UNIVERSIDAD PANAMERICANA
FACULTAD DE INGENIERÍA

Marzo – Junio 2020

MAESTRÍA EN CIENCIA DE DATOS
ESTADÍSTICA

24 DE MARZO DE 2020
AGENDA

- Encuadre del curso
- Temario y avisos en Moodle
- Plataforma para trabajo en línea
  - Asignación de licencias individuales de uso
    - Revisión de actividades en la plataforma
- Software para prácticas
    - Instalación de R y R Studio
    - Tutorial de programación en R

DISTRIBUCIONES PROBABILÍSTICAS

1. DISTRIBUCIÓN DE BERNOULLI
2. DISTRIBUCIÓN BINOMIAL
3. DISTRIBUCIÓN DE POISSON
4. DISTRIBUCIÓN HIPERGEOMÉTRICA

VARIABLES ALEATORIAS DISCRETAS

- Se seleccionan al azar
- Tienen valores enteros

DISTRIBUCIÓN DE BERNOULLI

- Un experimento binomial tiene las siguientes características:
  - Consiste en *n* intentos idénticos
  - Cada intento produce uno de dos posibles resultados. El resultado con valor 1 se denomina éxito y el otro, con valor 0, se llama fracaso
  - La probabilidad de éxito en un solo intento es igual a p. La probabilidad de fracaso es igual a 1-p
  - Los intentos son independientes
  - "x" es el número de éxitos observado durante los *n* intentos, para x=0,1,2,…,*n*.

$X \sim \text{Bernoulli}(p)$

La variable aleatoria X tiene una distribución de Bernoulli con parámetro p

Si $X=1$ entonces decimos que es un éxito

Si $X=0$ entonces decimos que es un fracaso

$p(0)= P(X=0)=1-p$

$p(1)= P(X=1)=p$

$p(x)=0$ para toda x diferente de 0 y 1

EJEMPLO 1

Una moneda tiene probabilidad 0.5 de caer en cara cuando se lanza.

Entonces

$X=1$ si la moneda cae en cara

$X=0$ si la moneda cae en águila

¿Cuál es la distribución de X?

La probabilidad de éxito ($X=1$) es:

$p(X=1)= 0.5$

$X \sim \text{Bernoulli}(0.5)$

EJEMPLO 2

- Se tiene un dado. La probabilidad de que caiga en 6 es 1/6
- Entonces X=1 si el dado cae en 6 y X= 0 si cae en cualquier otro valor
- $P(X=6)=1/6$
- $X \sim \text{Bernoulli}(1/6)$

EJEMPLO 3

Si 10% de los componentes fabricados en un proceso están defectuosos

Entonces $X=1$ si el componentes está defectuoso, $X=0$ si no está defectuoso

$P(X=1)=0.10$

$X \sim \text{Bernoulli}(0.10)$

MEDIA Y VARIANZA

$$\mu_x = (0)(1-p) + 1p = p$$

media o valor esperado

$$\sigma_x^2 = (0-p)^2(1-p) + (1-p)^2 p = p(1-p)$$

La varianza de una variable es una medida de dispersión definida como la esperanza del cuadrado de la desviación de dicha variable respecto a su media

…PARA EL EJEMPLO 3

X~Bernoulli(0.1)

p=0.1

$\mu_x = 0.1$

$\sigma_x^2 = 0.1(1\text{-}0.1) = 0.09$

. After scoring a touchdown, a football team may elect to attempt a two-point conversion, by running or passing the ball into the end zone. If successful, the team scores two points. For a certain football team, the probability that this play is successful is 0.40.

a. Let $X = 1$ if successful, $X = 0$ if not. Find the mean and variance of X.

b. If the conversion is successful, the team scores 2 points; if not the team scores 0 points. Let Y be the number of points scored. Does Y have a Bernoulli distribution? If so, find the success probability. If not, explain why not.

c. Find the mean and variance of Y.

(a) $\mu_X = 0.4$, $\sigma_X^2 = 0.24$

(b) No, a Bernoulli random variable has possible values 0 and 1. The possible values of Y are 0 and 2.

(c) $\mu_Y = 0.8$, $\sigma_Y^2 = 0.96$

a) $\mu_x = 0.4$, $\sigma_x^2 = 0.24 = 0.4(1-0.4)$

b) No, sólo se pueden tener valores de 0 ó 1.

c) $\mu_y = 0.4(2) = 0.8$

$= (0-0.8)^2(1-p) + (1-0.8)^2(p) = 0.96$

$= (x-\mu_x)^2(0.6) + (1.2)^2(0.4)$

When a certain glaze is applied to a ceramic surface, the probability is 5% that there will be discoloration, 20% that there will be a crack, and 23% that there will be either discoloration or a crack, or both. Let $X = 1$ if there is discoloration, and let $X = 0$ otherwise. Let $Y = 1$ if there is a crack, and let $Y = 0$ otherwise. Let $Z = 1$ if there is either discoloration or a crack, or both, and let $Z = 0$ otherwise.

a. Let p_X denote the success probability for X. Find p_X.

b. Let p_Y denote the success probability for Y. Find p_Y.

c. Let p_Z denote the success probability for Z. Find p_Z.

d. Is it possible for both X and Y to equal 1?

e. Does $p_Z = p_X + p_Y$?

f. Does $Z = X + Y$? Explain.

(a) 0.05

(b) 0.20

(c) 0.23

(d) Yes

(e) No

(f) No. If the surface has both discoloration and a crack, then $X = 1$, $Y = 1$, and $Z = 1$, but $X + Y = 2$.

DISTRIBUCIÓN BINOMIAL

Es una **distribución** de probabilidad discreta que cuenta el número de éxitos en una secuencia de n ensayos de Bernoulli independientes entre sí, con una probabilidad fija p de ocurrencia del éxito entre los ensayos.

Se dice que X tiene una distribución binomial con parámetros *n* y *p*

$$X \sim Bin(n,p)$$

X es una variable aleatoria discreta

$$X=0,1,2,\ldots,n$$

EJEMPLO 1

Una moneda se lanza 10 veces, X es el número de caras que aparecen. ¿Cuál es la distribución de X?

10 lanzamientos

p=0.5 para cada lanzamiento

$X \sim Bin(10,0.5)$

EJEMPLO 2

Un lote contiene varios miles de componentes. 10% de ellos están defectuosos. Se muestrean 7 componentes del lote.

X representa el número de componentes defectuosos de la muestra

¿Cuál es la distribución de X?

La muestra es menor que el 5% del lote, entonces el número de éxitos (componentes defectuosos) sigue una distribución binomial.

$X \sim \text{Bin}\ (7, 0.1)$

P$(X = x) =$ (NÚMERO DE ARREGLOS DE ÉXITOS EN n ENSAYOS) $p^x \ (1-p)^{x-1}$

Para el problema de la moneda

X~Bin(3,0.6)

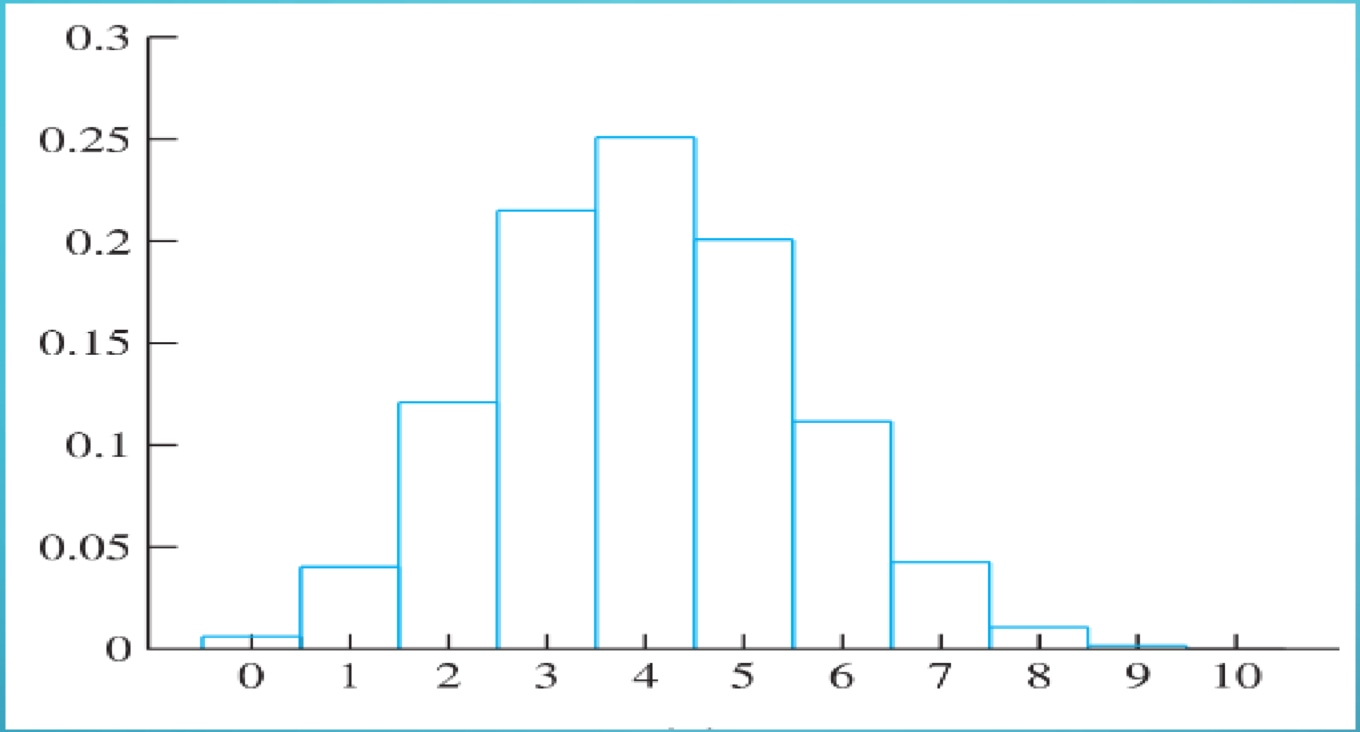
X=1 si cae en cara (H)

Se hacen 3 lanzamientos, con 3 combinaciones de 2 caras (H).

$$\begin{aligned} P(X=2) &= P(\text{HHT or HTH or THH}) \\ &= P(\text{HHT}) + P(\text{HTH}) + P(\text{THH}) \\ &= (0.6)^2(0.4)^1 + (0.6)^2(0.4)^1 + (0.6)^2(0.4)^1 \\ &= 3(0.6)^2(0.4)^1 \end{aligned}$$

If $X \sim \text{Bin}(n, p)$, the probability mass function of X is

$$p(x) = P(X = x) = \begin{cases} \dfrac{n!}{x!\,(n-x)!} p^x (1-p)^{n-x} & x = 0, 1, \dots, n \\ 0 & \text{otherwise} \end{cases}$$



¿Cuál es la función de probabilidad de la variable aleatoria X=5, si X~ Bin(10,0.4) ?

$$p(x) = \begin{cases} \dfrac{10!}{x!\,(10-x)!}(0.4)^x(0.6)^{10-x} & x = 0, 1, \ldots, 10 \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} P(X = 5) \quad = \quad p(5) \quad &= \quad \frac{10!}{5!\,(10-5)!}(0.4)^5(0.6)^{10-5} \\ &= \quad 0.2007 \end{aligned}$$

Si se lanza un dado 8 veces

¿Cuál es la probabilidad de que caiga la cara con seis punto no más de dos veces?

$$
\begin{aligned}
P(X \le 2) &= P(X = 0 \text{ or } X = 1 \text{ or } X = 2) \\
&= P(X = 0) + P(X = 1) + P(X = 2) \\
&= \frac{8!}{0!\,(8-0)!}\left(\frac{1}{6}\right)^{0}\left(\frac{5}{6}\right)^{8-0} + \frac{8!}{1!(8-1)!}\left(\frac{1}{6}\right)^{1}\left(\frac{5}{6}\right)^{8-1} \\
&\quad + \frac{8!}{2!\,(8-2)!}\left(\frac{1}{6}\right)^{2}\left(\frac{5}{6}\right)^{8-2} \\
&= 0.2326 + 0.3721 + 0.2605 \\
&= 0.8652
\end{aligned}
$$

MEDIA Y VARIANZA

If $X \sim \text{Bin}(n, p)$, then the mean and variance of X are given by

$$\mu_X = np$$

$$\sigma_X^2 = np(1 - p)$$

ESTIMACIÓN DE LOS VALORES DE PROBABILIDAD DE ÉXITO

$$\hat{p} = \frac{numero\ de\ \acute{E}xitos}{numero\ de\ intentos} = \frac{X}{n}$$

EJEMPLO 3

Se tienen 20 recipientes, 3 de ellos no están llenos
¿Cuál es la probabilidad de que la máquina no llene un recipiente?

$$\hat{p} = 3/20 = 0.15$$

Sesgo $\mu_{\hat{p}} - p$

$$\mu_{\hat{p}} = \mu_{X/n} = \frac{\mu_X}{n} = \frac{np}{n} = p$$

No hay sesgo

If $X \sim \text{Bin}(n, p)$, then the sample proportion $\hat{p} = X/n$ is used to estimate the success probability p.

- $\hat{p}$ is unbiased.
- The uncertainty in $\hat{p}$ is

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

In practice, when computing $\sigma_{\hat{p}}$, we substitute $\hat{p}$ for p, since p is unknown.

TAMAÑO DE MUESTRA PARA AJUSTAR A LA INCERTIDUMBRE DESEADA

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

$n = 40$, X=4

$$\hat{p} = \frac{4}{40}$$

$$\sigma_x^2 = \sqrt{\frac{0.10 * 0.90}{40}} = 0.047$$

Si se busca $\sigma_x^2 = 0.02$

$$0.02 = \sqrt{\frac{0.1 * 0.9}{n}}$$

n= 225

Propagación del error

$$\sigma_u = abs\left(\frac{dU}{dX}\right)\sigma_x$$

Se tienen 100 llantas y de ellas 7 tienen fallas.
Estimar la probabilidad de que NINGUNA de las cuatro llantas
Seleccionadas para un auto tenga fallas. Encuentra la incertidumbre.

$$\hat{p} = \frac{93}{100}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{(0.93)(0.07)}{100}} = 0.0255$$

Incertidumbre para $\hat{p}$

La probabilidad para las cuatro llantas es p^4

$$\begin{aligned} \sigma_{\hat{p}} &\approx \left| \frac{d}{d\hat{p}} \hat{p}^4 \right| \sigma_{\hat{p}} \\ &= 4\hat{p}^3 \sigma_{\hat{p}} \\ &= 4(0.93)^3 (0.0255) \\ &= 0.082 \end{aligned}$$

$$p^4 \cong \hat{p}^4 = 0.93^4$$

Incertidumbre para $\hat{p}^4$