

Parcial punto 3 y 5

Eddy Herrera Daza

25 de agosto de 2018

Punto 3

El rendimiento de una cosecha de cereal se considera bueno si es superior a 15 kg por área de cultivo y malo si no llega a dicha cantidad. Se hacen determinaciones en parcelas donde se ha sembrado cereales de tipo A y cereales tipo B y se clasifican como se muestra en la tabla ¿Son igualmente efectivos para el cultivo los cereales A y B?. justifique su respuesta

```
datos3<-matrix(c(40,98,50,75),nrow=2,byrow=T)
dimnames(datos3)<-list(c("M","B"), c("A","B"))
datos3
```

```
##      A  B
## M 40 98
## B 50 75
```

H_0 : Clase Independiente del tipo de cereal

H_0 : La clase No es independiente del cereal

La prueba que se aplicará es Chi.cuadrado, el resultado de la prueba con p-value = 0.08011 mayor que el valor de significancia α indica que No se rechaza la hipótesis nula, osea que no hay evidencias significativas que indiquen que hay dependencia entre los factores

```
prueba3=chisq.test(datos3)
prueba3
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  datos3
## X-squared = 3.0626, df = 1, p-value = 0.08011
```

Punto 5

Cargar el paquete MASS

```
library(MASS)
```

```
## Warning: package 'MASS' was built under R version 3.4.4
```

Cargar el data **swiss**.

**** Swiss Fertility and Socioeconomic Indicators (1888) Data****

Medición estandarizada de la fecundidad e indicadores socioeconómicos para cada una de las 47 provincias de Suiza alrededor de 1888.

```
swiss
```

##	Fertility	Agriculture	Examination	Education	Catholic
## Courtelary	80.2	17.0	15	12	9.96
## Delemont	83.1	45.1	6	9	84.84
## Franches-Mnt	92.5	39.7	5	5	93.40
## Moutier	85.8	36.5	12	7	33.77

## Neuveville	76.9	43.5	17	15	5.16
## Porrentruy	76.1	35.3	9	7	90.57
## Broye	83.8	70.2	16	7	92.85
## Glane	92.4	67.8	14	8	97.16
## Gruyere	82.4	53.3	12	7	97.67
## Sarine	82.9	45.2	16	13	91.38
## Veveyse	87.1	64.5	14	6	98.61
## Aigle	64.1	62.0	21	12	8.52
## Aubonne	66.9	67.5	14	7	2.27
## Avenches	68.9	60.7	19	12	4.43
## Cossonay	61.7	69.3	22	5	2.82
## Echallens	68.3	72.6	18	2	24.20
## Grandson	71.7	34.0	17	8	3.30
## Lausanne	55.7	19.4	26	28	12.11
## La Vallee	54.3	15.2	31	20	2.15
## Lavaux	65.1	73.0	19	9	2.84
## Morges	65.5	59.8	22	10	5.23
## Moudon	65.0	55.1	14	3	4.52
## Nyone	56.6	50.9	22	12	15.14
## Orbe	57.4	54.1	20	6	4.20
## Oron	72.5	71.2	12	1	2.40
## Payerne	74.2	58.1	14	8	5.23
## Paysd'enhaut	72.0	63.5	6	3	2.56
## Rolle	60.5	60.8	16	10	7.72
## Vevey	58.3	26.8	25	19	18.46
## Yverdon	65.4	49.5	15	8	6.10
## Conthey	75.5	85.9	3	2	99.71
## Entremont	69.3	84.9	7	6	99.68
## Herens	77.3	89.7	5	2	100.00
## Martigny	70.5	78.2	12	6	98.96
## Monthey	79.4	64.9	7	3	98.22
## St Maurice	65.0	75.9	9	9	99.06
## Sierre	92.2	84.6	3	3	99.46
## Sion	79.3	63.1	13	13	96.83
## Boudry	70.4	38.4	26	12	5.62
## La Chauxdfnd	65.7	7.7	29	11	13.79
## Le Locle	72.7	16.7	22	13	11.22
## Neuchatel	64.4	17.6	35	32	16.92
## Val de Ruz	77.6	37.6	15	7	4.97
## ValdeTravers	67.6	18.7	25	7	8.65
## V. De Geneve	35.0	1.2	37	53	42.34
## Rive Droite	44.7	46.6	16	29	50.43
## Rive Gauche	42.8	27.7	22	29	58.33
##	Infant.Mortality				
## Courtelary	22.2				
## Delemont	22.2				
## Franches-Mnt	20.2				
## Moutier	20.3				
## Neuveville	20.6				
## Porrentruy	26.6				
## Broye	23.6				
## Glane	24.9				
## Gruyere	21.0				
## Sarine	24.4				

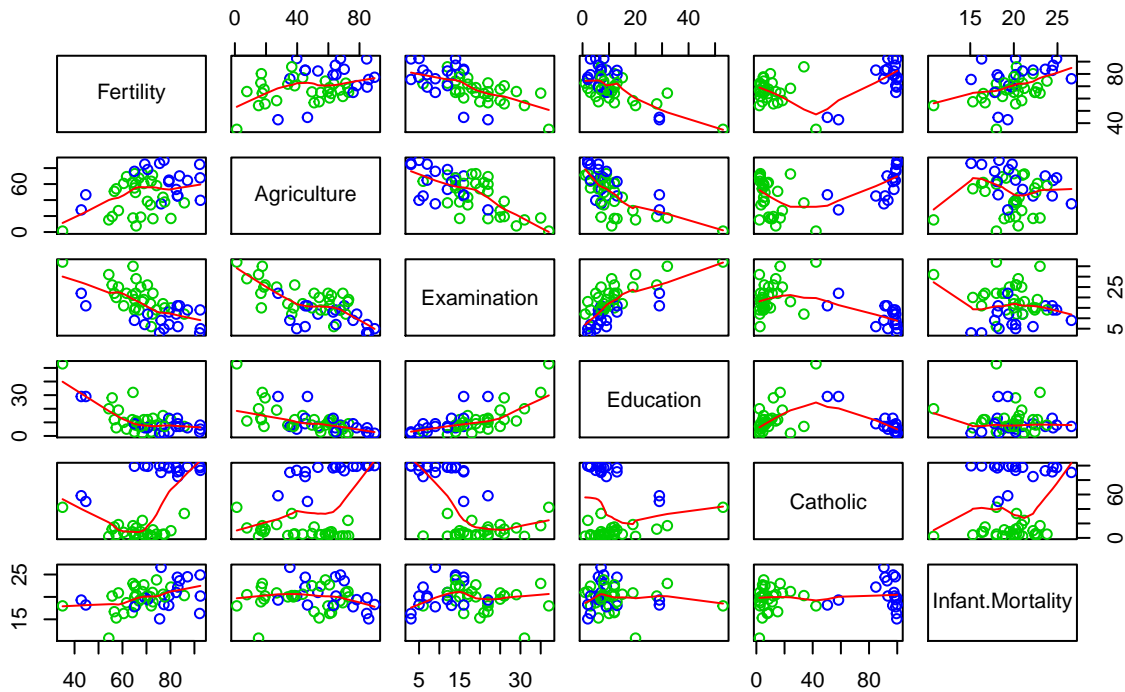
## Veveyse	24.5
## Aigle	16.5
## Aubonne	19.1
## Avenches	22.7
## Cossonay	18.7
## Echallens	21.2
## Grandson	20.0
## Lausanne	20.2
## La Vallee	10.8
## Lavaux	20.0
## Morges	18.0
## Moudon	22.4
## Nyone	16.7
## Orbe	15.3
## Oron	21.0
## Payerne	23.8
## Paysd'enhaut	18.0
## Rolle	16.3
## Vevey	20.9
## Yverdon	22.5
## Conthey	15.1
## Entremont	19.8
## Herens	18.3
## Martigwy	19.4
## Monthey	20.2
## St Maurice	17.8
## Sierre	16.3
## Sion	18.1
## Boudry	20.3
## La Chauxdfnd	20.5
## Le Locle	18.9
## Neuchatel	23.0
## Val de Ruz	20.0
## ValdeTravers	19.5
## V. De Geneve	18.0
## Rive Droite	18.2
## Rive Gauche	19.3

a. Calcule e interprete la matriz de correlación

Primero veamos si las variables podrían estar relacionadas linealmente con la variable respuesta: Fertilida

```
require(stats); require(graphics)
pairs(swiss, panel = panel.smooth, main = "swiss data",
      col = 3 + (swiss$Catholic > 50))
```

swiss data



Para el cálculo de la matriz, se hace necesario dos pasos:

- Esto es para que las magnitudes no afecten el resultado.
- Verificar si los datos se ajustan a una distribución normal. Esto es para determinar si el coeficiente a utilizar "Pearson" ó "Spearman"

Por ejemplo, la variable agricultura no se distribuye normal, ya que se rechaza la hipótesis nula

```
shapiro.test(swiss$Agriculture)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  swiss$Agriculture
## W = 0.96643, p-value = 0.193
```

i. Matriz de correlación (coeficiente de Spearman) con datos estandarizados

```
cor(scale(swiss), method = "spearman")
```

```
##          Fertility Agriculture Examination  Education
## Fertility    1.0000000    0.2426643 -0.66090300 -0.44325769
## Agriculture    0.2426643    1.0000000 -0.59885994 -0.65046381
## Examination  -0.6609030   -0.5988599  1.00000000  0.67460383
## Education    -0.4432577   -0.6504638  0.67460383  1.00000000
## Catholic      0.4136456    0.2886878 -0.47505753 -0.14441631
## Infant.Mortality 0.4371367  -0.1521287 -0.05915436 -0.01898137
##
##          Catholic Infant.Mortality
## Fertility    0.41364556    0.43713670
## Agriculture    0.28868781   -0.15212866
## Examination  -0.47505753   -0.05915436
```

```
## Education      -0.14441631    -0.01898137
## Catholic       1.00000000     0.06611714
## Infant.Mortality 0.06611714     1.00000000
```

**** Pruebas de Correlación ****

```
cor.test(swiss$Fertility, swiss$Agriculture, method = "spearman")
```

```
## Warning in cor.test.default(swiss$Fertility, swiss$Agriculture, method =
## "spearman"): Cannot compute exact p-value with ties

##
## Spearman's rank correlation rho
##
## data:  swiss$Fertility and swiss$Agriculture
## S = 13099, p-value = 0.1003
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.2426643
```

Luego, la correlación entre fertilidad y agricultura No es significativa

```
cor.test(swiss$Fertility, swiss$Education, method = "spearman")
```

```
## Warning in cor.test.default(swiss$Fertility, swiss$Education, method =
## "spearman"): Cannot compute exact p-value with ties

##
## Spearman's rank correlation rho
##
## data:  swiss$Fertility and swiss$Education
## S = 24963, p-value = 0.001806
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.4432577
```

Correlación significativa entre fertilidad y educación

```
cor.test(swiss$Fertility, swiss$Infant.Mortality, method = "spearman")
```

```
## Warning in cor.test.default(swiss$Fertility, swiss$Infant.Mortality, method
## = "spearman"): Cannot compute exact p-value with ties

##
## Spearman's rank correlation rho
##
## data:  swiss$Fertility and swiss$Infant.Mortality
## S = 9735.3, p-value = 0.002124
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.4371367
```

Correlación es significativa entre fertilidad y con mortalidad infantil

```
cor.test(swiss$Fertility, swiss$Catholic, method = "spearman")
```

```
## Warning in cor.test.default(swiss$Fertility, swiss$Catholic, method =
```

```
## "spearman"): Cannot compute exact p-value with ties
##
## Spearman's rank correlation rho
##
## data:  swiss$Fertility and swiss$Catholic
## S = 10142, p-value = 0.003851
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.4136456
```

Correlación significativa

```
cor.test(swiss$Fertility, swiss$Examination, method = "spearman")
```

```
## Warning in cor.test.default(swiss$Fertility, swiss$Examination, method =
## "spearman"): Cannot compute exact p-value with ties
##
## Spearman's rank correlation rho
##
## data:  swiss$Fertility and swiss$Examination
## S = 28727, p-value = 4.282e-07
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.660903
```

Correlación significativa

**** Modelos de Regresión Múltiple****

b. Corra el modelo de regresión y decida cuál es el mejor modelo porque

```
modelo5a=summary(lm(Fertility~Examination+Education,data=swiss))
modelo5a
```

```
##
## Call:
## lm(formula = Fertility ~ Examination + Education, data = swiss)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.9935  -6.8894  -0.3621   7.1640  19.2634
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  85.2533     3.0855  27.630  <2e-16 ***
## Examination  -0.5572     0.2319  -2.402   0.0206 *
## Education    -0.5395     0.1924  -2.803   0.0075 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.982 on 44 degrees of freedom
## Multiple R-squared:  0.5055, Adjusted R-squared:  0.483
## F-statistic: 22.49 on 2 and 44 DF,  p-value: 1.87e-07
```

Este modelo tiene un R^2 de 0.483 y las variables examination y education son significativas

```
modelo5b=summary(lm(Fertility ~ . , data = swiss))
modelo5b
```

```
##
## Call:
## lm(formula = Fertility ~ . , data = swiss)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.2743  -5.2617   0.5032   4.1198  15.3213
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    66.91518    10.70604     6.250 1.91e-07 ***
## Agriculture    -0.17211     0.07030    -2.448  0.01873 *
## Examination    -0.25801     0.25388    -1.016  0.31546
## Education      -0.87094     0.18303    -4.758 2.43e-05 ***
## Catholic        0.10412     0.03526     2.953  0.00519 **
## Infant.Mortality 1.07705     0.38172     2.822  0.00734 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.165 on 41 degrees of freedom
## Multiple R-squared:  0.7067, Adjusted R-squared:  0.671
## F-statistic: 19.76 on 5 and 41 DF,  p-value: 5.594e-10
```

En este modelo las variables son todas significativas, excepto la variable examination y hay un R^2 de 0.671

```
modelo5c=summary(lm(Fertility~Infant.Mortality+Education+Agriculture+Catholic ,data=swiss))
modelo5c
```

```
##
## Call:
## lm(formula = Fertility ~ Infant.Mortality + Education + Agriculture +
##      Catholic, data = swiss)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.6765  -6.0522   0.7514   3.1664  16.1422
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    62.10131     9.60489     6.466 8.49e-08 ***
## Infant.Mortality 1.07844     0.38187     2.824  0.00722 **
## Education      -0.98026     0.14814    -6.617 5.14e-08 ***
## Agriculture    -0.15462     0.06819    -2.267  0.02857 *
## Catholic        0.12467     0.02889     4.315 9.50e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.168 on 42 degrees of freedom
## Multiple R-squared:  0.6993, Adjusted R-squared:  0.6707
## F-statistic: 24.42 on 4 and 42 DF,  p-value: 1.717e-10
```

Como se puede ver el modelo las variables son todas significativas, hay un R^2 de 0.6707

Modelo Seleccionado

$$Y = 62.1 - 0.15X_1 - 0.98X_3 + 0.12X_4 + 1.08X_5$$

- c. Verifique que los residuos cumplen los supuestos de normalidad y de homocedasticidad generar los residuos del modelo seleccionado 5c y probar que se distribuyen normal y que las varianzas son constantes

```
shapiro.test(modelo5c$residuals)
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  modelo5c$residuals  
## W = 0.97657, p-value = 0.459
```

Como el valor de p es mayor que alfa entonces, los residuos se distribuyen normal

**** Prueba de Heterocedasticidad****

Esta prueba de varianzas constantes, para esto se carga la libreria de “car”

```
library(car)
```

```
## Warning: package 'car' was built under R version 3.4.4
```

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

```
ncvTest(lm(Fertility~Infant.Mortality+Education+Agriculture+Catholic ,data=swiss))
```

```
## Non-constant Variance Score Test  
## Variance formula: ~ fitted.values  
## Chisquare = 0.4687214, Df = 1, p = 0.49358
```

Por lo tanto, los residuales cumplen el principio de heterocedasticidad