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

Handed out: Monday, March 1, 2021

Return date: Friday, March 12, 2021, at ELEARNING's **Lab02Submit** link.

Grading: This lab counts 13 % towards your final grade

Task 1. Partial Regression Coefficient [3 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 in a regression model.

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

```
model.1 <- lm(water81~income+water80+educat,data = concord)
summary(model.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*? [1 point]

[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**

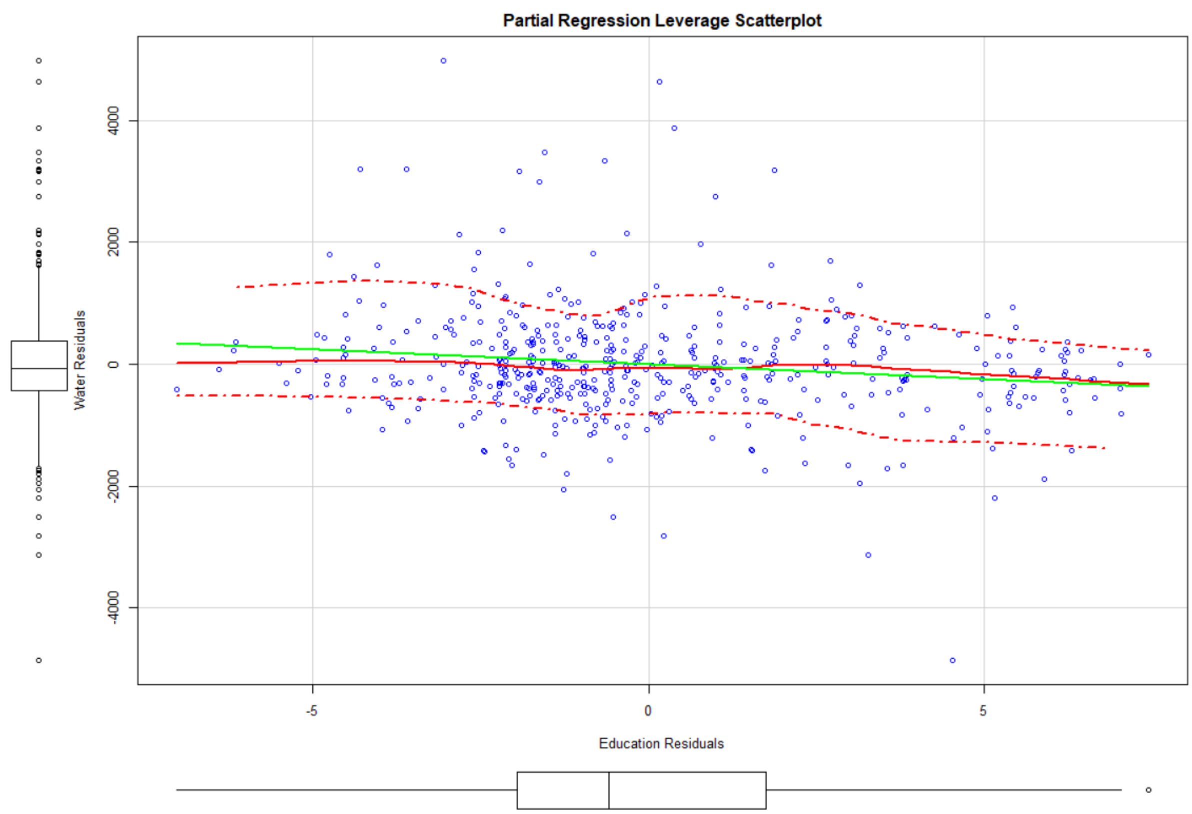
```
lm3 <- lm(educat~income+water80, data=concord)
summary(lm3)
Call:
lm(formula = educat ~ income + water80, data = concord)

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
```

<u>Comment:</u> 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",
pch=1, smooth=list(span = 0.35,lty.smooth=1, col.smooth="red",
col.var="red"),regLine=list(col="green"))
```



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*? [1 point]

```
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)  -4.989e+01  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 [7 points]

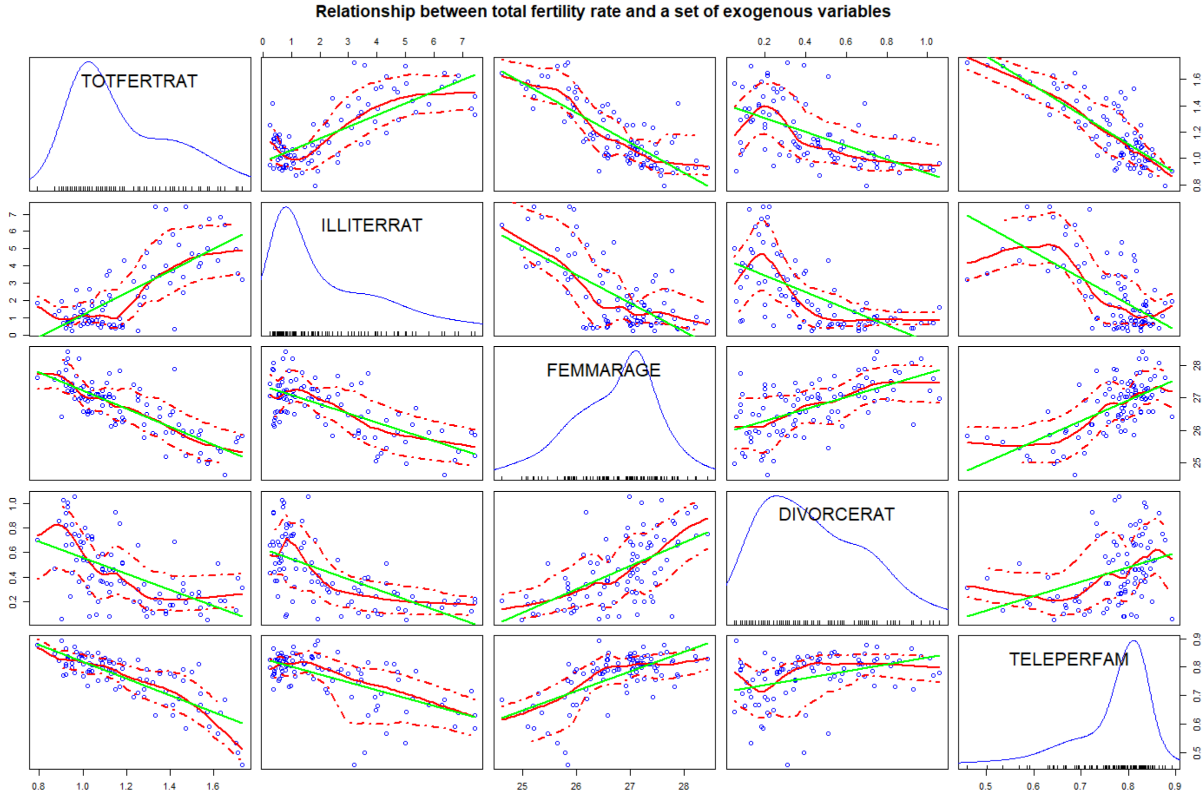
Task 2.1: Use common sense arguments ***how*** these four metric variables will influence the provincial fertility rates. Use one or two sentences per explanation, and formulate preferably *one* sided null and alternative hypotheses based on your explanation. The statistical hypotheses should be *type-set* properly, for instance, as $H_0: \beta_{VarXYZ} \leq 0$ against $H_1: \beta_{VarXYZ} > 0$. Format everything in the table shown below. [1 point]

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. Also generate a boxplot of the fertility rate against the regions. *Briefly interpret the scatterplot matrix.* [1 point]

```
provItaly <- foreign::read.dbf("provinces.dbf")
car::scatterplotMatrix(~TOTFERTRAT+ILLITERRAT+FEMMARAGE+DIVORCERAT+TE
LEPERFAM, 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"))
```



Comment:
[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. [1 point]

```
lm1<- lm(TOTFERTRAT~ ILLITERRAT+ FEMMARAGE+ DIVORCERAT+ TELEPERFAM,
data=provItaly)
summary(lm1)
Call:
lm(formula = TOTFERTRAT ~ ILLITERRAT + FEMMARAGE + DIVORCERAT +
    TELEPERFAM, data = provItaly)

Residuals:
```

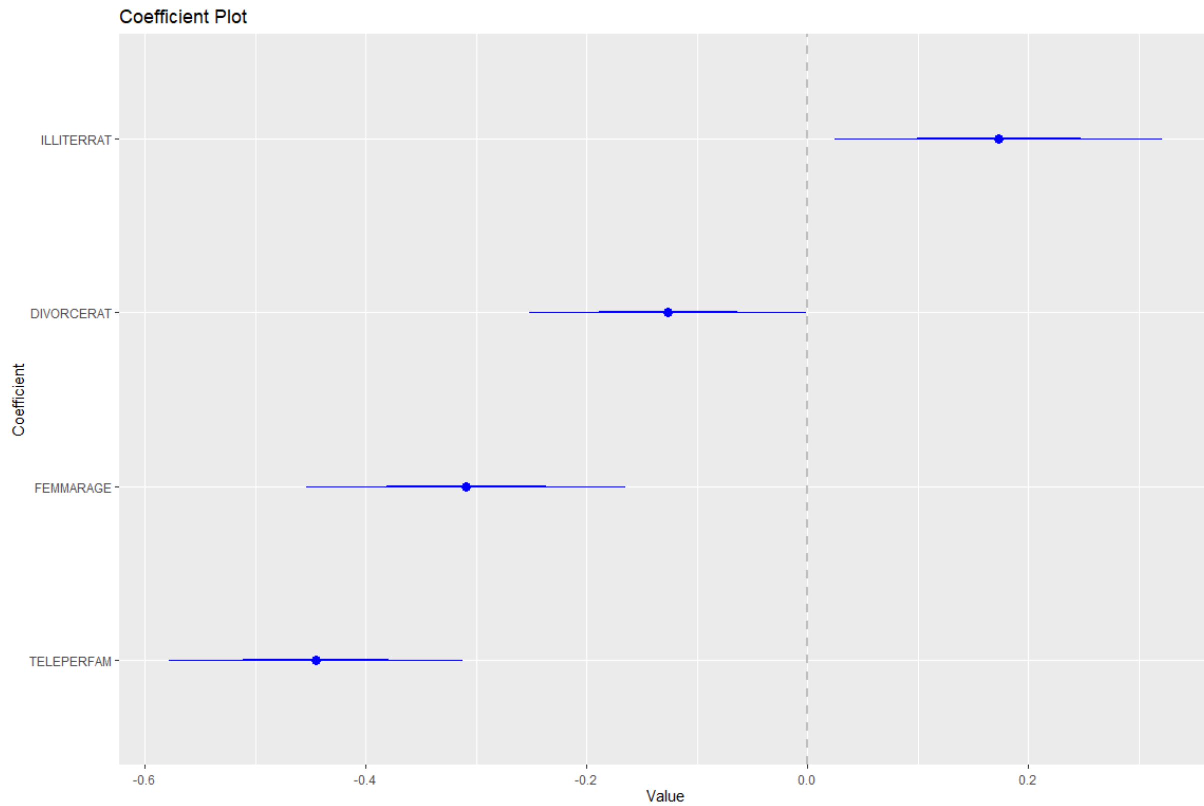
```
     Min       1Q   Median       3Q      Max
-0.21906 -0.06267 -0.00966  0.05425  0.41272

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)
```

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 as well as [b] the dependent fertility rate using the factor **REGION** as independent variable.
Does the **REGION** factor *explain variation* of the four independent variables as well as the fertility rate, i.e., what are the factor's R^2's? [1 point]

Hint: To calibrate all five models with one function call you can use the regression formula syntax **cbind(TOTFERTRAT,ILLITERRAT,FEMMARAGE,DIVORCERAT,TELEPERFAM)~REGION**.
The **summary** function gives you the results for all five models.

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

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

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 = provItaly)

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 = provItaly)

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 = provItaly)

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 = provItaly)

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? [1 point]

```
lm3 <-
lm(TOTFERTRAT~FEMMARAGE+DIVORCERAT+ILLITERRAT+TELEPERFAM+REGION,data=
provItaly)
summary(lm3)
Call:
lm(formula = TOTFERTRAT ~ FEMMARAGE + DIVORCERAT + ILLITERRAT +
    TELEPERFAM + REGION, data = provItaly)

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
```

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. [1 point]

That is, test the null hypothesis: $H_0: \beta_{Region\ 1} = \beta_{Region\ 2} = \cdots = \beta_{Region\ J} = 0$ against the alternative hypothesis is $H_0: \beta_{Region\ j} \neq 0$ *for at least one* $j \in \{1,2, \dots, J\}$.

```
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
```

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 [3 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:

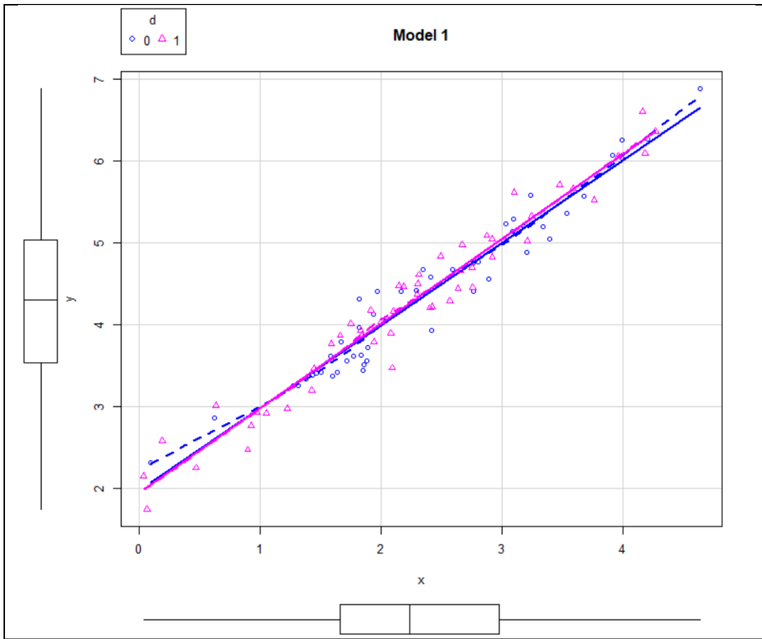
Name	Models Structure
Full interaction model	**`lm(y~g+x+g:x, data=mod?)`** ⇔ **`lm(y~g*x, data=mod?)`**
Intercept model	**`lm(y~g+x, data=mod?)`**
Slope model	**`lm(y~g:x, data=mod?)`**
Means model	**`lm(y~g, data=mod?)`**
Plain regression model	**`lm(y~x, data=mod?)`**

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. By visual inspection extend one of the regression lines to $x = 0$ to check if both lines share an identical intercept. If several competing model structures seem to be reasonably relevant, then try to eliminate inferior models by looking at their $R^2_{adjusted}$ or the t-values of the most elaborated model **`lm(y~g*x, data=mod?)`**.

Task 3.1: Identify the underlying model structure for **`mod1`**. [0.5 points]



Plain regression model

```
lm(formula = y ~ x, data = mod)
Residuals:
    Min      1Q  Median      3Q     Max
-3.0288 -0.9416  0.0601  0.8972  3.3632
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.82815    0.19726   9.268   <2e-16 ***
x            0.62359    0.03476  17.941   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1
Residual standard error: 1.237 on 198 degrees of freedom
Multiple R-squared:  0.6192,  Adjusted R-squared:  0.6172
F-statistic: 321.9 on 1 and 198 DF,  p-value: < 2.2e-16
```

Means model

```
lm(formula = y ~ g, data = mod)
Residuals:
    Min      1Q  Median      3Q     Max
-4.2071 -1.3745  0.0404  1.3072  5.1967
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.4710     0.1685  26.534  < 2e-16 ***
gB            1.4107     0.2752   5.127 6.99e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1
Residual standard error: 1.884 on 198 degrees of freedom
Multiple R-squared:  0.1172,  Adjusted R-squared:  0.1127
F-statistic: 26.28 on 1 and 198 DF,  p-value: 6.992e-07
```

Intercept model

```
lm(formula = y ~ g + x, data = mod)
Residuals:
    Min      1Q  Median      3Q     Max
-3.0335 -0.9385  0.0567  0.9010  3.3602

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.82866    0.19799   9.236   <2e-16 ***
gB           0.01061    0.20094   0.053    0.958
x            0.62270    0.03864  16.114   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1
Residual standard error: 1.24 on 197 degrees of freedom
Multiple R-squared:  0.6192,  Adjusted R-squared:  0.6153
F-statistic: 160.1 on 2 and 197 DF,  p-value: < 2.2e-16
```

Full interaction model

```
lm(formula = y ~ g * x, data = mod)
Residuals:
    Min      1Q  Median      3Q     Max
-2.9930 -0.9452  0.0734  0.9012  3.4968

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

Slope model

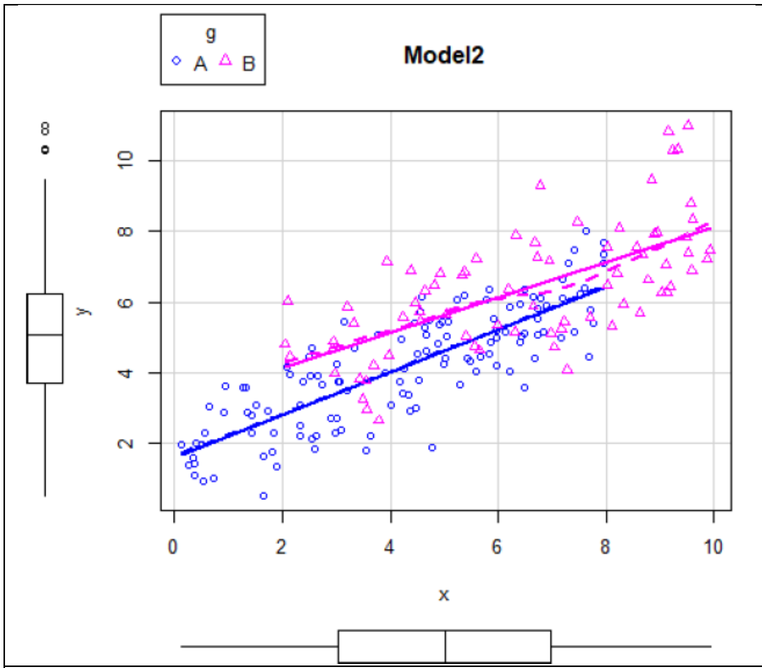
```
lm(formula = y ~ g:x, data = mod)
Residuals:
    Min      1Q  Median      3Q     Max
-2.9907 -0.9422  0.0874  0.8856  3.4047
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.80316    0.20760   8.686 1.44e-15 ***
```

```
(Intercept)  1.68697    0.23808   7.086 2.43e-11 ***
gB           0.48488    0.48637   0.997    0.320
x            0.65609    0.04965  13.215  < 2e-16 ***
gB:x        -0.08462    0.07904  -1.071    0.286
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 1.24 on 196 degrees of freedom
Multiple R-squared:  0.6214,  Adjusted R-squared:  0.6156
F-statistic: 107.2 on 3 and 196 DF,  p-value: < 2.2e-16
```

```
gA:x         0.63466    0.04475  14.183  < 2e-16 ***
gB:x         0.62179    0.03513  17.701  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 1.24 on 197 degrees of freedom
Multiple R-squared:  0.6195,  Adjusted R-squared:  0.6156
F-statistic: 160.3 on 2 and 197 DF,  p-value: < 2.2e-16
```

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**. [0.5 points]



Plain regression model

```
lm(formula = y ~ x, data = mod)
Residuals:
    Min      1Q  Median      3Q     Max
-2.8549 -0.8909 -0.1586  0.6597  3.2390
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.78315    0.19262   9.257   <2e-16 ***
x            0.63243    0.03394  18.634   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 1.208 on 198 degrees of freedom
Multiple R-squared:  0.6368,  Adjusted R-squared:  0.635
F-statistic: 347.2 on 1 and 198 DF,  p-value: < 2.2e-16
```

Means model

```
lm(formula = y ~ g, data = mod)
Residuals:
    Min      1Q  Median      3Q     Max
-3.4098 -1.1140  0.0327  1.0594  4.2119
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.9984     0.1366   29.28   <2e-16 ***
gB            2.6709     0.2230   11.98   <2e-16 ***
---
```

Intercept model

```
lm(formula = y ~ g + x, data = mod)
Residuals:
     Min       1Q   Median       3Q      Max
-2.45860 -0.76067  0.04597  0.73026  2.72341
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.85683    0.16047  11.571   <2e-16 ***
gB           1.53614    0.16286   9.432   <2e-16 ***
x            0.50470    0.03132  16.114   <2e-16 ***
```

<table>
<tr>
<td>Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.527 on 198 degrees of freedom
Multiple R-squared: 0.4201, Adjusted R-squared: 0.4172
F-statistic: 143.4 on 1 and 198 DF, p-value: < 2.2e-16</td>
<td>---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

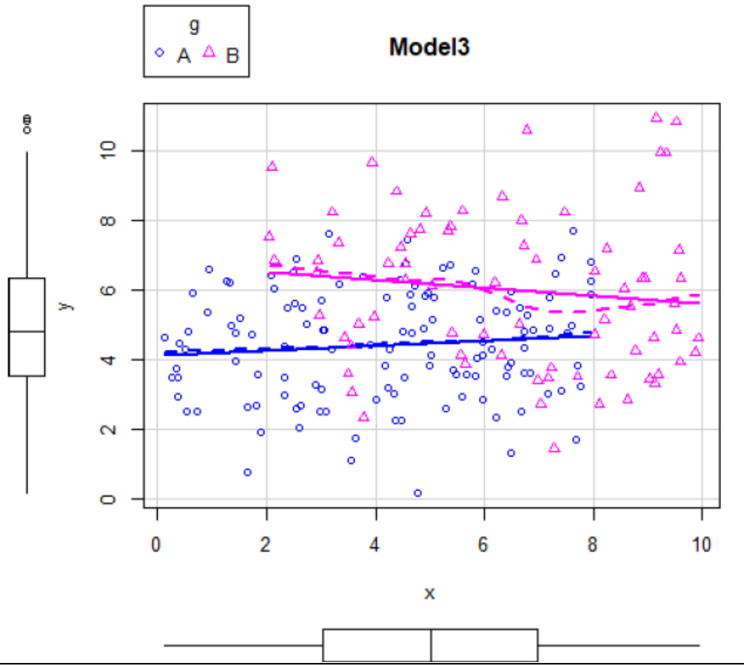
Residual standard error: 1.005 on 197 degrees of freedom
Multiple R-squared: 0.7498, Adjusted R-squared: 0.7473
F-statistic: 295.2 on 2 and 197 DF, p-value: < 2.2e-16</td>
</tr>
<tr>
<td>**Full interaction model**
lm(formula = y ~ g * x, data = mod)
Residuals:
Min 1Q Median 3Q Max
-2.42583 -0.76610 0.05948 0.73039 2.83412
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.74199 0.19296 9.028 < 2e-16 ***
gB 1.92054 0.39420 4.872 2.27e-06 ***
x 0.53176 0.04024 13.215 < 2e-16 ***
gB:x -0.06859 0.06406 -1.071 0.286

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 1.005 on 196 degrees of freedom
Multiple R-squared: 0.7513, Adjusted R-squared: 0.7475
F-statistic: 197.3 on 3 and 196 DF, p-value: < 2.2e-16</td>
<td>**Slope model**
lm(formula = y ~ g:x, data = mod)
Residuals:
Min 1Q Median 3Q Max
-2.4167 -0.7978 -0.0585 0.7776 2.6942
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.20217 0.17771 12.39 <2e-16 ***
gA:x 0.44685 0.03830 11.67 <2e-16 ***
gB:x 0.66249 0.03007 22.03 <2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 1.061 on 197 degrees of freedom
Multiple R-squared: 0.7212, Adjusted R-squared: 0.7183
F-statistic: 254.8 on 2 and 197 DF, p-value: < 2.2e-16</td>
</tr>
</table>

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`. [0.5 points]



Plain regression model

```
lm(formula = y ~ x, data = mod)
Residuals:
    Min      1Q  Median      3Q     Max
-4.8010 -1.5912 -0.1239  1.2518  5.4102
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.26213    0.31423   13.56  < 2e-16 ***
x            0.14507    0.05537    2.62  0.00947 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 1.971 on 198 degrees of freedom
Multiple R-squared:  0.03351, Adjusted R-squared:  0.02863
F-statistic: 6.864 on 1 and 198 DF,  p-value: 0.009475
```

Means model

```
lm(formula = y ~ g, data = mod)
Residuals:
    Min      1Q  Median      3Q     Max
-4.4144 -1.3540  0.0893  1.2991  4.8614
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.3163     0.1608  26.838  < 2e-16 ***
gB            1.8232     0.2626   6.942 5.41e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 1.798 on 198 degrees of freedom
Multiple R-squared:  0.1957,  Adjusted R-squared:  0.1917
F-statistic: 48.19 on 1 and 198 DF,  p-value: 5.407e-11
```

Intercept model

```
lm(formula = y ~ g + x, data = mod)
Residuals:
    Min      1Q  Median      3Q     Max
-4.4080 -1.3638  0.0824  1.3093  4.8828
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.350440   0.287703  15.121  < 2e-16 ***
gB           1.841244   0.291993   6.306 1.85e-09 ***
x           -0.008042   0.056154  -0.143    0.886
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 1.803 on 197 degrees of freedom
Multiple R-squared:  0.1958,  Adjusted R-squared:  0.1877
F-statistic: 23.99 on 2 and 197 DF,  p-value: 4.755e-10
```

Full interaction model

```
lm(formula = y ~ g * x, data = mod)
Residuals:
    Min      1Q  Median      3Q     Max
-4.3493 -1.3736  0.1066  1.3095  5.0813
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.14455    0.34596  11.980  < 2e-16 ***
gB           2.53043    0.70676   3.580 0.000433 ***
```

Slope model

```
lm(formula = y ~ g:x, data = mod)
Residuals:
    Min      1Q  Median      3Q     Max
-4.3372 -1.4391 -0.0118  1.3361  4.6391
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.75086    0.31060  15.296  < 2e-16 ***
gA:x        -0.07140    0.06695  -1.066 0.287507
```

<table>
<tr>
<td>

```
x            0.04048    0.07214   0.561 0.575380
gB:x        -0.12297    0.11485  -1.071 0.285622
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1
Residual standard error: 1.802 on 196 degrees of freedom
Multiple R-squared:  0.2005,  Adjusted R-squared:  0.1883
F-statistic: 16.38 on 3 and 196 DF,  p-value: 1.533e-09
```

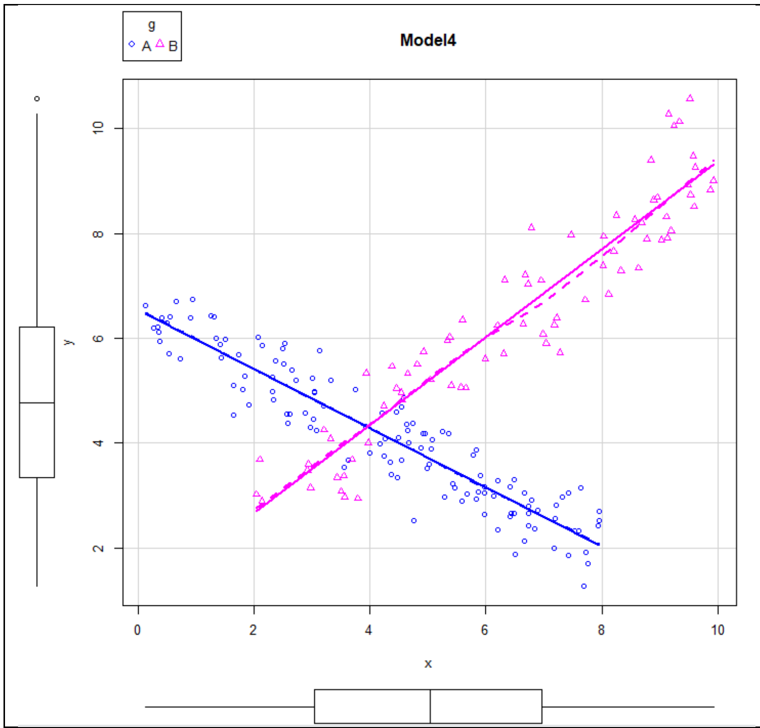
</td>
<td>

```
gB:x         0.18012    0.05255   3.427 0.000742 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1
Residual standard error: 1.855 on 197 degrees of freedom
Multiple R-squared:  0.1482,  Adjusted R-squared:  0.1396
F-statistic: 17.14 on 2 and 197 DF,  p-value: 1.373e-07
```

</td>
</tr>
</table>

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_{adj2}-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**. [0.5 points]



Plain regression model

```
lm(formula = y ~ x, data = mod)
Residuals:
    Min      1Q  Median      3Q     Max
-4.2000 -1.4723  0.0146  1.5496  4.7610
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.07158    0.31104   13.09  < 2e-16 ***
x            0.18253    0.05481    3.33  0.00103 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1
Residual standard error: 1.951 on 198 degrees of freedom
Multiple R-squared:  0.05305, Adjusted R-squared:
0.04826
F-statistic: 11.09 on 1 and 198 DF,  p-value: 0.001034
```

Means model

```
lm(formula = y ~ g, data = mod)
Residuals:
    Min      1Q  Median      3Q     Max
-3.5374 -1.2077 -0.0586  1.4302  4.1353
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.1388     0.1489  27.794   <2e-16 ***
gB            2.2964     0.2432   9.444   <2e-16 ***
```

Intercept model

```
lm(formula = y ~ g + x, data = mod)
Residuals:
    Min      1Q  Median      3Q     Max
-3.5825 -1.1893 -0.0542  1.4131  4.1667
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.18284    0.26636  15.703  < 2e-16 ***
gB           2.31972    0.27034   8.581  2.8e-15 ***
```

<table>
<tr>
<td>

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1
Residual standard error: 1.665 on 198 degrees of freedom
Multiple R-squared:  0.3105,  Adjusted R-squared:  0.3071
F-statistic: 89.18 on 1 and 198 DF,  p-value: < 2.2e-16
```

</td>
<td>

```
x            -0.01037    0.05199  -0.199    0.842
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1
Residual standard error: 1.669 on 197 degrees of freedom
Multiple R-squared:  0.3107,  Adjusted R-squared:  0.3037
F-statistic:  44.4 on 2 and 197 DF,  p-value: < 2.2e-16
```

</td>
</tr>
<tr>
<td>

Full interaction model

```
lm(formula = y ~ g * x, data = mod)
Residuals:
     Min       1Q   Median       3Q      Max
-1.37465 -0.43413  0.03371  0.41389  1.60602
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   6.5333     0.1094   59.75   <2e-16 ***
gB           -5.5481     0.2234  -24.84   <2e-16 ***
x            -0.5643     0.0228  -24.75   <2e-16 ***
gB:x          1.4038     0.0363   38.67   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1
Residual standard error: 0.5695 on 196 degrees of freedom
Multiple R-squared:  0.9201,  Adjusted R-squared:  0.9189
F-statistic: 752.7 on 3 and 196 DF,  p-value: < 2.2e-16
```

</td>
<td>

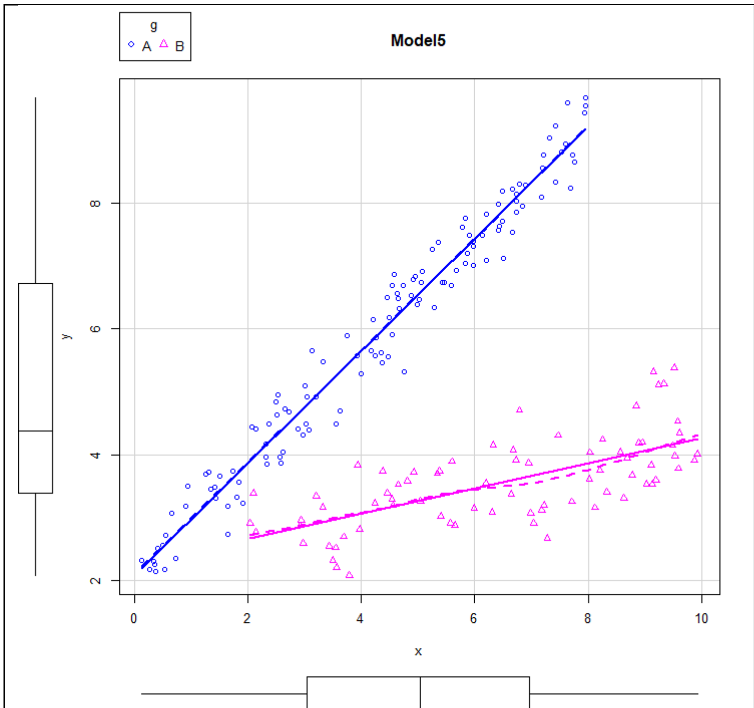
Slope model

```
lm(formula = y ~ g:x, data = mod)
Residuals:
    Min      1Q  Median      3Q     Max
-3.2603 -0.5942  0.0696  0.7969  2.8553
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.20392    0.19368  26.868  < 2e-16 ***
gA:x        -0.31899    0.04175  -7.641 9.20e-13 ***
gB:x         0.26375    0.03277   8.048 7.77e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1
Residual standard error: 1.157 on 197 degrees of freedom
Multiple R-squared:  0.6688,  Adjusted R-squared:  0.6654
F-statistic: 198.9 on 2 and 197 DF,  p-value: < 2.2e-16F-
statistic: 13.82 on 1 and 98 DF,  p-value: 0.0003357
```

</td>
</tr>
</table>

The R_{adj}^2 of the full interaction model is much higher than the other models, so we should choose the **full interaction model** for mod4.

Task 3.5: Identify the underlying model structure for **mod5**. [0.5 points]



Plain regression model

```
lm(formula = y ~ x, data = mod)
Residuals:
    Min      1Q  Median      3Q     Max
-3.0119 -1.4703 -0.2183  1.6869  3.7961
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.42187    0.29412  11.634  < 2e-16 ***
x            0.31026    0.05182   5.987  9.9e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 1.845 on 198 degrees of freedom
Multiple R-squared:  0.1533,  Adjusted R-squared:  0.149
F-statistic: 35.84 on 1 and 198 DF,  p-value: 9.904e-09F-statistic:  1853 on 1 and 98 DF,  p-value: < 2.2e-16
```

Means model

```
lm(formula = y ~ g, data = mod)
Residuals:
   Min     1Q Median     3Q    Max
-3.723 -0.950  0.028  1.012  3.797
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.8647     0.1486   39.46   <2e-16 ***
gB           -2.3058     0.2427   -9.50   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 1.662 on 198 degrees of freedom
Multiple R-squared:  0.3131,  Adjusted R-squared:  0.3096
F-statistic: 90.25 on 1 and 198 DF,  p-value: < 2.2e-16
```

Intercept model

```
lm(formula = y ~ g + x, data = mod)
Residuals:
    Min      1Q  Median      3Q     Max
-1.7420 -0.7779  0.0354  0.6191  2.5462
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.24470    0.14108   23.00   <2e-16 ***
gB          -3.69405    0.14319  -25.80   <2e-16 ***
x            0.61744    0.02754   22.42   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.884 on 197 degrees of freedom
Multiple R-squared:  0.8066,  Adjusted R-squared:  0.8047
F-statistic: 410.9 on 2 and 197 DF,  p-value: < 2.2e-16
```

Full interaction model

```
lm(formula = y ~ g * x, data = mod)
Residuals:
     Min       1Q   Median       3Q      Max
-1.04777 -0.33090  0.02569  0.31547  1.22412
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.09042    0.08334  25.082   <2e-16 ***
```

Slope model

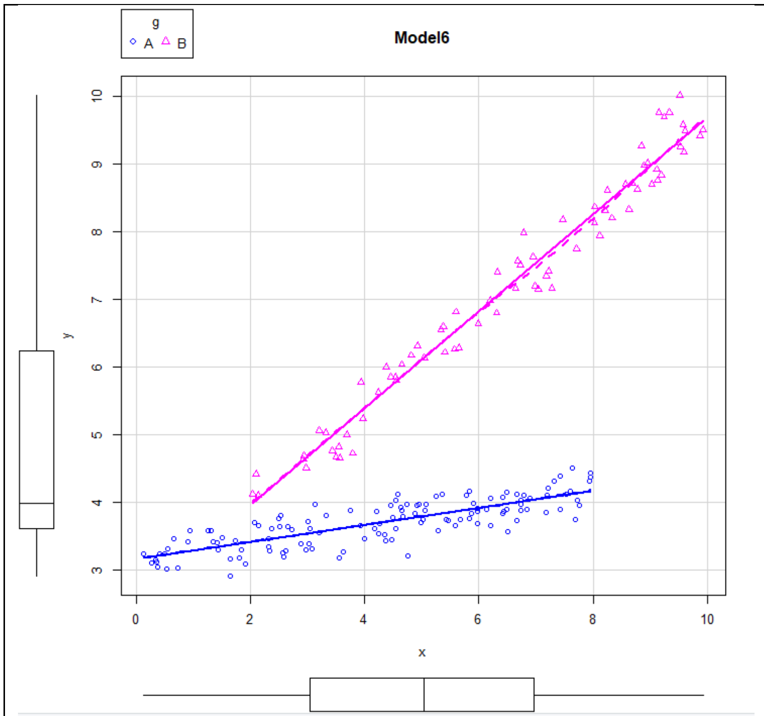
```
lm(formula = y ~ g:x, data = mod)
Residuals:
     Min       1Q   Median       3Q      Max
-1.04696 -0.32982  0.03061  0.31001  1.19186
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.13109    0.07268   29.32   <2e-16 ***
```

```
gB           0.16974    0.17026   0.997      0.32
x            0.88946    0.01738  51.178    <2e-16 ***
gB:x        -0.68941    0.02767 -24.917    <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1
Residual standard error: 0.4341 on 196 degrees of freedom
Multiple R-squared:  0.9536,  Adjusted R-squared:  0.9529
F-statistic:  1343 on 3 and 196 DF,  p-value: < 2.2e-16
```

```
gA:x         0.88196    0.01566   56.30   <2e-16 ***
gB:x         0.21767    0.01230   17.70   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1
Residual standard error: 0.4341 on 197 degrees of freedom
Multiple R-squared:  0.9534,  Adjusted R-squared:  0.9529
F-statistic:  2014 on 2 and 197 DF,  p-value: < 2.2e-16
```

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 mod5.

Task 3.6: Identify the underlying model structure for **mod6**. [0.5 points]



Plain regression model

```
lm(formula = y ~ x, data = mod)
Residuals:
   Min     1Q Median     3Q    Max
-2.766 -1.109  0.190  1.155  2.433
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.04497    0.21740   9.407   <2e-16 ***
x            0.58096    0.03831  15.166   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1
Residual standard error: 1.364 on 198 degrees of freedom
Multiple R-squared:  0.5374,  Adjusted R-squared:  0.5351
F-statistic:   230 on 1 and 198 DF,  p-value: < 2.2e-16
```

Means model

```
lm(formula = y ~ g, data = mod)
Residuals:
     Min       1Q   Median       3Q      Max
-3.06502 -0.41165  0.02258  0.39965  2.83671
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.69633    0.09629   38.39   <2e-16 ***
gB           3.47647    0.15724   22.11   <2e-16 ***
---
```

Intercept model

```
lm(formula = y ~ g + x, data = mod)
Residuals:
     Min       1Q   Median       3Q      Max
-1.50376 -0.49922 -0.02274  0.59316  1.74898
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.17300    0.11181   19.43   <2e-16 ***
gB           2.66932    0.11347   23.52   <2e-16 ***
x            0.35899    0.02182   16.45   <2e-16 ***
---
```

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1
‘ ’ 1
Residual standard error: 1.077 on 198 degrees of freedom
Multiple R-squared:  0.7117,  Adjusted R-squared:  0.7103
F-statistic: 488.8 on 1 and 198 DF,  p-value: < 2.2e-16
```

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1
‘ ’ 1
Residual standard error: 0.7005 on 197 degrees of freedom
Multiple R-squared:  0.8785,  Adjusted R-squared:  0.8773
F-statistic: 712.5 on 2 and 197 DF,  p-value: < 2.2e-16
```

Full interaction model

```
lm(formula = y ~ g * x, data = mod)
Residuals:
     Min       1Q   Median       3Q      Max
-0.57701 -0.18223  0.01415  0.17373  0.67413
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.159606   0.045898  68.840  < 2e-16 ***
gB          -0.633212   0.093766  -6.753  1.6e-10 ***
x            0.126485   0.009571  13.215  < 2e-16 ***
gB:x         0.589262   0.015237  38.673  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1
‘ ’ 1
Residual standard error: 0.2391 on 196 degrees of freedom
Multiple R-squared:  0.9859,  Adjusted R-squared:  0.9857
F-statistic:  4577 on 3 and 196 DF,  p-value: < 2.2e-16
```

Slope model

```
lm(formula = y ~ g:x, data = mod)
Residuals:
     Min       1Q   Median       3Q      Max
-0.74575 -0.16762 -0.01297  0.18221  0.81227
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.007883   0.044323   67.86   <2e-16 ***
gA:x        0.154481   0.009554   16.17   <2e-16 ***
gB:x        0.650031   0.007500   86.67   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1
‘ ’ 1
Residual standard error: 0.2647 on 197 degrees of freedom
Multiple R-squared:  0.9827,  Adjusted R-squared:  0.9825
F-statistic:  5580 on 2 and 197 DF,  p-value: < 2.2e-16
```

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 mod6.