Statistics Review

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

Variable Aleatoria Discreta

- Número de personas que se forman en una fila en 1 hora
- Número de águilas que se obtienen al lanzar una moneda 5 veces.

Variable Aleatoria Continua

- Tiempo entre dos llamadas de teléfono
- Temperatura.
- Estatura
- Peso

Distribuciones de Probabilidad

- La distribución de probabilidad de una variable aleatoria describe cómo se distribuyen las probabilidades entre los valores de la variable aleatoria.
- La distribución de probabilidad está definida por una función de probabilidad, denotada por f(x)

Distribución de Probabilidad Discreta

- Histograma
- Una vez que se conoce la distribución de probabilidad, es relativamente fácil determinar la probabilidad de diversos eventos

$$f(x) \geq 0$$

$$\sum f(x) = 1$$

Valor Esperado:

$$E(x) = \mu = \sum f(x) * x$$

Varianza:

$$Var(x) = \sigma^2 = \sum f(x) * (x - \mu)^2$$

Ejemplos de Distribuciones Discretas

- Binomial:
  - n etapas idénticas e independientes
  - Cada ensayo tiene 2 posibles resultados, éxito o fracaso
  - Probabilidad de éxito para cada ensayo es p, de fracaso es 1-p
  - x denota al número de éxitos que queremos obtener

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}; \qquad \binom{n}{x} = \frac{n!}{x!\,(n-x)!}$$

$$f(x) = probabilidad\ de\ x\ \text{é}xitos\ en\ n\ ens\ ayos$$

  - Tiras una moneda 5 veces

- Poisson:
  - Número de veces que un evento sucede en un tiempo determinado.

$$f(x) = \frac{\mu^x e^{-\mu}}{x!} = probabilidad\ de\ x\ ocurrencias\ en\ un\ intervalo$$

Distribución de Probabilidad Continua

- **Normal (campana de gauss)**
- **t-student**
- **χ^2**
- **F**
- **Log-Normal**
- **Logística**
- **Exponencial**

$$f(x) \geq 0$$

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

$f(x)$ No es una probabilidad, es una densidad.

$$\int_5^6 f(x)\,dx = Probabilidad\ de\ observar\ valores\ de\ x\ entre\ 5\ y\ 6 = P(5 \leq X \leq 6)$$

$$\int_{-\infty}^{x} f(t)\,dt = F(x) = P(X \leq x) = Funcion\ de\ Probabilidad\ Acumulada$$

$$F(6) - F(5) = \int_5^6 f(x)\,dx = Probabilidad\ de\ observar\ valores\ de\ x\ entre\ 5\ y\ 6$$

Distribución de Probabilidad Normal

- La distribución de probabilidad más usada para describir variables aleatorias continuas es la distribución de probabilidad normal
- Peso, estatura, puntuaciones de exámenes, etc.

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

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

- Dos parámetros: μ y σ^2.
- Media=Moda=Mediana
- Simétrica (coeficiente de asimetría = 0) y Mesocrática (curtosis = 3)
- Probabilidades:
  - Área debajo de media es 50%
  - 68.3% de los valores de una variable aleatoria normal se encuentran más o menos una desviación estándar de la media
  - 95.4% a dos desviaciones estándar
  - 99.7% a tres desviaciones estándar.

Distribución de Probabilidad Normal Estándar

- Una variable aleatoria que tiene una distribución normal con media cero y desviación estándar de uno.
- Se puede estandarizar cualquier variable aleatoria con distribución normal y las áreas entre la media y la desviación estándar no cambian.

$$z = \frac{x - \mu}{\sigma}$$

Ejercicios

$X \sim N(5,4)$, cuál es la probabilidad de que X sea menor o igual a 7?

$$z = \frac{7-5}{\sqrt{4}} = \frac{2}{2} = 1$$

84.13%

$X \sim N(5,4)$, cuál es la probabilidad de que X sea mayor o igual a 7?

15.87%

$X \sim N(5,4)$, cuál es la probabilidad de que X sea se ubique entre 3 y 7?

$$z_1 = \frac{3-5}{\sqrt{4}} = \frac{-2}{2} = -1$$

$$z_2 = \frac{7-5}{\sqrt{4}} = \frac{2}{2} = 1$$

68.27%

Estimación Puntual

Parámetros Poblacionales

Media Poblacional = $\mu = E(x)$

Varianza Poblacional = $\sigma^2 = E[(x-\mu)^2] = E[x^2] - 2\mu E[x] + \mu^2 = E[x^2] - \mu^2$

Desviación Estándar Poblacional = σ

Estadísticos Muestrales

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

Varianza Muestral $s^2 = \frac{\sum(x_i-\bar{x})^2}{n-1}$

Desviación Estándar Muestral $s = \frac{\sum x_i}{n}$

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

Si seleccionar una muestra aleatoria es un experimento, la media muestral $\bar{x}$ es un resultado de ese experimento. Por lo tanto, cualquier $\bar{x}$ puede ser considerado como una variable aleatoria. Por lo tanto, como cualquier otra variable aleatoria $\bar{x}$ también cuenta con:

- Valor esperado
- Desviación estándar → Error Estandar
- Distribución de probablidad.

Esto permite hacer inferencias acerca del valor de la media poblacional μ, mismo que desconocemos

$$E(\bar{x}) = E\left[\frac{\sum x_i}{n}\right] = \frac{1}{n}\sum E[x_i] = \frac{1}{n}\sum \mu = \frac{1}{n} * (\mu * n) = \mu$$

$$var(\bar{x}) = var\left[\frac{\sum x_i}{n}\right] = \frac{1}{n^2}\sum var[x_i] = \frac{1}{n^2}\sum \sigma^2 = \frac{1}{n} * (\sigma^2 * n) = \frac{\sigma^2}{n}$$

$$sd(\bar{x}) = \frac{\sigma}{\sqrt{n}}$$

Distribución de $\bar{x}$

- Cuando la población tiene distribución normal, la distribución de $\bar{x}$ esta también distribuida normalmente.
- Cuando la población no tiene distribución normal podemos usar el teorema del límite central y así deducir que $\bar{x}$ también se distribuye de forma normal.
  - **Teorema del Límite Central**: Cuando se seleccionan muestras aleatorias simples de tamaño n de una población, la distribución muestral de la media muestral puede aproximarse mediante una distribución normal a medida que el tamaño de la muestra se hace grande.
  - **Para que el Teorema del Límite Central sea válida se necesita un tamaño de muestra igual o mayor a 30**.

Distribución de $\bar{p}$

$$\bar{p} = \frac{x}{n}; \quad x\ es\ el\ numero\ de\ elementos\ de\ la\ muestra\ que\ cumplen\ con\ el\ criterio$$

$$E[\bar{p}] = p$$

$$sd(\bar{p}) = \sqrt{\frac{p(1-p)}{n}}$$

- La distribución muestral de se aproxima mediante una distribución normal siempre que $np \geq 5$ y $n(1-p) \geq 5$.

<u>**Ejercicio**</u>

Se sabe que la media poblacional de altura es de 1.75 y que la desviación estándar es de 0.15. Se toma una muestra de la altura de 225 personas y se obtiene la altura de cada una.

a) Calcula el error estándar de $\bar{x}$.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.15}{\sqrt{225}} = 0.01$$

b) ¿Cuál es la probabilidad de que la media muestral de mi encuesta sea 1.7 o menos que 1.7?

0.00003%

c) ¿Cuál es la probabilidad de que la media muestral de mi encuesta sea igual a 1.7 o menor que 1.8?

99.99997%

d) ¿Cuál es la probabilidad de que la media muestral de mi encuesta se ubique entre 1.70 y 1.80?

$$P(1.70 < \bar{x} < 1.80) = P(\bar{x} < 1.80) - P(\bar{x} < 1.70)$$

99.99994%