Tema 9. Contrastes de hipótesis no paramétricos

A. Hermoso Carazo
Universidad de Granada

Curso 2020/2021

En todos los problemas de inferencia considerados hasta ahora (inferencia paramétrica) se ha supuesto que la distribución de la variable bajo estudio es conocida salvo, quizás, el valor de uno o varios parámetros. En la práctica, sin embargo, no siempre ocurre esto y es necesario usar métodos de inferencia libres de esta suposición.

Un problema de inferencia no paramétrica consiste en obtener conclusiones sobre la distribución de una o varias variables, a partir de observaciones de ellas, cuando no se conoce su forma funcional explícita. El conocimiento sobre las distribuciones puede ser nulo, o bien referirse a características de tipo general: las variables son de tipo discreto o de tipo continuo, tienen momentos de todos los órdenes, son simétricas, etc.

Aunque existen procedimientos no paramétricos de estimación, aquí nos centramos exclusivamente en algunos problemas típicos de contraste de hipótesis. En concreto, en los denominados problemas de *bondad de ajuste* y *localización*, a partir de una muestra de una variable aleatoria, el problema de *independencia de dos variables* y, por último, el *problema de homogeneidad o igualdad de distribuciones.*

9.1. Problema de bondad de ajuste

Se trata de decidir, a partir de la información proporcionada por una muestra de una variable, si la distribución de dicha variable es una específica ($\mathcal{N}(0,1)$, $exp(3)$...), o bien si corresponde a un determinado modelo (normal, exponencial...).

Presentamos aquí las dos soluciones más frecuentes para resolver este problema, el *test χ^2 de Pearson* y el *test de Kolmogorov-Smirnov*.

9.1.1. Test χ^2 de bondad de ajuste (K. Pearson)

Cronológicamente, este test es la primera solución al problema de bondad de ajuste.

- *Específico para variables cualitativas, o categóricas, no necesariamente numéricas* (clasifican a los individuos de una población en un número finito de categorías o modalidades).
- Puede usarse, aunque con reservas, para ajustar la distribución de cualquier variable aleatoria, discreta o continua.

Test χ^2 para variables cualitativas:

Supongamos que los individuos de una población pueden clasificarse, de acuerdo a una determinada característica X, en k categorías, clases o modalidades, $A_1, \ldots, A_k$, exhaustivas y mutuamente excluyentes.

Ya que la distribución de X está especificada por las probabilidades de las diferentes categorías, se trata de contrastar las siguientes hipótesis:

$$H_0 : P(X = A_i) = p_i^0, \quad \forall i = 1, \cdots, k$$

$$H_1 : P(X = A_i) \neq p_i^0, \quad \text{para algún } i = 1, \cdots, k.$$

Para ello, se eligen, de forma aleatoria e independiente, n individuos de la población, se les mide la característica X y se consideran las siguientes variables aleatorias:

N_i : *Número de observaciones muestrales en la categoría* A_i, $i = 1, \ldots, k$.

El test se basa en el siguiente estadístico:

Estadístico de contraste: $\chi^2(N_1, \ldots, N_k) = \sum_{i=1}^{k} \frac{(N_i - np_i^0)^2}{np_i^0}$.

Justificación: Al tomar una muestra aleatoria simple de individuos de la población es esperable que el número de individuos en cada modalidad sea proporcional a la probabilidad de dicha modalidad. Por tanto, si H_0 es cierta, es esperable que el número de observaciones muestrales en A_i sea np_i^0, $i = 1, \ldots, k$.

Así, $\chi^2(N_1, \ldots, N_k)$ proporciona una medida de la *discrepancia* entre las *frecuencias observadas* en cada clase (N_i) y las *frecuencias esperadas* bajo H_0 (np_i^0). El factor de ponderación $1/np_i^0$ se usa para dar más importancia a las diferencias en las clases con menos frecuencia esperada (por ejemplo, una diferencia de 2 en una clase donde se esperan 20 observaciones es menos importante que en una clase donde sólo se esperan 5).

Ya que el estadístico mide la discrepancia entre lo observado y lo esperado, *se rechazará H_0 si su valor es grande, $\chi^2(N_1, \ldots, N_k) \geq c$, y se aceptará en caso contrario, $\chi^2(N_1, \ldots, N_k) < c$.*

El punto crítico c, que delimita las regiones de aceptación y rechazo, se obtiene imponiendo el tamaño o nivel de significación requerido, para lo que es preciso conocer la distribución del estadístico bajo H_0. Aunque esta no es conocida explícitamente, sí lo es su distribución asintótica, que fue obtenida por Pearson.

Distribución asintótica de $\chi^2(N_1, \ldots, N_k)$:

$$\text{Bajo } H_0: \quad \chi^2(N_1, \ldots, N_k) \xrightarrow[n \to +\infty]{L} \chi^2(k-1).$$

Nota: Los grados de libertad indican el número de variables N_i, $i = 1, \ldots, k$, que toman valores libremente. Ya que $\sum_{i=1}^{k} N_i = n$, cualquiera de ellas está determinada por el resto. Así, los grados de libertad son el número de variables menos el número de restricciones.

Test χ^2 de tamaño α (aproximado): Para n grande, el test de Pearson de tamaño α es:

$$\varphi(N_1,\dots,N_k)=\begin{cases}1, & \chi^2(N_1,\dots,N_k)\geq \chi^2_{k-1;\ \alpha}\\ 0, & \chi^2(N_1,\dots,N_k)< \chi^2_{k-1;\ \alpha}\end{cases}$$

donde $\chi^2_{k-1;\ \alpha}$ es el valor que deja a la derecha probabilidad α en la distribución $\chi^2(k-1)$.

Si no se especifica un tamaño o nivel de significación, se calcula el *p-nivel asociado a los datos:*

$$P\left(\chi^2(k-1)\geq \chi^2_{exp}\right),\quad \chi^2_{exp}=\chi^2(n_1,\dots,n_k),$$

y se rechaza H_0 si el p-nivel es pequeño.

Nota: Al ser un test asintótico y, por tanto, aproximado, el número de observaciones debe ser suficientemente grande para que la aproximación sea aceptable. Esto va en relación con las frecuencias esperadas; usualmente, el test se considera aceptable si $np_i^0\geq 5,\ \ \forall\, i=1,2,\dots,k.$

Aplicación al ajuste de distribuciones de variables aleatorias:

Supongamos que X es una variable aleatoria con distribución desconocida, y se quiere contrastar que dicha distribución es una específica, P_0:

$$P(X\in A)=P_X(A)=P_0(A),\quad \forall A\in\mathcal{B}.$$

Para ello, se particiona el conjunto de valores de la distribución hipotética P_0 en k subconjuntos no nulos, $\chi_0=A_1\cup\cdots\cup A_k$, $A_i\cap A_j=\emptyset$, $i\neq j$, $P_0(A_i)=p_i^0>0$, $i=1,\dots,k$, y se aplica el test de Pearson para contrastar:

$$H_0: P(X\in A_i)=p_i^0,\quad i=1,\dots,k.$$

Nota: La hipótesis contrastada, H_0, es más débil que la hipótesis original ($P_X=P_0$) de manera que si P_1 es cualquier otra distribución que asigna las mismas probabilidades que P_0 a los conjuntos A_i, $i=1,\dots,k$, el test no discrimina entre P_1 y P_0. Esto es, si una determinada muestra conduce al rechazo (aceptación) de P_0, también conduce al rechazo (aceptación) de P_1. Por este motivo, debe usarse el mayor número posible de conjuntos en la partición (es aconsejable usar, como mínimo, 5 conjuntos, sin olvidar que cada uno tenga frecuencia esperada mayor o igual que 5).

Aplicación al ajuste de distribuciones con parámetros desconocidos:

En muchas ocasiones, la distribución a contrastar tiene una forma funcional conocida, pero depende de parámetros desconocidos. La forma de proceder en tal caso es estimar los parámetros, y contrastar el ajuste a la distribución con los parámetros estimados. Sin embargo, este procedimiento no puede hacerse de forma arbitraria ya que la distribución del estadístico de contraste depende de cómo se hayan obtenido las estimaciones:

- Si se estiman los parámetros a partir de observaciones independientes de las que se van a usar para el contraste, el test de Pearson se aplica sin ninguna modificación.
- Sin embargo, lo usual es estimar los parámetros con las mismas observaciones usadas en el test. En este caso, las probabilidades estimadas, $\widehat{p}_i^0$, dependen de los datos muestrales, y la distribución del estadístico de contraste varía. Lo usual es estimar los parámetros por *máxima verosimilitud* en la distribución a contrastar. En tal caso, el estadístico de contraste es:

$$\widehat{\chi}^2(N_1,\ldots,N_k) = \sum_{i=1}^{k} \frac{(N_i - n\widehat{p}_i^0)^2}{n\widehat{p}_i^0},$$

y su *distribución asintótica bajo* H_0, si se han estimado q parámetros, es $\chi^2(k-q-1)$.

Ejemplo 9.1.1: *Se recoge una muestra aleatoria simple de 30 tornillos producidos por cierta máquina y se mide su longitud, obteniéndose:*

10.39	10.66	10.12	10.32	10.25	10.52	10.83	10.72	10.28	10.35
10.46	10.54	10.23	10.18	10.62	10.49	10.61	10.64	10.29	10.78
10.81	10.34	10.75	10.41	10.53	10.31	10.47	10.43	10.57	10.74

Contrastar si estos datos avalan que la distribución de la longitud de los tornillos es normal.

Se trata de un problema de ajuste, en el que la variable aleatoria observada y la hipótesis a contrastar son:

$$X = \textit{Longitud de los tornillos.} \quad H : X \to \{\mathcal{N}(\mu,\sigma^2);\ \mu \in \mathbb{R},\ \sigma^2 > 0\}.$$

Ya que la distribución hipotética depende de parámetros no especificados, para aplicar el test debemos estimarlos por máxima verosimilitud:

$$\widehat{\mu} = \overline{x} = 10.488, \qquad \widehat{\sigma}^2 = \frac{\sum_{i=1}^{30}(x_i - \overline{x})^2}{30} = 0.038204 \quad (\widehat{\sigma} = 0.195458).$$

Ahora adaptamos el test χ^2 para contrastar la hipótesis

$$H_0 : X \to \mathcal{N}(10.488,\ 0.038204).$$

En primer lugar, debemos particionar el conjunto de valores de la distribución hipotética ($\mathbb{R}$) en k subconjuntos, $A_1, \ldots, A_k$, tales que el número esperado de observaciones en cada uno de ellos sea, al menos, 5.

Así, si $\widehat{p}_i^0 = P_{H_0}(X \in A_i)$, $i = 1, \ldots, k$, ha de ser $30\widehat{p}_i^0 \geq 5$ o, equivalentemente, $\widehat{p}_i^0 \geq 1/6$, $i = 1, \ldots, k$. Entonces, con objeto de tener el mayor número de subconjuntos, tomamos 6 intervalos con probabilidad $1/6$, que se determinan a partir de los cuantiles $Q_{1/6}$, $Q_{2/6}, \cdots, Q_{5/6}$ de la distribución a contrastar.

Notando $Z \to \mathcal{N}(0,1)$:

$$P_{H_0}\left(X \leq Q_{i/6}\right) = P\left(Z \leq \frac{Q_{i/6} - 10.488}{0.195458}\right) = \frac{i}{6}, \quad i = 1, \cdots, 5.$$

La siguiente tabla presenta los cuantiles, los intervalos de la partición, y las frecuencias esperadas y observadas en cada uno:

$Q_{i/6}$	A_i	$n\widehat{p}_i^0$	N_i
10.2989	$(-\infty, 10.2989]$	5	6
10.4038	$(10.2989, 10.4038]$	5	5
10.4880	$(10.4038, 10.4880]$	5	4
10.5722	$(10.4880, 10, 5722]$	5	5
10.6771	$(10.5722, 10.6771]$	5	4
	$(10.6771, +\infty)$	5	6

Tema 9. Contrastes de hipótesis no paramétricos A. Hermoso Carazo
└9.1. Problema de bondad de ajuste
└9.1.1. Test χ^2 de bondad de ajuste (K. Pearson)

Ahora, calculamos el valor experimental del estadístico:

$$\widehat{\chi}^2_{exp} = \sum_{i=1}^{6} \frac{(N_i - n\widehat{p}_i^0)^2}{n\widehat{p}_i^0} = \frac{(6-5)^2}{5} + \frac{(5-5)^2}{5} + \cdots + \frac{(6-5)^2}{5} = 0.8.$$

Ya que no se especifica nivel de significación, calculamos el p-nivel asociado a los datos, teniendo en cuenta que la distribución teórica del estadístico de contraste bajo H_0 es $\chi^2(3)$ (estamos trabajando con seis conjuntos y se han estimado dos parámetros):

$$p - nivel = P_{H_0}\left(\widehat{\chi}^2(N_1, \ldots, N_6) \geq 0.8\right) \approx 0.85.$$

Puesto que este valor es muy grande, se acepta H_0:

Los datos no aportan evidencia para rechazar la hipótesis de que la longitud de los tornillos tiene distribución normal.

Inconvenientes de los tests χ^2:

- Es un *test asintótico* y, por tanto, aproximado, por lo que debe aplicarse con reservas si el tamaño de muestra no es suficientemente grande.
- Si se aplica para ajustar la distribución de una variable aleatoria, *no trata las observaciones individualmente, sino que las agrupa en clases* y, por tanto, no usa toda la información proporcionada por los datos.

  Esto, como se ha comentado anteriormente, puede ser un grave inconveniente, sobre todo en distribuciones de tipo continuo, ya que el test *discretiza tales distribuciones*. Sin embargo, el inconveniente puede aliviarse si el tamaño muestral es muy grande y es posible usar muchas categorías.

  Aún así, para distribuciones de tipo continuo debe usarse el test que describimos a continuación, que no agrupa los datos, sino que los trata individualmente y es un test exacto (no asintótico).

9.1.2. Test de Kolmogorov-Smirnov

Se basa en el *teorema de Glivenko-Cantelli*, que asegura la convergencia casi segura uniforme de la función de distribución muestral a la función de distribución de una variable aleatoria:

> *Si $\{X_n\}_{n\in\mathbb{N}}$ es una sucesión de variables independientes e idénticamente distribuidas con función de distribución F_X, las funciones de distribución muestrales $F^*_{X_1,\ldots,X_n}$ convergen casi segura y uniformemente a F_X:*
>
> $$P\left\{\lim_{n\to+\infty}\sup_{x\in\mathbb{R}}\left|F^*_{X_1,\ldots,X_n}(x)-F_X(x)\right|=0\right\}=1.$$

El test se aplica exclusivamente para ajustar distribuciones de tipo continuo totalmente especificadas. Así, el planteamiento del problema es:

Si X es una variable aleatoria de tipo continuo, se desea contrastar:

$$H_0: P_X = P_0$$
$$H_1: P_X \neq P_0$$

El test se basa en una muestra aleatoria simple de X, $(X_1, \ldots, X_n)$, para la que se considera el siguiente estadístico:

Estadístico de contraste (de Kolmogorov-Smirnov):

$$D(X_1, \ldots, X_n) = \sup_{x \in \mathbb{R}} \left| F^*_{X_1,\ldots,X_n}(x) - F_0(x) \right|,$$

donde F_0 es la función de distribución correspondiente a P_0.

Justificación: $D(X_1, \ldots, X_n)$ proporciona una medida de discrepancia entre $F^*_{X_1,\ldots,X_n}$ y F_0, de forma que si H_0 es cierta, por el *teorema de Glivenko-Cantelli* es de esperar que su valor sea pequeño.

Así, *valores grandes de $D(X_1, \ldots, X_n)$ deben conducir al rechazo de H_0, y la región crítica del test es de la forma $D(X_1, \ldots, X_n) \geq d$,* donde d se calcula imponiendo el tamaño o nivel de significación requerido.

Distribución de $D(X_1, \ldots, X_n)$ bajo H_0: Es conocida y está tabulada, así como su distribución asintótica, conocida como *distribución Z de Kolmogorov*, que fue tabulada por Smirnov.

Test de tamaño α:

$$\varphi(X_1, \ldots, X_n) = \begin{cases} 1, & D(X_1, \ldots, X_n) \geq d_\alpha \\ 0, & D(X_1, \ldots, X_n) < d_\alpha \end{cases} \qquad P_{H_0}\left(D(X_1, \ldots, X_n) \geq d_\alpha\right) = \alpha$$

Si no se especifica el nivel de significación, se calcula el *p-nivel asociado a los datos observados:*

$$P_{H_0}\left(D(X_1, \ldots, X_n) \geq D_{exp}\right), \quad D_{exp} = D(x_1, \ldots, x_n),$$

y se rechaza H_0 si este valor es pequeño.

Cálculo práctico de $D(x_1, \ldots, x_n)$: Si las n observaciones, $x_1, \cdots, x_n$, son distintas (en caso contrario se eliminan las repetidas), notando $F^*(x)$ a $F^*_{x_1,\ldots,x_n}(x)$, se tiene la siguiente expresión:

$$D(x_1, \cdots, x_n) = \max\left\{ \max_{x_i} \left[F^*(x_i) - F_0(x_i)\right], \ \max_{x_i} \left[F_0(x_i) - F^*(x_i^-)\right] \right\}.$$

Para aplicarla, se disponen los datos en orden creciente, $x_{(1)} < \cdots < x_{(n)}$, de manera que $F^*(x_{(i)}) = i/n$, $F^*(x^-_{(i)}) = (i-1)/n$. Así, sólo deben calcularse los valores $F_0(x_{(i)})$ y hacer las diferencias indicadas.

Existe otra expresión para el caso de observaciones igualadas, aunque esta es una situación improbable por ser la distribución de tipo continuo; si ocurre, por redondeos, se eliminan.

Nota: El test puede aplicarse al caso en que la distribución a contrastar dependa de parámetros, estimando estos previamente, pero sólo si las estimaciones se obtienen de datos independientes a los que se usan para el contraste. Si las estimaciones se obtienen de la propia muestra, la distribución del estadístico se altera de manera impredecible y se debe tratar cada caso de manera individual. Por ello, en tales casos se usa el test χ^2, como en el Ejemplo 9.1.1.

Ejemplo 9.1.2: *Se supone que el tiempo de reacción a un determinado compuesto se distribuye según una* $\mathcal{N}(10.5;\ 0.15^2)$. *Contrastar si los siguientes datos, obtenidos en un muestreo aleatorio simple de 10 individuos a los que se ha administrado el compuesto, proporcionan evidencia para rechazar esta hipótesis:*

$$10.39 \quad 10.66 \quad 10.12 \quad 10.32 \quad 10.25 \quad 10.52 \quad 10.83 \quad 10.72 \quad 10.28 \quad 10.35$$

La variable observada y la hipótesis a contrastar son:

$$X = \textit{Tiempo de reacción al compuesto.} \qquad H_0 : X \to \mathcal{N}(10.5;\ 0.15^2).$$

Puesto que la distribución hipotética es de tipo continuo y está totalmente especificada, podemos aplicar el test de Kolmogorov-Smirnov.

Para ello, disponemos los datos en orden creciente, $x_{(1)} < \cdots < x_{(10)}$, y calculamos las diferencias $F^*(x_{(i)}) - F_0(x_{(i)})$ y $F_0(x_{(i)}) - F^*(x_{(i)}^-)$, que proporcionan el valor del estadístico de contraste:

- $F_0(x_{(i)}) = P_{H_0}(X \leq x_{(i)}) = P\left(Z \leq \dfrac{x_{(i)} - 10.5}{0.15}\right), \quad Z \to \mathcal{N}(0,1),$
- $F^*(x_{(i)}) = \dfrac{i}{10},$
- $F^*(x_{(i)}^-) = \dfrac{i-1}{10}.$

$x_{(i)}$	$F^*(x_{(i)})$	$F_0(x_{(i)})$	$F^*(x_{(i)}) - F_0(x_{(i)})$	$F_0(x_{(i)}) - F^*(x_{(i)}^-)$
10.12	0.1	0.00565	0.09435	0.00565
10.25	0.2	0.04779	0.15221	−0.05221
10.28	0.3	0.07123	0.22877	−0.12877
10.32	0.4	0.11507	0.28493	−0.18493
10.35	0.5	0.15865	0.34135	−0.24135
10.39	0.6	0.23168	**0.36832**	−0.26832
10.52	0.7	0.55303	0.14697	−0.04697
10.66	0.8	0.85694	−0.05694	**0.15694**
10.72	0.9	0.92877	−0.02877	0.12877
10.83	1	0.98609	0.01391	0.08609

$$D_{exp} = \text{máx}\left\{ \underset{x_i}{\text{máx}}\,[F^*(x_i) - F_0(x_i)],\ \underset{x_i}{\text{máx}}\,[F_0(x_i) - F^*(x_i^-)] \right\} = 0.36832.$$

Tema 9. Contrastes de hipótesis no paramétricos A. Hermoso Carazo
└9.1. Problema de bondad de ajuste
└9.1.2. Test de Kolmogorov-Smirnov

La tabla del estadístico de Kolmogorov-Smirnov para $n = 10$ proporciona el p-nivel asociado a los datos:

$$P_{H_0}\big(D(X_1, \ldots, X_{10}) \geq 0.36866\big) = 0.1$$

$$\Downarrow$$

$$p - nivel = P_{H_0}\big(D(X_1, \ldots, X_{10}) \geq 0.36832\big) > 0.1,$$

y concluimos que:

Los datos no proporcionan evidencia para rechazar la hipótesis de que el tiempo de reacción es $\mathcal{N}(10.5,\ 0.15^2)$, para cualquier nivel de significación menor o igual que 0.1.

9.2. Problema de localización

Los tests de localización resuelven problemas de contraste relativos a medidas de posición; concretamente, a los cuantiles de la distribución considerada. Aquí nos centraremos en contrastes sobre la mediana.

Consideraremos una muestra aleatoria simple $(X_1, \ldots, X_n)$ de una variable aleatoria X con distribución desconocida, y notando M_X a la mediana, se trata de resolver los siguientes problemas de contraste:

$$\begin{array}{lll} H_0 : M_X = m & \qquad H_0 : M_X = m & \qquad H_0 : M_X = m \\ H_1 : M_X \neq m & \qquad H_1 : M_X > m & \qquad H_1 : M_X < m \end{array}$$

Para ello, describiremos el *test de los signos de Fisher* y el *test de los rangos signados de Wilcoxon*, ambos aplicables sólo a *distribuciones de tipo continuo.*

9.2.1. Test de los signos (Fisher)

Dada una muestra aleatoria simple, $(X_1, \ldots, X_n)$, de una variable X con *distribución de tipo continuo*, se pretende resolver los problemas:

$$\begin{array}{lll} H_0 : M_X = m & \qquad H_0 : M_X = m & \qquad H_0 : M_X = m \\ H_1 : M_X \neq m & \qquad H_1 : M_X > m & \qquad H_1 : M_X < m \end{array}$$

El test se basa en el hecho de que si P_X es una distribución de tipo continuo, $P(X = M_X) = 0$ y $P(X < M_X) = P(X > M_X) = 1/2$. Por tanto, al tomar una muestra aleatoria simple de X, cabe esperar que *aproximadamente* la mitad de los datos sean menores que M_X y la mitad mayores.

Para resolver los problemas de contraste mencionados, con $H_0 : M_X = m$, supondremos que todas las observaciones muestrales son distintas del hipotético valor de la mediana, m. Esta suposición es razonable porque, al ser X de tipo continuo, $P(X = m) = 0$. En caso de que, por redondeo, alguna observación sea igual a m, no se considera y se rehace la muestra.

Estadístico de contraste:

$T(X_1, \ldots, X_n) =$ *Número de observaciones muestrales X_i mayores que m*.

Justificación de los tests: El estadístico $T(X_1, \ldots, X_n)$ toma valores de 0 a n, y si H_0 es cierta, es esperable que tome un valor intermedio, próximo a $n/2$. Por tanto, *valores grandes o pequeños de $T(X_1, \ldots, X_n)$ deben conducir al rechazo de H_0*.

En concreto, la región de rechazo del test depende de la alternativa:

- $H_1 : M_X > m \rightarrow P_{H_1}(X > m) > P_{H_1}(X > M_X) = 1/2$.

  Por tanto, si H_1 es cierta, se espera que el número de observaciones muestrales mayores que m sea grande. Entonces, se rechaza H_0 a favor de H_1 si $T_{exp} = T(x_1, \ldots, x_n)$ es grande.
- $H_1 : M_X < m \rightarrow P_{H_1}(X > m) < P_{H_1}(X > M_X) = 1/2$.

  Similarmente, ahora se rechaza H_0 si T_{exp} es pequeño.
- $H_1 : M_X \neq m$.

  Se rechaza H_0 para valores grandes o pequeños de T_{exp}.

En cualquier caso, el valor crítico que determina la región de rechazo se calcula imponiendo un nivel de significación o tamaño, a partir de la distribución del estadístico bajo H_0.

Alternativa	*Región de rechazo (n.s. α)* $T \rightarrow B(n, 1/2)$	$p-nivel$ $T \rightarrow B(n, 1/2)$
$H_1 : M_X > m$	$T_{exp} \geq k$ $P(T \geq k) \leq \alpha$	$P(T \geq T_{exp})$
$H_1 : M_X < m$	$T_{exp} \leq k$ $P(T \leq k) \leq \alpha$	$P(T \leq T_{exp})$
$H_1 : M_X \neq m$	$T_{exp} \leq k$ ó $T_{exp} \geq n-k$ $P(T \leq k) \leq \alpha/2$	$2P(T \leq T_{exp})$, si $T_{exp} \leq n/2$ $2P(T \geq T_{exp})$, si $T_{exp} \geq n/2$

Ejemplo 9.2.1: *Una empresa que comenzaba su actividad diaria a las 9h. ha cambiado su horario para abrir a las 8h. y se pregunta si ello ha afectado significativamente al retraso de sus empleados. Es aceptable pensar que la forma de la distribución de los retrasos no ha variado con el cambio de horario, pero se teme que se haya desplazado a la derecha, lo que supondría un incremento del tiempo perdido. Se sabe que la mediana de los retrasos de los empleados era inicialmente de 5 minutos. Con el nuevo horario se selecciona a 12 empleados y se observa, en determinados días, los siguientes retrasos (en minutos):*

$$2.5, \quad 1.2, \quad 7, \quad 1.8, \quad 8.3, \quad 6.8, \quad 5.2, \quad 3.4, \quad 4.7, \quad 6.2, \quad 9.1, \quad 5.2$$

A partir de estos datos, contrastar la hipótesis de que la distribución de los retrasos no ha variado con el cambio de horario.

Ya que las distribuciones de los retrasos tienen la misma forma antes y después del cambio de horario, si hay diferencia entre ellas, será debida a su localización, que está determinada por el valor de la mediana. Se trata, por tanto de un problema de localización de la mediana de la variable:

$$X: \textit{ Retraso de los empleados con el cambio de horario.}$$

Ya que el temor del empresario es que la distribución se haya desplazado a la derecha con el cambio de horario, el problema de contraste es:

$$H_0: M_X = 5$$
$$H_1: M_X > 5.$$

Tema 9. Contrastes de hipótesis no paramétricos A. Hermoso Carazo
9.2. Problema de localización
9.2.1. Test de los signos (Fisher)

Como la variable X es de tipo continuo, puede aplicarse el *test de los signos* que, para el problema planteado, *rechaza H_0 si hay un alto número de observaciones muestrales mayores que 5*:

$$T(X_1, \ldots, X_{12}): \text{ N}^o \textit{ de observaciones muestrales mayores que } 5 \underset{H_0}{\longrightarrow} B(12, 1/2).$$

El valor de este estadístico para la muestra observada es $T_{exp} = 7$ y, por tanto, el p-nivel de los datos observados es:

$$p - nivel = P_{H_0}\big(T(X_1, \ldots, X_{12}) \geq 7\big) = 0.3872.$$

Puesto que este valor es grande, se acepta H_0:

Los datos no aportan evidencia para concluir que el cambio de horario desplaza a la derecha la distribución de los retrasos.

9.2.2. Test de Wilcoxon de los rangos signados

El test de los signos ignora las distancias de las observaciones al valor propuesto de la mediana, m, considerando únicamente si cada observación es mayor o menor que m. Para distribuciones, además de continuas, *simétricas*, Wilcoxon propuso un test que sí tiene en cuenta dichas distancias y, al usar más información, es más potente que el anterior.

Se trabaja, como es habitual, con $(X_1, \ldots, X_n)$, una muestra aleatoria simple de una variable aleatoria X con *distribución de tipo continuo y simétrica*, y se pretende resolver los siguientes problemas:

$$H_0 : M_X = m \qquad H_0 : M_X = m \qquad H_0 : M_X = m$$
$$H_1 : M_X \neq m \qquad H_1 : M_X > m \qquad H_1 : M_X < m$$

El test de Wilcoxon se fundamenta en el hecho de que, al ser M_X el centro de simetría de la distribución, es esperable que los valores muestrales $X_1, \ldots, X_n$ se distribuyan simétricamente en torno a M_X. Así, bajo $H_0 : M_X = m$, es esperable que para cada dato a la derecha de m haya otro a la izquierda *aproximadamente* a la misma distancia.

Para contrastar este hecho, se trabaja con el siguiente estadístico.

Estadístico de contraste: Se supondrá, como en el test de Fisher, que todas las observaciones muestrales son distintas del hipotético valor de la mediana, m (si alguna es igual se elimina de la muestra) y se consideran las diferencias respecto de dicho valor, $D_i = X_i - m \neq 0, \; i = 1, \ldots, n$.

- Se disponen en orden creciente los valores absolutos, $|D_i|, \; i = 1, \ldots, n$.
- Se anota el rango (de 1 a n) de cada $|D_i|$ en la ordenación. Si hay empates, se asigna a cada uno el promedio de los rangos correspondientes.

Estadísticos de Wilcoxon:

$T^+(X_1, \ldots, X_n) =$ *Suma de los rangos de las D_i positivas.*

$T^-(X_1, \ldots, X_n) =$ *Suma de los rangos de las D_i negativas.*

Ya que $T^+ + T^- = \sum_{i=1}^{n} i = n(n+1)/2$, cada uno está determinado por el otro, y los tests se basan indistintamente en T^+ o T^-. Aquí los describimos en términos de $T^+ = 0, \ldots, n(n+1)/2$.

Justificación de los tests: Bajo $H_0 : M_X = m$, es esperable que para cada D_i positiva haya una negativa, aproximadamente del mismo tamaño. Así, los valores absolutos serán aproximadamente iguales y ocuparán lugares consecutivos en la ordenación. Los signos del conjunto estarán alternados.

Por lo tanto, bajo H_0, es esperable que T^+ y T^- sean aproximadamente iguales o, lo que es lo mismo, que T^+ tome un valor intermedio, ni grande ni pequeño. Esto es, *valores grandes o pequeños del estadístico T^+ deben conducir al rechazo de H_0*.

La región de rechazo concreta depende de la hipótesis alternativa:

- $H_1 : M_X > m \to P_{H_1}(X > m) > P_{H_1}(X > M_X) = 1/2$.

  Entonces, si H_1 es cierta, se esperan más diferencias positivas que negativas y, a la vez, se espera que las positivas sean de mayor tamaño. Por tanto, se rechaza H_0 si $T^+_{exp} = T^+(x_1, \ldots, x_n)$ es un valor grande.

- $H_1 : M_X < m \to P_{H_1}(X > m) < P_{H_1}(X > M_X) = 1/2$.

  En este caso, se rechaza H_0 si T^+_{exp} es pequeño.

- $H_1 : M_X \neq m$:

  Se rechaza H_0 si el valor experimental, T^+_{exp}, es grande o pequeño.

Distribución de $T^+(X_1, \ldots, X_n)$ bajo H_0: Es simétrica y está tabulada. Para valores grandes de n, la distribución es aproximadamente normal con

$$E[T^+] = \frac{n(n+1)}{4} \text{ y } Var[T^+] = \frac{n(n+1)(2n+1)}{24}.$$

Tema 9. Contrastes de hipótesis no paramétricos A. Hermoso Carazo
9.2. Problema de localización
9.2.2. Test de Wilcoxon de los rangos signados

Alternativa	Región de rechazo (n.s. α) $T^+ \equiv T^+(X_1, \ldots, X_n)$	$p-nivel$ $T^+ \equiv T^+(X_1, \ldots, X_n)$
$H_1 : M_X > m$	$T^+_{exp} \geq k$ $P_{H_0}(T^+ \geq k) \leq \alpha$	$P_{H_0}(T^+ \geq T^+_{exp})$
$H_1 : M_X < m$	$T^+_{exp} \leq k$ $P_{H_0}(T^+ \leq k) \leq \alpha$	$P_{H_0}(T^+ \leq T^+_{exp})$
$H_1 : M_X \neq m$	$T^+_{exp} \leq k$ ó $T^+_{exp} \geq \frac{n(n+1)}{2} - k$ $P_{H_0}(T^+ \leq k) \leq \alpha/2$	$2P_{H_0}(T^+ \leq T^+_{exp})$, si $T^+_{exp} \leq \frac{n(n+1)}{4}$ $2P_{H_0}(T^+ \geq T^+_{exp})$, si $T^+_{exp} \geq \frac{n(n+1)}{4}$

Ejemplo 9.2.2: *A partir de los datos del Ejemplo 9.2.1, y suponiendo que la distribución de los retrasos es simétrica, contrastar la hipótesis de que ésta no varía con el cambio de horario.*

La simetría en la distribución permite usar el *test de los rangos signados*:

$$T^+(X_1, \ldots, X_{12}): \text{ Suma de los rangos correspondientes a } D_i \text{ positivas.}$$

Ya que $H_1 : M_X > 5$, se rechazará H_0 si los datos proporcionan un alto valor de $T^+(X_1, \ldots, X_{12})$. Calculamos dicho valor, notando $d_i = x_i - 5$:

d_i :	-2.5	-3.8	2	-3.2	3.3	1.8	0.2	-1.6	-0.3	1.2	4.1	0.2
$\|d_i\|$:	2.5	3.8	2	3.2	3.3	1.8	0.2	1.6	0.3	1.2	4.1	0.2

$\|d_i\|$ ord.	**0.2**	**0.2**	0.3	**1.2**	1.6	**1.8**	**2**	2.5	3.2	**3.3**	3.8	**4.1**
Signo	+	+	−	+	−	+	+	−	−	+	−	+
Rango	**1**	**2**	3	**4**	5	**6**	**7**	8	9	**10**	11	**12**

$$T^+_{exp} = 1 + 2 + 4 + 6 + 7 + 10 + 12 = 42.$$

Calculamos ahora el p-nivel, usando la tabla de Wilcoxon para $n = 12$:

$$p - nivel = P_{H_0}\left(T^+(X_1, \ldots, X_{12}) \geq 42\right) = 0.425.$$

Se concluye, como con el test de los signos, que *los datos no dan evidencia para decidir que el cambio de horario aumenta la mediana de los retrasos.*

Nota: Los tests de localización se usan a menudo para contrastar la homogeneidad de las distribuciones (continuas) correspondientes a dos *muestras apareadas* (observaciones bidimensionales en cada individuo de la muestra; por ejemplo, antes y después de un tratamiento), siempre que se tenga constancia de que ambas distribuciones tienen la misma forma funcional, pero pueden estar desplazadas.

En este caso, la igualdad de las distribuciones equivale a la igualdad de las medianas o, lo que es lo mismo, a que la variable diferencia, $D = X - Y$, tiene mediana cero:

$$P(D > 0) = P(X > Y) = P(X < Y) = P(D < 0)$$
$$\Leftrightarrow P(D > 0) = P(D < 0) = 1/2.$$

El problema se resuelve tomando la muestra formada por las diferencias y contrastando que la mediana de esta distribución es nula.

Problema 6

La siguiente tabla presenta las presiones sanguíneas sistólicas de 10 individuos antes y después de haber dejado la bebida.

A	140	165	160	160	175	190	170	175	155	160
D	145	150	150	160	170	175	160	165	145	170

¿Se puede afirmar a partir de los datos que el abandono de la bebida no disminuye la presión sanguínea? ¿Bajo qué hipótesis?

Se trata de un contraste de igualdad de las distribuciones correspondientes a dos muestras apareadas. Suponiendo que la forma de la distribución de la presión es la misma antes y depués de dejar la bebida, el problema se puede resolver mediante un test de localización de la mediana de las diferencias.

$X = A - D$*: Diferencia de presión antes y después de dejar la bebida.*

$$H_0 : M_X = 0$$
$$H_1 : M_X > 0.$$

Aunque las medidas suelan darse exactas por redondeo, la presión (y, por tanto, X) es una variable continua y podemos aplicar los tests de localización.

$$H_0 : M_X = 0$$
$$H_1 : M_X > 0.$$

Test de los signos:

Observaciones de $X = A - D$ (eliminamos el 0, hipotético valor de M_X):

$$-5,\ 15,\ 10,\ 5,\ 15,\ 10,\ 10,\ 10,\ -10 \longrightarrow n = 9$$

$T(X_1, \ldots, X_9) =$ *Número de observaciones muestrales mayores que 0* $\underset{H_0}{\longrightarrow} B(9, 1/2).$

$$T_{exp} = 7 \longrightarrow p - nivel = P_{H_0}(T \geq 7) = 0.0898.$$

Se acepta H_0 *a niveles de significación inferiores a 0.0898, y se rechaza a niveles superiores.*

$$H_0 : M_X = 0$$
$$H_1 : M_X > 0.$$

Test de los rangos signados: $D = X - 0 = X \longrightarrow$ *simétrica.*

$T^+(X_1, \ldots, X_n)$ = *Suma de rangos correspondientes a diferencias positivas.*

$d_i = -5,\ 15,\ 10,\ 0,\ 5,\ 15,\ 10,\ 10,\ 10,\ -10 \longrightarrow$ se elimina el valor 0.

$\lvert d_i \rvert$	5	5	10	10	10	10	10	15	15
Signo	+	−	+	+	+	+	−	+	+
Rango	1.5	1.5	5	5	5	5	5	8	9

$$T^+_{exp} = 1.5 + 5 + 5 + 5 + 5 + 8 + 9 = 38.5.$$

$$p - nivel = P(T^+ \geq 38.5) = P(T^+ \geq 39) = 0.027.$$

Se acepta H_0 a niveles de significación inferiores a 0.027, y se rechaza a niveles superiores.

Nota: Para niveles de significación entre 0.027 y 0.0898 el test de los rangos signados conduce al rechazo de H_0 y el test de los signos a la aceptación. Ya que el primero es más potente, si se está razonablemente seguro de que la distribución de las diferencias es simétrica, éste es el que debe usarse.

9.3. Problema de independencia

A menudo se presenta el problema de estudiar conjuntamente dos variables en los individuos de una población y analizar si se comportan de forma independiente o, por el contrario, existe algún tipo de relación entre ellas.

Aunque existen diversos tests para resolver este problema, aquí sólo describimos el test χ^2, para variables categóricas, que puede adaptarse, como el de bondad de ajuste, para contrastar la independencia de variables aleatorias, sin más que particionar sus rangos de valores en un número finito de subconjuntos.

9.3.1. Test χ^2 de independencia

Supongamos que los individuos de una población se clasifican de acuerdo a dos características, X e Y, en m y k clases exhaustivas y mutuamente excluyentes, $A_1, \ldots, A_m$ y $B_1, \ldots, B_k$, respectivamente.

Se pretende contrastar si dichas características se comportan de forma independiente o, por el contrario, el valor de cualquiera de ellas condiciona de alguna forma el valor de la otra. La hipótesis de independencia se especifica como:

$$H_0 : P(X = A_i,\ Y = B_j) = P(X = A_i)P(Y = B_j), \quad \forall i = 1, \ldots, m;\ j = 1 \ldots, k.$$

Para resolver el problema se toma una muestra aleatoria simple de n individuos y denotamos:

N_{ij}: *número de individuos que presentan la modalidad A_i de X y la modalidad B_j de Y*, $i = 1, \ldots, m;\ j = 1, \ldots, k.$

La clasificación de los individuos por sus modalidades se suele presentar en la denominada *tabla de contingencia*.

X \ Y	B_1	$\cdots$	B_j	$\cdots$	B_k	Totales
A_1	N_{11}	$\cdots$	N_{1j}	$\cdots$	N_{1k}	$N_{1.}$
$\vdots$	$\cdots$		$\cdots$		$\cdots$	$\vdots$
A_i	N_{i1}	$\cdots$	N_{ij}	$\cdots$	N_{ik}	$N_{i.}$
$\vdots$	$\cdots$		$\cdots$		$\cdots$	$\vdots$
A_m	N_{m1}	$\cdots$	N_{mj}	$\cdots$	N_{mk}	$N_{m.}$
Totales	$N_{.1}$	$\cdots$	$N_{.j}$	$\cdots$	$N_{.k}$	n

$$N_{i.} = \sum_{j=1}^{k} N_{ij}$$
$$N_{.j} = \sum_{i=1}^{m} N_{ij}.$$

El test χ^2 de independencia usa una medida de discrepancia entre las *frecuencias observadas* para cada par de modalidades (N_{ij}) y las *frecuencias esperadas* en caso de independencia (E_{ij}) similar a la del test de ajuste:

$$\sum_{i=1}^{m}\sum_{j=1}^{k}\frac{(N_{ij}-E_{ij})^2}{E_{ij}}.$$

Cálculo de las frecuencias esperadas: Si H_0 es cierta, la frecuencia esperada en el cruce de modalidades (A_i, B_j) es

$$E_{ij} = nP_{H_0}(X = A_i,\ Y = B_j) = nP(X = A_i)P(Y = B_j).$$

Ahora bien, como las probabilidades marginales son desconocidas, se estiman por máxima verosimilitud a partir de los datos:

- $P(X = A_i)$, $i = 1, \ldots, m$, se estima por $N_{i.}/n$, frecuencia relativa observada de individuos en A_i.[1] Puesto que $\sum_{i=1}^{m} P(X = A_i) = 1$, sólo hay que estimar $m-1$ de estas probabilidades.
- $P(Y = B_j)$, $j = 1, \ldots, k$, se estima por la frecuencia relativa observada en B_j, $N_{.j}/n$. El número de probabilidades estimadas es $k-1$.

Por tanto, E_{ij} se estima por $n\frac{N_{i.}}{n}\frac{N_{.j}}{n} = \frac{N_{i.}N_{.j}}{n}$, y *el estadístico de contraste y su distribución asintótica bajo H_0 son:*

$$\chi^2(N_{ij}) = \sum_{i=1}^{m}\sum_{j=1}^{k}\frac{\left(N_{ij}-\frac{N_{i.}N_{.j}}{n}\right)^2}{\frac{N_{i.}N_{.j}}{n}} \xrightarrow[n\to+\infty]{L} \chi^2\big((m-1)(k-1)\big) \quad (bajo\ H_0).$$

Nota: Los grados de libertad de $\chi^2(N_{ij})$ se deben a que las variables N_{ij} tienen $mk-1$ grados de libertad (mk variables y una restricción, $\sum_{i=1}^{m}\sum_{j=1}^{k} N_{ij} = n$) y se han estimado $(m-1)+(k-1)$ parámetros.

[1] $N_{i.} \to B(n, p_i)$, $p_i = P(X = A_i)$. Entonces, el e.m.v. de p_i es $N_{i.}/n$.

Ya que $\chi^2(N_{ij})$ mide la discrepancia entre frecuencias observadas y frecuencias esperadas (estimadas) bajo H_0, *se rechazará H_0 si su valor es grande, aceptándose en caso contrario.*

Test de tamaño α (aproximado): Para valores de n grandes, el test de tamaño α es:

$$\varphi(N_{ij}) = \begin{cases} 1, & \chi^2(N_{ij}) \geq \chi^2_{(m-1)(k-1);\ \alpha} \\ 0, & \chi^2(N_{ij}) < \chi^2_{(m-1)(k-1);\ \alpha}. \end{cases}$$

Si no se especifica un tamaño o nivel de significación, se calcula el *p-nivel asociado a los datos,*

$$P\left(\chi^2((m-1)(k-1)) \geq \chi^2_{exp}\right), \quad \chi^2_{exp} = \chi^2(n_{ij}),$$

y se rechaza H_0 si este valor es pequeño.

Nota: Ya que el test es asintótico, para que sea aceptable se aconseja tomar n lo suficientemente grande para que todas las frecuencias esperadas sean mayores o iguales que 2 y no más del 20 % menores que 5.

Aplicación al contraste de independencia de variables aleatorias:

El test puede aplicarse para contrastar la independencia de dos variables aleatorias, sin más que agrupar sus valores en un número finito de categorías, aunque esto, como ya comentamos en el test de ajuste, supondrá una pérdida de información.

Concretamente, si se quiere contrastar la independencia de dos variables aleatorias X, Y:

$$P(X \in A,\ Y \in B) = P(X \in A)P(Y \in B), \quad \forall A, B \in \mathcal{B},$$

se particiona el rango de valores de X en m subconjuntos, $A_1, \cdots, A_m$, y el de Y en k subconjuntos, $B_1, \cdots, B_k$, todos de probabilidad no nula. Clasificando las observaciones muestrales según las distintas modalidades, se aplica el test χ^2 para contrastar:

$$H_0 : P(X \in A_i,\ Y \in B_j) = P(X \in A_i)P(Y \in B_j), \quad \forall i = 1, \ldots, m;\ j = 1, \ldots, k.$$

Ejemplo 9.3.1: *Para estudiar si el grupo sanguíneo de los individuos tiene relación con la predisposición a la diabetes, se han seleccionado al azar 400 sujetos a los que se ha determinado el grupo sanguíneo y el nivel de glucosa en sangre en idénticas condiciones experimentales. Clasificando la segunda medida en tres niveles, los resultados han sido:*

Grupo \ *Nivel de glucosa*	*Bajo*	*Medio*	*Alto*
O	137	86	35
A	42	23	11
B	19	17	7
AB	14	7	2

Contrastar, al nivel de significación 0.05, si las variables son independientes.

Se aplica el test χ^2 para contrastar la independencia de las variables *grupo sanguíneo* y *nivel de glucosa en sangre*.

Puesto que las variables están clasificadas en 4 y 3 categorías, respectivamente, el estadístico de contraste tiene distribución (aproximada) $\chi^2(6)$.

Por lo tanto, al nivel especificado, se rechazará H_0 si los datos dan un valor $\chi^2_{\text{exp}} \geq \chi^2_{6;\ 0.05} = 12.5916$.

Construimos la tabla de frecuencias observadas y esperadas $(n_{i.}n_{.j}/n)$:

Grupo \ *Nivel de glucosa*	*Bajo*	*Medio*	*Alto*	Totales
O	137 136.74	86 85.785	35 35.475	258
A	42 40.28	23 25.27	11 10.45	76
B	19 22.79	17 14.2975	7 5.9125	43
AB	14 12.19	7 7.6475	2 3.1625	23
Totales	212	133	55	400

Observamos que todas las frecuencias esperadas son mayores que 2 y sólo una (menos del 20 %) menor que 5; por lo tanto, podemos aplicar el test χ^2.

$$\chi^2_{exp} = \frac{(137-136.74)^2}{136.74} + \frac{(86-85.785)^2}{85.785} + \cdots + \frac{(2-3.1625)^2}{3.1625} = 2.41.$$

Puesto que $\chi^2_{exp} < \chi^2_{6;\ 0.05}$, se deduce que *los datos no dan evidencia para rechazar la hipótesis de independencia entre el grupo sanguíneo y el nivel de glucosa al nivel de significación 0.05.*

De hecho, el p-nivel es $P\left(\chi^2(N_{ij}) > 2.41\right) \in (0.85, 0.9)$, lo que lo que avala perfectamente la hipótesis nula.

9.4.1. Test χ^2 de homogeneidad

Supongamos que la característica que se desea estudiar presenta k modalidades, $A_1, \ldots, A_k$, y sean $X_1, \ldots, X_m$ las variables que describen dicha característica en las m poblaciones.

Se pretende contrastar la hipótesis de que la característica se distribuye de igual forma en las m poblaciones, lo que se especifica como:

$$H_0 : P(X_1 = A_j) = P(X_2 = A_j) = \cdots = P(X_m = A_j), \quad j = 1, \ldots, k.$$

Para resolver este problema, se toma una muestra aleatoria de n_i individuos en cada población, *todas independientes*. Clasificando cada individuo muestreado según su modalidad, denotamos:

N_{ij} : número de individuos de la muestra i-ésima que presentan la modalidad A_j; $i = 1, \ldots, m$, $j = 1, \ldots, k$.

Estas frecuencias se presentan en una tabla de contingencia en la que las filas representan las m poblaciones y las columnas las k modalidades:

El test de homogeneidad usa la medida usual de discrepancia de los tests χ^2 entre frecuencias observadas (N_{ij}) y esperadas bajo H_0 (E_{ij}):

$$\sum_{i=1}^{m} \sum_{j=1}^{k} \frac{(N_{ij} - E_{ij})^2}{E_{ij}}.$$

Cálculo de las frecuencias esperadas: Si H_0 es cierta, y p_j denota la probabilidad, común en todas las poblaciones, de que un individuo presente la modalidad A_j, el número esperado de individuos de la muestra i-ésima con la modalidad A_j es:

$$E_{ij} = n_{i.} P_{H_0}(X_i = A_j) = n_{i.} p_j.$$

Ahora bien, como p_j es desconocida, se sustituye por $\dfrac{N_{.j}}{n}$, que es su estimador máximo verosímil,[2] y, por tanto, E_{ij} se estima por $n_{i.}\dfrac{N_{.j}}{n}$.

El estadístico de contraste y su distribución asintótica bajo H_0, cuando todos los tamaños muestrales tienden a infinito, son:

$$\boxed{\chi^2(N_{ij}) = \sum_{i=1}^{m} \sum_{j=1}^{k} \frac{\left(N_{ij} - \frac{n_{i.}N_{.j}}{n}\right)^2}{\frac{n_{i.}N_{.j}}{n}}} \xrightarrow[\substack{n_{i.} \to +\infty \\ i=1,\ldots,m}]{L} \chi^2((m-1)(k-1)).$$

[2] Bajo H_0, $N_{.j} \to B(n, p_j)$ y su e.m.v es $N_{.j}/n$.

Nota: Como ya se ha indicado, las mk variables N_{ij} tienen una restricción por cada fila y, por tanto, $mk - m$ grados de libertad. Los grados de libertad del estadístico se obtienen restando a éstos el número de parámetros estimados, $k-1$ (ya que $\sum_{j=1}^{k} p_j = 1$, sólo es preciso estimar $k-1$ de ellos).

El estadístico $\chi^2(N_{ij})$ mide la discrepancia entre frecuencias observadas y frecuencias esperadas (estimadas) bajo H_0, de manera que *se rechazará H_0 si su valor es grande, aceptándose en caso contrario.*

Test de tamaño α (aproximado): Para valores de $n_1, \ldots, n_m$ grandes, el test de tamaño α es:

$$\varphi(N_{ij}) = \begin{cases} 1, & \chi^2(N_{ij}) \geq \chi^2_{(m-1)(k-1);\ \alpha} \\ 0, & \chi^2(N_{ij}) < \chi^2_{(m-1)(k-1);\ \alpha}. \end{cases}$$

Si no se especifica el tamaño o nivel de significación, se calcula el *p-nivel asociado a los datos,*

$$P\left(\chi^2((m-1)(k-1)) \geq \chi^2_{exp}\right), \quad \chi^2_{exp} = \chi^2(n_{ij}),$$

Tema 9. Contrastes de hipótesis no paramétricos A. Hermoso Carazo
└9.4. Problema de homogeneidad
└9.4.1. Test χ^2 de homogeneidad

Nota 1: Debe tomarse al menos 20 observaciones de cada variable ($n_{i.} \geq 20$), todas las frecuencias esperadas deben ser mayores o iguales que 2 y no más del 20 % menores que 5.

Nota 2: La tabla de contingencia usada para resolver el problema es similar a la del test de independencia y el estadístico de contraste se calcula igual en ambos tests. Sin embargo, la interpretación es distinta, ya que corresponden a distintos planteamientos (el diseño de la experiencia es distinto; en el problema de independencia se toma una muestra y se miden las dos características, y en el problema de homogeneidad se toma una muestra de cada población y se mide una misma característica).

Aplicación al contraste de igualdad de distribuciones de variables aleatorias:

El test puede aplicarse para contrastar si un conjunto de m variables aleatorias, $X_1, \ldots, X_m$, tienen la misma distribución:

$$P(X_1 \in A) = \cdots = P(X_m \in A), \quad \forall A \in \mathcal{B}.$$

Para ello, se particiona el conjunto de valores de las variables (común a todas) en k subconjuntos, $A_1, \ldots, A_k$, de probabilidad no nula bajo la distribución hipotéticamente común, y se contrasta la hipótesis:

$$H_0 : P(X_1 \in A_j) = \cdots = P(X_m \in A_j), \quad \forall j = 1, \ldots, k.$$

Tomando una muestra aleatoria simple de cada una de las variables, *todas independientes*, se clasifica cada observación muestral en cada una de las clases $A_1, \ldots, A_k$ y se aplica el test descrito.

Ejemplo 9.4.1: *Contrastar, a partir de los resultados de la siguiente tabla, si los distintos grupos sanguíneos se presentan con la misma frecuencia en tres grupos étnicos diferentes:*

Grupo sanguíneo / *Raza*	*O*	*A*	*B*	*AB*
1	32	11	7	2
2	47	13	17	9
3	23	7	9	6

Se trata de contrastar la homogeneidad de las tres poblaciones frente a la variable *grupo sanguíneo* (categórica, con tres categorías). Debe suponerse que las muestras han sido obtenidas aleatoria e independientemente en cada población y que las tres son independientes.

Puesto que tenemos una tabla de cuatro filas y tres columnas, el estadístico de contraste tiene distribución (aproximada) $\chi^2(6)$.

Como no se especifica tamaño ni nivel de significación, calcularemos el p-nivel asociado a los datos.

Comenzamos construyendo la tabla de contingencia con los valores observados, los totales y los valores esperados bajo la hipótesis de homogeneidad, $n_{i.}n_{.j}/n$.

Raza \ *Grupo sanguíneo*	0	A	B	AB	Totales
1	32 28.98	11 8.8	7 9.377	2 4.83	52
2	47 47.93	13 14.57	17 15.5	9 7.99	86
3	23 25.08	7 7.62	9 8.15	6 4.18	45
Totales	102	31	33	17	183

Observamos que todas las frecuencias esperadas son mayores que 2 y sólo dos (menos del 20 %) menores que 5. Por lo tanto, podemos aplicar el test χ^2.

$$\chi^2_{exp} = \frac{(32-28.98)^2}{28.98} + \frac{(11-8.8)^2}{8.8} + \cdots + \frac{(6-4.18)^2}{4.18} = 4.691$$

y determinamos el p-nivel:

$$0.55 < P\left(\chi^2(N_{ij}) > 4.691\right) < 0.6.$$

Por tanto, se acepta la hipótesis de homogeneidad; esto es, *los datos avalan que las tres razas presentan los distintos grupos sanguíneos con la misma frecuencia.*

A_i	0	1	2	3	4	5	6	≥ 7
p_i^0	0.1353	0.2707	0.2707	0.1804	0.0902	0.0361	0.0120	0.0045
$200p_i^0$	27.0671	54.1341	54.1341	36.0894	18.0447	7.2179	2.406	0.9068

$\downarrow$ Agrupamos las tres últimas clases

$A_i,\ i=1,\ldots 6$	0	1	2	3	4	≥ 5
$F.E. = 200p_i^0$	27.0671	54.1341	54.1341	36.0894	18.0447	10.5306
$F.O. = n_i$	22	53	58	39	20	8

$$\Downarrow$$

$$\chi^2_{exp} = \sum_{i=1}^{6} \frac{(n_i - np_i^0)^2}{np_i^0} = \frac{(22-27.0671)^2}{27.0671} + \ldots + \frac{(8-10.5306)^2}{10.5306} = 2.30315.$$

$$\downarrow \chi^2(N_1,\ldots,N_6) \underset{H_0}{\simeq} \chi^2(5)$$

$$p-\text{nivel} = P_{H_0}\left(\chi^2(N_1,\ldots,N_6) > 2.30315\right) \in (0.8, 0.85).$$

Se concluye que los datos no proporcionan evidencia para rechazar la hipótesis de que el número de accidentes diarios sigue una distribución $\mathcal{P}(2)$.

Problema 2

Una tela cuadrada tiene 60 defectos de fabricación. Con objeto de analizar la distribución de los defectos en la superficie de la tela, se ha dividido ésta en 9 zonas cuadradas exactamente iguales, observándose los siguientes defectos en cada zona:

8	7	3
5	9	11
6	4	7

Contrastar, a partir de estos datos, si los defectos se distribuyen uniformemente en toda la superficie o, por el contrario, siguen algún patrón de ocurrencia.

Se trata de contrastar que la distribución de la variable *localización de un defecto en la superficie de la tela* es uniforme. Es una variable continua cuyos valores se han clasificado en 9 categorías (cuadrículas). Así, el problema se resuelve contrastando que la probabilidad de que un defecto caiga en cada cuadrícula específica, $C_1, \dots, C_9$, es $1/9$.

Relación 9. Contrastes de hipótesis no paramétricos A. Hermoso Carazo
└ Problema 2

X : cuadrícula en la que cae un defecto

$$H_0 : P(X \in C_i) = 1/9, \ i = 1, \dots, 9$$
$$H_1 : \exists i = 1, \dots 9, \ / \ P(X \in C_i) \neq 1/9$$

$\downarrow$

Variable categórica $\longrightarrow$ *Test* χ^2

Frecuencias esperadas $(n = 60) \longrightarrow \ np_i^0 = 60 \times 1/9 \ (\geq 5), \ \ i = 1, \dots, 9.$

$$\Downarrow$$

$$\chi^2_{exp} = \sum_{i=1}^{9} \frac{(n_i - np_i^0)^2}{np_i^0} = \sum_{i=1}^{9} \frac{(n_i - 60/9)^2}{60/9} = 7.5.$$

$$\downarrow \chi^2(N_1, \dots, N_9) \underset{H_0}{\simeq} \chi^2(8)$$

$$p - \text{nivel} = P_{H_0}\left(\chi^2(N_1, \dots, N_9) > 7.5\right) \in (0.45, 0.5).$$

Se concluye que los datos no proporcionan evidencia para rechazar la hipótesis de que los defectos se distribuyen uniformemente por toda la superficie.

Problema 3

Un modelo genético indica que la distribución de una población de hombres y mujeres, daltónicos o no, se ajusta a probabilidades de la forma:

	Hombres	Mujeres
No daltónicos	$(1-p)/2$	$(1-p^2)/2$
Daltónicos	$p/2$	$p^2/2$

Para comprobar esta teoría se examinaron 2000 individuos de la población, elegidos al azar, obteniéndose los siguientes resultados:

	Hombres	Mujeres
No daltónicos	894	1015
Daltónicos	81	10

Contrastar, mediante el test de la χ^2, si esta muestra concuerda con el modelo teórico.

Puesto que el modelo depende del parámetro p, desconocido, comenzamos estimándolo a partir de la información muestral, dada por el valor de las siguientes variables:

- N_1: *Número de hombres no daltónicos en una muestra aleatoria de 2000.*
- N_2: *Número de hombres daltónicos en una muestra aleatoria de 2000.*
- N_3: *Número de mujeres no daltónicas en una muestra aleatoria de 2000.*
- N_4: *Número de mujeres daltónicas en una muestra aleatoria de 2000.*

$$(N_1, N_2, N_3) \underset{H_0}{\longrightarrow} M\big(2000;\ (1-p)/2,\ p/2,\ (1-p^2)/2\big).$$

Estimación máximo verosímil de p bajo H_0: $n_1 = 894,\ n_2 = 81,\ n_3 = 1015$

$$* \ L_{n_1,n_2,n_3}(p) = PR_{2000}^{894,81,1015,10}\left(\frac{1-p}{2}\right)^{894}\left(\frac{p}{2}\right)^{81}\left(\frac{1-p^2}{2}\right)^{1015}\left(\frac{p^2}{2}\right)^{10}$$
$$\propto p^{101}(1-p)^{894}(1-p^2)^{1015}.$$

$$* \ \ln L_{n_1,n_2,n_3}(p) \propto 101\ln p + 894\ln(1-p) + 1015\ln(1-p^2).$$

$$* \ \frac{d\ln L_{n_1,n_2,n_3}(p)}{dp} = \frac{101}{p} - \frac{894}{1-p} - \frac{2030p}{1-p^2} = 0 \longrightarrow 3025p^2 + 894p - 101 = 0.$$

$$\Rightarrow p = \begin{cases} 0.08723 \\ -0.38277 \end{cases} \Rightarrow \widehat{p} = 0.08723.$$

Frecuencias esperadas: $2000 \times \widehat{p}_i^0, \quad i = 1,2,3,4$

	Hombres	*Mujeres*
No daltónicos	$2000 \times \dfrac{1-0.08723}{2} = 912.77$	$2000 \times \dfrac{1-0.08723^2}{2} = 992.39$
Daltónicos	$2000 \times \dfrac{0.08723}{2} = 87.23$	$2000 \times \dfrac{0.08723^2}{2} = 7.61$

$$\Downarrow$$

$$\widehat{\chi}^2_{exp} = \sum_{i=1}^{4} \frac{(n_i - n\widehat{p}_i^0)^2}{n\widehat{p}_i^0}$$

$$= \frac{(894-912.77)^2}{912.77} + \frac{(81-87.23)^2}{87.23} + \frac{(1015-992.39)^2}{992.39} + \frac{(10-7.61)^2}{7.61} = 2.0973$$

$$\downarrow \widehat{\chi}^2(N_1,\ldots,N_4) \underset{H_0}{\simeq} \chi^2(2)$$

$$p - \text{nivel} = P_{H_0}\left(\chi^2(N_1,\ldots,N_4) > 2.09731\right) \in (0.35, 0.4).$$

Se concluye que los datos no proporcionan evidencia para rechazar el modelo teórico.

Problema 4

Un laboratorio farmacéutico afirma que uno de sus productos confiere inmunidad a la picadura de insectos durante un tiempo exponencial de media 2.5 horas. Probado el producto en 20 sujetos, en un ambiente con gran número de mosquitos, los instantes (en horas) en que recibieron la primera picadura fueron:

0.01,	0.02	0.03,	0.23,	0.51,	0.74,	0.96,	1.17,	1.46,	1.62,
2.18,	2.25,	2.79,	3.45,	3.83,	3.92,	4.27,	5.43,	5.79,	6.34.

Usando el test de Kolmogorov-Smirnov, contrastar, a partir de estos datos, si puede aceptarse la afirmación del laboratorio.

X : *Tiempo de inmunidad a la picadura* $\quad H_0 : X \to exp(1/2.5)$, $\quad H_1 : X \nrightarrow exp(1/2.5)$

Test de K-S $\longrightarrow$ $D(X_1,\ldots,X_n)=\sup\limits_{x\in\mathbb{R}}\left|F^*_{X_1,\ldots,X_n}(x)-F_0(x)\right|$

$$\downarrow\ x_i\neq x_j,\quad i\neq j$$

$$D_{exp}=\text{máx}\left\{\,\text{máx}_{x_i}\left[F^*(x_i)-F_0(x_i)\right],\ \text{máx}_{x_i}\left[F_0(x_i)-F^*(x_i^-)\right]\,\right\}.$$

$$x_{(1)}<\cdots<x_{(20)}$$

- $F_0(x_{(i)})=P_{H_0}(X\leq x_{(i)})=1-e^{-x_{(i)}/2.5}$,
- $F^*(x_{(i)})=\dfrac{i}{20}$,
- $F^*(x_{(i)}^-)=\dfrac{i-1}{20}$.

$x_{(i)}$	$F^*(x_{(i)})$	$F_0(x_{(i)})$	$F^*(x_{(i)})-F_0(x_{(i)})$	$F_0(x_{(i)})-F^*(x_{(i)}^-)$
0.01	0.05	0.00399	0.04601	0.00399
0.02	0.10	0.00797	0.09203	−0.04203
0.03	0.15	0.01193	0.13807	−0.08807
0.23	0.20	0.08789	0.11211	−0.06211
0.51	0.25	0.18454	0.06546	−0.01546
0.74	0.30	0.25621	0.04379	0.00621
0.96	0.35	0.31887	0.03113	0.01887
1.17	0.40	0.37375	0.02625	0.02375
1.46	0.45	0.44234	0.00766	0.04234
1.62	0.50	0.47691	0.02309	0.02691
2.18	0.55	0.58189	−0.03189	0.08189
2.25	0.60	0.59343	0.00657	0.04343
2.79	0.65	0.67241	−0.02241	0.07241
3.45	0.70	0.74842	−0.04842	0.09842
3.83	0.75	0.78390	−0.03390	0.08390
3.92	0.80	0.79154	0.00846	0.04154
4.27	0.85	0.81877	0.03123	0.01877
5.43	0.90	0.88605	0.01395	0.03605
5.79	0.95	0.90133	0.04867	0.00133
6.34	1	0.92082	0.07918	−0.02918

$$D_{exp}=0.13807\ \Rightarrow\ p-nivel=P_{H_0}\left(D(X_1,\ldots,X_{20})>0.13807\right)$$

$$P_{H_0}\left(D(X_1,\ldots,X_{20})>0.23156\right)=0.2\Rightarrow P_{H_0}\left(D(X_1,\ldots,X_{20})>0.13807\right)>0.2.$$

Los datos no dan evidencia para rechazar H_0. Se acepta que proceden de una distribución exponencial de media 2.5.

Problema 5

Cierta comunidad ha modificado la procedencia del agua destinada al consumo doméstico. Se sabe que, con el antiguo suministro, la distribución de la cantidad de sodio por unidad de volumen de sangre de sus habitantes es simétrica alrededor de 3.24 gr. Tras cierto tiempo, se quiere comprobar si la modificación ha afectado a la concentración de sodio, en el sentido de que su distribución se haya trasladado o no. Para ello, se han realizado 15 análisis con los siguientes resultados (en gr. por unidad):

2.37	2.95	3.4	2.64	3.66	3.18	2.72	3.61
3.87	1.97	1.66	3.72	2.10	1.83	3.03	

¿Se puede afirmar, al nivel de significación 0.1, que la distribución de la cantidad de sodio no ha variado?

X : Gramos de sodio por unidad de volumen de sangre con el nuevo suministro

$$H_0 : M_X = 3.24 \qquad H_1 : M_X \neq 3.24.$$

X simétrica $\longrightarrow$ Test de los rangos signados $T^+ \equiv T^+(X_1, \ldots, X_n)$

$$\varphi(X_1,\ldots,X_n) = \begin{cases} 1, & T^+ \leq k \text{ o } T^+ \geq \frac{n(n+1)}{2} - k \\ 0, & k < T^+ < \frac{n(n+1)}{2} - k \end{cases} \longrightarrow \text{Tamaño} = 2P_{H_0}(T^+ \leq k)$$

$$\downarrow \quad n = 15 \rightarrow \frac{n(n+1)}{2} = 120$$

Nivel de significación $0.1 \rightarrow 2P_{H_0}(T^+ \leq k) \leq 0.1$.

$$\varphi(x_1,\ldots,x_{15}) = \begin{cases} 1, & T^+_{exp} \leq k \text{ o } T^+_{exp} \geq 120 - k \\ 0, & k < T^+_{exp} < 120 - k \end{cases} \qquad P_{H_0}(T^+ \leq k) \leq 0.05.$$

$\downarrow$ Tablas $(n = 15) \rightarrow P_{H_0}(T^+ \leq 30) = 0.047, \quad P_{H_0}(T^+ \leq 31) = 0.053 \Rightarrow k = 30.$

$$\varphi(x_1,\ldots,x_{15}) = \begin{cases} 1, & T^+_{exp} \leq 30 \text{ o } T^+_{exp} \geq 90 \\ 0, & 30 < T^+_{exp} < 90. \end{cases}$$

$d_i = x_i - 3.24$

$= -0.87, -0.29, 0.16, -0.6, 0.42, -0.06, -0.52, 0.37, 0.63, -1.27, -1.58, 0.48, -1.14, -1.41, -0.21$

$\lvert d_i \rvert$	0.06	0.16	0.21	0.29	0.37	0.42	0.48	0.52	0.6	0.63	0.87	1.14	1.27	1.41	1.58
Signo	−	+	−	−	+	+	+	−	−	+	−	−	−	−	−
Rango	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

$$T^+_{exp} = 2 + 5 + 6 + 7 + 10 = 30.$$

Se rechaza H_0 al nivel de significación 0.1. El cambio en el suministro sí afecta a la cantidad de sodio en sangre.

El rechazo de H_0 es debido a un valor pequeño de T^+. Por tanto, parece lógico pensar que la mediana de la distribución ha disminuido. Nos planteamos entonces el siguiente problema:

$$H_0 : M_X = 3.24$$
$$H_1 : M_X < 3.24$$

El test para resolver este problema, al nivel de significación 0.1, es

$$\varphi(x_1, \ldots, x_{15}) = \begin{cases} 1, & T^+_{exp} \leq k \\ 0, & T^+_{exp} > k \end{cases} \rightarrow P_{H_0}(T^+ \leq k) \leq 0.1 \longrightarrow \varphi = \begin{cases} 1, & T^+_{exp} \leq 36 \\ 0, & T^+_{exp} > 36. \end{cases}$$

A partir de los datos se concluye que, al nivel de significación 0.1, *se rechaza la hipótesis nula a favor de que el cambio en el suministro ha disminuido la cantidad de sodio en sangre.*