Trabalho 1 - Econometria Aplicada 2023

Prof.: Adalto Acir Althaus Juniore

William Borges e Rafael Buttini

ATIVIDADE A

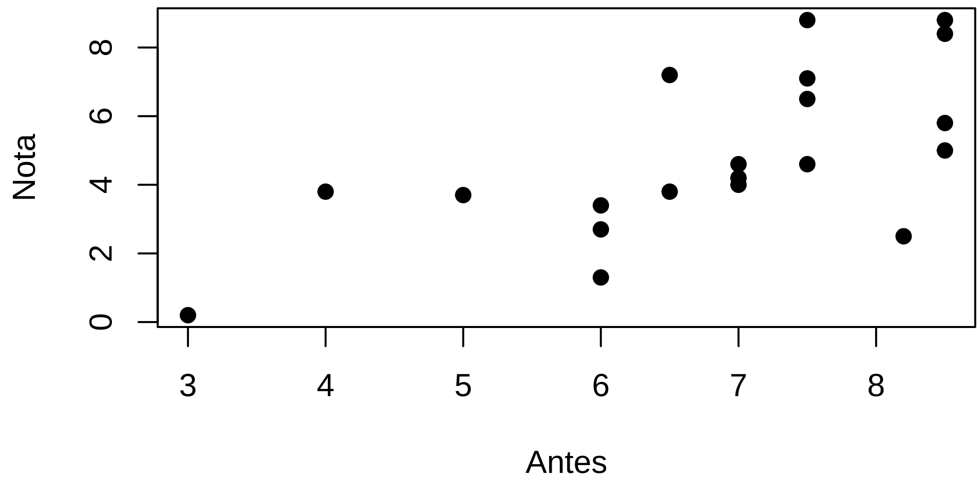
A planilha exemplo1.xls contém informações referentes às seguintes variáveis: - NOTA – nota obtida na P1 por cada aluno da turma A de TPE no semestre passado - ANTES – nota esperada por cada aluno antes de ver a prova - APOS – nota esperada por cada aluno após a realização da prova Utilize excel e responda

1 – Mostre em um diagrama de dispersão a relação entre NOTA e ANTES. Ao rodar uma regressão de NOTA em ANTES, que valores você esperaria para beta0 e beta1?

Resp: para beta0 eu espero um valor negativo, e para beta1 eu espepro um valor positivo.

```r
exemplo_1 <- read.csv2("exemplo_1.csv")

plot(exemplo_1$antes,exemplo_1$nota, pch=19, ylab="Nota",xlab="Antes")
```



2 – Realize a regressão citada no item anterior de 2 formas distintas:

- (i) "manualmente", isto é, calculando explicitamente os termos presentes na fórmula do estimador de MQO;

Resp:

```
vetor_betas = function(vetor_x,vetor_y){
  beta1=sum((vetor_x-mean(vetor_x))*(vetor_y-mean(vetor_y)))/
        sum((vetor_x-mean(vetor_x))^2)
  beta0=mean(vetor_y)-beta1*mean(vetor_x)
  return(data.frame("Parâmetros"=c("Beta 0","Beta 1"),
                    "Estimativas"=c(beta0,beta1)))
}

vetor_betas(exemplo_1$antes,exemplo_1$nota)
```

```
  Parâmetros Estimativas
1     Beta 0   -2.163015
2     Beta 1    1.025406
```

- (ii) usando o comando interceptação e inclinação em fórmulas estatísticas. Os valores estimados dos coeficientes deveriam, evidentemente, ser iguais para ambos os métodos. Tais coeficientes estão de acordo com o esperado no item 1?

Resp: sim, são os mesmos coeficientes, numéricamente falando.

```r
lm(exemplo_1$nota~exemplo_1$antes)|>summary()
```

```

Call:
lm(formula = exemplo_1$nota ~ exemplo_1$antes)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.7453 -0.9494 -0.6458  1.6411  3.2725 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)   
(Intercept)      -2.1630     1.9959  -1.084  0.29280   
exemplo_1$antes   1.0254     0.2865   3.579  0.00214 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.883 on 18 degrees of freedom
Multiple R-squared:  0.4158,	Adjusted R-squared:  0.3833 
F-statistic: 12.81 on 1 and 18 DF,  p-value: 0.002144
```

3 – Calcule os resíduos da regressão e verifique que sua média é zero (a menos de erros de arredondamento). Obtenha uma estimativa da variância (e, portanto, do desvio padrão) do erro aleatório U do modelo.

Resp:

```r
# Imagino que aqui devemos fazer "a mão" essa conta

preditor = function(vetor_x,vetor_y){
  betas= vetor_betas(exemplo_1$antes,exemplo_1$nota)$Estimativas
  vetores_x= data.frame(rep(1,times=length(vetor_x)),
                        vetor_x)
  return(apply(vetores_x,1,function(x){betas%*%x}))
}

y_hat=preditor(exemplo_1$antes,exemplo_1$nota)

residuos=exemplo_1$nota-y_hat
```

```r
print(paste("Média: ", round(mean(residuos),4),
            ", Variância: ",round(var(residuos),4),
            ", Desvio Padrão: ",round(var(residuos)^(1/2),4)))
```

```
[1] "Média:  0 , Variância:  3.3605 , Desvio Padrão:  1.8332"
```

3a – Refaça os itens 2 e 3 utilizando a função do excel: Dados -> Análise de dados -> Regressão

```r
residuos=lm(exemplo_1$nota~exemplo_1$antes)$residuals

print(paste("Média: ", round(mean(residuos),4),
            ", Variância: ", round(var(residuos),4),
            ", Desvio Padrão: ", round(var(residuos)^(1/2),4)))
```

```
[1] "Média:  0 , Variância:  3.3605 , Desvio Padrão:  1.8332"
```

4 – Calcule o coeficiente de correlação amostral entre NOTA e ANTES de 2 formas distintas:

- (i) "manualmente", isto é, aplicando explicitamente a fórmula adequada (note que a maior parte dos cálculos já foi feita no item 2.i acima);

```r
correlacao_pearson = function(vetor_x,vetor_y){
  cov_xy = cov(vetor_x,vetor_y)
  desvpad_x = var(vetor_x)^(1/2)
  desvpad_y = var(vetor_y)^(1/2)

  return(cov_xy/(desvpad_x*desvpad_y))
}

correlacao_pearson(exemplo_1$antes,exemplo_1$nota)
```

```
[1] 0.6448134
```

- (ii) usando a função estatística CORREL. Verifique que o R2 da regressão do item anterior corresponde ao quadrado desse coeficiente de correlação.

```r
cor(exemplo_1$antes,exemplo_1$nota)
```

```
[1] 0.6448134
```

```r
cor(exemplo_1$antes,exemplo_1$nota)^(2)
```

```
[1] 0.4157843
```

5 – Realize a regressão de NOTA (variável dependente) em ANTES (variável independente) supondo que o intercepto seja zero (ou seja, excluindo o termo constante do modelo). Calcule a soma dos resíduos da regressão e compare com o resultado obtido no item 3.

```r
com_intercepto=lm(exemplo_1$nota~exemplo_1$antes)

sem_intercepto=lm(exemplo_1$nota~exemplo_1$antes-1)

data.frame("Modelo"=c("Com intercepto","Sem intercecpto"),
           "Soma dos resíduos"=c(sum(com_intercepto$residuals)|>round(4),
                                 sum(sem_intercepto$residuals)|>round(4)),
           "Média dos resíduos"=c(mean(com_intercepto$residuals)|>round(4),
                                  mean(sem_intercepto$residuals)|>round(4)),
           "Desv.Pad. dos resíduos"=c(var(com_intercepto$residuals)^(1/2)|>round(4),
                                      var(sem_intercepto$residuals)^(1/2)|>round(4)))
```

```
           Modelo Soma.dos.resíduos Média.dos.resíduos Desv.Pad..dos.resíduos
1  Com intercepto             0.000             0.0000                 1.8332
2 Sem intercecpto            -1.926            -0.0963                 1.8895
```

6 – Um teste da hipótese de racionalidade das expectativas se basearia na hipótese nula H0:beta0=0 e beta1=1. Com base nos valores estimados, gostaríamos de testar tal hipótese. Veremos formalmente no curso como testar hipóteses conjuntas como essa. Informalmente, porém, já podemos dizer alguma coisa a respeito dessa hipótese? Ela parece razoável dados os betas e seus respectivos desvios padrões estimados nos modelos com e sem intercepto acima?

Resp: para o modelo com intercepto, as duas hipóteses parecem razoáveis a um nível de 95% de confiança. Repare que os valores da hipótese nula de cada coeficiente beta pertencecm aos

$IC_{95\%}$ de cada beta estimado. Já com o modelo sem o intercepto, a hipótese de que $\beta_1 = 1$ é rejeitada a um nível de 95% de confiança (i.e: o valor de 1 não pertence ao intervalo).

```
com_intercepto|>confint()
```

```
                    2.5 %   97.5 %
(Intercept)     -6.3563414 2.030311
exemplo_1$antes  0.4235091 1.627303
```

```
sem_intercepto|>confint()
```

```
                   2.5 %    97.5 %
exemplo_1$antes 0.5948217 0.8490252
```

7 – A nota esperada por cada aluno reflete diversos fatores, em particular: (i) grau de dificuldade esperado da prova; (ii) nível esperado de exigência na correção; (iii) nível de conhecimento da matéria percebido pelo aluno. Os desvios da nota efetiva em relação à esperada refletem, assim, erros referentes a cada uma dessas expectativas. Qual seria, então, a diferença entre o modelo estimado acima e um segundo modelo, no qual incluíssemos como regressor adicional a variável (APOS – ANTES)? Realize essa regressão (usando Análise de Dados) e compare com os resultados acima.

Resp: os resultados melhoraram considerávelemente! E desta vez, não rejeitamos a hipótese nula para o $\beta_0 = 0$.

```
apos_menos_antes=exemplo_1$apos-exemplo_1$antes
lm(exemplo_1$nota~exemplo_1$antes+apos_menos_antes)|>summary()
```

```

Call:
lm(formula = exemplo_1$nota ~ exemplo_1$antes + apos_menos_antes)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.00026 -1.10121 -0.05337  1.04021  2.74784 

Coefficients:
```

```
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)       -2.4127     1.4514  -1.662 0.114763    
exemplo_1$antes    1.3022     0.2186   5.956 1.56e-05 ***
apos_menos_antes   1.4405     0.3483   4.135 0.000692 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.368 on 17 degrees of freedom
Multiple R-squared:  0.7088,	Adjusted R-squared:  0.6745 
F-statistic: 20.69 on 2 and 17 DF,  p-value: 2.793e-05
```

8 – Realize agora a regressão de NOTA contra ANTES e APOS e compare com os resultados do item 7.

Resp: mesmíssimo coeficiente de determinação. E desta vez, a variável "antes" perdeu a sua representatividade no modelo e falhamos em rejeitar a hipótese nula de que $\beta_1 = 0$. No entanto, quando vemos a correlação entre o "antes" e o "após", vemos uma correlação de 80%. Logo, podemos supor que existe colinearidade em tal modelo, e por isso a variância dos estimadores também é sobrestimada. O melhor é utilizar o modelo do item 7.

```r
lm(nota~.,data=exemplo_1)|>summary()
```

```
Call:
lm(formula = nota ~ ., data = exemplo_1)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.00026 -1.10121 -0.05337  1.04021  2.74784 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -2.4127     1.4514  -1.662 0.114763    
antes        -0.1384     0.3500  -0.395 0.697519    
apos          1.4405     0.3483   4.135 0.000692 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.368 on 17 degrees of freedom
Multiple R-squared:  0.7088,	Adjusted R-squared:  0.6745 
F-statistic: 20.69 on 2 and 17 DF,  p-value: 2.793e-05
```

```r
# Ex 8
cor(exemplo_1$antes,exemplo_1$apos)
```

```
[1] 0.8039884
```

```r
# Ex 7
cor(exemplo_1$antes,apos_menos_antes)
```

```
[1] -0.3060926
```