**ESTADÍSTICA**

**VARIABLE ESTADÍSTICA UNIDIMENSIONAL**

POBLACIÓN: Conjunto de elementos sobre el que se realiza un estudio estadístico.

MUESTRA: Parte de la población que estudiamos. **E**l tamaño de la muestra es **n.**

MUESTREO: Es el conjunto de técnicas que aplicamos para extraer muestras representativas de una población

* MUESTREO ALEATORIO: Todos tienen la misma probabilidad de ser seleccionados. Se puede corregir aumentando la muestra.
	+ SIMPLE: Se numeran y se extra al azar.
	+ SISTEMÁTICO: Se numeran, se calcula h= N/n y se escoge cada h en la muestra.
	+ ESTRATIFICADO: Realizar grupos homogéneos. Muestra simple de cada uno.
	+ CONGLOMERADOS: Grupos heterogéneos, pero parecidos entre sí. Muestra simple.
* MUESTREO NO ALEATORIO: No tienen la misma probabilidad. Se produce un sesgo en la estadística y no se puede corregir.

INDIVIDUO: Cada uno de los elementos de la población o de la muestra.

VARIABLES ESTADÍSTICAS: Cada una de las propiedades o características que podemos estudiar en una población o muestra.

* CUALITATIVAS : Los valores que toma no son números, sino cualidades y no se pueden cuantificar.
* CUANTITATIVAS: Los valores que toma son números y se pueden cuantificar.
	+ DISCRETA: En cada intervalo la variable solo puede tomar un número finito de valores.
	+ CONTINUA: La variable puede tomar tantos valores como queramos por pequeño que sea el intervalo. En este caso , se utiliza para hacer los cálculos la marca de clase, que es el punto medio del intervalo.

FRECUENCIAS:

- Frecuencia Absoluta : ($f\_{i})$ De un dato es el número de veces que aparece en la muestra. Recuento.

- Frecuencia Relativa : ($h\_{i})$ De un dato, es el cociente entre la frecuencia absoluta y el número total de datos.$\left(h\_{i}=\frac{f\_{i}}{n}\right)$

- Frecuencia absoluta acumulada: ($F\_{i})$ De un dato es la suma de todas las frecuencias absolutas de los datos menores o iguales que él. ($F\_{i}=\sum\_{j=1}^{i}f\_{i}$)

- Frecuencia relativa acumulada ($H\_{i}$). De un dato, es la suma de todas las frecuencias relativas de los valores menores o iguales que él.$ (H\_{i}=\sum\_{j=1}^{i}h\_{i}$)

Para determinar el porcentaje que representa un dato respecto del total tenemos que considerar sus frecuencias relativa o relativa acumulada y multiplicarlas por 100.

TIPOS DE GRÁFICOS**.**

* Barras ( Variable discreta)
* Histograma ( Variable continua)
* Poligonal de frecuencias, se hace sobre el de barras o el histograma.
* Sectores. Con los porcentajes.
* Cajas y bigotes.

MEDIDAS DE CENTRALIZACIÓN**.**

MEDIA ARITMÉTICA.$\overbar{x}= \frac{\sum\_{}^{}x\_{i}f\_{i}}{n}$

MODA. La variable que tiene mayor frecuencia absoluta. (Puede haber más de una, en ese caso se dice que es bimodal, trimodal … , si todas tienen la misma frecuencia, se dice que no tiene moda). En el caso de una variable continua se habla de intervalo modal.

MEDIANA:Es el valor que ocupa la posición central de los datos, después de ordenarlos, o la media de los valores centrales, si el número de datos es par. Si la variable es continua, hablamos de intervalo mediano.

 CUARTILES**:** Es el valor que ocupa la posición ¼, 2/4 y 3/4 de los datos, después de ordenarlos, o la media de los valores centrales, si el número de datos es par.

MEDIDAS DE DISPERSIÓN

RANGO o RECORRIDO: Es la diferencia entre el mayor y el menor valor de la variable.

 R= Máx - Mín

DESVIACIÓN MEDIA $DM= \frac{\sum\_{i=1}^{n}f\_{i}∙\left|x\_{i}-\overbar{x}\right|}{n}$

VARIANZA $σ^{2}= \frac{\sum\_{}^{}x\_{i}^{2}f\_{i}}{n}- \overbar{x}\_{i}^{2}$

DESVIACIÓN TÍPICA $σ= \sqrt{σ^{2}}$

INTERVALOS$\left(\overbar{x}-σ,\overbar{x}+σ\right)$ **= 68%**

$\left(\overbar{x}-2σ,\overbar{x}+2σ\right)$ **= 95%**

$\left(\overbar{x}-3σ,\overbar{x}+3σ\right)$ **= 99%**

COEFICIENTE DE VARIACIÓN **:** $ CV=\frac{ σ}{\overbar{x}}$

|  |  |  |
| --- | --- | --- |
|  | POBLACIONAL | MUESTRAL |
| MEDIA | µ | $$\overbar{x}$$ |
| DESVIACIÓN TÍPICA | σ | s |

**VARIABLE ESTADÍSTICA BIDIMENSIONAL**

Tabla de doble entrada, en ella se establecen dos entradas, una para los datos de cada variable y colocamos las frecuencias absolutas de cada par de daros en las casillas centrales.

Tabla de frecuencias marginales, se obtienen al estudiar por separado cada una de las variables que forman la variable bidimensional.

DIAGRAMA DE DISPERSIÓN

* + Dependencia funcional
	+ Dependencia aleatoria o correlación:
		- Positiva (Directa)
		- Negativa (Inversa)
		- Nula

CENTRO DE GRAVEDAD $\left(\overbar{x}, \overbar{y}\right)$

COVARIANZA $σ\_{xy}=\frac{\sum\_{}^{}x\_{i}y\_{i}}{n}- \overbar{x}\overbar{y}$

COEFICIENTE DE CORRELACIÓN $r=\frac{σ\_{xy}}{σ\_{x}σ\_{y}}$$ -1\leq r\leq 1$

RECTA DE REGRESIÓN DE Y SOBRE X $y-\overbar{y}=m \left(x-\overbar{x}\right) m=\frac{σ\_{xy}}{σ\_{x}^{2}}$

RECTA DE REGRESIÓN DE X SOBRE Y x$-\overbar{x}=m \left(y-\overbar{y}\right) m=\frac{σ\_{xy}}{σ\_{y}^{2}}$

**DISTRIBUCIÓN BINOMIAL**

 X$≡B(n, p)$ **n**= nº veces que se hace el experimento y **p**= probabilidad de que ocurra el suceso A. Las n veces que se realiza el experimento son independientes, lo que ocurra en una de ellas no influye en el resto de experimentos.

P (X=$x\_{i})= \left(\genfrac{}{}{0pt}{}{n}{i}\right)p^{i}\left(1-p\right)^{n-i}$

Media $μ=np$

Varianza $σ^{2}=np (1-p)$

**DISTRIBUCIÓN NORMAL**

 X$≡N (μ,σ)$

Tiene que ser una variable aleatoria continua

Tipificar x : $x\rightarrow \frac{x-μ}{σ} \rightarrow N(0,1)$

PROPIEDADES

P$\left(z\geq -a\right)=P\left(z\leq a\right)$

P$\left(z\geq a\right)=1-P\left(z\leq a\right)$

P$\left(z\leq -a\right)=P\left(z\geq a\right)=1- P\left(z\leq a\right)$

P($a\leq z\leq b)$=$P\left(z\leq b\right)-P(z\leq a)$

**INTERVALOS CARACTERÍSTICOS DE UNA DISTRIBUCIÓN NORMAL**

 $( μ-z\_{\frac{α}{2}}∙σ , μ+z\_{\frac{α}{2}}∙σ)$

Se llama valor crítico a $z\_{\frac{α}{2}}$

Ejemplo: p= 90%=0,90 α= 1-0,90=0,10 → α/2= 0,05 → 1- α/2=1-0,05=0,95

 $P( z<z\_{\frac{α}{2}}$)= 0,95 → $z\_{\frac{α}{2}}=1,645$

90% → $z\_{\frac{α}{2}}=1,645$

95% → $z\_{\frac{α}{2}}=1,96$

99% → $z\_{\frac{α}{2}}=2,58$

Si n es suficientemente grande, se puede aproximar la distribución binomial a una normal de media µ= np y$ σ= \sqrt{np (1-p)}$

**TEOREMA CENTRAL DEL LÍMITE N** $ $

Si generalizamos la aproximación de la binomial a la normal, cuando n es suficientemente grande, se cumple que $ X≡N(n∙μ,σ∙\sqrt{n})$

**DISTRIBUCIÓN MEDIA MUESTRAL**

 n>30 $\overbar{X}≡N(μ,\frac{σ}{\sqrt{n}}$)

**DISTRIBUCIÓN PROPORCIONAL MUESTRAL**

 N= ( p, $\sqrt{\frac{pq}{n}})$

**INTERVALOS DE CONFIANZA PARA LA MEDIA POBLACIONAL**

 n$\geq 30 $ Distribución normal $\overbar{ X}\rightarrow N(μ,\frac{σ}{\sqrt{n}}$)

Nivel de confianza 1-α $\left(\overbar{x}-z\_{\frac{α}{2}}∙\frac{σ}{\sqrt{n} }, \overbar{x}+z\_{\frac{α}{2}}∙\frac{σ}{\sqrt{n}} \right)$

Amplitud del intervalo de confianza es $L=2z\_{\frac{α}{2}} \frac{σ}{\sqrt{n}} $

**ERROR MÁXIMO ADMISIBLE**

 $E= z\_{\frac{α}{2}}∙\frac{σ}{\sqrt{n} }$

**TAMAÑO DE LA MUESTRA**

 $n= \left[\frac{z\_{\frac{α}{2 }}∙ σ}{E}\right]^{2}$

**TAMAÑO DEL INTERVALO**

 $\left(\overbar{x}+z\_{\frac{α}{2}}∙\frac{σ}{\sqrt{n} }\right)-( \overbar{x}-z\_{\frac{α}{2}}∙\frac{σ}{\sqrt{n}} )$