

Trabajo Práctico 1: Estadística Descriptiva

Descripción de datos univariados

Ejercicio 1

Clasificar las siguientes variables:

- *Estado nutricional de una cría de camélido* evaluado como "Bueno", "Regular" o "Deficiente"
- *Superficie foliar de plantas de coirón* medida en cm^2
- *Velocidad de reacción* en una reacción química de descomposición medida en mol/min
- *Peso corporal de conejos* medido en kilogramos
- *Tipo de reacción química*
- *Nivel educativo alcanzado* evaluado como 1="Primario Incompleto", 2="Primario Completo", 3="Secundario Incompleto", 4="Secundario Completo", 5="Estudio Superior Incompleto" o 6="Estudio Superior Completo".
- *Nivel educativo alcanzado* evaluado como el número de años que la persona estuvo dentro del sistema educativo.
- *Número de impactos de partículas (He^{+2}) en una placa radiográfica* de 100 cm^2
- *Presencia o ausencia de efecto antrópico* en una parcela de campo.
- *Número de alumnos aprobados en un examen de Estadística.*
- *Masa de producto en cierto momento de una reacción química de síntesis.*
- *Sexo del charito (cría del ñandú) recién nacido.*
- Longitud de onda de un haz electromagnético.
- *Número de rebotes por minuto* de una pelota lanzada en condiciones homogéneas

Ejercicio 2

Considere los datos del ejercicio 1.21 del libro *Probabilidad y Estadística para Ingeniería y Ciencias* de Walpole, página 31.

- ¿Cuál es la variable en estudio?
- ¿A qué tipo corresponde la variable?
- Construir un conjunto de clases adecuado para estos pesos.
- Encontrar las frecuencias absolutas y relativas.
- Encontrar las frecuencias acumuladas (absolutas y relativas).
- Construir una tabla de distribución de frecuencias.
- Dibujar un histograma con los datos.
- ¿Por qué se ha utilizado un histograma para representar estos datos, en lugar de una gráfica de barras?
- Dibujar un polígono de frecuencias con los datos del apartado.
- Dibujar un polígono de frecuencias acumuladas con los datos del apartado.

Ejercicio 3

Se considera que una falla eléctrica es de “corta duración” si dura 30 minutos o menos y de “larga duración” si supera los 30 minutos.

- Según este punto de corte (30 minutos) redefinir la variable del ejercicio 2.
- Construir un conjunto de clases adecuado para esta nueva variable.
- Encontrar las frecuencias absolutas y relativas.
- Construir una tabla de distribución de frecuencias.
- Dibujar un diagrama de barras con los datos del apartado.
- Dibujar un diagrama de sectores con los datos del apartado.

Ejercicio 4

Para los datos del ejercicio 2 calcular:

- Las medidas de tendencia central: media, mediana y moda.
- Las medidas de dispersión: rango, varianza, desviación estándar y coeficiente de variación.
- Coeficiente de asimetría y coeficiente de curtosis.
- Medidas de posición: mediana, primer cuartil, tercer cuartil y los percentilos P_5 , P_{10} , P_{90} y P_{95}

Ejercicio 5

Para la variable del ejercicio 3 calcular:

- La proporción de fallas de corta duración.
- La proporción de fallas de larga duración.
- La disparidad de fallas de duración larga respecto de duración corta.
- Interprete sus resultados.

Ejercicio 6

Los siguientes datos corresponden al recuento de cierto tipo de células en pacientes afectados de tuberculosis. El recuento de células se realizó "antes" y "después" de la aplicación de cierto tratamiento específico.

Paciente	Antes	Después	Paciente	Antes	Después
1	46	127	13	124	124
2	405	123	14	434	138
3	228	87	15	411	21
4	84	129	16	15	15
5	87	235	17	244	244
6	98	323	18	96	358
7	155	96	19	48	44
8	141	413	20	336	662
9	442	379	21	394	394
10	255	370	22	87	266
11	198	420	23	479	423
12	98	142			

- ¿Cuál es la variable en estudio?
- ¿A qué tipo corresponde la variable?
- Construya un conjunto de clases adecuado para los datos “Antes” y “Después” por separado.
- Encuentre las frecuencias absolutas y relativas de las clases que construyó, para los datos “Antes” y “Después” por separado.
- Encuentre las frecuencias acumuladas (absolutas y relativas) de las clases que construyó, para los datos “Antes” y “Después” por separado.
- Construya tablas de distribución de frecuencias.
- Dibuje histogramas.
- Dibuje polígonos de frecuencias.
- Dibuje polígonos de frecuencias acumuladas (ojivas).

Ejercicio 7

Respecto de los datos del ejercicio anterior, se considera que el tratamiento tuvo éxito en un paciente si el recuento de células disminuyó luego de la aplicación.

- Según este criterio defina la variable en estudio y diga de qué tipo es.
- Construya un conjunto de clases adecuado.
- Encuentre las frecuencias absolutas y relativas de las clases que construyó.
- Construya una tabla de distribución de frecuencias.
- Dibuje un diagrama de barras.
- Dibuje un diagrama de sectores.

Ejercicio 8

Para los datos del ejercicio 2 calcular las siguientes medidas (para el recuento celular "antes" y "después" del tratamiento):

- Medidas de tendencia central: promedio, mediana y moda.
- Medidas de dispersión: rango, varianza, desviación estándar y coeficiente de variación.
- Medidas de posición: mediana, primer cuartil, tercer cuartil y los percentiles P_5 , P_{10} , P_{90} y P_{95} .
- Coeficiente de asimetría y coeficiente de curtosis.
- Interpretar los resultados obtenidos a fin de caracterizar la muestra de pacientes incluidos en el tratamiento.

Ejercicio 9

Para los datos del ejercicio 6, basándose en las medidas calculadas en el ejercicio 8, construya un diagrama de caja y bigotes para los datos “Antes” y uno para los datos “Después”. Compare ambos diagramas y comente sus observaciones.

Ejercicio 10

Si a cada dato se le resta el promedio, se obtienen los *datos centrados*. Si a cada dato centrado se lo divide por la desviación estándar se obtienen los *datos estandarizados* o *datos tipificados*. Considere los datos del ejercicio 2.

- Obtenga los datos centrados. Calcule promedio y desviación estándar.
- Obtenga los datos estandarizados. Calcule promedio y desviación estándar.
- Complete la siguiente tabla, observe y comente.

Tipo de dato	Promedio	Desviación Estándar
Datos Crudos		
Datos Centrados		
Datos Estandarizados		

Ejercicio 11

Para un conjunto de datos $\{x_1, x_2, x_3, \dots, x_n\}$ se define su promedio como

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Y la varianza como $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$.

Además, podemos expresar como $y_i = x_i - \bar{x}$ a los datos centrados y como $z_i = \frac{x_i - \bar{x}}{s}$

a los datos estandarizados. Basado en lo anterior demuestre los siguientes hechos:

- Los datos centrados tienen promedio cero
- Los datos estandarizados tienen promedio cero
- Los datos centrados tienen la misma varianza que los datos crudos
- Los datos estandarizados tienen varianza 1.

Ejercicio 12

Para un conjunto de datos $\{x_1, x_2, x_3, \dots, x_n\}$ se define la función $v(M) = \frac{1}{n-1} \sum_{i=1}^n (x_i - M)^2$

siendo m un número real cualquiera.

- Pruebe que la función $v(M)$ alcanza un mínimo local si $M = \bar{x}$.
- Calcule el valor del mínimo de $v(M)$ e interprete.