

UNIVERSIDAD NACIONAL DE LA PLATA
Facultad de Ciencias Agrarias y Forestales



CÁLCULO ESTADÍSTICO Y BIOMETRÍA

Curso 2019

GUÍA DE TRABAJOS PRÁCTICOS

DOCENTES

Profesora Titular: Ing. Agr. María Inés Urrutia

Profesor Adjunto: Ing. Agr. Martín E. Delucis

Jefe de Trabajos Prácticos: Lic. Pablo Rodrigo Altamirano
Lic. Humberto Chávez Zegarra

Ayudantes Diplomados: Dra. Noelia Ferrando
Dr. Adrian Jáuregui
Mgr. Laura Maly
Dra. Marina Pifano

TRABAJO PRÁCTICO N° 1

Estadística Descriptiva

Ejercicios de resolución manual

EJERCICIO 1.1: Los siguientes datos fueron obtenidos observando el número de lesiones en la piel de una muestra de 30 manzanas:

2, 3, 5, 3, 2, 6, 3, 2, 2, 1, 0, 1, 0, 4, 1, 2, 1, 4, 5, 0, 3, 4, 2, 1, 2, 2, 0, 1, 0, 2.

- a) Defina la variable en estudio, tipo y recorrido.
- b) Complete la tabla de distribución de frecuencias:

| X_i | f_i | h_i | F_i | H_i |
|-----------|-------|-------|-------|-------|
| 0 | | | | |
| 1 | | | | |
| 2 | | | | |
| 3 | | | | |
| 4 | | | | |
| 5 | | | | |
| 6 | | | | |
| $\Sigma=$ | | | | |

- c) Realice gráficos de barras con las frecuencias absolutas (f_i) y otro con las frecuencias relativas acumuladas (H_i).
- d) Indique el valor de: $f_{x=2}$, $h_{x=5}\%$, $F_{x=4}$ luego interprete cada resultado.
- e) Calcule el número medio de golpes por manzana.
- f) Calcule la varianza y el desvío estándar
- g) Describa los datos a partir de un diagrama de cajas (Box-plot)

EJERCICIO 1.2: Se midió la circunferencia en centímetros de 46 árboles que crecen en la Facultad, obteniéndose:

39,6 44,4 43,9 42,9 47,9 46,9 48,3 40,4 45,8 44,2 39,1 46,3 41,9 51,2 50,2 47,7 43,9
 47,1 42,6 46,6 44,8 43,7 47,2 41,3 37,4 43,1 43,9 48,9 41,5 42,5 48,2 40,7 41,8 46,7
 42,3 43,3 44,1 49,8 40,6 44,0 44,0 44,5 45,1 45,4 45,2 45,7

- a) Defina la variable en estudio, tipo y recorrido.
- b) Construya la tabla de distribución de frecuencias con los datos agrupados en 8 intervalos de amplitud 2 a partir del valor 37.
- c) Grafique las frecuencias relativas (h_i) y frecuencias relativas acumuladas (H_i)
- d) A partir del gráfico de las frecuencias relativas acumuladas obtenga de forma aproximada el valor de la mediana, primer y tercer cuartil.

- d) Calcule la media de los datos originales y la media de los datos agrupados en intervalos. ¿Coinciden ambos resultados? ¿Por qué?
- e) Calcule el desvío estándar y el coeficiente de variación de los datos agrupados.
- f) Describa los datos a partir de un diagrama de cajas (Box-plot).

EJERCICIO 1.3: Utilizando los datos del EJERCICIO 1.2, considere los primeros 20 datos como correspondientes a una especie A y los restantes a una especie B.

- a) Realice diagramas de cajas para visualizar posibles diferencias en la circunferencia originadas al dividir los datos en dos grupos.
- b) Hallar el porcentaje de árboles con diámetro entre 43 y 47 cm. para cada conjunto de datos.

Ejercicio de resolución utilizando software estadístico

El crecimiento de las plantas en las etapas iniciales de su vida resulta crucial para su supervivencia y crecimiento posterior. En muchas especies tanto herbáceas como leñosas, las plántulas que presentan mayor crecimiento presentan mayores tasas de supervivencia y mayor crecimiento en las etapas posteriores. Se realizaron ensayos de crecimiento de ejemplares jóvenes de una especie vegetal con semillas de diferentes muestras provenientes de 4 localidades distintas. Se obtuvieron los siguientes datos de crecimiento en mm / día, representados en la siguiente tabla:

Tabla: Crecimiento de plántulas en mm/día:

| <i>Localidad 1</i> | <i>Localidad 2</i> | <i>Localidad 3</i> | <i>Localidad 4</i> |
|--------------------|--------------------|--------------------|--------------------|
| 49.0 | 17.6 | 12.3 | 17.4 |
| 15.3 | 17.8 | 12.3 | 17.2 |
| 14.1 | 16.3 | 13.5 | 15.1 |
| 15.2 | 17.8 | 13.2 | 17.4 |
| 16.6 | 18.7 | 13.1 | 17.1 |
| 16.0 | 18.5 | 13.9 | 16.5 |
| 14.8 | 17.0 | 12.3 | 17.4 |
| 15.4 | 17.5 | 13.7 | 16.5 |
| 13.6 | 16.0 | 14.1 | 16.7 |
| 15.5 | 17.7 | 14.5 | 15.5 |
| 17.4 | 19.6 | 14.4 | 15.0 |
| 16.1 | 18.8 | 15.5 | 15.3 |
| 16.6 | 18.1 | 15.1 | 14.9 |
| 14.9 | 16.4 | 15.5 | 14.6 |
| 15.7 | 17.8 | 21.5 | 13.3 |
| 15.0 | 16.4 | 16.4 | 12.8 |

| | | | |
|------|------|------|------|
| 15.6 | 17.1 | 17.6 | 11.7 |
| 14.0 | 16.7 | 18.9 | 10.9 |
| 16.6 | 18.1 | 19.0 | 16.3 |
| 15.2 | 17.4 | 20.9 | 17.3 |

EJERCICIO 1.4: Utilizando los datos de la tabla:

- a) Calcular los estadísticos descriptivos de las muestras provenientes de las localidades 2 y 3.
 - b) ¿Cuáles son los estadísticos que más difieren entre las muestras 2 y 3? Interpretar que significan las diferencias.
 - c) Realizar los gráficos de caja y los histogramas para las localidades 2 y 3. ¿Cómo se reflejan las diferencias identificadas en el inciso anterior?
 - d) Calcular los estadísticos descriptivos del resto de las localidades y generar los gráficos de caja e histogramas. ¿Si hubiera que elegir una localidad como fuente de semillas para una siembra cual sería preferible y por qué?
 - f) ¿Cuáles son las dos muestras con crecimiento medio más alto?
 - g) ¿Que estadísticos de posición y dispersión permiten diferenciar las cuatro muestras? ¿Cómo se reflejan en los gráficos?
 - h) Se sabe que las plántulas con crecimiento menor de cierto valor tienen muy bajas probabilidades de supervivencia al ser plantadas. Indicar cuáles serían las dos localidades con mayor probabilidad de originar plantas establecidas si el crecimiento mínimo requerido es mayor a:
 - Caso 1: 13 mm/día
 - Caso 2: 17 mm/día
- Explicar. (Para su resolución se aconseja utilizar la Tabla de Frecuencias o el Histograma)
- i) ¿Qué diferencias existen entre las muestras 3 y 4? ¿Cómo se reflejan en los gráficos?
 - j) Se desea realizar un experimento para determinar diferencias de crecimiento en diferentes suelos. Se requiere que el crecimiento de las plántulas utilizadas sea homogéneo. ¿Qué plantas utilizaría?
 - k) Se tomó una nueva muestra de la localidad 2 compuesta de 30 semillas. Los crecimientos de las plántulas tuvieron un crecimiento medio = 15 y un desvío estándar = 0,946, calcular el error estándar. Comparar con el valor de la muestra anterior y explicar la diferencia.
 - l) Se tomaron nuevas muestras de 20 semillas para todas las localidades. Indicar las localidades para las que se espera la mayor y la menor diferencia respecto de la media de la muestra anterior. Explicar por qué.

Ejercicios Teóricos y/o Complementarios

EJERCICIO 1.5: Utilizando los datos del EJERCICIO 1.1:

a) Indique el valor e interprete el resultado de $F_{x=6}-F_{x=3}$, $H_{x=4}\%$, $100-H_{x=3}\%$.

b) Calcular la mediana y la moda. Rango y coeficiente de variación.

Interpretar complementando con los cálculos realizados en los incisos e) y f).

c) A partir de los datos originales realizar las siguientes operaciones:

c1) Multiplicar los datos por 2 y volver a calcular la media y el desvío estándar. Al comparar estos resultados con los obtenidos con los datos originales, que propiedades se verifican?

c2) Sumar a todos los datos el número 3, calcular la media y el desvío estándar, comparando con los resultados originales, que propiedades se verifican?.

c3) A los datos originales restarle la media y a esa diferencia dividirla por el desvío estándar, calcular ahora media y desvío estándar de la nueva variable transformada. Le sorprende los valores que obtuvo? como se llama la transformación de la variable que realizó ?

EJERCICIO 1.6: Sea la variable X con media 14.6 y varianza 6.5. Si con ella generamos la variable W tal que $W = 4.2 X + 5.8$, cual es la media, la varianza y el desvío estándar de W?.

EJERCICIO 1.7: Enumere y explique al menos 3 propiedades de la media aritmética.

EJERCICIO 1.8: Con los datos del **ejercicio 1.3** realice una descripción del crecimiento de las plántulas de las localidades 1 y 4. A modo de ejemplo se describe el crecimiento de las plántulas de la **localidad 3**:

“El crecimiento medio de las plántulas provenientes de la localidad 3 es de 15,39 mm/día., mientras la mediana es 14,45 mm/día y el crecimiento más frecuente de 12,3 mm/día. Que la media sea mayor que la mediana me indica que se trata de una distribución asimétrica positiva. Esto se aprecia visualmente en los histogramas y en los gráficos de caja. La asimetría también puede evaluarse analizando el estadístico que mide específicamente esa característica y que en esta oportunidad toma el valor de 0,98, que no hace más que confirmar la asimetría positiva. En cuanto al grado de achatamiento no vemos que se diferencie de una distribución mesocúrtica (curtosis $-0,03$).

Si bien los valores varían entre 12,3 y 21,5 (rango = 9,2), el 50 % de los valores se encuentran entre 13,28 y 16,29, es decir en un intervalo de 3,01 mm/día (rango intercuartílico). El desvío estándar indica que el alejamiento promedio de las observaciones respecto de la media es aproximadamente 2,81 mm/día. El error de la media es aproximadamente 0,63 indicando que la variabilidad esperada entre las medias de un conjunto de muestras no es muy grande. El coeficiente de variación está expresando una alta variabilidad representando el desvío un 18,25 % de la media.

El análisis del histograma muestra que la clase más frecuente comprende el intervalo de 13 a 15 mm/día. El diagrama de caja muestra el desplazamiento de la caja hacia la barra izquierda y la diferencia entre la media y la mediana, marcando ambas características la asimetría positiva de la distribución.

La tabla de frecuencia entre otras cosas nos permite ver rápidamente que el 75% de las observaciones son iguales o menores a 17 mm/día.

EJERCICIO 1.9: Clasificar las siguientes variables en continuas o discretas:

- Número de semillas de alfalfa por metro lineal.
- Diámetro a la altura del pecho (DAP) de 50 árboles de *Eucaliptus* sp.
- Temperaturas registradas cada media hora en un laboratorio, durante una semana
- Nº de pulgones por hoja.
- Milímetros de precipitación mensual registrados en La Plata durante un año.
- Número de semillas germinadas en cajas de 50 semillas.
- Cociente entre el DAP y la altura de árboles de Pino.

EJERCICIO 1.10: Los siguientes datos corresponden a la ganancia de peso por día (expresada en gramos), de 40 novillos sometidos a una dieta experimental.

| | | | | |
|-----|-----|-----|-----|-----|
| 704 | 890 | 986 | 806 | 798 |
| 995 | 706 | 915 | 720 | 960 |
| 876 | 705 | 801 | 807 | 858 |
| 606 | 893 | 705 | 896 | 830 |
| 798 | 906 | 758 | 910 | 690 |
| 708 | 660 | 809 | 800 | 708 |
| 780 | 918 | 825 | 700 | 969 |
| 615 | 804 | 697 | 880 | 895 |

- Calcular los estadísticos descriptivos de posición, dispersión y forma para describir la variable ganancia de peso por día.
- Realizar el gráfico de caja.

EJERCICIO 1.11: En un estudio en un monte del Chaco árido se midieron los perímetros basales de troncos de plantas de quebracho blanco (en centímetros) y se obtuvo la siguiente información.

138 164 150 132 144 125 149 157 146 158 140 147 136 148 152 144 168 126 138 176 163
119 154 165 146 173 142 147 135 153 140 135 161 145 135 161 145 142 150 156 145 128

- Construir la tabla de distribución de frecuencias.
- Realizar el histograma.
- Obtener los estadísticos de posición y dispersión.