

ESTADÍSTICA
2º curso de Biología
Curso 2013-2014

ESTADÍSTICA DESCRIPTIVA

1. Estamos interesados en la variable X = “Tiempo de vida (en días)” de una especie de insectos.

a) En una muestra pequeña de 11 insectos, los resultados muestrales fueron:

20, 25, 13, 18, 32, 25, 20, 15, 28, 40, 27

Hallar el tiempo medio de vida, el tiempo mediano de vida, y la desviación típica.

b) En una muestra grande, los resultados obtenidos se resumen de la siguiente forma:

Percentil	30	50	70	100
Tiempo de vida	18	22	26	30

Hallar el tiempo medio de vida (indicando previamente las clases, marcas de clase y frecuencias proporcionadas por la información muestral).

2. Con el fin de controlar la contaminación de un río, todas las semanas se hace una medición del nivel de ácido úrico.

a) Las mediciones durante 9 semanas fueron:

13 10 7 5 12 7 9 5 5

Hallar el nivel medio de ácido úrico, el nivel mediano, y la desviación típica.

b) En un estudio más completo, las mediciones semanales de ácido úrico se resumieron de la siguiente forma:

Percentil	20	40	70	100
Nivel de ácido úrico	6	8	12	18

Hallar el nivel medio de ácido úrico y dibujar el histograma (indicando previamente las clases, marcas de clase y frecuencias proporcionadas por la información muestral).

3. Los babuinos constituyen un género de primates catarrinos de la familia *Cercopithecidae*. Se lleva a cabo un amplio estudio con 100 babuinos macho y los resultados sobre su altura hasta los hombros se resumen en la siguiente tabla:

Altura en cm	Entre 40 y 50	Entre 50 y 55	Entre 55 y 60	Entre 60 y 70
Número de individuos	19	28	30	23

(a) Calcular la media muestral de estos datos.

(b) Representar los datos en un histograma.

(b) Calcular (aproximadamente) el primer cuartil y la mediana de los datos.

4. Una de las variedades de calabaza más utilizada para el consumo es la *cucurbita maxima* que puede alcanzar un peso de hasta 30 Kg. En una plantación se recolectan una serie de ejemplares para llevar a cabo un estudio estadístico, con los siguientes resultados:

40 calabazas tienen un peso entre 20 y 23 Kg.

80 calabazas tienen un peso entre 23 y 25 Kg.

35 calabazas tienen un peso entre 25 y 27 Kg.

45 calabazas tienen un peso entre 27 y 30 Kg.

(a) Halla (aproximadamente) el peso medio de las calabazas recogidas y representa los datos en un histograma.

(b) Halla (aproximadamente) la mediana del peso de los ejemplares recogidos.

5. Se piensa que la frecuencia del canto de los grillos (en una hora) tiene una estrecha relación con la temperatura. Para poder expresar la frecuencia del canto (Y) en función de la temperatura en grados Fahrenheit (X) se recogen cinco pares de datos que se resumen a continuación:

$$\sum x_i = 394 \quad \sum y_i = 84 \quad \sum x_i^2 = 31316 \quad \sum y_i^2 = 1426 \quad \sum x_i y_i = 6672$$

- (a) Obtener la recta de regresión para expresar la frecuencia del canto en función de la temperatura.
 (b) Evaluar el ajuste. La asociación entre X e Y , ¿es positiva o negativa?

6. El peso de los animales salvajes es siempre más difícil de determinar que su altura. Para poder expresar (aproximadamente) el peso de los ñues azules (*connochaetes taurinus*) en función de su altura, se apresan 10 ejemplares adultos, y se anota su altura en cm (X) y su peso en Kg (Y). Los resultados se resumen a continuación:

$$\sum x_i = 1255 \quad \sum y_i = 2275 \quad \sum x_i^2 = 158325 \quad \sum y_i^2 = 519875 \quad \sum x_i y_i = 286850$$

- (a) Con los datos anteriores, halla la recta de regresión del peso (Y) sobre la altura (X).
 (b) ¿Cuál sería el peso aproximado de un ñu cuya altura es de 125 cm?

7. Se quiere calibrar una nueva técnica experimental indirecta para medir presiones en relación con un método estándar directo. Para esto, se han realizado 9 tomas de presión (en mm de Hg) por el método estándar directo (X) y por la nueva técnica experimental indirecta (Y). Los resultados obtenidos se resumen a continuación:

$$\sum x_i = 343 \quad \sum y_i = 325 \quad \sum x_i^2 = 17693 \quad \sum y_i^2 = 16367 \quad \sum x_i y_i = 16992$$

- a) Calcular la recta de regresión de Y sobre X . Para una presión de 55 mm de Hg, medida con el método estándar, ¿qué presión cabría esperar con la nueva técnica?
 b) ¿Qué podemos decir del ajuste de la recta de regresión a nuestros datos?

8. (Ordenador) En 1778, H. Cavendish realizó una serie de 29 experimentos con objeto de medir la densidad de la tierra. Sus resultados, tomando como unidad la densidad del agua, fueron:

5'50 5'61 4'88 5'07 5'26 5'55 5'36 5'29 5'58 5'65
 5'57 5'53 5'62 5'29 5'44 5'34 5'79 5'10 5'27 5'39
 5'42 5'47 5'63 5'34 5'46 5'30 5'75 5'68 5'85

- a) Representa los datos por medio de un diagrama de tallos y hojas.
 b) Representa los datos por medio de un diagrama de caja y bigotes.
 c) Halla la media y la desviación típica.

9. (Ordenador) Para evaluar la viabilidad de un proyecto de reforestación de una zona sometida a una fuerte actividad turística, se analiza la composición en mg por cm^3 de desechos orgánicos del territorio. Los datos que se obtienen son:

10.87	9.01	22.50	12.35	17.39	31.05	17.19	16.74	20.33
19.32	23.18	25.15	15.49	20.30	2.38	13.55	9.33	22.72
10.96	25.90	27.66	9.74	18.65	9.31	24.60	17.41	24.86
15.34	23.34	22.81	17.86	30.72	32.60	8.96	32.71	15.86
16.71	5.48	8.25	20.57	4.57	2.30	32.56	7.92	4.84
4.57	26.45	23.58	19.27	9.79	3.03	19.40	23.92	22.45
22.05	21.18	18.85	8.38	15.01	18.12	4.24	3.39	7.17
22.71	22.44	15.89	24.20	24.75	28.08	19.73	13.22	17.69
5.53	11.42	5.58	3.15	14.06	5.83	19.42	21.13	18.32
23.31	11.89	23.95	19.30	12.22	21.45	9.84	4.78	38.63
12.65	13.89	23.82	16.91	28.09	15.73	12.53	16.52	9.48
4.08								

Efectuar un estudio descriptivo como en el ejercicio anterior.

10. (Ordenador) El maíz es un alimento importante para los animales. De todas formas, este alimento carece de algunos aminoácidos que son esenciales. Un grupo de científicos desarrolló una nueva variedad que sí contenía niveles apreciables de dichos aminoácidos. Para comprobar la utilidad de esta nueva variedad para la alimentación animal se llevó a cabo el siguiente experimento: a un grupo de 20 pollos de 1 día se les suministró un pienso que contenía harina de maíz de la nueva variedad. A otro grupo de 20 pollos (grupo de control) se le alimentó con un pienso que sólo se diferenciaba del anterior en que no contenía harina de la variedad mejorada de maíz. Los resultados que se obtuvieron sobre las ganancias de peso de los pollos (en gramos) al cabo de 21 días de alimentación fueron los siguientes:

- *Variedad normal*

380 321 366 356 283 349 402 462 356 410 329 399 350 384 316 272 345 455 360 431

- *Variedad mejorada*

361 447 401 375 434 403 393 426 406 318 467 407 427 420 477 392 430 339 410 326

- a) Para comparar las dos distribuciones, representa los dos diagramas de caja y bigotes en un mismo gráfico. ¿Qué se puede deducir de estos diagramas?
- b) ¿Cuáles son las medias y desviaciones típicas de los datos de ambos grupos? ¿Qué diferencias hay entre ambos?
11. (Ordenador) Las autoridades sanitarias de un municipio están interesadas en evaluar la calidad del agua para consumo en términos de colonias de bacterias tróficas en un acuífero próximo a la ciudad. Se consideran dos zonas diferentes del acuífero y se obtienen los siguientes resultados (número de colonias por 1000 mm de agua):

zona 1	194	199	191	202	215	214	197
	204	199	202	230	193	194	209
zona 2	158	161	143	174	220	156	156
	156	198	161	188	139	147	116

Realizar un estudio comparativo de la calidad del agua en ambas zonas, utilizando resúmenes numéricos y diagramas de cajas. ¿Se puede considerar que ambas zonas son similares?

12. (Ordenador) Se tienen dos métodos, *A* y *B*, para determinar el calor latente de fusión del hielo. La siguiente tabla da los resultados obtenidos (en calorías por gramo de masa para pasar de -0.72°C a 0°C) utilizando ambos métodos independientemente:

Método <i>A</i>	79.98	80.04	80.02	80.04	80.03	80.03	80.04	79.97	80.05	80.03	80.02	80.00	80.02
Método <i>B</i>	80.02	79.94	79.98	79.97	79.97	80.03	79.95	79.97					

- a) Hacer una comparación descriptiva de los dos métodos.
- b) Si por un error al transcribir los datos, se modifica el primer dato de cada método y se escribe 799.8 y 800.2, ¿cómo varían la media y la mediana?
13. (Ordenador) Los manatíes son unos animales grandes y dóciles que viven a lo largo de la costa de Florida. Cada año las lanchas motoras hieren o matan muchos de ellos. A continuación se presenta una tabla que contiene, para cada año, el número de licencias para motoras (expresado en miles de licencias) expedidas en Florida y el número de manatíes muertos en los años 1977 a 1990.
- a) Queremos analizar la relación entre el número de licencias expedidas anualmente en Florida y el número de manatíes muertos. ¿Cuál es la variable explicativa? Dibuja un diagrama de dispersión con esos datos. ¿Qué nos dice el diagrama sobre la relación entre esas dos variables? Las variables ¿están asociadas positiva o negativamente?
- b) Halla la recta de regresión para expresar el número de manatíes muertos en función del número de licencias.

c) Describe la fuerza de la relación. ¿Se puede predecir con precisión el número de manatíes muertos cada año conociendo el número de licencias expedidas ese año? Si Florida decidiera congelar el número de licencias en 700.000, ¿cuántos manatíes matarían, aproximadamente, las lanchas motoras?

Año	Licencias	Manatíes	Año	Licencias	Manatíes
1977	447	13	1984	559	34
1978	460	21	1985	585	33
1979	481	24	1986	614	33
1980	498	16	1987	645	39
1981	513	24	1988	675	43
1982	512	20	1989	711	50
1983	526	15	1990	719	47

14. (Ordenador) Los corredores buenos dan más pasos por segundo a medida que aumentan la velocidad. He aquí el promedio de pasos por segundo de un grupo de corredoras de élite a distintas velocidades. La velocidad se expresa en metros por segundo.

Velocidad (m/s)	4,83	5,14	5,33	5,67	6,08	6,42	6,74
Pasos por segundo	3,05	3,12	3,17	3,25	3,36	3,46	3,55

a) Se quiere predecir el número de pasos por segundo a partir de la velocidad. En primer lugar, dibuja un diagrama de dispersión.

b) Halla la recta de regresión del número de pasos por segundo con relación a la velocidad.

c) Halla el coeficiente de correlación lineal.

15. (Ordenador) La tabla siguiente presenta tres conjuntos de datos preparados por el estadístico Frank Anscombe para ilustrar los peligros de hacer cálculos sin antes representar los datos. *Los tres conjuntos de datos tienen la misma correlación y la misma recta de regresión.*

Conjunto de datos A:

X	10	8	13	9	11	14	6	4	12	7	5
Y	8,04	6,95	7,58	8,81	8,33	9,96	7,24	4,26	10,84	4,82	5,68

Conjunto de datos B:

X	10	8	13	9	11	14	6	4	12	7	5
Y	9,14	8,14	8,74	8,77	9,26	8,10	6,13	3,10	9,13	7,26	4,74

Conjunto de datos C:

X	8	8	8	8	8	8	8	8	8	8	19
Y	6,58	5,76	7,71	8,84	8,47	7,04	5,25	5,56	7,91	6,89	12,50

a) Calcular la correlación y la recta de regresión para los tres conjuntos de datos y comprobar que son iguales.

b) Dibujar un diagrama de dispersión para cada uno de los conjuntos de datos con las rectas de regresión correspondientes.

c) ¿En cuál de los tres casos utilizaríamos la recta de regresión para predecir Y dado $X = 14$.

Conclusión: REPRESENTA SIEMPRE TUS DATOS.