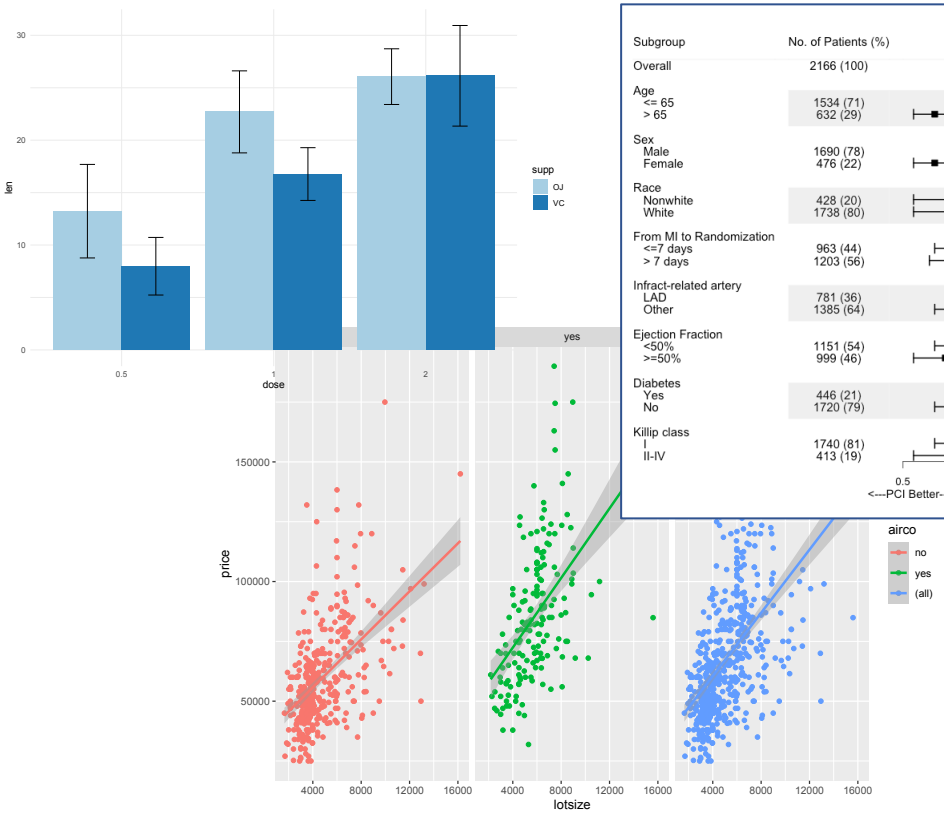
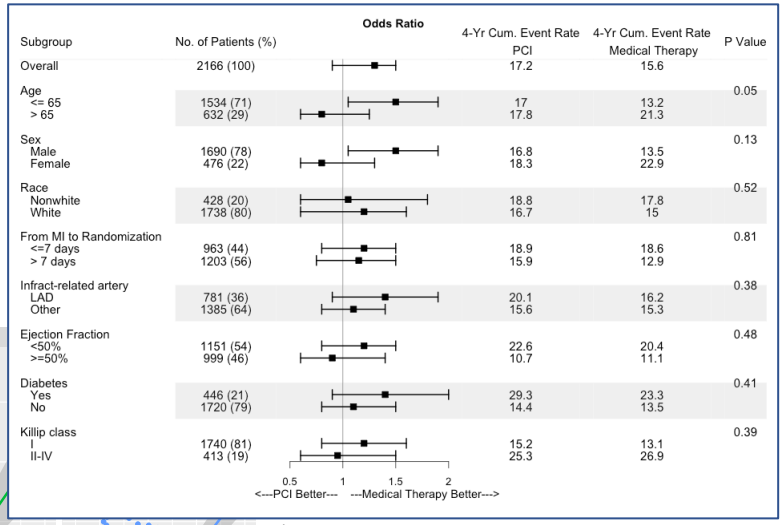
Análisis Estadístico con





8. Regresión logística

Índice

8.1. Introducción

8.2. Odds ratios

8.3. Predicción

8. Regresión logística

8.1.Introducción

- Outcome binarios
- Método de estimación de parámetros; Máxima verosimilitud
- Aquellos parámetros que optimizan un criterio de ajuste.
- Pertenece a los modelos lineales generalizados (en este caso la variable respuesta sigue una distribución binomial)
- **glm**(model , family= binomial)

$$P(Y_i = y_i) = \binom{n_i}{y_i} p_i^{y_i} (1 - p_i)^{n_i - y_i}$$

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_q x_{iq} = X_i \beta$$

8. Regresión logística

8.1.Introducción

mtcars

[, 1] mpg Miles/(US) gallon
[, 2] cyl Number of cylinders
[, 3] disp Displacement (cu.in.)
[, 4] hp Gross horsepower
[, 5] drat Rear axle ratio
[, 6] wt Weight (1000 lbs)
[, 7] qsec 1/4 mile time
[, 8] vs V/S
[, 9] am Transmission (0 = automatic, 1 = manual)
[,10] gear Number of forward gears
[,11] carb Number of carburetors

```
> modelo.logistico<- glm(vs ~ am, data=mtcars, family=binomial)
> modelo.logistico

Call:  glm(formula = vs ~ am, family = binomial, data = mtcars)

Coefficients:
(Intercept)           am
    -0.5390       0.6931

Degrees of Freedom: 31 Total (i.e. Null);  30 Residual
Null Deviance:      43.86
Residual Deviance: 42.95        AIC: 46.95
> summary(modelo.logistico)

Call:
glm(formula = vs ~ am, family = binomial, data = mtcars)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.2435  -0.9587  -0.9587   1.1127   1.4132

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.5390     0.4756  -1.133    0.257
am            0.6931     0.7319   0.947    0.344

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 43.860  on 31  degrees of freedom
Residual deviance: 42.953  on 30  degrees of freedom
AIC: 46.953

Number of Fisher Scoring iterations: 4
```

8. Regresión logística

8.1.Introducción

Wald test

```
> modelo.logistico<- glm(vs ~ as.factor(am), data=mtcars, family=binomial)
> summary(modelo.logistico)

Call:
glm(formula = vs ~ as.factor(am), family = binomial, data = mtcars)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.2435  -0.9587  -0.9587   1.1127   1.4132

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   -0.5390     0.4756  -1.133    0.257
as.factor(am)1   0.6931     0.7319   0.947    0.344

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 43.860  on 31  degrees of freedom
Residual deviance: 42.953  on 30  degrees of freedom
AIC: 46.953

Number of Fisher Scoring iterations: 4
```

```
> confint.default(modelo.logistico)
                    2.5 %    97.5 %
(Intercept)    -1.4711453 0.3931523
as.factor(am)1 -0.7413995 2.1276939
```

Confidence intervals likelihood ratio test

```
> confint(modelo.logistico)
Waiting for profiling to be done...
                    2.5 %    97.5 %
(Intercept)    -1.5274824 0.3716003
as.factor(am)1 -0.7334126 2.1696774
```

```
> anova(modelo.logistico,test="Chisq")
Analysis of Deviance Table

Model: binomial, link: logit

Response: vs

Terms added sequentially (first to last)


             Df Deviance Resid. Df Resid. Dev Pr(>Chi)
NULL                            31     43.860
as.factor(am)  1   0.9071        30     42.953   0.3409
```

```
> table(mtcars$vs,mtcars$am)

   0  1
0 12  6
1  7  7
```

```
> prop.test(c(7,7),c(19,13),correct=F)

        2-sample test for equality of proportions without continuity
        correction

data:  c(7, 7) out of c(19, 13)
X-squared = 0.90688, df = 1, p-value = 0.3409
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.517146  0.177065
sample estimates:
   prop 1    prop 2
0.3684211 0.5384615
```

8. Regresión logística

8.2.Odds ratios

	Diseased	Healthy
Exposed	D_E	H_E
Not exposed	D_N	H_N

$$OR = \frac{D_E/H_E}{D_N/H_N}$$

Odds de desarrollar enfermedad en el grupo de expuestos

Odds de desarrollar enfermedad en el grupo de no expuestos

8. Regresión logística

8.2.Odds ratios

```
> table(mtcars$vs,mtcars$am)

   0  1
0 12  6
1  7  7
```

Odds am_1= 7/6

Odds am_0=7/12

OR=2

```
> modelo.logistico<- glm(vs ~ as.factor(am), data=mtcars, family=binomial)
> summary(modelo.logistico)

Call:
glm(formula = vs ~ as.factor(am), family = binomial, data = mtcars)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.2435  -0.9587  -0.9587   1.1127   1.4132

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   -0.5390     0.4756  -1.133    0.257
as.factor(am)1   0.6931     0.7319   0.947    0.344

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 43.860  on 31  degrees of freedom
Residual deviance: 42.953  on 30  degrees of freedom
AIC: 46.953

Number of Fisher Scoring iterations: 4
```

exp(0.6931..)= 2

8. Regresión logística

8.2.Odds ratios

exp(cbind(OR=coef(modelo.logistico),confint(modelo.logistico)))

```
                      OR     2.5 %   97.5 %
(Intercept)    0.5833333 0.2170815 1.450053
as.factor(am)1 2.0000000 0.4802672 8.755459
```

8. Regresión logística

8.3.Predicción

- predict(modelo.logisto,type="response")
- Importante para generar matriz de confusión

```
> cbind(mtcars$vs,predict(modelo.logistico,type="response"))
                     [,1]      [,2]
Mazda RX4               0 0.5384615
Mazda RX4 Wag           0 0.5384615
Datsun 710              1 0.5384615
Hornet 4 Drive          1 0.3684211
Hornet Sportabout       0 0.3684211
Valiant                 1 0.3684211
Duster 360              0 0.3684211
Merc 240D               1 0.3684211
Merc 230                1 0.3684211
Merc 280                1 0.3684211
Merc 280C               1 0.3684211
Merc 450SE              0 0.3684211
Merc 450SL              0 0.3684211
Merc 450SLC             0 0.3684211
Cadillac Fleetwood      0 0.3684211
Lincoln Continental     0 0.3684211
Chrysler Imperial       0 0.3684211
Fiat 128                1 0.5384615
Honda Civic             1 0.5384615
Toyota Corolla          1 0.5384615
Toyota Corona           1 0.3684211
Dodge Challenger        0 0.3684211
AMC Javelin             0 0.3684211
Camaro Z28              0 0.3684211
Pontiac Firebird        0 0.3684211
Fiat X1-9               1 0.5384615
Porsche 914-2           0 0.5384615
Lotus Europa            1 0.5384615
Ford Pantera L          0 0.5384615
Ferrari Dino            0 0.5384615
Maserati Bora           0 0.5384615
Volvo 142E              1 0.5384615
```

8. Regresión logística

Otras cosas

Chequeo de los modelos (diagnóstico)

Curva ROC (sensibilidad frente a especificidad)

Modelo multivariable (ver ejercicios)

EJERCICIOS
"ejercicios.8.regresion. Logistica.R"