# Econometría I (EC402)-II semestre de 2013
# Clase #3 - Revisión conceptos de inferencia estadística

Andrés M. Castaño

Ingeniería Comercial
Universidad Católica del Norte
Agosto 12 de 2013

# Revisión: inferencia estadística

- **Inferencia:** estudia la relación entre muestra y población, permite obtener conclusiones sobre población a partir de una muestra aleatoria.
- **Población:** todos los posibles resultados de un fenómeno.
- **Muestra:** subconjunto de esa población.
- Para inferir, la muestra debe ser aleatoria.
- La inferencia es la estimación de los parámetros y el contraste de hipótesis.

# Variables aleatorias

- Variable cuyo valor está determinado por el resultado de un experimento al azar.
- Variable aleatoria discreta: toma un número finito de valores (lanzamiento de dos datos)
- Variable aleatoria continua: toma cualquier valor dentro de un intervalo de valores (Estatura, temperatura)

## Función de Densidad de Probabilidad para una VA discreta (FDP)

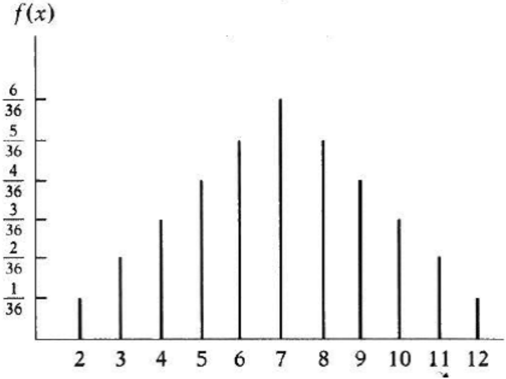
- Sea X una VA discreta que toma valores diferentes $x_1, x_2, ..., x_n$ entonces, la función:

$$f(x) = P(X = x_i) \forall i = 1, 2, ...., n \tag{1}$$

# Ejemplo VA discreta

En el lanzamiento de dos dados, considere la VA X, como la suma de los números que aparecen en dos dados:

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| X= | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| f(x)= | $\frac{1}{36}$ | $\frac{2}{36}$ | $\frac{3}{36}$ | $\frac{4}{36}$ | $\frac{5}{36}$ | $\frac{6}{36}$ | $\frac{5}{36}$ | $\frac{4}{36}$ | $\frac{3}{36}$ | $\frac{2}{36}$ | $\frac{1}{36}$ |



# VA continua

## Función de Densidad de Probabilidad para una VA continua (FDP)

- Sea X una VA continua. Entonces, se dice que $f(x)$ es la FDP de X si se satisfacen las siguientes condiciones:

$$f(x) \geq 0$$

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

$$\int_{a}^{b} f(x)dx = P(a \leq x \leq b)$$

Donde $f(x)dx$ es la probabilidad asociada a un pequeño intervalo de una variable continua y $P(a \leq x \leq b)$ indica la probabilidad de que X se encuentre en el intervalo que va desde a hasta b.

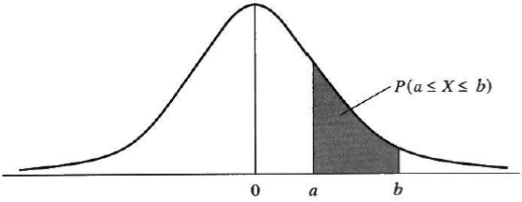
# Ejemplo VA continua

Considere la siguiente función de densidad:

$$f(x) = \frac{1}{9}x^2$$

$$0 \leq x \leq 3$$

$$\int_0^3 \frac{1}{9}x^2 dx = \frac{1}{27}x^3|_0^3 = 1$$



# Características de las distribuciones de probabilidad

## Valor esperado E(X)

- Valor esperado VA discreta:

$$E(X) = \sum_{x} x f(x)$$

$$E(X) = 2\frac{1}{36} + 3\frac{2}{36} + 4\frac{3}{36} + ..... + 12\frac{1}{36}$$

- Valor esperado VA continua:

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

$$E(X) = \int_{0}^{3} x(\frac{x^2}{9}) dx$$

$$= \frac{1}{9}(\frac{x^4}{4})|_0^3 = \frac{9}{4}$$

# Características de las distribuciones de probabilidad

## Varianza

- Mide la dispersión de los valores de X alrededor del valor esperado:

$$var(X) = \sigma_x^2 = E(X - \mu)^2$$

- Caso si X es discreta:

$$var(X) = \sum_x (x - \mu)^2 f(x)$$

- Caso si X es continua:

$$var(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

# Ejemplo

Caso VA discreta:

| x | -2 | 1 | 2 |
|---|---|---|---|
| f(x) | $\frac{5}{8}$ | $\frac{1}{8}$ | $\frac{2}{8}$ |

$$E(X) = -2\frac{5}{8} + 1\frac{1}{8} + 2\frac{2}{8}$$

$$= -\frac{5}{8}$$

$$E(X^2) = 4\frac{5}{8} + 1\frac{1}{8} + 4\frac{2}{8}$$

$$= \frac{29}{8}$$

$$var(X) = E(X^2) - (E(X))^2$$

$$var(X) = \frac{29}{8} - (-\frac{5}{8})^2 = \frac{207}{64} = 3{,}23$$

# Ejemplo

Caso VA continua:

$$E(X^2) = \int_0^3 x^2(\frac{x^2}{9})dx$$

$$= \frac{1}{9}(\frac{x^5}{5})|_0^3 = \frac{243}{45}$$

$$var(X) = E(X^2) - (E(X))^2$$

$$var(X) = \frac{243}{45} - (\frac{9}{4})^2 = 0{,}34$$

# Estimación

- Procedimiento por medio del cual obtenemos información sobre el valor de un parámetro poblacional a partir de información muestral.
- Se trata de obtener el valor del parámetro poblacional (desconocido) a partir de datos muestrales, a partir de una muestra aleatoria simple (m.a.s.) de tamaño n.

## Problema de estimación

- Estimación puntual o de un punto; y
- Estimación por intervalos o de un intervalo.

**Estimador:** un estimador de un parámetro poblacional es una variable aleatoria que depende de la información de la muestra y cuyas realizaciones proporcionan aproximaciones al valor desconocido del parámetro poblacional.

**Estimación:** es una realización específica de esta variable aleatoria.

# Estimación puntual: ejemplo

- La media muestral $\bar{X}$ es una variable aleatoria y la llamamos estadístico (genéricamente). Un estadístico es una medida descriptiva muestral y es una variable aleatoria porque su valor depende de la muestra que extraigamos.
- Si con $\bar{X}$ queremos estimar $\mu_x$ (parametro poblacional), llamamos a $\bar{X}$ estimador es una v.a.
- En una muestra, su valor particular se llama estimación.

## Claridad importante

- **Estimador puntual:** un estimador puntual de un parámetro poblacional es una función de la muestra que da como resultado un único valor.
- **Estimación puntual:** es una realización (un valor particular obtenido de una m.a.s.) del estimador puntual.

# Estimación puntual: ejemplo

Para simplificar: no importa cómo llamemos a la función de datos muestrales, si estadístico, estimador o estimador puntual. Solo tendremos en cuenta que son variables aleatorias, cuyo valor depende de la muestra y un valor particular (es decir, una realización) se llamará estimación puntual o estimación de un parámetro poblacional dado.

# Propiedades de los buenos estimadores

## ¿Cómo elegir un buen estimador para obtener una estimación puntual lo más "fiable" posible del parámetro poblacional?

- sentido común.
- propiedades del estimador.

## Linealidad

Un estimador es lineal si es una función lineal de las observaciones muestrales.

## Insesgadez

Un estimador es insesgado si la media de la distribución muestral es igual al parámetro correspondiente. $\theta$ : parámetro y $\hat{\theta}$ : estimador.

$$E(\hat{\theta}) = \theta$$

# Propiedades de los buenos estimadores

## Varianza Mínima

Un estimador de $\theta$ es de varianza mínima si su varianza es menor que la de cualquier otro estimador de $\theta$.

## Eficiencia

Dados dos o más estimadores insesgados, un estimador es eficiente cuando tiene la varianza menor. Ejemplo:

$$E(\bar{X}) = \mu_x$$

$$E(med.muestral) = \mu_x$$

$$Var(\bar{X}) = \frac{\sigma^2}{n}$$

$$var(med.muestral) = \frac{\pi\sigma^2}{2n} \simeq 1{,}57 var(\bar{X})$$

# Propiedades de los buenos estimadores

Tengo cuidado con los conceptos de:

- Distribución leptocurtica (Más apuntada y con colas más anchas que la normal)
- Distribución mesocurtica (la distribución normal es mesocúrtica, curtosis de 3)
- Distribución platicurtica, (menos apuntada y con colas menos anchas que la normal)

Una de las formulas es:

$$G_2 = \frac{\mu_4}{\sigma^4} - 3$$

## Consistencia

Un estimador es consistente si, a medida que aumenta el tamaño de la muestra, las estimaciones se aproximan al valor del parámetro. ** Si un estadístico NO es consistente no sirve tomar muestras más grandes para obtener estimaciones puntuales más precisas.

$$lim\hat{\beta}_i = \beta_i$$

# Propiedades de los buenos estimadores

## Suficiencia

Un estimador es suficiente si utiliza toda la información disponible en la muestra. La suficiencia es una condición necesaria para la eficiencia.

## Estimador lineal insesgado óptimo

- Es un estimador lineal, insesgado y tiene la varianza mínima entre todos los estimadores lineales insesgados del parámetro.
- Aclaración: Para comparar dos estimadores sesgados utilizamos el concepto de error cuadrático medio (ECM) de un estimador.

$$ECM(\hat{\beta}) = var(\hat{\beta}) + [sesgo(\hat{\beta})]^2$$

# Estimación por intervalos..

El uso de una estimación puntual tiene inconvenientes:

- Si el estimador no es consistente, por más que tomemos una muestra "grande", no estaremos seguros de la confiabilidad de esa estimación puntual.
- Tampoco tenemos una medida del error posible en la estimación puntual. (Hemos visto que para distintas muestras el valor posible cambia...el error posible en la estimación puntual se mide en términos de la variabilidad de la distribución de estimador).
- Un estimador por intervalos de un parámetro poblacional es una regla (basada en información muestral) para determinar un rango, o un intervalo aleatorio, en el cual posiblemente se encuentre dicho parámetro.

# Estimación por intervalos..

- Ahora, en vez de tener un estimador que es una variable aleatoria, tendremos un intervalo aleatorio pues los límites dependerán de la muestra...
- Para la estimación por intervalos se consideran, en esencia: 1. un estimador puntual, una función del estimador puntual y del parametro a estimar, y la distribución de probabilidades asociada a esta función.
- El intervalo se especifíca para un nivel de confianza dado respecto de la precisión de la estimación, por eso se denomina INTERVALO DE CONFIANZA.

# Estimación por intervalos..

- Hasta ahora hemos dicho que los estimadores por intervalos contienen posiblemente. al verdadero, aunque desconocido, parámetro poblacional. Para formalizar, es necesario hablar de esas posibilidades en términos de probabilidades.

- Para la estimación por intervalos se consideran, en esencia: 1. un estimador puntual, una función del estimador puntual y del parametro a estimar, y la distribución de probabilidades asociada a esta función.

- Sea $\theta$ el parámetro a estimar. El estimador por intervalos es un intervalo aleatorio tal que la probabilidad de que el parámetro esté entre LI y LS es $1 - \alpha$. LI y LS son v.a.

$$P(LI < \theta < LS) = 1 - \alpha$$

# Estimación por intervalos..

- Una vez que extraemos la muestra, calculamos los valores particulares de LI y LS (li y ls) y tendremos un intervalo de confianza:

$$li < \theta < ls$$

.

- Es decir, el parámetro está entre li y ls con un nivel de confianza de $(1 - \alpha)100\,\%$. (Interprete $\alpha\,\%$).

- **Atención:** en una muestra particular, NO podemos hablar de probabilidad; el intervalo contiene o no al verdadero parámetro. Por ello hablamos de nivel de confianza.

# Historia del análisis de regresión.....