

PRÁCTICA 4

- a) Calcular e interpretar las siguientes medidas descriptivas para la tasa de alfabetización: mínimo, máximo, media, percentil 30, coeficiente de asimetría y coeficiente de apuntamiento.

Estadísticos → Resúmenes → Resúmenes numéricos

Variables: alfabet

	mean	skewness	kurtosis	min	25%	50%	75%	100%	30%	n	NA
	78.33645	-0.9802928	-0.2077413	18	63	88	98	100	72.8	107	2
media		Simetría	Apuntamiento						P.30		

- b) ¿Qué datos presentan una mayor dispersión relativa, los correspondientes a la esperanza de vida femenina o los datos de la tasa de alfabetización?

Estadísticos → Resúmenes → Resúmenes numéricos

Variables: alfabet y espvidaf

Resúmenes numéricos

Datos Estadísticos

☐ Media
☒ Desviación típica
☐ Error típico de la media
☐ Rango Intercuartílico
☐ Coeficiente de variación
☐ Recuentos binarios de frecuencias
☐ Asimetría ☒ Tipo 1
☐ Apuntamiento ☐ Tipo 2
☐ Tipo 3
☐ Cuantiles: 0, .25, .5, .75, 1, .30

Ayuda Reiniciar **Aceptar** Cancelar Aplicar

```

          sd   n NA
alfabet 22.88319 107 2
espvidaf 10.57178 109 0

```

Cogeríamos la de la esperanza de vida de la mujer, ya que es la menor

- c) Recodificar la variable *urbana* en otra variable de nombre *urbana_rec*, en cuatro intervalos igual amplitud, comenzando en 0 y terminando en 100. Construir tablas de contingencia adecuadas, para las variables *urbana_rec* y *región* para responder a las siguientes preguntas:

Datos → Modificar variables del conjunto de datos activo → Recodificar variables

Recodificar Variables

Variables a recodificar (elige una o más)

pib_cap
 poblac
 región
 región_cod
 religión
urbana

Nuevo nombre o prefijo para variables múltiples recodificadas:

☒ Convertir cada nueva variable en factor

Introducir directrices de recodificación

0:25=" [0, 25] "
 25:50=" (25, 50] "
 50:75=" (50, 75] "
 75:100=" (75, 100] "

Ayuda Reiniciar **Aceptar** Cancelar Aplicar

c1) ¿Qué porcentaje de países, de los que tienen un porcentaje en ciudades entre el 25 % y el 50 %, se encuentran en América Latina?

Estadísticos → Tablas de contingencia → Tabla de doble entrada

R Tabla de doble entrada

Datos Estadísticos

Variable de fila (elegir una)

país
región
religión
urbana_rec

Variable de columna (elegir una)

país
región
religión
urbana_rec

Expresión de selección

< todos los casos válidos

Ayuda Reiniciar **Aceptar** Cancelar Aplicar

R Tabla de doble entrada

Datos Estadísticos

Calcular porcentajes

☒ Porcentajes por filas → Cuando pide muchas cosas, es por filas

☐ Porcentajes por columnas → Cuando pide muchas cosas, es por filas

☐ Porcentajes totales

☐ Sin porcentajes

Test de hipótesis

☐ Test de independencia Chi-cuadrado

☐ Componentes del estadístico Chi-cuadrado

☐ Imprimir las frecuencias esperadas

☐ Test exacto de Fisher

Ayuda Reiniciar **Aceptar** Cancelar Aplicar

Frequency table:

	región					
urbana_rec	OCDE	Europa	Oriental	Asia / Pacífico	África	Oriente Medio
(25,50]	1		1	6	9	3
(50,75]	9		12	3	0	6
(75,100]	11		0	3	0	7
[0,25]	0		0	5	10	1

Row percentages:

	región								
urbana_rec	OCDE	Europa	Oriental	Asia / Pacífico	África	Oriente Medio	América Latina	Total	Count
(25,50]	3.7		3.7	22.2	33.3	11.1	25.9	99.9	27
(50,75]	22.5		30.0	7.5	0.0	15.0	25.0	100.0	40
(75,100]	44.0		0.0	12.0	0.0	28.0	16.0	100.0	25
[0,25]	0.0		0.0	31.2	62.5	6.2	0.0	99.9	16

25.9 (SOLUCION)

c2) ¿Cuál es el porcentaje de países de Oriente medio en los que más del 50 % de la población vive en ciudades?

Estadísticos → Tablas de contingencia → Tabla de doble entrada

Tabla de doble entrada

Datos Estadísticos

Variable de fila (elegir una)

- país
- región
- religión
- urbana_rec

Variable de columna (elegir una)

- país
- región
- religión
- urbana_rec

Expresión de selección

< todos los casos válidos >

Ayuda Reiniciar Aceptar Cancelar Aplicar

Tabla de doble entrada

Datos Estadísticos

Calcular porcentajes

☐ Porcentajes por filas

☒ Porcentajes por columnas → Cuando pide una sola cosa, es por columnas

☐ Porcentajes totales

☐ Sin porcentajes






Test de hipótesis

☐ Test de independencia Chi-cuadrado

☐ Componentes del estadístico Chi-cuadrado

☐ Imprimir las frecuencias esperadas

☐ Test exacto de Fisher

 Ayuda  Reiniciar  Aceptar  Cancelar  Aplicar

Frequency table:

	región					
urbana_rec	OCDE	Europa Oriental	Asia / Pacífico	África	Oriente Medio	América Latina
(25,50]	1	1	6	9	3	7
(50,75]	9	12	3	0	6	10
(75,100]	11	0	3	0	7	4
[0,25]	0	0	5	10	1	0

Column percentages:

	región					
urbana_rec	OCDE	Europa Oriental	Asia / Pacífico	África	Oriente Medio	América Latina
(25,50]	4.8	7.7	35.3	47.4	17.6	33.3
(50,75]	42.9	92.3	17.6	0.0	35.3	47.6
(75,100]	52.4	0.0	17.6	0.0	41.2	19.0
[0,25]	0.0	0.0	29.4	52.6	5.9	0.0
Total	100.1	100.0	99.9	100.0	100.0	99.9
Count	21.0	13.0	17.0	19.0	17.0	21.0

Sumamos 35.3 + 41.2, ya que nos piden que sea mayor que el 50%. Por lo tanto, el intervalo 50-75 y 75-100. **Da: 76.5**

c3) ¿Qué porcentaje de países pertenecen a la OCDE y tienen un porcentaje de habitantes que habitan en ciudades entre el 25 % y el 75 %?

Frequency table:

	región					
urbana_rec	OCDE	Europa Oriental	Asia / Pacífico	África	Oriente Medio	América Latina
(25,50]	1	1	6	9	3	7
(50,75]	9	12	3	0	6	10
(75,100]	11	0	3	0	7	4
[0,25]	0	0	5	10	1	0

Column percentages:

	región					
urbana_rec	OCDE	Europa Oriental	Asia / Pacífico	África	Oriente Medio	América Latina
(25,50]	4.8	7.7	35.3	47.4	17.6	33.3
(50,75]	42.9	92.3	17.6	0.0	35.3	47.6
(75,100]	52.4	0.0	17.6	0.0	41.2	19.0
[0,25]	0.0	0.0	29.4	52.6	5.9	0.0
Total	100.1	100.0	99.9	100.0	100.0	99.9
Count	21.0	13.0	17.0	19.0	17.0	21.0

Exactamente igual que el C2, pero sumando el intervalo de 25-50 y 50-75. **Da: 4.8 + 42.9 = 47,7**

d) Construir tablas de estadísticas adecuadas, que tengan como factores las variables *región* y *urbana_rec*, que permitan responder a las siguientes cuestiones sobre la ingesta diaria de calorías:

d1) ¿Cuál es la media de las ingestas diarias de calorías por habitante para los países de la OCDE en los que más del 75 % de los habitantes vive en ciudades?

Como anteriormente ya habíamos recodificado la variable urbana, ya no es necesario (APARTADO C1).

Una vez eso: Estadístico → Resúmenes → Tablas de estadísticas

	(25, 50]	(50, 75]	(75, 100]	[0, 25]
OCDE	NaN	3542.111	3329.222	NA
Europa Oriental	NaN	3399.500	NA	NA
Asia / Pacífico	2553.400	NaN	3077.000	2184.000
África	2253.714	NA	NA	2073.444
Oriente Medio	3336.000	2984.500	3131.000	NaN
América Latina	2396.167	2466.333	2732.250	NA

d2) Si consideramos los países en los que el porcentaje de habitantes en ciudades se encuentra entre el 25 % y el 50 %, cuál es la región en la que la ingesta diaria por habitante presenta una mayor variabilidad?

Lo mismo, pero con la Desviación típica:

Tabla de estadísticas

Factores (elegir uno o más): país, región, religión, urbana_rec

Variables explicadas (elegir una o más): alfabet, calorías, densidad, espvidaf, espvidam, fertilid

Estadístico:

☐ Media

☐ Mediana

☒ Desviación típica

☐ Rango intercuartílico

☐ Otro (especificar):

Ayuda Reiniciar Aceptar Cancelar Aplicar

	(25,50]	(50,75]	(75,100]	[0,25]
OCDE	NA	191.2410	223.3152	NA
Europa Oriental	NA	345.7752	NA	NA
Asia / Pacífico	240.6788	NA	171.1198	124.7905
África	147.1028	NA	NA	221.5362
Oriente Medio	NA	279.4167	231.7261	NA
América Latina	316.5807	332.4808	256.0617	NA

e) Construir un gráfico de sectores para la variable *religión*. Determinar el porcentaje de países que corresponde a cada sector.

Para ello nos vamos a Gráficas → Gráficas de sectores

Gráfica de sectores

Variable (elegir una): país, región, religión, urbana_rec

Selección de Color:

☒ Por defecto

☐ Desde la paleta de colores

Dibujar etiquetas:

Etiqueta del eje x: <auto>

Etiqueta del eje y: <auto>

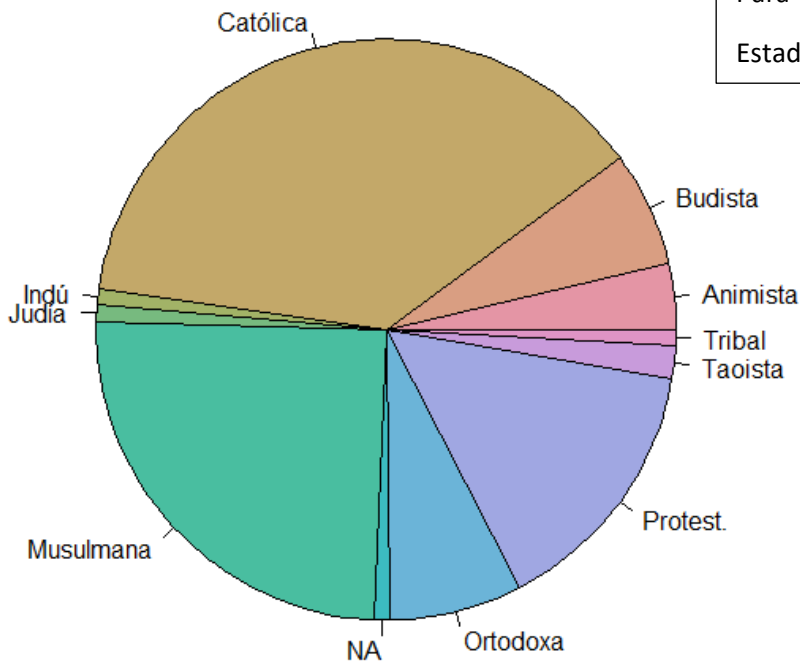
Título del gráfico: <auto>

Ayuda Reiniciar Aceptar Cancelar Aplicar

religión

Para ver los porcentajes:

Estadísticos → Resúmenes → Distribución de frecuencias



R Distribuciones de frecuencias

Variables (seleccione una o más)

país
región
religión
urbana_rec

☐ Test Chi-cuadrado de bondad de ajuste (sólo para una variable)

Ayuda Reiniciar **Aceptar** Cancelar Aplicar

counts:

religión

Animista	Budista	Católica	Indú	Judía	Musulmana	NA	Ortodoxa	Protest.
4	7	41	1	1	27	1	8	16
Taoista	Tribal							
2	1							

percentages:

religión

Animista	Budista	Católica	Indú	Judía	Musulmana	NA	Ortodoxa	Protest.
3.67	6.42	37.61	0.92	0.92	24.77	0.92	7.34	14.68
Taoista	Tribal							
1.83	0.92							

f) Construir un grafico de barras que muestre el número de países en los que cada religión es mayoritaria, agrupandolos según las distintas regiones. Elegir en la pestaña Opciones, de la ventana Gráfica de barras, la opción Lado a lado y comparar los resultados que se obtienen para los distintos grupos.

Gráficas → Gráficas de barras

R Gráfica de barras

Datos Opciones

Variable (elegir una)

país
región
religión
urbana_rec

Gráfica según: religión

R Gráfica de barras

Datos Opciones

Escala de los ejes

☒ Recuentos de frecuencias
☐ Porcentajes

Selección de Color

☒ Por defecto
☐ Desde la paleta de colores

Estilo del grupo de barras

☐ Dividido (apilado)
☒ Lado a lado (paralelo)

Porcentajes para Barras Agrupadas

☒ Condicional
☐ Total

Posición del texto

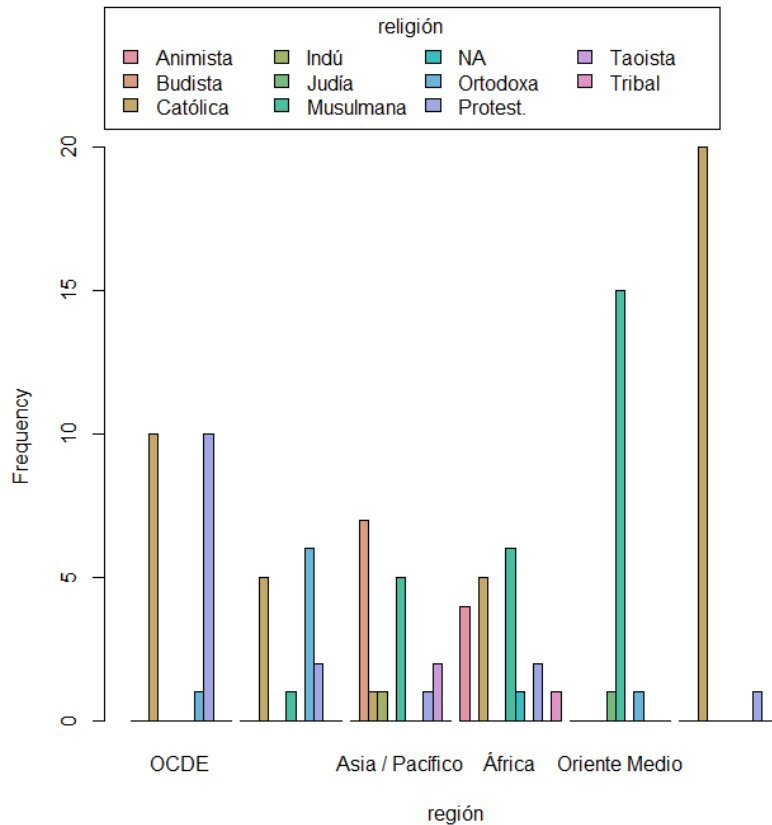
☒ Encima del gráfico
☐ Derecha
☐ Centro
☐ Izquierda

Dibujar etiquetas

Etiqueta del eje x <auto>
< >

Etiqueta del eje y <auto>
< >

Título del gráfico <auto>
< >

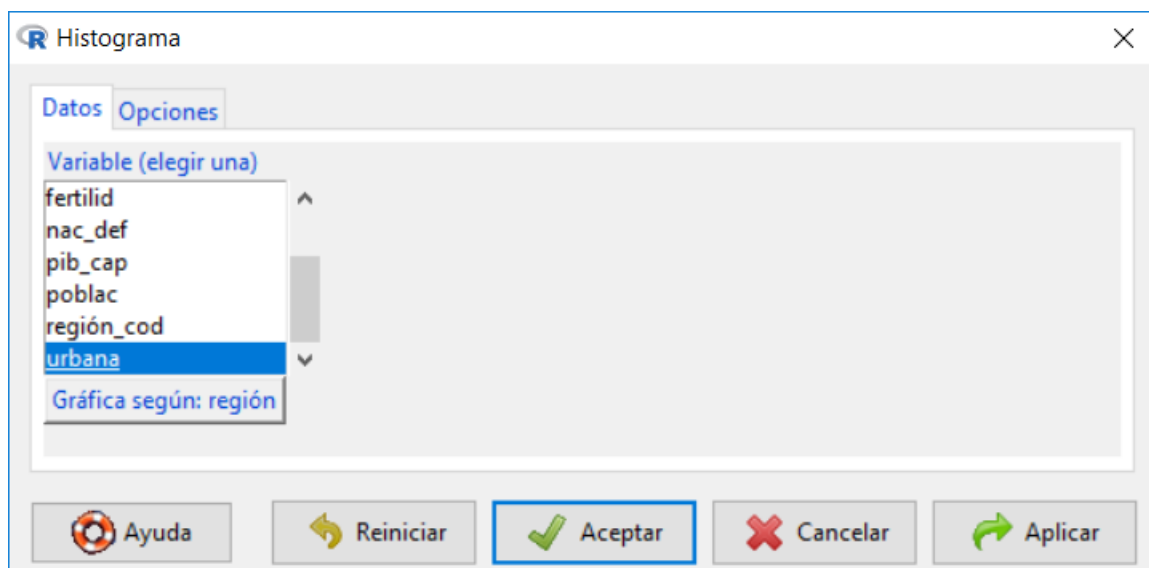


g) Construir histogramas que permitan estudiar el porcentaje de habitantes en ciudades, agrupando los datos según las distintas regiones. Utilizar cuatro intervalos para la construcción de los histogramas que comiencen en 0 y terminen en 100.

Podemos escribir en la terminal directamente: `with(Datos, Hist(urbana, groups=región, scale="percent", breaks=c(0,25,50,75,100), col="darkgray"))` → **TODO EN LA MISMA LÍNEA DE CÓDIGO**, y le damos a **Ejecutar**

O podemos también hacerlo por pasos:

Gráficas → Histograma



Histograma

Datos

Opciones

Opciones gráfica

Número de clases: <auto>

Escala de los ejes

☐ Recuentos de frecuencias
 ☒ Porcentajes
 ☐ Densidades

Dibujar etiquetas

Etiqueta del eje x: <auto>

Etiqueta del eje y: <auto>

Título del gráfico: <auto>

Ayuda

Reiniciar

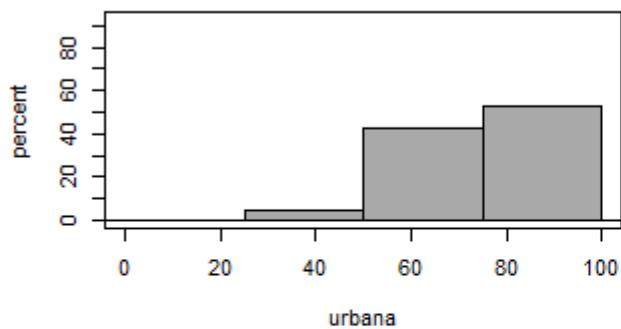
Aceptar

Cancelar

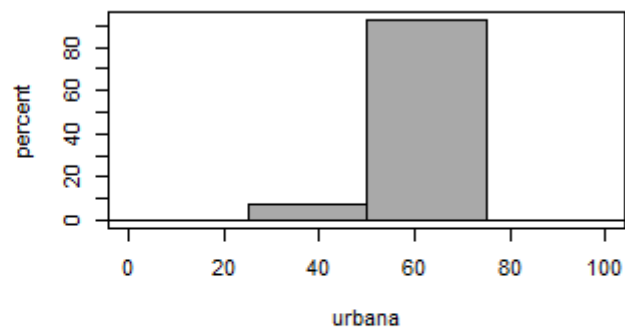
Aplicar

Ahora editamos en la consola, el apartado breaks, por los 4 intervalos que nos piden y le damos a Ejecutar:

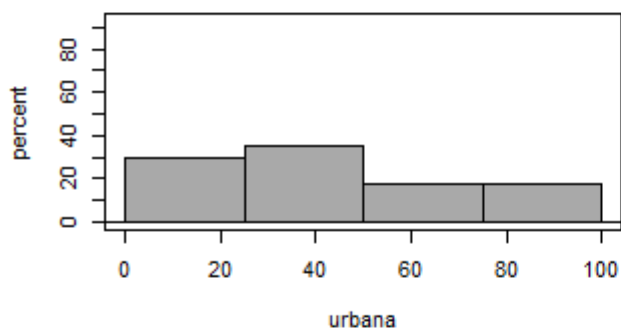
región = OCDE



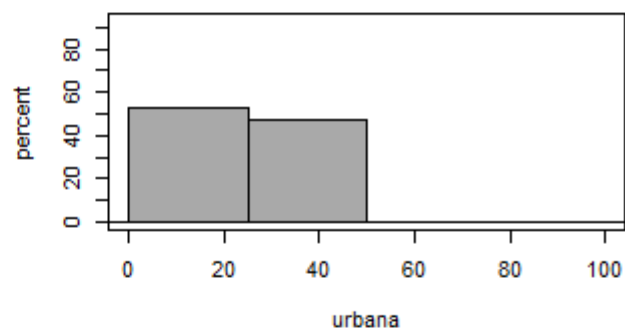
región = Europa Oriental



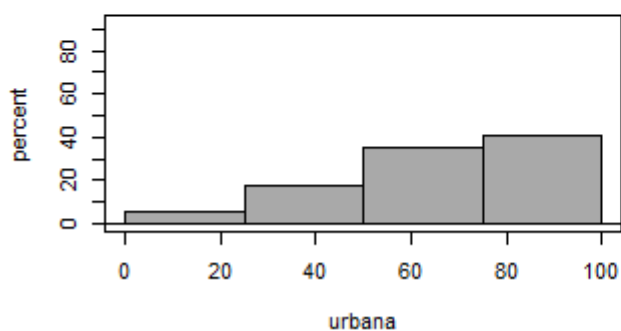
región = Asia / Pacífico



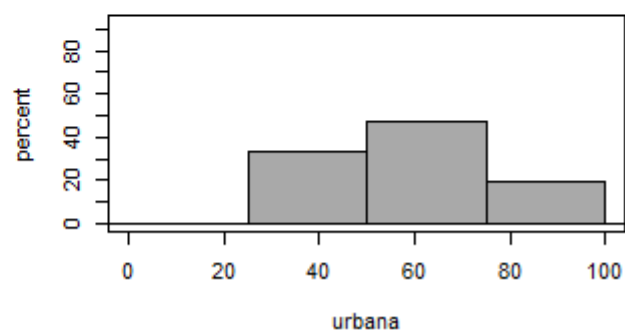
región = África



región = Oriente Medio



región = América Latina



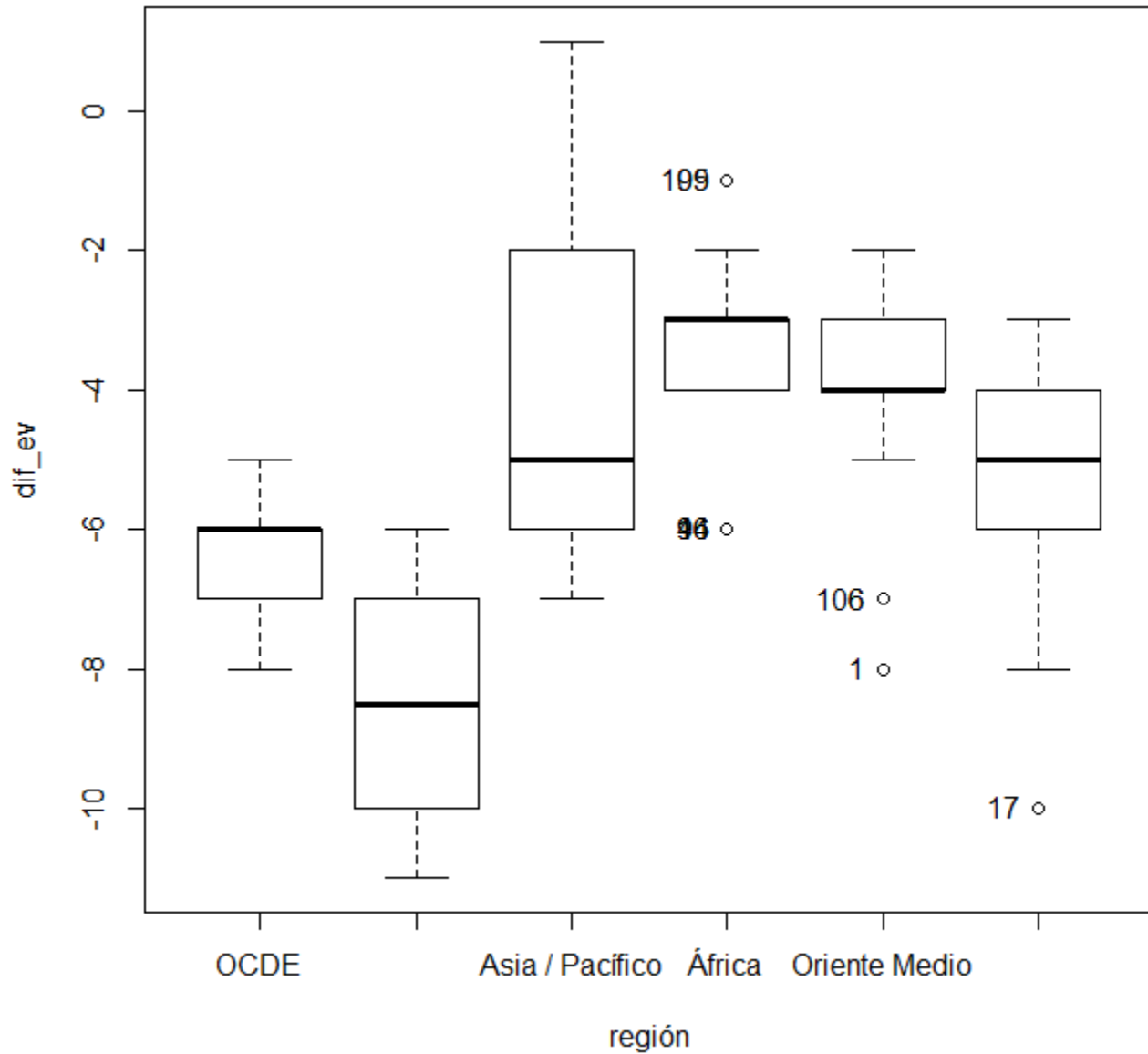
- h) Construir una variable de nombre *dif_ev* como la diferencia entre las variables *espvidam* y *espvidaf*. Representar e interpretar gráficos de cajas para la nueva variable, según las distintas regiones.

Para crear la nueva variable: Datos → Modificar variables del conjunto de datos activo → Crear nueva variable

The screenshot shows the 'Calcular una nueva variable' (Calculate a new variable) dialog box. It has a title bar with the R logo and a close button. The main area is divided into two sections. The top section, 'Variables actuales (doble clic para enviar a la expresión)', contains a list box with the following variables: 'alfabet', 'calorías', 'densidad', 'espvidaf' (which is highlighted), 'espvidam', and 'fertilid'. Below this list is a text field labeled 'Nombre de la nueva variable' containing the text 'dif_ev'. The bottom section, 'Expresión a calcular', contains a text field with the expression 'espvidam- espvidaf'. At the bottom of the dialog are five buttons: 'Ayuda' (Help), 'Reiniciar' (Reset), 'Aceptar' (Accept), 'Cancelar' (Cancel), and 'Aplicar' (Apply). The 'Aceptar' button is highlighted with a blue border.

Para la gráfica: Gráficas → Diagrama de Caja

The screenshot shows the 'Diagrama de caja' (Boxplot) dialog box. It has a title bar with the R logo and a close button. The dialog has two tabs: 'Datos' (Data) and 'Opciones' (Options). The 'Datos' tab is active. It contains a section labeled 'Variable (elegir una)' with a list box showing the following variables: 'alfabet', 'calorías', 'densidad', 'dif_ev' (which is highlighted), 'espvidaf', and 'espvidam'. Below the list box is a text field labeled 'Gráfica según: región'. At the bottom of the dialog are five buttons: 'Ayuda' (Help), 'Reiniciar' (Reset), 'Aceptar' (Accept), 'Cancelar' (Cancel), and 'Aplicar' (Apply). The 'Aceptar' button is highlighted with a blue border.



- i) Construir un gráfico XY para estudiar la esperanza de vida femenina en función de la tasa de fertilidad, eligiendo como condición el valor de la variable *región*. Interpretar los resultados obtenidos.

Gráficas → Gráficas XY

Gráfica XY

Datos Opciones

Variables explicativas (elegir una o más) Variables explicadas (elegir una o más)

espvda
 espvdam
fertilid
 nac_def
 pib_cap
 poblac

alfabet
 calorías
 densidad
 dif_ev
espvda
 espvdam

Condiciones 'l' (elija alguna o ninguna) Grupos 'groups=' (elija alguno o ninguno)

país
región
 religión

país
 región
 religión

Ayuda Reiniciar Aceptar Cancelar Aplicar

