Unidad 2: Estadística Basica y Aplicada

Inferencia estadistica

Nicolas Sidicaro

Octubre 2025

El para que

Caso: queremos saber el ingreso promedio de los argentinos. Pero no podemos preguntarle a todos. Entonces tomamos una muestra (EPH)... ¿como llevamos las conclusiones de la muestra a la poblacion?

Con la inferencia estadística

Conceptos: población, muestra, parámetro, estadístico

En el caso: los trabajadores argentinos; la EPH; el ingreso promedio de la poblacion (μ); el ingreso promedio de la muestra ($\overline{x}$)

¿Como estudiamos la muestra?

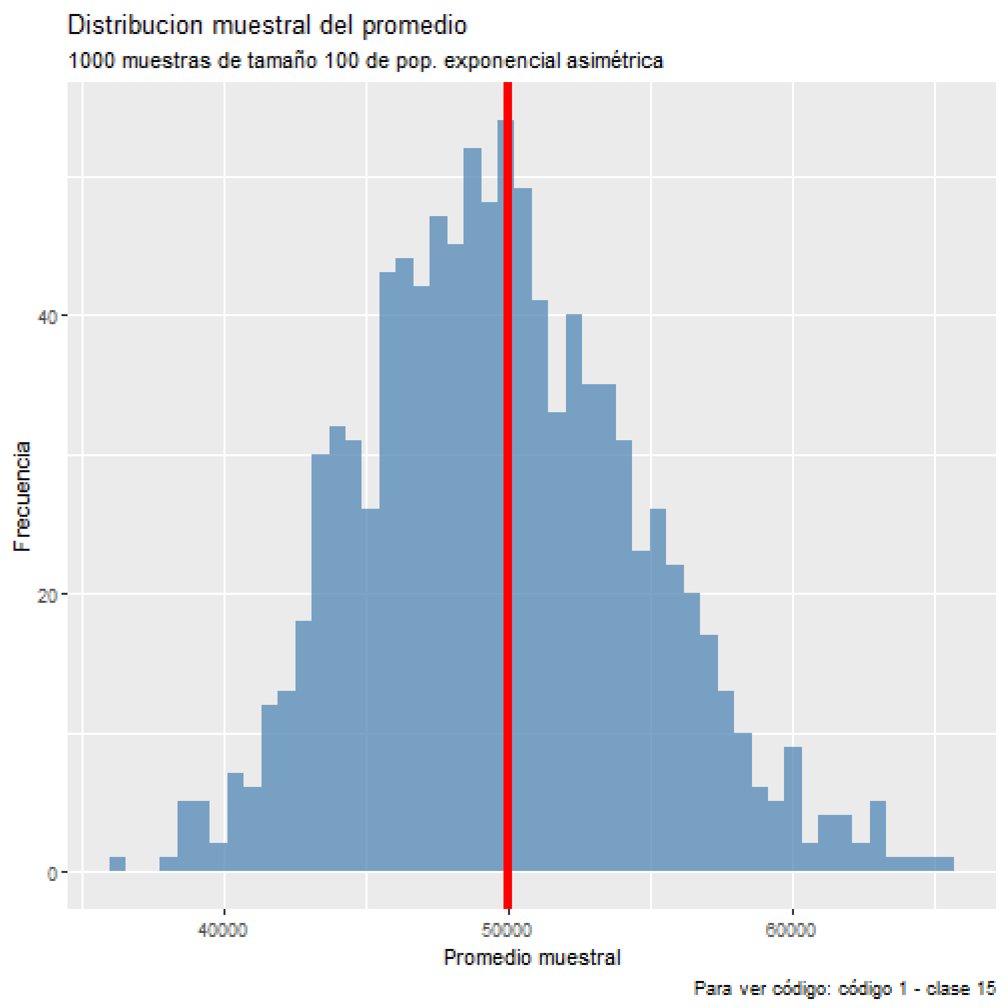
Tomemos 1000 muestras de nuestra población de tamaño 100.

1. Cada muestra tendrá un promedio diferente
2. Esos promedios forman una **distribucion muestral**
3. Esta distribucion tiene propiedades predecibles

¿Gracias a quién? Al Teorema Central del Límite

- Para muestras suficientemente grandes $\overline{x}$ es aproximadamente normal
- Con media = μ
- Con error estandar = $\frac{\sigma}{\sqrt{n}}$

Dada una poblacion exponencial



Intervalos de Confianza

¿Qué es un intervalo de confianza?

Definición: Rango de valores plausibles para un parámetro poblacional.

Interpretación correcta (95% de confianza):

- Si repetimos el muestreo infinitas veces, el 95% de los intervalos calculados contendrán el verdadero parámetro

Interpretación INCORRECTA:

✗ "Hay 95% de probabilidad de que μ esté en este intervalo" ✗ "El 95% de los datos están en este intervalo"

Fórmula general para desvío conocido

$$IC = \bar{x} \pm z * \frac{\sigma}{\sqrt{n}}$$

Factores que afectan el ancho del IC

```r
# Simulación: Impacto del tamaño muestral
tamaños <- c(10, 30, 100, 500)
resultados <- map_df(tamaños, function(n) {
  muestra <- rnorm(n, mean = 50, sd = 15)
  ic <- t.test(muestra)$conf.int
  tibble(
    n = n,
    ancho_ic = ic[2] - ic[1]
  )
})

print(resultados)
```

```
## # A tibble: 4 × 2
##       n ancho_ic
##   <dbl>    <dbl>
## 1    10    31.8
## 2    30    12.5
## 3   100     6.10
## 4   500     2.66
```

Conclusión: A mayor n, menor ancho del IC (más precisión).

Test de hipótesis

Caso: Decimos que el salario promedio de Argentina es de 800 mil pesos, pero al tomar una muestra representantiva nos da 700 mil ¿la diferencia es por azar o hay evidencia suficiente para rechazar la afirmación inicial?

Componentes del test

1. Hipótesis Nula (H_0): Lo que asumimos verdadero hasta probar lo contrario
2. Hipótesis Alternativa (H_1): Lo que queremos probar
3. Estadístico de prueba: Número que resume la evidencia
4. Nivel de significancia (α): Tolerancia al error (típicamente 0.05)
5. P-valor: Probabilidad de obtener nuestros datos si H_0 es cierta
6. Decisión: Rechazar o no rechazar H_0

Tipos de error

	H_0 **es verdadera**	H_0 **es falsa**
No rechazamos H_0	Decisión correcta	Error tipo II (β)
Rechazamos H_0	Error tipo I (α)	Decisión correcta (Poder)

P-value: interpretación correcta

¿Qué es el p-value?

Probabilidad de observar un estadístico tan extremo (o más) como el observado, asumiendo que H_0 es verdadera.

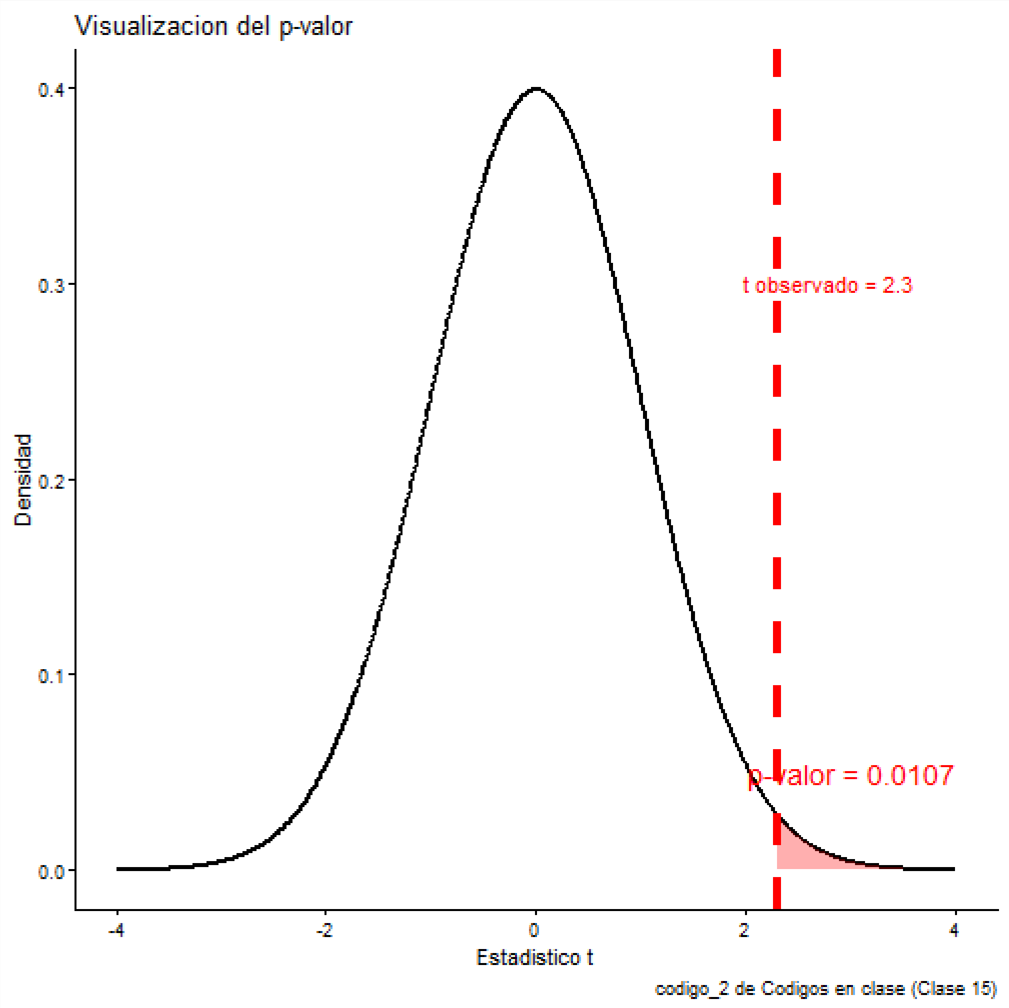
Es decir, si H_0 fuera cierta ¿qué tan raro sería obtener estos datos?

Nos ayuda a decidir sin ver los estadísticos

- Si $p - value < \alpha$ => se rechaza H_0
- Si $p - value > \alpha$ => No se rechaza H_0

Siendo un α = {0.1,0.05,0.01}

Visualizacion de p-value



Supuestos de los tests paramétricos

Antes de aplicar tests como el t-test, debemos verificar:

1. **Normalidad**: Los datos (o residuos) siguen una distribución normal
2. **Homogeneidad de varianzas**: Las varianzas son similares entre grupos
3. **Independencia**: Las observaciones son independientes

¿Por qué importa?

- Si se violan estos supuestos, los p-valores pueden ser incorrectos
- Existen alternativas no paramétricas cuando no se cumplen

Herramientas para verificar normalidad:

- Test de Shapiro-Wilk
- Gráficos Q-Q

El rol del TCL

El Teorema Central del Límite nos protege… pero no siempre

Tamaño muestral	**¿Verificar normalidad?**	**Razón**
n < 30	✅ **Sí, siempre**	TCL no aplica efectivamente
30 ≤ n < 100	⚠️ **Considerar**	TCL aplica, pero puede ser débil con asimetrías severas
n ≥ 100	✗ **Generalmente NO**	TCL garantiza normalidad de $\overline{x}$

Matiz importante:

- El t-test evalúa la **media muestral** ($\overline{x}$), no los datos individuales
- Con n grande, $\overline{x}$ es normal **incluso si los datos no lo son** (gracias al TCL)
- Con n pequeño, necesitamos que los **datos originales** sean aproximadamente normales

Test de Shapiro-Wilk

¿Qué evalúa?

Contrasta si una muestra proviene de una distribución normal.

- H_0: Los datos provienen de una distribución normal
- H_1: Los datos NO provienen de una distribución normal

¿Cuándo es realmente necesario?

- **Crucial** para muestras pequeñas (n < 30)
- **Opcional** para muestras medianas (30-100)
- **Menos relevante** para muestras grandes (n > 100) por el TCL

Test de Shapiro-Wilk

```r
# Datos que parecen normales
datos_normales <- rnorm(50, mean = 100, sd = 15)
shapiro.test(datos_normales)
```

```
##
##      Shapiro-Wilk normality test
##
## data:  datos_normales
## W = 0.98504, p-value = 0.7733
```

```r
# Datos claramente NO normales (exponencial)
datos_exponencial <- rexp(50, rate = 0.1)
shapiro.test(datos_exponencial)
```

```
##
##      Shapiro-Wilk normality test
##
## data:  datos_exponencial
## W = 0.84853, p-value = 0.00001412
```

Limitaciones del Test de Shapiro

Problema 1: Sensible al tamaño muestral

- Con **muestras pequeñas** ($n < 30$): Poco poder, puede no detectar desviaciones importantes
- Con **muestras grandes** ($n > 200$): Demasiado sensible, rechaza H_0 por desviaciones triviales que **no afectan** al t-test (el TCL ya protege)

```r
# Muestra grande de datos casi normales
set.seed(123)
datos_grandes <- rnorm(1000, mean = 50, sd = 10)
datos_grandes[1:10] <- datos_grandes[1:10] + 5  # Pequeña perturbación
shapiro.test(datos_grandes)  # Probablemente rechace H0
# Pero el t-test seguirá siendo válido por el TCL
```

Limitaciones del Test de Shapiro

Problema 2: No es suficiente por sí solo

- Un p-valor > 0.05 NO garantiza normalidad perfecta
- Siempre complementar con visualización (Q-Q plot)

Recomendación: Shapiro + Q-Q plot + considerar el tamaño muestral (TCL)

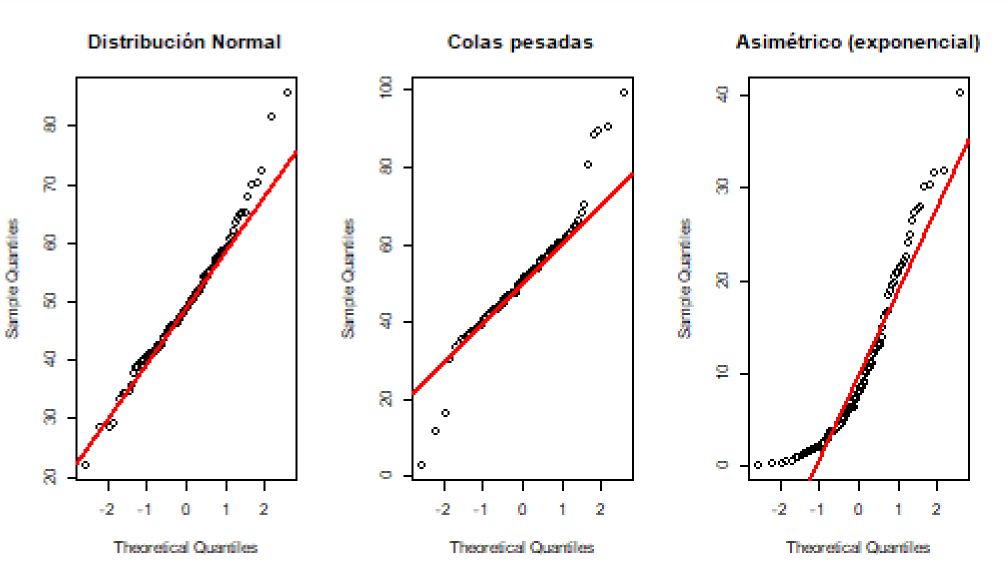
Gráfico Q-Q (Quantile-Quantile)

¿Qué muestra?

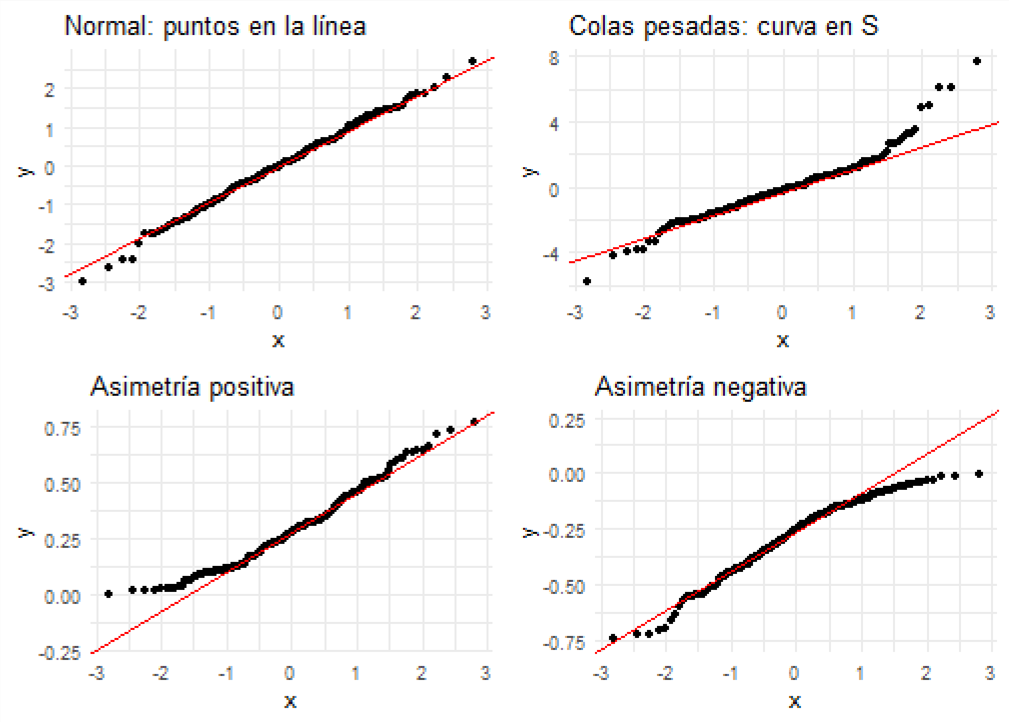
Compara los cuantiles de nuestros datos vs. los cuantiles teóricos de una distribución normal.

Interpretación:

- **Puntos sobre la línea**: Los datos son aproximadamente normales
- **Puntos se desvían sistemáticamente**: Desviación de normalidad



Interpretación de patrones en Q-Q plot



Test contra un valor

Test t para una muestra

Pregunta: ¿El promedio poblacional es igual a un valor específico?

Ejemplo: ¿El salario promedio es $50,000?

```
salarios ← c(48, 52, 45, 55, 49, 51, 47, 53, 46, 54)
# H0: μ = 50 (mil pesos) # H1: μ ≠ 50 (test de dos colas)
resultado ← t.test(salarios, mu = 50)
print(resultado)
```

```
##
##      One Sample t-test
##
## data:  salarios
## t = 0, df = 9, p-value = 1
## alternative hypothesis: true mean is not equal to 50
## 95 percent confidence interval:
##  47.49909 52.50091
## sample estimates:
## mean of x
##        50
```

Test de homogeneidad de varianzas

¿Por qué importa?

El t-test clásico (Student) combina las varianzas de ambos grupos en un único error estándar:

$$SE_{pooled} = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Problema: Esta fórmula asume $\boldsymbol{s_1^2 \approx s_2^2}$

Si las varianzas son MUY diferentes:

- El error estándar está mal calculado
- El p-valor es incorrecto (más errores tipo I o menos poder)

Test de Levene: verifica homogeneidad

Test de Levene (más robusto que el F-test):

- H_0: Las varianzas son homogéneas (iguales)
- H_1: Las varianzas son diferentes

¿Cómo funciona?

1. Calcula desviaciones absolutas de cada dato respecto a la mediana de su grupo
2. Aplica ANOVA sobre esas desviaciones
3. Usa la **mediana** $\rightarrow$ robusto a outliers (ventaja vs F-test)

Test de Levene: verifica homogeneidad

```r
library(car)  # Para test de Levene

# Crear dos grupos con varianzas diferentes
grupo_a <- rnorm(50, mean = 50, sd = 5)   # Varianza pequeña
grupo_b <- rnorm(50, mean = 52, sd = 15)  # Varianza grande

# Test de Levene
leveneTest(c(grupo_a, grupo_b) ~ rep(c("A", "B"), each = 50))
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.

## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value        Pr(>F)
## group  1    33.6 0.00000008247 ***
##       98
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interpretando Levene y eligiendo el test

Si p-valor > 0.05 (NO rechazamos H_0):

- ✅ Las varianzas son similares
- Puedes usar t-test clásico con `var.equal = TRUE` (más poder)

Si p-valor < 0.05 (Rechazamos H_0):

- ❌ Las varianzas son diferentes
- **Solución**: Usar corrección de Welch con `var.equal = FALSE`

```
# Test t con corrección de Welch (por defecto en R)
t.test(grupo_a, grupo_b, var.equal = FALSE)  # ← DEFAULT en R

# Test t clásico (solo si Levene confirma varianzas iguales)
t.test(grupo_a, grupo_b, var.equal = TRUE)
```

Buena noticia: R usa Welch por defecto

¿Qué hace la corrección de Welch?

En lugar de combinar varianzas, calcula errores estándar **separados**:

T-test clásico (Student):

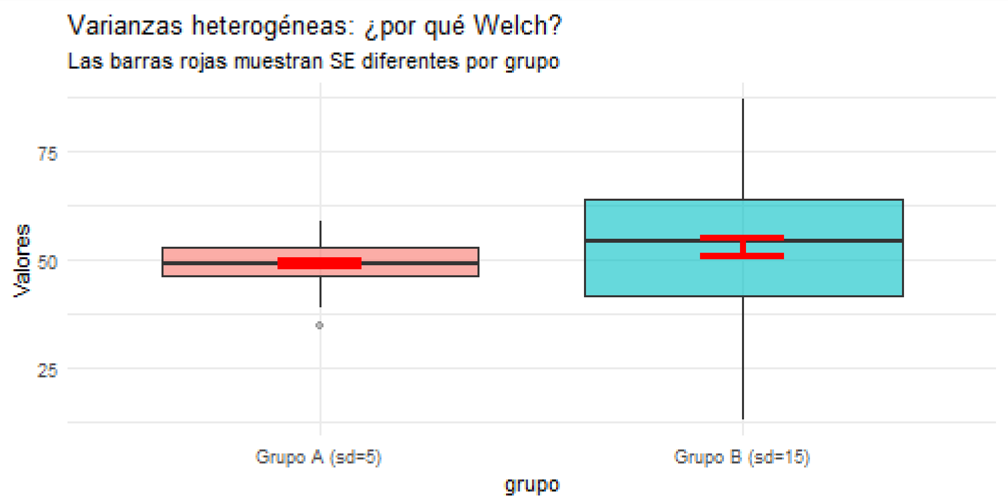
- Un solo SE para ambos grupos → asume $s_1 = s_2$

T-test con Welch:

- SE separado para cada grupo → permite $s_1 \neq s_2$
- Ajusta los grados de libertad automáticamente

¿Qué hace la corrección de Welch?

Comparación visual:



Levene vs F-test: ¿cuál usar?

Aspecto	Test F (var.test)	Test de Levene
Supuesto	Normalidad estricta	Menos exigente
Outliers	Muy sensible	Más robusto (usa mediana)
Uso recomendado	Solo si datos muy normales	En la mayoría de casos

```r
# Test F (más antiguo, más restrictivo)
var.test(grupo_a, grupo_b)

# Test de Levene (preferido actualmente)
leveneTest(c(grupo_a, grupo_b) ~ rep(c("A", "B"), each = 50))
```

Regla práctica:

1. Siempre verifica homogeneidad cuando compares grupos
2. Si Levene rechaza H_0 $\rightarrow$ usa Welch (`var.equal = FALSE`)
3. En R, como Welch es el default, ¡estás protegido automáticamente!

Dos muestras independientes

Pregunta: ¿Hay diferencia entre dos grupos?

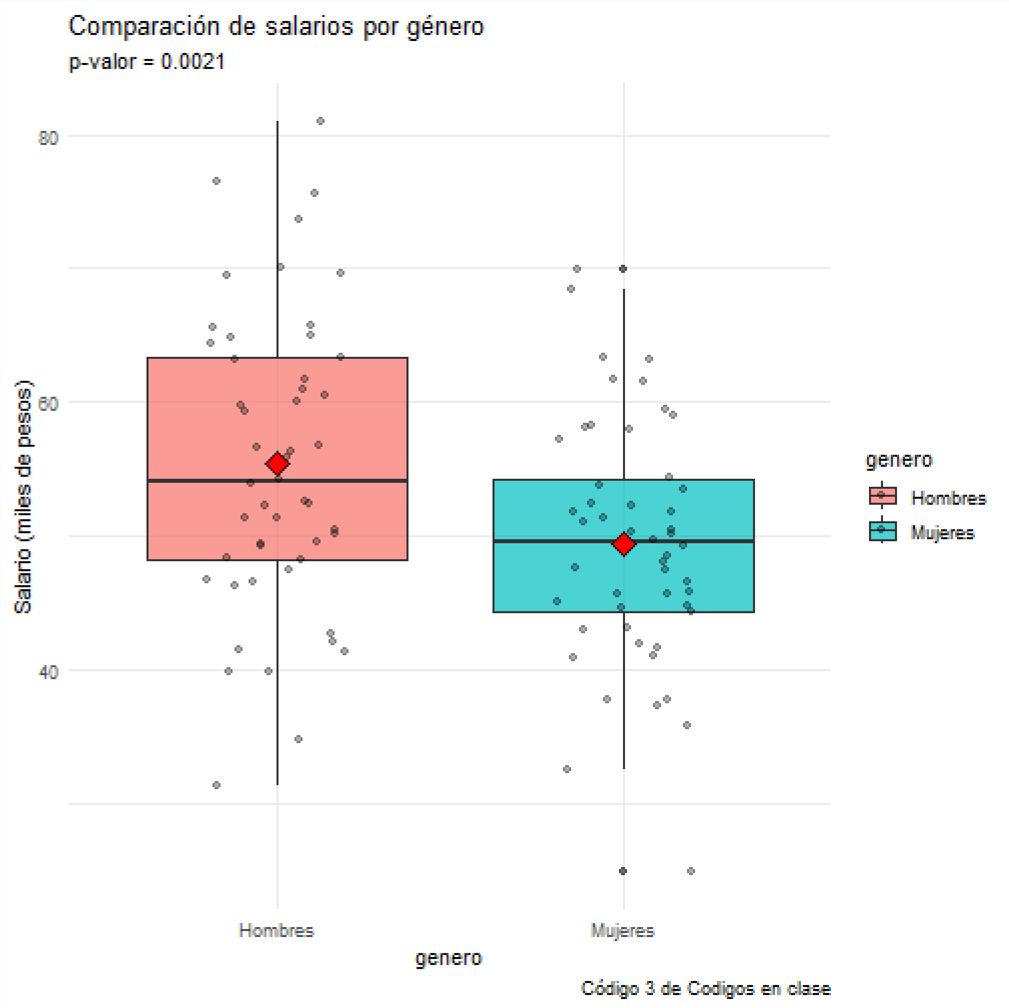
Ejemplo: ¿Los hombres ganan más que las mujeres?

```r
set.seed(123)
salarios_hombres <- rnorm(50, mean = 55, sd = 12)
salarios_mujeres <- rnorm(50, mean = 48, sd = 10)
# H0: μ_hombres = μ_mujeres (no hay diferencia)
# H1: μ_hombres > μ_mujeres (test de una cola)
resultado <- t.test(salarios_hombres, salarios_mujeres,
                    alternative = "greater")
print(resultado)
```

Dos muestras independientes

```
##
##      Welch Two Sample t-test
##
## data:  salarios_hombres and salarios_mujeres
## t = 2.9348, df = 94.165, p-value = 0.002096
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  2.581605      Inf
## sample estimates:
## mean of x mean of y
##  55.41284  49.46408
```

Dos muestras independientes



Wilcoxon: no paramétrica

¿Cuándo usar Test de Wilcoxon en lugar de t-test?

1. Los datos NO son normales (y n es pequeño)
2. Hay outliers extremos
3. Los datos son ordinales (ej: escalas de Likert)

Diferencias clave con el t-test:

Aspecto	**Test T**	**Test de Wilcoxon**
Supuesto	Normalidad	No requiere normalidad
Compara	Medias	Medianas/rangos
Sensible a	Outliers	Más robusto
Poder	Mayor (si supuestos OK)	Menor (pero más seguro)

Ejemplo comparativo: T-test vs Wilcoxon

```r
# Datos con outliers extremos
set.seed(456)
grupo_control <- c(rnorm(20, mean = 100, sd = 10), 250)  # Un outlier
grupo_tratamiento <- rnorm(21, mean = 110, sd = 10)

# Test T (sensible a outliers)
t.test(grupo_tratamiento, grupo_control)$p.value
```

```
## [1] 0.6004376
```

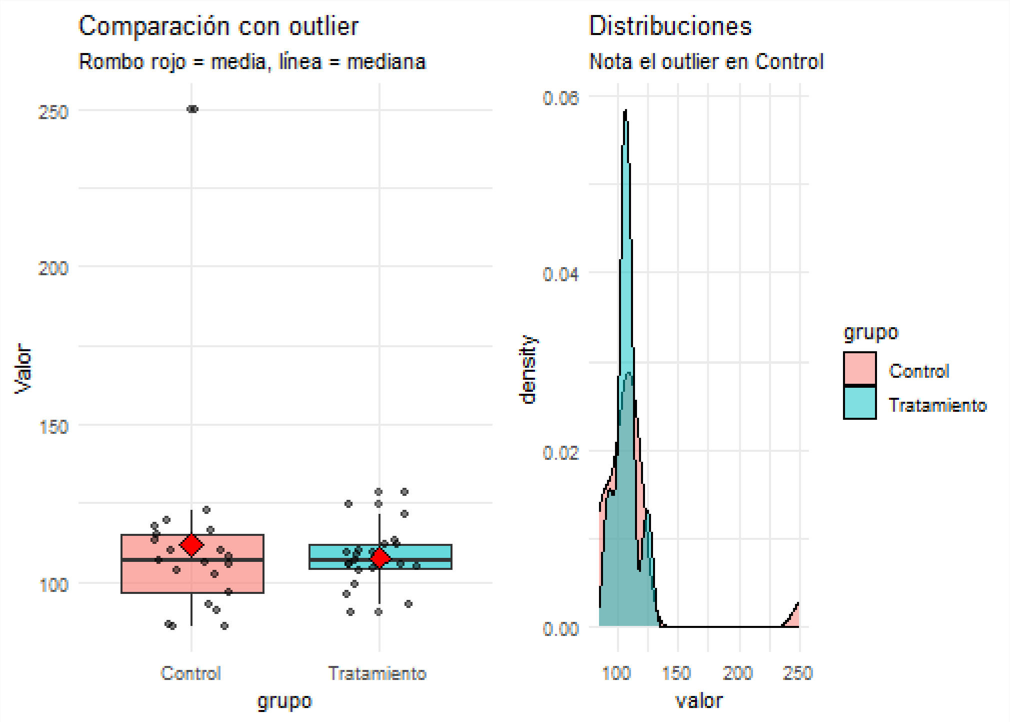
```r
# Test de Wilcoxon (robusto a outliers)
wilcox.test(grupo_tratamiento, grupo_control)$p.value
```

```
## [1] 0.9207121
```

Interpretación:

- El t-test puede verse afectado por el outlier (250)
- Wilcoxon se basa en rangos, más robusto

Visualización: T-test vs Wilcoxon



Test pareado

- ¿Qué es?
    - Una prueba estadística para comparar las medias de dos grupos *relacionados* o *dependientes*.
    - Evalúa si la diferencia promedio entre las dos observaciones pareadas es significativamente distinta de cero.
    - Los datos provienen de los **mismos sujetos**, medidos en dos momentos o condiciones diferentes.
- Aplicaciones comunes
    - **"Antes y Después":** Medir el rendimiento de un grupo de empleados antes y después de un curso de capacitación.
    - **Comparación de tratamientos:** Probar la efectividad de una política pública en empresas similares

Test pareado

- Hipótesis
  - H_0: La diferencia media entre los pares es cero. (No hay efecto)
  - H_1: La diferencia media entre los pares no es cero. (Hay un efecto)
- Cómo funciona
  1. Se calculan las diferencias individuales para cada par.
  2. Se aplica un t-test sobre estas diferencias.
- Conclusión
  - Si el **p-valor < 0.05**:
    - El resultado es estadísticamente significativo.
    - Se rechaza la hipótesis nula.
    - Existe una diferencia real entre las dos mediciones.

Ejemplo: en codigo_4 de Códigos en clase (Clase 15)

Test pareado de Wilcoxon

Cuando los datos pareados NO son normales, usamos el test de rangos con signo de Wilcoxon:

Ventaja: Más apropiado para escalas ordinales o datos con distribuciones asimétricas.

```
# Ejemplo: satisfacción antes y después de capacitación (escala 1-10, datos ordinales)
antes <- c(5, 6, 4, 7, 5, 6, 4, 5, 6, 7)
despues <- c(7, 8, 6, 8, 7, 7, 5, 6, 8, 9)
# Test t pareado (asume normalidad)
t.test(despues, antes, paired = TRUE)$p.value
```

```
## [1] 0.000004239739
```

```
# Test de Wilcoxon pareado (no asume normalidad)
wilcox.test(despues, antes, paired = TRUE)$p.value
```

```
## Warning in wilcox.test.default(despues, antes, paired = TRUE): cannot
## compute exact p-value with ties
```

```
## [1] 0.004565114
```

¿Y la regresión lineal?

Ya vamos a llegar...

Nos sirve para potenciar este tipo de test: controlamos por más variables al mismo tiempo para determinar los efectos.

Test de proporciones

Pregunta: ¿Una proporción poblacional es igual a un valor específico?

Ejemplo: ¿El 60% de los consumidores prefieren nuestro producto?

El desvío de este test surge de la hipótesis nula, a diferencia de los anteriores que surgen de la propia muestra.

Es decir:

- media muestral: $\overline{p}$
- media poblacional (a testear): p
- error estandar de la media (desvio / n): $\sqrt{\frac{p*(1-p)}{n}}$

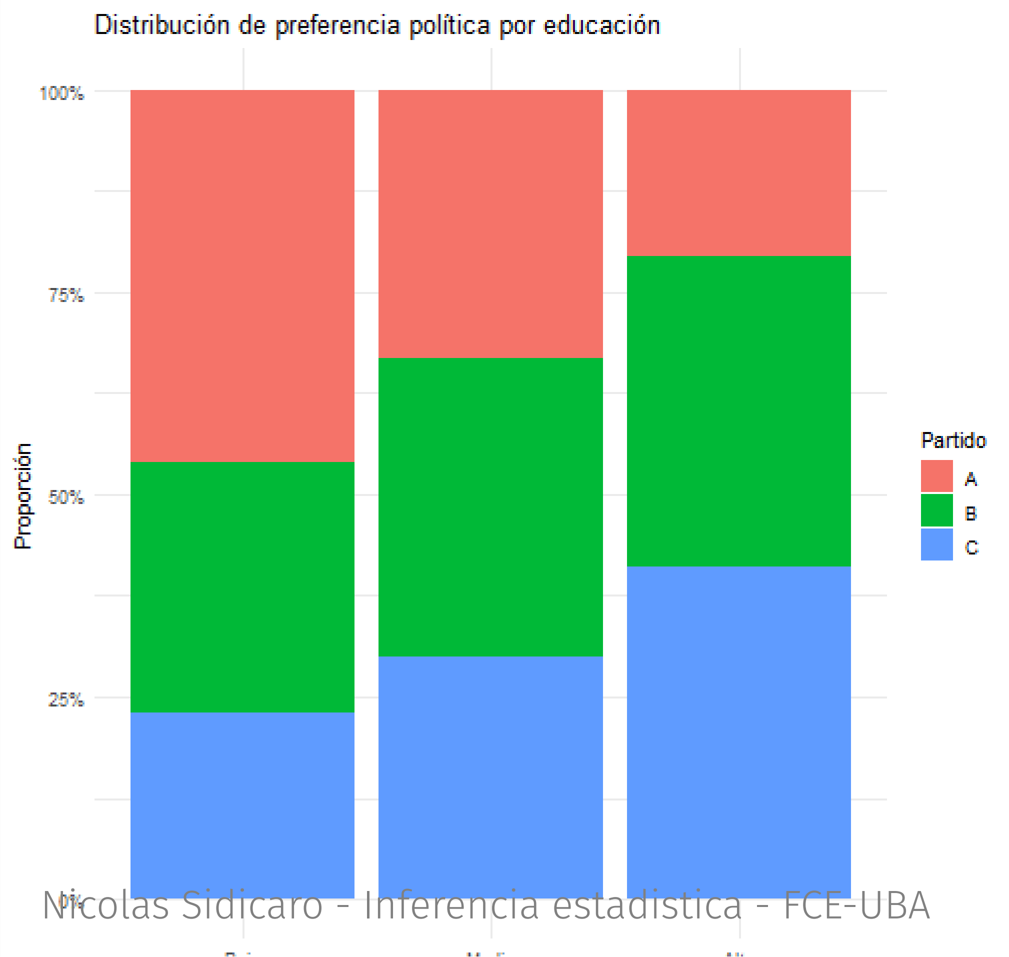
¿Por qué usamos p en lugar de $\overline{p}$? Porque el test pregunta: "Si H_0 fuera verdadera (p = 0.6) ¿qué tan raro sería observar un $\overline{p} = 0.7$?

Entonces se calcula considerando la variabilidad esperada bajo H_0

Test Chi de Independencia

Pregunta: ¿Dos variables categóricas son independientes?

Ejemplo: ¿El nivel educativo está relacionado con la preferencia política?



Test Chi de Independencia

```
##
##      Pearson's Chi-squared test
##
## data:  tabla
## X-squared = 48.192, df = 4, p-value = 0.0000000008605
```

Rechazamos H0 => las variables están relacionadas

Código 5

Resumen: ¿Qué test usar?

Situacion	**Test**
1 muestra vs valor ($n \geq 30$)	t.test(x, mu = valor)
1 muestra vs valor ($n < 30$ o NO normal)	wilcox.test(x, mu = valor)
2 muestras independientes ($n \geq 30$)	t.test(x, y)
2 muestras independientes ($n < 30$ o NO normal)	wilcox.test(x, y)
2 muestras pareadas ($n \geq 30$)	t.test(x, y, paired = TRUE)
2 muestras pareadas ($n < 30$ o NO normal)	wilcox.test(x, y, paired = TRUE)
Verificar normalidad (crucial si $n < 30$)	shapiro.test() + Q-Q plot
Verificar varianzas iguales	leveneTest() o var.test()

Regla general: Con $n < 30$, verificar supuestos. Con $n \geq 30$, el TCL protege al t-test.