

Estadística I

David Gabriel Corzo Mcmath

2020-Jan-07 11:57:29

Índice general

1. Clase introductoria	5
1.1. Clase introductoria	5
2. Clase - 2020-01-09	7
2.1. Notas	7
2.2. Audit.xlsx	7
3. Clase - 2020-01-14	9
3.1. Pasos de ordenamiento en Excel	9
3.2. Medidas de localización o tendencia central	9
3.2.1. Media:	9
3.2.2. Mediana:	9
3.2.3. Moda:	10
3.2.4. Percentiles:	10
3.2.5. Cuartiles:	10
3.2.6. Observaciones	10
4. Clase - 2020-01-16	11
4.1. Dudas	11
4.2. Medidas de localización	11
4.3. Medidas de variabilidad	11
4.3.1. Rango	11
4.3.2. Varianza muestral:	12
4.3.3. Desviación estándar	12
4.4. Excel	12
4.5. Ejemplo	12
5. Clase - 2020-01-21	13
5.1. Continuación	13
6. Clase - 2020-01-28	15
6.1. Otras medidas de localización	15
6.2. Excel steps	15
7. Clase - 2020-02-04	17
7.1. Combinaciones y permutaciones	17
7.1.1. En general	17
7.1.2. Combinaciones	17
7.1.3. Permutaciones	17
7.2. Explicación de número de placas en GT	18
7.2.1. Combinaciones de letras	18
7.2.2. Permutaciones de letras	18

8. Clase - 2020-02-06	19
8.1.	19
8.1.1. Ley de adición	19
8.1.2. Probabilidad condicional	19
9. Clase - 2020-02-13	21
9.1. Ejemplo del coronavirus	21
10. Clase - 2020-02-20	23
10.1. Solución del parcial	23
10.1.1. 1	23
10.1.2. 2	24

Capítulo 1

Clase introductoria

1.1. Clase introductoria

- Hay dos tipos de datos en estadística;
 1. Cualitativo: el cualitativo es por
 2. Cuantitativo:
- Distribución de frecuencias: nos dice qué tan frecuente es la distribución de los datos en un set.

Capítulo 2

Clase - 2020-01-09

2.1. Notas

- Tabla de frecuencias: Con todos los datos, la suma de todo es lo que se pone.
- Tabla de frecuencias relativas: cuando la suma de todo es uno.
- Tabla de frecuencias porcentual: cuando la suma de todo es 100 %.

2.2. Audit.xlsx

- Las diferentes categorías que se agrupan se les da el nombre de clase, mientras más peculiaridades se tengan por clase se tendrán más clases.
- La cantidad total de datos \equiv número de observaciones.
- El número de observaciones se le llama “n”.
- Si queremos 5 clases cada clase debe de tener el mismo ancho, esta para dar uniformidad a todos los intervalos para “comparar peras con peras”.
- Al ancho de clase que salga de la fórmula hay **que redondearlo para arriba**.
- Los histogramas:
 - Sólo se pueden hacer para variables cuantitativas, para números.
 - Cuando en el eje-x están intervalos son números.
 - Las barras estarán pegadas sin ningún gap entre ellas.
 - Son números enteros osea **discretos**.
 - En Excel: Seleccionar una barra \rightarrow click derecho \rightarrow Dar formato de serie de datos \rightarrow Ancho de rango \rightarrow bajarlo a 0 %.

Capítulo 3

Clase - 2020-01-14

3.1. Pasos de ordenamiento en Excel

- ↓ Seleccionar todos los datos
- ↓ Orden personalizado
- ↓ Advertencia antes de ordenad → Ampliar la selección → ordenar
- ↓ Ordenar por → Valores de celda → A a Z

3.2. Medidas de localización o tendencia central

3.2.1. Media:

- La media aritmética
- La media ponderada
- La notación que se utilizará sera x-barra:

$$\bar{x} = \frac{\sum_{i=1}^n (x_i)}{n} = \frac{(x_1 + x_2 + x_3 + \dots + x_n)}{n}$$

- n siendo el número de observaciones.

Definición de “media”: es un número.

- Cuando se agregan valores abnormales al promedio hace un cambio para arriba o para abajo que es significativo.

3.2.2. Mediana:

- Es un dato que denota cuánto mide la persona que está cabal en medio.
- Es el valor que parte a la mitad todos los datos.
- Cuando hay una cantidad impar va a haber uno, cuando es par pueden haber dos.
- La mediana no se ve afectada por los valores que están debajo de ella ni arriba de ella.

- Mediana = $\frac{n}{2}$

- Los datos de la media tienen que estar en orden ascendente para poder calcularse.
- El número que salga de mediana se usa la parte entera como límite inferior y el número redondeado para arriba es el límite superior, en los números pares.
- Si el valor es un número impar se agarra el del medio, si el valor es un numero par se agarra el de floor(enmedio) y el roundedUp(floor(enmedio)).

3.2.3. Moda:

- Es el número que más se repite en un set de datos.
- No hay fórmula.

3.2.4. Percentiles:

- Es un número que nos dice qué porcentaje de los datos esta debajo de él.
- La mediana es igual al percentil 50.
- Es el límite superior en el cual el porcentaje dicta, el percentil 20 se interpreta como el 20 % de datos están debajo de él.
- Percentil:

$$i = \left(\frac{p}{100} \right) \times n$$

donde p es percentil deseado, e i es el índice.

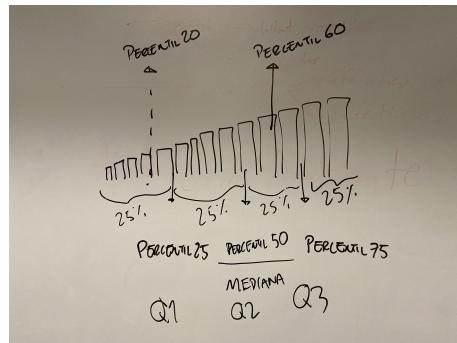


Figura 3.1:

3.2.5. Cuartiles:

- El cuartil son percentiles partidos en cuatro a lo largo del numero 1-100.
- **Ejemplo:** El percentil 25 % es el primer cuartil, el 50 % es el segundo cuartil, etcétera.

3.2.6. Observaciones

$$\text{Media} < \text{Mediana} < \text{Moda}$$

- Posiblemente hay más personas de baja estatura, por la media ser menor a la mediana.

$$\text{Media} > \text{Mediana} > \text{Moda}$$

Capítulo 4

Clase - 2020-01-16

4.1. Dudas

- **Nos preguntamos:** ¿cuál es la diferencia entre población y muestra?
 - Las muestras son parciales, la población es el total.
 - Población es el concepto de todo lo que existe, existe y va a existir en algún predeterminado lugar.

4.2. Medidas de localización

- Las medidas de localización dan una idea de lo que está pasando en un set de observaciones.

4.3. Medidas de variabilidad

4.3.1. Rango

Problemas de la media & solución es es la introducción del **rango**.

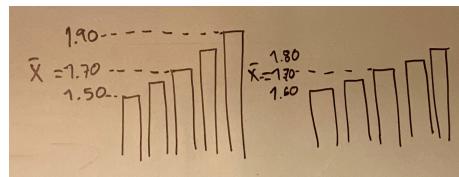


Figura 4.1: Misma media, diferente rango de datos

- Los datos están confusos ya que a pesar de tener la misma media los datos varían.
- Entonces se introduce el rango que se calcula como:

$$\text{Rango} = \text{Máximo} + \text{Mín}$$

Rango intercuartílico

La diferencia entre el cuartil tres y el cuartil uno.

$$R_{\text{Intercuartílico}} = Q_3 - Q_1$$

Entre Q_3 y el Q_1 , **Nos preguntamos:** ¿qué diferencias hay? es el 50 % de todos los que se parecen entre sí.

4.3.2. Varianza muestral:

- Definición de “Varianza Muestral”: cuánto varian en promedio los datos respecto a la media.
- Se denota por una S^2

■

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{x})}{n - 1}$$

- Este en el ejemplo esta expresado en centímetros².

4.3.3. Desviación estándar

- Definición de “Desviación estándar”: cuánto varían los datos respecto a la media.
- Es la raíz cuadrada de la varianza, se denota por solo S .
-
- $\sqrt{S^2} = S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{x})}{n - 1}}$
- En el ejemplo que tenemos está expresado en centímetros.
- Desviación estándar significa que en promedio los datos difieren respecto a la media.
- Cuando hay una desviación estándar está alta nos dice qué tanto se parecen los datos, qué tan variable es el grupo de datos.
- DE: 15 respecto a otra de DV:7, dice que hay más variación en la desviación estándar de 15.

4.4. Excel

- Para fijar una celda usar la letra de la columna encerrada por signos de dólar, así: \$E\$56
- Para desviación estándar usar fórmula: =DESVEST.M(¡TodosLosDatosOriginales¡) .

4.5. Ejemplo

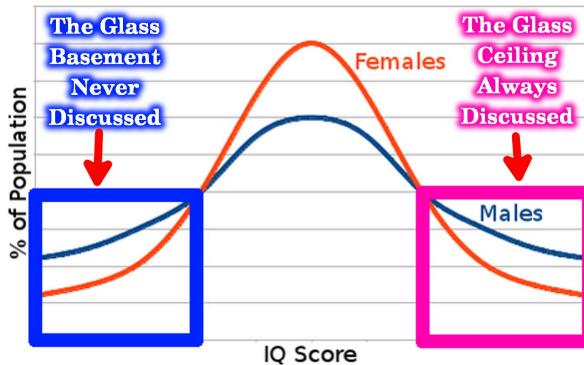


Figura 4.2: Niveles de IQ entre hombres y mujeres

- Las deducciones son que los hombres tienen una desviación estándar mayor.

Capítulo 5

Clase - 2020-01-21

Preliminares

- RAPORT

5.1. Continuación

- Puntos z :

$$\text{Punto} Z_i = \frac{x_i - \bar{x}}{s}$$

- *Ejemplo:* Las estaturas de la UFM & UVG.

UFM	UVG
$\bar{x} = 172\text{cm}$ MAX: 194cm	$\bar{x} = 1,68$ MAX: 197cm
$S = 9,71\text{cm}$ MIN: 156cm	$S = 10,9\text{CM}$ MIN: 145cm

- Punto z del ejemplo:

$$Z_{\text{MAX UFM}} = \frac{194 - 172}{9,71} = 2,26 \text{ Desviación estándar del promedio}$$

$$Z_{\text{MIN UFM}} = \frac{156 - 172}{9,71} = -1,65 \text{ Son negativos por que son menores a la media}$$

$$Z_{\text{MAX UVG}} = \frac{197 - 168}{10,9} = 2,66 \text{ Desviaciones estándar de su media}$$

$$Z_{\text{MIN UVG}} = \frac{145 - 168}{10,9} = 2,11 \text{ Está a ciertas desviaciones estándar de la media}$$

$$Z_{172} = \frac{172 - 172}{9,74} = \frac{0}{9,71}$$

- *Definición de “punto Z”:* es cuando

- La gráfica Z está centrada en el número cero, el cero lo que quiere decir es que de ahí parte la distribución Z, es decir el punto Z es el que esta exactamente a 0 de desviación estándar.
- Volver un número un punto Z se le llama estandarizar.
- Interpretemos: La gran mayoría de datos están concentrados en el centro. Ver el punto crítico y el punto de inflexión.

$$\pm z = 160 \% \pm z = 295 \%$$

Figura 5.1:

-
- Distribución en la forma de montaña o de campana: la distribución presentada de una manera línea continua contempla una cantidad infinitesimal de clases.

Capítulo 6

Clase - 2020-01-28

6.1. Otras medidas de localización

- Media ponderada:

-

$$\bar{x} = \frac{\sum_{i=1}^n W_i \cdot X_i}{\sum_{i=1}^n W_i}$$

- Donde W es “weight” o el peso que le vamos a delegar a cada clase.
- Toma en cuenta del peso de lo que estamos midiendo y a base de eso resulta la media ponderada. La media aritmética no toma en cuenta el peso.
- X_i es la observación.

- *Definición de “media ponderada”:*

- Media a partir de datos agrupados:

-

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

- La media a partir de datos agrupados se tiene un resultado con un nivel moderado de incertidumbre.
- *Citación: “Cuando agrupamos datos perdemos información.”*
- La media aritmética no va a ser exactamente igual que la media a partir de datos agrupados.

- Varianza a partir de datos agrupados:

-

$$S^2 = \frac{\sum f_i (M_i - \bar{x})^2}{n - 1}$$

- M_i es la media de las clases, si tengo un clase de 80-90 $M_i = \frac{(80+90)}{2}$.
-

-

6.2. Excel steps

- To show empty data fields:

↓ right click en row labels (random)

- ↓ Configuración de campo ...
- ↓ Diseño e impresión
- ↓ Mostrar elementos sin datos (cheque)

Capítulo 7

Clase - 2020-02-04

7.1. Combinaciones y permutaciones

7.1.1. En general

- N es el número total de resultados posibles, y n la muestra.
- Combinaciones: $\binom{N}{n}$: de un conjunto de N elementos cuántos

7.1.2. Combinaciones

- Combinaciones:

$$C_N^n =_N C_n = \binom{N}{n} = \frac{N!}{n!(N-n)!}$$

- La más usual es $_N C_n$ o $\binom{N}{n}$
- Cuenta la cantidad de los resultados muestrales tomando en cuenta cada combinación sin orden.
- Ejemplo:

456 – HBJ

$$C_3^7 = \frac{10!}{3!7!} = \frac{10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{(3 \times 2 \times 1)(7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1)} = \frac{720}{6} = 120$$

7.1.3. Permutaciones

-

$$P_N^n =_N P_n = n! \binom{N}{n} = \frac{N!}{(N-n)!}$$

- Ejemplo:

$$n! \binom{N}{n} = \frac{10!}{7!} = \frac{10 \times 9 \times \dots \times 1}{7 \times 6 \times \dots \times 1} = 720$$

7.2. Explicación de número de placas en GT

7.2.1. Combinaciones de letras

$$\frac{21!}{3!(21-3)!} = \frac{21!}{3!18!} = 1330$$

7.2.2. Permutaciones de letras

$$\begin{aligned}\frac{21!}{18!} &= 21 \times 20 \times 19 = 7,980 \\ 720 \times 7980 &= 5,745,600\end{aligned}$$

Capítulo 8

Clase - 2020-02-06

8.1.

- Las probabilidades marginales están en los márgenes o en los totales.
- Las probabilidades conjuntas, son las que están agrupadas.
- Ejemplo:

$A = \text{Reggaeton}$

$$P(A) = 0,4231 \quad \# \text{Cuál es la probabilidad que ocurra el evento "que le gusta el reggaeton"}$$
$$P(A^c) = 1 - P(A) = 0,5769 \quad \# \text{Cual es la probabilidad que ocurra el complemento del evento A.}$$

8.1.1. Ley de adición

- La ley de la adición:
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
- Se tiene que restar $P(A \cap B)$ porque hay instancias que la probabilidad está por arriba del 100 %.

8.1.2. Probabilidad condicional

- $P(\text{Rock}|\text{Fut})$ dado que les gusta el fut, a cuántos les gusta el Rock.
- $P(\text{Fut}|\text{Rock})$ dado que les gusta el rock, a cuántos les gusta el fut.
- $$P(A|B) = \frac{P(A \cap B)}{P(B)}$$
- Si la $P(A) = P(B)$ los eventos son independientes, cuando $P(A) \neq P(B)$ ó $P(A) \neq P(B)$ son dependientes.
 - Ser dependiente significa que las probabilidades de la música respecto de, fut varía proporcionalmente.
 - Si alguien le gusta el fútbol tiene que ver con la música que le gusta.

Capítulo 9

Clase - 2020-02-13

9.1. Ejemplo del coronavirus

- Que las personas infectadas del coronavirus se mueran:

Si esto sube el porcentaje sube

$$P(\text{ Morir} \mid \text{Estoy_enfemo}) = \frac{\overbrace{42,000}^{\text{Si esto sube el porcentaje sube}}}{60,000} = 0,03$$

Capítulo 10

Clase - 2020-02-20

10.1. Solución del parcial

10.1.1. 1

- Prevalencia:

$$\text{Prevalencia} = \frac{1,500}{100,000} = \underbrace{0,015}_{\text{Probabilidades apriori}}$$

- De una muestra, resultó positiva para 965 de 1,000 personas que se sabe que sí tenían la enfermedad:

$$P(+|E) = 0,965$$

- De una muestra resultó ser negativa para 97 de 100 personas que se sabe que no tenían la enfermedad:

$$P(-|NE) = 0,97$$

Probabilidades apriori	Probabilidades condicionales	Probabilidades posteriores	
$P(E) = 0,015$	$P(+ E) = 0,965$ $P(- E) = 0,035$	$P(E)P(+ E) = 0,014$ $P(E)P(- E) = 0,000525$	$P(E +)$ $\underline{P(E)P(+)}$ $\frac{0,014}{0,04403}$
$P(NE) = 0,985$	$P(+ NE) = 0,03$ $P(- NE) = 0,97$	$P(NE)P(+ NE) = 0,02955$ $P(NE)P(- NE) = 0,95545$	

- Probabilidad que sea un falso negativo:

- $P(-|E) = 0,035$

- Probabilidad de salir positivo:

$$P(+) = \frac{P(E)P(+|E)}{P(E)P(+|E) + P(NE)P(+|NE)} = \frac{0,014}{0,014 + 0,02955} = 0,04403$$

- # Tomar en cuenta que para sacar $P(+)$ tomo todos los escenarios en los que se calcula la probabilidad de salir positivo.
- Probabilidad de salir negativo:

$$P(-) = \frac{P(E)P(-|E)}{P(E)}$$

10.1.2. 2

■