Tests para igualdad de varianzas

En los problemas de comparaciones de dos o más tratamientos que hemos tratado hemos supuesto que, si existe una diferencia entre tratamientos, ésta se traduce en un corrimiento. Obviamente hay otras alternativas posibles, como por ejemplo que la diferencia entre tratamientos se traduzca en un cambio en la varianza. El primer test que veremos trabaja con los rangos tal como los definimos en el problema de posición.

Test de Mood para igualdad de varianzas: Sean $X_1,...,X_n$ e $Y_1,...,Y_m$ muestras aleatorias de distribuciones $F((x-\mu)/\sigma)$ y $F((x-\mu)/\tau)$ respectivamente, con μ común y conocida. Supongamos que la hipótesis nula de interés es H_o: $\sigma = \tau$.

Sea $R(Y_i)$ el rango de Y_i en la muestra agrupada de tamaño $N = n+m$, y definamos

$$M = \sum_{i=1}^{m} \left(R(Y_i) - \frac{N+1}{2} \right)^2$$

Valores grandes de *M* indicarían que $\tau > \sigma$ y valores pequeños de *M* indicarían que $\tau < \sigma$. Por lo tanto, se pueden construir tests para las diferentes alternativas basados en M.

Tablas exactas de la distribución de *M* bajo H_o se hallan en Laubscher, Steffens y DeLange (1968). Si n y m sin suficientemente grandes se puede usar la aproximación normal, basada en

$$E(M) = m\frac{N^2-1}{12}$$

$$Var(M) = nm\frac{(N+1)(N^2-4)}{180}$$

Hipótesis a testear y región de rechazo:

A. H_o: $\sigma = \tau$ vs H_1: $\sigma > \tau$

Rechazamos H_o si $M < m\frac{N^2-1}{12} - z_\alpha \sqrt{Var(M)}$

B. H_o: $\sigma = \tau$ vs H_1: $\sigma < \tau$

Rechazamos H_o si $M > m\frac{N^2-1}{12} + z_\alpha \sqrt{Var(M)}$

C. H_o: $\sigma = \tau$ vs H_1: $\sigma \neq \tau$

Rechazamos H_o si $\left| M - m\frac{N^2-1}{12} \right| > z_{\alpha/2} \sqrt{Var(M)}$

Notar que se supone que los parámetros de posición de ambas muestras son conocidos e idénticos. Si no se puede suponer que son idénticos, debe realizarse previamente algún tipo de ajuste. Si los valores que se utilizan en ese ajuste son estimaciones alejadas de los verdaderos valores de los parámetros, es posible que el rechazo de H_o no implique necesariamente que las varianzas son distintas.

Eficiencia: La eficiencia el test de Mood fue estudiada por Hollander (1963). Respecto del test F y en el caso Normal es 0.76.

Ejemplo: Se desea testear si el orden en que las preguntas son formuladas influye en los resultados en un examen multiple choice. 21 alumnos que tomaron un curso de Inferencia Estadística fueron divididos en dos grupos aleatoriamente y se les tomó un examen que constaba de 20 preguntas. A 11 alumnos (Grupo 1) las preguntas se les presentaron en el orden en que se habían desarrollado los temas en clase y al resto (Grupo 2) en un orden aleatorio. Los resultados (número de respuestas correctas) se presentan en la siguiente tabla, juntamente con los rangos:

Grupo 1 (X)	**Grupo 2 (Y)**	**Rango X**	**Rango Y**
	3		1
	6		2
7		3	
	8		4
9		6	
9		6	
	9		6
10		8	
11		10.5	
11		10.5	
	11		10.5
	11		10.5
12		13	
	13		15
	13		15
	13		15
14		17	
15		18.5	
	15		18.5
17		20	
19		21	

Supongamos que se puede suponer que ambas poblaciones tienen el mismo centro y que se desea testear H_o: $\sigma = \tau$ vs H_1: $\sigma \neq \tau$

E(M) = 366.67 y Var(M) = 5875.22. El valor del estadístico M = 359.75 y por lo tanto el p-valor aproximado será

$$2\left[\phi\left(\frac{359.75-366.67}{76.65}\right)\right]=0.93$$

y no se rechaza H_o.

A continuación se presentan los resultados obtenidos al procesar este ejemplo en R.

```
grupo1<-c(7,9,9,10,11,11,12,14,15,17,19)
grupo2<-c(3,6,8,9,11,11,13,13,13,15)

mood.test(grupo1,grupo2)

        Mood two-sample test of scale

data:  grupo1 and grupo2
Z = 0.021, p-value = 0.9833
alternative hypothesis: two.sided
```

Nota: En caso de empates, R utiliza un ajuste propuesto por Mielke (1967).

Test de rangos al cuadrado (Conover): Sean $X_1,...,X_n$ e $Y_1,...,Y_m$ muestras aleatorias de distribuciones $F((x-\mu_1)/\sigma)$ y $F((x-\mu_2)/\tau)$ respectivamente. Supongamos que la hipótesis nula de interés es

$$H_o: \sigma = \tau$$

Transformamos los datos a desviaciones absolutas respecto de su media, es decir, definimos

$$U_i = |X_i - \mu_1| \,, \quad i = 1,...,n \qquad \text{y} \qquad V_j = |Y_j - \mu_2| \,, \quad j = 1,...,m$$

Sean $R(U_i)$ y $R(V_j)$ los rangos de U_i y V_j respectivamente en la muestra agrupada de tamaño $N=n+m$. Si no hay empates, el estadístico del test será

$$T = \sum_{i=1}^{n} \left[R(U_i)\right]^2$$

que, bajo H_o, satisface

$$E(T) = n\frac{1}{N}\sum_{i=1}^{N} i^2 = \frac{n}{N}\frac{N(N+1)(2N+1)}{6} = \frac{n(N+1)(2N+1)}{6}$$

$$Var(T) = \frac{nm}{(N-1)N}\left\{\sum_{i=1}^{N} i^4 - N\frac{(N+1)^2(2N+1)^2}{36}\right\} = \frac{nm(N+1)(2N+1)(8N+11)}{180}$$

Si hay empates, se utiliza el estadístico

$$T_1 = \frac{T - n\,\overline{R}_2}{\left[\frac{nm}{N(N-1)}\sum_{i=1}^{N} R_i^4 - \frac{nm}{N-1}\left(\overline{R}_2\right)^2\right]^{1/2}}$$

donde $\overline{R}_2 = \frac{1}{N}\left\{\sum_{i=1}^{n}\left[R(U_i)\right]^2 + \sum_{j=1}^{m}\left[R(V_j)\right]^2\right\}$ es el promedio de los rangos al cuadrado de ambas muestras combinadas y $\sum_{i=1}^{N} R_i^4 = \sum_{i=1}^{n}\left[R(U_i)\right]^4 + \sum_{j=1}^{m}\left[R(V_j)\right]^4$.

Los cuantiles de la distribución exacta de T se encuentran tabulados (Tabla A9 de Conover) para n ≤ 10, m ≤ 10 y en el caso en que no hay empates. Para tamaños mayores de muestra o si hay empates se utiliza la aproximación Normal.

Supongamos que deseamos testear las hipótesis:

A. H_o: $\sigma = \tau$ vs H_1: $\sigma > \tau$

Si n y m son menores o iguales que 10 y no hay empates, se rechaza H_o si T > $w_{1-\alpha}$, siendo $w_{1-\alpha}$ el cuantil 1- α de la distribución exacta.

Si n y m son grandes y no hay empates, se rechaza H_o si $\frac{T - E(T)}{\sqrt{Var(T)}} > z_\alpha$.

Si hay empates, se rechaza H_o si

$$T_1 = \frac{T - n\,\overline{R}_2}{\left[\frac{nm}{N(N-1)}\sum_{i=1}^{N} R_i^4 - \frac{nm}{N-1}\left(\overline{R}_2\right)^2\right]^{1/2}} > z_\alpha$$

De manera análoga se aplica el test a las otras dos alternativas.

<u>Eficiencia</u>: La eficiencia de este test respecto al test F es similar a la del test de Mood. Es $15/(2\ \pi^2)=0.76$ en el caso Normal, 1 en el caso uniforme y 1.08 en el caso doble exponencial.

En caso de que las medias sean desconocidas, deben reemplazarse por estimaciones. En este caso la distribución exacta deja de ser válida y debe usarse la aproximación normal.

Extensión a más de dos muestras: Este test se puede extender fácilmente a más de dos muestras. Sean

$$X_{ij} \sim F\left(\frac{x-\mu_i}{\sigma_i}\right) \qquad 1 \leq j \leq n_i,\ 1 \leq i \leq k$$

Si μ_i es conocido para todo i, se considera $U_{ij} = \left|X_{ij} - \mu_i\right|$. Si no se conoce, se estima y se considera $U_{ij} = \left|X_{ij} - \hat{\mu}_i\right|$.

Una vez construidos los U_{ij} se procede como en el test de Kruskal-Wallis, definiendo $R_{ij} = R(U_{ij})$ = rango de U_{ij} en la muestra total ordenada de tamaño $N = \sum_{i=1}^{k} n_i$.

Sea $S_i = \sum_{j=1}^{n_i} \left(R_{ij}\right)^2$ (en el caso k=2, S_1 = T)

Nos interesa testear

H_o: $\sigma_1 = \sigma_2 = ... = \sigma_k$ vs H_1: existe al menos un par $i \neq j$ tal que $\sigma_i \neq \sigma_j$

Como en el test de Kruskal-Wallis, se construye un estadístico que, bajo H_o tiene distribución asintótica χ^2_{k-1}.

Si no hay empates, el estadístico es

$$T = \frac{180}{N(N+1))(2N+1)(8N+11)}\left[\sum_{j=1}^{k} \frac{S_j^2}{n_j} - \frac{N(N+1)^2(2N+1)^2}{36}\right]$$

Si hay empates, se deben corregir la esperanza y la varianza y se utiliza

$$T = \frac{1}{D^2}\left(\sum_{i=1}^{k} \frac{S_i^2}{n_i} - N\,\overline{S}^2\right)$$

donde $\overline{S} = \frac{1}{N}\sum_{i=1}^{k} S_i$ y $D^2 = \frac{1}{N-1}\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} R_{ij}^4 - N\,\overline{S}^2\right)$.

Se rechaza H_o a nivel α si T > $\chi^2_{k-1,\alpha}$.

<u>Comparaciones múltiples</u>: Diremos que σ_i es significativamente distinta de σ_j si

$$\left|\frac{S_i}{n_i}-\frac{S_j}{n_j}\right| > t_{N-k,\alpha/2}\left(D^2\,\frac{N-1-T}{N-k}\right)^{1/2}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)^{1/2}$$

donde $t_{N-k,\alpha/2}$ es el valor crítico α/2 de la distribución t con N-k grados de libertad.

Test de Siegel-Tukey: Sean $X_1,...,X_n$ e $Y_1,...,Y_m$ muestras aleatorias de distribuciones $F((x-\mu)/\sigma)$ y $F((x-\mu)/\tau)$ respectivamente. Supongamos que las hipótesis de interés son

$$H_o: \sigma = \tau \qquad \text{vs} \qquad H_1: \sigma > \tau$$

Se asignan rangos de la siguiente manera:

- rango 1 a la menor observación en la muestra agrupada
- rango 2 a la mayor observación
- rango 3 a la segunda mayor observación
- rango 4 a la segunda menor observación y así sucesivamente.

Si la muestra X tiene mayor dispersión que la muestra Y, los rangos correspondientes a la muestra Y deberían ser mayores que los de la muestra X. Llamemos S_1, S_2,..., S_m a los rangos de la muestra Y en la muestra agrupada, el estadístico del test es

$$T_v = \sum_{i=1}^{m} S_i$$

y se rechazará H_o a favor de H_1 si $T_v > c$. ¿Cómo se halla el valor c?

Se puede probar que T_V tiene la misma distribución que el estadístico del test de Mann-Whitney, y por lo tanto si no hay empates y los tamaños de muestra son menores o iguales que 20, se puede utilizar la Tabla A7 de Conover. Si n y m son grandes se rechaza H_o si

$$T_v > m(N+1) + z_\alpha \sqrt{\frac{mn(N+1)}{12}}$$

Este test se basa en el supuesto de que las medias de ambas muestras son iguales. Si ésto no ocurriese, podrían reemplazarse las medias por valores estimados y trabajar con $X_i - \hat{\mu}_X$ y $Y_j - \hat{\mu}_Y$. En este caso, debe usarse la distribución aproximada.

<u>Eficiencia</u>: Este test tiene menor eficiencia que los tests anteriores. Su eficiencia respecto al test F bajo normalidad es $6/\pi^2 = 0.608$ y bajo la distribución doble exponencial es 0.94. Otro problema adicional de este test es su falta de simetría, es decir que si se invierte el orden de ranqueo, el p-valor obtenido no será el mismo.

Ejemplo Este test aplicado a los datos del ejemplo usado previamente da por resultado

SIEGEL-TUKEY TEST
[Sum of scores from population < 1 >]

Min	Max	Mean	Std-dev	Observed	Standardized
57.33	162.3	110.0	13.95	109.3	-0.04778

Asymptotic Inference:
One-sided p-value: Pr { Test Statistic .LE. Observed } = 0.4809
Two-sided p-value: 2 * One-sided = 0.9619

Test de Klotz: Es un test basado en scores normales, con alta eficiencia si la distribución es simétrica. Bajo normalidad su eficiencia respecto al test F es 1. Sean $X_1,...,X_n$ e $Y_1,...,Y_m$ muestras aleatorias de distribuciones $F((x-\mu)/\sigma)$ y $F((x-\mu)/\tau)$.

El estadístico del test de Klotz es

$$K = \sum_{i=1}^{m} \left[\Phi^{-1}\left(\frac{R(Y_i)}{N+1} \right) \right]^2$$

Klotz (1962) calculó la distribución exacta, pero si los tamaños de muestra son grandes puede utilizarse la aproximación Normal y

$$E(K) = \frac{m}{N} \sum_{i=1}^{N} \left[\Phi^{-1}\left(\frac{i}{N+1} \right) \right]^2$$

$$Var(K) = \frac{nm}{N(N-1)} \sum_{i=1}^{N} \left[\Phi^{-1}\left(\frac{i}{N+1} \right) \right]^4 - \frac{[E(K)]^2 n}{m(N-1)}$$

El test de Klotz aplicado a los datos del ejemplo anterior, da por resultado

KLOTZ TEST
[Sum of scores from population < 1 >]

Min	Max	Mean	Std-dev	Observed	Standardized
1.347	14.28	7.624	2.022	7.613	-0.005233

Asymptotic Inference:
One-sided p-value: Pr { Test Statistic .LE. Observed } = 0.4979
Two-sided p-value: 2 * One-sided = 0.9958

Correlación de rangos y asociación

Introduciremos nociones de correlación basadas en rangos. Consideraremos el caso en que las observaciones consisten en pares $(X_1,Y_1),...,(X_n,Y_n)$ y supondremos que $(X_i,Y_i) \sim F(x,y)$ con F absolutamente continua con marginales F_X y F_Y en Ω_o.

Antes de proponer métodos basados en rangos recordemos que el coeficiente de correlación lineal, o coeficiente de Pearson, se define como

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\left[\sum_{i=1}^{n}(X_i - \overline{X})^2 \sum_{i=1}^{n}(Y_i - \overline{Y})^2\right]^{1/2}}$$

que es el cociente entre la covarianza muestral entre *X* e *Y* y el producto de sus desvíos standard muestrales. Este coeficiente es una v.a. que mide la asociación lineal entre *X* e *Y* y su función de distribución depende de la distribución de (*X,Y*).

Presentaremos algunas medidas de correlación basadas en rangos cuya distribución no depende de la distribución de (*X,Y*) si ésta es absolutamente continua.

Coeficiente Rho de Spearman: Es el coeficiente de Pearson aplicado a los rangos. Sean $R(X_i)$ el rango de X_i entre las *X*'s y $R(Y_i)$ el rango de Y_i entre las *Y*'s, entonces el coeficiente de correlación de Spearman (1904) se define como

$$\rho = \frac{\sum_{i=1}^{n} R(X_i)R(Y_i) - n\left(\frac{n+1}{2}\right)^2}{\left(\sum_{i=1}^{n}\left[R(X_i)\right]^2 - n\left(\frac{n+1}{2}\right)^2\right)^{1/2}\left(\sum_{i=1}^{n}\left[R(Y_i)\right]^2 - n\left(\frac{n+1}{2}\right)^2\right)^{1/2}}$$

Si no hay empates, se reduce a

$$\rho = 1 - \frac{6\sum_{i=1}^{n}\left[R(X_i) - R(Y_i)\right]^2}{n(n^2-1)}$$

Si suponemos, sin pérdida de generalidad que $X_1 < X_2 < ... X_n$, entonces el coeficiente de correlación de Spearman puede expresarse como

$$\rho = \frac{\sum_{i=1}^{n}\left(i - \frac{n+1}{2}\right)\left(R(Y_i) - \frac{n+1}{2}\right)}{\left(\sum_{i=1}^{n}\left(i - \frac{n+1}{2}\right)^2\right)^{1/2}\left(\sum_{i=1}^{n}\left(R(Y_i) - \frac{n+1}{2}\right)^2\right)^{1/2}}$$

Muchas veces el coeficiente de Spearman se utiliza para testear independencia de variables aleatorias (la hipótesis nula es la independencia de *X* e *Y*). En rigor, se debería decir que se utiliza para testear correlación entre variables aleatorias. Bajo la hipótesis nula de independencia, los rangos de Y se distribuyen uniformemente sobre los primeros n enteros y

$$E(\rho) = 0 \qquad\qquad V(\rho) = \frac{1}{n-1}$$

A su vez, la distribución asintótica es:

$$\sqrt{n-1}\ \rho \xrightarrow{d} N(0,1)$$

La distribución exacta de ρ se ha tabulado para $n \leq 30$ (Tabla A10 de Conover) y para valores mayores de n, el cuantil p se aproxima mediante:

$$w_p \cong \frac{z_p}{\sqrt{n-1}}$$

donde z_p es el cuantil p de la distribución Normal standard. Los cuantiles inferiores se obtienen a partir de la relación $w_p = -w_{1-p}$.

Supongamos que se desea testear, por ejemplo, las hipótesis

H_o: X e Y son no correlacionadas

H_1: valores grandes de X tienden a estar apareados con valores grandes de Y y valores pequeños de X tienden a estar apareados con valores pequeños de Y.

Se rechazará H_o si $\rho > w_{1-\alpha}$, siendo $w_{1-\alpha}$ el percentil 1-α de la distribución exacta (Tabla A10 de Conover) o el percentil aproximado. El p-valor aproximado se obtiene en la forma

$$\text{p-valor} \cong P(Z > \rho\sqrt{n-1}) = 1 - \Phi(\rho\sqrt{n-1})$$

Ejemplo: Para estudiar la relación entre los scores obtenidos en los exámenes de ingreso a un Master y el promedio general obtenido al finalizar los estudios, se analizaron los datos correspondientes a 12 graduados del Master. En la siguiente tabla se presentan los datos y sus correpondientes rangos:

Graduado	Ingreso (X)	Promedio (Y)	R(X)	R(Y)
1	710	4.0	12	11.5
2	610	4.0	9.5	11.5
3	640	3.9	11	10
4	580	3.8	8	9
5	545	3.7	3	8
6	560	3.6	5	7
7	610	3.5	9.5	5
8	530	3.5	1	5
9	560	3.5	5	5
10	540	3.3	2	3
11	570	3.2	7	1.5
12	560	3.2	5	1.5

El valor del coeficiente de Spearman ρ es 0.59.

Cálculo del coeficiente usando R:

```
ingreso<-c(710,610,640,580,545,560,610,530,560,540,570,560)
promedio<-c(4,4,3.9,3.8,3.7,3.6,3.5,3.5,3.5,3.3,3.2,3.2)
cor(rank(promedio),rank(ingreso))

 0.5900188
```

Coeficiente Tau de Kendall (1938): Se basa en la definición de concordancia. Dadas dos observaciones bivariadas (X_1,Y_1) y (X_2,Y_2) diremos que son concordantes si

$$(X_1 > X_2 \ \wedge\ Y_1 > Y_2) \qquad \text{ó} \qquad (X_1 < X_2 \ \wedge\ Y_1 < Y_2)$$

o sea, si $sg(X_2 - X_1)\, sg(Y_2 - Y_1) = 1$. Análogamente, diremos que son discordantes si $sg(X_2 - X_1)\, sg(Y_2 - Y_1) = -1$.

Sea N_c el número de pares concordantes entre los $\binom{n}{2}$ pares posibles y N_d el número de pares discordantes. Por ahora suponemos que no hay empates.

El exceso de concordancia sobre discordancia será

$$N_c - N_d = \sum_{i<j} sg(X_j - X_i)\, sg(Y_j - Y_i)$$

Observemos que esta diferencia toma valores entre $-\frac{n(n-1)}{2}$ y $\frac{n(n-1)}{2}$. El coeficiente de Kendall se define como

$$\tau = \frac{N_c - N_d}{max(N_c - N_d)} = \frac{2(N_c - N_d)}{n(n-1)}$$

De esta manera, si todos los pares son concordantes τ = 1 y si todos los pares son discordantes τ = -1. Bajo independencia E(τ) = 0 pues E(sg($X_j - X_i$) sg($Y_j - Y_i$)) = 0

¿Cómo se procede si hay empates? Sólo se comparan los pares con $X_i \neq X_j$, es decir se descartan los empates en X, y se utiliza la siguiente forma del coeficiente:

$$\tau = \frac{N_c - N_d}{N_c + N_d}$$

donde los empates en Y se cuentan como ½ en N_c y ½ en N_d. Es decir:

- si $X_i = X_j$, se excluye ese par
- si sg($X_j - X_i$) sg($Y_j - Y_i$) = 1, se suma 1 a N_c
- si sg($X_j - X_i$) sg($Y_j - Y_i$) = -1, se suma 1 a N_d
- si sg($X_j - X_i$) sg($Y_j - Y_i$) = 0, se suma 1/2 a N_c y 1/2 a N_d

Goodman y Kruskal (1963) propusieron esta modificación del coeficiente de Kendall y algunos autores la denominan coeficiente gamma.

<u>Ejemplo</u>: En el ejemplo anterior, los números de pares concordantes y discordantes son N_c = 44.5 y N_d = 17.5, por lo cual

$$\tau = \frac{N_c - N_d}{N_c + N_d} = \frac{44.5 - 17.5}{44.5 + 17.5} = 0.4355$$

También el coeficiente tau de Kendall puede ser utilizado para testear la hipótesis de independencia (o mejor expresado “no correlación) entre X e Y. Se utiliza como estadístico del test, si no hay empates

$$T = N_c - N_d$$

y si hay empates el coeficiente τ. La distribución exacta bajo Ho (independencia) de T y τ está tabulada para n $\leq$ 60 (Tabla A11 de Conover). Para valores mayores de n o en presencia de empates, se aproxima el cuantil de τ mediante:

$$w_p = z_p \frac{\sqrt{2(2n+5)}}{3\sqrt{n(n-1)}}$$

y el cuantil de T mediante:

$$w_p = z_p \sqrt{n(n-1)(2n+5)/18}$$