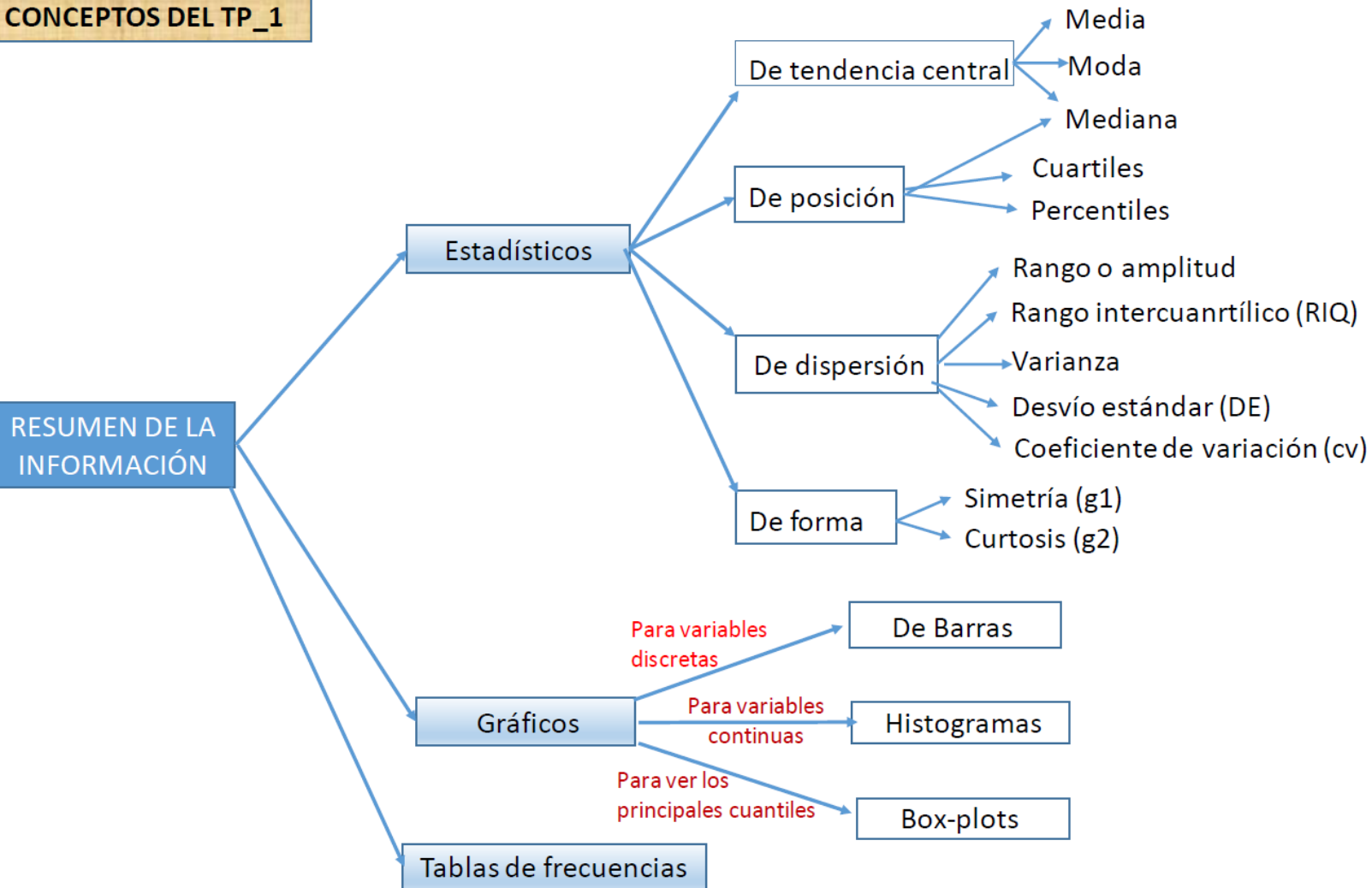


# TP 1 - Manejo de datos y estadística descriptiva

## CONCEPTOS DEL TP\_1



# TP 1 - Manejo de datos y estadística descriptiva

## Problema 3

La araucaria o pehuén (*Araucaria araucana*) es una especie arbórea endémica de los bosques del extremo noroeste de la Patagonia argentina y chilena, distribuida en zonas muy restringidas de la cordillera de los Andes. Es una especie de alto valor de conservación en ambos países y desde 2001 se encuentra incluida en el listado de CITES (Convención Internacional de Tráfico de Especies Silvestres), principalmente debido a la disminución de su área de distribución.

Se realizaron ensayos prospectivos en laboratorio con el fin de evaluar la capacidad germinativa de semillas colectadas en la localidad de Caviahue (Neuquén). Para ello se sembraron las semillas en bandejas con solo seis lugares (una semilla por lugar) usando un suelo de tipo arcilloso como sustrato. Al cabo de un tiempo **se tomó una muestra aleatoria de 50 bandejas** y se contó la cantidad de semillas germinadas por bandeja con los siguientes resultados:

1	3	2	0	5	1	5	2	5	3
2	4	3	1	2	3	4	0	2	3
6	3	4	3	6	4	3	2	1	4
4	3	3	3	1	5	4	3	2	5
0	6	5	3	6	3	2	4	3	4



## Problema 3. Araucaria

**3.1.-** Indicar la variable aleatoria y clasificarla, la unidad de análisis, la observación, la muestra, muestra estadística, la población y la población estadística.

**u.e.** Bandeja (con 6 semillas sembradas)

**V.a.** número de semillas germinadas por bandeja (Cuantitativa discreta)

**Observación** cada valor que toma la v.a. en cada u. e.

**Muestra** Conjunto de 50 bandejas (con 6 semillas sembradas)

**Muestra estadística** Conjunto de **los valores de la v.a.** de las 50 bandejas

**Población** Todas las bandejas del laboratorio (con 6 semillas sembradas/bandeja)

**Población estadística** **todos** los valores de la v.a. de las bandejas del laboratorio

**3.2.-** Construir una base de datos que permita analizar estas observaciones en R. Ayuda: En Excel, disponer los datos en una columna. Escribir un encabezado (p.ej. semillas). Ir a ‘Archivo’ y seleccionar la opción ‘Guardar como...’. En el cuadro desplegable seleccionar la opción: Texto (delimitado por tabulaciones) (con la que tendrá extensión .txt) o CSV (delimitado por comas), (con la que tendrá extensión .csv). Ambos formatos pueden ser leídos desde R.

```
Muestra$sem.germ
```

```
[1] 1 3 2 0 5 1 5 2 5 3 2 4 3 1 2 3 4 0 2 3 6 3 4 3 6 4 3 2 1 4 4 3 3 3 1 5 4 3 2  
5 0 6 5 3 6 3 2 4 3 4
```

**3.3.- Agrupar los datos en una tabla de frecuencias. ¿Qué número de semillas germinadas fue el más frecuente? ¿Qué nombre recibe dicho estadístico?**

```
table(muestra)
0 1 2 3 4 5 6
3 5 8 15 9 6 4
```

```
FA <- table(muestra)
tabla_frec_completa <- transform(FA, FAA = cumsum(Freq),
                                FR= prop.table(Freq), FRA = cumsum(prop.table(Freq)))
tabla_frec_completa
```

Consultar *Script\_TP1.r*

```
> tabla_frec_completa
  v.a. Freq FAA   FR   FRA
1    0   3    3 0.06 0.06
2    1   5    8 0.10 0.16
3    2   8   16 0.16 0.32
4    3  15   31 0.30 0.62
5    4   9   40 0.18 0.80
6    5   6   46 0.12 0.92
7    6   4   50 0.08 1.00
```



3.4.- ¿En qué porcentaje de bandejas germinaron a lo sumo 2 semillas? ¿Y en qué porcentaje germinó alguna semilla?

# Semillas germinadas por bandeja	> tabla_frec_completa				
	v.a.	Freq	FAA	FR	FRA
1	0	3	3	0.06	0.06
2	1	5	8	0.10	0.16
3	2	8	16	0.16	0.32
4	3	15	31	0.30	0.62
5	4	9	40	0.18	0.80
6	5	6	46	0.12	0.92
7	6	4	50	0.08	1.00

Respuesta  
32 %

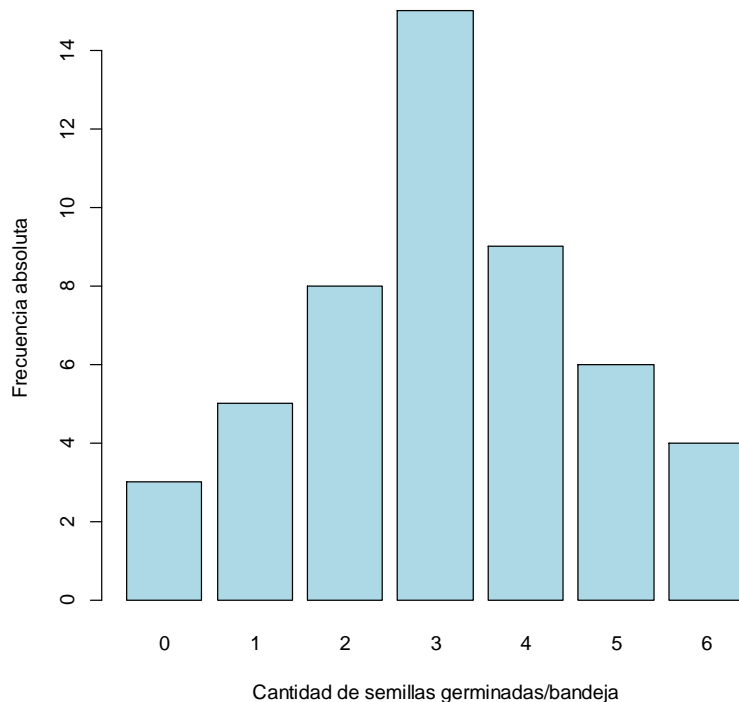
Germinó alguna semilla:  $1 - FR(va=0) = 1 - 0.06$

└──────────┘

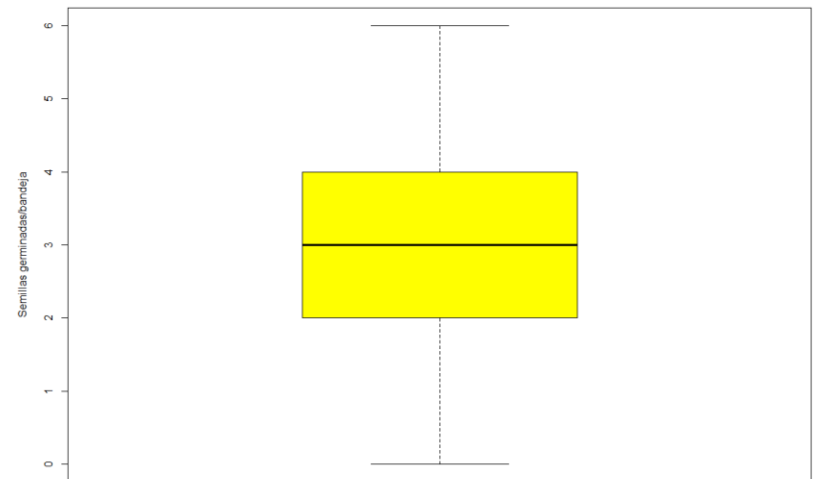
Respuesta  
94 %

**3.5.-** Señalar a través de un gráfico la forma de la distribución de esta variable (tipo de simetría). Luego, en un polígono de frecuencias acumuladas señalar el P75 e interpretar el valor obtenido.

```
barras_FA <- barplot(FA, ylab="Frecuencia absoluta", xlab="Cantidad de semillas germinadas/bandeja", col = "lightblue")
```



```
boxplot(muestra, ylab="Semillas germinadas/bandeja", col = "yellow")
```



```
summary(muestra$sem.ger)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00	2.00	3.00	3.12	4.00	6.00

## Polígono de frecuencias acumuladas y ubicar el percentil 75 (P75).

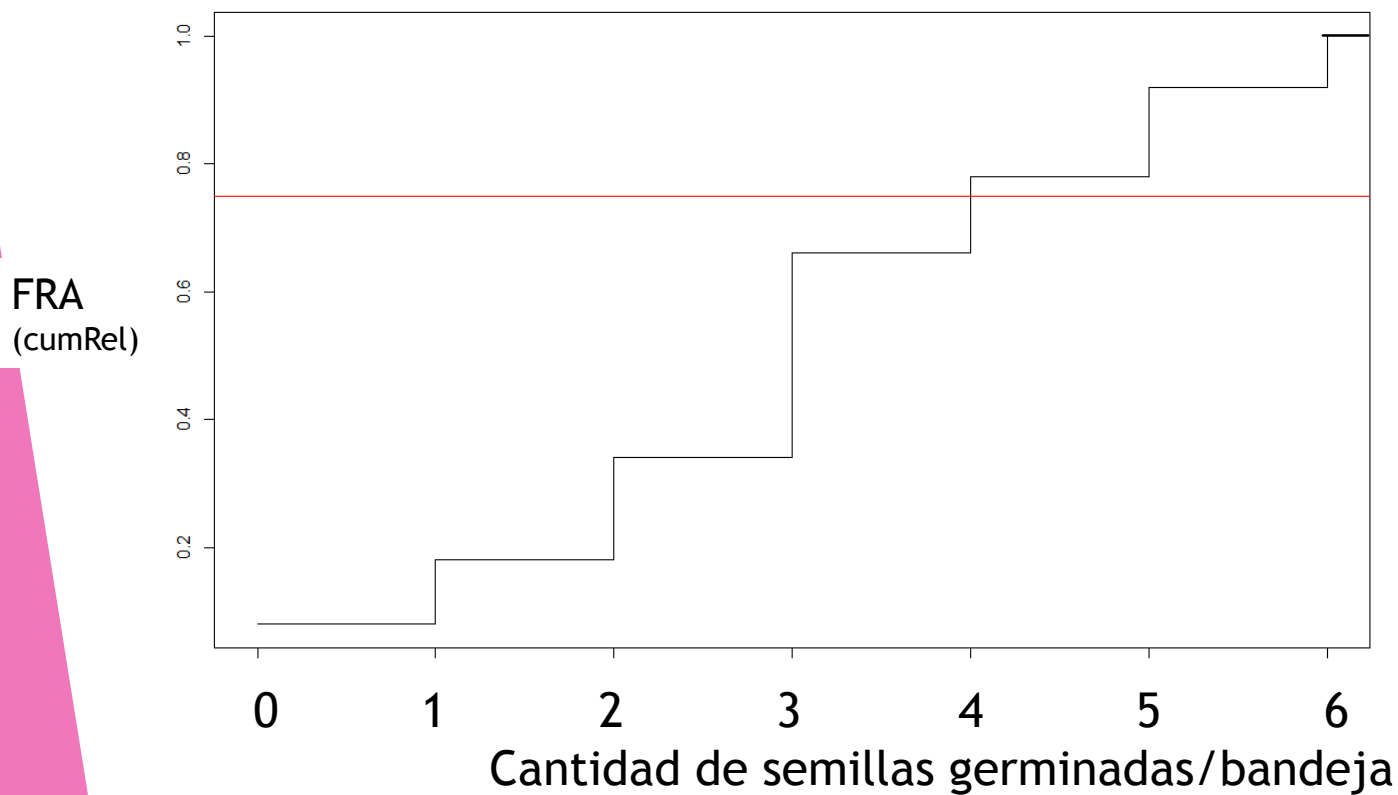
```
> tabla_frec_completa
```

v.a.	Freq	FAA	FR	FRA
0	3	3	0.06	0.06
1	5	8	0.10	0.16
2	8	16	0.16	0.32
3	15	31	0.30	0.62
4	9	40	0.18	0.80
5	6	46	0.12	0.92
6	4	50	0.08	1.00

```
plot(tabla_frec_completa$FRA, type = "s", ab = "Cantidad de  
semillas germinadas/bandeja", ylab = "FRA")
```

```
abline(h=0.75, col="red")
```

```
quantile(muestra, 0.75)
```





**3.6.-** ¿Cuál será el número medio de semillas germinadas por bandeja y su desvío estándar?

```
mean(muestra$semi.ger)
# [1] 3.12
sd (muestra$semi.ger))      semillas germinadas / bandeja
# [1] 1.586
```

**3.7.-** Indicar bajo qué valor de la variable se encuentra el 30% de las observaciones. ¿Qué nombre recibe ese estadístico?

```
quantile(muestra$semi.ger, 0.3)
```

2 semillas germinadas por bandeja