

# Taller ley de Benford

## Contenidos

<b>La distribución de probabilidad de Benford</b>	<b>1</b>
El primer dígito de las potencias de 2 y de 5 . . . . .	2
El primer dígito de las potencias de 2 . . . . .	2
El primer dígito de las potencias de 5 . . . . .	4

## La distribución de probabilidad de Benford

La [ley de Benford](#) es una distribución discreta que siguen las frecuencias de los primeros dígitos significativos (de 1 a 9) de algunas series de datos que os pueden resultar sorprendentes como los precios, los valores de la contabilidad de una empresa, algunas tablas mediciones científicas o resultados de funciones trigonométricas.

Sea una v.a.  $X$  con dominio  $D_X = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$  diremos que sigue una ley de Benford si

$$P(X = x) = \log_{10} \left( 1 + \frac{1}{x} \right) = \log_{10}(1 + x) - \log_{10}(x) \text{ para } x \in \{1, 2, 3, 4, 5, 6, 7, 8, 9\}.$$

Concretamente la función de probabilidad de la ley Benford es (se muestran 4 decimales):

	Díg. 1	Díg. 2	Díg. 3	Díg. 4	Díg. 5	Díg. 6	Díg. 7	Díg. 8	Díg. 9
probs_Benford	0.301	0.1761	0.1249	0.0969	0.0792	0.0669	0.058	0.0512	0.0458

Su función de probabilidad y de distribución son

$$P_X(x) = P(X = x) = \begin{cases} 0.301 & \text{si } x = 1 \\ 0.1761 & \text{si } x = 2 \\ 0.1249 & \text{si } x = 3 \\ 0.0969 & \text{si } x = 4 \\ 0.0792 & \text{si } x = 5 \\ 0.0669 & \text{si } x = 6 \\ 0.058 & \text{si } x = 7 \\ 0.0512 & \text{si } x = 8 \\ 0.0458 & \text{si } x = 9 \\ 0 & \text{en otro caso.} \end{cases}$$

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & \text{si } x < 1 \\ 0.301 & \text{si } 1 \leq x < 2 \\ 0.4771 & \text{si } 2 \leq x < 3 \\ 0.6021 & \text{si } 3 \leq x < 4 \\ 0.699 & \text{si } 4 \leq x < 5 \\ 0.7782 & \text{si } 5 \leq x < 6 \\ 0.8451 & \text{si } 6 \leq x < 7 \\ 0.9031 & \text{si } 7 \leq x < 8 \\ 0.9542 & \text{si } 8 \leq x < 9 \\ 1 & \text{si } 9 \leq x \end{cases}$$

El valor esperado y la varianza son

$$E(X) = \sum_{x=1}^9 x \cdot \log_{10} \left( 1 + \frac{1}{x} \right) = 3.4402.$$

$$Var(X) = \sum_{x=1}^9 x^2 \cdot \log_{10} \left( 1 + \frac{1}{x} \right) - E(X)^2 = 17.8917 - E(X)^2 = 6.0565.$$

## El primer dígito de las potencias de 2 y de 5

Por ejemplo el primer dígito de las potencias de 2 y las de 5 sigue una ley Benford

## El primer dígito de las potencias de 2

Calculemos las potencia enteras de 2 de 1 a 200 ( la potencia máxima para R es  $2^{1023}$ ,  $2^{1024}$  da Inf.)

```
# Calculemos las potencia de 2 de 1 a 200
# podéis calcularlas hasta 2023.
potencias_muestra2=2^c(1:200)
# Las transformamos a character
d1=as.character(potencias_muestra2)
# Extraemos el primer dígito
d1=substr(d1,1,1)
# Calculamos las frecuencias absolutas de cada dígito
frec_absolutas_observadas=table(d1)
frec_absolutas_observadas
```

```
## d1
##  1  2  3  4  5  6  7  8  9
## 60 36 24 20 16 13 11 11  9
```

```
# Calculamos las teóricas según la ley de Benford
prob_teoricas=log10(1+1/c(1:9))
prob_teoricas
```

```
## [1] 0.30103000 0.17609126 0.12493874 0.09691001 0.07918125 0.06694679 0.05799195
## [8] 0.05115252 0.04575749
```

```
prop_observadas= prop.table(frec_absolutas_observadas)
prop_observadas
```

```
## d1
##      1      2      3      4      5      6      7      8      9
## 0.300 0.180 0.120 0.100 0.080 0.065 0.055 0.055 0.045
```

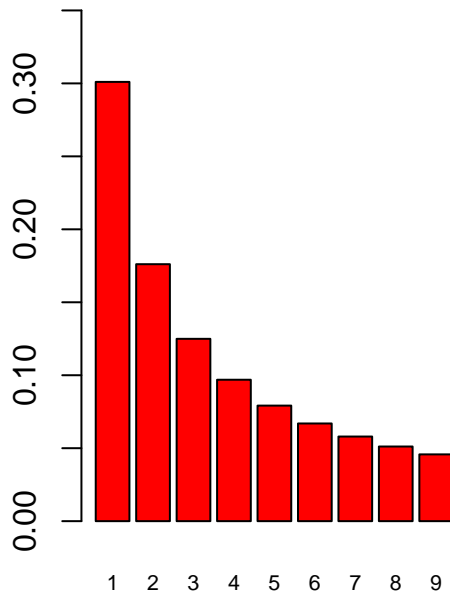
Guardemos todo en un data frame

```
dig_pot_2= data.frame(dig=1:9,
                      prob_teoricas=prob_teoricas,
                      prob_observadas=as.numeric(prop_observadas)
                      )
dig_pot_2$dif=dig_pot_2$prob_teoricas-dig_pot_2$prob_observadas
knitr::kable(dig_pot_2,digits = 4)
```

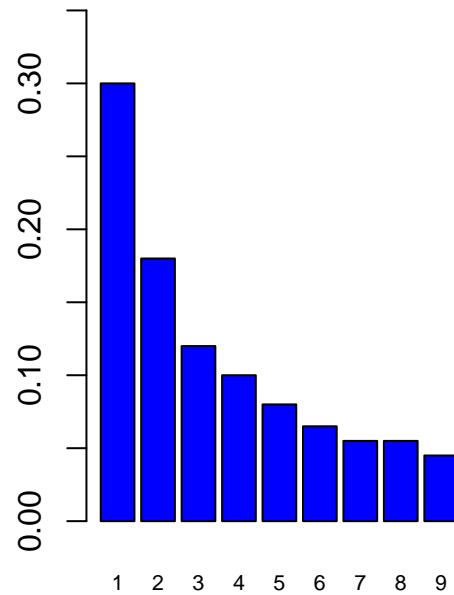
dig	prob_teoricas	prob_observadas	dif
1	0.3010	0.300	0.0010
2	0.1761	0.180	-0.0039
3	0.1249	0.120	0.0049
4	0.0969	0.100	-0.0031
5	0.0792	0.080	-0.0008
6	0.0669	0.065	0.0019
7	0.0580	0.055	0.0030
8	0.0512	0.055	-0.0038
9	0.0458	0.045	0.0008

```
par(mfrow=c(1,2))
bar1=dig_pot_2[,2]
names(bar1)=1:9
barplot(bar1,
        ylim=c(0,0.35),
        col=c("red"),
        width =0.2,
        main="Probabilidades observadas",
        cex.names=0.7)
bar2=dig_pot_2[,3]
names(bar2)=1:9
barplot(bar2,
        ylim=c(0,0.35),
        col=c("blue"),
        width=0.2,
        main="Probabilidades teóricas",
        cex.names=0.7)
```

**Probabilidades observadas**



**Probabilidades teóricas**



```
par(mfrow=c(1,1))
```

## El primer dígito de las potencias de 5

Para las potencias de 5 también podemos hacer lo mismo

```
# Calculamos las potencias de 5 de 1 a 200
potencias_muestra5=5^c(1:200)
# Las transformamos a character
d1=as.character(potencias_muestra5)
# Extraemos el primer dígito
d1=substr(d1,1,1)
# Calculamos las frecuencias absolutas de cada dígito
frec_absolutas_observadas=table(d1)
frec_absolutas_observadas
```

```
## d1
##  1  2  3  4  5  6  7  8  9
## 60 35 25 19 16 14 12 10  9
```

```
# Calculamos las teóricas según la ley de Benford
prob_teoricas=log10(1+1/c(1:9))
prob_teoricas
```

```
## [1] 0.30103000 0.17609126 0.12493874 0.09691001 0.07918125 0.06694679 0.05799195
## [8] 0.05115252 0.04575749
```

```
prop_observadas= prop.table(frec_absolutas_observadas)
```

Guardemos todo en un data frame

```
dig_pot_5= data.frame(dig=1:9,
                      prob_teoricas=prob_teoricas,
                      prob_observadas=as.numeric(prop_observadas)
                      )
knitr::kable(dig_pot_5,digits = 4)
```

dig	prob_teoricas	prob_observadas
1	0.3010	0.300
2	0.1761	0.175
3	0.1249	0.125
4	0.0969	0.095
5	0.0792	0.080
6	0.0669	0.070
7	0.0580	0.060
8	0.0512	0.050
9	0.0458	0.045

```
x=round(10*runif(1000,0.1,0.9),0)
table(x)
```

```
## x
##  1  2  3  4  5  6  7  8  9
## 66 137 107 138 126 124 120 116 66
```