



Comparación de procesos

Cristóbal Honores

Resumen clase anterior

1. Estadísticos descriptivos.
2. Tabulación de resultados.
3. Control de calidad.
4. Test para detectar anomalías.



Estadísticos descriptivos

1. De centro:

1. Media
2. Mediana
3. Moda

2. De Posición:

1. Cuartiles
2. Percentiles

3. Variabilidad:

1. Rango
2. RIC
3. Desviación estándar
4. Varianza
5. Error estándar
6. CV



Diagrama de Pareto (qcc)

- El gráfico de Pareto es un gráfico de barras verticales en el que los valores se representan en orden decreciente de frecuencia relativa de izquierda a derecha.

```
pareto.chart(x, ylab = ,ylab2 =, xlab, cumperc, ylim, main, col  
= plot = TRUE, ...)
```



Diagrama de causa-efecto (qcc)

- Ayuda a identificar las posibles causas de un efecto, problema o resultado indeseable mientras las clasifica en categorías.

```
cause.and.effect(cause, effect, title =, font = c(1, 3, 2))
```

Grafico de dispersión

- Los gráficos de distribución ayudan a determinar si existe correlación entre los conjuntos de datos

```
plot(x, y, main, xlab, ylab, pch, col, )
```

```
ggplot(data, aes(x=..., y=...)) + geom_point()
```

```
plot_ly(data = ..., x = ~..., y = ~..., type = "scatter")
```

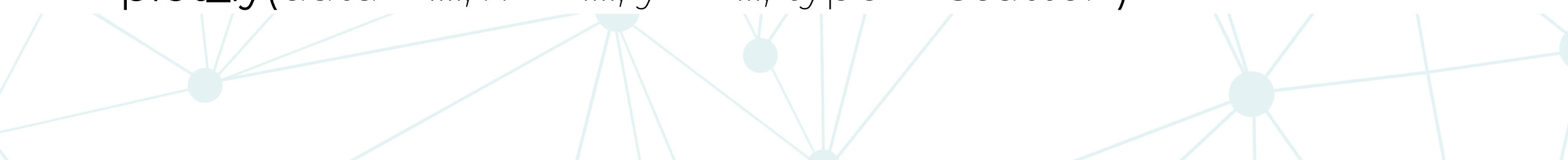


Gráfico de caja

- **Permite visualizar la distribución de los datos a través de los cuartiles.**

```
boxplot(x, horiz=..., width=..., pch=..., ....)
```

```
ggplot(data, aes(x=..., y=...)) + geom_boxplot ()
```

```
plot_ly(data = ..., x = ~..., y = ~..., type = "box")
```

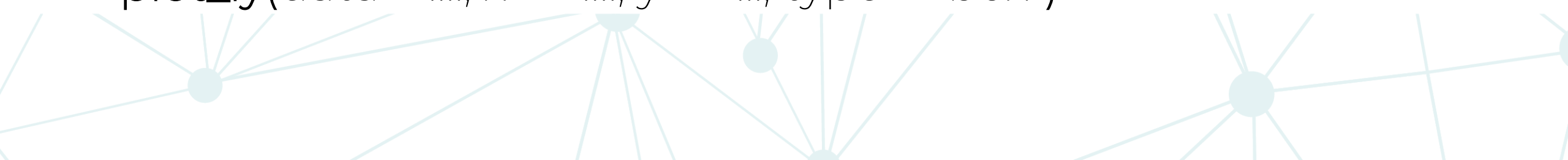


Gráfico de violín

- Muestra la distribución de los datos atreves del grafico de cajas y del grafico de densidad.

```
Vioplot(x, ylim=..., xlim=..., col..., ...)
```

```
ggplot(data, aes(x=..., y=...)) + geom_violin () + geom_boxplot  
()
```

```
plot_ly(data = ..., x = ~..., y = ~..., type = "violin")
```



Grafico de control (qcc)

- Muestran si las muestras de productos o procesos cumplen con las especificaciones previstas y, de no ser así, el grado en que varían de esas especificaciones.

`qcc(data, type, newdata, plot,)`



Test Rosner (EnvStats)

- **Paquete: EnvStats**
- Realiza la prueba de Rosner para hasta k valores atípicos potenciales en un conjunto de datos, suponiendo que los datos sin valores atípicos provienen de una distribución normal (gaussiana).

```
rosnerTest(x, k = 3, alpha = 0.05, warn = TRUE)
```



Tabulación de resultados

- Las tablas de frecuencias se utilizan con el con el comando:

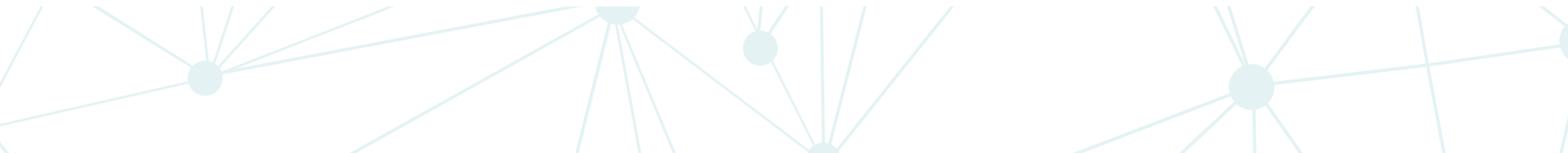
`table(base$..... , base$.....)`

- `addmargins()` se utiliza para agregar los totales por filas o por columnas a una tabla de frecuencia.
- `prop.table()` se utiliza para tablas de frecuencia relativa a partir de tablas de frecuencia absoluta.




Propósito de la clase

Aprender técnicas para ser capaz de comparar entre dos o mas procesos y reconocer si existe una diferencia significativa entre ellos.



Comparación de procesos

1. Estimación de parámetros.
 2. Test de hipótesis.
 3. Intervalos de confianza.
 4. Comparación de dos muestras.
 5. Comparación de varias muestras..
- 
- A decorative network diagram at the bottom of the slide, consisting of several light blue circular nodes connected by thin, light blue lines, forming a web-like structure.

Test de hipótesis

- Una prueba de hipótesis es un método estadístico inferencial para la toma de decisiones sobre una población en base a la información proporcionada por los datos de una muestra.
- Se plantean dos hipótesis:
 - $H_0 \leftarrow$ nula
 - $H_1 \leftarrow$ hipótesis alternativa

Planteamientos de hipótesis

- Test de igualdad:

$$H_0: \mu = \mu_0$$

$$H_1: \mu \neq \mu_0$$

“two.sided”

- Test de mayor que:

$$H_0: \mu = \mu_0$$

$$H_1: \mu < \mu_0$$

“greater”

- Test de menor que:

$$H_0: \mu = \mu_0$$

$$H_1: \mu > \mu_0$$

“less”

Tipos de errores

- Tanto la aceptación como el rechazo de H_0 dependen de los resultados de evidencia empírica muestral, lo cual puede estar sujeta a posibles errores asociados.

Error tipo I: $P(\text{rechazo } H_0 | H_0 \text{ es verdadera})$

Error tipo II: $P(\text{no rechazo } H_0 | H_0 \text{ es falsa})$

La probabilidad real de cometer Error Tipo I se conoce como valor-p.

Valor p

El valor p nos entrega una probabilidad de que los resultados obtenidos provengan de una muestra de población donde se cumple H_0

Test para la media

- `t.test(x, y = NULL, alternative = c("two.sided", "less", "greater"), mu = 0, paired = FALSE, var.equal = FALSE, conf.level = 0.95, ...)`
- `x`: vector numérico con los datos.
- `alternative`: tipo de hipótesis alterna. Los valores disponibles son "two.sided" cuando la hipótesis alterna es \neq , "less" para el caso $<$ y "greater" para $>$.
- `mu`: valor de referencia de la prueba.
- `conf.level`: nivel de confianza para reportar el intervalo de confianza asociado (opcional).

Test para la media

- **Varianza desconocida:**

```
library(TeachingDemos)
```

```
z.test(X,sd = sigma, alternative = "two.sided",  
conf.level = 0.95,mu=...)
```

Test Chi-Cuadrado

- `prop.test(x, n, p = NULL, alternative = c("two.sided", "less", "greater"), conf.level = 0.95, correct = TRUE)`
- `x`: número de éxitos en la muestra.
- `n`: número de observaciones en la muestra.
- `alternative`: tipo de hipótesis alterna. Los valores disponibles son "two.sided" cuando la alterna es \neq , "less" para el caso $<$ y "greater" para $>$.
- `p`: valor de referencia de la prueba.
- `correct`: valor lógico para indicar si se usa la corrección de Yates.
- `conf.level`: nivel de confianza para reportar el intervalo de confianza asociado (opcional).

Test para una varianza

- `sigma.test(x, alternative = "two.sided", null.value = 1, conf.level = 0.95)`
- `x`: vector numérico con los datos.
- `alternative`: tipo de hipótesis alterna. Los valores disponibles son "two.sided" cuando la alterna es \neq , "less" para el caso $<$ y "greater" para $>$.
- `null.value`: valor de referencia de la prueba.
- `conf.level`: nivel de confianza para reportar el intervalo de confianza asociado (opcional).

Intervalos de confianza

- Cuando inferimos no tenemos garantía de que la conclusión que obtenemos sea exactamente la correcta.
- Sin embargo, es posible cuantificar el error asociado a la estimación.

IC en R

Del test de hipótesis que se realizó previamente:

- Si se selecciona la variable `conf.int` entrega el valor del intervalo de confianza con el nivel de confianza deseado.
- De querer solo el valor número se coloca delante de este la función `as.numeric()`

```
as.numeric(TEST$conf.int)
```

IC en R

Ojo!!! El IC solo te obtiene cuando el test de hipótesis se realiza “two.sided”

Comparación de dos muestras

- Test de igualdad:

$$H_0: \mu_A - \mu_B = \mu_0$$

$$H_1: \mu_A - \mu_B \neq \mu_0$$

- Test de mayor que:

$$H_0: \mu_A - \mu_B = \mu_0$$

$$H_1: \mu_A - \mu_B < \mu_0$$

- Test de menor que:

$$H_0: \mu_A - \mu_B = \mu_0$$

$$H_1: \mu_A - \mu_B > \mu_0$$

Comparación de 2 varianzas

- `var.test(x, y, alternative = "two.sided", null.value = 1, conf.level = 0.95)`
- `x`: vector numérico con los datos.
- `alternative`: tipo de hipótesis alterna. Los valores disponibles son "two.sided" cuando la alterna es \neq , "less" para el caso $<$ y "greater" para $>$.
- `null.value`: valor de referencia de la prueba.
- `conf.level`: nivel de confianza para reportar el intervalo de confianza asociado (opcional).

Comparación 2 medias

Es necesario hacer un test de varianzas antes de hacer comparación de medias.

- Varianzas iguales:

```
t.test(x=T1, y=T2, alternative="two.sided", mu=0, paired =  
FALSE, var.equal=TRUE, conf.level=0.97)
```

- Varianzas distintas:

```
t.test(x=T1, y=T2, alternative="two.sided", mu=0, paired =  
FALSE, var.equal=FALSE, conf.level=0.97)
```

Comparación 2 medias

Medias pareadas:

En este caso, se realizan dos mediciones a una misma muestra, ambas medias obtenidas son no independientes.

- Medias pareadas:

```
t.test(x=T1, y=T2, alternative="two.sided", mu=0, paired =  
TRUE, var.equal=FALSE, conf.level=0.97)
```

Comparación 2 medias

- Al igual que en test de hipótesis para la media, se puede seleccionar si es mayor o menor solo cambiando en el comando `alternative`.
- De igual manera se puede ver si una muestras es menor o mayor a la otra en un numero especifico cambiando el valor de μ .

Comparación de Proporciones

- Para realizar el test de proporciones, se utiliza el mismo comando `prop.test()` de igual manera que para una sola proporción.
- Sin embargo, a la hora de colocar los éxitos y el numero de casos totales, se coloca como vector.

$$X = C(E_1, E_2) ; n = c(n_1, n_2)$$

Test de Normalidad

Con el fin de comprobar la normalidad de los datos se realizan los siguientes test:

1. Q-q plot

`qqnorm(datos) ; qqline(datos)`

2. Kolmogorov-Smirnov

`ks.test(datos)`

3. Shapiro-Wilk

`shapiro.test(datos)`

OJO! Shapiro-Wilk solo puede realizarse con tamaños de muestra menores a 5000.

Comparación varias medias

- La comparación de varias medias tiene como objetivo ver si todas las medias son iguales, o existe a lo menos una distinta.

$$H_0: \mu_A = \mu_B = \mu_C = \mu_D = \dots$$

$$H_1: \text{A lo menos una distinta}$$

- Existen distintos test para comprobar esto dependiendo de si los datos tienen una distribución normal o no.

Análisis de la varianza. (ANOVA)

- Esta prueba estadística es de las más utilizadas para poder comparar más de dos muestras poblacionales, las cuales deben cumplir los siguientes requisitos:
 - Que las variables sean independientes
 - Que tengan una distribución normalizada
 - Que sus varianzas no difieran significativamente.

```
Anova <- aov(valores ~ categoría)
summary(Anova)
plot(TukeyHSD(Anova))
```

Análisis de Kruskal-Wallis

- Es similar al ANOVA pero cuando se cuentan con datos que nos siguen una distribución normal.

```
kruskal.test(valores ~ categoría, data = ...)
```

```
pairwise.wilcox.test(valores ~ categoría, data = ...,  
                     p.adjust.method = "holm" )
```



Comparación de procesos

Cristóbal Honores