# Trabajo Práctico 1. Estadística descriptiva

1. .

**Problema 3:** A continuación, se describen distintos sistemas estudiados. Defina la población, muestra y unidad muestral para cada uno de ellos. Identifique la variable aleatoria en estudio y clasifíquela.

1. En el estudio de la activación de caminos intracelulares en corteza cerebral de ratas Wistar se determinó la longitud del axón de las células neuronales luego de aplicar una corriente de 100 pA a una rata Wistar elegida al azar.
2. En el estudio de cultivos celulares in vitro de la línea N2A se determinó la velocidad de crecimiento en 20 placas de dichos cultivos.
3. En un estudio realizado sobre el crecimiento del Pez Palo (*Percophis brasiliensis*) del Atlántico Sur se observaron 135 ejemplares de dicha especie elegidos al azar en el Golfo San Matías y se determinó el sexo, largo total, diámetro del otolito y concentración de plomo.
4. Se realizaron estudios a cerdos tomados de diferentes establecimientos de cría porcina del partido bonaerense de Navarro elegidos al azar, para determinar la presencia de triquinosis.
5. En un control realizado en campos de la cuenca del río Salado se ha observado un crecimiento en el número de casos de resistencia parasitaria en bovinos y ovinos. Una muestra aleatoria de 30 rodeos ovinos mostró que el 60% de los mismos posee resistencia a alguna droga antiparasitaria.

**Problema 4:** Grafique la siguiente información de la manera que considere más adecuada.

1. Los rendimientos en tubérculos por unidad de superficie indican la superioridad del *Tropaeolum tuberosum* sobre la *Solanum tuberosum* (33,25 y 30,35 t/ha respectivamente) respecto a la *Oxalis tuberosa* que presentó el rendimiento más bajo (21,97 t/ha). Datos en archivo: TP1.4tuber.IDB2
2. En un estudio relacionado con la *diabetes mellitus* se analizó la factibilidad de experimentación con una mutación de la cepa Wistar de rata (BB), observando que el 40% de los ejemplares desarrolla un síndrome agudo de hiperglucemia con hipoinsulinemia y cetoacidosis, y otro 30% presenta intolerancia a la glucosa sin progresión a cetosis. El restante de los individuos raramente desarrolla síntomas de la diabetes. Datos en archivo: TP1.4diabetes.IDB2
3. En un estudio de citotoxicidad sobre la línea células A549, se les internalizaron nanopartículas magnéticas (NPM) en cuatro concentraciones distintas: 1) control (sin internalizar), 2) 10 µg de NPM internalizadas por cultivo,3) 20 µg de NPM internalizadas por cultivo y 4) 50 µg de NPP internalizadas por cultivo. Luego se registró la cantidad de células viables (cv), células necróticas (cn) y células apoptópicas (ca) en cada cultivo. Datos en archivo: TP1.4A549.IDB2

*Ayuda: Aquí las categorías serían el estado de la célula luego del experimento, y habría cuatro muestras distintas.*

**Problema 5:El** caraguatá (*Eringium horridum*) es una planta que suele crecer en los establecimientos ganaderos de la provincia de Entre Ríos y cuyo crecimiento se desea controlar porque compite con los pastos que sirven de alimento para los animales. La siguiente tabla de datos representa la cantidad de rizomas presentes en 23 plantas de caraguatá tomadas al azar de un establecimiento ganadero en la Provincia de Entre Ríos, luego de tratarlas con un herbicida para evaluar su efecto sobre el crecimiento (TP1.5.IDB2).

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 3 | 1 | 2 | 3 | 2 | 8 | 3 | 5 | 1 | 4 | 5 |
| 4 | 5 | 9 | 5 | 4 | 5 | 6 | 6 | 7 | 8 | 6 |  |

1. ¿Cómo diseñaría este experimento?
2. ¿Cuál es la variable de estudio y tipo? Indique la unidad experimental, el tamaño de la muestra y la población.
3. Elabore una tabla de frecuencias absolutas y frecuencias absolutas acumuladas.
4. Represente gráficamente las frecuencias absolutas simples y acumuladas. ¿Qué forma tiene la distribución?
5. Calcule la media. ¿Cuál es el valor de la mediana muestral? ¿Qué representan estos valores?
6. ¿Cuál es la cantidad de rizomas más frecuente?
7. Compare los valores de la media, mediana y moda; ¿Qué sugieren con respecto a la forma de la distribución de esta variable?
8. Calcule la varianza, el desvío estándar y el coeficiente de variación. ¿Son muy variables las plantas respecto a la cantidad de rizomas?
9. ¿Qué porcentaje de las plantas poseen como máximo 4 rizomas? Y ¿menos de 4?
10. Obtenga el cuartil 1, el percentil 25 y el percentil 75; interprételos.

**Problema 6:** En la tabla siguiente se presenta la cantidad de mercurio (µg/g de músculo fresco) presente en una muestra aleatoria de percas extraídas del lago Nahuel Huapi, producto de la contaminación relacionada con la actividad volcánica. Niveles inferiores a 0,5 µg/g están considerados como valores aptos para el consumo humano.

1. Usando las definiciones de frecuencia acumulada y relativa, complete los datos faltantes de la tabla. *FA* y *FR* son la frecuencia absoluta y relativa respectivamente, y *FAA* la frecuencia acumulada.
2. Realice el histograma correspondiente e indique qué tipo de simetría tiene.
3. Estime el porcentaje de peces con niveles de mercurio no aptos para consumo humano.

|  |  |  |  |
| --- | --- | --- | --- |
| Intervalo | *FA* | *FR* | *FAA* |
| 0 – 0.1 | 60 | f1 | 60 |
| 0.1 – 0.2 | n2 | 0.4 | N2 |
| 0.2 – 0.3 | 30 | f3 | 170 |
| 0.3 – 1 | n4 | 0.1 | N4 |
| 1 – 2 | n5 | f5 | 200 |

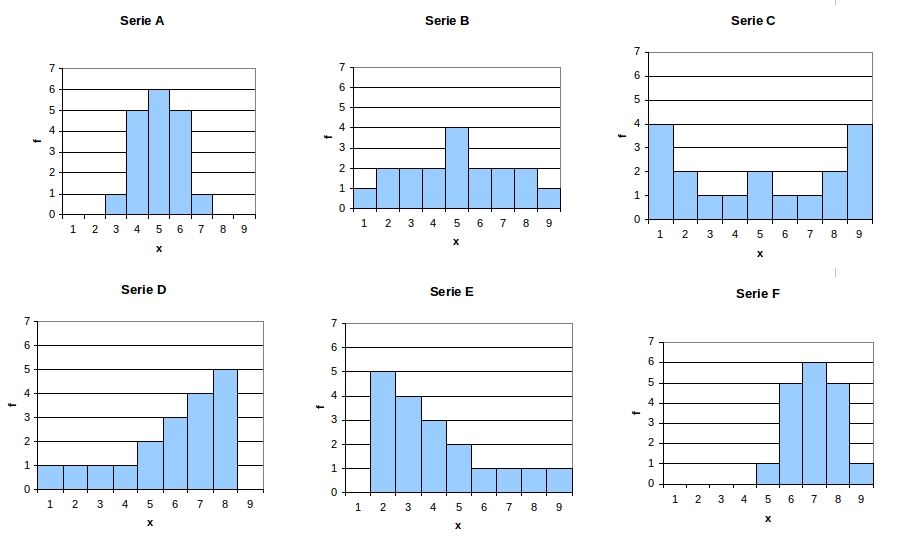
**Problema 7:** La siguiente información corresponde a pulsaciones cardíacas de 30 personas adultas (entre 30 y 50 años) que participaron de un programa de entrenamiento (en latidos por minuto:TP1.7.IDB2).

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 52 | 76 | 66 | 56 | 76 | 66 | 60 | 64 | 80 | 60 |
| 64 | 84 | 92 | 64 | 66 | 72 | 64 | 66 | 69 | 72 |
| 66 | 60 | 78 | 72 | 60 | 64 | 88 | 60 | 64 | 78 |

1. Grafique y determine la forma de la distribución.
2. Realice una tabla de frecuencias acumuladas y grafique un polígono de frecuencias acumuladas.
3. Obtenga la media aritmética, la varianza y el percentil 25. Interprete.
4. Represente los datos mediante un diagrama de caja e interprete.
5. Antes del entrenamiento los participantes fueron sometidos a la misma prueba. Si sus pulsaciones fueron, en general, mayores y más variables, ¿qué diferencias esperaría encontrar entre el histograma de las pulsaciones previas al entrenamiento y el efectuado en el punto 1?

**Problema 8:** Las distribuciones de frecuencias presentadas en la siguiente figura están realizadas sobre el mismo número de observaciones.

1. Sin efectuar cálculos, ordénelas en orden creciente según su media, mediana, moda y desvío estándar.
2. Arme las tablas de frecuencias para cada gráfico. Calcule la media, mediana, moda y desvío estándar. Recuerde que para datos agrupados, la distribución de datos dentro de cada clase se supone uniforme.



**Problema 9:** En un vivero experimental se siembran semillas de una especie de araucaria en bandejas con sólo seis lugares (una semilla por lugar). Al cabo de un tiempo se toma una muestra aleatoria de 50 bandejas y se cuenta la cantidad de semillas germinadas por bandeja con los siguientes resultados (TP1.9.IDB2):

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 2 | 6 | 4 | 0 | 3 | 4 | 3 | 3 | 6 |
| 2 | 3 | 4 | 3 | 5 | 0 | 1 | 3 | 3 | 3 |
| 5 | 2 | 6 | 1 | 6 | 1 | 3 | 4 | 5 | 3 |
| 5 | 4 | 3 | 4 | 2 | 2 | 0 | 2 | 3 | 4 |
| 5 | 2 | 1 | 2 | 3 | 3 | 3 | 4 | 5 | 4 |

1. Indique la unidad experimental, la variable, la población, la muestra, y la población estadística. ¿De qué tipo de variable se trata?
2. Agrupe los datos en una tabla de frecuencias. ¿En qué porcentaje de bandejas germinaron a lo sumo 2 semillas? ¿Y en qué porcentaje germinó alguna semilla?
3. Grafique los datos mediante un diagrama de frecuencias relativas y un polígono de frecuencias acumuladas.
4. Obtener los estadísticos de tendencia central y de variabilidad.
5. Obtener el rango de las medidas, realizar un diagrama de caja e interpretar.
6. Calcular el percentil 30. ¿Qué información provee este percentil?

**Problema 10:** En el siguiente conjunto de datos se proporciona el tiempo de desarrollo hasta su madurez de una variedad de papa (cultivada en la puna de los andes centrales y meridionales), indicado en días después de la siembra (DDS). La medición del nivel de desarrollo viene dada por la cantidad de fitomasa producido por una planta, alcanzando su madurez cuando no existe aumento de fitomasa (TP1.10.IDB2).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 179 | 173 | 169 | 172 | 171 | 169 | 168 | 181 | 174 | 173 | 177 | 179 | 178 | 176 | 178 | 175 |
| 165 | 175 | 167 | 175 | 174 | 176 | 174 | 174 | 172 | 177 | 175 | 176 | 176 | 167 | 176 | 177 |
| 183 | 164 | 181 | 175 | 173 | 178 | 173 | 178 | 175 | 176 | 173 | 171 | 174 | 175 | 174 | 175 |
| 173 | 178 | 175 | 176 | 174 | 178 | 179 | 180 |

1. Construir una tabla de frecuencias relativas y absolutas de estos tiempos.
2. Hallar las frecuencias relativas acumuladas.
3. Calcular la mediana y la moda.
4. Realizar un diagrama para las frecuencias relativas acumuladas. ¿Qué fracción madura antes de los 175 DDS?
5. Encuentre cuántos días tarda en alcanzar la madurez el 25%, 50% y 75% de la población.
6. Utilice el rango y los valores hallados en el item anterior para realizar un diagrama de caja.

**Problema 11:**

La frutilla es una de las frutas más perecederas, susceptible a daño mecánico, deterioro fisiológico, ataque por hongos, etc. Por otro lado, gran parte de la fruta se congela y es necesario evaluar el efecto de este proceso sobre la calidad de la fruta. Se desea comparar la resistencia al congelamiento de la variedad Camarosa con la de la variedad Sweet Charly. Para ello, en muestras de 100 g de frutillas de cada variedad congeladas a –20°C durante de 3 meses, se midió el volumen de líquido perdido luego del descongelamiento. Este es un parámetro importante en el procesamiento industrial no solo porque permite evaluar la calidad del proceso de congelado-descongelado sino que además permite estudiar la pérdida de resistencia mecánica en el cambio de estado. Los resultados fueron (en ml/100 g de fruta) (TP1.11IDB2):

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 10,96 | 8,15 | 10,69 | 12,46 | 9,71 | 10,99 |
| 9,78 | 11,55 | 9,19 | 10,27 | 13 | 9,84 |
| 10,51 | 9,84 | 10,03 | 8,85 | 11,91 | 10,69 |
| 7,57 | 8,49 | 10,22 | 12,69 | 11,04 | 8,85 |
| 11,68 | 11,62 | 12,24 | 13,04 | 6,75 | 10,24 |
| 11,58 | 11,52 | 9,63 | 11,75 | 9,99 | 7,95 |
| 10,03 | 9,36 | 10,53 | 10,95 |  |  |

Camarosa Sweet Charly

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 20,04 | 21,56 | 22,13 | 20,76 | 19,09 | 21,65 |
| 19,8 | 19,61 | 24,38 | 22,62 | 22,34 | 21,24 |
| 18,84 | 17,61 | 19,45 | 23,12 | 21,47 | 16,57 |
| 18,37 | 17,78 | 20,42 | 15,71 | 19,72 | 17,6 |
| 21,91 | 21,11 | 20,82 | 22,68 | 19,84 | 17,06 |
| 20,42 | 21,91 | 18,86 | 19,78 | 16,87 | 21,78 |
| 15,38 | 20,07 | 19,73 | 19,34 |  |  |

Para ambos conjuntos de datos:

1. Definir y clasificar la variable, identificar la unidad experimental, muestra y población.
2. Construir la distribución de frecuencias considerando 6 intervalos. Graficar mediante un histograma.
3. Construir el polígono de frecuencias acumuladas relativas. Ubicar gráficamente a la mediana.
4. Calcular la media, la mediana, la moda, el desvío estándar y el coeficiente de variación.
5. Calcular el percentil 75 e interpretar el resultado.
6. Utilizando los gráficos y estadísticos obtenidos comparar ambos conjuntos de datos, ¿para qué variedad se obtiene un mejor proceso de congelado? ¿Cuál presenta mayor variabilidad?

**Problema 12:** Los datos de la tabla corresponden al salario mensual de un grupo de 50 obreros de la industria textil, a mediados del 2009 (TP1.12.IDB2):

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2530 | 2230 | 2310 | 2320 | 2810 | 2340 | 2310 | 2420 | 2340 | 1960 |
| 2460 | 2780 | 2350 | 2520 | 2780 | 2380 | 2520 | 2460 | 2370 | 2510 |
| 2920 | 2380 | 2370 | 2420 | 2380 | 2970 | 2200 | 2260 | 2170 | 2360 |
| 6310 | 2440 | 2980 | 2370 | 2180 | 2250 | 2390 | 2360 | 2440 | 2400 |
| 2450 | 2380 | 2410 | 2480 | 2330 | 2450 | 2740 | 7100 | 2370 | 1980 |

1. Calcular las medidas de tendencia central. ¿Cuál de las tres medidas es la menos representativa de los datos? ¿Por qué?
2. Calcular el desvío estándar y la varianza.
3. Calcule el primer y tercer cuartil. Interprete el intervalo [Q1; Q3].
4. Calcule el salario promedio mensual de los trabajadores suponiendo un aumento de $300 en el sueldo de cada uno de ellos.
5. Si un grupo de 50 obreros chilenos de la industria textil tiene un sueldo mensual promedio de $311000 y un desvío típico de $41000, ¿cuál de los dos grupos es más homogéneo? ¿Por qué?
6. Realice una tabla de frecuencias, agrupando los datos en intervalos de manera adecuada.
7. Para los datos agrupados, calcular la media, mediana y moda. ¿Por qué se obtienen distintos valores que los calculados en 1?
8. Grafique un histograma para los datos agrupados y comente la forma de la distribución.

**Problema 13:** Se determinó mediante un nuevo método la concentración de fosfolípidos en 36 alícuotas de aceite crudo de girasol (de 10 ml cada una), extraídas de un lote. Los resultados (en g/100g) fueron (TP1.13.IDB2):

|  |  |  |  |
| --- | --- | --- | --- |
| 0,2 | 0,35 | 0,29 | 0,28 |
| 0,39 | 0,23 | 0,18 | 0,15 |
| 0,45 | 0,42 | 0,38 | 0,21 |
| 0,26 | 0,27 | 0,17 | 0,31 |
| 0,1 | 0,25 | 0,26 | 0,3 |
| 0,44 | 0,35 | 0,28 | 0,29 |
| 0,24 | 0,28 | 0,3 | 0,27 |
| 0,38 | 0,37 | 0,29 | 0,32 |
| 0,16 | 0,24 | 0,22 | 0,28 |

1. Identifique y clasifique la variable, la unidad de observación, la muestra y la población en estudio.
2. Grafique los datos. ¿Identifica algún dato atípico? ¿Qué tipo de simetría tiene la distribución?
3. Calcule el promedio y el desvío estándar.
4. Calcule la mediana de la concentración de fosfolípidos.
5. La presencia de fosfolípidos causa siempre inconvenientes y pérdidas en las distintas etapas de refinación del aceite, ya que favorece la formación de emulsiones. Los aceites con contenidos superiores a 0,40 g/100 g son considerados de calidad inferior. ¿Qué porcentaje representan estas muestras?
6. Calcule el percentil 10 e interprete el valor obtenido.
7. El método tradicional para la determinación de fosfolípidos aplicado sobre las mismas 36 muestras, arrojó un promedio de 0,25 g/100 g con una varianza de 0,01 (g/100 g)2. ¿Cuál de los dos métodos considera más preciso?

**Problema 14:** Creando histogramas sin datos.

Para las siguientes variables, determine qué forma de la distribución esperaría y explique brevemente por qué:

1. Tiempo en completar un examen difícil (el tiempo máximo es de 2 horas)
2. Las notas de un examen fácil
3. Edad de la población de la Ciudad de Buenos Aires en 2010 (según los resultados del Censo Nacional, el 50% tenía a lo sumo 36,8 años)
4. Edad de la población de Misiones en 2010 (según los resultados del Censo Nacional, el 50% tenía a lo sumo 23,9 años)
5. Contenido neto de azúcar en paquetes de 1 kg
6. Edad en el momento de la defunción para la Argentina en 2012
7. Ingresos mensuales de los jugadores de fútbol profesionales