Variable Aleatoria

- Una variable aleatoria es una variable cuyo valor está sujeto a variaciones que dependen de la aleatoriedad.
- Deben tomar valores numéricos, que dependen del resultado del experimento
- Puede ser continua o discreta

Propiedades de esperanzas, sumatorias, varianzas y covarianzas

a y b son constantes

X y Y son Variables Aleatorias

- $\mu_X = E[X]$
- $\sigma^2_{X,X} = VAR[X] = E[(X-\mu)^2] = E[X^2] - \mu^2$
- $\sigma^2_{X,Y} = COV[X,Y] = E[(X-\mu_x)(Y-\mu_Y)] = E[XY] - \mu_x\mu_Y$
- $COV[X,X] = VAR[X]$

Esperanza

- $E[b] = b$
- $E[aX+b] = aE[X] + b$
- Si X y Y son VA independientes:
    - $E[XY] = E[X]E[Y]$
    - $E[X/Y] \neq E[X]/E[Y]$
- Si X y Y son VA NO independientes:
    - $COV(X,Y) = E[XY] - E[X]E[Y]$
    - $E[XY] = COV(X,Y) + E[X]E[Y]$

Ejercicio:

$$Y = 2X + 3$$
$$E[Y] = E[2X+3] = 2E[X] + 3$$

Varianza

- $VAR[X] = E[(X-\mu)^2] = E[X^2] - \mu^2$
- $VAR[b] = 0$
- $VAR[aX+b] = a^2\, VAR(X)$

- Si X y Y son VA independientes:
  - $VAR[X+Y] = VAR(X) + VAR(Y)$
  - $VAR[X-Y] = VAR(X) + VAR(Y)$
  - $VAR[aX+bY] = a^2VAR(X) + b^2VAR(Y)$
  - $VAR[aX-bY] = a^2VAR(X) + b^2VAR(Y)$
- Si X y Y son VA no independientes:
  - $VAR[X+Y] = VAR(X) + VAR(Y) + 2COV(X,Y)$
  - $VAR[X-Y] = VAR(X) + VAR(Y) - 2COV(X,Y)$
  - $VAR[aX+bY] = a^2VAR(X) + b^2VAR(Y) + 2ab\ COV(X,Y)$
  - $VAR[aX-bY] = a^2VAR(X) + b^2VAR(Y) - 2ab\ COV(X,Y)$

Ejercicio:

$$Y = 2X + 3$$
$$VAR[Y] = VAR[2X+3] = 4VAR[X]$$

$Y = 2X_1 + 3X_2 + 1$: X_1 y X_2 son independientes

$$VAR[Y] = VAR[2X_1 + 3X_2 + 1] = VAR[2X_1 + 3X_2]$$
$$VAR[Y] = 4VAR[X_1] + 9VAR[X_2]$$

<u>Covarianza</u>

- $COV[X,Y] = E[(X-\mu_x)(Y-\mu_Y)] = E[XY] - \mu_x\mu_Y$
- Si X y Y son VA independientes:
  - $COV[X,Y] = 0$
    - $COV[X,Y] = E[XY] - \mu_x\mu_Y = E[X]E[Y] - \mu_x\mu_Y = 0$
- $COV[a,X] = 0$
- $Cov(aX,bY) = a\ b\ Cov(X,Y)$
- $COV[a+bX, c+dY] = bd\ COV(X,Y)$
- $Cov(X+Y,\ Z) = Cov(X,Z) +\ Cov(Y,Z)$
- $Cov(aX+bY,\ cZ) = a\ c\ Cov(X,Z) + b\ c\ Cov(Y,Z)$

$$Y = 2X + 3$$
$$COV[Y,X] = COV[2X+3, X] = COV[2X, X] = 2COV(X,X) = 2VAR(X)$$

$Y = 2X_1 + 3X_2 + 1$: X_1 y X_2 no son independientes

$$COV[Y, X_1] = COV[2X_1 + 3X_2 + 1, X_1]$$
$$COV[Y, X_1] = COV[2X_1, X_1] + COV[3X_2, X_1] + COV[1, X_1]$$
$$COV[Y, X_1] = 2VAR[X_1] + 3COV[X_2, X_1]$$

Sumatoria

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$$
$$\sum_{i=1}^{n} k = nk$$
$$\sum_{i=1}^{n} kx_i = k\sum_{i=1}^{n} x_i$$
$$\sum_{i=1}^{n}(a + bx_i) = \sum_{i=1}^{n}(a) + \sum_{i=1}^{n}(bx_i) = na + b\sum_{i=1}^{n}(x_i)$$
$$\sum_{i=1}^{n}(x_i + y_i) = \sum_{i=1}^{n}(x_i) + \sum_{i=1}^{n}(y_i)$$

Ejercicios

- $\sum_{i=1}^{N}(X_i - \bar{X}) = 0$

Operador de Producto

$$\prod_{i=1}^{n} x_i = x_1 * x_2 * \ldots * x_n$$
$$log\left(\prod_{i=1}^{n} x_i\right) = \log(x_1 * x_2 * \ldots * x_n) = \log(x_1) + \log(x_2) + \cdots + \log(x_n) = \sum_{i=1}^{n} \log(x_i)$$

Ejercicio

- $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$

$$Var(Z) = E[(Z - E(Z))^2] \quad \text{donde} \quad \mu = E(X)$$

$$Var(X+Y) = E\left[\left((X+Y) - E(X+Y)\right)^2\right]$$

$$Var(X+Y) = E[(X+Y-E(X)-E(Y))^2]$$

$$Var(X+Y) = E\left[\left(X+Y-\mu_x-\mu_y\right)^2\right]$$

$$Var(X+Y) = E\left[X^2 + 2XY - 2X\mu_x - 2X\mu_y + Y^2 - 2Y\mu_x - 2Y\mu_y + \mu_x^2 + 2\mu_x\mu_y + \mu_y^2\right]$$

$$Var(X+Y) = E\left[(X^2 - 2X\mu_x + \mu_x^2) + \left(Y^2 - 2Y\mu_y + \mu_y^2\right) + \left(2XY - 2X\mu_y - 2Y\mu_x + 2\mu_x\mu_y\right)\right]$$

Nota. Dado que:

$(X-\mu_x)^2 = X^2 - 2X\mu_x + \mu_x^2$

$\left(Y-\mu_y\right)^2 = Y^2 - 2Y\mu_y + \mu_y^2$

$(2)(X-\mu_x)\left(Y-\mu_y\right) = 2XY - 2X\mu_y - 2Y\mu_x + 2\mu_x\mu_y$

$$Var(X+Y) = E\left[(X-\mu_x)^2 + \left(Y-\mu_y\right)^2 + (2)(X-\mu_x)\left(Y-\mu_y\right)\right]$$

$$Var(X+Y) = E[(X-\mu_x)^2] + E\left[\left(Y-\mu_y\right)^2\right] + 2 * E\left[(X-\mu_x)\left(Y-\mu_y\right)\right]$$

$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$$

Variables Discretas

$$\sum_{i=1}^{n} f(x_i) = 1$$

$$f(x) = PMF = P(X = x)$$

$$E[X] = \mu = \sum_{i=1}^{n} [x_i * f(x_i)]$$

$VAR[X] = E[(X - \mu)^2] = E[X^2] - \mu^2$

$$VAR[X] = \sigma^2 = \sum_{i=1}^{n} [(x_i - \mu)^2 * f(x_i)]$$

Continuas

$$\int_{-\infty}^{+\infty} f(x)\, dx = 1$$

$$f(x) = PDF \neq Probabilidad$$

$$\int_{2}^{4} f(x)\, dx = Prob\ de\ que\ X\ este\ entre\ 2\ y\ 4$$

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\, dt = Prob\ de\ que\ X\ sea\ menor\ que\ x$$

$$E[X] = \mu = \int_{-\infty}^{+\infty} x\, f(x)\, dx$$

$$VAR[X] = \sigma^2 = \int_{-\infty}^{+\infty} (x - \mu)^2\, f(x)\, dx$$

Bernoulli

$$f(x) = p^x (1 - p)^{1-x} \qquad x \,\epsilon\, \{0,1\}$$

$$E[X] = \mu = \sum_{i=1}^{n} [x_i * f(x_i)] = 1 * p + 0 * (1 - p) = p$$

$$VAR[X] = \sigma^2 = \sum_{i=1}^{n} [(x_i - \mu)^2 * f(x_i)] = (1 - p)^2 * p + (0 - p)^2 * (1 - p)$$

$$VAR[X] = (1 - 2p + p^2) * p + p^2(1 - p)$$

$$VAR[X] = p - 2p^2 + p^3 + p^2 - p^3$$

$$VAR[X] = p - p^2 = p * (1 - p)$$

- Si tiro una moneda 1 vez, ¿cuál es la probabilidad de que sea cara?

Binomial

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x \, \epsilon \, \{0,1\}$$

$$\binom{n}{x} = \frac{n!}{x!\,(n-x)!}$$

$$E[X] = n * p$$

$$VAR[X] = n * p * (1 - p)$$

- Si tiro una moneda 5 veces (n=5), ¿cuál es la probabilidad de obtener 2 caras?

Poisson

- En promedio llegan 20 personas a la cola de Starbucks cada horas. Cuál es la probabilidad de que lleguen 18 personas en un intervalo de 1 hora?

$$f(x) = \frac{e^{-\lambda} * \lambda^x}{x!}$$

$$E[X] = \lambda$$

$$VAR[X] = \lambda$$

Más de Variables Continuas

- Si $f(x) = 1;\ \ x \, \epsilon \, \left(0, \sqrt{2}\right)$ and $f(x) = 0$ otherwise.

$$\int_{-\infty}^{0} f(x)\, dx = \int_{-\infty}^{0} 0\, dx = 0$$

$$\int_{0}^{\sqrt{2}} f(x)\, dx = \int_{0}^{\sqrt{2}} x\, dx = \left(\frac{x^2}{2}\right) - \left(\frac{x^2}{2}\right)\Bigg|_{0}^{\sqrt{2}} = \left(\frac{\sqrt{2}^2}{2}\right) - \left(\frac{0^2}{2}\right) = 1 - 0 = 1$$

$$\int_{\sqrt{2}}^{+\infty} f(x)\, dx = \int_{0}^{\sqrt{2}} 0\, dx = 0$$

$$\int_{-\infty}^{+\infty} f(x)\, dx = 1$$

En variables continuas $P(X = x) = 0$

$$\int_1^1 x\,dx = \left(\frac{x^2}{2}\right) - \left(\frac{x^2}{2}\right)\Big|_1^1 = \left(\frac{1^2}{2}\right) - \left(\frac{1^2}{2}\right) = .5 - .5 = 0$$

Exponencial

$$f(x) = \begin{cases} \lambda * e^{-\lambda x} & para\ x \geq 0 \\ 0 & de\ otro\ modo \end{cases}$$

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & para\ x \geq 0 \\ 0 & de\ otro\ modo \end{cases}$$

$$E[X] = \frac{1}{\lambda}$$

$$VAR[X] = \frac{1}{\lambda^2}$$

- En promedio llegan 20 personas a la cola de Starbucks cada hora. Cuál es la probabilidad de que la siguiente persona llegue en los próximos 6 minutos?

Normal

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$E[X] = \mu$$

$$VAR[X] = \sigma^2$$

$$X \sim N(\mu, \sigma^2)$$

Normal Estándar

$$\mathrm{Z} = \frac{X-\mu}{\sigma}\ ; \quad X \sim N(\mu, \sigma^2)$$

$$E[Z] = E\left[\frac{X-\mu}{\sigma}\right] = \frac{1}{\sigma}E[X-\mu] = \frac{1}{\sigma}(E[X] - E[\mu]) = \frac{1}{\sigma}(\mu - \mu) = 0$$

$$VAR[Z] = VAR\left[\frac{X-\mu}{\sigma}\right] = \frac{1}{\sigma^2}VAR[X-\mu] = \frac{1}{\sigma^2}VAR[X] = \frac{\sigma^2}{\sigma^2} = 1$$

Relación de Distribución Z y Chi^2

$$Z^2 = \chi_1^2$$

Donde 1 es el grado de libertad de la distribución Chi^2.

Distribución Chi^2

$$Z_1^2 + Z_2^2 + \; .... \; + \; Z_n^2 = \chi_n^2$$

Definición de t-student

$$t_n = \frac{Z}{\sqrt{\frac{\chi_n^2}{n}}}$$

Donde n son los grados de libertad de la muestra.

Definición de F:

$$F_{(n_1,n_2)} = \frac{\frac{\chi_{n_1}^2}{n_1}}{\frac{\chi_{n_2}^2}{n_2}}$$

Dónde: n_1 son los grados de libertad de la muestra 1.

n_2 son los grados de libertad de la muestra 2.

Estimadores Puntuales

La media muestral $\bar{x}$ es el estimador puntual de la media poblacional μ

La varianza muestral s^2 es el estimador puntual de la varianza poblacional σ^2

En regresiones la media muestral $\hat{\beta}$ es el estimador puntual del parámetro β

$$X \sim N(\mu, \sigma^2)$$

$$\bar{x} = \frac{\sum x_i}{n}$$

$$\bar{x} \sim N \text{ si } X \sim N \text{ o si } n > 30$$

$$E[\bar{x}] = \mu$$

$$VAR[\bar{x}] = \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

$$SD[\bar{x}] = \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$$\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$$

Ejercicio

Se sabe que la media poblacional en un examen de Estadística es de 70 y que la varianza es de 324. Se toma una muestra aleatoria a 81 alumnos y se registra su calificación. Asume que las calificaciones se distribuyen de forma normal.

- ¿Cuál es la probabilidad de que al escoger aleatoriamente una de las calificaciones de mi muestra esta se ubique entre 85 y 90?
  18.86%
- Calcula el error estándar de la media muestral.
  2
- ¿Cuál es la probabilidad de que la media muestral de los 81 alumnos se ubique entre 85 y 90?
  30.854%

Ejemplo:
En un estudio de 100 estudiantes de la Anáhuac, se preguntaron las calificaciones de ética y se obtuvo $\bar{x} = 82$. Se sabe que $\sigma = 20$.

Concluimos que $\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$
Además, sabemos que $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{100}} = 2$

Nosotros queremos construir un intervalo de confianza del 95%.

- IC= Intervalo de Confianza
- α =Nivel de significancia = 1 – IC

$$IC = \bar{x} \pm Z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}$$

$$Intervalo = 82 \pm 1.96 * 2 = 78.08\ y\ 85.92$$