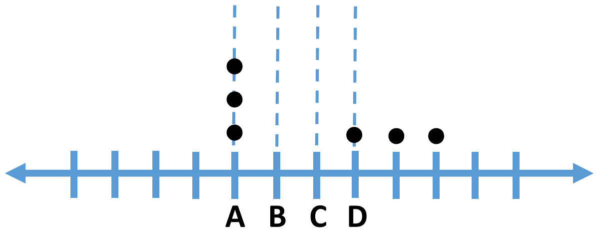
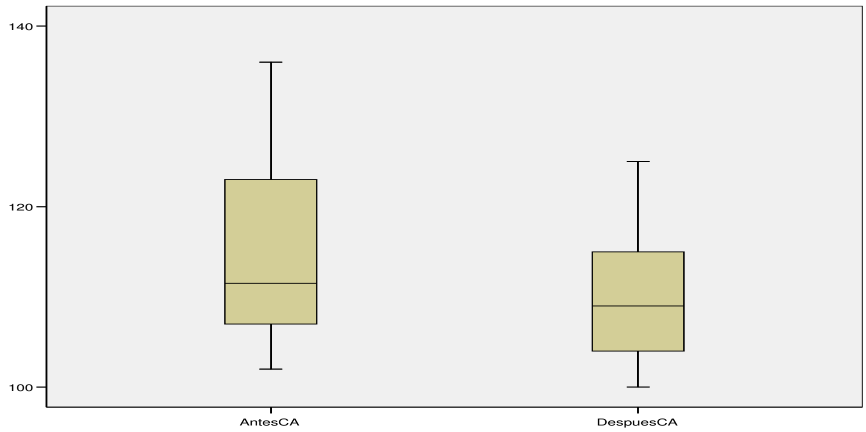
Relación de ejercicios propuestos

1. La media muestral. Si la muestra está formada por los puntos que se ven en la figura, ¿en cuál de las posiciones A, B, C, D está la media?



2. Media de clientes de un restaurante. Se ha registrado el número de clientes diarios de un restaurante de comida rápida durante 30 días, tanto en fin de semana como de lunes a viernes. Para los fines de semana (8 días) se obtuvo un número de clientes medio de 389,56, mientras que para los días entre semana (22 días) se obtuvo una media de 402,19. Calcula el número medio de clientes globales para los 30 días.

3. Consumo de calcio y tensión arterial. Para determinar el efecto del consumo de calcio sobre la tensión arterial, se midió la tensión arterial sistólica de 10 personas antes y después de recibir un suplemento de calcio en su dieta durante 12 semanas. La figura siguiente recoge los diagramas de cajas correspondientes a los datos registrados antes y después del tratamiento:



Determina razonadamente si las siguientes afirmaciones son verdaderas o falsas:

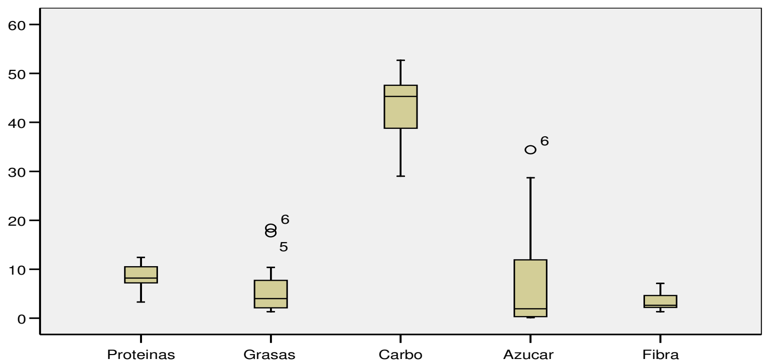
(a) Al menos 6 de las tensiones registradas después del tratamiento son inferiores a 120.

(b) El rango intercuartílico de las tensiones registradas antes del tratamiento es inferior al rango intercuartílico de las tensiones registradas después.
(c) Al menos el 75 % de las tensiones registradas antes del tratamiento es inferior a la máxima tensión registrada después.

4. **El estudio de la obesidad: el índice de Quetelet**. El *Índice de Masa Corporal* de una persona se define como $Q = P/E^2$, donde P es el peso en kilos y E es la estatura en metros. Este índice se llama también *Índice de Quetelet*, en honor del matemático y científico social belga Adolphe Quetelet (1796-1874) que fue el primero en proponerlo y estudiarlo. Se ha medido este índice en una muestra aleatoria de 125 personas. Los datos se encuentran en el fichero `imc.omv`.
(a) Resume gráficamente estos datos en un histograma.
(b) Calcula las medidas descriptivas básicas.

5. **Propiedades nutricionales de los alimentos de una panadería**. En un estudio sobre las propiedades nutricionales de los alimentos de una panadería se ha analizado la información contenida en las etiquetas de 25 productos. En primer lugar se ha llevado a cabo un análisis descriptivo básico de las siguientes variables relativas a 100 g de cada producto: `Proteinas` (proteínas en g), `Grasas` (grasas en g), `Carbo` (carbohidratos en g), `Azucar` (azúcares en g) y `Fibra` (fibra en g). Los resultados fueron los siguientes:

		Proteinas	Grasas	Carbo	Azucar	Fibra
N	Válidos	25	25	25	25	25
	Perdidos	0	0	0	0	0
Media		8,3720	5,6480	43,4800	8,0360	3,3720
Mediana		8,2000	4,0000	45,3000	1,9000	2,6000
Desv. típ.		2,68631	4,65789	5,87927	10,93114	1,66871
Mínimo		3,30	1,30	29,00	,10	1,30
Máximo		12,40	18,40	52,70	34,40	7,10
Percentiles	25	6,4500	2,1000	38,5500	,2500	2,2000
	50	8,2000	4,0000	45,3000	1,9000	2,6000
	75	10,5000	8,0000	47,7500	16,1000	4,7000



Determina razonadamente si las siguientes afirmaciones son verdaderas o falsas:
(a) En los diagramas de cajas se observa que valores altos en carbohidratos están asociados con valores bajos en fibra, por lo que la correlación entre ambas variables es negativa.
(b) Con la información disponible podemos afirmar que todas las variables tienen una distribución bastante simétrica.
(c) Con la información disponible podemos concluir que el producto etiquetado como 5 es un

dato atípico para todas las variables.

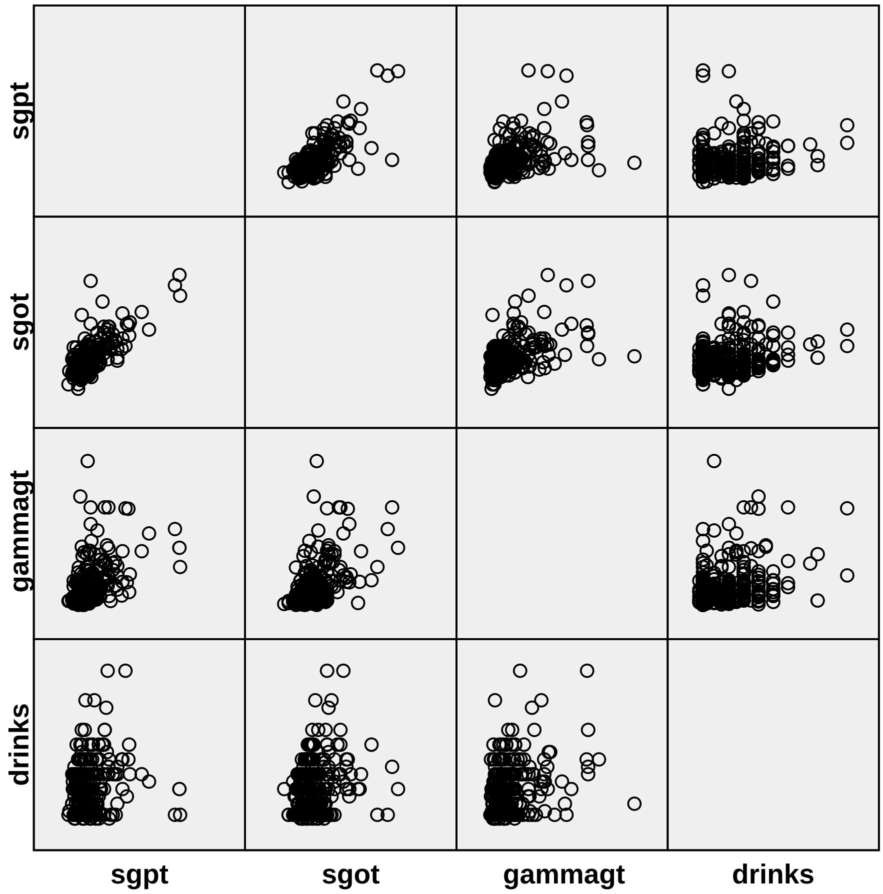
Analiza las relaciones existentes entre las variables representando la matriz de diagramas de dispersión y calculando las correlaciones. Los datos se encuentran en el fichero `panaderia.omv`.

6. **Afecciones hepáticas e ingesta de alcohol.** Se realizó un estudio sobre afecciones hepáticas que pueden aparecer a causa de la ingesta excesiva de alcohol midiendo 3 variables de interés en la analítica de un total de 345 varones.

- x_1 = `sgpt` (alanina aminotransferasa).
- x_2 = `sgot` (aspartato aminotransferasa).
- x_3 = `gammagt` (gamma-glutamil transpeptidasa).

Además de estas 3 variables en la analítica de la sangre se consideró la variable

- x_4 = `drinks` (número de medias pintas de cerveza o equivalentes en bebidas alcohólicas consumidas al día).



Estadísticos

		sgpt	sgot	gammagt	drinks
N	Válidos	345	345	345	345
	Perdidos	0	0	0	0
	Media	30,4058	24,643	38,284	3,4551
	Mediana	26,0000	23,000	25,000	3,0000
	Desv. típ.	19,51231	10,0645	39,2546	3,33784
	Mínimo	4,00	5,0	5,0	,00
	Máximo	155,00	82,0	297,0	20,00
Percentiles	25	19,0000	19,000	15,000	,5000
	50	26,0000	23,000	25,000	3,0000
	75	34,0000	27,000	46,500	6,0000
	90	52,0000	35,000	82,800	8,0000
	95	67,7000	44,400	115,000	10,0000
	99	131,9000	71,780	203,000	16,0000

Contesta razonadamente a las siguientes preguntas:

(a) Verdadero o falso: Aproximadamente 259 personas de entre las encuestadas beben el equivalente a 6 medias pintas o más de cerveza al día.

(b) Verdadero o falso: La persona que más alcohol consume al día (lo correspondiente a 20 medias pintas de cerveza) tiene un nivel de gamma-glutamil transpeptidasa en sangre de 297,0.

(c) ¿Qué variables tienen la mayor correlación entre ellas?

(d) Verdadero o falso: Todas las variables consideradas son simétricas.

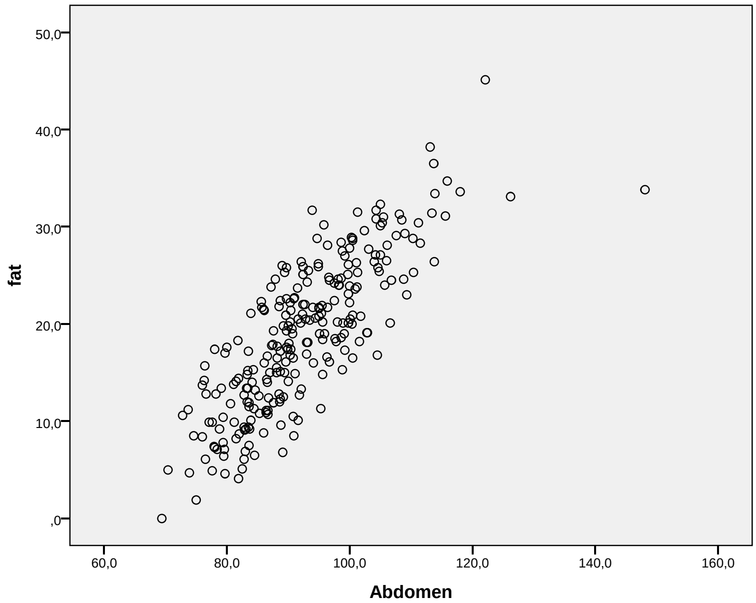
7. Porcentaje de grasa corporal y perímetro abdominal. Gran número de estudios afirman que el porcentaje de grasa corporal es un buen indicador de salud. En este ejercicio se pretende analizar la relación entre el porcentaje de grasa corporal y el perímetro abdominal. Se calculó de forma precisa el porcentaje de grasa corporal (variable `fat`) de 252 hombres mediante una técnica de pesado bajo el agua y utilizando la fórmula de Brozek. Además, se midió su perímetro abdominal (variable `Abdomen`) en centímetros desde el ombligo y al nivel de la cresta ilíaca.

Los resultados de los análisis estadísticos de estos datos se encuentran después de las preguntas.

(a) Calcula la varianza de la variable `Abdomen`. ¿En qué unidades se expresa?

(b) ¿Aproximadamente cuántos hombres de la muestra tienen un porcentaje de grasa corporal superior al $24,6\,\%$?

(c) Identifica cada una de las observaciones 39, 41 y 216 en el diagrama de dispersión de las variables `Abdomen` y `fat`.

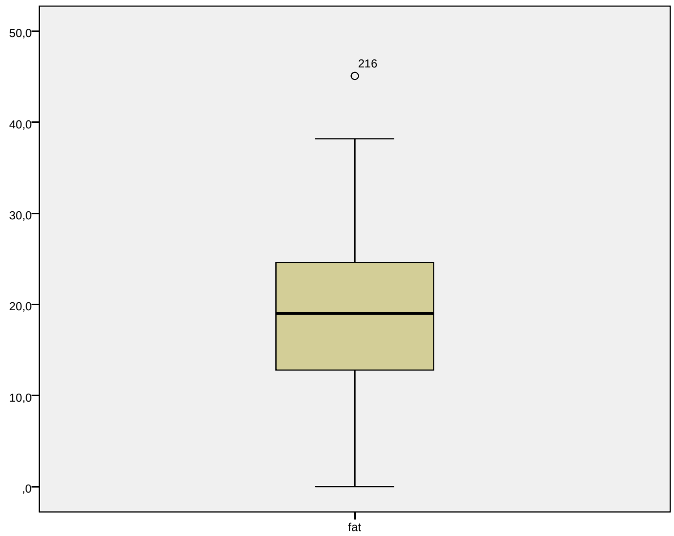


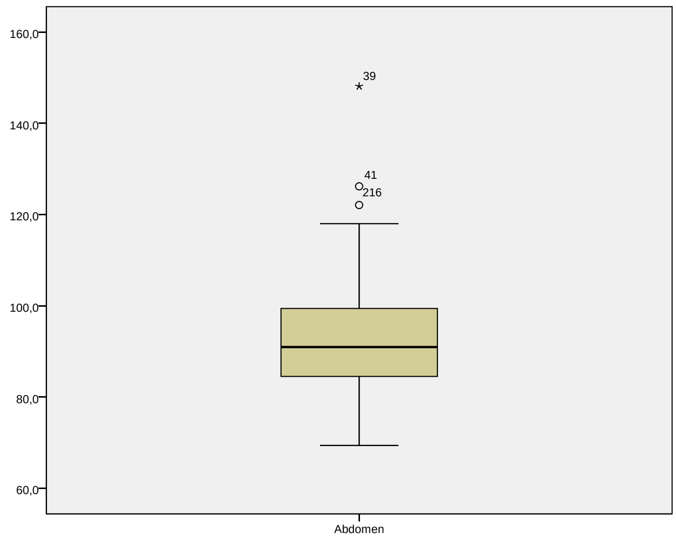
Estadísticos

		fat	Abdomen
N	Válidos	252	252
	Perdidos	0	0
Media		18,938	92,556
Mediana		19,000	90,950
Desv. típ.		7,7509	10,7831
Mínimo		,0	69,4
Máximo		45,1	148,1
Percentiles	25	12,800	84,525
	50	19,000	90,950
	75	24,600	99,575

Correlaciones

		fat	Abdomen
Correlación de Pearson	fat	1,000	,814
	Abdomen	,814	1,000
Sig. (unilateral)	fat	.	,000
	Abdomen	,000	.
N	fat	252	252
	Abdomen	252	252





8. Medidas adimensionales. Consideramos la siguiente lista de medidas utilizadas en estadística:

coeficiente de correlación; varianza; media; cuartil 1; coeficiente de variación; covarianza; mediana; rango intercuartílico; rango; desviación típica.

(a) Clasifica las cantidades de la lista anterior en alguno de los tres grupos siguientes:
1. Medidas de posición de la distribución de un conjunto de datos.
2. Medidas de dispersión de la distribución de un conjunto de datos.
3. Cantidades no utilizadas para medir ni la posición ni la dispersión.

(b) De las medidas de la lista, enumera todas aquellas cuyo valor no dependa de las unidades en las que se expresen los datos (es decir, las medidas adimensionales).

9. Contaminación por mercurio en la caza (cálculo de probabilidades con el modelo normal). En 1969 se descubrió que los faisanes de Montana (EE.UU) padecían una apreciable contaminación por mercurio que podía deberse a que habían comido semillas de plantas que fueron tratadas durante su crecimiento con metilo de mercurio. Sea X el nivel de mercurio de un pájaro en partes por millón. Supongamos que X tiene distribución normal con media 0.25 y desviación típica 0.08. Calcula $P(X \leq 3)$, $P(X \geq 0{,}17)$, $P(0{,}2 \leq X \leq 0{,}4)$ y $P(0{,}01 \leq X \leq 0{,}49)$.

10. Contenido en calcio del agua (cálculo de probabilidades con el modelo normal). El contenido en calcio de las aguas de una cierta marca de agua mineral sigue una distribución normal con media 320 mg/l y desviación típica 10 mg/l.
(a) Si se elige al azar una botella de agua mineral de esa marca ¿cuál es la probabilidad de que la concentración de calcio en el agua de esa botella sea menor que 300 mg/l?
(b) Si se eligen al azar 10 botellas de esa marca ¿cuál es la probabilidad de la concentración promedio de las 10 botellas sea menor que 310 mg/l?
(c) Se eligen al azar dos botellas de esa marca. ¿Cuál es la probabilidad de que la diferencia entre sus concentraciones de calcio sea menor que 10 mg/l?

11. Consumo diario de calorías (cálculo de probabilidades con el modelo normal) Según un estudio realizado en Hong-Kong, la distribución del consumo diario de kcal. en la dieta de la población de los hombres adultos jóvenes (menores de 34 años) es aproximadamente normal con una media de 2557 y una desviación típica de 723.
(a) ¿Qué porcentaje de hombres adultos jóvenes consumen diariamente más de 2700 kcal?
(b) Determina el valor C que cumple la condición de que exactamente el 10 % de esta población tiene un consumo diario de kcal mayor que C.
(c) Se seleccionan aleatoriamente 100 hombres en esta población. Calcula la probabilidad aproximada de que más de la mitad de ellos consuma más de 2700 kcal.

12. Estimación de la ingesta media diaria de calorías Utilizando una muestra pequeña se ha estimado que el consumo medio diario de calorías en mujeres jóvenes es de 2300 kcal y la desviación típica estimada es de 237 kcal. Se desea repetir el estudio para obtener una estimación más precisa del consumo medio. En concreto, se desea obtener un estimación que, con probabilidad 0.99, difiera del valor verdadero en, a lo sumo, 80 kcal. ¿Qué tamaño debe tener esta muestra? [Basado en Tuschl et al. (1990), *American Journal of Clinical Nutrition*, julio de 1990, 81-86].

13. Estimación mediante intervalos de confianza de la ingesta media diaria de vitamina D. La ingesta diaria media de vitamina D en una muestra de 36 escolares de educación primaria ha resultado ser de 4.5 μg y la correspondiente cuasi-desviación típica es de 2.02.
(a) Calcula un intervalo de confianza aproximado de nivel 0.95 para estimar la ingesta media de vitamina D en la población de escolares de primaria.
(b) Calcula el tamaño muestral necesario para tener una probabilidad 0.95, de cometer un error menor que 0.3 en la estimación de esta ingesta media.

14. Determinación del tamaño muestral en una encuesta sobre hábitos nutricionales. Se va a realizar un estudio para estimar la proporción p de personas que tienen hábitos alimenticios inadecuados en una gran ciudad, en lo que respecta al consumo de grasas animales y de frutas y verduras. En una pequeña muestra piloto de 20 personas seleccionadas al azar ha resultado que 6 tenían hábitos alimenticios inadecuados.
(a) ¿Qué tamaño muestral habría que elegir para estimar p cometiendo un error menor que 0.02 con una probabilidad de 0.95?
(b)Suponiendo que en una muestra de 500 personas ha resultado que 124 tenían hábitos alimenticios inadecuados, calcula un intervalo de confianza de nivel 0.99 para estimar p.

15. Determinación del tamaño muestral para estimar la media poblacional Se desea estimar la contaminación (en partes por millón) por metales pesados en el pescado que se vende en una lonja. Para ello se ha seleccionado una pequeña muestra de 12 unidades resultando que la contaminación media observada en ellos fue de 0.79 con una cuasi-desviación típica de 0.13. ¿Qué tamaño muestral habría que utilizar para estimar la contaminación media con un error máximo de 0.01 a un nivel de confianza de 0.95?

16. Intervalo de confianza para la media de una distribución normal. Se desea estimar el contenido medio de grasas (en gramos por cada 100 gr.) de la carne de cerdo. Supongamos que se dispone de los siguientes resultados correspondientes a la carne de 12 animales elegidos al azar:

24.1, 24.7, 25.3, 25.8, 26.3, 23.4, 25.2, 25.9, 24.7, 23.8, 24.4, 25.6.

Obtén un intervalo de confianza de nivel 0.95 para el contenido medio de grasas. Indica claramente las suposiciones previas necesarias para garantizar la validez del procedimiento empleado.

17. Consumo de carne en Estados Unidos. Los siguientes datos corresponden al consumo de carne (libras por persona consumidas en un año) en una muestra de 16 personas elegidas al azar en barrios de clase media de pequeñas ciudades norteamericanas:

118, 110, 117, 120, 119, 126, 115, 112, 112, 113, 122, 125, 130, 115, 118, 123

(a) Da un intervalo de confianza de nivel 0.95 para el consumo medio.
(b) Con estos datos, ¿se puede afirmar con un nivel de significación de 0.05, que el consumo medio de carne es superior a 115 libras por persona?

18. **Comparación de los niveles de contaminación en la pesca procedente de dos ríos cercanos.** Se han analizado con *jamovi* los datos del fichero `mercurio.omv` con el fin de analizar si el nivel medio de contaminación por mercurio en los dos ríos es o no diferente. La salida obtenida ha sido la siguiente:

Independent Samples T-Test

Independent Samples T-Test

		statistic	df	p
CONC	Student's t	−1.69 [a]	169	0.046

Note. H_a 0 < 1

[a] Levene's test is significant (p < .05), suggesting a violation of the assumption of equal variances

Group Descriptives

	Group	N	Mean	Median	SD	SE
CONC	0	73	1.08	0.910	0.649	0.0759
	1	98	1.28	1.05	0.829	0.0838

(a) ¿Existe evidencia estadística para afirmar al nivel $\alpha = 0{,}05$ que el nivel medio de concentración en el río Wacamaw (1) es superior al nivel medio en el río Lumber (0)?
(b) Indica las suposiciones previas necesarias para garantizar la validez del procedimiento empleado.

19. **Sobre las propiedades antioxidantes del vino tinto.** La aterosclerosis coronaria (AC) es una de las principales causas de mortalidad en los países occidentales. Se cree que la oxidación del colesterol de baja densidad [ver, por ejemplo, Steinberg et al. (1989), *New Engl. J. Med.*, 320, 915-924] es un importante mecanismo en el desarrollo de la AC. Hay una interesante polémica científica sobre los supuestos efectos antioxidantes (y, por tanto, beneficiosos contra la AC) de las bebidas alcohólicas consumidas en cantidades moderadas. Algunos autores han puesto en duda estos efectos cardioprotectores, otros los han atribuido al alcohol en sí mismo y, por último, otros [ver, por ejemplo, Gorinstein et al., 2002, *Nutrition Research* 22, 659-666, para una visión más amplia de este tema] atribuyen la mayor parte de los efectos beneficiosos a las sustancias fenólicas que están contenidas en el vino tinto en mucha mayor medida que en otras bebidas alcohólicas. Según los defensores de esta última teoría, el vino tinto (consumido siempre en cantidades muy moderadas) sería mucho más cardiosaludable que las demás bebidas alcohólicas.

Se han determinado los valores de epicatequina (una sustancia fenólica) en 10 muestras de vino tinto, encontrando que la media muestral era 195.1 mg/l y el error típico 10.051. Los correspondientes valores para 10 muestras de cerveza fueron 65.5 mg/L y 3.4184. En vista de estos resultados ¿puede decirse que son significativamente diferentes los contenidos medios de esta sustancia en el vino tinto y en la cerveza?

Responde a la misma pregunta para el caso del ácido ferúlico en el que los resultados obtenidos fueron, para el vino tinto, 7.2 mg/l (media) y 0.4541 mg/l (error típico) (tamaño muestral = 10) y para la cerveza 6.8 mg/l (media) y 0.3571 mg/l (error típico) (tamaño muestral = 10).

[Datos adaptados de Gorinstein et al. (2000), *Nutrition Research*, 20, 131-139]

20. **Comparación de algunos efectos nutricionales de la carne y el pescado.** Se realiza un experimento para comparar los incrementos en los niveles plasmáticos de insulina producidos por la ingesta de carne y de pescado. Para ello se midieron los incrementos (medidos en picomoles por litro) producidos en la concentración de insulina en la sangre de 6 voluntarios, 90 minutos después de comer un bistec de 250 gr. Dos días más tarde se realizó de nuevo el experimento con las mismas 6 personas, después de consumir un filete de pescado. En la siguiente tabla se indican los respectivos incrementos en la concentración de insulina producidos por la carne y por el pescado:

Persona:	1	2	3	4	5	6
Resultados con la carne:	109	106	111	105	110	108
Resultados con el pescado:	100	95	105	106	80	88

(a) ¿Proporcionan estos datos suficiente evidencia estadística, al nivel de significación 0.05, para afirmar que el incremento medio en la concentración de insulina producido por el consumo de carne es mayor que el producido por el consumo de pescado? Responde a la misma pregunta anterior utilizando un nivel de significación de 0.01. Indica las suposiciones previas necesarias para garantizar la validez del procedimiento empleado.
(b) Calcula el tamaño muestral que sería necesario para estimar el incremento medio producido por el consumo de carne, de manera que se tenga una probabilidad 0.95 de cometer, como máximo, un error de 0.2 unidades.

21. **El efecto del consumo de ajo en los niveles de colesterol y triglicéridos.** En un estudio [R. Aouadi et al., (2000) *Nutrition Research* **20**, 273-280] sobre el efecto de la ingesta de ajo en los niveles de colesterol se analizaron cuatro grupos, G1-G4, de ratas de forma que en cada uno de estos grupos se suministró a los animales una dieta distinta durante un período de 12 semanas. A continuación se midieron diversas variables y, en particular, los niveles de colesterol y triglicéridos.

Las ratas del grupo control (G1) recibieron una dieta estándar, las del grupo G2 recibieron la dieta estándar complementada con un 10 % de ajo, las del grupo G3 recibieron la dieta estándar con un 2 % de colesterol y las del grupo G4 recibieron dos complementos de un 10 % de ajo y de un 2 % de colesterol en la dieta estándar.

Los resultados obtenidos al final del estudio en lo relativo a los niveles plasmáticos de colesterol y triglicéridos, se resumen (expresados en la forma media ± desviación típica) en la siguiente tabla:

Grupo	n^0 de ratas	Colesterol en mg/dl	Triglicéridos en mg/dl
G1	8	72.00 ± 8.32	59.89 ± 6.55
G2	8	63.21 ± 7.24	50.25 ± 4.40
G3	8	94.33 ± 24.50	70.65 ± 10.43
G4	7	65.03 ± 10.57	72.97 ± 11.02

(a) Construye intervalos de confianza de nivel 0.95 para los niveles medios de colesterol en los cuatro grupos.
(b) ¿Es significativamente más bajo el nivel de colesterol en los animales del grupo G2 que en los del grupo G1? (utiliza $\alpha = 0{,}05$)

Indica en cada caso cuáles son las suposiciones previas necesarias para garantizar la validez de los métodos empleados.

22. Consumo de productos lácteos en la población universitaria española. En un estudio sobre el consumo de productos lácteos en la población universitaria española, realizado a partir de una muestra de 380 mujeres y 120 hombres, se ha determinado que 260 mujeres y 87 hombres de la muestra consumen leche entera.
(a) Calcula, a partir de los datos anteriores, un intervalo de confianza de nivel 95 % para la proporción poblacional de hombres que consumen leche entera.
(b) ¿Permiten los datos afirmar a nivel $\alpha = 0{,}05$ que, en la población, la proporción de hombres que consumen leche entera es diferente a la de mujeres?

23. Sobre el efecto de las estatinas y el aceite de pescado en el nivel de colesterol.
(a) En un estudio realizado recientemente en Suecia (Liu et al. *Nutrition Resarch*, 2003, 23, pp. 1023-1034) se analizó el efecto del aceite de pescado y el de la Simvastatina (un fármaco para reducir el colesterol) sobre los niveles de colesterol de baja densidad (el llamado "colesterol malo", LDL). Para ello se suministró una dosis de 10 mg diarios de Simvastatina a una muestra de 18 pacientes diagnosticados de hiperlipemia y se observó que, en promedio, su concentración de LDL (en mmol/L) después de 12 semanas de tratamiento fue de 3.05 y la (cuasi-) desviación típica fue 0.70. En otra muestra control de 22 pacientes, que no tomaron ningún medicamento y sólo siguieron unas recomendaciones dietéticas básicas de bajo consumo de grasas, los correspondientes valores fueron 4.04 y 0.84. ¿Proporcionan estos datos suficiente evidencia estadística, al nivel de significación 0.05, para afirmar que los pacientes tratados con Simvastatina tienen una concentración media de LDL menor que la de los pacientes no tratados?. Indica las suposiciones previas necesarias para garantizar la validez del procedimiento empleado.
(b) En el mismo estudio se incorporaron 10 ml diarios de aceite de pescado a la dieta de una muestra de 29 pacientes y se observó que la reducción media (concentración media antes del tratamiento menos concentración media después) en la concentración de LDL fue 0.32 y la (cuasi-) desviación típica de las disminuciones observadas fue 0.75. ¿Proporcionan estos datos suficiente evidencia estadística, al nivel de significación 0.05, para afirmar que el aceite de pescado ha reducido la concentración media de LDL? Responde a la misma pregunta tomando el nivel de significación 0.01. Indica las suposiciones previas necesarias para garantizar la validez del procedimiento empleado.

24. Relación entre el tamaño de los peces y su nivel de contaminación. Utilizando los datos del fichero `mercurio.omv`, se ha estudiado la regresión lineal entre la longitud de los peces y el nivel de contaminación por mercurio. Los resultados obtenidos con *jamovi* han sido los siguientes:

Linear Regression

Model Fit Measures

Model	R	R^2
1	0.650	0.422

Model Coefficients

Predictor	Estimate	SE	t	p
Intercept	−1.1316	0.21361	−5.30	< .001
LONG	0.0581	0.00523	11.12	< .001

(a) ¿Cuál es el coeficiente de correlación entre la longitud del pez y la concentración de mercurio?
(b) ¿Existe evidencia estadística para afirmar a nivel $\alpha = 0{,}05$ que un incremento de la longitud del pez produce un incremento en el nivel medio de concentración de mercurio?
(c) Ajusta el modelo de regresión lineal:

$$\log(\texttt{CONC}) = \beta_0 + \beta_1 \log(\texttt{PESO}) + \epsilon.$$

Determina los estimadores de los parámetros y contrasta la hipótesis $\mathrm{H}_0 : \beta_1 = 0$ a nivel $\alpha = 0{,}05$.

25. **Estudio de la efectividad de un tratamiento.** Los siguientes datos corresponden al peso, en libras, de 10 pacientes de diabetes cuando son diagnosticados (variable x) y al cabo de un año de tratamiento (variable y).

Paciente	1	2	3	4	5	6	7	8	9	10
x_i	140	160	180	120	132	146	190	200	120	145
y_i	115	130	135	125	112	130	160	160	123	143

(a) Evalúa el grado de asociación lineal entre las variables x e y y comentar brevemente el resultado.
(b) Supongamos que el peso de un paciente al ser diagnosticado fuera de $x_0 = 150$ lb. Utiliza un modelo de regresión lineal para predecir el peso de este paciente después de un año de tratamiento.

26. **Estudio, mediante regresión lineal, de la capacidad del organismo humano para absorber hierro y plomo.** Se ha realizado un estudio por medio de técnicas de seguimiento de marcadores radiactivos para analizar la capacidad del organismo para absorber hierro y plomo. Participan en el estudio diez personas. A cada una de ellas se le da una dosis oral idéntica de hierro (en forma sulfato ferroso) y de plomo (cloruro de plomo-203). Después de doce días se mide la cantidad de cada componente retenida en el organismo y, a partir de ésta, se determina el porcentaje absorbido por el cuerpo. Se obtuvieron los siguientes datos (donde x_i e y_i representan, respectivamente, los porcentajes de hierro y plomo absorbidos por el individuo i-ésimo de la muestra considerada):

x_i	17	22	35	43	80	85	91	92	96	100
y_i	8	17	18	25	58	59	41	30	43	58

(a) Ajusta un modelo de regresión lineal y discutir su idoneidad para describir la relación entre x e y.
(b) Si el modelo lineal resulta apropiado, utilízalo para predecir el porcentaje de plomo absorbido por un individuo cuyo organismo absorbe el 15 % del hierro ingerido.
(c) ¿Hay suficiente evidencia estadística (al nivel 0.05) para afirmar que hay una correlación lineal positiva entre el porcentaje de plomo y el de hierro?

27. **Producción de insulina y péptido C en sangre.** Una manera de controlar la correcta producción de insulina consiste en medir la concentración de péptido C en la sangre, puesto que éste se libera cuando se produce insulina. En un estudio sobre diabetes se ajustó un modelo de regresión lineal a una muestra de 43 individuos con el fin de estudiar el logaritmo de la concentración de péptido C (pmol/ml) en función de la edad. Los datos se analizaron con *jamovi* con los siguientes resultados:

Model Fit Measures

Model	R	R^2
1	0.464	0.215

Model Coefficients

Predictor	Estimate	SE	t	p
Intercept	3.9960	0.2446	16.34	< .001
edad	0.0831	0.0248	3.35	0.002

(a) Escribe la ecuación de la recta de mínimos cuadrados y predice el valor de la variable respuesta para un individuo de 2 años de edad.
(b) Calcula un intervalo de confianza para la pendiente de la recta de regresión. ¿Podemos afirmar a nivel $\alpha = 0{,}05$ que hay relación lineal entre la edad y la concentración de péptido C?

28. **Niveles de beta-caroteno, ingesta de grasas e IMC.** En un estudio[1] se analizaron los factores que determinan la cantidad de algunos micronutrientes en la sangre mediante una muestra de 314 pacientes. Con los datos de este estudio se han ajustado con *jamovi* dos modelos de regresión simple en los que se estudia cómo varía el logaritmo de la cantidad de β-caroteno en el plasma sanguíneo (en ng/ml) en función de la ingesta diaria de grasas (en g, variable *fat*) y en función del índice de masa corporal (variable *Quet*), respectivamente. Los resultados obtenidos son los siguientes (algunos datos han sido suprimidos de la salida):

[1] Nierenberg *et al.* (1989). Determinants of plasma levels of beta-carotene and retinol. *American Journal of Epidemiology* , 130, 511–521.

Model Coefficients

Predictor	Estimate	SE	t	p
Intercept	5.17095	0.10495	49.27	< .001
Fat	-0.00275	0.00125	-2.20	0.029

Model Coefficients

Predictor	Estimate	SE	t	p
Intercept	5.8690	0.18117	32.39	< .001
Quet	AA	0.00675	-5.15	< .001

(a) A partir de estos datos, determina si es posible afirmar a nivel $\alpha = 0{,}05$ que la ingesta diaria de grasas está relacionada linealmente con el logaritmo del nivel de β-caroteno.
(b) Calcula el valor señalado como `AA` en la salida anterior, así como un intervalo de confianza de nivel 95 % para la pendiente de la recta de regresión que relaciona el β-caroteno con el índice de masa corporal.
(c) Utilizando los datos del mismo estudio del ejercicio anterior se compararon los logaritmos de los niveles de β-caroteno en individuos fumadores y no fumadores, obteniéndose los siguientes valores para cada grupo:

	F/NF	logBeta
N	Fumadores	156
	No Fumadores	158
Mean	Fumadores	5.07
	No Fumadores	4.85
Std. error mean	Fumadores	AA
	No Fumadores	BB
Standard deviation	Fumadores	0.714
	No Fumadores	0.766

¿Permiten los datos afirmar a nivel $\alpha = 0{,}05$ que la media del logaritmo de la cantidad de β-caroteno es más alta en individuos fumadores que en individuos que no fuman? Enumera las condiciones bajo las cuales es válida tu respuesta. Calcula el valor de `AA` y `BB` en la tabla anterior.

29. **Beneficios de la lactancia materna.** Las recomendaciones acerca de cuánto tiempo debería alimentarse sólo con lactancia materna a los bebés de países en vías de desarrollo son controvertidas. Si la calidad nutricional de la leche materna es inadecuada porque las madres están malnutridas, entonces existe el riesgo de que el bebé también esté malnutrido. Por otro lado, la introducción de otros alimentos acarrea el riesgo de infección por contaminación. La situación se complica más todavía si se tiene en cuenta que las compañías que producen leche y otros productos para lactantes se benefician cuando hay muchos consumidores de estos alimentos. Como parte de un estudio acerca de la ingesta de energía en lactantes alimentados de diferentes maneras, se midió la ingesta de energía (en kcal/día) en lactantes alimentados sólo con leche materna durante 4, 5 ó 6 meses. Los datos aparecen a continuación.

Lactancia materna durante		
4 meses	5 meses	6 meses
499	490	585
620	395	647
469	402	477
485	177	445
660	475	485
588	617	703
675	616	528
517	587	465
649	528	
209	518	
404	370	
738	431	
628	518	
609	639	
617	368	
704	538	
558	519	
653	506	
548		

Al hacer un análisis de la varianza obtenemos los siguientes resultados:

ANOVA de un factor

Energia

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	71288,325				,078
Intra-grupos					
Total	622099,200				

Informe

Energia

Meses	Media	N	Desv. típ.
4	570,00	19	122,958
5	483,00	18	112,948
6	541,88	8	93,963
Total	530,20	45	118,906

(a) ¿Es razonable utilizar un modelo en el que las varianzas son iguales?
(b) Completa la tabla ANOVA.
(c) ¿Hay evidencia estadística a favor de la existencia de diferencias entre la ingesta de energía según la duración del período de lactancia?

30. **Sobre el contraste de igualdad de medias.** Considera los dos conjuntos de datos siguientes:

Réplica	$T1$	$T2$	$T3$	$T1$	$T2$	$T3$
1	20	22	24	45	8	15
2	19	22	24	0	30	44
3	20	22	23	10	38	2
4	21	22	25	25	12	35
Medias	20	22	24	20	22	24

Calcula las dos tablas de análisis de la varianza correspondientes. ¿Qué valores coinciden y qué valores difieren? ¿Cambian las conclusiones del contraste de igualdad de las tres medias (a nivel $\alpha = 0{,}05$).

31. **Niveles de beta-caroteno, ingesta de grasas e IMC (continuación).** Se compararon los logaritmos de los niveles de β-caroteno en individuos fumadores y no fumadores, obteniéndose los siguientes valores para cada grupo:

	F/NF	logBeta
N	Fumadores	156
	No Fumadores	158
Mean	Fumadores	5.07
	No Fumadores	4.85
Std. error mean	Fumadores	AA
	No Fumadores	BB
Standard deviation	Fumadores	0.714
	No Fumadores	0.766

¿Permiten los datos afirmar a nivel $\alpha = 0{,}05$ que la media del logaritmo de la cantidad de β-caroteno es diferente en individuos que no fuman que en individuos fumadores? Responde a la cuestión calculando la tabla ANOVA correspondiente a los datos (para el caso de un factor con solo dos niveles, ser fumador o no serlo). Compara el resultado con el que se obtendría mediante un contraste de igualdad de medias.

32. **Capacidad pulmonar.** Se quiere comparar la capacidad pulmonar en niños, adultos y ancianos, obteniéndose los siguientes resultados:

Niños	8,4	7,6	7,9	8,0	8,1
Adultos	8,7	8,1	8,5	8,2	8,0
Ancianos	7,4	7,8	7,3	7,6	8,0

Calcula la tabla ANOVA correspondiente a estos datos. Contrasta si la edad influye en la capacidad pulmonar media.

33. **Antiinflamatorios y riesgo de infarto.** Investigaciones recientes apuntan a que el uso continuado de antiinflamatorios no esteroides puede aumentar el riesgo de infarto de miocardio. La agregación plaquetaria es uno de los indicadores de riesgo de infarto, por lo que se lleva a cabo un estudio con un total de 30 pacientes a los que se mide la agregación plaquetaria (en unidades adecuadas) después de un tratamiento prolongado. De los 30 pacientes, 10 han sido tratados con Diclofenac, otros 10 con Ibuprofeno y, finalmente, el último grupo de 10 pacientes ha recibido placebo. Se obtienen los siguientes resultados:

Agregación plaquetaria

					Intervalo de confianza para la media al 95%	
	N	Media	Desviación típica	Error típico	Límite inferior	Límite superior
Diclofenac	10	907,50	15,138	4,787	896,67	918,33
Ibuprofeno	10	897,50	13,591	4,298	887,78	907,22
Placebo	10	879,00	24,922	7,881	861,17	896,83
Total	30	894,67	21,573	3,939	886,61	902,72

ANOVA

Agregación plaquetaria

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	4181,667	2	2090,833	6,060	,007
Intra-grupos	9315,000	27	345,000		
Total	13496,667	29			

(a) ¿Se detecta en estos datos una influencia significativa del producto empleado sobre las agregaciones plaquetarias medias? Da una respuesta al nivel de significación 0.05.¿Bajo qué supestos es válida la conclusión obtenida?
(b) Calcula la media global $\bar{Y}_{..}$.
(c) Calcula la cuasivarianza global

$$S^2 = \frac{\sum_{i=1}^{3}\sum_{j=1}^{10}(Y_{ij} - \bar{Y}_{..})^2}{29}.$$

¿Es S^2 un estimador adecuado de la varianza poblacional σ^2?