

Untitled

Yasmin Al Omar y Jorge Cano

2024-10-26

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0    Min.   :  2.00
##  1st Qu.:12.0    1st Qu.: 26.00
##  Median :15.0    Median : 36.00
##  Mean   :15.4    Mean   : 42.98
##  3rd Qu.:19.0    3rd Qu.: 56.00
##  Max.   :25.0    Max.   :120.00
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

```
#Lectura de la base de datos
affairs <- read.csv("http://www-eio.upc.edu/~pau/cms/rdata/csv/COUNT/affairs.csv")
head(affairs)
```

```
##      X naffairs kids vryunhap unhap avgmarr hapavg vryhap antirel notrel slghtrel
## 1 1      0      0      0      0      0      1      0      0      0      1
## 2 2      0      0      0      0      0      1      0      0      0      0
## 3 3      3      0      0      0      0      1      0      0      0      1
## 4 4      0      1      0      0      0      1      0      1      0      0
## 5 5      3      1      0      0      0      0      1      0      0      1
## 6 6      0      1      0      0      0      0      1      0      0      0
##      smerel vryrel yrsmarr1 yrsmarr2 yrsmarr3 yrsmarr4 yrsmarr5 yrsmarr6
## 1      0      0      0      0      0      0      1      0
## 2      1      0      0      0      1      0      0      0
## 3      0      0      0      1      0      0      0      0
## 4      0      0      0      0      0      0      0      1
## 5      0      0      0      0      1      0      0      0
## 6      0      1      0      0      0      0      0      1
```

```
#2.Realizad un modelo de conteos de infidelidades. ¿Son las mismas variables las que afectan respecto a
formula <- as.formula('naffairs ~ kids + vryunhap + unhap + avgmarr + hapavg + vryhap + antirel + notre)
modelo_lineal <- glm(formula = formula, data = affairs, family = gaussian)
summary(modelo_lineal)
```

```
##
## Call:
## glm(formula = formula, family = gaussian, data = affairs)
##
## Coefficients: (3 not defined because of singularities)
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.12821    0.55158   2.045 0.041261 *
## kids        -0.37361    0.36729  -1.017 0.309471
## vryunhap     2.53143    0.81219   3.117 0.001918 **
## unhap       2.85816    0.44641   6.403 3.13e-10 ***
## avgmarr      0.51183    0.38272   1.337 0.181631
## hapavg       0.31315    0.30814   1.016 0.309923
## vryhap       NA         NA       NA      NA
## antirel      2.10421    0.58630   3.589 0.000360 ***
## notrel       0.91703    0.45075   2.034 0.042359 *
## slghtrel     1.21060    0.46381   2.610 0.009282 **
## smerel       0.04198    0.43389   0.097 0.922964
## vryrel       NA         NA       NA      NA
## yrsmarr1    -1.51762    0.56627  -2.680 0.007568 **
## yrsmarr2    -1.77008    0.48609  -3.641 0.000295 ***
## yrsmarr3    -1.07193    0.39872  -2.688 0.007384 **
## yrsmarr4    -0.62657    0.40905  -1.532 0.126118
## yrsmarr5     0.04547    0.43262   0.105 0.916333
## yrsmarr6     NA         NA       NA      NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 9.431891)
##
##      Null deviance: 6529.1  on 600  degrees of freedom
## Residual deviance: 5527.1  on 586  degrees of freedom
## AIC: 3071.1
##
## Number of Fisher Scoring iterations: 2

formula <- as.formula('naffairs ~ kids + vryunhap + unhap + avgmarr + hapavg + vryhap + antirel + notrel')
modelo_lineal1 <- glm(formula = formula, data = affairs, family = gaussian)
summary(modelo_lineal1)

##
## Call:
## glm(formula = formula, family = gaussian, data = affairs)
##
## Coefficients: (2 not defined because of singularities)
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.01523    0.53226   1.907 0.056955 .
## kids        -0.33313    0.36617  -0.910 0.363328
## vryunhap     2.58118    0.80540   3.205 0.001425 **
## unhap       2.82231    0.44414   6.355 4.19e-10 ***
## avgmarr      0.52106    0.38255   1.362 0.173695
## hapavg       0.33406    0.30796   1.085 0.278470
## vryhap       NA         NA       NA      NA
## antirel      2.03338    0.58464   3.478 0.000543 ***
## notrel       0.83250    0.44586   1.867 0.062374 .
```

```
## slghtrel      1.14598      0.46179      2.482 0.013357 *
## smerel       -0.03705      0.43112     -0.086 0.931541
## vryrel        NA          NA          NA      NA
## yrsmarr1     -1.35159      0.53810     -2.512 0.012279 *
## yrsmarr2     -1.59775      0.45208     -3.534 0.000441 ***
## yrsmarr3     -0.92177      0.36483     -2.527 0.011779 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 9.442479)
##
## Null deviance: 6529.1 on 600 degrees of freedom
## Residual deviance: 5552.2 on 588 degrees of freedom
## AIC: 3069.8
##
## Number of Fisher Scoring iterations: 2
```

```
#Calculad la frecuencia de infidelidades de una persona con más de 10 años de matrimonio, no-religioso,
valores_10matr_norel_sinhijos <- data.frame(kids = 0, vryunhap = 0, unhap = 0, avgmarr = 0, hapavg = 0,
prediccion_frecuencia_affairs <- predict(modelo_lineal1, newdata = valores_10matr_norel_sinhijos , type = "response")
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```
prediccion_frecuencia_affairs
```

```
##      1
## 3.881108
## attr(,"non-estim")
## 1
## 1
```

```
mean'affairs$naffairs)
```

```
## [1] 1.455907
```

```
#¿Podríaais calcular el 95% de confianza de la estimación anterior?
```

```
prediccion_frecuencia_affairs <- predict(modelo_lineal1, newdata = valores_10matr_norel_sinhijos , se.fit = TRUE)
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = residual.scale, type =
## if (type == : prediction from rank-deficient fit; attr(*, "non-estim") has
## doubtful cases
```

```
log_fit <- prediccion_frecuencia_affairs$fit
log_se <- prediccion_frecuencia_affairs$se.fit
limite_inferior <- exp(log_fit - 1.96 * log_se)
limite_superior <- exp(log_fit + 1.96 * log_se)
print(limite_inferior)
```

```
##      1
```

```
## 11.11089
## attr(,"non-estim")
## 1
## 1
```

```
print(limite_superior)
```

```
##      1
## 211.514
## attr(,"non-estim")
## 1
## 1
```

6. Calculad, bajo qué nivel de confianza, los residuos de la distribución pueden considerarse normales.

Primero calculamos los residuos y los valores ajustados

Mediante el test de shapiro vemos si rechazamos los residuos al 95% de confianza

```
##
## Shapiro-Wilk normality test
##
## data: residuals
## W = 0.77093, p-value < 2.2e-16
```

```
## Se rechaza la normalidad de los residuos al 95% de confianza.
```

y finalmente lo contrastamos con la representación gráfica

Residuos vs. Valores ajustados

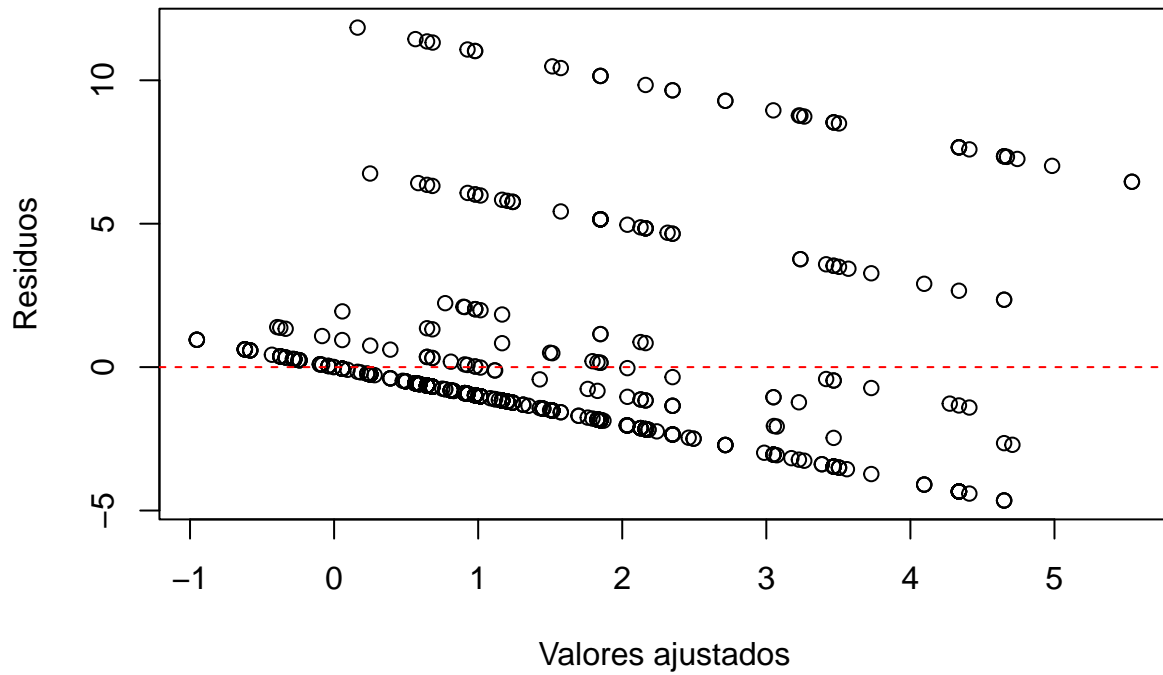
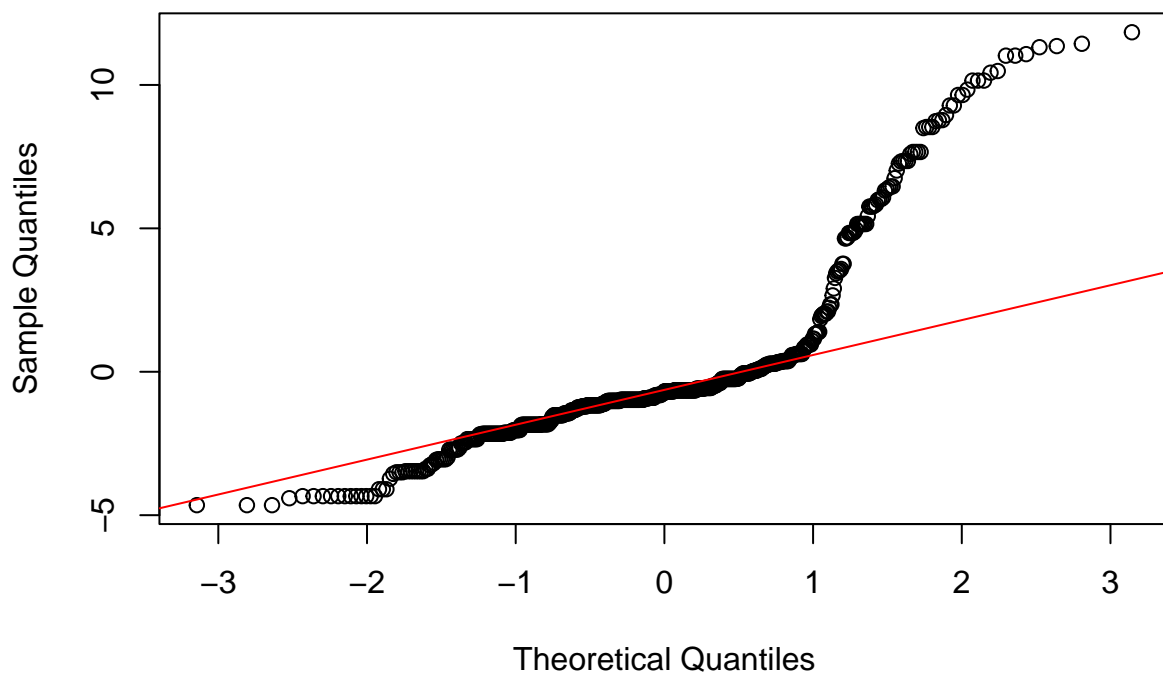


Gráfico Q-Q de los residuos



En el gráfico de Q-Q (quantile-quantile) de los residuos, podemos observar cómo se comparan los cuantiles de los residuos con los cuantiles de una distribución normal teórica. Visualmente podemos contrastar que los extremos presentan una fuerte desviación general del ajuste normal.

7. Calculad si la combinación de Años de Matrimonio e Hijos da nueva información a nuestro modelo.

Para calcular el efecto combinado de tener hijos con cada rango de años de matrimonio aplicaremos un modelo que itere la variable Kids con el resto de variable de años de matrimonio. Hay riesgo de multicolinealidad? lo veremos con los datos...

```
##              df      AIC
## modelo_lineal1    14 3069.798
## modelo_interaccion 21 3070.375
```

Podemos observar un modelo de AIC muy elevado, como consecuencia de la multicolinealidad (Si hay alta correlación entre los predictores, el modelo podría estar estimando coeficientes con poca precisión)

8. Teniendo la combinación de Años de Matrimonio e Hijos metido en el modelo, ¿cuál sería el cambio en infidelidades de no tener hijos a tener hijos?

```
##
## Call:
## glm(formula = formula_interaccion, family = gaussian, data = affairs)
##
## Coefficients: (4 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.00182    1.18052   2.543 0.011256 *
## kids          -2.21828    1.12152  -1.978 0.048409 *
## yrs marr1      -3.74182    1.20532  -3.104 0.001999 **
## yrs marr2      -3.76417    1.17682  -3.199 0.001456 **
## yrs marr3      -2.57906    1.21633  -2.120 0.034398 *
## yrs marr4      -1.54393    1.44055  -1.072 0.284272
## yrs marr5      -1.69467    1.60476  -1.056 0.291394
## yrs marr6           NA           NA      NA      NA
## vry unhap       2.35202    0.81602   2.882 0.004094 **
## unhap          2.86696    0.44620   6.425 2.74e-10 ***
## avg marr       0.50642    0.38184   1.326 0.185269
## hap avg        0.30789    0.30703   1.003 0.316381
## vry hap         NA           NA      NA      NA
## antirel        2.03111    0.58475   3.473 0.000552 ***
```

```

## notrel          0.83808    0.45053    1.860 0.063360 .
## slghtrel        1.09008    0.46354    2.352 0.019024 *
## smerel          -0.08252    0.43422   -0.190 0.849353
## vryrel          NA         NA         NA         NA
## kids:yrsmarr1   4.38255    1.58770    2.760 0.005957 **
## kids:yrsmarr2   2.69354    1.35574    1.987 0.047416 *
## kids:yrsmarr3   1.44436    1.28433    1.125 0.261223
## kids:yrsmarr4   0.86310    1.50058    0.575 0.565393
## kids:yrsmarr5   1.80032    1.66026    1.084 0.278656
## kids:yrsmarr6   NA         NA         NA         NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 9.345175)
##
##    Null deviance: 6529.1  on 600  degrees of freedom
## Residual deviance: 5429.5  on 581  degrees of freedom
## AIC: 3070.4
##
## Number of Fisher Scoring iterations: 2

## kids:yrsmarr1 kids:yrsmarr2 kids:yrsmarr3 kids:yrsmarr4 kids:yrsmarr5
##    2.1642696    0.4752590   -0.7739238   -1.3551830   -0.4179671
## kids:yrsmarr6
##          NA

## [1] -2.218284

```


El coeficiente de kids indica que, en promedio, tener hijos reduce las infidelidades en 2.218284.

Sin embargo, los términos de interacción (kids:yrs marrX) muestran que este efecto varía según la duración del matrimonio:

Para relaciones más cortas (menos de 1.5 años), el efecto de tener hijos es menor, y en algunos casos casi no afecta las infidelidades.

A medida que la duración del matrimonio aumenta, el efecto de tener hijos sobre las infidelidades se vuelve más fuerte, especialmente para parejas con más de 4 o 7 años de matrimonio.

9. Calculad una variable que convierta las dummies de años de matrimonio en numérica. Calculad también esta misma variable al cuadrado. ¿En teoría hay alguna edad de matrimonio en la que cada año adicional suponga un descenso de las infidelidades?

En primer lugar calculamos una variable numerica para los años de matrimonio y otra que suponga el cuadrado de esta misma.

Generamos una nueva regresión lineal incluyendo `naffairs ~ kids + years_married + years_married_squared`. Vemos que la variable es significativa en el modelo con un p-valor inferior a 0.05.

```
##
## Call:
## lm(formula = formula_2, data = affairs)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.3098 -1.8995 -0.9993 -0.4089 11.7929
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.211186   0.359611   0.587   0.557
## kids          -0.217247   0.381622  -0.569   0.569
## years_married   0.291844   0.129918   2.246   0.025 *
## years_married_squared -0.010129   0.007135  -1.420   0.156
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.244 on 597 degrees of freedom
## Multiple R-squared:  0.03799,    Adjusted R-squared:  0.03316
## F-statistic: 7.859 on 3 and 597 DF,  p-value: 3.773e-05
```

Por último calculamos un punto de inflexión que nos informa que el valor en los años de matrimonio en el cual el efecto de los años adicionales sobre las infidelidades cambia de dirección. En este caso, alrededor de los 14 años de matrimonio, el efecto de los años adicionales pasa de un posible aumento en infidelidades a una posible reducción.

```
coef_modelo <- coef(modelo_continuo)
punto_inflexion <- -coef_modelo["years_married"] / (2 * coef_modelo["years_married_squared"])
punto_inflexion
```

```
## years_married
##      14.40641
```