EXAMEN DE ESTADÍSTICA

1. La superficie (en miles de km^2) de los 7 países más grandes del mundo es la que se indica en la tabla adjunta.

a) (1 punto) Da las extensiones en miles de km^2 en números índice, tomando como base la extensión de Australia. (Redondea a las décimas).

b) (1,5 puntos) Sabiendo que la superficie habitable (tierra firme) del mundo es 133.342.000 km^2, representa en un diagrama de sectores las superficies de los 7 países anteriores junto con la del resto del mundo.

País	**Superficie (miles de km^2)**
Rusia	17.075
Canadá	9.976
China	9.561
Estados Unidos	9.373
Brasil	8.512
Australia	7.687
India	3.288

2. (1 punto) Los gastos del año pasado en comida, vivienda, educación y cultura, y varios, para una familia fueron, en porcentajes respectivos, de 30, 35, 20 y 15. Si este año, los precios de los apartados considerados aumentaron en el 4%, 12%, 8% y 6%, respectivamente, ¿en qué porcentaje aumentó el gasto medio de esa familia?

3. Los rendimientos medios (en kilogramos por hectárea) en España, para los cereales que se indican, fueron:

Año	1999	2000	2001	2002	2003
Trigo	2150	3100	2300	2830	2840
Maíz	9450	9220	9720	9510	9110

a) (0,75 puntos) Halla los rendimientos medios para el quinquenio de cada cereal.

b) (0,75 puntos) Para cada cereal halla el coeficiente de variación. ¿Qué cereal es más fiable?

4. a) (0,75 puntos) Explica brevemente el concepto de correlación. Pon dos ejemplos: uno con correlación directa y otro con correlación inversa.

b) (0,75 puntos) Explica el significado del coeficiente de correlación lineal. ¿Entre qué valores se mueve?

5. En una población, la media de los pesos de sus habitantes es de 65 kg y la de la estatura es 170 cm, siendo sus desviaciones típicas de 5 kg y 10 cm, respectivamente. Se sabe además que el coeficiente de correlación lineal entre ambas variables es de 0,8. (Se designa por X la variables peso; y por Y, la variable estatura).

a) (1,25 puntos) Halla la recta de regresión de Y sobre X.

b) (0,25 puntos) ¿Cuánto se estima que medirá una persona que pesa 75 kg?

6. Una compañía de seguros sospecha que el número de accidentes está en función de la edad del conductor. Para ello elige 100 personas de cada grupo de edad y contabiliza los accidentes totales del último año. Los datos fueron:

Edad (X)	20	25	30	35	40	45
Nº de accidentes (Y)	10	11	9	7	4	5

a) (1,25 puntos) ¿Existe correlación entre las variables estudiadas? Determina el coeficiente de correlación lineal entre ellas.

b) (1,25 puntos) Halla la recta de regresión de Y sobre X. ¿Podrá estimarse el número de accidentes para los grupos de 23, de 55 y de 60 años? Justifica tu respuesta.

Alcalá de Henares, 7 de mayo de 2015. JoséMMM

Soluciones

1. La superficie (en miles de km^2) de los 7 países más grandes del mundo es la que se indica en la tabla adjunta.

c) (1 punto) Da las extensiones en miles de km^2 en números índice, tomando como base la extensión de Australia. (Redondea a las décimas).

d) (1,5 puntos) Sabiendo que la superficie habitable (tierra firme) del mundo es 133.342.000 km^2, representa en un diagrama de sectores las superficies de los 7 países anteriores junto con la del resto del mundo.

País	Superficie (miles de km^2)
Rusia	17.075
Canadá	9.976
China	9.561
Estados Unidos	9.373
Brasil	8.512
Australia	7.687
India	3.288

Solución:

a) Si se asigna a Australia el valor 100, multiplicando la superficie de cada país por $\frac{100}{7.687}$ se obtiene:

Rusia: 222,1 Canadá: 129,8 China: 124,4
Estados Unidos: 121,9 Brasil: 110,7 India: 42,8

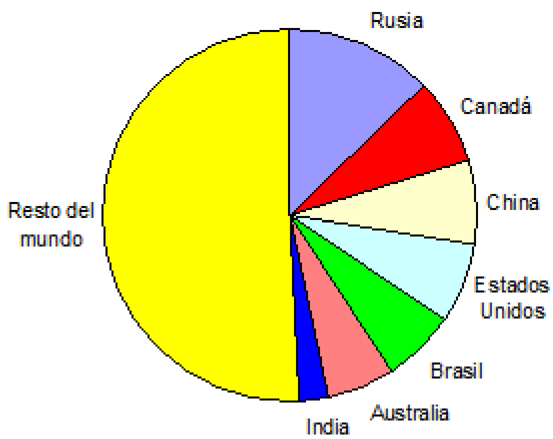
b) La superficie del resto del mundo es: 67.870.000 km^2.
La asignación de sectores, que se obtiene por proporcionalidad (regla de tres), para cada superficie es:

Rusia: 46,1º Canadá: 26,9º
China: 25,8º Estados Unidos: 25,3º
Brasil: 23º Australia: 20,8º
India: 8,9º Resto del mundo: 183,2º

El detalle para Rusia es:

si a 133.342.000 $km^2 \to 360^\circ$
a 17.075.000 $km^2 \to x \Rightarrow$

$$\Rightarrow x = \frac{17075 \cdot 360}{133342} = 46{,}1^\circ$$



2. (1 punto) Los gastos del año pasado en comida, vivienda, educación y cultura, y varios, para una familia fueron, en porcentajes respectivos, de 30, 35, 20 y 15. Si este año, los precios de los apartados considerados aumentaron en el 4%, 12%, 8% y 6%, respectivamente, ¿en qué porcentaje aumentó el gasto medio de esa familia?

Solución:

El aumento medio del gasto hay que ponderarlo, pues los pesos de cada apartado son diferentes. Así:

$$\text{Aumento medio} = \frac{4 \cdot 30 + 12 \cdot 35 + 8 \cdot 20 + 6 \cdot 15}{100} = 7{,}9\%$$

3. Los rendimientos medios (en kilogramos por hectárea) en España, para los cereales que se indican, fueron:

Año	1999	2000	2001	2002	2003
Trigo	2150	3100	2300	2830	2840
Maíz	9450	9220	9720	9510	9110

a) (0,75 puntos) Halla los rendimientos medios para el quinquenio de cada cereal.
b) (0,75 puntos) Para cada cereal halla el coeficiente de variación. ¿Qué cereal es más fiable?
Solución:
a) T representa al trigo y M al maíz.

$$\overline{x}(T)=\frac{2150+3100+2300+2830+2840}{5}=2644 \text{ kg/ha}; \quad s(T)=358,7$$

$$\overline{x}(M)=\frac{9450+9220+9720+9510+9110}{5}=9402 \text{ kg/ha}; \quad s(M)=216,1$$

b) $CV(T)=\dfrac{358,7}{2644}=0,1357\to 13,6\%$; $CV(\mathrm{M})=\dfrac{216,1}{9402}=0,02298\to 2,3\%$
Como el coeficiente de variación para el maíz es menor, es más fiable el maíz.

4. a) (0,75 puntos) Explica brevemente el concepto de correlación. Pon dos ejemplos: uno con correlación directa y otro con correlación inversa.
b) (0,75 puntos) Explica el significado del coeficiente de correlación lineal. ¿Entre qué valores se mueve?
Solución:
a) Cuando al estudiar distribuciones bidimensionales se observa que los cambios en una de las variables acarrean cambios en la otra se dice que las variables están correlacionadas. Estas variaciones pueden observarse al dibujar el diagrama de dispersión, la nube de puntos.
Si la nube de puntos adopta una forma alargada, la correlación puede calificarse de lineal.

Cuando las variables crecen conjuntamente, la correlación es *directa*. Si, por el contrario, al aumentar una de ellas disminuye la otra, la correlación será *inversa*.
La correlación puede calificarse como *fuerte* cuando el grado de dependencia es alto; y como *débil* en caso contrario.
Ejemplos:
→ La correlación entre el número de zapato y la estatura de las personas es directa y fuerte. A mayor número de zapato suele corresponder una mayor estatura de la persona.
→ Las variables temperatura y número de enfermos de gripe están inversamente correlacionadas: a menor temperatura más enfermos de gripe. Quizá se trate, también, de una correlación fuerte.

b) El **coeficiente de correlación lineal** es el criterio que se utiliza para medir la fuerza de la correlación entre dos variables.

Este coeficiente, denotado por *r*, se define así: $r=\dfrac{s_{xy}}{s_x \cdot s_y}$

Esto es, la razón entre la covarianza de las variables X e Y y el producto de sus desviaciones típicas marginales.
El valor de r está entre –1 y +1: $-1\le r\le 1$
→ Si r toma valores cercanos a –1, la correlación es fuerte (e inversa).
→ Si r toma valores cercanos a +1, la correlación es fuerte (y directa).

5. En una población, la media de los pesos de sus habitantes es de 65 kg y la de la estatura es 170 cm, siendo sus desviaciones típicas de 5 kg y 10 cm, respectivamente. Se sabe además que el coeficiente de correlación lineal entre ambas variables es de 0,8. (Se designa por *X* la variables peso; y por *Y*, la variable estatura).
a) (1,25 puntos) Halla la recta de regresión de *Y* sobre *X*.
b) (0,25 puntos) ¿Cuánto se estima que medirá una persona que pesa 75 kg?
Solución:

a) El coeficiente de correlación es $r = \dfrac{s_{xy}}{s_x s_y}$, siendo s_{xy} la covarianza y s_x y s_y las desviaciones típicas de la variable *X* (peso) e *Y* (estatura).

Como $r = 0,8$, $s_x = 5$ y $s_y = 10 \Rightarrow 0,8 = \dfrac{s_{xy}}{5 \cdot 10} \Rightarrow s_{xy} = 40$.
La ecuación de la recta de regresión de *Y* sobre *X* es:

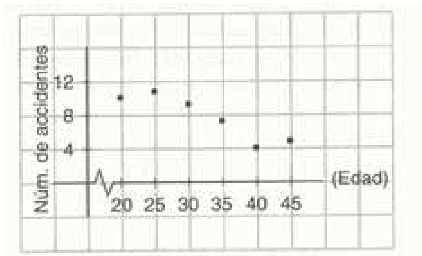
$$y - \overline{y} = \frac{s_{xy}}{s_x^2}\left(x - \overline{x}\right) \Rightarrow y - 170 = 1,6(x - 65) \Rightarrow y = 1,6x + 66$$

b) Para $x = 75$ se obtiene $y = 1,6 \cdot 75 + 66 = 186$ cm.

6. Una compañía de seguros sospecha que el número de accidentes está en función de la edad del conductor. Para ello elige 100 personas de cada grupo de edad y contabiliza los accidentes totales del último año. Los datos fueron:

Edad (*X*)	20	25	30	35	40	45
Nº de accidentes (*Y*)	10	11	9	7	4	5

a) ¿Existe correlación entre las variables estudiadas? Determina el coeficiente ce correlación lineal entre ellas.
b) Halla la recta de regresión de *Y* sobre *X*. ¿Podrá estimarse el número de accidentes para los grupos de 23, de 55 y de 60 años? Justifica tu respuesta.
Solución:
a) Se traza el diagrama de dispersión (figura adjunta). En él se observa que los valores dados están bastante alineados.
Con la calculadora o con ayuda de una tabla se halla el coeficiente de correlación.
Se obtiene:



$\overline{x} = 32,5$; $\overline{y} = 7,667$; $s_x = 8,539$; $s_y = 2,56$; $s_{xy} = -20$; $r = -0,915$.

Este coeficiente indica una correlación muy fuerte y negativa, que sugiere una fuerte dependencia entre la edad y el número de accidentes.

b) La recta de regresión es $y = -0,274x + 16,571$.
Esta ecuación puede utilizarse para estimar valores de *Y* siempre que la variable edad no esté lejos de los valores considerados. Así, puede usarse para $x = 23$, obteniéndose $y = 10,270$, pero no puede usarse para $x = 55$ o $x = 60$, datos para los que 1,511 y 0,141, respectivamente. (Incluso podría sospecharse que el número de accidente empieza a crecer cuando $x > 45$.)