

## Tema 2. Introducción al Jackknife y a los tests de permutaciones

basado en

B. Efron, R. Tibshirani (1993). An Introduction to the bootstrap.

O. Kirchkamp (2019). Resampling methods.

Curso 2022/2023

# Introducción al Jackknife

- ▶ El método de **jackknife** fue propuesto por Quenouille alrededor de 1950 en principio para estimar errores estándar y sesgos.
- ▶ En esencia, dado un conjunto de datos  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  la muestra  $i$ -ésima  $\mathbf{x}_{(i)}$  se define como el conjunto de datos original **menos** la observación  $i$ -ésima

$$\mathbf{x}_{(i)} = (x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$$

para  $i = 1, 2, \dots, n$ .

- ▶ La réplica  $i$ -ésima  $\hat{\theta}_{(i)}$  del estadístico  $\theta = s(\mathbf{x})$  es el valor de  $s(\cdot)$  evaluado en  $\mathbf{x}_{(i)}$

$$\hat{\theta}_{(i)} = s(\mathbf{x}_{(i)})$$

para  $i = 1, 2, \dots, n$ .

# Jackknife

- ▶ El estimador jackknife del sesgo se define como

$$\widehat{\text{Sesgo}}_{jack} = (n - 1) \left( \hat{\theta}_{(\cdot)} - \hat{\theta} \right)$$

donde

$$\hat{\theta}_{(\cdot)} = \frac{\sum_{i=1}^n \hat{\theta}_{(i)}}{n}$$

- ▶ Se puede definir también el estimador jackknife del error estándar

$$\hat{s}e_{jack} = \sqrt{\frac{n-1}{n} \sum_{i=1}^n \left( \hat{\theta}_{(i)} - \frac{1}{n} \sum_{i=1}^n \hat{\theta}_{(i)} \right)^2}$$

- ▶ El jackknife funciona bien si el estadístico es *suave* (smooth). Por ejemplo, la mediana **no** lo es...

# Jackknife

- ▶ **Justificación de la fórmula:**

En el jackknife se consideran  $n$  muestras fijas  $\mathbf{x}_{(1)}, \dots, \mathbf{x}_{(n)}$  que se obtienen eliminando una observación en cada caso.

- ▶ Se observa un *factor de inflación*  $(n - 1)$  que multiplica a los habituales términos  $1/(n - 1)$  o  $1/n$ .

- ▶ Se necesita poner ya que las desviaciones jackknife  $\left(\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)}\right)^2$  son bastante pequeñas al ser más similares a los datos originales  $\mathbf{x}$ .

- ▶ La expresión exacta del factor  $(n - 1)/n$  se obtiene considerando el caso especial de  $\hat{\theta} = \bar{x}$ .

# Jackknife

- ▶ Intuitivamente, la idea de suavidad (*smoothness*) implica que pequeños cambios en los datos originan solo pequeños cambios en el estadístico.
- ▶ Esto NO se produce, por ejemplo, en el caso de la mediana.
- ▶ Por ejemplo consideramos los datos del grupo control de ratones

10   27   31   40   46   50   52   104   146

- ▶ La mediana es 46.

# Jackknife

- ▶ Supongamos que se incrementa el valor del cuarto elemento mayor  $x = 40$ .
- ▶ La mediana no cambia hasta que  $x$  sea mayor que 46 y luego permanece igual hasta que  $x$  excede 50.
- ▶ Esto implica que la mediana no es una función suave.
- ▶ Por esto el estimador jackknife del error estándar es inconsistente para la mediana, es decir, cuando el tamaño muestral  $n \rightarrow \infty$  no converge al verdadero error estándar.

# Ejemplo sobre farmacología

- ▶ Se tiene una serie de datos sobre el efecto de unos parches hormonales sobre 8 personas. Dichos parches difunden un medicamento en la sangre.
- ▶ Se mide el nivel de la hormona que aparece después de usar tres parches diferentes: un parche *placebo* (sin hormona), un parche *viejo* y uno *nuevo*.
- ▶ Se trata de estudiar su *bioequivalencia*.
- ▶ El criterio que se utiliza en la agencia estatal norteamericana de medicamentos (*FDA*) es que

$$\frac{|E(\text{nuevo}) - E(\text{viejo})|}{E(\text{viejo}) - E(\text{placebo})} \leq 0,20$$

# Ejemplo sobre farmacología

- ▶ Es decir, la *FDA* exige que el nuevo tipo de parche se ajuste a la cantidad de hormona que liberaba el antiguo (respecto al placebo) en no más del 20 %.
- ▶ Se denomina como
$$\mathbf{z} \equiv (\text{medida parche viejo} - \text{medida placebo})$$
$$\mathbf{y} \equiv (\text{medida parche nuevo} - \text{medida parche viejo})$$
- ▶ Se asume que los valores  $\mathbf{x}_i = (z_i, y_i)$  se obtienen a partir de una distribución bivalente  $F \rightarrow \mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_8)$ .



## Ejemplo sobre farmacología

- ▶ El parámetro en este caso es entonces

$$\theta = t(F) = \frac{E_F(\mathbf{y})}{E_F(\mathbf{z})} = \frac{E(\text{nuevo}) - E(\text{viejo})}{E(\text{viejo}) - E(\text{placebo})}$$

- ▶ En este caso  $t(\cdot)$  es una función que tiene como objeto las parejas de  $\mathbf{X}$  y da como resultado el ratio de las esperanzas.
- ▶ El estimador *plug-in* de  $\theta$  es

$$\hat{\theta} = t(\hat{F}) = \frac{\bar{y}}{\bar{z}} = \frac{\sum_i y_i / 8}{\sum_i z_i / 8}$$

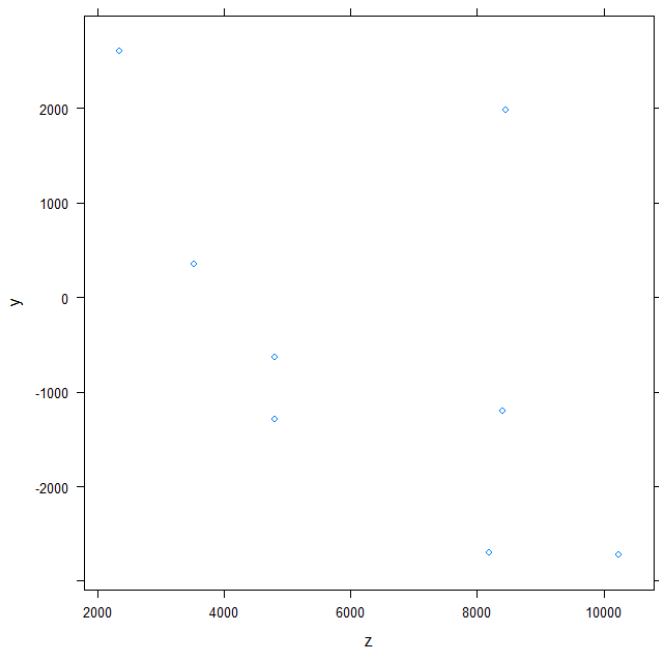
# Ejemplo sobre farmacología

- ▶ En el ejemplo,

$$\frac{E(nuevo) - E(viejo)}{E(viejo) - E(placebo)} = \frac{E(\mathbf{y})}{E(\mathbf{z})} \approx \frac{\bar{y}}{\bar{z}} = \frac{-452,25}{6342,375} \approx -0,0713$$

```
data(patch, package="bootstrap")  
lattice::xyplot(y ~ z, data=patch)
```

- ▶ Se puede aplicar el método de jackknife en este estadístico.



# Introducción a los contrastes de permutaciones

- ▶ Supongamos que se diseña un nuevo medicamento que supuestamente inhibiría la capacidad de un ratón para atravesar un laberinto.
- ▶ Los científicos diseñan un experimento en el que tres ratones son elegidos al azar para recibir el medicamento y otros tres ratones sirven como controles al ingerir un placebo.
- ▶ El tiempo que tarda cada ratón en atravesar un laberinto se mide en segundos.
- ▶ Supongamos que se obtienen los siguientes resultados:

**Con medicamento:** 30, 25, 20

**Control:** 18, 21, 22

# Introducción a los contrastes de permutaciones

- ▶ El tiempo medio para el grupo con medicación es de 25s, y el del grupo de control es 20.33s. Así, la diferencia media en tiempos es  $25 - 20,33 = 4,67s$ .
- ▶ El tiempo medio para los ratones que reciben el medicamento es mayor que el tiempo promedio para el grupo de control, pero esto podría deberse a una variabilidad aleatoria en lugar de un verdadero efecto del medicamento.  
No podemos decir **con certeza** si hay un efecto real.
- ▶ Lo que se puede hacer es estimar si es posible que el azar puro produzca una diferencia tan grande.
- ▶ Si esa probabilidad es pequeña, entonces concluimos que hay algo más que pura casualidad, y se debería concluir que hay un efecto real del medicamento.

# Introducción a los contrastes de permutaciones

- ▶ Si el medicamento realmente no influye en los tiempos, la división de las seis observaciones en dos grupos fue esencialmente aleatoria.
- ▶ Los resultados podrían distribuirse también como

**Con medicamento:** 30, 25, 18

**Control:** 20, 21, 22

- ▶ En este caso la diferencia media es
$$((30 + 25 + 18) / 3) - ((20 + 21 + 22) / 3) = 3,33.$$
- ▶ Hay  $\binom{6}{3} = 20$  formas de distribuir 6 números en dos conjuntos de tamaño 3, sin tener en cuenta el orden dentro de cada conjunto.

# Introducción a los contrastes de permutaciones

- ▶ La idea central de la significación estadística o la prueba de hipótesis clásica es calcular con qué frecuencia el azar puro daría un efecto tan grande como el observado en los datos, en ausencia de cualquier efecto real.
- ▶ Si esa probabilidad es lo suficientemente pequeña, concluimos que los datos proporcionan la evidencia de un efecto real.
- ▶ Si la probabilidad no es pequeña, no llegamos a esa conclusión.

# Introducción a los contrastes de permutaciones

- ▶ Esto no es lo mismo que concluir que no hay ningún efecto; es solo que los datos disponibles no proporcionan evidencia convincente de que haya un efecto.
- ▶ En la práctica, puede haber muy pocos datos para proporcionar evidencia convincente.
- ▶ Si el efecto del fármaco es pequeño, es posible distinguir el efecto del ruido aleatorio con 60 ratones, pero no con 6.



# Contrastes de hipótesis

- ▶ Se puede considerar el lenguaje de los tests de hipótesis o contrastes de significación.
- ▶ **Definición:** La hipótesis nula denotada por  $H_0$  es la afirmación que corresponde al hecho de que no haya efecto real. Este es el *status quo* que se mantiene excepto que los datos sugieran evidencia de lo contrario.
- ▶ La hipótesis alternativa que se denota como  $H_A$  indica que existe un efecto real.
- ▶ Los datos obtenidos sugieren o no la evidencia de que esta hipótesis es cierta.

# Contrastes de hipótesis

- ▶ Una hipótesis involucra una afirmación acerca de un parámetro (o parámetros) de una población que se puede denotar como  $\theta$ .
- ▶ La hipótesis nula es  
 $H_0 : \theta = \theta_0$  para algún valor real  $\theta_0$ .
- ▶ Una hipótesis alternativa de una cola es de la forma  
 $H_A : \theta > \theta_0$  ó  $H_A : \theta < \theta_0$
- ▶ Una hipótesis alternativa de dos colas es de la forma  
 $H_A : \theta \neq \theta_0$

# Contrastes de hipótesis

- ▶ Se puede considerar el ejemplo anterior de los ratones.
- ▶ Sea  $\mu_d$  el tiempo medio real que tarda un ratón seleccionado al azar que recibió el medicamento en atravesar el laberinto;
- ▶ Sea  $\mu_c$  el tiempo medio para un ratón de control.
- ▶ Entonces  $H_0 : \mu_d = \mu_c$
- ▶ Es decir, en promedio, no hay diferencia en los tiempos medios entre los ratones que reciben el medicamento y los ratones en el grupo de control.

# Contrastes de hipótesis

- ▶ La hipótesis alternativa es  $H_A : \mu_d > \mu_c$
- ▶ Es decir, en promedio, los ratones que reciben el medicamento tienen tiempos más lentos (valores más grandes) que los ratones en el grupo de control.
- ▶ Las hipótesis pueden ser re-escritas como
$$H_0 : \mu_d - \mu_c \leq 0$$
$$H_A : \mu_d - \mu_c > 0$$
- ▶ Así  $\theta = \mu_d - \mu_c$   
(cualquier función de los parámetros es en sí misma un parámetro).

# Contrastes de hipótesis

- ▶ Un estadístico de contraste es una función numérica de los datos cuyo valor determina el resultado del contraste.
- ▶ La función en sí misma generalmente se denota  $T = T(\mathbf{X})$  donde  $\mathbf{X}$  representa los datos.
- ▶ Por ejemplo,  $T = T(X_1, X_2, \dots, X_n)$  en un problema de una muestra, y
- ▶  $T = T(X_1, X_2, \dots, X_m, Y_1, \dots, Y_n)$  en un problema de dos muestras.

# Contrastes de hipótesis

- ▶ Después, cuando se evalúa para los datos concretos de la muestra, el resultado es un estadístico del test *observado* (un número) y se escribe en minúscula  $t = T(x)$ .
- ▶ **Definición.** El  $p$ -valor es la probabilidad de que por azar se obtenga un estadístico de contraste tan extremo como el observado asumiendo que la hipótesis nula era cierta.
- ▶ En el ejemplo de ratones, el estadístico de contraste es la diferencia en medias,  $T = T(X_1, X_2, X_3, Y_1, Y_2, Y_3) = \bar{X} - \bar{Y}$  con el valor observado  $t = \bar{x} - \bar{y} = 4,67$ .

# Contrastes de hipótesis

- ▶ Valores grandes del estadístico de contraste apoyan la hipótesis alternativa, entonces el p-valor es  $P(T \geq 4,67) = 3/20$ .
- ▶ **Definición.** Un resultado es *estadísticamente significativo* si ocurre rara vez por azar.
- ▶ **¿Cuánto?** Depende de un contexto, pero, por ejemplo, un p-valor de 0.0002 indicaría que si se asume como cierta la hipótesis nula, el resultado observado ocurriría solo 2 de cada 10000 veces por casualidad, lo que en la mayoría de las circunstancias resulta bastante extraño.
- ▶ Esto llevaría a concluir que la evidencia apoya la hipótesis alternativa.

# Contrastes de hipótesis

- ▶ Cuanto más pequeño se fije el p-valor para afirmar que el resultado es estadísticamente significativo, más conservador será el contraste, ya que se exige una evidencia más sólida para rechazar el *status quo* (la hipótesis nula).
- ▶ En lugar de simplemente calcular la probabilidad, a menudo comenzamos respondiendo una pregunta más amplia:  
¿Cuál es la distribución del estadístico del test cuando no hay un efecto real?



# Contrastes de hipótesis

- ▶ La **distribución nula** es la distribución del estadístico de test si la hipótesis nula es **verdadera**.
- ▶ Se puede pensar en la distribución nula como una distribución de referencia: comparamos el estadístico de test observado con esta referencia para determinar cómo es de extraño el estadístico de test observado.
- ▶ Hay diferentes formas de calcular distribuciones nulas exactas o aproximadas y p-valores. Una de ellas es mediante los **tests de permutaciones**.

# Tests de Permutaciones

- ▶ Se trata de un test de significación estadística para contrastar la diferencia entre grupos. Fue desarrollado por Fisher y Pitman en 1930.
- ▶ La distribución del estadístico estudiado (media, mediana...) se obtiene calculando el valor de dicho estadístico para todas las posibles reorganizaciones de las observaciones en los distintos grupos.
- ▶ Dado que implica calcular todas las posibles situaciones se trata en principio de un test exacto.

# Tests de Permutaciones

- ▶ Por ejemplo, supongamos que un conjunto de sujetos se distribuye en dos grupos,  $A$  y  $B$ , de tamaños  $n_A$  y  $n_B$ , cuyas medias muestrales tras el experimento resultan ser  $\bar{x}_A$  y  $\bar{x}_B$ .
- ▶ Se desea determinar si existe una diferencia significativa entre las medias de los dos grupos.
- ▶ O sea comprobar si hay evidencias en contra de la hipótesis nula de que la diferencia observada es debida únicamente a la asignación al azar de los sujetos a los dos grupos y que ambas muestras proceden de la misma población.
- ▶ En primer lugar se calcula la diferencia entre las medias de los dos grupos, a lo que se conoce como diferencia observada.

# Tests de Permutaciones

- ▶ Todas las observaciones se combinan juntas sin tener en cuenta el grupo al que pertenecían.
- ▶ Se calculan todas las posibles permutaciones en las que las observaciones pueden ser agrupadas en dos grupos de tamaño  $n_A$  y  $n_B$ .
- ▶ Para cada permutación se calcula la diferencia entre medias. El conjunto de valores calculados forman la distribución exacta de las posibles diferencias siendo cierta la hipótesis nula.
- ▶ El p-valor de dos colas se calcula como la proporción de permutaciones muestrales en las que el valor absoluto de la diferencia calculada es mayor o igual al valor absoluto de la diferencia observada.

# Tests de Permutaciones

- ▶ La condición necesaria para un test de permutaciones se conoce como **intercambiabilidad**, según la cual todas las posibles permutaciones tienen la misma probabilidad de ocurrir siendo cierta la hipótesis nula.
- ▶ Las conclusiones de un test de permutaciones solo son aplicables a diseños de tipo experimental, es decir, en los que tras haber elegido los sujetos del estudio, se realiza una asignación aleatoria de los sujetos a los diferentes grupos.
- ▶ Los test de permutaciones son test de significación y por lo tanto se emplean para calcular p-valores, no para intervalos de confianza.

# Tests de Permutaciones

- ▶ Se observan dos muestras aleatorias independientes  $\mathbf{x}$  e  $\mathbf{y}$  extraídas de posiblemente dos distribuciones de probabilidad diferentes:

$$F \rightarrow \mathbf{x} = (x_1, x_2, \dots, x_n)$$

$$G \rightarrow \mathbf{y} = (y_1, y_2, \dots, y_m)$$

- ▶ Se considera la hipótesis nula

$$H_0 : F = G$$

- ▶ La igualdad  $F = G$  significa que  $F$  y  $G$  asignan igual probabilidad a todos los conjuntos

$$P_F \{A\} = P_G \{A\},$$

donde  $A$  es un subconjunto del espacio muestral conjunto de las v.a.  $x$  e  $y$ .

# Tests de Permutaciones

- ▶ Si  $H_0$  es cierta, entonces no hay diferencias entre el comportamiento probabilístico de las variables  $\mathbf{x}$  e  $\mathbf{y}$ .
- ▶ **Ejemplo:** En el caso de los datos de ratones del tema 1 (bajo tratamiento para prolongar su supervivencia después de cirugía invasiva), hay pocas observaciones, de modo que el tamaño muestral del caso tratamiento es  $n = 7$  y el del caso del control es  $m = 9$ .
- ▶ La diferencia entre las medias es

$$\hat{\theta} = \bar{x} - \bar{y} = 30,63$$

- ▶ Parece deducirse que la distribución del tratamiento ( $F$ ) provoca tiempos de supervivencia mayores que la distribución del control ( $G$ ).

# Tests de Permutaciones

- ▶ Pero, si no se puede rechazar categóricamente la posibilidad de que  $H_0$  sea cierta, entonces eso **NO** significa que se haya **demostrado** la hipótesis alternativa.
- ▶ En realidad, un test de hipótesis es un método formal que se plantea para decidir si los datos rechazan *decisivamente* la hipótesis nula  $H_0$ .
- ▶ En el caso de los ratones se calcula el estadístico  $\hat{\theta}$  de modo que, intuitivamente se esperan valores altos si  $H_0$  es falsa.
- ▶ Cuanto mayor sea el valor observado de  $\hat{\theta}$  mayor evidencia se tendrá en contra de  $H_0$ .



# Tests de Permutaciones

- ▶ Se define el **nivel de significación alcanzado** (*ASL*) como la probabilidad de observar al menos un valor tan grande del estadístico cuando la hipótesis nula es cierta

$$ASL = P_{H_0} \left\{ \hat{\theta}^* \geq \hat{\theta} \right\}$$

- ▶ Cuanto menor sea el valor de *ASL*, mayor es la evidencia en contra de  $H_0$ .
- ▶ El valor  $\hat{\theta}$  es un valor fijo y observado. La variable  $\hat{\theta}^*$  tiene la distribución que se asume bajo  $H_0$ .
- ▶ La notación *estrella* diferencia entre la observación real del estadístico  $\hat{\theta}$  y la distribución de  $\hat{\theta}^*$  generada de acuerdo con la hipótesis nula  $H_0$ .

# Tests de Permutaciones

- ▶ La validación de  $H_0$  se hace calculando el valor de  $ASL$  para ver si es pequeño en relación con ciertos límites.
- ▶ Formalmente si se toma una significación de  $\alpha$ , por ejemplo igual a 0,05, entonces se rechaza  $H_0$  si  $ASL$  es menor que  $\alpha$ .
- ▶ En el caso de los ratones, un **test tradicional** asumiría que  $F$  y  $G$  se distribuyen como una normal con medias posiblemente diferentes:

$$F = N(\mu_x, \sigma^2)$$

$$G = N(\mu_y, \sigma^2)$$

- ▶ De este modo, la hipótesis nula  $H_0$  es equivalente a decir que  $\mu_x = \mu_y$ .

# Tests tradicionales

- ▶ Si se cumple  $H_0$  entonces

$$\hat{\theta} \sim N\left(0, \sigma^2 \left(\frac{1}{n} + \frac{1}{m}\right)\right).$$

- ▶ Entonces

$$\begin{aligned} \text{p-valor} = P_{H_0} \left\{ \hat{\theta} \geq 0 \right\} &= P \left\{ Z \geq \frac{\hat{\theta}}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} \right\} \\ &= 1 - \Phi \left( \frac{\hat{\theta}}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} \right) \end{aligned}$$

- ▶ De modo que bajo la suposición de normalidad,  $\hat{\theta}$  sigue así una distribución *conocida*.

# Tests tradicionales

- ▶ En general, en muestras pequeñas, se aplica el test de la  $t$  de Student de modo que, en el ejemplo,

$$p - \text{valor} = P \left\{ t_{14} \geq \frac{30,63}{54,21 \sqrt{\frac{1}{9} + \frac{1}{7}}} \right\} = 0,141$$

- ▶ Con lo que se concluye que este valor no permite rechazar  $H_0$  ni siquiera con una significación  $\alpha = 0,10$ .
- ▶ Pero esta solución solo es válida si se asume normalidad de los datos.

# Procedimiento de remuestreo

- ▶ Se juntan las dos muestras en una sola  $(\mathbf{x}, \mathbf{y})$  y de esta nueva muestra se toman aleatoriamente dos nuevas submuestras  $\mathbf{x}'$  e  $\mathbf{y}'$  de tamaño  $n$  y  $m$  respectivamente.
- ▶ Así se elimina la asociación previa de los valores con las distribuciones originales  $F$  y  $G$ .
- ▶ Se generan  $B$  submuestras independientes  $\mathbf{x}'$  e  $\mathbf{y}'$   
**SIN reemplazamiento** de la muestra conjunta  $(\mathbf{x}, \mathbf{y})$ .

# Procedimiento de remuestreo

- ▶ Se calcula  $\hat{\theta}^*(b)$  para cada muestra.
- ▶ Después se calcula

$$\widehat{ASL}_{perm} = \frac{\# \left\{ \hat{\theta}^*(b) \geq \hat{\theta} \right\}}{B}$$

- ▶ De manera ideal se deberían tomar

$$B = \binom{N}{n}$$

diferentes submuestras para calcular el valor *exacto* de *ASL*.

# Ejemplo

Del libro *Comparing Groups Randomization and Bootstrap Methods Using R*

- ▶ Se trata de determinar si la participación en actividades extraescolares aumenta la capacidad empática de los estudiantes.
- ▶ Así el instituto ofrece un programa en el que cada participante se asigna de forma aleatoria a un grupo *control*, que no recibe clases extraescolares, o a un grupo *tratamiento* que sí las recibe.
- ▶ Al final del año todas las personas del estudio realizan un examen que determina su capacidad empática.
- ▶ A la vista de los resultados ¿Se puede considerar que las clases extraescolares tienen algún impacto en comportamiento social de los estudiantes?

## Ejemplo tablas de contingencia

Creencias Religiosas				
Educación	Fundamentalista	Moderada	Liberal	Total
< Secundaria	178	138	108	424
Secundaria	570	648	442	1660
Graduado	138	252	252	642
Total	886	1038	802	2726

```
tabla = as.table(rbind(c(178, 138, 108), c(570, 648, 442),  
                      c(138,252, 252)))  
  
(res = chisq.test(tabla))
```

Pearson's Chi-squared test

```
data:  tabla.array  
X-squared = 69.1568, df = 4, p-value = 3.42e-14
```



## Ejemplo tablas de contingencia

- ▶ Con la función `chisq.test` también se pueden hacer contrastes por simulación Montecarlo, es decir, calculando el estadístico de la chi cuadrado para todas las posibles tablas con las mismas sumas marginales por filas y columnas de la tabla original.

```
chisq.test(tabla, sim=T, B=2000)
```

```
Pearson's Chi-squared test with simulated p-value  
(based on 2000 replicates)
```

```
data:  tabla.array  
X-squared = 69.1568, df = NA, p-value = 0.0004998
```

# Ejemplo tablas de contingencia usando la librería coin

- Se consideran los datos de relación entre enfermedades degenerativas cerebrales y el consumo de tabaco

```
library(coin)  
print(alzheimer)
```

	smoking	disease	gender
1	None	Alzheimer	Female
2	None	Alzheimer	Female
...	.....	.....	.....
537	>20	Other diagnoses	Male
538	>20	Other diagnoses	Male

```
mosaicplot(~disease + smoking + gender,  
data=alzheimer, main="Enfermedad",  
col=c("pink","blue"), off=c(5,5,5,5))
```

# Enfermedad

