



Apuntes clase 23 febrero Diseño de Experimentos

Esteban Padilla

Instituto Tecnológico de Costa Rica

27 febrero 2022



Administrativos

- Se realizo el quiz 2 al inicio de la clase.
- Dado la entrega de primera tarea en semana 5, se cambia a ver principios de estadísticas, la semana que viene se empieza con R.
- El inicio de la clase estuvo pixeleado, luego mejoro.



Repaso de estadística

Hay muchas intuiciones que están equivocadas en conceptos de estadística a fin de aclararlos la clase consistió en un repaso que comprende:

- Definiciones.
- Mediciones de tendencia central.
- Distribuciones de probabilidad.
- Muestreo.
- Inferencia estadística y prueba de hipótesis.



Definiciones

Estadística:

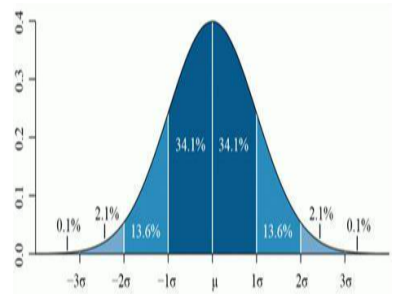
La estadística es una disciplina científica que se ocupa de la recolección, organización e interpretación de los datos. Hay dos clases de estadística:

- Descriptiva, la cual presenta organiza y resume datos, su fin es comprender los datos de una mejor manera.
- Inferencial, hace conclusiones de una población basado en los datos observados de una muestra de dicha población.

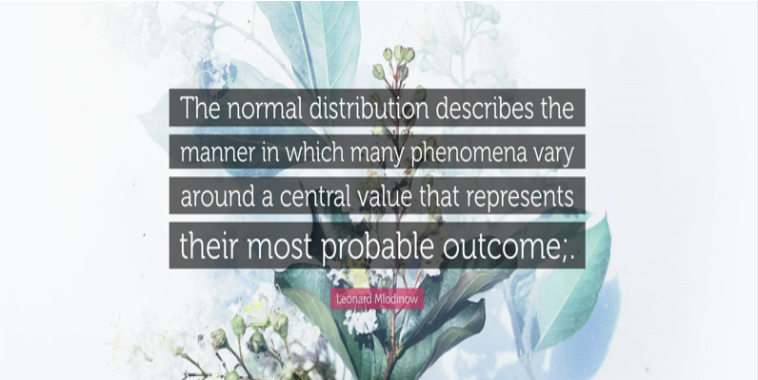
Estadística descriptiva

Resume, describe y caracteriza la población estudiada, determina la distribución de la muestra, se busca que sea normalmente distribuida ya que la mayoría de la estadística requiere que la muestra presente una distribución normal. Determina si la muestra es comparable a la población, es desplegada a través de gráficos y tablas, y es reportada como mediciones de tendencia central.

- Distribución normal es común en la naturaleza.
- Distribución normal permite pruebas T de welch.
- La mayoría de la estadística requiere de una distribución normal.
- Se emplean muestras ya que existen poblaciones muy grandes.
- Hay que ser cuidadoso en la recolección de la muestra, debe ser representativa de la población.



Estadística descriptiva



The normal distribution describes the manner in which many phenomena vary around a central value that represents their most probable outcome;.

Leonard Mlodinow

Experimento aleatorio y espacio muestral

Experimento aleatorio: Experimento que puede dar distintos resultados cada vez que se lleva a cabo a pesar de que sea repetido de la misma manera cada vez. contrario a experimentos deterministas en donde los resultados son los mismos.

Espacio Muestral: Conjunto de posibles resultados para un experimento aleatorio.

¿Existen realmente los eventos aleatorios?

- Interpretación bayesiana: Ciertos fenómenos que se creen aleatorios con el suficiente conocimiento se vuelven determinísticos.
- Pensamiento frecuentista: Define la probabilidad en términos de la experimentación, define un fenómeno en base a como se comporta.

Probabilidad

Es la medida de la certidumbre o posibilidad asociada a un evento, se expresa como un numero entre 0 y 1.

- Para un sistema aleatorio la probabilidad esta normalmente asignada con base a un modelo razonable del sistema bajo estudio.
- La probabilidad de un evento es la suma de la probabilidad de los posibles resultados de dicho evento. Ejemplo la probabilidad de que un dado de 3 o un 5 es la suma de dichas probabilidades, $1/6 + 1/6 = 1/3$.



Promedio, Mediana y Moda

Promedio

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

Se puede encontrar para toda la población, es sensible a valores extremos en la muestra

$$\bar{x} = \frac{2+3+4+6+6+5+2+3+3+6+6}{11}$$

Mediana

Valor que esta exactamente a la mitad de los datos si estos se ordenan de menor a mayor

Es preferible utilizar la Mediana cuando se esta en presencia de valores extremos

15 elem
2, 3, 4, 4, 6, 7, 7, 7, 8, 8, 8, 8, 9, 14, 17
Mediana = 7

14 elem
2, 3, 4, 4, 6, 7, 7, 8, 8, 8, 8, 9, 14, 17
Mediana = $(7 + 8) / 2$
= 7,5

Moda

Es el valor que se repite mas veces en una muestra

6, 3, 9, 6, 6, 5, 9, 3



Varianza y desviación estándar

Varianza: Es una medida de la variabilidad o de la dispersión en la distribución.

$$\sigma^2 = \frac{\sum i^n (X_i - \bar{X})^2}{n - 1} \quad (1)$$

Sin embargo, se emplea en mayor medida la desviación estándar ya que esta en las mismas unidades de la variable por lo que es más fácil de interpretar.

$$\sigma = \sqrt{\sigma^2} \quad (2)$$

Distribución de probabilidad

- La distribución de probabilidad es la descripción de las probabilidades asociada a los posibles valores de una variable aleatoria X , si la variable es discreta se especifica como una lista de los posibles valores de la variable aleatoria al lado de su respectiva probabilidad.
- Algunas veces la distribución de probabilidad se especifica como una formula.
- Existen distribuciones que presentan comportamientos pre-establecidos como la distribución de Bernoulli, la distribución uniforme, distribución binomial, la distribución normal, la distribución de poisson y la distribución exponencial.

Distribuciones de probabilidad

Distribución de Bernoulli: Cada evento tiene solamente dos posibles resultados 0 y 1. Las probabilidades de los resultados no necesariamente tienen que ser iguales.

$$P(X) = \begin{cases} \phi & x = 1 \\ 1 - \phi & x = 0 \end{cases} \quad (3)$$

Distribución uniforme: Todos los posibles resultados tienen la misma probabilidad, se le conoce también como distribución rectangular. un ejemplo de este tipo de distribución es el tiro de un dado.

$$P(x) = \frac{1}{b - a} \quad (4)$$

Donde b es el valor maximo y a el valor minimo.

Distribuciones de probabilidad

Distribución binomial: La distribución binomial es una distribución de probabilidad discreta que cuenta el número de éxitos en una secuencia de n ensayos.

$$P(x) = \frac{n!}{(n-x)!x!} p^x q^{(n-x)} \quad (5)$$

En donde x es el numero de éxitos, p es la probabilidad del éxito y q es la probabilidad del fallo.

Distribución normal: Se le llama distribución normal porque surge con normalidad en múltiples fenómenos en la naturaleza. Es un modelo descriptivo que describe situaciones del mundo real. Si una variable es normalmente distribuida se pueden hacer inferencias acerca de los valores de dicha variable.

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (6)$$

En donde x es la variable, μ es el promedio y σ es la desviación estándar.

Distribuciones de probabilidad

Distribución Poisson: Es aplicable a situaciones en donde los eventos ocurren en puntos aleatorios en el tiempo y el espacio, el interés es el numero de ocurrencias del evento.

$$P(x) = \frac{e^{-\lambda} \lambda^k}{x!} \quad (7)$$

En donde x es el numero de eventos en un intervalo, λ es la tasa de ocurrencia de los eventos y k es el largo del intervalo de tiempo.

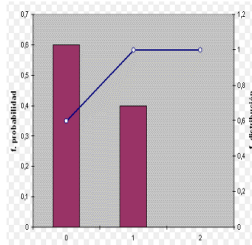
Distribución exponencial: Si en la distribución de poisson tenemos eventos, la distribución exponencial modela el intervalo de tiempo entre los eventos.

$$P(x) = \lambda e^{-\lambda x} \quad (8)$$

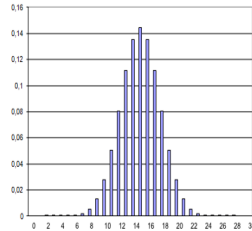
En donde λ es conocida como la tasa de eventos en cualquier tiempo t , la inversa de λ es conocido como tiempo medio para el fallo MTTF por sus siglas en ingles.

Distribuciones de probabilidad

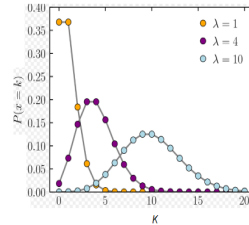
Bernoulli



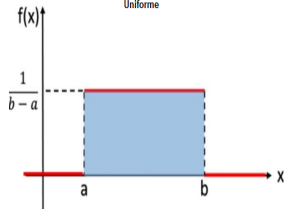
Binomial



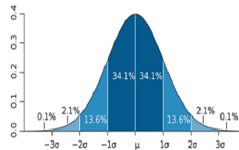
Poisson



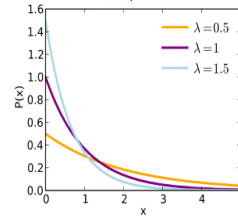
Uniforme



Normal



Exponencial



Conceptos

Se tiene una población sobre la cual se quieren establecer conclusiones, de esta población se extrae una muestra empleando métodos de muestreo ya sea probabilísticos o no probabilísticos.

métodos probabilísticos

- **Muestreo aleatorio simple:** Cualquier miembro de la población tiene la misma probabilidad de ser seleccionado. se debe incluir toda la población para la selección.
- **Muestreo sistemático:** Toda la población tiene un numero y se escogen elementos en intervalos regulares. hay que evitar que en la lista de selección exista un patrón que genere un sesgo.
- **Muestreo estratificado:** Se divide la población en sub-poblaciones según una característica importante, basado en las proporciones de las sub-poblaciones de determina cuantos elementos de cada una debe contener la muestra para que asi cada subgrupo debe estar propiamente representado.
- **Muestreo por grupos:** Se divide la población en subgrupos, cada subgrupo debe tener características similares a toda la muestra, se seleccionan subgrupos enteros de manera aleatoria, es un método bueno para poblaciones largas y dispersas.

Tamaño de la muestra

Existe una formula para calcular el tamaño de la muestra dada por:

$$n = \frac{Z^2 N p q}{e^2 (N - 1) + Z^2 p q} \quad (9)$$

En donde Z indica el nivel de confianza (1.96 para un 95%), e es el error de muestreo deseado, p y q son proporciones de individuos con y sin una característica.

Tamaño de la muestra

De la ecuación anterior se pueden crear tablas con el tamaño de la muestra deseada, a continuación se muestra una tabla creada con una confianza de 95%.

Si se tiene una población de 10000 y se requiere un error de 2%, el tamaño de la muestra debe de ser de 2000 elementos/individuos .

Tamaño de la población	+/- 1%	+/- 2%	+/- 3%	+/- 4%	+/- 5%	+/- 6%
500					222	83
1000				385	286	91
1500			638	441	316	94
2000			714	476	333	95
2500		1250	760	500	345	97
3000		1364	811	517	353	98
3500		1458	843	530	359	98
4000		1538	870	541	364	98
4500		1607	891	519	36	98
5000		1667	909	556	370	98
6000		1765	938	568	375	98
7000		1842	949	574	378	98
8000		1905	976	580	381	99
9000		1957	989	584	383	99
10000	5000	2000	1000	588	385	99
15000	6000	2143	1034	600	390	99
20000	6667	2222	1053	606	392	100
25000	7143	2273	1064	610	394	100
50000	8333	2381	1087	617	397	100
100000	9091	2439	1099	621	398	100
---	10000	2500	1111	625	400	100

Muestreo no probabilístico

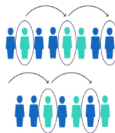
En clase se vieron cuatro tipos:

- **Conveniencia:** Las muestras o los individuos se seleccionan porque están convenientemente disponibles.
- **Respuesta Voluntaria:** La muestra se obtiene de los que voluntariamente participan o dieron datos, presenta sesgos ya que muchas veces solo los disconformes opinan.
- **Muestreo con propósito:** Busco solo los que cumplen ciertos requisitos.
- **Bola de nieve:** Encuestar y pedir a otros que compartan la encuesta.

Simple random sample



Systematic sample



Stratified sample



Cluster sample

