

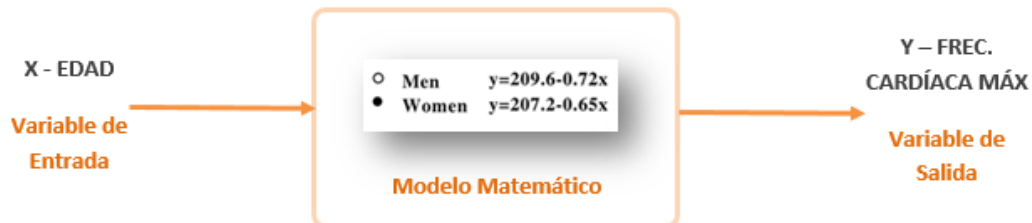
INTRODUCCIÓN A LOS MODELOS

Empezamos con los modelos predictivos.

Es un poco diferente a lo que has visto hasta ahora sobre los test estadísticos pero muy cercano a los test de tablas ANOVA.

Recuerda que una tabla ANOVA es un modelo lineal.

Y que un modelo se puede entender cómo entrada y salida:



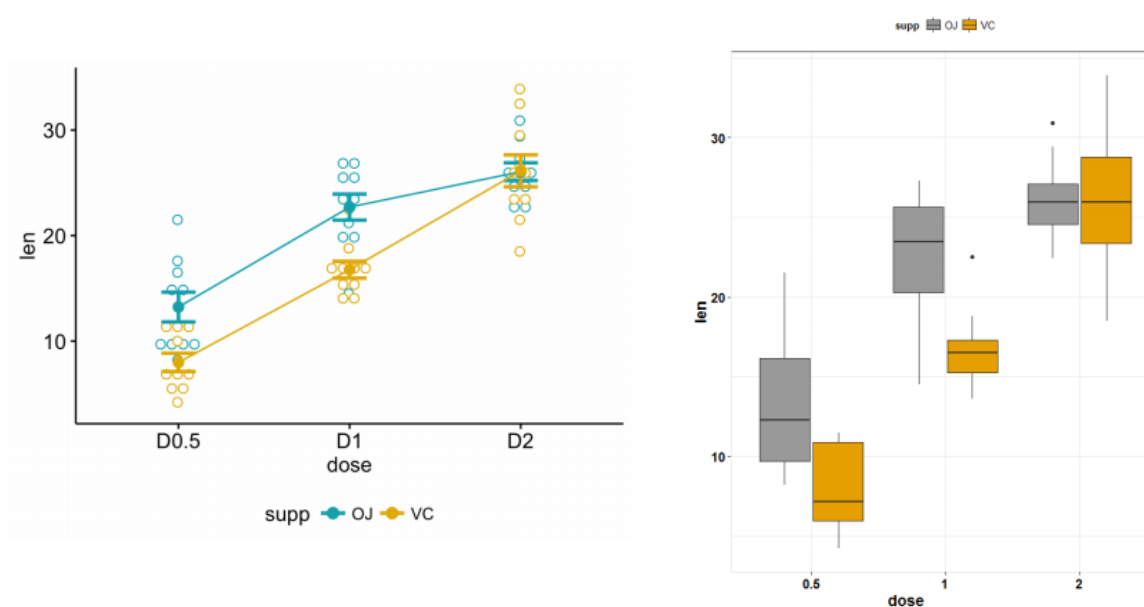
ANOVA COMO UN MODELO LINEAL

Vamos a repasar el ejemplo sobre los ratones de laboratorio.

Queremos saber cómo influye la cantidad suministrada ("Dose") y el tipo de sustancia ("Supp") en el crecimiento de los dientes:

- medida □ Log. del diente "len"
- 2 factores □
 - Tipo de Suplemente "Supp"
 - Dosis suministrada "Dose"

De forma descriptiva:



INTRODUCCIÓN A LOS MODELOS

```

      Df Sum Sq Mean Sq F value    Pr(>F)
supp    1  205.4    205.4    14.02 0.000429 ***
dose    2 2426.4   1213.2    82.81 < 2e-16 ***
Residuals 56  820.4     14.7
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

La estructura de la tabla que ves arriba es la que verás en todos los modelos.

Ahora coloca las variables en la tabla de los modelos según la variable de entrada y salida:

Variable de Entrada Variable Independientes Variable de Estudio -Eje horizontal (son sinónimos)	SUPP Dose
Variable de Salida Variable Dependiente Variable respuesta -Eje vertical (son sinónimos)	Longitud del diente

¿Qué conclusiones te aporta este modelo ANOVA?

El modelo ANOVA demuestra que tanto el factor dosis como el SUPP influyen sobre la longitud del diente, siendo esta relación significativa en ambos factores con p pequeñas

PRACTICANDO LA ESTRUCTURA DE UN MODELO

Queremos saber la influencia causa-efecto entre el consumo y el peso de vehículo.

El resultado del modelo es el siguiente:

INTRODUCCIÓN A LOS MODELOS

```

call:
lm(formula = consumo ~ peso, data = varEstudio)

Residuals:
    Min       1Q   Median       3Q      Max
-0.0179837 -0.0044301  0.0009413  0.0045354  0.0116583

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.006169   0.004695   1.314   0.199
peso         0.014938   0.001398  10.685 9.57e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.007616 on 30 degrees of freedom
Multiple R-squared:  0.7919,    Adjusted R-squared:  0.785
F-statistic: 114.2 on 1 and 30 DF,  p-value: 9.566e-12

```

Según la tabla coloca las variables en la tabla de los modelos según la variable/s de entrada y salida:

Variable/s de Entrada Variable Independientes Variable de Estudio -Eje horizontal (son sinónimos)	Peso
Variable de Salida Variable Dependiente Variable respuesta -Eje vertical (son sinónimos)	Consumo

¿Qué conclusiones te aporta este modelo?

El modelo explica que existe relación entre el peso del vehículo y su consumo, siendo esta relación directa, a mas peso mas consumo, y significativa con una $p < 0.05$

Imagínate que tienes datos de personas que tienen problemas de apnea del sueño. Por si no lo sabes, son los problemas respiratorios al dormir. Dejan de respirar durante unos segundos por un problema en el músculo el cuello. Y lo repiten durante la noche con frecuencia.

No descansan bien y a la larga pueden tener repercusiones importantes en la salud del corazón.

INTRODUCCIÓN A LOS MODELOS

A ti te interesa analizar estas personas y ver cuando la enfermedad es más grave o cuando no.

Hay una variable que indica la situación del paciente es grave o no. Es el número de apneas en una hora. Si es más de 30 la cosa es peligrosa. Si no pasa este número se puede considerar un enfermo no tan grave.

Ahora lo que quieres es tener un modelo para predecir la probabilidad de estar grave sabiendo la presión diastólica (la presión baja) del paciente.

Una variable numérica, la presión máxima o sistólica.

Una categórica dicotómica, IAH30

- 1 = Es grave
- 0 = No es grave

Este es el resultado del modelo:

```
call:
glm(formula = IAH30 ~ TAS_m, family = "binomial", data = varEstudio)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9949  -1.0487   0.6106   1.0163   1.7737

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -6.16228    1.47788  -4.170 3.05e-05 ***
TAS_m        0.04681    0.01093   4.284 1.84e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 218.12  on 157  degrees of freedom
Residual deviance: 196.42  on 156  degrees of freedom
(1 observation deleted due to missingness)
AIC: 200.42

Number of Fisher Scoring iterations: 4
```

Para este caso crearás una regresión lineal de dos variables numéricas. ¿Y cómo? La variable la vas a convertir a probabilidad.

- La variable Y es la probabilidad de estar grave. Ya has convertido la variable categórica a una numérica.
- La variable X es la variable Presión diastólica (la baja)

Sería algo así:

$$\text{logit}(P(\text{grave})) = a + b \cdot \text{Presión Diastólica}$$

INTRODUCCIÓN A LOS MODELOS

Según la tabla coloca las variables en la tabla de los modelos según la variable/s de entrada y salida:

Variable/s de Entrada Variable Independientes Variable de Estudio -Eje horizontal (son sinónimos)	Presión diastólica
Variable de Salida Variable Dependiente Variable respuesta -Eje vertical (son sinónimos)	IAH30 (gravedad)

¿Cómo es la ecuación del modelo?

$\text{logit}(P(\text{grave})) = a + b \cdot \text{Presión Diastólica}$

¿Qué valor es el de a y el de b?

A: -6.16

B: 0.04

$-6.16 + 0.04 \cdot \text{Presión diastólica}$

¿Qué conclusiones te aporta este modelo?

El modelo muestra que existe relación entre la gravedad de la apnea y la presión diastólica y esta relación es significativa y que a mayor presión diastólica, mayor probabilidad de tener mayor número de apneas y ser clasificado como grave.

Poco a poco construimos las bases para entender la esencia de los modelos ;)

¡Enhorabuena y sigue avanzando!