Estadística I

David Gabriel Corzo Mcmath

2020-Jan-07 11:57:29

Índice general

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 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 → click derecho → Dar formato de serie de datos → Ancho de rango → bajarlo a 0 %.

Capítulo 3

Clase - 2020-01-14

3.1. Pasos de ordenamiento en Excel

- $\downarrow$ Seleccionar todos los datos
- $\downarrow$ Orden personalizado
- $\downarrow$ Advertencia antes de ordenad $\rightarrow$ Ampliar la selección $\rightarrow$ ordenar
- $\downarrow$ Ordenar por $\rightarrow$ Valores de celda $\rightarrow$ 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:

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

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

- 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.

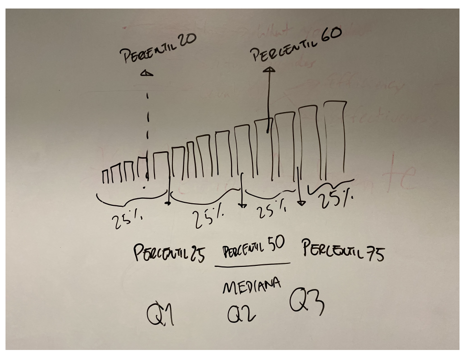

Figura 3.1:

-

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**.

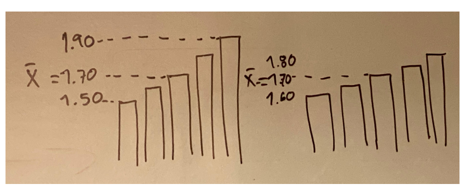

Figura 4.1: Misma media, diferente rango de datos

- 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 centímetros^2.

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: E56
- Para desviación estándar usar fórmula: =DESVEST.M(¡TodosLosDatosOriginales¿) .

4.5. Ejemplo

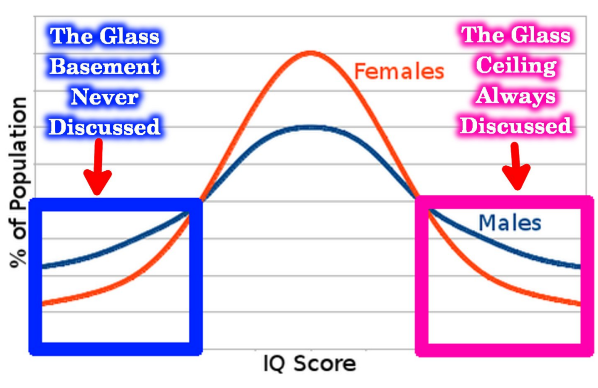

Figura 4.2: Niveles de IQ entre hombres y mujeres

- 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
$\bar{x} = 172cm$ MAX: 194cm $S = 9{,}71cm$ MIN: 156cm	$\bar{x} = 1{,}68$ MAX: 197cm $S = 10{,}9CM$ 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 $_NC_n$ o $\binom{N}{n}$
- Cuenta la cantidad de las resultados muestrales tomando en cuenta cada combinación sin orden.
- Ejemplo:

$$456 - \text{ HBJ}$$

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

7.1.3. Permutaciones

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

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

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

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

Capítulo 8

Clase - 2020-02-06

8.1.

- Las probabilidades marginales están en los márgenes o en lso grand 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 $P(A \cap B)$ por que hay instancias que la probabilidad está por arriba del 100 %.

8.1.2. Probabilidad condicional

- $P(Rock|Fut)$ dado que les gusta el fut, a cuántos les gusta el Rock.
- $P(Fut|Rock)$ 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) ¡P(B) ó P(A) ¿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:

$$P(\text{ Morir } | \text{ Estoy\_enfemo }) = \frac{\overbrace{42,000}^{\text{Si esto sube el porcentage sube}}}{60,000} = 0{,}03$$

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

- 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

- Permutaciones:

Letra 1	Letra 2	Letra 3	Dígito 1
26 opciones	25 opciones	24 opciones	

Una permutación es un proceso sin remplazo. No se repiten.
Fórmula:

$$_{N}P_n = \frac{N!}{(N-n)!}$$

Para el inciso en letras:

$$_{26}P_3 = \frac{26!}{(26-3)!} = 15,600$$

Para el inciso en números:

$$_{10}P_3 = \frac{10!}{(10-3)!} = 720$$

Si multiplico las dos permutaciones:

$$15,600 \cdot 720 = 11,232,000$$

Saco la cantidad de permutaciones para sacar el total de posibilidades con repetición:

$$(26)^3 = 17,576$$
$$(10)^3 = 1,000$$
$$\text{\# Multiplico:} \quad 17,576 \times 1,000 = 17,576,000$$

Calculo la probabilidad dividiendo:

$$\frac{_{26}P_3 \times_{10} P_3}{(26)^3 \times (10)^3} \approx 0{,}6390$$

10.1.3. 3

- Para calcular independencia:

$$P(A) \times P(B) = P(A \cap B)$$

Capítulo 11

Clase - 2020-02-25

11.1. Variables aleatorias

- ***Definición de "Experimento":*** *Proceso que genera resultados posibles. Los resultados de un experimento pueden ser cualitativos y cuantitativos*
- ***Definición de "Variable aleatoria":*** *una variable aleatoria es una descripción numérica del resultado de un experimento.* Puede ser discreta o contínua. ***Recordar lo siguiente:*** *Discreta es que siempre tiene un número positivo. Contínua es que puede tener decimales infinitesimales.*
- ***Ejemplo:*** *Si un hombre planea sacar a bailar a 10 mujeres, las variables aleatorias o posibles resultados del experimento son un número entre 0 y 10.*
- ***Interesante:****En econometría, Una variable dummy es cuando se pone un resultado a algo, 0 a hombres y 1 a mujeres por ejemplo.*

11.2. Distribución de probabilidad discreta

- ***Definición de "Distribución uniforme discreta":*** *Cuando probabilidad de cada variable aleatoria es la misma.*

11.3. Valor esperado y varianzas

- Se puede calcular el valor esperado, el valor esperado es el promedio de todos los valores posibles.

$$E(x) = \underbrace{\mu}_{\text{Parámetro poblacional } \approx \bar{x}} = \sum xf(x)$$

- Cálculo de un valor esperado de una distribución discreta.

Clase	UMAS	f(x)
Análisis matemático	4.5	0.43
Estadística	3	0.29
Historia	3	0.29
	10.5	$\sum(xf(x)) = 1{,}50$

- Varianza de valor ponderado:

$$\omega^2 = \sum(x + \mu)^2 f(x)$$

- Desviación estándar:

$$\omega = \sqrt{\sum (x-\mu)^2 f(x)}$$

11.4. Experimento binomial

- ***Definición de "Evento binomial":*** *es un evento que sólo puede tener dos resultados posibles.*
- Propiedades de un experimento binomial:
  1. El experimento consiste en una serie de n ensayos idénticos.
  2. En cada ensayo hay dos resultados posibles. Resultados posibles: éxito o fracaso.
  3. La probabilidad de éxito p no cambia de ensayo a otro. La probabilidad de fracaso se denota como $1-p$, tampoco cambia de un ensayo a otro.
  4. Los ensayos son **independientes**.
- Experimento: Lanzar 3 monedas:

$x =$ Cara $P = 0{,}50$	$\neg x =$ Escudo $1 - P = 0{,}5$

- ***Interesante:*** *Niños y niñas*
  - De cada 100 niñas hay 106 niños:

$$\text{Niños } = 106 = \frac{106}{206} = 0{,}53$$

$$\text{Niñas } = 100 = \frac{100}{206} = 0{,}47$$

  - Probabilidad que salgan 7 niños en varones:

$$P(\text{ 7 niños }) = (0{,}53)^7 = 0{,}011$$

- ***Interesante:*** *Monedas*

C	C	C	CCC $\rightarrow$ 3
		E	CCE $\rightarrow$ 2
	E	C	CEC $\rightarrow$ 2
		E	CEE $\rightarrow$ 1
E	C	C	ECC $\rightarrow$ 2
		E	ECE $\rightarrow$ 1
	E	C	EEC $\rightarrow$ 1
		E	EEE $\rightarrow$ 0

- Función de probabilidad binomial:

$$f(x) = \underbrace{\binom{n}{x}}_{\text{Combinaciones}} p^x (1+p)^{(n-1)}$$

- $f(x)$ = Probabilidad de x éxitos en n ensayos.
- n = número de ensayos
- $\binom{n}{x} = \frac{n!}{x!(n-x)!}$: Combinaciones.
- p = probabilidad de un éxito en cualquiera de los ensayos.
- $1 - p$ = probabilidad de un fracaso en cualquiera de los ensayos.

▪ Ejemplo con la función de probabilidad binomial:

$f(x)$	$\binom{n}{x}$	p^x	$(1-p)^{n-x}$	=
$f(0)$	1	$(0{,}50)^0$	$(0{,}50)^3$	= 0,125
$f(1)$	3	$(0{,}50)^1$	$(0{,}50)^2$	= 0,375
$f(2)$	3	$(0{,}50)^2$	$(0{,}50)^1$	= 0,377
$f(3)$	1	$(0{,}50)^3$	$(0{,}50)^0$	= 0,125

▪ Para sacar la probabilidad binomial en Excel:

↓ =DIST.BINOMIAL(num_exit,0.5,FALSO), el núm_éxito es cuántos éxitos quiero, 0.5 → la probabilidad de éxito, FALSO.

↓ Abrir f(x)

Capítulo 12

Clase - 2020-02-27

12.1. Probabilidad de Poisson

- Función de probabilidad de Poisson:
$$f(x) = \frac{\mu e^x}{x!}$$
    - f(d): probabilidad de x ocurrencias en un intervalo
    - μ: valor esperado o número medio de ocurrencias en un intervalo.
    - e: número de Euler.
    - Pensar μ como $\mu = \frac{\text{Eventos}}{\text{Tiempo}}$
- Sirve para predecir cosas como:
    - Las muertes de los magistrados en los estados unidos.
    - Patrones en los huracanes.
- Ejemplo de cajeros automáticos de un banco:
- $\mu = \frac{10 \text{ carros}}{15 \text{ minutos}} = 0.\overline{66} = \frac{2}{3}$ e$^{-}x/x! {-} 10 : 10xy$