

Prueba de Chi Cuadrado y Correlaciones

Aleksander Dietrichson, PhD

17 de junio 2021

Agenda

1. Prueba Chi Cuadrado
2. Correlacion Pearson
3. Correlación Spearman y Kendall
4. Ejemplos

Prueba Chi Cuadrado

- ▶ Permite realizar un test de independencia cuando las variables son *nominales*
- ▶ Funciona sobre frecuencias (no proporciones o porcentajes)
- ▶ Responde la pregunta de si los valores encontrados difieren de manera significativo de los esperados según su definición:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

¿cómo saber cual es el valor esperado?

- ▶ Son valores teóricos, p.ej. Esperamos que sexo se distribuye mitad y mitad (+/-)
- ▶ Se calculan basados en la misma muestra

Ejemplo

Table 1: Opinión sobre la Ley del Aborto

	A favor	En contra
Mujeres	762	468
Hombres	484	477

$$E_{1,1} = \frac{1230 \times 1246}{2191} = 699,48$$

Ejemplo

Si lo calculamos para todas las celdas obtenemos:

Table 2: Valores esperados: opiniones sobre la ley del aborto.

	A favor	En contra	total
Mujeres	699,49	530,51	1230
Hombres	546,51	414,49	961
total	1246,00	945,00	2191

Ejemplo

Table 3: Diferencias entre valores observados y esperados.

	A favor	En contra
Mujeres	62,51	-62,51
Hombres	-62,51	62,51

$$\chi^2 = \sum \frac{(O - E)^2}{E} = \frac{62,52^2}{699,49} + \frac{-62,52^2}{530,51} + \frac{62,52^2}{546,51} + \frac{-62,52^2}{414,49} = 29,53.$$

Ejemplo

O hacerlo con R:

```
M <- as.table(rbind(c(762, 468), c(484, 477)))  
dimnames(M) <- list(gender = c("Mujeres", "Hombres"),  
                     party = c("A favor", "En contra"))
```

```
chisq.test(M)
```

```
##
```

```
## Pearson's Chi-squared test with Yates' continuity correction
```

```
##
```

```
## data: M
```

```
## X-squared = 29.06, df = 1, p-value = 7.019e-08
```


Comparar con valores conocidos

```
x <- c(Mujeres=123,Hombres=96)
chisq.test(x) # Assume 50-50
```

```
##
## Chi-squared test for given probabilities
##
## data:  x
## X-squared = 3.3288, df = 1, p-value = 0.06808
```

Definir probabilidades

```
chisq.test(x, p = c(0.50,0.50))
```

```
##
```

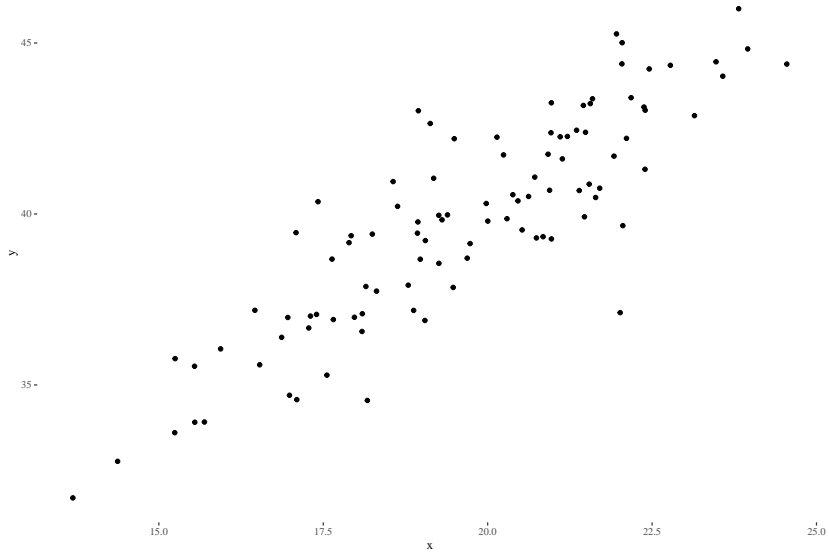
```
##  Chi-squared test for given probabilities
```

```
##
```

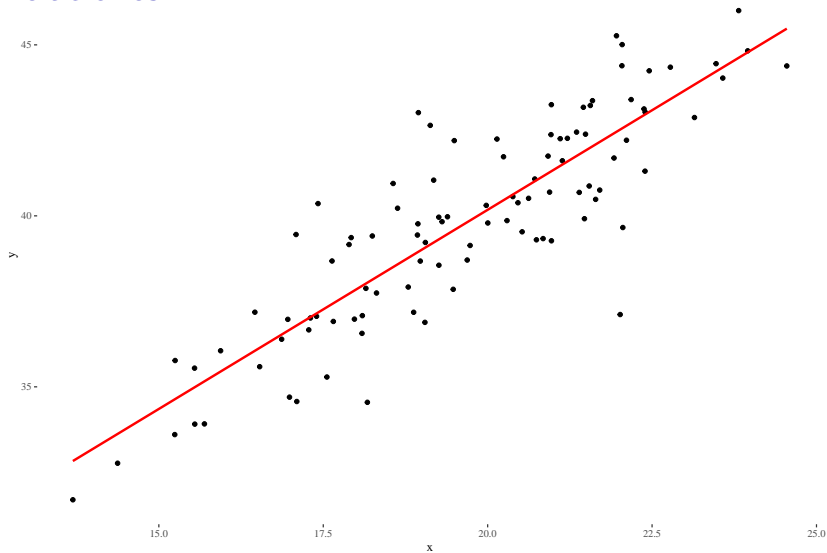
```
## data:  x
```

```
## X-squared = 3.3288, df = 1, p-value = 0.06808
```

Correlaciones



Correlaciones

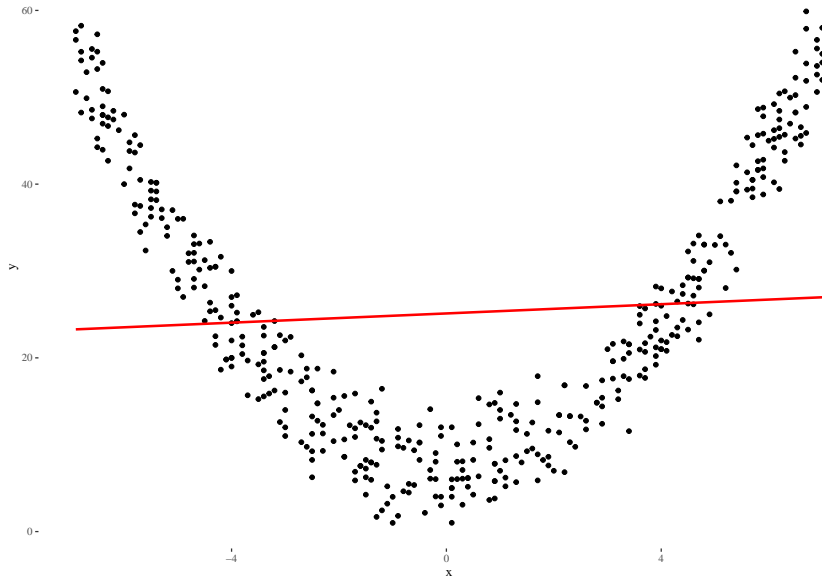


Correlaciones

60 -



Correlaciones



Correlación Pearson

- ▶ Mide correlación entre dos variables
- ▶ Deben ser de intervalo o razón
- ▶ Toma en cuenta la *magnitud relativa*

Correlación Pearson

Table 4: Resultados de un examen de traducción (x) y de comprensión (y) de ingles.

Estudiante	x	y
1	17	15
2	13	13
3	12	8
4	14	17
5	15	16
6	8	9
7	9	14
8	13	10
9	11	16
10	14	13
11	12	14
12	16	17

Correlación Pearson

```
datos<- data.frame(  
  Estudiante = c( 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12),  
  x = c( 17, 13, 12, 14, 15, 8, 9, 13, 11, 14, 12, 16),  
  y = c( 15, 13, 8, 17, 16, 9, 14, 10, 16, 13, 14, 17)  
)  
cor(datos$x, datos$y)
```

```
## [1] 0.5031258
```

► ¿Interpretación?

Correlación Spearman

- ▶ Cuando los datos son *Ordinales*
- ▶ Cuando son de ratio/intervalo y hay valores extremos
- ▶ Funciona sobre el *rango* u *orden*

Correlación Spearman

```
rankings <- data.frame(  
  Cafe.A = c(7, 6, 4, 5, 8, 7, 10, 3, 9, 2),  
  Cafe.B = c(5, 4, 5, 6, 10, 7, 9, 2, 8, 1)  
)  
  
# Llamamos función  
cor(rankings$Cafe.A, rankings$Cafe.B, method = "spearman")  
  
## [1] 0.875
```

Correlación Kendall

- ▶ Alternativa de Spearman
- ▶ Diferencias teóricas

Correlación Kendall

```
cor(rankings$Cafe.A, rankings$Cafe.B, method = "kendall")
```

```
## [1] 0.75
```

Correlación Spearman con Valores Extremos

```
datos<- data.frame(  
  Estudiante = c( 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12),  
  x = c( 17, 13, 120, 14, 15, 8, 9, 13, 11, 14, 120, 160),  
  y = c( 16, 13, 8, 17, 16, 9, 14, 10, 16, 13, 14, 17)  
)  
cor(datos$x,datos$y)
```

```
## [1] -0.008842766
```

```
cor(datos$x,datos$y, method = "spearman")
```

```
## [1] 0.3024988
```