Modelos de Regresión II

Modelos Probabilísticos y Análisis Estadístico

Carlos Ricardo Bojacá

Departamento de Ciencias Básicas y Modelado
Facultad de Ciencias Naturales e Ingeniería
Universidad Jorge Tadeo Lozano

El modelo de regresión lineal compuesta

La variable de respuesta y puede verse influenciada por más de una variable independiente

Por ejemplo, el rendimiento de un cultivo puede depender de la cantidad de fertilizantes a base de N, P y K que se le hayan aplicado.

Un modelo lineal que relaciona la variable de respuesta y con varias variables predictoras tiene la forma:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots \beta_k x_k + \epsilon$$

El modelo de regresión lineal compuesta

1. Los parámetros $\beta_0, \beta_1, ..., \beta_k$ representan los parámetros o coeficientes de regresión del modelo

2. La variable aleatoria ϵ representa la variación aleatoria en y que no es explicada por el conjunto de variables explicatorias x

El modelo para la i-ésima observación estará dado por:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... \beta_k x_{ik} + \epsilon, i = 1, 2, ..., n$$

Cuadrados mínimos

Modelo lineal compuesto

En notación matricial, la matriz X es una matriz $n \times k$ en la que cada columna contiene n observaciones de la k-ésima variable independiente X_k.

Los parámetros mediante el método de cuadrados mínimos pueden ser estimados de acuerdo con la siguiente fórmula:

$$\hat{\beta} = (X'X)^{-1}X'y$$

Cuadrados mínimos

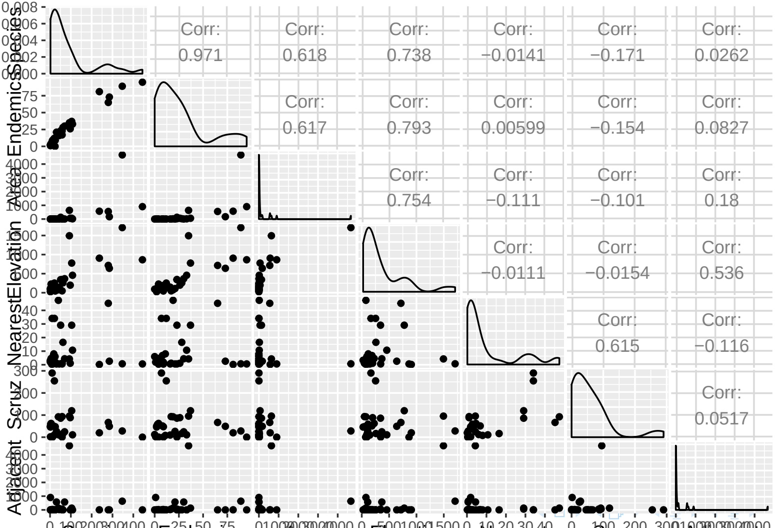
Ejemplo - Implementación

Datos *gala*: reporte de número de especies de tortugas en 30 de las Islas Galápagos

- Species: número de especies de tortugas encontradas en la isla
- Endemics: número de especies endémicas
- Area: área de la isla (km2)
- Elevation: altura máxima de la isla (m)
- Nearest: distancia a la isla más cercana (km)
- Scruz: distancia a la isla de Santa Cruz (km)
- Adjacent: área de la isla adyacente (km2)

Cuadrados mínimos

Ejemplo - Implementación



Modelo lineal compuesto

Ejemplo - Implementación

```
modcomp <- lm(Species ~ Area + Elevation + Nearest + Scruz + Adjacent,
              data = gala)
summary(modcomp)
```

```
##
## Call:
## lm(formula = Species ~ Area + Elevation + Nearest + Scruz + Adjacent,
##     data = gala)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -111.679  -34.898   -7.862   33.460  182.584
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.068221  19.154198   0.369 0.715351
## Area        -0.023938   0.022422  -1.068 0.296318
## Elevation    0.319465   0.053663   5.953 3.82e-06 ***
## Nearest      0.009144   1.054136   0.009 0.993151
## Scruz       -0.240524   0.215402  -1.117 0.275208
## Adjacent    -0.074805   0.017700  -4.226 0.000297 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 60.98 on 24 degrees of freedom
## Multiple R-squared:  0.7658,Adjusted R-squared:  0.7171
## F-statistic:  15.7 on 5 and 24 DF,  p-value: 6.838e-07
```

Modelo lineal compuesto

Estimación de Parámetros

$$\hat{\beta} = (X'X)^{-1}X'y$$

```
x <- model.matrix(~ Area + Elevation + Nearest + Scruz + Adjacent,
                  data = gala)
y <- gala$Species
solve(crossprod(x), crossprod(x, y))

##                      [,1]
## (Intercept)  7.068220709
## Area        -0.023938338
## Elevation    0.319464761
## Nearest      0.009143961
## Scruz       -0.240524230
## Adjacent    -0.074804832
```

Modelo lineal compuesto

Estimación de Parámetros

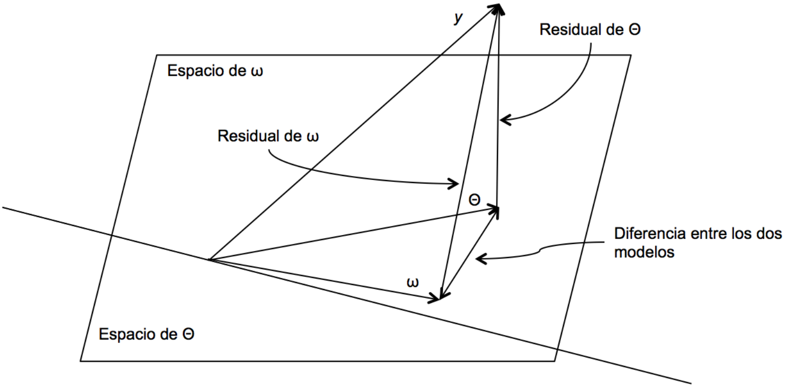
```
##
## Call:
## lm(formula = Species ~ Area + Elevation + Nearest + Scruz + Adjacent,
##     data = gala)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -111.679  -34.898   -7.862   33.460  182.584
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.068221  19.154198   0.369 0.715351
## Area        -0.023938   0.022422  -1.068 0.296318
## Elevation    0.319465   0.053663   5.953 3.82e-06 ***
## Nearest      0.009144   1.054136   0.009 0.993151
## Scruz       -0.240524   0.215402  -1.117 0.275208
## Adjacent    -0.074805   0.017700  -4.226 0.000297 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 60.98 on 24 degrees of freedom
## Multiple R-squared:  0.7658,Adjusted R-squared:  0.7171
## F-statistic:  15.7 on 5 and 24 DF,  p-value: 6.838e-07
```

Pruebas de hipótesis para comparar modelos

Dado un grupo de variables independientes, es consecuente definir si todas las variables del grupo son necesarias para explicar la respuesta de la variable dependiente

Considere un modelo grande, θ, y uno más pequeño, ω, que consiste en un subgrupo de predictores contenidos en θ. Dependiendo del grado de ajuste entre los dos modelos, qué modelo preferiría?

Pruebas de hipótesis para comparar modelos



Pruebas de hipótesis para comparar modelos

Potencialmente, un buen estadístico para determinar cuál de los dos modelos es mejor sería:

$$\frac{RSS_\omega - RSS_\theta}{RSS_\theta}$$

Suponga que la dimensión (el número de parámetros) de θ es p y la dimensión de ω es q, entonces el estadístico teórico de la distribución sería:

$$F = \frac{(RSS_\omega - RSS_\theta)/(p-q)}{RSS_\theta/(n-p)} \sim F_{p-q,n-p}$$

En consecuencia se rechazaría la hipótesis nula si:

$$F > F^{(\alpha)}_{p-q,n-p}$$

Distribución F

Distribución utilizada para estudiar las varianzas dentro de una población

Propiedades básicas:

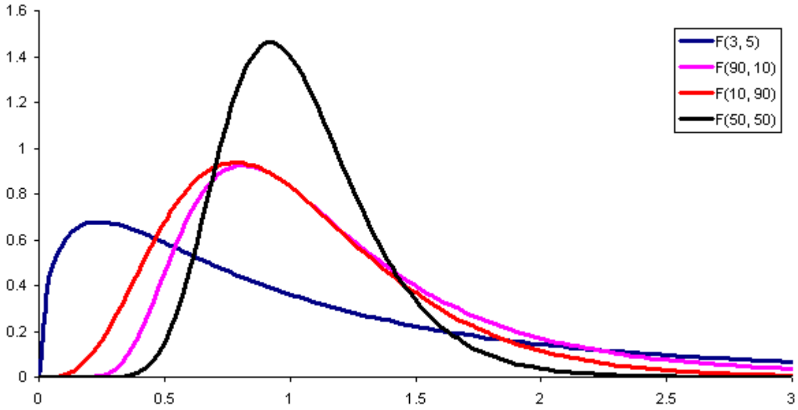
1. La distribución F es una distribución de familias
2. La distribución F es 0 o positiva pero no negativa
3. La distribución F es sesgada a la derecha

Distribución F

Usos:

1. Probar si dos muestras independientes han sido seleccionadas de poblaciones normales con la misma varianza
2. Probar si dos valores independientes de la varianza de la población son homogéneos o no

Distribución F



Prueba para todas las variables predictoras

Siendo el modelo θ igual a $y = X\beta + \epsilon$ donde X es una matriz $n \times p$ y el modelo reducido (ω) igual a $y = \mu\epsilon$. La hipótesis nula es:

$$H_0 : \beta_1 = ...\beta_{p-1} = 0$$

donde el estadístico sería:

$$F = \frac{(TSS - RSS)/(p-1)}{RSS/(n-p)}$$

con

$$RSS = RSS_\theta = (y - X\hat{\beta})'(y - X\hat{\beta})$$

y

$$RSS_\omega = (y - \bar{y})'(y - \bar{y}) = TSS$$

Prueba para todas las variables predictoras

```
tss <- sum((gala$Species - mean(gala$Species))^2)
regss <- sum((fitted(modcomp) - mean(gala$Species))^2)
rss <- tss - regss
(fstat <- ((tss - rss)/(ncol(x) - 1))/(rss/
                                    (nrow(x) - ncol(x))))

## [1] 15.69941

1 - pf(fstat, ncol(x) - 1, nrow(x) - ncol(x))

## [1] 6.837893e-07
```

Prueba para todas las variables predictoras

```
##
## Call:
## lm(formula = Species ~ Area + Elevation + Nearest + Scruz + Adjacent,
##     data = gala)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -111.679  -34.898   -7.862   33.460  182.584
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.068221  19.154198   0.369 0.715351
## Area        -0.023938   0.022422  -1.068 0.296318
## Elevation    0.319465   0.053663   5.953 3.82e-06 ***
## Nearest      0.009144   1.054136   0.009 0.993151
## Scruz       -0.240524   0.215402  -1.117 0.275208
## Adjacent    -0.074805   0.017700  -4.226 0.000297 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 60.98 on 24 degrees of freedom
## Multiple R-squared:  0.7658,Adjusted R-squared:  0.7171
## F-statistic:  15.7 on 5 and 24 DF,  p-value: 6.838e-07
```

Prueba para una sola variable independiente

Puede ser una variable independiente suprimida del modelo? La hipótesis nula sería:

$$H_0 : \beta_i = 0$$

En este caso, RSS_θ será el RSS del modelo con todas las variables explicatorias de interés que tiene p parámetros y RSS_ω será el RSS para el modelo con las mismas variables independientes excepto el predictor i

El estadístico F se calcula utilizando la fórmula estándar. Una alternativa es utilizar el estadístico t para probar esta hipótesis:

$$t_i = \hat{\beta}_i / se(\hat{\beta}_i)$$

y se verifica la significancia utilizando una distribución t con $n - p$ grados de libertad

Prueba para una sola variable independiente

Es posible omitir la variable Area del modelo?

```
modcomp2 <- lm(Species ~ Elevation + Nearest + Scruz + Adjacent,
               data = gala)
regss2 <- sum((fitted(modcomp2) - mean(gala$Species))^2)
rss2 <- tss - regss2
(fstat <- ((rss2 - rss)/1)/(rss/(nrow(x) - ncol(x))))

## [1] 1.139792

1 - pf(fstat, 1, nrow(x) - ncol(x))

## [1] 0.296318
```

Prueba para una sola variable independiente

El valor de p obtenido a partir del estadístico t será:

```
sqrt(fstat)

## [1] 1.067611

(tstat <- summary(modcomp)$coef[2, 3])

## [1] -1.067611

2 * (1 - pt(sqrt(fstat), nrow(x) - ncol(x)))

## [1] 0.296318
```

Prueba para una sola variable independiente

Una manera más conveniente de comparar dos modelos anidados es:

```
anova(modcomp, modcomp2)

## Analysis of Variance Table
##
## Model 1: Species ~ Area + Elevation + Nearest + Scruz + Adjacent
## Model 2: Species ~ Elevation + Nearest + Scruz + Adjacent
##   Res.Df   RSS Df Sum of Sq      F Pr(>F)
## 1     24 89231
## 2     25 93469 -1   -4237.7 1.1398 0.2963
```

Qué variables explicatorias debería incluir un modelo simplificado para las especies de tortuga?

Modelos con variables cualitativas

Los modelos lineales compuestos permiten la inclusión de variables cualitativas (categóricas) dentro de su estructura

R utiliza un esquema de códigos dummy donde compara cada nivel de la variable con respecto a un nivel de referencia fijo.

Modelos con variables cualitativas

```
data(iris)
head(iris)

##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

Modelos con variables cualitativas

```
##
## Call:
## lm(formula = Sepal.Length ~ Sepal.Width + Petal.Length + Petal.Width +
##     Species, data = iris)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.79424 -0.21874  0.00899  0.20255  0.73103
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        2.17127    0.27979   7.760 1.43e-12 ***
## Sepal.Width        0.49589    0.08607   5.761 4.87e-08 ***
## Petal.Length       0.82924    0.06853  12.101  < 2e-16 ***
## Petal.Width       -0.31516    0.15120  -2.084  0.03889 *
## Speciesversicolor -0.72356    0.24017  -3.013  0.00306 **
## Speciesvirginica  -1.02350    0.33373  -3.067  0.00258 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3068 on 144 degrees of freedom
## Multiple R-squared:  0.8673,Adjusted R-squared:  0.8627
## F-statistic: 188.3 on 5 and 144 DF,  p-value: < 2.2e-16
```

Modelos con variables cualitativas

La variable *Species* tiene tres niveles: *setosa*, *versicolor*, *virginica*

```
x <- model.matrix(irislm)
head(x)
```

```
##   (Intercept) Sepal.Width Petal.Length Petal.Width Speciesversicolor
## 1           1         3.5          1.4         0.2                 0
## 2           1         3.0          1.4         0.2                 0
## 3           1         3.2          1.3         0.2                 0
## 4           1         3.1          1.5         0.2                 0
## 5           1         3.6          1.4         0.2                 0
## 6           1         3.9          1.7         0.4                 0
##   Speciesvirginica
## 1                0
## 2                0
## 3                0
## 4                0
## 5                0
## 6                0
```

```
contr.treatment(3)
```

```
##   2 3
## 1 0 0
## 2 1 0
## 3 0 1
```

Estimación de parámetros

La estimación de parámetros sigue el mismo procedimiento. Para cada nivel de la variable categórica se estima su correspondiente parámetro, excepto para el primer nivel cuyo valor siempre será 0

```
y <- iris$Sepal.Length
solve(crossprod(x), crossprod(x, y))

##                          [,1]
## (Intercept)         2.1712663
## Sepal.Width         0.4958889
## Petal.Length        0.8292439
## Petal.Width        -0.3151552
## Speciesversicolor -0.7235620
## Speciesvirginica  -1.0234978
```

Predicción de la variable de respuesta

```
predict(irislm,
        newdata = data.frame(Sepal.Width = 3.5, Petal.Length = 1.6,
                             Petal.Width = 0.3, Species = "setosa"))

##        1
## 5.139121

coef(irislm)[1] +
  coef(irislm)[2] * 3.5 +
  coef(irislm)[3] * 1.6 +
  coef(irislm)[4] * 0.3 +
  0 +
  coef(irislm)[5] * 0 +
  coef(irislm)[6] * 0

## (Intercept)
##    5.139121
```

Predicción de la variable de respuesta

```
predict(irislm,
        newdata = data.frame(Sepal.Width = 3.5, Petal.Length = 1.6,
                             Petal.Width = 0.3, Species = "versicolor"))
```

```
##        1
## 4.415559
```

```
coef(irislm)[1] +
  coef(irislm)[2] * 3.5 +
  coef(irislm)[3] * 1.6 +
  coef(irislm)[4] * 0.3 +
  0 +
  coef(irislm)[5] * 1 +
  coef(irislm)[6] * 0
```

```
## (Intercept)
##    4.415559
```

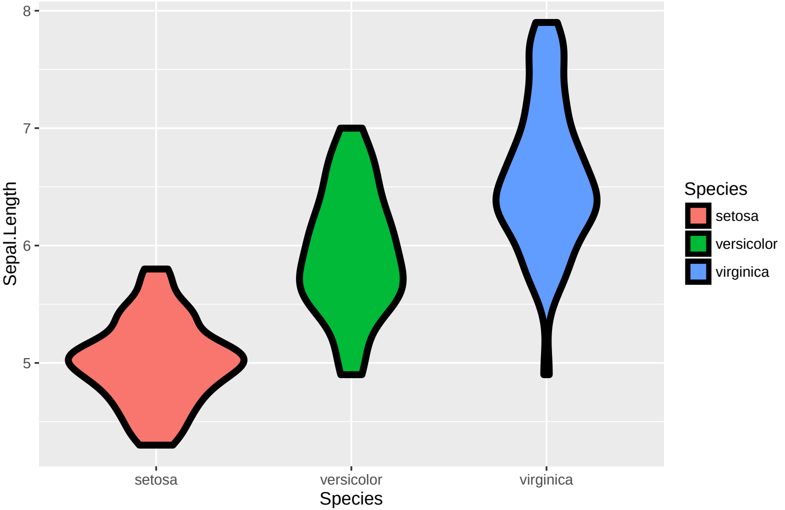
Predicción de la variable de respuesta

```
predict(irislm,
        newdata = data.frame(Sepal.Width = 3.5, Petal.Length = 1.6,
                             Petal.Width = 0.3, Species = "virginica"))

##        1
## 4.115623

coef(irislm)[1] +
  coef(irislm)[2] * 3.5 +
  coef(irislm)[3] * 1.6 +
  coef(irislm)[4] * 0.3 +
  0 +
  coef(irislm)[5] * 0 +
  coef(irislm)[6] * 1

## (Intercept)
##    4.115623
```

Modelos con variables cualitativas

```
##
## Call:
## lm(formula = Sepal.Length ~ Species, data = iris)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.6880 -0.3285 -0.0060  0.3120  1.3120
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)         5.0060     0.0728  68.762  < 2e-16 ***
## Speciesversicolor   0.9300     0.1030   9.033 8.77e-16 ***
## Speciesvirginica    1.5820     0.1030  15.366  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5148 on 147 degrees of freedom
## Multiple R-squared:  0.6187,Adjusted R-squared:  0.6135
## F-statistic: 119.3 on 2 and 147 DF,  p-value: < 2.2e-16
```

Modelos con variables cualitativas



Probando un subespacio

Suponga que usted desea predecir la calificación final de una asignatura (y) a partir de un conjunto de variables explicatorias entre las que se encuentran las calificaciones de los componentes teórico (X_j) y práctico (X_k) de la asignatura, los cuales son aditivos

La pregunta sería si se necesitan las dos calificaciones por separado o si se pueden reemplazar por el total $X_j + X_k$. El modelo original sería:

$$y = \beta_0 + ... + \beta_j X_j + \beta_k X_k + ... + \epsilon$$

Probando un subespacio

El modelo simplificado sería:

$$y = \beta_0 + ... + \beta_l (X_j + X_k) + ... + \epsilon$$

el cual requiere que $\beta_j = \beta_k$ para que esta simplificación sea posible. Luego la hipótesis nula será:

$$H_0 : \beta_j = \beta_k$$

Esto define un subespacio en el que el procedimiento de la prueba de F aplica.

Probando un subespacio

Ejemplo

Los datos *savings* contienen los promedios de variables económicas de 50 países de 1960 a 1970. *dpi* es el ingreso per cápita en dólares, *ddpi* es la tasa de cambio en el ingreso per cápita, *sr* es el ahorro personal dividido por el ingreso. También incluye el porcentaje de población menor a 15 ($pop15$) y mayor a 75 años ($pop75$)

```
## 'data.frame': 50 obs. of  5 variables:
##  $ sr   : num  11.43 12.07 13.17 5.75 12.88 ...
##  $ pop15: num  29.4 23.3 23.8 41.9 42.2 ...
##  $ pop75: num  2.87 4.41 4.43 1.67 0.83 2.85 1.34 0.67 1.06 1.14 ...
##  $ dpi  : num  2330 1508 2108 189 728 ...
##  $ ddpi : num  2.87 3.93 3.82 0.22 4.56 2.43 2.67 6.51 3.08 2.8 ...
```

Probando un subespacio

Ejemplo

En este caso, se puede plantear que el efecto de la gente joven y mayor es el mismo:

$$H_0 : \beta_{pop15} = \beta_{pop75}$$

En este caso el modelo tomaría la forma:

$$y = \beta_0 + \beta_{dep}(pop15 + pop75) + \beta_{dpi}dpi + \beta_{ddpi}ddpi + \epsilon$$

Probando un subespacio

Ejemplo

```
savlm <- lm(sr ~ pop15 + pop75 + dpi + ddpi, savings)
summary(savlm)
```

```
##
## Call:
## lm(formula = sr ~ pop15 + pop75 + dpi + ddpi, data = savings)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.2422 -2.6857 -0.2488  2.4280  9.7509
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 28.5660865  7.3545161   3.884 0.000334 ***
## pop15       -0.4611931  0.1446422  -3.189 0.002603 **
## pop75       -1.6914977  1.0835989  -1.561 0.125530
## dpi         -0.0003369  0.0009311  -0.362 0.719173
## ddpi         0.4096949  0.1961971   2.088 0.042471 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.803 on 45 degrees of freedom
## Multiple R-squared:  0.3385,Adjusted R-squared:  0.2797
## F-statistic: 5.756 on 4 and 45 DF,  p-value: 0.0007904
```

Probando un subespacio

Ejemplo

```
savlm2 <- lm(sr ~ I(pop15 + pop75) + dpi + ddpi, savings)
summary(savlm2)
```

```
##
## Call:
## lm(formula = sr ~ I(pop15 + pop75) + dpi + ddpi, data = savings)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -7.787 -2.767 -0.125  1.744 10.342
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      21.6093051  4.8833633   4.425 5.87e-05 ***
## I(pop15 + pop75) -0.3336331  0.1038679  -3.212  0.00241 **
## dpi              -0.0008451  0.0008444  -1.001  0.32212
## ddpi              0.3909649  0.1968714   1.986  0.05302 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.827 on 46 degrees of freedom
## Multiple R-squared:  0.3152,Adjusted R-squared:  0.2705
## F-statistic: 7.056 on 3 and 46 DF,  p-value: 0.0005328
```

Probando un subespacio

Ejemplo

```
anova(savlm2, savlm)

## Analysis of Variance Table
##
## Model 1: sr ~ I(pop15 + pop75) + dpi + ddpi
## Model 2: sr ~ pop15 + pop75 + dpi + ddpi
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     46 673.63
## 2     45 650.71  1    22.915 1.5847 0.2146
```

Definición de parámetros a priori

Suponga que se desea probar si el valor de uno de los coeficientes de la regresión se puede definir de antemano. Por ejemplo:

$$H_0 : \beta_{ddpi} = 0.5$$

El modelo nulo tomaría la siguiente forma:

$$y = \beta_0 + \beta_{pop15} pop15 + \beta_{pop75} pop75 + \beta_{dpi} dpi + 0.5 ddpi + \epsilon$$

El valor fijo del parámetro *ddpi* en la ecuación de regresión se denomina un *offset*

Definición de parámetros a priori

Suponga que se desea probar si el valor de uno de los coeficientes de la regresión se puede definir de antemano. Por ejemplo:

$$H_0 : \beta_{ddpi} = 0.5$$

El modelo nulo tomaría la siguiente forma:

$$y = \beta_0 + \beta_{pop15} pop15 + \beta_{pop75} pop75 + \beta_{dpi} dpi + 0.5 ddpi + \epsilon$$

El valor fijo del parámetro *ddpi* en la ecuación de regresión se denomina un *offset*

Probando un subespacio

Ejemplo

```
savlm3 <- lm(sr ~ pop15 + pop75 + dpi + offset(0.5 * ddpi),
             savings)
anova(savlm3, savlm)

## Analysis of Variance Table
##
## Model 1: sr ~ pop15 + pop75 + dpi + offset(0.5 * ddpi)
## Model 2: sr ~ pop15 + pop75 + dpi + ddpi
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     46 653.78
## 2     45 650.71  1    3.0635 0.2119 0.6475
```

Ejercicio a resolver

Una firma de abogados ha sido consultada para representar a un grupo de mujeres que pretenden demandar a su empleador por discriminación de género, especialmente por el pago recibido. Las mujeres argumentan que los incrementos salariales son de manera consistente y considerable más bajos que los aumentos que obtienen los hombres.

De otro lado la compañía argumenta que los incrementos están basados enteramente en el rendimiento en el trabajo, el cual es medido por un supervisor imparcial a través de una evaluación que incluye una serie de indicadores de rendimiento. Se ha pedido que la firma de abogados realice una evaluación preliminar de los méritos de la demanda.

Ejercicio a resolver

Situación de algunos empleados de la compañía seleccionados al azar:

Individuo	Sexo	Calificación calidad	Años	División	Incremento
1	F	10	9	Producción	21000
2	F	90	1	Producción	96000
3	F	20	4	Producción	47000
4	F	80	1	Producción	128000
5	F	30	4	Investigación	64000
6	F	70	1	Investigación	52000
7	F	10	4	Ventas	73000
8	F	15	7	Producción	19000
9	M	20	6	Investigación	128000
10	M	80	3	Ventas	474000
11	M	50	3	Investigación	342000
12	M	70	2	Ventas	330000
13	M	30	7	Ventas	185000
14	M	70	7	Ventas	331000
15	M	40	1	Ventas	267000
16	M	90	6	Producción	517000
17	M	50	8	Producción	390000

Ejercicio a resolver

- Sería admitible la demanda de las mujeres?
- Cuáles son los factores que definen el incremento de salarios en la compañía?
- Cómo afecta el género el incremento de salarios en la compañía?