ANOVA y test chi-cuadrado

Sesión 4. Curso: "Análisis estadístico aplicado con R Commander"

Carlos Vergara Hernández
vergara_car@gva.es

Una breve introducción

Estadística paramétrica vs. no paramétrica

En la clase previa se abordaron algunas pruebas básicas de contraste de hipótesis bivariado.

Estadística paramétrica vs. no paramétrica

En la clase previa se abordaron algunas pruebas básicas de contraste de hipótesis bivariado.

- Algunas presentan requisitos que los datos deben cumplir.
- Principalmente, los requisitos que nos encontramos con más frecuencia son:
  - normalidad (test de Shapiro-Wilk),
  - homogeneidad de varianzas (test Levene),
  - independencia en las observaciones (test de autocorrelación –p.ej.,Durbin-Watson–).

Estadística paramétrica vs. no paramétrica

En la clase previa se abordaron algunas pruebas básicas de contraste de hipótesis bivariado.

- Algunas presentan requisitos que los datos deben cumplir.
- Principalmente, los requisitos que nos encontramos con más frecuencia son:
    - normalidad (test de Shapiro-Wilk),
    - homogeneidad de varianzas (test Levene),
    - independencia en las observaciones (test de autocorrelación –p.ej.,Durbin-Watson–).

¿Qué pasa si se incumplen los requisitos?

El resultado será la obtención de estimaciones sesgadas.

Estadística paramétrica vs. no paramétrica

En la clase previa se abordaron algunas pruebas básicas de contraste de hipótesis bivariado.

- Algunas presentan requisitos que los datos deben cumplir.
- Principalmente, los requisitos que nos encontramos con más frecuencia son:
  - normalidad (test de Shapiro-Wilk),
  - homogeneidad de varianzas (test Levene),
  - independencia en las observaciones (test de autocorrelación –p.ej.,Durbin-Watson–).

¿Qué pasa si se incumplen los requisitos?

El resultado será la obtención de estimaciones sesgadas.

Solución

Pruebas no paramétricas (sin –apenas– requisitos), aunque a costa de una menor potencia estadística.

¿Qué prueba usar?

Prueba	Tipo respuesta	Variable independiente	Tipo de test
Test t	Númerica independiente	2 categorías	Paramétrico
Test t pareado	Númerica pareada	2 categorías	Paramétrico
Mann-Whitney	Ordinal independiente	2 categorías	No paramétrico
Mann-Whitney pareado	Ordinal pareada	2 categorías	No paramétrico
ANOVA una vía	Númerica independiente	≥ 2 categorías	Paramétrico
ANOVA medidas repetidas	Númerica pareada	≥ 2 categorías	Paramétrico
Kruskal-Wallis	Ordinal independiente	≥ 2 categorías	No paramétrico
Friedman	Ordinal pareada	≥ 2 categorías	No paramétrico
Chi-cuadrado	Categórica independiente	≥ 2 categorías	Nominal
Test de Fisher	Categórica independiente	≥ 2 categorías	Nominal

Organización de la sesión

En esta sesión se presentarán 6 pruebas diferentes:

- Pruebas paramétricas:
  - ANOVA,

Organización de la sesión

En esta sesión se presentarán 6 pruebas diferentes:

- Pruebas paramétricas:
  - ANOVA,
- Pruebas no paramétricas:
  - test de Mann-Whitney,
  - test de Mann-Whitney pareado,
  - test de Kruskal-Wallis,

Organización de la sesión

En esta sesión se presentarán 6 pruebas diferentes:

- Pruebas paramétricas:
  - ANOVA,
- Pruebas no paramétricas:
  - test de Mann-Whitney,
  - test de Mann-Whitney pareado,
  - test de Kruskal-Wallis,
- Pruebas nominales:
  - test χ^2.
  - test de Fisher.

Variable cuantitativa y categórica con 2 categorías

Test de Mann-Whitney (1/3)

Recuerdo

En la clase previa se vio el contraste de dos medias cuando se cumplen los requisitos de normalidad y homogeneidad de varianzas y ahora, empezaremos viendo qué opciones hay cuando alguno se incumple.

Test de Mann-Whitney (1/3)

Recuerdo

En la clase previa se vio el contraste de dos medias cuando se cumplen los requisitos de normalidad y homogeneidad de varianzas y ahora, empezaremos viendo qué opciones hay cuando alguno se incumple.

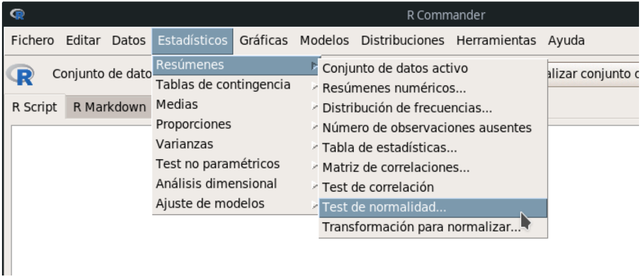
El contraste de hipóstesis se plantea sobre las medianas:

$$H_0 : Me_1 = Me_2$$

$$H_1 : Me_1 \neq Me_2$$

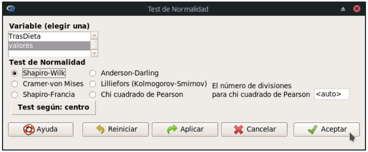
Test de Mann-Whitney (2/3)

1. Valorar los requisitos del test t



Test de Mann-Whitney (2/3)

1. Valorar los requisitos del test t



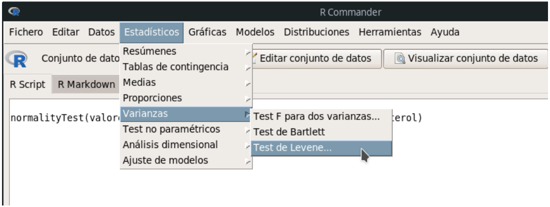
Test de Mann-Whitney (2/3)

1. Valorar los requisitos del test t

```
        Shapiro-Wilk normality test

data:  valores
W = 0.9448, p-value = 0.2949

 --------

 p-values adjusted by the Holm method:
  unadjusted adjusted
1 0.41568    0.58984
2 0.29492    0.58984
```

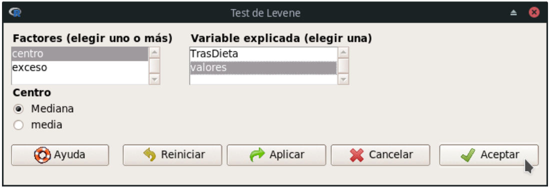
Test de Mann-Whitney (2/3)

1. Valorar los requisitos del test t



Test de Mann-Whitney (2/3)

1. Valorar los requisitos del test t



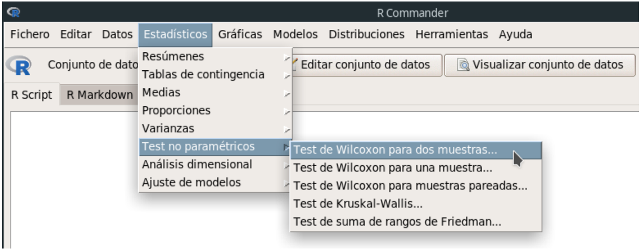
Test de Mann-Whitney (2/3)

1. Valorar los requisitos del test t

```
> leveneTest(valores ~ centro, data = colesterol, center = "median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  1  0.0581 0.8108
      38
```

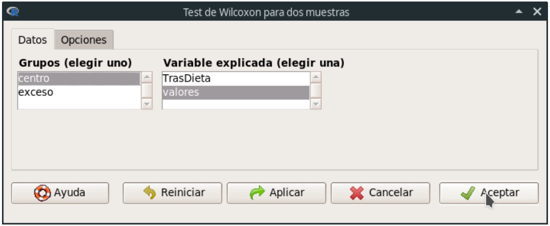
Test de Mann-Whitney (3/3)

2. Plantear su alternativa no paramétrica: U de Mann-Whitney



Test de Mann-Whitney (3/3)

2. Plantear su alternativa no paramétrica: U de Mann-Whitney



Test de Mann-Whitney (3/3)

2. Plantear su alternativa no paramétrica: U de Mann-Whitney

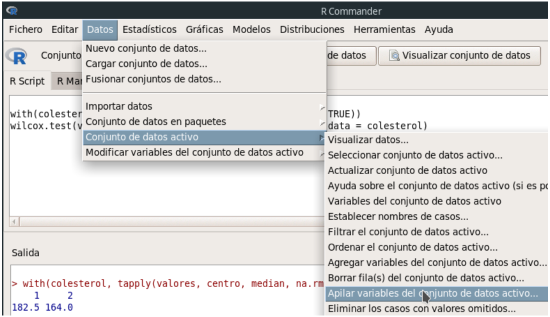
```
> with(colesterol, tapply(valores, centro, median, na.rm = TRUE))
    1     2
182.5 164.0

> wilcox.test(valores ~ centro, alternative = "two.sided", data = colesterol)

        Wilcoxon rank sum test with continuity correction

data:  valores by centro
W = 289.5, p-value = 0.016
alternative hypothesis: true location shift is not equal to 0
```

Test de Mann-Whitney pareado (1/2)

Manejo de datos

Para poder trabajar con esta prueba se requiere que los datos estén apilados, lo que implica un pre-procesado de los mismos.

1. Introducir el formato correcto de los datos

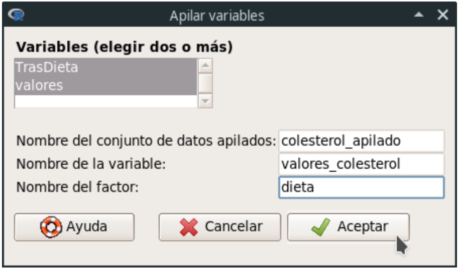


Test de Mann-Whitney pareado (1/2)

Manejo de datos

Para poder trabajar con esta prueba se requiere que los datos estén apilados, lo que implica un pre-procesado de los mismos.

1. Introducir el formato correcto de los datos



Test de Mann-Whitney pareado (1/2)

Manejo de datos

Para poder trabajar con esta prueba se requiere que los datos estén apilados, lo que implica un pre-procesado de los mismos.

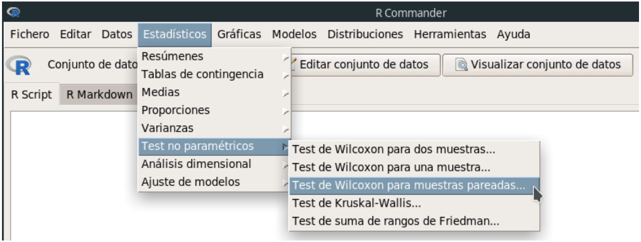
1. Introducir el formato correcto de los datos

colesterol_apilad

	valores_colesterol	dieta
1	157	TrasDieta
2	172	TrasDieta
3	157	TrasDieta
4	205	TrasDieta
5	173	TrasDieta
6	150	TrasDieta
7	182	TrasDieta
8	189	TrasDieta
9	181	TrasDieta
10	168	TrasDieta
11	202	TrasDieta
12	175	TrasDieta
13	159	TrasDieta
14	120	TrasDieta
15	200	TrasDieta
16	179	TrasDieta
17	168	TrasDieta
18	184	TrasDieta
19	189	TrasDieta
20	181	TrasDieta
21	190	TrasDieta

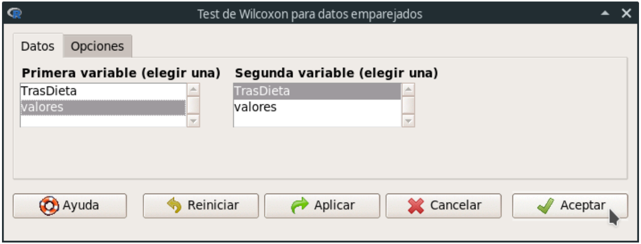
Test de Mann-Whitney pareado (2/2)

2. Plantear su alternativa no paramétrica: U de Mann-Whitney con datos pareados



Test de Mann-Whitney pareado (2/2)

2. Plantear su alternativa no paramétrica: U de Mann-Whitney con datos pareados



Test de Mann-Whitney pareado (2/2)

2. Plantear su alternativa no paramétrica: U de Mann-Whitney con datos pareados

```
> with(colesterol, median(valores - TrasDieta, na.rm = TRUE))  # median difference
[1] 5

> with(colesterol, wilcox.test(valores, TrasDieta, alternative = "two.sided", paired = TRUE))

        Wilcoxon signed rank test with continuity correction

data:  valores and TrasDieta
V = 732, p-value = 0.00001519
alternative hypothesis: true location shift is not equal to 0
```

Variable cuantitativa y categórica con > 2 categorías

ANOVA de un factor (1/5)

Más de dos grupos

Hasta este momento se ha trabajado con variables independientes con únicamente dos categorías (p.ej., 'sí' vs 'no'), pero ahora se introduce una extensión de esos contrastes para variables con dos o más categorías (p.ej., 'No fumador', 'Ex-fumador' y 'Fumador').

ANOVA de un factor (1/5)

Más de dos grupos

Hasta este momento se ha trabajado con variables independientes con únicamente dos categorías (p.ej., 'sí' vs 'no'), pero ahora se introduce una extensión de esos contrastes para variables con dos o más categorías (p.ej., 'No fumador', 'Ex-fumador' y 'Fumador').

- El ANOVA realiza una comparación entre la varianza de la variable dependiente atribuible a la variable independiente, y aquella varianza no explicada (residual), asumiendo además de la independencia en las observaciones:
  - normalidad en la variable dependiente en los distintos grupos (con una $n > 30$ esto se relaja): hay que aplicar el test de Shapiro-Wilk,
  - homogeneidad de las varianzas de la variable dependiente en los grupos: hay que aplicar el test de Levene.

ANOVA de un factor (2/5)

El contraste de hipóstesis que se plantea es:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_j$$
$$H_1 : \exists (i, j). \quad \mu_i \neq \mu_j$$

ANOVA de un factor (2/5)

El contraste de hipóstesis que se plantea es:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_j$$
$$H_1 : \exists(i,j). \quad \mu_i \neq \mu_j$$

El resultado de aplicar un ANOVA es la clásica tabla de ANOVA con la que todos debéis estar familiarizados:

	gl	**Suma de cuadrados**	**Media de cuadrados**	**Valor F**	**Pr($>$ F)**
Variable independiente	2	366	183.244	12.615	< 0.001
Residuo	6321	91816	14.526		

ANOVA de un factor: ejemplo (3/5)

El problema

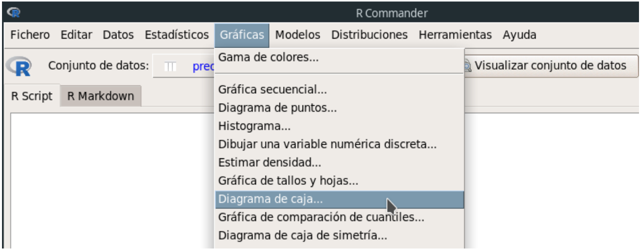
En la base de datos ‘predimed‘ se quiere estudiar la posible diferencia entre los grupos de intervención (3 categorías) respecto a su IMC.

ANOVA de un factor: ejemplo (3/5)

El problema

En la base de datos ‘predimed‘ se quiere estudiar la posible diferencia entre los grupos de intervención (3 categorías) respecto a su IMC.

1. Valorar los requisitos del ANOVA



ANOVA de un factor: ejemplo (3/5)

El problema

En la base de datos ‘predimed‘ se quiere estudiar la posible diferencia entre los grupos de intervención (3 categorías) respecto a su IMC.

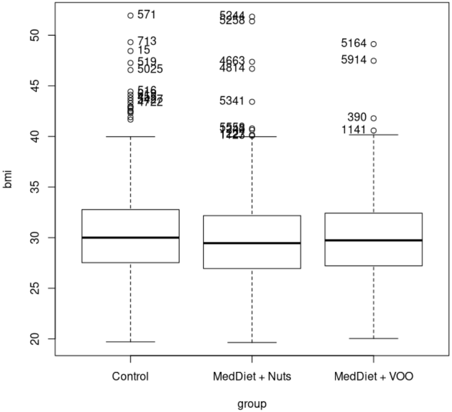
1. Valorar los requisitos del ANOVA

ANOVA de un factor: ejemplo (3/5)

El problema

En la base de datos ‘predimed‘ se quiere estudiar la posible diferencia entre los grupos de intervención (3 categorías) respecto a su IMC.

1. Valorar los requisitos del ANOVA

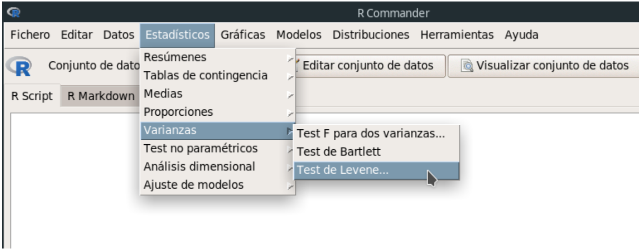


ANOVA de un factor: ejemplo (3/5)

El problema

En la base de datos ‘predimed‘ se quiere estudiar la posible diferencia entre los grupos de intervención (3 categorías) respecto a su IMC.

1. Valorar los requisitos del ANOVA

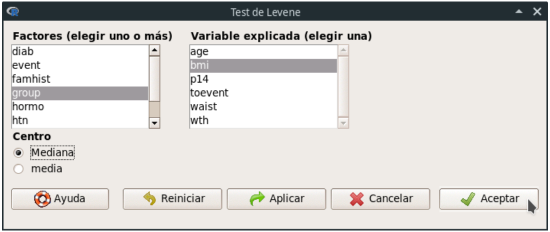


ANOVA de un factor: ejemplo (3/5)

El problema

En la base de datos ‘predimed‘ se quiere estudiar la posible diferencia entre los grupos de intervención (3 categorías) respecto a su IMC.

1. Valorar los requisitos del ANOVA



ANOVA de un factor: ejemplo (3/5)

El problema

En la base de datos ‘predimed‘ se quiere estudiar la posible diferencia entre los grupos de intervención (3 categorías) respecto a su IMC.

1. Valorar los requisitos del ANOVA

Salida

Ejecutar

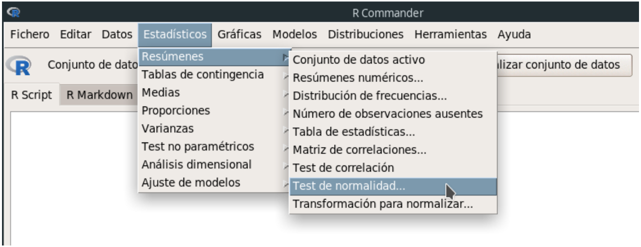
```
> with(predimed, tapply(bmi, group, var, na.rm = TRUE))
      Control MedDiet + Nuts  MedDiet + VOO
     15.71287       14.18556       13.74159

> leveneTest(bmi ~ group, data = predimed, center = "median")
Levene's Test for Homogeneity of Variance (center = "median")
        Df F value Pr(>F)
group    2  1.9506 0.1423
      6321
```

ANOVA de un factor: ejemplo (3/5)

El problema

En la base de datos ‘predimed‘ se quiere estudiar la posible diferencia entre los grupos de intervención (3 categorías) respecto a su IMC.

1. Valorar los requisitos del ANOVA

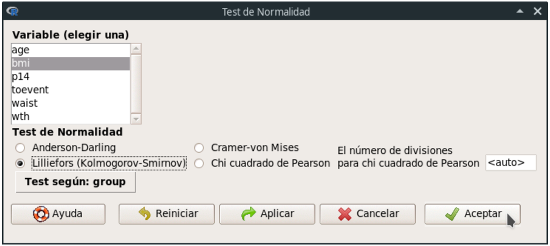


ANOVA de un factor: ejemplo (3/5)

El problema

En la base de datos ‘predimed‘ se quiere estudiar la posible diferencia entre los grupos de intervención (3 categorías) respecto a su IMC.

1. Valorar los requisitos del ANOVA



ANOVA de un factor: ejemplo (3/5)

El problema

En la base de datos ‘predimed‘ se quiere estudiar la posible diferencia entre los grupos de intervención (3 categorías) respecto a su IMC.

1. Valorar los requisitos del ANOVA

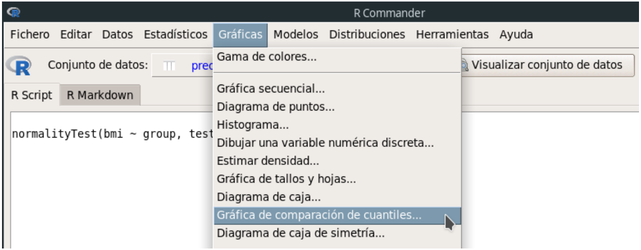
```
 --------
 group = MedDiet + VOO

        Lilliefors (Kolmogorov-Smirnov) normality test

data:  bmi
D = 0.027587, p-value = 0.0005723

 --------

 p-values adjusted by the Holm method:
              unadjusted  adjusted
Control        0.000072102 0.00021631
MedDiet + Nuts 0.00016648  0.00033296
MedDiet + VOO  0.00057233  0.00057233
```

ANOVA de un factor: ejemplo (3/5)

El problema

En la base de datos 'predimed' se quiere estudiar la posible diferencia entre los grupos de intervención (3 categorías) respecto a su IMC.

1. Valorar los requisitos del ANOVA

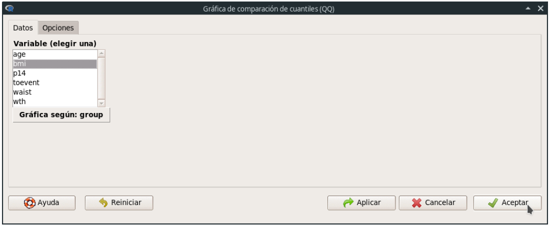


ANOVA de un factor: ejemplo (3/5)

El problema

En la base de datos ‘predimed‘ se quiere estudiar la posible diferencia entre los grupos de intervención (3 categorías) respecto a su IMC.

1. Valorar los requisitos del ANOVA

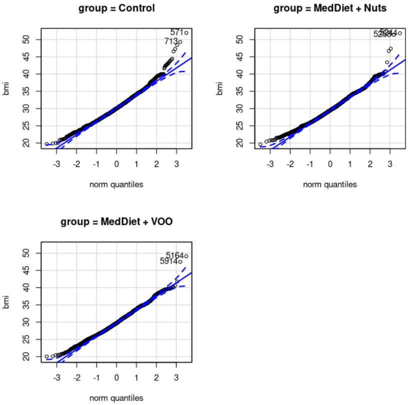


ANOVA de un factor: ejemplo (3/5)

El problema

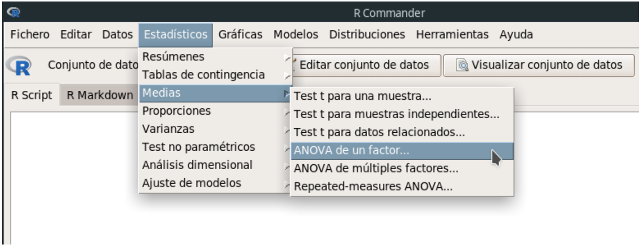
En la base de datos ‘predimed‘ se quiere estudiar la posible diferencia entre los grupos de intervención (3 categorías) respecto a su IMC.

1. Valorar los requisitos del ANOVA



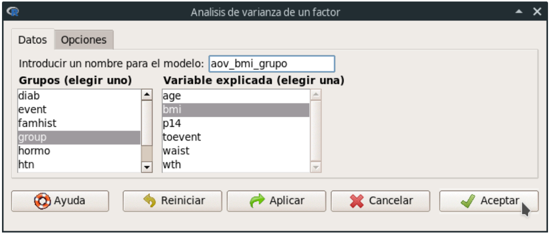
ANOVA de un factor: ejemplo (4/5)

2. Ejecución y resultados



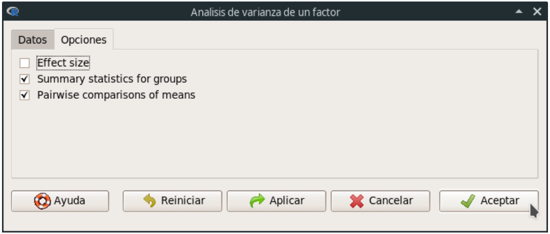
ANOVA de un factor: ejemplo (4/5)

2. Ejecución y resultados



ANOVA de un factor: ejemplo (4/5)

2. Ejecución y resultados



ANOVA de un factor: ejemplo (4/5)

2. Ejecución y resultados

```
> AnovaModel.4 <- aov(bmi ~ group, data = predimed)

> .myAnova <- summary(AnovaModel.4)

> .myAnova
              Df Sum Sq Mean Sq F value     Pr(>F)    
group          2    366  183.24   12.62 0.00000341 ***
Residuals   6321  91816   14.53                       
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ANOVA de un factor: ejemplo (4/5)

2. Ejecución y resultados

```
> # NUMERIC SUMMARY OF GROUPS

> with(predimed, numSummary(bmi, groups = group, statistics = c("mean", "sd")))
                     mean       sd data:n
Control          30.28044 3.963947   2042
MedDiet + Nuts   29.68725 3.766372   2100
MedDiet + VOO    29.94050 3.706966   2182
```

ANOVA de un factor: ejemplo (5/5)

¿Y ahora qué?

Hay diferencias significativas, vale... ¿entre qué grupos? En función de sus medias, parece que entre el grupo control y los otros dos...
¿Es posible realizar comparaciones dos a dos filtrando por la pertenencia a cada grupo (test t para grupos 1 vs. 2, grupos 2 vs. 3 y grupos 1 vs. 3)?
Las respuesta es un **NO** rotundo: hay que ajustar el error que se comete al realizar comparaciones múltiples.

ANOVA de un factor: ejemplo (5/5)

¿Y ahora qué?

Hay diferencias significativas, vale... ¿entre qué grupos? En función de sus medias, parece que entre el grupo control y los otros dos...
¿Es posible realizar comparaciones dos a dos filtrando por la pertenencia a cada grupo (test t para grupos 1 vs. 2, grupos 2 vs. 3 y grupos 1 vs. 3)?
Las respuesta es un **NO** rotundo: hay que ajustar el error que se comete al realizar comparaciones múltiples.

```
         Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts


Fit: aov(formula = bmi ~ group, data = predimed)

Linear Hypotheses:
                                Estimate Std. Error t value Pr(>|t|)
MedDiet + Nuts - Control == 0    -0.5932     0.1184  -5.008  <0.0001 ***
MedDiet + VOO - Control == 0     -0.3399     0.1173  -2.897   0.0106 *
MedDiet + VOO - MedDiet + Nuts == 0  0.2532     0.1165   2.174   0.0758 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)
```

ANOVA de un factor: ejemplo (5/5)

¿Y ahora qué?

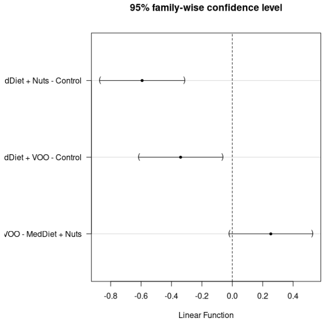
Hay diferencias significativas, vale... ¿entre qué grupos? En función de sus medias, parece que entre el grupo control y los otros dos...
¿Es posible realizar comparaciones dos a dos filtrando por la pertenencia a cada grupo (test t para grupos 1 vs. 2, grupos 2 vs. 3 y grupos 1 vs. 3)?
Las respuesta es un **NO** rotundo: hay que ajustar el error que se comete al realizar comparaciones múltiples.

```
         Simultaneous Confidence Intervals

Multiple Comparisons of Means: Tukey Contrasts


Fit: aov(formula = bmi ~ group, data = predimed)

Quantile = 2.3441
95% family-wise confidence level


Linear Hypotheses:
                                  Estimate lwr      upr
MedDiet + Nuts - Control == 0      -0.59319 -0.87085 -0.31553
MedDiet + VOO - Control == 0       -0.33994 -0.61502 -0.06487
MedDiet + VOO - MedDiet + Nuts == 0  0.25325 -0.01986  0.52635

       Control MedDiet + Nuts  MedDiet + VOO
           "b"            "a"            "a"
```

ANOVA de un factor: ejemplo (5/5)

¿Y ahora qué?

Hay diferencias significativas, vale... ¿entre qué grupos? En función de sus medias, parece que entre el grupo control y los otros dos...
¿Es posible realizar comparaciones dos a dos filtrando por la pertenencia a cada grupo (test t para grupos 1 vs. 2, grupos 2 vs. 3 y grupos 1 vs. 3)?
Las respuesta es un **NO** rotundo: hay que ajustar el error que se comete al realizar comparaciones múltiples.



Test de Kruskal-Wallis (1/3)

- Al igual que ocurría con el test t, es bastante habitual incumplir el requisito de homogeneidad de varianzas entre grupos o disponer de una n pequeña, en cuyo caso habrá que optar por la alternativa no-paramétrica: el test de Kruskal-Wallis.

Test de Kruskal-Wallis (1/3)

- Al igual que ocurría con el test t, es bastante habitual incumplir el requisito de homogeneidad de varianzas entre grupos o disponer de una n pequeña, en cuyo caso habrá que optar por la alternativa no-paramétrica: el test de Kruskal-Wallis.

El contraste de hipóstesis se plantea sobre las medianas:

$$H_0 : Me_1 = Me_2 = \cdots = Me_j$$

$$H_1 : \exists(i, j).Me_i \neq Me_j$$

Test de Kruskal-Wallis (2/3)

El problema

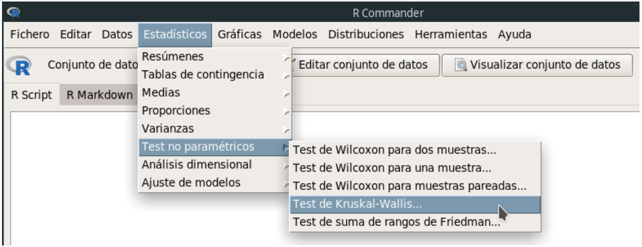
En la base de datos ‘predimed‘ se quiere estudiar la posible diferencia entre los grupos de intervención (3 categorías) respecto a su IMC.

Test de Kruskal-Wallis (2/3)

El problema

En la base de datos ‘predimed‘ se quiere estudiar la posible diferencia entre los grupos de intervención (3 categorías) respecto a su IMC.

1. Plantear alternativa no paramétrica: Kruskal-Wallis

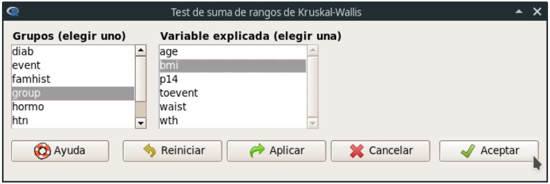


Test de Kruskal-Wallis (2/3)

El problema

En la base de datos ‘predimed‘ se quiere estudiar la posible diferencia entre los grupos de intervención (3 categorías) respecto a su IMC.

1. Plantear alternativa no paramétrica: Kruskal-Wallis



Test de Kruskal-Wallis (2/3)

El problema

En la base de datos ‘predimed‘ se quiere estudiar la posible diferencia entre los grupos de intervención (3 categorías) respecto a su IMC.

1. Plantear alternativa no paramétrica: Kruskal-Wallis

Salida

```
> with(predimed, tapply(bmi, group, median, na.rm = TRUE))
      Control MedDiet + Nuts  MedDiet + VOO
       30.000         29.455         29.735

> kruskal.test(bmi ~ group, data = predimed)

       Kruskal-Wallis rank sum test

data:  bmi by group
Kruskal-Wallis chi-squared = 21.8, df = 2, p-value = 0.00001846
```

Test de Kruskal-Wallis (3/3)

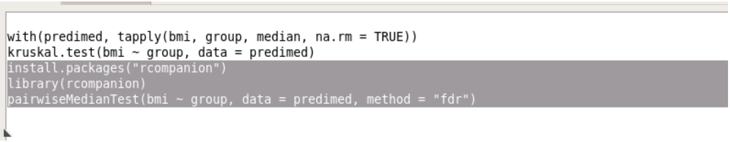
La posibilidad de ejecutar comparaciones múltiples dos a dos tras efectuar el test de Kruskal-Wallis todavía no está disponible en R Commander. Para ello se requiere del siguiente código:

```
install.packages("rcompanion")
rcompanion::pairwiseMedianTest(bmi ~ group, data = predimed)
```

Test de Kruskal-Wallis (3/3)

La posibilidad de ejecutar comparaciones múltiples dos a dos tras efectuar el test de Kruskal-Wallis todavía no está disponible en R Commander. Para ello se requiere del siguiente código:

```
install.packages("rcompanion")
rcompanion::pairwiseMedianTest(bmi ~ group, data = predimed)
```



Test de Kruskal-Wallis (3/3)

La posibilidad de ejecutar comparaciones múltiples dos a dos tras efectuar el test de Kruskal-Wallis todavía no está disponible en R Commander. Para ello se requiere del siguiente código:

```r
install.packages("rcompanion")
rcompanion::pairwiseMedianTest(bmi ~ group, data = predimed)
```

```
                        Comparison   p.value  p.adjust
1       Control - MedDiet + Nuts = 0 0.0001922 0.0005766
2        Control - MedDiet + VOO = 0   0.04217 0.0632600
3 MedDiet + Nuts - MedDiet + VOO = 0    0.0924 0.0924000
```

Asociación de dos variables categóricas (≥ 2 categorías)

Test χ^2 (1/3)

Cuando tenemos dos variables categóricas pero alguna o ambas tengan más de dos categorías, es posible utilizar el test χ^2. En base a una tabla de frecuencias, se comparan las frecuencias observadas con las esperadas en cada celda:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Test χ^2 (1/3)

Cuando tenemos dos variables categóricas pero alguna o ambas tengan más de dos categorías, es posible utilizar el test χ^2. En base a una tabla de frecuencias, se comparan las frecuencias observadas con las esperadas en cada celda:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

- El contraste de hipóstesis se plantea como:
  - H_0 : independencia entre variables
  - H_1 : dependencia entre variables

Test χ^2 (1/3)

Cuando tenemos dos variables categóricas pero alguna o ambas tengan más de dos categorías, es posible utilizar el test χ^2. En base a una tabla de frecuencias, se comparan las frecuencias observadas con las esperadas en cada celda:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

- El contraste de hipóstesis se plantea como:
  - H_0 : independencia entre variables
  - H_1 : dependencia entre variables
- Tiene como requisitos que la proporción de celdas con menos de 5 casos sea menor 0.2.

Test χ^2 (2/3)

El problema

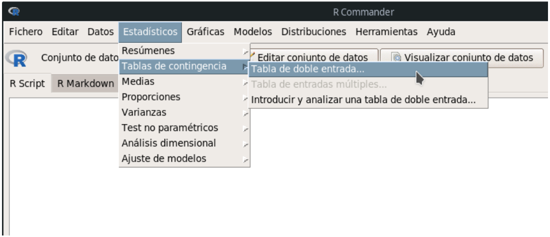
Se dispone de una base de datos en la que se quiere estudiar la posible asociación entre tabaquismo (4 categorías) y mortalidad (2 categorías) en 2000 personas.

Test χ^2 (2/3)

El problema

Se dispone de una base de datos en la que se quiere estudiar la posible asociación entre tabaquismo (4 categorías) y mortalidad (2 categorías) en 2000 personas.

1. Valorar los requisitos de la prueba χ^2 y resultados

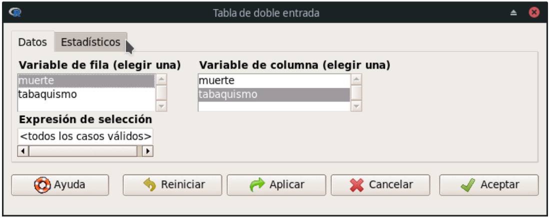


Test χ^2 (2/3)

El problema

Se dispone de una base de datos en la que se quiere estudiar la posible asociación entre tabaquismo (4 categorías) y mortalidad (2 categorías) en 2000 personas.

1. Valorar los requisitos de la prueba χ^2 y resultados

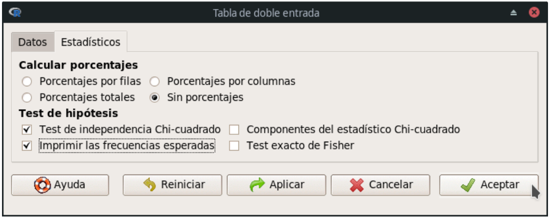


Test χ^2 (2/3)

El problema

Se dispone de una base de datos en la que se quiere estudiar la posible asociación entre tabaquismo (4 categorías) y mortalidad (2 categorías) en 2000 personas.

1. Valorar los requisitos de la prueba χ^2 y resultados



Test χ^2 (2/3)

El problema

Se dispone de una base de datos en la que se quiere estudiar la posible asociación entre tabaquismo (4 categorías) y mortalidad (2 categorías) en 2000 personas.

1. Valorar los requisitos de la prueba χ^2 y resultados

```
Frequency table:
      tabaquismo
muerte No fumador < 10 cig/día 10-20 cig/día > 20 cig/día
    No        687          145           336          113
    Sí        360           72           179           92
```

Test χ^2 (2/3)

El problema

Se dispone de una base de datos en la que se quiere estudiar la posible asociación entre tabaquismo (4 categorías) y mortalidad (2 categorías) en 2000 personas.

1. Valorar los requisitos de la prueba χ^2 y resultados

```
Expected counts:
      tabaquismo
muerte No fumador < 10 cig/día 10-20 cig/día > 20 cig/día
    No   676.0116    140.10938      332.5176    132.36139
    Sí   370.9884     76.89062      182.4824     72.63861
```

Test χ^2 (2/3)

El problema

Se dispone de una base de datos en la que se quiere estudiar la posible asociación entre tabaquismo (4 categorías) y mortalidad (2 categorías) en 2000 personas.

1. Valorar los requisitos de la prueba χ^2 y resultados

```
        Pearson's Chi-squared test

data:  .Table
X-squared = 9.0816, df = 3, p-value = 0.02823
```

Test de Fisher

Cuando no se cumple el requisito de la prueba χ^2 se puede recurrir a correcciones de la misma, estudios de simulación o al test de Fisher (que es lo que aquí se muestra).

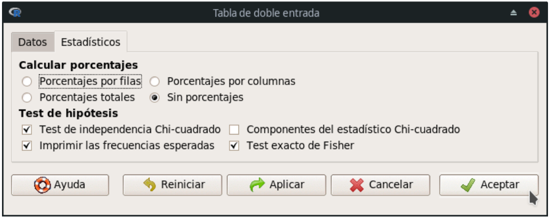
Test de Fisher

Cuando no se cumple el requisito de la prueba χ^2 se puede recurrir a correcciones de la misma, estudios de simulación o al test de Fisher (que es lo que aquí se muestra).

El problema

Imaginemos que disponemos de la misma base de datos que en el problema previo, pero esta es una selección aleatoria de la anterior con solo 100 personas.

1. Valorar los requisitos de la prueba χ^2 y plantear su alternativa no paramétrica: test de Fisher



Test de Fisher

Cuando no se cumple el requisito de la prueba χ^2 se puede recurrir a correcciones de la misma, estudios de simulación o al test de Fisher (que es lo que aquí se muestra).

El problema

Imaginemos que disponemos de la misma base de datos que en el problema previo, pero esta es una selección aleatoria de la anterior con solo 100 personas.

1. Valorar los requisitos de la prueba χ^2 y plantear su alternativa no paramétrica: test de Fisher

```
        tabaquismo
muerte No fumador < 10 cig/día 10-20 cig/día > 20 cig/día
    No         39            6            12            9
    Sí         18            4            10            2
```

Test de Fisher

Cuando no se cumple el requisito de la prueba χ^2 se puede recurrir a correcciones de la misma, estudios de simulación o al test de Fisher (que es lo que aquí se muestra).

El problema

Imaginemos que disponemos de la misma base de datos que en el problema previo, pero esta es una selección aleatoria de la anterior con solo 100 personas.

1. Valorar los requisitos de la prueba χ^2 y plantear su alternativa no paramétrica: test de Fisher

```
Expected counts:
      tabaquismo
muerte No fumador < 10 cig/día 10-20 cig/día > 20 cig/día
    No      37.62          6.6         14.52         7.26
    Sí      19.38          3.4          7.48         3.74
```

Test de Fisher

Cuando no se cumple el requisito de la prueba χ^2 se puede recurrir a correcciones de la misma, estudios de simulación o al test de Fisher (que es lo que aquí se muestra).

El problema

Imaginemos que disponemos de la misma base de datos que en el problema previo, pero esta es una selección aleatoria de la anterior con solo 100 personas.

1. Valorar los requisitos de la prueba χ^2 y plantear su alternativa no paramétrica: test de Fisher

```
        Pearson's Chi-squared test

data:  .Table
X-squared = 2.8222, df = 3, p-value = 0.4199
```

Test de Fisher

Cuando no se cumple el requisito de la prueba χ^2 se puede recurrir a correcciones de la misma, estudios de simulación o al test de Fisher (que es lo que aquí se muestra).

El problema

Imaginemos que disponemos de la misma base de datos que en el problema previo, pero esta es una selección aleatoria de la anterior con solo 100 personas.

1. Valorar los requisitos de la prueba χ^2 y plantear su alternativa no paramétrica: test de Fisher

```
        Fisher's Exact Test for Count Data

data:  .Table
p-value = 0.439
alternative hypothesis: two.sided
```