

## Actividad N° 4

**Estudiante:** Waldir Yeison Velasquez Quispe

**Profesor:** Fred Torres Cruz

**Codigo:** 236887

### Técnicas de Escalado para Datos Heterogéneos

## 1. Tests para Tendencia Central - Una Muestra

### 1.1. One-sample t-test: `t.test(x, mu = valor)`

**Variables:** Peso corporal (kg)

**Caso de ejemplo:** Un nutricionista quiere verificar si el peso promedio de sus pacientes ( $n=30$ ) difiere significativamente del peso ideal recomendado de 70 kg.

```
1 # Datos: pesos de 30 pacientes
2 pesos <- c(68.5, 72.1, 69.8, 71.2, 67.9, 73.4, 68.2, 70.5,
3           69.7, 71.8, 66.9, 72.3, 68.8, 70.1, 69.4, 71.6,
4           67.5, 72.8, 69.2, 70.9, 68.1, 71.5, 69.9, 70.7,
5           67.8, 72.0, 68.6, 71.1, 69.5, 70.3)
6 t.test(pesos, mu = 70)
```

### 1.2. One-sample z-test: `z.test()` (paquete BSDA)

**Variables:** Puntaje de CI (Coeficiente Intelectual)

**Caso de ejemplo:** Un psicólogo educativo evalúa si el CI promedio de estudiantes de una escuela ( $n=100$ ) difiere del promedio poblacional conocido de 100 puntos ( $\sigma = 15$ ).

```
1 library(BSDA)
2 # Muestra de CIs de estudiantes
3 ci_estudiantes <- c(95, 108, 112, 89, 103, 97, 115, 92, 101, 106,
4                   88, 109, 118, 94, 102, 99, 113, 91, 104, 107)
5 z.test(ci_estudiantes, mu = 100, sigma.x = 15)
```

### 1.3. Wilcoxon signed-rank test: `wilcox.test(x, mu = valor)`

**Variables:** Tiempo de reacción (milisegundos)

**Caso de ejemplo:** Un investigador quiere probar si el tiempo de reacción promedio de gamers difiere de 250 ms, pero los datos no siguen distribución normal.

```
1 tiempo_reaccion <- c(220, 245, 280, 195, 300, 240, 215, 265,
2                   290, 185, 310, 235, 225, 275, 285, 205,
3                   295, 245, 230, 270)
4 wilcox.test(tiempo_reaccion, mu = 250)
```

### 1.4. Sign test: `binom.test()` o `SignTest()` (DescTools)

**Variables:** Satisfacción del cliente (escala 1-10)

**Caso de ejemplo:** Una empresa quiere saber si la satisfacción de sus clientes es significativamente mayor que el punto medio (5.5) de la escala.

```
1 satisfaccion <- c(6, 7, 4, 8, 5, 9, 3, 7, 6, 8, 5, 7, 4, 9, 6,  
2                 8, 3, 7, 5, 8, 6, 9, 4, 7, 5, 8, 6, 7, 4, 9)  
3 library(DescTools)  
4 SignTest(satisfaccion, mu = 5.5)
```

## 2. Tests Comparando Dos Grupos

### 2.1. Grupos Independientes

#### 2.1.1. Independent samples t-test: `t.test(x, y)`

**Variables:** Rendimiento académico (calificaciones 0-20)

**Caso de ejemplo:** Comparar el rendimiento académico entre estudiantes que usan método tradicional vs. método digital de enseñanza.

```
1 grupo_tradicional <- c(14, 16, 13, 15, 17, 12, 18, 14, 16, 15,  
2                      13, 17, 12, 16, 14, 15, 13, 18, 16, 14)  
3 grupo_digital <- c(16, 18, 15, 17, 19, 14, 20, 16, 18, 17,  
4                 15, 19, 14, 18, 16, 17, 15, 20, 18, 16)  
5 t.test(grupo_tradicional, grupo_digital)
```

#### 2.1.2. Welch's t-test: `t.test(x, y, var.equal = FALSE)`

**Variables:** Salario mensual (soles)

**Caso de ejemplo:** Comparar salarios entre empleados de sector público vs. privado, asumiendo varianzas diferentes.

```
1 salario_publico <- c(3500, 4200, 3800, 4000, 3600, 4100, 3700,  
2                 3900, 3550, 4050, 3750, 3850, 3650, 4150)  
3 salario_privado <- c(5500, 7200, 6800, 8000, 9200, 6500, 7800,  
4                 8500, 9800, 7000, 8200, 9500, 6800, 8800)  
5 t.test(salario_publico, salario_privado, var.equal = FALSE)
```

#### 2.1.3. Mann-Whitney U test: `wilcox.test(x, y)`

**Variables:** Nivel de dolor (escala 1-10)

**Caso de ejemplo:** Comparar la efectividad de dos analgésicos midiendo el nivel de dolor reportado por pacientes.

```
1 analgesico_A <- c(3, 2, 4, 1, 3, 2, 5, 1, 4, 2, 3, 1, 4, 2, 3,  
2                 5, 2, 3, 1, 4)  
3 analgesico_B <- c(6, 5, 7, 4, 6, 5, 8, 4, 7, 5, 6, 4, 7, 5, 6,  
4                 8, 5, 6, 4, 7)  
5 wilcox.test(analgesico_A, analgesico_B)
```

### 2.2. Grupos Relacionados/Pareados

#### 2.2.1. Paired t-test: `t.test(x, y, paired = TRUE)`

**Variables:** Presión arterial sistólica (mmHg)

**Caso de ejemplo:** Evaluar el efecto de un medicamento comparando la presión arterial antes y después del tratamiento en los mismos pacientes.

```

1 presion_antes <- c(140, 135, 150, 145, 138, 142, 148, 136, 144, 139,
2                   147, 141, 149, 137, 143, 146, 134, 151, 140, 145)
3 presion_despues <- c(125, 120, 135, 130, 125, 128, 133, 122, 129, 124,
4                     132, 126, 134, 123, 128, 131, 119, 136, 125, 130)
5 t.test(presion_antes, presion_despues, paired = TRUE)

```

### 2.2.2. Wilcoxon signed-rank (pareado): wilcox.test(x, y, paired = TRUE)

**Variables:** Horas de sueño

**Caso de ejemplo:** Comparar las horas de sueño antes y después de implementar técnicas de relajación, cuando los datos no siguen distribución normal.

```

1 sue o_antes <- c(5.5, 6.0, 4.8, 5.2, 6.5, 5.8, 4.5, 5.9, 5.3, 6.2,
2                 4.9, 5.7, 5.1, 6.3, 5.6, 4.7, 6.1, 5.4, 5.0, 6.0)
3 sue o_despues <- c(7.2, 7.8, 6.5, 7.0, 8.2, 7.5, 6.8, 7.9, 7.1, 8.0,
4                   6.7, 7.6, 6.9, 8.1, 7.4, 6.6, 7.8, 7.2, 6.8, 7.7)
5 wilcox.test(sue o_antes, sue o_despues, paired = TRUE)

```

## 3. Tests Comparando Múltiples Grupos

### 3.1. ANOVA y Equivalentes No Paramétricos

#### 3.1.1. One-way ANOVA: aov(variable ~ grupo)

**Variables:** Productividad (unidades por hora)

**Caso de ejemplo:** Comparar la productividad entre trabajadores de tres turnos diferentes (mañana, tarde, noche).

```

1 productividad <- c(25, 27, 23, 28, 22, 20, 18, 19, 24, 26, 25, 21,
2                   29, 24, 26, 17, 19, 20, 23, 25, 27, 18, 20, 22)
3 turno <- factor(c(rep("mañana", 8), rep("tarde", 8), rep("noche", 8)))
4 modelo <- aov(productividad ~ turno)
5 summary(modelo)
6 # Post-hoc si es significativo
7 TukeyHSD(modelo)

```

#### 3.1.2. Kruskal-Wallis test: kruskal.test(variable ~ grupo)

**Variables:** Ranking de preferencia (1-5)

**Caso de ejemplo:** Comparar la preferencia de consumidores entre cuatro marcas de smartphones usando una escala ordinal.

```

1 preferencia <- c(4, 5, 3, 4, 2, 1, 3, 2, 5, 4, 3, 1, 4, 3, 5,
2                 2, 3, 4, 1, 2, 5, 3, 4, 2, 1, 3, 5, 4, 2, 3)
3 marca <- factor(c(rep("Apple", 8), rep("Samsung", 8),
4                   rep("Huawei", 7), rep("Xiaomi", 7)))
5 kruskal.test(preferencia ~ marca)
6 # Post-hoc si es significativo
7 library(dunn.test)
8 dunn.test(preferencia, marca)

```

### 3.1.3. Repeated measures ANOVA: `aov(variable ~ grupo + Error(sujeto/grupo))`

**Variables:** Concentración en plasma (mg/dl)

**Caso de ejemplo:** Evaluar el efecto de tres dosis diferentes de un medicamento midiendo la concentración en plasma de los mismos pacientes en tres momentos.

```

1 concentracion <- c(2.1, 3.5, 4.2, 1.8, 3.2, 4.0, 2.3, 3.8, 4.5,
2                   2.0, 3.4, 4.1, 1.9, 3.3, 4.3, 2.2, 3.7, 4.4,
3                   2.4, 3.6, 4.0, 1.7, 3.1, 4.2, 2.1, 3.5, 4.3,
4                   2.3, 3.4, 4.1)
5 dosis <- factor(rep(c("baja", "media", "alta"), 10))
6 paciente <- factor(rep(1:10, each = 3))
7 modelo_rm <- aov(concentracion ~ dosis + Error(paciente/dosis))
8 summary(modelo_rm)

```

### 3.1.4. Friedman test: `friedman.test()`

**Variables:** Evaluación de sabor (escala 1-7)

**Caso de ejemplo:** Comparar la evaluación de sabor de cuatro recetas diferentes por los mismos catadores.

```

1 # Matriz donde filas = catadores, columnas = recetas
2 sabor_matrix <- matrix(c(5, 6, 4, 7,
3                          4, 5, 3, 6,
4                          6, 7, 5, 7,
5                          3, 4, 2, 5,
6                          5, 6, 4, 6,
7                          4, 5, 3, 7,
8                          6, 7, 5, 6,
9                          5, 6, 4, 7,
10                         4, 5, 3, 6,
11                         5, 6, 4, 7,
12                         3, 4, 2, 5,
13                         6, 7, 5, 6),
14                       nrow = 12, byrow = TRUE)
15 friedman.test(sabor_matrix)
16 # Post-hoc si es significativo
17 library(PMCMRplus)
18 frdAllPairsNemenyiTest(sabor_matrix)

```

## 4. Tests para Varianza

### 4.1. F-test para igualdad de varianzas: `var.test(x, y)`

**Variables:** Tiempo de espera en consulta (minutos)

**Caso de ejemplo:** Comparar la variabilidad en tiempos de espera entre dos clínicas médicas.

```

1 clinica_A <- c(15, 20, 18, 22, 16, 25, 12, 19, 21, 17, 23, 14,
2              18, 20, 16, 24, 13, 19, 22, 15)
3 clinica_B <- c(30, 35, 28, 40, 32, 38, 29, 33, 37, 31, 36, 27,
4              34, 39, 30, 35, 28, 32, 38, 33)
5 var.test(clinica_A, clinica_B)

```

## 4.2. Levene's test: `leveneTest()` (paquete `car`)

**Variables:** Ingresos familiares (soles)

**Caso de ejemplo:** Evaluar si la variabilidad de ingresos es similar entre tres distritos de Lima.

```
1 library(car)
2 ingresos <- c(2500, 3200, 2800, 3000, 2700, 3100, 2900, 3300,
3             4500, 6200, 5800, 5200, 4800, 5600, 5000, 5900,
4             8500, 9200, 7800, 8800, 9000, 8200, 7900, 8600)
5 distrito <- factor(c(rep("San Juan", 8), rep("Miraflores", 8),
6                   rep("San Isidro", 8)))
7 leveneTest(ingresos ~ distrito)
```

## 4.3. Bartlett's test: `bartlett.test(variable ~ grupo)`

**Variables:** Puntuación en examen (0-100)

**Caso de ejemplo:** Verificar si la variabilidad en puntuaciones es similar entre estudiantes de diferentes carreras universitarias.

```
1 puntuacion <- c(85, 78, 92, 88, 76, 82, 90, 94, 89, 84,
2             72, 68, 75, 79, 71, 77, 73, 69, 74, 76,
3             91, 87, 95, 89, 93, 86, 92, 88, 90, 85)
4 carrera <- factor(c(rep("Medicina", 10), rep("Ingenier a", 10),
5                   rep("Psicolog a", 10)))
6 bartlett.test(puntuacion ~ carrera)
```

# 5. Consideraciones Prácticas

## 5.1. Proceso de Selección del Test Apropriado

### 1. Verificar supuestos:

- Normalidad: `shapiro.test()`, `qqnorm()`, `qqline()`
- Homogeneidad de varianzas: `leveneTest()`, `bartlett.test()`

### 2. Considerar el diseño del estudio:

- ¿Una muestra, dos grupos o múltiples grupos?
- ¿Grupos independientes o relacionados?
- ¿Datos paramétricos o no paramétricos?

### 3. Aplicar tests post-hoc cuando sea necesario:

- Después de ANOVA significativo: Tukey HSD
- Después de Kruskal-Wallis significativo: Dunn's test
- Después de Friedman significativo: Nemenyi test

## 5.2. Ejemplo de Flujo de Decisión

```

1 # 1. Evaluar normalidad
2 shapiro.test(datos)
3
4 # 2. Si normal y dos grupos independientes
5 if(p_value > 0.05 && grupos_independientes) {
6   # Verificar igualdad de varianzas
7   resultado_var <- var.test(grupo1, grupo2)
8   # Aplicar t-test apropiado
9   if(resultado_var$p.value > 0.05) {
10     t.test(grupo1, grupo2, var.equal = TRUE)
11   } else {
12     t.test(grupo1, grupo2, var.equal = FALSE)
13   }
14 }
15
16 # 3. Si no normal, usar alternativa no paramétrica
17 if(p_value <= 0.05 && grupos_independientes) {
18   wilcox.test(grupo1, grupo2)
19 }

```

## 5.3. Tabla Resumen de Selección de Tests

Situación	Paramétrico	No Paramétrico	Post-hoc
Una muestra	t-test	Wilcoxon signed-rank	-
Dos grupos indep.	t-test/Welch	Mann-Whitney U	-
Dos grupos pareados	Paired t-test	Wilcoxon pareado	-
Múltiples grupos indep.	One-way ANOVA	Kruskal-Wallis	Tukey/Dunn
Múltiples grupos pareados	RM-ANOVA	Friedman	Tukey/Nemenyi
Varianzas (2 grupos)	F-test	-	-
Varianzas (>2 grupos)	Bartlett/Levene	-	-