Actividad grupal: test de hipótesis bayesiano en R.

- Alumnos
  - Pardavila Herrero, Saroi
  - Esmerado Vela, Javier
  - Boltà Chumillas, Jaume
  - Jiménez Magallanes, Misael Sebastián
  - Saavedra Infante, Carlos Daniel

Antes de empezar, instalamos las librerías que usaremos a lo largo de la actividad.

```
library(ggplot2)
library(plotly)
library(BayesFactor)
library(rstanarm)
library(see)
library(bayestestR)
```

---

1. Indica si la correlación entre las dos variables numéricas es significativa empleando el coeficiente de correlación de Pearson. Genera una visualización que muestre la dispersión de las dos variables junto a su línea de regresión.

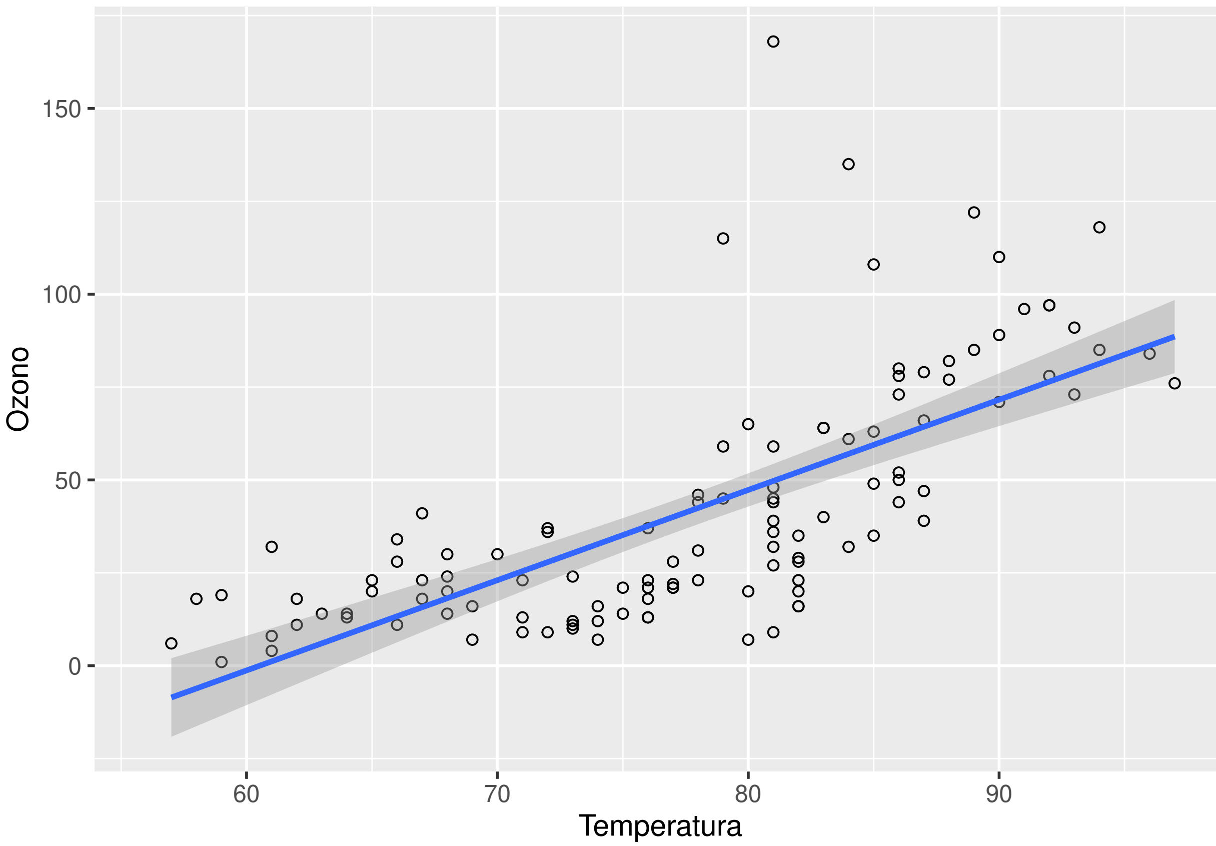
Empezaremos calculando la correlación muestral de las variables temperatura y ozono.

```
cor(airquality$Temp, airquality$Ozone, use = "complete.obs")
```

```
## [1] 0.6983603
```

Generamos una visualización que muestra las dos variables junto a su línea de regresión.

```
ggplot(data = airquality, aes(x=Temp, y=Ozone)) +
  # Agregamos los puntos
  geom_point(
    shape=1
  ) +
  # Agregamos la recta (lm: linear model)
  geom_smooth(
    method = lm
  ) +
  # Agregamos las etiquetas
  labs(
    x = "Temperatura",
    y = "Ozono"
  )
```



Finalmente podemos realizar el test clásico de la correlación.

```
cor.test(airquality$Temp, airquality$Ozone)
```

```
##
##  Pearson's product-moment correlation
##
## data:  airquality$Temp and airquality$Ozone
## t = 10.418, df = 114, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.5913340 0.7812111
## sample estimates:
##       cor
## 0.6983603
```

Podemos observar que el P-Value es menor que 2,2e-16, órdenes de magnitud menor que 0,05. Esto nos indica que los resultados obtenidos a favor de la hipótesis alternativa son estadísticamente significativos, es decir, la correlación poblacional entre las dos variables es distinta de 0.

---

2. Calcula ahora la correlación bayesiana de las dos mismas variables y calcula sus posteriores. (Tip: usa la librería BayesFactor).

Calculamos el factor de Bayes para correlaciones lineales. Guardaremos el resultado en la variable "result" para poder reutilizarlo más eficientemente.

```
result <- correlationBF(airquality$Temp, airquality$Ozone)
result
```

```
## Bayes factor analysis
## --------------
## [1] Alt., r=0.333 : 1.149457e+15 ±0%
##
## Against denominator:
##   Null, rho = 0
## ---
## Bayes factor type: BFcorrelation, Jeffreys-beta*
```

```
describe_posterior(result)
```

```
## Summary of Posterior Distribution
##
## Parameter | Median |       95% CI |   pd |          ROPE | % in ROPE |       BF |         Prior
## ---------------------------------------------------------------------------------------------------
## rho       |   0.68 | [0.57, 0.76] | 100% | [-0.05, 0.05] |        0% | 1.15e+15 | Beta (3 +- 3)
```

Obtenemos una descripción de la distribución posterior. De aquí podemos observar que la correlación bayesiana entre las dos variables da 0,68 (estimando a partir de la mediana).

---

3. Calcula ahora el factor de Bayes. Como sabrás, una prueba de correlación en realidad compara dos hipótesis, una nula (ausencia de efecto) con una alternativa (presencia de efecto). El factor de Bayes (BF) permite la misma comparación y determina bajo cuál de dos modelos los datos observados son más probables. Con el factor de Bayes, puedes medir la evidencia en contra y a favor del nulo.

Del mismo modo que antes, guardamos el resultado en la variable "bf" para su posterior utilización (al graficarlo).

```
bf <- bayesfactor(result)
bf
```

```
## Bayes Factors for Model Comparison
##
##     Model            BF
## [2] (rho != 0) 1.15e+15
##
## * Against Denominator: [1] (rho = 0)
## *   Bayes Factor Type: JZS (BayesFactor)
```

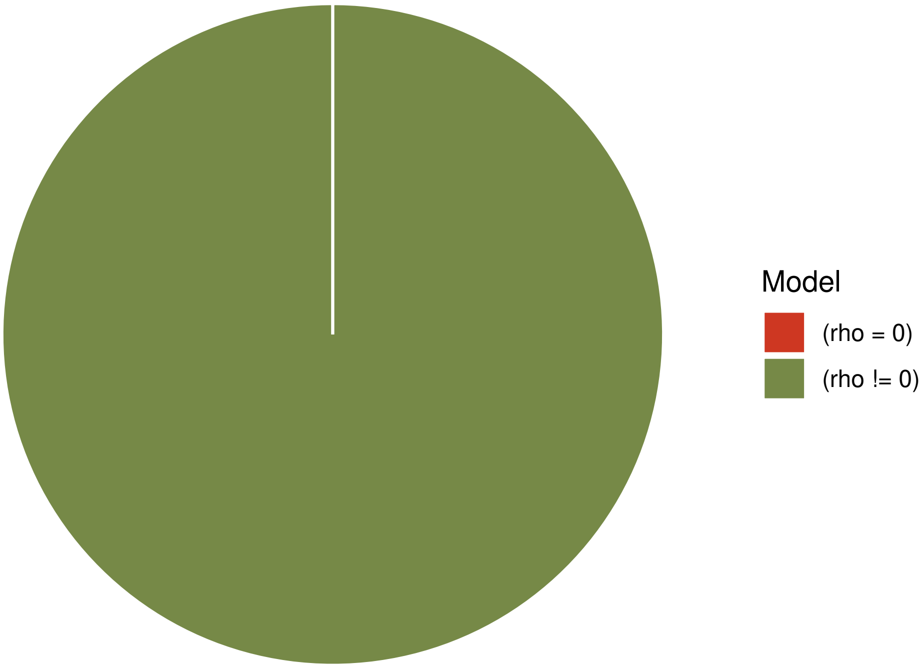
Podemos observar que el factor de Bayes es del orden de 1,15e+15, órdenes de magnitud mayor de 100, con lo cual según la escala de interpretación de Jeffreys la evidencia a favor de M1 es decisiva.

---

4. Genera ahora una visualización del factor de Bayes que te permita sacar conclusiones del estudio de manera clara y sencilla.

Emplearemos el típico gráfico de tarta.

```r
plot(bf) + scale_fill_pizza()
```



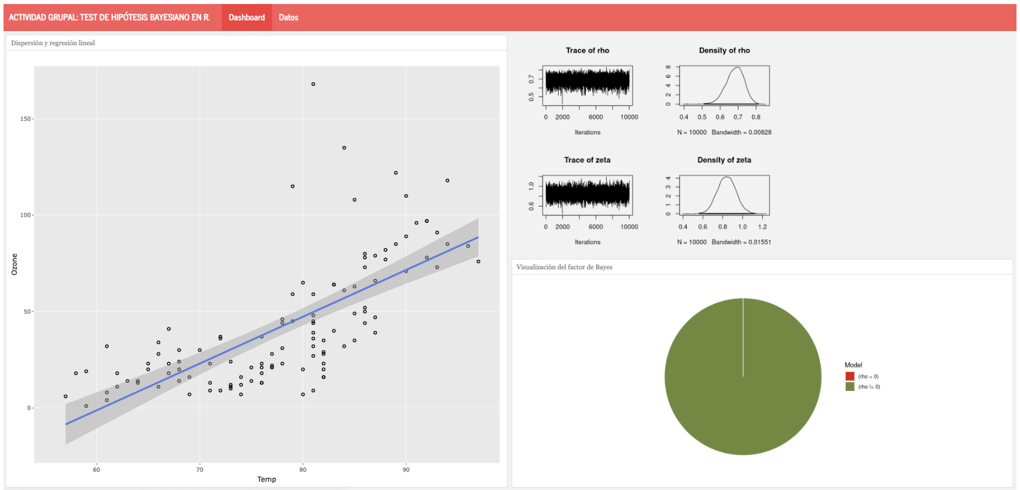
Vemos que no se parece demasiado a los gráficos de tarta que estamos acostumbrados a ver. Esto es porque al ser la evidencia a favor de rho != 0 órdenes de magnitud más grande que a favor de rho = 0, esta ocupa prácticamente todo el gráfico.

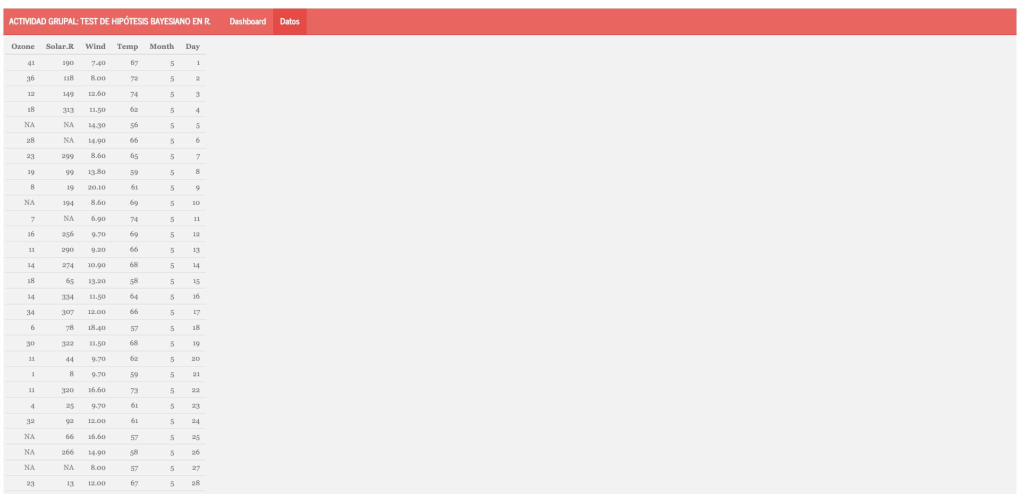
---

5. Concluye si las dos variables tienen correlación o no sobre la base de los estudios realizados en los puntos anteriores.

Tras todos los estudios realizados, tanto por Pearson como por Bayes podemos concluir que la correlación poblacional entre las dos variables es distinta de cero. Concretamente, existe una fuerte correlación positiva entre ambas.

Adicionalmente, hemos decidido crear un dashboard similar al de la actividad pasada mostrando los diagramas que hemos generado a lo largo de la actividad a demás del dataset con el que hemos estado trabajando.





ACTIVIDAD GRUPAL: TEST DE HIPÓTESIS BAYESIANO EN R. Dashboard Datos

Ozone	Solar.R	Wind	Temp	Month	Day
41	190	7.40	67	5	1
36	118	8.00	72	5	2
12	149	12.60	74	5	3
18	313	11.50	62	5	4
NA	NA	14.30	56	5	5
28	NA	14.90	66	5	6
23	299	8.60	65	5	7
19	99	13.80	59	5	8
8	19	20.10	61	5	9
NA	194	8.60	69	5	10
7	NA	6.90	74	5	11
16	256	9.70	69	5	12
11	290	9.20	66	5	13
14	274	10.90	68	5	14
18	65	13.20	58	5	15
14	334	11.50	64	5	16
34	307	12.00	66	5	17
6	78	18.40	57	5	18
30	322	11.50	68	5	19
11	44	9.70	62	5	20
1	8	9.70	59	5	21
11	320	16.60	73	5	22
4	25	9.70	61	5	23
32	92	12.00	61	5	24
NA	66	16.60	57	5	25
NA	266	14.90	58	5	26
NA	NA	8.00	57	5	27
23	13	12.00	67	5	28
45	252	14.90	81	5	29