seem to be reasonably relevant then try to eliminate inferior models using by looking for statistically superior $R^2_{adjusted}$, non-significant coefficients' t-tests and ***nested*** ***partial F-tests***. Provide a ***rational*** for your model selection.

Task 3.1: Identify the underlying model structure for **`mod1`**. [1 point]



Plain regression model

```
Call:
lm(formula = y ~ x, data = mod1)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.1845 -0.9699  0.0432  0.9636  3.2781 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   1.8645     0.2009   9.282   <2e-16 ***
x             0.6164     0.0354  17.416   <2e-16 ***
---
Signif. codes:  
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.26 on 198 degrees of freedom
Multiple R-squared:  0.605,    Adjusted R-squared:  0.603 
F-statistic: 303.3 on 1 and 198 DF,  p-value: < 2.2e-16
```

Means model

```
Call:
lm(formula = y ~ g, data = mod1)

Residuals:
   Min     1Q Median     3Q    Max 
-4.207 -1.401  0.050  1.272  5.203 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   4.5461     0.1714  26.519  < 2e-16 ***
gB            1.2105     0.2799   4.324 2.43e-05 ***
---
Signif. codes:  
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Intercept model

```
Call:
lm(formula = y ~ x + g, data = mod1)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.0890 -0.9401  0.0485  0.9205  3.3399 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.85417    0.20106   9.222   <2e-16 ***
x            0.63438    0.03924  16.166   <2e-16 ***
gB          -0.21584    0.20405  -1.058    0.291    
---
Signif. codes:
```

```
Residual standard error: 1.917 on 198 degrees of freedom
Multiple R-squared:  0.08629, Adjusted R-squared:
0.08167
F-statistic:  18.7 on 1 and 198 DF,  p-value: 2.425e-05
```

```
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.26 on 197 degrees of freedom
Multiple R-squared:  0.6073,  Adjusted R-squared:  0.6033
F-statistic: 152.3 on 2 and 197 DF,  p-value: < 2.2e-16
```

Full interaction model

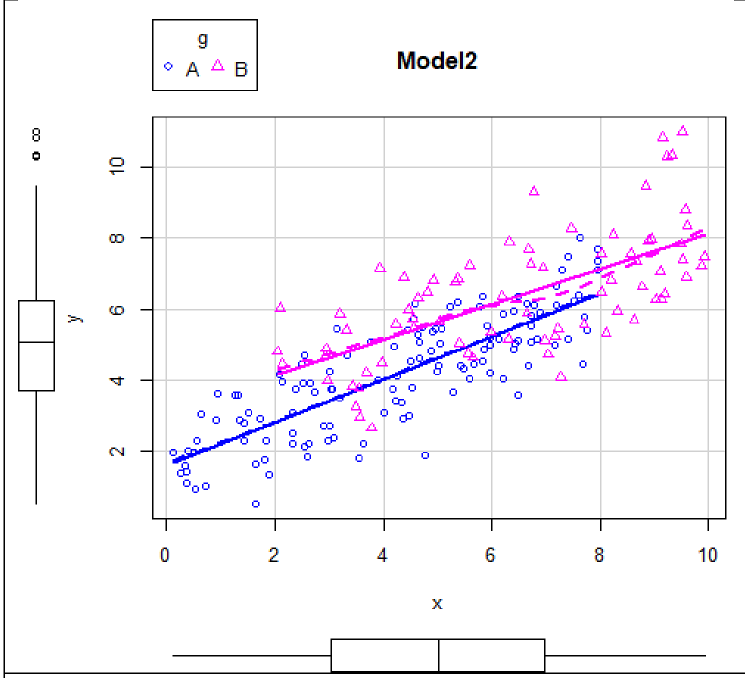
```
Call:
lm(formula = y ~ x * g, data = mod1)

Residuals:
    Min      1Q  Median      3Q     Max
-3.0298 -0.9569  0.0743  0.9122  3.5398

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.64686    0.24101   6.833 1.02e-10 ***
x            0.68324    0.05026  13.595  < 2e-16 ***
gB           0.47813    0.49235   0.971    0.333
x:gB        -0.12382    0.08001  -1.548    0.123
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.255 on 196 degrees of freedom
Multiple R-squared:  0.612,   Adjusted R-squared:
0.6061
F-statistic: 103.1 on 3 and 196 DF,  p-value: < 2.2e-16
```

Slope model

```
Call:
lm(formula = y ~ x:g, data = mod1)

Residuals:
    Min      1Q  Median      3Q     Max
-3.0276 -0.9537  0.0884  0.8988  3.4489

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.76142    0.21013   8.383 9.75e-15 ***
x:gA         0.66210    0.04529  14.619  < 2e-16 ***
x:gB         0.60904    0.03555  17.130  < 2e-16 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.255 on 197 degrees of freedom
Multiple R-squared:  0.6101,  Adjusted R-squared:  0.6062
F-statistic: 154.2 on 2 and 197 DF,  p-value: < 2.2e-16
```

Comment: Except the mean model, the relatively R^2_{adj} values of the rest four models are very close. After conducting the nested partial *F*-test, we can conclude the plain regression model is not significantly different from the other three models due to such large *p*-value. Based on the parsimony rule, we should choose the ***plain regression*** model for mod1.

Task 3.2: Identify the underlying model structure for **mod2**. [1 point]



Plain regression model

```
Call:
lm(formula = y ~ x, data = mod2)

Residuals:
    Min      1Q  Median      3Q     Max
-2.8831 -0.9128 -0.0825  0.8239  3.2223

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.74876    0.18895   9.255   <2e-16 ***
x            0.63919    0.03329  19.199   <2e-16 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.185 on 198 degrees of freedom
Multiple R-squared:  0.6505,  Adjusted R-squared:  0.6488
F-statistic: 368.6 on 1 and 198 DF,  p-value: < 2.2e-16
```

Means model

```
Call:
lm(formula = y ~ g, data = mod2)

Residuals:
```

Intercept model

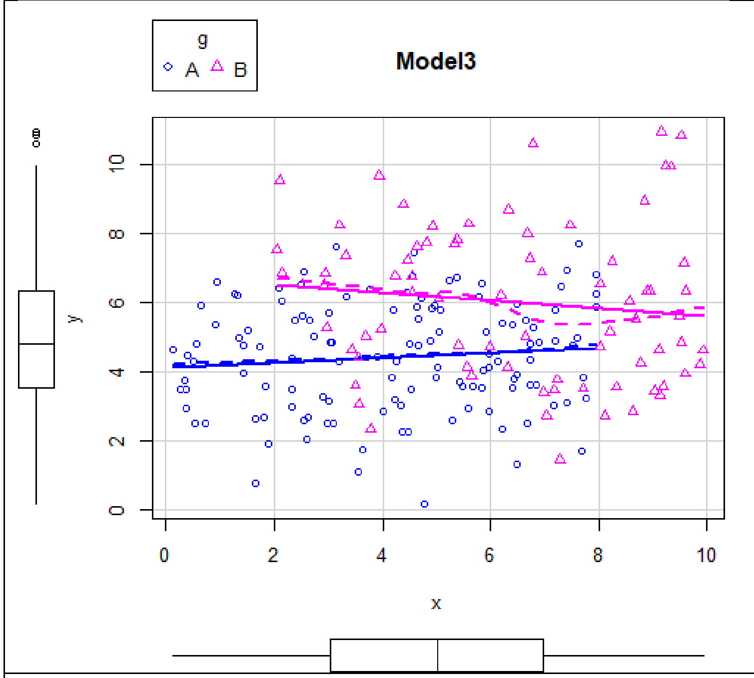
```
Call:
lm(formula = y ~ x + g, data = mod2)

Residuals:
```

<pre>   Min   1Q Median   3Q   Max
-3.7182 -1.2386 0.0442 1.1243 4.5981
 	
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.1731   0.1515 27.545 < 2e-16 ***
gB   2.2051   0.2474   8.913 3.27e-16 ***
--- 	
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
 	
Residual standard error: 1.694 on 198 degrees of freedom
Multiple R-squared: 0.2863, Adjusted R-squared:
0.2827
F-statistic: 79.45 on 1 and 198 DF, p-value: 3.271e-16</pre>
	<pre> Min 1Q Median 3Q Max
	-2.72996 -0.83083 0.04285 0.81349 2.95172
	
	Coefficients:
	            Estimate Std. Error t value Pr(>|t|)
	(Intercept) 1.79407 0.17769 10.097 < 2e-16 ***
	x 0.56065 0.03468 16.166 < 2e-16 ***
	gB 0.94458 0.18034 5.238 4.15e-07 ***
	--- 	
	Signif. codes:
	0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
	
	Residual standard error: 1.113 on 197 degrees of freedom
	Multiple R-squared: 0.6933, Adjusted R-squared: **0.6901**
	F-statistic: 222.6 on 2 and 197 DF, p-value: < 2.2e-16</pre>

Full interaction model<pre>Call:
lm(formula = y ~ x * g, data = mod2)
 	
Residuals:
     Min   1Q   Median   3Q   Max
-2.67768 -0.84564 0.06566 0.80622 3.12836
 	
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.61085   0.21299   7.563 1.49e-12 ***
x   0.60383   0.04442 13.595 < 2e-16 ***
gB   1.55788   0.43513   3.580 0.000433 ***
x:gB   -0.10943   0.07071 -1.548 0.123325
--- 	
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
 	
Residual standard error: 1.109 on 196 degrees of freedom
Multiple R-squared: 0.697,   Adjusted R-squared:
0.6923
F-statistic: 150.3 on 3 and 196 DF, p-value: < 2.2e-16</pre>
	Slope model<pre>Call:
	lm(formula = y ~ x:g, data = mod2)
	
	Residuals:
	     Min 1Q Median 3Q Max
	-2.67024 -0.88598 -0.00724 0.82261 2.85609
	
	Coefficients:
	            Estimate Std. Error t value Pr(>|t|)
	(Intercept) 1.98413 0.19122 10.38 <2e-16 ***
	x:gA 0.53495 0.04122 12.98 <2e-16 ***
	x:gB 0.65608 0.03236 20.28 <2e-16 ***
	--- 	
	Signif. codes:
	0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
	
	Residual standard error: 1.142 on 197 degrees of freedom
	Multiple R-squared: 0.6771, Adjusted R-squared: 0.6739
	F-statistic: 206.6 on 2 and 197 DF, p-value: < 2.2e-16</pre>

Comment: The intercept and full interaction models have relatively the highest R^2_{adj} values among all models. After conducting the nested partial *F*-test, we can conclude the intercept model is not significantly different from the full interaction model. Based on the parsimony rule, we should choose the ***intercept model*** for mod2.

Task 3.3: Identify the underlying model structure for **mod3**. [1 point]



Plain regression model

```
Call:
lm(formula = y ~ x, data = mod3)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.8123 -1.5236 -0.1378  1.3752  5.3784 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  4.34500    0.31539  13.777   <2e-16 ***
x            0.12877    0.05557   2.317   0.0215 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.978 on 198 degrees of freedom
Multiple R-squared:  0.0264,	Adjusted R-squared:  0.02149 
F-statistic:  5.37 on 1 and 198 DF,  p-value: 0.02151
```

Means model

```
Call:
lm(formula = y ~ g, data = mod3)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.5585 -1.3911  0.0671  1.3607  4.9206 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   4.4107     0.1658  26.605  < 2e-16 ***
gB            1.5714     0.2707   5.804 2.53e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.854 on 198 degrees of freedom
Multiple R-squared:  0.1454,	Adjusted R-squared:  
0.1411 
F-statistic: 33.69 on 1 and 198 DF,  p-value: 2.534e-08
```

Intercept model

```
Call:
lm(formula = y ~ x + g, data = mod3)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.5566 -1.3868  0.0715  1.3578  4.9268 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  4.42062    0.29658  14.905  < 2e-16 ***
x           -0.00233    0.05789  -0.040    0.968    
gB           1.57662    0.30101   5.238 4.15e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.858 on 197 degrees of freedom
Multiple R-squared:  0.1454,	Adjusted R-squared:  0.1367 
F-statistic: 16.76 on 2 and 197 DF,  p-value: 1.896e-07
```

Full interaction model

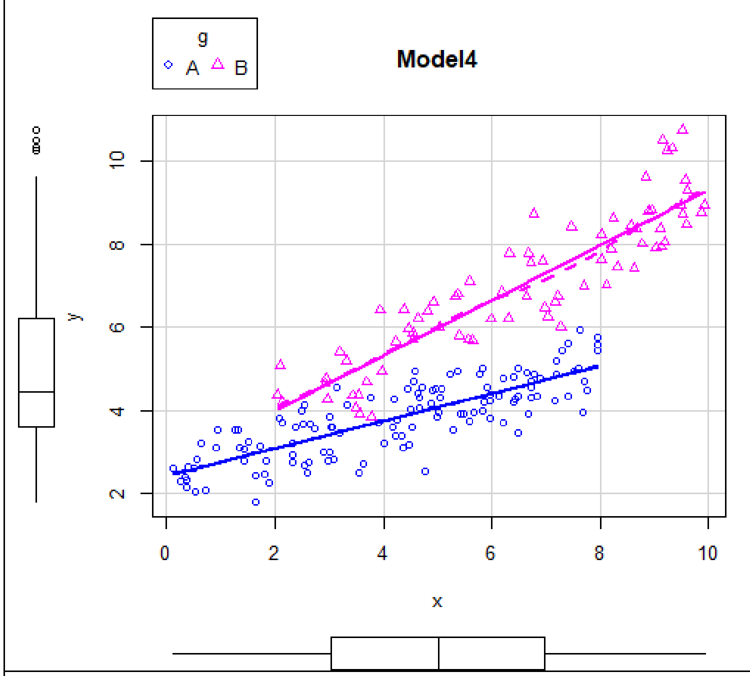
```
Call:
lm(formula = y ~ x * g, data = mod3)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.4694 -1.4115  0.1096  1.3457  5.2216 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  4.11480    0.35551  11.574  < 2e-16 ***
x            0.06974    0.07413   0.941 0.348009    
gB           2.60030    0.72628   3.580 0.000433 ***
x:gB        -0.18265    0.11802  -1.548 0.123325    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.852 on 196 degrees of freedom
Multiple R-squared:  0.1557,	Adjusted R-squared:  
0.1428 
F-statistic: 12.05 on 3 and 196 DF,  p-value: 2.834e-07
```

Slope model

```
Call:
lm(formula = y ~ x:g, data = mod3)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.4570 -1.4788 -0.0121  1.3730  4.7672 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  4.73786    0.31917  14.844  < 2e-16 ***
x:gA        -0.04523    0.06880  -0.657  0.51169    
x:gB         0.15695    0.05400   2.906  0.00408 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.906 on 197 degrees of freedom
Multiple R-squared:  0.1005,	Adjusted R-squared:  0.09139 
F-statistic: 11.01 on 2 and 197 DF,  p-value: 2.939e-05
```

Comment: The mean and full interaction models have relatively the highest R^2_{adj}-values among all models. Furthermore, the mean and the intercept model have similar R^2_{adj}-values, however, the slope coefficient of the intercept model is not significant. After conducting the nested partial *F*-test, we can conclude the mean model is not significantly different from the full interaction model. Based on the parsimony rule, we should choose the ***mean*** model for mod3.

Task 3.4: Identify the underlying model structure for **mod4**. [1 point]



Plain regression model

```
Call:
lm(formula = y ~ x, data = mod4)

Residuals:
     Min       1Q   Median       3Q      Max
-2.72136 -0.97298  0.07819  0.79625  2.91364

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.7441     0.1884   9.255   <2e-16 ***
x             0.6401     0.0332  19.278   <2e-16 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.182 on 198 degrees of freedom
Multiple R-squared:  0.6524,  Adjusted R-squared:  0.6507
F-statistic: 371.7 on 1 and 198 DF,  p-value: < 2.2e-16
```

Means model

```
Call:
lm(formula = y ~ g, data = mod4)

Residuals:
    Min      1Q  Median      3Q     Max
-3.1446 -0.8339  0.0161  0.8261  3.7666

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.8120     0.1147   33.24   <2e-16 ***
gB            3.1680     0.1873   16.92   <2e-16 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.282 on 198 degrees of freedom
Multiple R-squared:  0.591,   Adjusted R-squared:  0.589
F-statistic: 286.1 on 1 and 198 DF,  p-value: < 2.2e-16
```

Intercept model

```
Call:
lm(formula = y ~ x + g, data = mod4)

Residuals:
     Min       1Q   Median       3Q      Max
-1.89189 -0.49908 -0.02244  0.48997  2.36285

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.84605    0.11616    15.89   <2e-16 ***
x           0.46330    0.02267    20.43   <2e-16 ***
gB          2.12638    0.11790    18.04   <2e-16 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7278 on 197 degrees of freedom
Multiple R-squared:  0.8689,  Adjusted R-squared:  0.8676
F-statistic: 652.8 on 2 and 197 DF,  p-value: < 2.2e-16
```

Full interaction model

```
Call:
lm(formula = y ~ x * g, data = mod4)

Residuals:
     Min       1Q   Median       3Q      Max
-1.50698 -0.47592  0.03695  0.45373  1.76062

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
```

Slope model

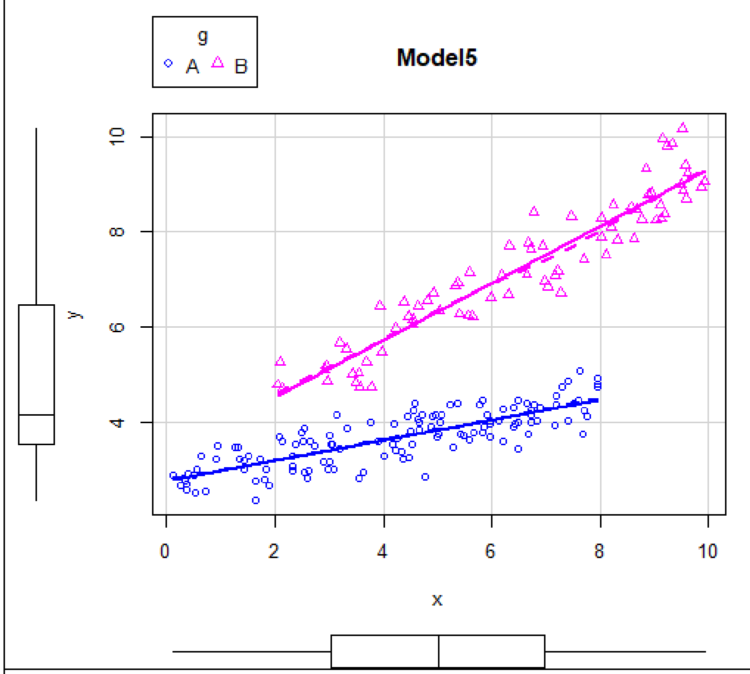
```
Call:
lm(formula = y ~ x:g, data = mod4)

Residuals:
     Min       1Q   Median       3Q      Max
-1.50585 -0.47436  0.04397  0.44703  1.71543

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)  2.41024    0.11987  20.107  < 2e-16 ***
x            0.33034    0.02500  13.215  < 2e-16 ***
gB           0.23781    0.24489   0.971    0.333
x:gB         0.33697    0.03979   8.468 5.84e-15 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6244 on 196 degrees of
freedom
Multiple R-squared:  0.904,   Adjusted R-squared:
0.9025
F-statistic: 615.3 on 3 and 196 DF,  p-value: < 2.2e-16
```

```
(Intercept)  2.46722    0.10451   23.61   <2e-16 ***
x:gA         0.31983    0.02253   14.20   <2e-16 ***
x:gB         0.69199    0.01768   39.13   <2e-16 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6243 on 197 degrees of freedom
Multiple R-squared:  0.9036,  Adjusted R-squared:  0.9026
F-statistic: 922.8 on 2 and 197 DF,  p-value: < 2.2e-16
```

Comment: The slope and full interaction models have relatively the highest R^2_{adj}-values among all models. The g variable is not significant in the full interaction model. Based on the parsimony rule, we should choose the **slope model** for mod4.

Task 3.5: Identify the underlying model structure for **mod5**. [1 point]



Plain regression model

```
Call:
lm(formula = y ~ x, data = mod5)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.7863 -1.1792  0.1347  1.0919  2.5444 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.98557    0.21221   9.357   <2e-16 ***
x            0.59264    0.03739  15.849   <2e-16 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.331 on 198 degrees of freedom
Multiple R-squared:  0.5592,  Adjusted R-squared:  0.557
F-statistic: 251.2 on 1 and 198 DF,  p-value: < 2.2e-16
```

Means model

```
Call:
lm(formula = y ~ g, data = mod5)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.52264 -0.65840  0.00285  0.58932  2.94900 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.67058    0.09143   40.15   <2e-16 ***
gB           3.54511    0.14930   23.75   <2e-16 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.022 on 198 degrees of freedom
Multiple R-squared:  0.7401,  Adjusted R-squared:
0.7388
F-statistic: 563.8 on 1 and 198 DF,  p-value: < 2.2e-16
```

Intercept model

```
Call:
lm(formula = y ~ x + g, data = mod5)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.50016 -0.39030 -0.02668  0.39717  1.83903 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.11610    0.09418   22.47   <2e-16 ***
x            0.36634    0.01838   19.93   <2e-16 ***
gB           2.72146    0.09558   28.47   <2e-16 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5901 on 197 degrees of freedom
Multiple R-squared:  0.9138,  Adjusted R-squared:  0.913
F-statistic:  1045 on 2 and 197 DF,  p-value: < 2.2e-16
```

Full interaction model

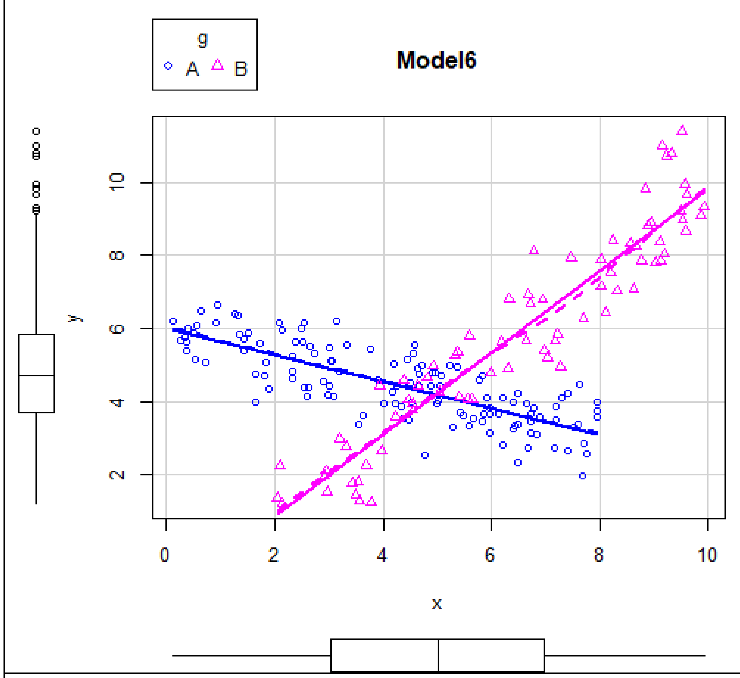
```
Call:
lm(formula = y ~ x * g, data = mod5)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.98031 -0.30959  0.02404  0.29516  1.14531 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2.75873    0.07798   35.38  < 2e-16 ***
x            0.21489    0.01626   13.21  < 2e-16 ***
gB           0.57035    0.15930    3.58 0.000433 ***
x:gB         0.38382    0.02589   14.83  < 2e-16 ***
---
Signif. codes:  
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4061 on 196 degrees of 
freedom
Multiple R-squared:  0.9594,  Adjusted R-squared:  
0.9588 
F-statistic:  1543 on 3 and 196 DF,  p-value: < 2.2e-16
```

Slope model

```
Call:
lm(formula = y ~ x:g, data = mod5)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.97759 -0.32436 -0.00265  0.30116  1.04563 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2.89539    0.07001   41.36   <2e-16 ***
x:gA         0.18967    0.01509   12.57   <2e-16 ***
x:gB         0.65790    0.01185   55.54   <2e-16 ***
---
Signif. codes:  
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4182 on 197 degrees of freedom
Multiple R-squared:  0.9567,  Adjusted R-squared:  0.9563 
F-statistic:  2178 on 2 and 197 DF,  p-value: < 2.2e-16
```

Comment: The slope and full interaction models have relatively the highest R^2_{adj}-values among all models. All the variables are significant in the full interaction model, and its R^2_{adj} is slightly higher than that of the slope model, so we should choose the **full interaction** model for mod5.

Task 3.6: Identify the underlying model structure for **`mod6`**. [1 point]



Plain regression model

```
Call:
lm(formula = y ~ x, data = mod6)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.7846 -1.4616 -0.1156  1.4534  5.0689 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  3.49558    0.29654  11.788  < 2e-16 ***
x            0.29577    0.05225   5.661 5.24e-08 ***
---
Signif. codes:  
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.86 on 198 degrees of freedom
Multiple R-squared:  0.1393,  Adjusted R-squared:  0.1349 
F-statistic: 32.04 on 1 and 198 DF,  p-value: 5.243e-08
```

Means model

```
Call:
lm(formula = y ~ g, data = mod6)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.6971 -0.9775 -0.0633  1.1408  5.4910 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
```

Intercept model

```
Call:
lm(formula = y ~ x + g, data = mod6)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.0476 -1.2747 -0.0646  1.3977  4.8299 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)    4.4662      0.1683    26.54   < 2e-16 ***
gB             1.4235      0.2748     5.18 5.45e-07 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.882 on 198 degrees of freedom
Multiple R-squared:  0.1193,  Adjusted R-squared:
0.1149
F-statistic: 26.83 on 1 and 198 DF,  p-value: 5.451e-07
```

```
(Intercept)  3.54033    0.29035  12.193  < 2e-16 ***
x            0.21819    0.05667   3.850  0.00016 ***
gB           0.93294    0.29468   3.166  0.00179 **
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.819 on 197 degrees of freedom
Multiple R-squared:  0.181,   Adjusted R-squared:  0.1726
F-statistic: 21.76 on 2 and 197 DF,  p-value: 2.888e-09
```

Full interaction model

```
Call:
lm(formula = y ~ x * g, data = mod6)

Residuals:
     Min       1Q   Median       3Q      Max
-1.82531 -0.57645  0.04476  0.54958  2.13253

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.01986    0.14519   41.46   <2e-16 ***
x           -0.36614    0.03028  -12.09   <2e-16 ***
gB          -7.36695    0.29662  -24.84   <2e-16 ***
x:gB         1.48092    0.04820   30.73   <2e-16 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7562 on 196 degrees of
freedom
Multiple R-squared:  0.8592,  Adjusted R-squared:  0.857
F-statistic: 398.6 on 3 and 196 DF,  p-value: < 2.2e-16
```

Slope model

```
Call:
lm(formula = y ~ x:g, data = mod6)

Residuals:
    Min      1Q  Median      3Q     Max
-4.3291 -0.7890  0.0924  1.0581  3.7913

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.25467    0.25718  16.543  < 2e-16 ***
x:gA        -0.04043    0.05543  -0.729    0.467
x:gB         0.35022    0.04352   8.048 7.77e-14 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.536 on 197 degrees of freedom
Multiple R-squared:  0.416,   Adjusted R-squared:  0.4101
F-statistic: 70.16 on 2 and 197 DF,  p-value: < 2.2e-16
```

Comment: The R^2_{adj} of the full interaction model is much higher than the other models, so we should choose the full interaction model for mod5.