



Madrid Internet
of Things Institute
Make-build-learn

05 de marzo de 2020

Estadística para Data Science

Sesión 4: Estadística Descriptiva (II)

Jesús Hernando Corrochano



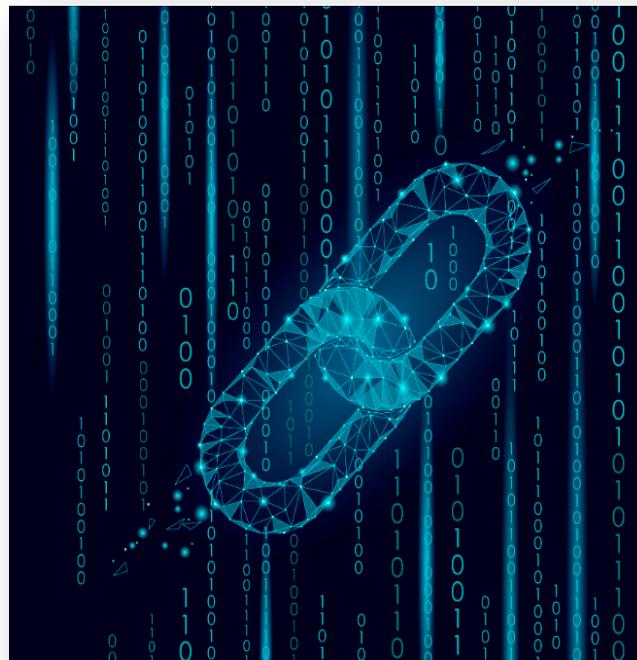
Estadística para Data Science

● Programa



	Sesión 1 13/2	Sesión 2 20/2	Sesión 3 27/2	Sesión 4 5/3	Sesión 5 12/3	Sesión 6 19/3	Sesión 7 26/3	Sesión 8 2/3
Introducción a la estadística								
Introducción a la combinatoria y la probabilidad								
Estadística descriptiva								
Regresión y correlación								
Estadística inferencial								
Probabilidad Total. Teorema de Bayes. Test A/B								

1. Tipificación
2. Geométrica
3. Hipergeométrica
4. Binomial Negativa
5. Poisson



Tipificación



Pensemos.....

Ejercicio:

En una clase de 20 alumnos y 15 alumnas:

La media y desviación típica del peso de los alumnos es de 64,7 kg y 4,2 respectivamente

La de las chicas es de 57,3kg y 5,7.

Luisa pesa 62kg y Carlos 71Kg

¿Cuál de ellos, dentro de su grupo, es mas pesado?

● Tipificación Ejercicio

Permite pasar de cualquier distribución normal en la que $\mu \neq 0$ o $\sigma \neq 1$ o las dos cosas, a la normal $N(0; 1)$

$$\text{tipificación: } N(\mu; \sigma) \rightarrow N(0; 1)$$

También existe el proceso opuesto

$$\text{destipificación: } N(0; 1) \rightarrow N(\mu; \sigma)$$

La tipificación y la destipificación permiten que cualquier normal pueda utilizar las tablas de la $N(0; 1)$

$$\text{tipificación: } z = \frac{x - \mu}{\sigma}$$

$$\text{destipificación: } x = \mu + z\sigma$$

● Teorema central del límite

No importa como se distribuya una población. Las medias de las muestras obtenidas de una población siempre se distribuyen según una distribución Normal.

Si la media de la población es μ y su desviación estándar es σ , entonces las medias de sus muestras se distribuirán con una normal

$$\bar{x} \rightarrow N(\mu ; \frac{\sigma}{\sqrt{n}})$$

n es el tamaño de la muestra o número de elementos que forman la muestra

Aunque lo anterior es cierto, solamente podemos utilizar la tabla de la Distribución Normal si:

- El tamaño de la muestra es grande (mayor o igual a 30) y
- Conocemos la desviación típica de la población

Si esto no es así hay que utilizar la tabla de la distribución t de Student

Estadística Descriptiva

● Geométrica

El proceso consta de un número no definido de pruebas o experimentos separados o separables.

El proceso concluirá cuando se obtenga por primera vez el resultado deseado (éxito).

- Cada prueba puede dar dos resultados mutuamente excluyentes : A y no A
- La probabilidad de obtener un resultado A en cada prueba es p y la de obtener un resultado no A es q siendo $(p + q = 1)$

La distribución **geométrica** es la probabilidad de la función masa:

$$P(X=k) = (1-p)^{k-1}p$$

La **distribución geométrica** puede ser definida como el número de intentos hasta conseguir el primer éxito.

Media: $1 - p / p$

Varianza: $1 - p / p^2$

● Geométrica / Ejemplos

Ejemplos:

Supongamos que Roberto está en una fiesta y comenzó a invitar chicas a bailar. Sea X el número de chicas que tuvo que invitar para encontrar una pareja de baile. Cuál es la probabilidad de que pueda bailar a la quinta solicitud.

Un barco proveniente de Japón llega a la aduana de Talavera con una carga de contenedores de productos eléctricos. Si los contenedores tienen una probabilidad del 5% de ser radioactivos ¿cuál es la probabilidad de que el 15 contenedor que se extraiga posea esta característica?

Hipergeométrica

Hipergeométrica: (discreta y sin reemplazo). Dada una población de N elementos de los cuales, d pertenecen a la categoría A y N-d a la B.

La distribución hipergeométrica mide la probabilidad de obtener x ($0 \leq x \leq d$) elementos de la categoría A en una muestra sin reemplazo de n elementos de la población original.

$$P(X = x) = \frac{\binom{d}{x} \binom{N-d}{n-x}}{\binom{N}{n}},$$

Media: $E[X] = \frac{nd}{N}$

Varianza: $Var[X] = npq \frac{N-n}{N-1}.$

La distribución hipergeométrica es aplicable a muestreos sin reemplazo y la binomial a muestreos con reemplazo. En situaciones en las que el número esperado de repeticiones en el muestreo es presumiblemente bajo, puede aproximarse la primera por la segunda. Esto es así cuando N es grande y el tamaño relativo de la muestra extraída, n/N , es pequeño

Hipergeométrica

Ejercicio:

En un grupo de 5 hombre y 5 mujeres ¿cuál es la probabilidad de formar un equipo de 5 con 3 mujeres y 2 hombres?

● Binomial Negativa

¿Cuál es la probabilidad de tener que repetir el experimento n veces para obtener de manera exacta k éxitos?

Así, dado un suceso de Bernoulli (Éxito/Fracaso). Cuando X representa el número de veces que se da F (ausencias, fallos, etc.) hasta que se produce r veces E, en una serie de repeticiones de la experiencia aleatoria en condiciones independientes, decimos que X sigue la distribución **Binomial negativa**.

Este modelo queda definido por dos parámetros p ($p = P(E)$) y r (el número de veces que debe producirse A para que detengamos la experiencia). La función de densidad viene dada por:

$$\binom{x-1}{k-1} p^k (1-p)^{x-k}$$

Cuando r = 1, tenemos exactamente el modelo geométrico.

$$BN(r,p) \rightarrow P(X=x)=P(Y_x=r-1)$$

Y_x es una $B(r+x-1, p)$

● Binomial Negativa / Ejemplo

Ejemplo:

Los empleados de una empresa que manufactura aislamientos están siendo examinados en busca de indicios de asbestos en sus pulmones. La empresa ha sido requerida para enviar tres empleados que tengan indicios positivos de asbestos a un centro médico para realizarles exámenes adicionales.

Si 40% de los empleados tienen indicios positivos de asbestos en sus pulmones, encuentre la probabilidad de que diez empleados deban ser examinados para hallar tres positivos.

Ejercicio:

Un componente electrónico tiene una probabilidad de 0,90 de pasar un control de calidad.

Calcular la probabilidad de que sean necesarios revisar 5 componentes para obtener que 3 pasen el control de calidad

Poisson

Distribución de probabilidad de variable aleatoria discreta, que nos mide la de que ocurra un determinado suceso un número de veces k en un determinado espacio de tiempo, longitud, área,...

$$P(x = k) = \frac{\mu^k \cdot e^{-\mu}}{k!} \quad e \approx 2.71828$$

Donde μ es el número medio de veces que ocurre el suceso en ese determinado espacio de tiempo, longitud, área,...

Concretamente, se especializa en la probabilidad de ocurrencia de sucesos con probabilidades muy pequeñas, o sucesos “raros”.

Se ha comprobado que estas asunciones son razonables en muchas situaciones diversas como:

- Ciertos estudios de patrones de tráfico del número de coches que pasan por un determinado punto en un intervalo de tiempo de longitud t .
- El número de llamadas de teléfono a una centralita en un intervalo de tiempo t .
- El número de fallos de un sistema complicado en un intervalo de tiempo t .
- El número de productos de un determinado tipo vendidos en un periodo de tiempo t .
- El número de suicidios en un intervalo de tiempo t .
- El número de bacterias en un determinado volumen t de líquido.
- El número de accidentes laborales en un intervalo de tiempo t .

Ejercicio:

En un hospital se está estudiando los nacimientos bebés varones. Se sabe que en una semana nacen una media de 7 varones.

Calcular:

- Probabilidad de que nazcan 3 varones en una semana
- Menos de tres varones a la semana

Resumen



● Resumen

- Distribución **uniforme**. Todos los valores tienen la misma probabilidad de ocurrir.
- Distribución **binomial**. Estudiamos la variable "numero de éxitos" (k) al realizar un cierto numero de ensayos (n) de Bernouilli. La binomial reparte la probabilidad entre los valores que puede tomar "k" que son cada uno de los valores comprendidos entre 0 y n.
- Distribución **geométrica**. Proporciona la probabilidad de obtener f fracasos antes de obtener el primer éxito al realizar repetidamente un ensayo de Bernouilli.
- Distribución **binomial negativa**. Otorga a cada numero de fracasos antes de obtener el k-esimo éxito al realizar repetidamente el ensayo de Bernouilli..
- Distribución **hipergeometrica**. Hay N elementos de los que B son de color blanco, y N-B no son de color blanco. Si se extraen de una vez (sin reemplazamiento) n elementos de entre esos N, ¿Cuál es la probabilidad, $p(x=b)$, de que haya b que sean de color blanco?
- Distribución de **Poisson**. La probabilidad se reparte entre valores de sucesos esporádicos. Por ejemplo, si definimos la variable: "número de rayos que caen en una cierta superficie durante una tormenta". La probabilidad se reparte entre la probabilidad de que caigan dos rayos, la probabilidad de uno, o de siete, etc.

Ejercicios Normal



Ejercicios Normal

Supongamos que las puntuaciones de prueba de un examen de admisión universitaria se ajustan a una distribución normal. Además, la puntuación media de la prueba es 72, y la desviación estándar es 15,2. ¿Cuál es el porcentaje de estudiantes que anotan 84 o más en el examen?

La temperatura durante septiembre está distribuida normalmente con media $18,7^{\circ}\text{C}$ y desviación standard 5°C .

- Calcule la probabilidad de que la temperatura durante septiembre esté por debajo de 21°C .
- Calcule la probabilidad de que la temperatura esté entre 18°C y 21°C

La media de los pesos de 5.000 estudiantes de un colegio es 70 kg y la desviación típica 3 kg. Suponiendo que los pesos se distribuyen normalmente, “hallar cuántos estudiantes” pesan menos de 60 kg

Se supone que los resultados de un examen siguen una distribución normal con media 78 y desviación típica 36. Se pide: ¿Cuál es la probabilidad de que una persona que se presenta el examen obtenga una calificación superior a 72?



Calle Rufino González 25
28037 Madrid
+34810527241
www.mioti.es

