Outline

Inferencia Estadística

- Hacer inferencia estadística consiste en aprender sobre el comportamiento de una población dada en base a una muestra de la población.
  - Ej: Cuál es el retorno promedio de un año adicional de educación en la población chilena? Usamos muestras sobre años de educación e ingresos para una muestra de la población, pues es costoso conseguir esa data a nivel poblacional.
- Primero definimos la **población de interés**; luego especificamos un **modelo** que nos permita predecir el comportamiento de la población – típicamente una **distribución de probabilidad** (pdf, fn de media condicional, etc), la cual a su vez depende **parametros poblacionales desconocidos** (e.g., el retorno a la educación es el parámetro de interés cuando estudiamos cómo la educación afecta a los salarios)
- Usamos la muestra de estudio para construir "**estadísticos**" (indicadores resúmen de la muestra), los cuales son utilizados para inferir el comportamiento de los parámetros poblacionales. Cada estadístico tiene una contraparte poblacional, es decir, utilizamos los estadísticos como estimadores de los parámetros poblacionales.

Obteniendo una muestra aleatoria de la población

- Sea Y_i una v.a. que representa a una poblacióin con una pdf $f(y;\theta)$, donde θ es el parámetro poblacional de interés.
- Dado que θ es desconocido, nuestro objetivo es obtener una muestra que nos permita aprender sobre el valor de θ.
- Muestreo Aleatorio: sea $Y_1, Y_2, ...Y_n$ v.a. independientes con una pdf $f(y;\theta)$; luego el vector de realizaciones aleatorias $\{y_1, y_2, ...y_n\}$es una *muestra aleatoria* de la población representada por $f(y;\theta)$. (usamos letras minúsculas para denotar realizaciones de las variables aleatorias)
  1. Diremos entonces que $y_1, y_2, ...y_n$son v.a. porque si bien cada observación es obtenida de la misma población, ex-ante no es posible determiar cuál va a ser la realización y_i correspondiente a cada Y_i.
  2. También diremos que $y_1, y_2, ...y_n$ son v.a. independientes e idénticamente distribuidas **(i.i.d)** de la población $f(y;\theta)$ ya que (i) las realizaciones provienen de una muestra aleatoria y por lo tanto son independientes la una de la otra; y (ii) todas ellas son obtenidas de la misma población $f(y;\theta)$ y por tanto son *Idénticas*

Obteniendo una muestra aleatoria de la población

- Más aún:
  - $\{y_1, y_2, ...y_n\}$ es una muestra aleatoria pues los resultados que se obtienen con cada muestra serán aleatorios (ej: la muestra 1 va a diferir de la muestra 2 en relación a $\overline{y}$)
  - Las muestras aleatorias se pueden obtener de cualquier función de distribución, sin embargo la decisión de qué función elegir dependerá de la naturaleza del parámetro poblacional de interés (ej: si la fenomeno a estudiar es binario, tipicamente se utiliza una distribución Bernoulli)
    - ⋆ Nótese que la muestra aleatoria es univariada si y_i es cada observación contiene una única variable, es bivariada si cada observación contiene información sobre dos variables, y es multivariada si cada observación alberga información sobre multiples variable.

"Estadísticos"

- Lo primero que hacemos con una muestra es construir diferentes "estadísticos" que nos permitan caracterizar los datos
  - Un "estadístico" es un valor que es calculado como una función de la muestra aleatoria
    - Promedio muestral: $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$
    - Desviación Estándar: $s_y = \sqrt{\frac{\sum(y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{(\sum_{i=1}^{n} y_i^2) - n \cdot \bar{y}^2}{n-1}}$
    - Covarianza: $s_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{(\sum_{i=1}^{n} x_i \cdot y_i) - n \cdot \bar{x} \cdot \bar{y}}{n-1}$, Correlación: $r_{xy} = \frac{s_{xy}}{s_x \cdot s_y}$
  - Otros estadísticos útiles: mediana, moda, oblicuidad, kurtosis, rango intercuartil, etc. También podemos usar gráficos de dispersión, histogramas, etc, todo lo cual nos ayuda a caracterizar el comportamiento de los datos.
  - Dado que cada estadístico tiene una contraparte poblacional, decimos entonces que los estadísticos juegan el rol de "estimadores" de los parámetros poblacionales.

Estadístico: Ejemplo

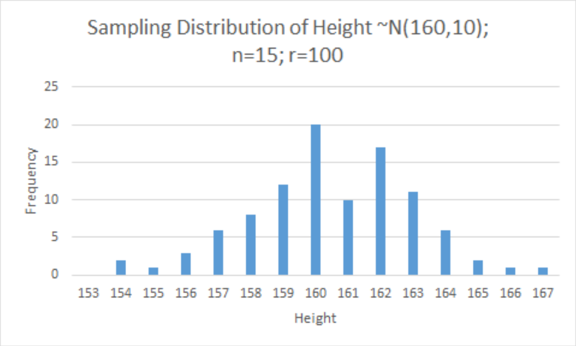
- *Ej 1*: Estudiemos la relación entre altura (Y) y peso (X) en una población ficticia.
  - Usando Excel, obtenemos una muestra aleatoria de 15 v.a. (observaciones) de una población que se distribuye $N \sim (\mu, \sigma^2)$. La altura (en cm.) se distribuye $N \sim (160, 10)$ y el peso (en kg.) se distribuye $N \sim (70, 10)$
  - Calcule $\bar{y}$, $\bar{x}$, s_y, s_x, s_{xy}, y r_{xy}.
- Cada "estadístico" será un "estimador" del parámetro poblacional de interés, y diremos que el valor del estadístico es un "estimado".
  - Importante: el "estimado" aproxima el parámetro poblacional, pero no es exactamente igual. De hecho, típicamente $\bar{y} \neq \mu_y$ y $s_y \neq \sigma_y$
    - ★ Pero cómo podemos saber "cuán bueno" es un estimado en cuanto a su nivel de aproximación al parámetro poblacional?

Distribución muestral

- Si obtuvieramos otra muestra aleatoria, obtendríamos un estimado diferente ya que las observaciones de la nueva muestra son v.a.
  - *Ej 2*: Obtengamos una nueva muestra aleatoria de 15 observaciones de altura y peso en Excel, bajo las mismas condiciones de distribución. Son los estimados de la nueva muestra aleatoria iguales que aquellos derivados de la primera muestra?
- Cada estadístico es una función de la muestra aleatoria obtenida, y por tanto los estimados son v.a. En efecto, diremos que cada estadístico tiene una distribución de probabilidad que llamamos *distribución muestral*.
  - La *distribución muestral* es la distribución de probabilidad de un "estadístico".
  - Sio pudieramos obtener multuples muestras aleatorias y para cada una de ellas computar el estadístico de interés, entonces podríamos obtener su distribución muestral: simplemente recolectamos los estimados muestrales y vemos como se distribuyen.
  - *Ej 3*: Usando la distribución poblacional de altura (Y), genere 100 muestras aleatorias (es decir, $r = 1, \ldots, 100$), cada una de tamaño $n = 15$. Calcule $\overline{y}_r$ para cada muestra r y luego haga un histograma de $\overline{y}_r$.

Distribución muestral

Figure: Distribución muestral de Altura.



Distribución muestral

- La *distribución muestral* se utiliza para hacer inferencia respecto del parametro poblacional de interés.
- Sin embargo, en realidad nunca observamos la distribución muestral (podríamos derivarla si conocemos la distribución poblacional de Y_i, pero esta típicamente es difícil de saber con exactitud por lo tanto lo que nos queda es elegir un modelo de esa distribución).
  - Dado que recoger datos es costoso, sólo tenemos una chance para obtener una muestra aleatoria de la población, y por tanto sólo una chance para aprender del parámetro poblacional de interés.
- Pero si no podemos observar la distribución muestral, cómo podemos evaluar que tan bueno es nuestro "estimado" como un proxy del parámetro poblacional de interés?
  - Necesitamos criterios para evaluar la bondad de los estimadores. Para ello estudiamos algunas carcaterísticas de las distribuciones muestrales de los estimadores y en base a ellas decidimos la calidad de un estimador.

Propiedades de los estimadores

- Definición: un estimador $\hat{\theta}$ de θ, es una regla que asigna a cada posible resultado de la muestra un valor aproximado de θ.
- En general, un estimador punto $\hat{\theta}$ de un parámetro θ puede ser expresado como $\hat{\theta} = h(Y_1, ...Y_n)$ y una *estimación puntual* de θ como $\hat{\theta} = h(y_1, ...y_n)$, donde h es la regla.
  - la regla es especificada antes de obtener la mnuestra aleatoria $\Rightarrow$ mientras cada muestra arroja una estimado distinto, la regla es siempre la misma. Ej: el promedio muestral, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, es un estimador de μ (el promedio poblacional) independiente de la muestra obtenida.
- Hay dos tipos de propiedas para evaluar la calidad de un estimador: propiedades de muestras finitas, y propiedades de muestras grandes. Por ahora nos concentraremos en las siguientes:
  - Propiedades en muestras finitas: insesgamiento, eficiencia, error cuadrático medio.
  - Propiedades en muestras grandes: Consistencia, eficiencia.

Outline

Insesgamiento

- Definición: $\hat{\theta}$ es un estimador insesgado de θ si el promedio de su distribución muestral es igual a θ, es decir, si $E[\hat{\theta}] = \theta$
- Esto es equivalente a decir que $E[\hat{\theta} - \theta] = Bias[\hat{\theta}|\theta] = 0$
- Ejemplo:
  - $E[\overline{y}] = E[\frac{1}{n}\sum_{i=1}^{n} y_i] = \frac{1}{n}\sum_{i=1}^{n} E[y_i] = \frac{1}{n} \cdot n\,\mu_y = \mu_y \Rightarrow E[\overline{y}]$ es un estimador insesgado de μ_y
  - $E[s_y^2] = \dfrac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2 = \frac{\sigma_y^2}{n-1}$. Por lo tanto, $\frac{s_y^2}{n-1}$ es un estimador insesgado de σ^2.
  - Ojo, algunos estimadores insesgados no son muy buenos estimadores de los parametros poblacionales. Supongamos un estimador de μ_y que es simplemente $\hat{\theta} = y_1$, i.e., la primera observación de una muestra aleatoria. Si hacemos la matemática, tenemos que $E[y_1] = E[y_i] = \mu_y \Rightarrow y_1$ es un estimador insesgado de μ_y. Sin embargo, estamos dejando de usar $n-1$ observaciones para estimar el promedio poblacional, un uso ineficiente de la información $\Rightarrow$ necesitamos un criterio adicional

Varianza muestral (error) de un estimador

- $\bar{y}$ nunca va a ser exactamente igual a $\mu \Rightarrow$ nos enfocamos en la distribución de $\bar{y}$ (su *varianza muestral*): mientras menor es la varianza de la distribución muestral es mas probable que la media muestral se acerque a μ, y por tanto mejor es la calidad del estimador.

$$
\begin{aligned}
Var(\bar{y}) &= E\left[(\bar{y} - E[\bar{y}])^2\right] = E\left[\left(\frac{1}{n}\sum_{i=1}^{n} y_i - E[\bar{y}]\right)^2\right] \leftarrow E[\bar{y}] = E[y_i] \\
&= \frac{1}{n^2} E\left[\sum_{i=1}^{n}(y_i - E[y_i])^2 + \sum_{i=1}^{n}\sum_{i \neq j}(y_i - E[y_i])(y_j - E[y_j])\right] \\
&= \frac{1}{n^2}\left\{\sum_{i=1}^{n} Var[y_i] + \sum_{i=1}^{n}\sum_{j \neq i} Cov(y_i, y_j)\right\} \leftarrow Cov(y_i, y_j) = 0 \\
&= \frac{\sigma_y^2}{n}
\end{aligned}
$$

Varianza muestral (error) de un estimador

Ejercicio: Simulando muestras

1. Generar muchas (digamos $r = 1 \dots R$) muestras aleatorias de tamaño n de una distribución $N \sim (\mu = 10, \sigma^2 = 25)$.
2. Calcula $\overline{y}_r$ para cada muestra r.
3. Plotear un histograma de $\overline{y}_r$.
4. Repote el ejercicio para distintos valores de n (10,30,100,300).

* Sugiero hacerlo en sus casas, ya sea en Excel, Stata, o R.

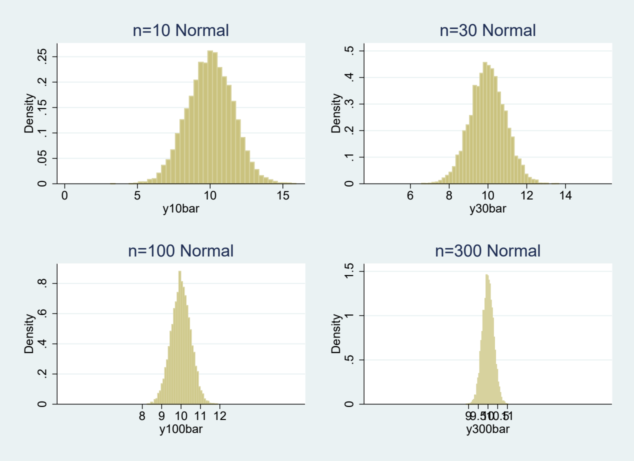
Varianza muestral (error) de un estimador

Código en STATA

```
**** nsample = 10
local nsample = 10 local
nstr = "n=10"
local mu = 10
local sigma = 5
global DIST Normal
** generate n draws
forvalues i = 1(1)‘nsample’{
gen y10_‘i’ = rnormal(‘mu’,‘sigma’)
}
** calculate row average
egen y10bar = rowmean(y10_1-y10_‘nsample’)
** histogram
hist y10bar, title(‘nstr’ $DIST) xsc(range(4 16))
graph save hist_‘nsample’.gph, replace
```

Resultados de la simulación

Figure: Distribución de la media muestral de una población distribuida normalmente para diferentes tamaños muestrales.



Calculando el Error Estándar de un estimador

- El *Error Estándar de un estimador* $\hat{\theta}$ viene dado por su desviación estándar, $\sqrt{Var[\hat{\theta}]}$. En el ejemplo,

$$S.E\,(\bar{y}) = \sqrt{Var[\bar{y}]} = \sqrt{\frac{\sigma_y^2}{n}} = \sigma_y/\sqrt{n}$$

- Nótese que el error estándar depende de un parámetro desconocido, σ_y . Por lo tanto tenemos que estimar σ_y antes de calcular el estándar error.
- Hemos demostrado que un estimador insesgado de σ_y^2 viene dado por:

$$S_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2$$

- Luego,

$$S.E\,(\bar{y}) = \sqrt{S_y^2/n}$$

Eficiencia

- Eficiencia: un estimador $\hat{\theta}$ es eficiente si, dentro de todos los estimadores insesgados, $\hat{\theta}$ es aquel con la menor varianza.
  - Ej: sabemos que $\hat{\theta} = y_1$ es insesgado ya que $E[y_1] = E[y_i] = \mu_y$. Sin embargo, $Var[y_1] = Var[y_i] = \sigma_y > Var[\overline{y}] = \frac{\sigma_y^2}{n}$ para cualquier $n > 1$. Luego, $\bar{y}$ eficiente relativo a $y_1$para estimar μ.

Error Cuadrático Medio

- Una manera de comparar la bondad de estimadores que no necesariamente son insesgados es computando el error cuadrático medio (MSE) de los estimadores.
$MSE[\hat{\theta}] = E[(\hat{\theta} - \theta)^2] = Var[\hat{\theta}^2] + (Bias[\hat{\theta}])^2$ (Demuestre usando nikita nipone$-E[\hat{\theta}] + E[\hat{\theta}]$)
- Tiene sentido usar el MSE para comparar dos estimadores insesgados?

Intervalos de Confianza

- Reportar $\bar{y}$ y $S.E(\bar{y})$ provee información valiosa para evaluar la precisión de un estimador. Sin embargo, no nos permite decir nada respecto de cuan cercano está $\bar{y}$ del parámetro poblacional μ_y.
- Dado que no conocemos μ_y, no podemos saber cuan cerca esta $\bar{y}$ de μ_y para una muestra en particular. No obstante, podemos decir algo respecto a la probabilidad de que μ_y se encuentre en un determinado rango, y con ello establecer cuan cerca está (probabilísticamente) $\bar{y}$ de μ_y.
- La lógica detrás de un *intervalo de confianza* radica en usar la muestra de datos para construir un intervalo $[inferior_{\mu_y}, superior_{\mu_y}]$, tal que con cierto *nivel de confianza (probabilidad)* sabemos que dicho intervalo contiene a μ_y .

Intervalo de confianza para poblaciones distribuidas Normalmente con σ conocida

- Supongamos que $Y \sim N(\mu_y, \sigma_y^2)$ y μ_y e σ_y^2 son parámetros conocidos. Luego, conocemos la media muestral $\bar{y} \sim N(\mu_y, \frac{\sigma_y^2}{n})$. Con esto podemos estandarizar $\bar{y}$ y obtener $Z = \frac{\bar{y}-\mu_y}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\bar{y}-\mu_y}{SE(\bar{y})}$.
- Notar que $\mu_y = \bar{y} - Z \cdot SE(\bar{y})$. Luego, el intervalo de confianza de la media poblacional μ_y se puede definir como $C(\mathbf{\bar{y}}) = \bar{y} \pm z_{1-\frac{\alpha}{2}} \cdot SE(\bar{y})$, con $z_{1-\frac{\alpha}{2}}$ el valor crítico z para el cual $P(z \leq z_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$.
- Notar también que $C(\mathbf{\bar{y}})$ es una variable aleatoria que varía de muestra en muestra. En cambio, μ_y es una constante. De hecho, vamos a decir que "la probabilidad de que $C(\mathbf{\bar{y}})$ contenga a μ_y es α". (Ojo: decir que "la probabilidad de que μ_y esté en el intervalo $C(\mathbf{\bar{y}})$ es α%" es incorrecto ya que μ_y es una constante!)
- α es "el nivel de confianza", i.e., por cada 100 muestras aleatorias que obtenemos, $C(\mathbf{\bar{y}})$ va a contener μ_y en un α% de ellas$\rightarrow$ $C(\mathbf{\bar{y}})$ nos entrega un rango de valores de μ_y con probabilidad "α", y con ello podemos inferir cuan cerca está $\bar{y}$ de μ_y, o cuán preciso es $\bar{y}$.

Ejemplo

- Sea $\{Y_1, ..., Y_n\}$ una muestra aleatoria obtenida de una población que se distribuye $\sim N(\mu_y, 1)$.
- Luego, sabemos que $\bar{y} \sim N(\mu_y, \sqrt{\frac{1}{n}})$ y que $Z = \frac{\bar{y}-\mu_y}{\sqrt{\frac{1}{n}}}$. Dado que Z proviene de una normal estándar, entonces sabemos que para $\alpha = 0.05$ y $z_{1-\frac{\alpha}{2}} = 1.96$, $P(-1.96 < Z < 1.96) = 0.95$
- Del mismo modo, $P(-1.96 < \frac{\bar{y}-\mu_y}{\sqrt{\frac{1}{n}}} < 1.96) = P(\bar{y} - 1.96 \cdot \sqrt{\frac{1}{n}} < \mu_y < \bar{y} + 1.96 \cdot \sqrt{\frac{1}{n}}) = 0.95$
- Supongamos que $n = 100$ y $\bar{y} = 20$. Luego, $P(20 - 1.96 \cdot \sqrt{\frac{1}{100}} < \mu_y < 20 + 1.96 \cdot \sqrt{\frac{1}{100}}) = P(19.804 < \mu_y < 20.196) = 0.95$

Intervalo de confianza para poblaciones distribuidas Normalmente con σ desconocida

- Cuando la σ es conocida, podemos computar un intervalo de confianza al 95% para un Y distribuido normalmente como $(\bar{y} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}, \bar{y} + 1.96 \cdot \frac{\sigma}{\sqrt{n}})$
- En cambio, si σ es desconocido, entonces primero debemos estimar σ y luego proceder vía una *t-student.*
- En efecto, supongamos que $Y \sim N(\mu_y, \sigma_y^2)$ donde μ_y y σ_y son parámetros desconocidos. Usamos $s_y = \sqrt{\frac{\sum(y_i - \bar{y})^2}{n-1}}$ como estimador de σ_y para construir $C(\bar{\mathbf{y}})$.
- Si llegamos e insertamos s_y, obtendríamos $C(\bar{\mathbf{y}}) = \bar{y} \pm z_{1-\frac{\alpha}{2}} \cdot s_y$, pero esto no va a contener μ_y con probabilidad .95 ya que la constante σ_y ha sido reemplazada por la variable aleatoria s_y
- Sin embargo, sabemos que $T = \frac{\bar{y} - \mu_y}{\frac{s_y}{\sqrt{n}}} \sim t_{n-1}$. Luego $C(\bar{\mathbf{y}}) = \bar{y} \pm t_{(n-1, 1-\frac{\alpha}{2})} \cdot \frac{s_y}{\sqrt{n}}$, con $t_{(n-1, 1-\frac{\alpha}{2})}$ el valor crítico t para el cual $P(t \leq t_{(n-1, 1-\frac{\alpha}{2})}) = 1 - \frac{\alpha}{2}$.

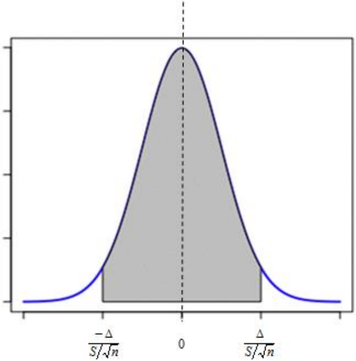


Figure: t-student distribution

Ejemplo

- Sea $\{Y_1, ..., Y_n\}$ una muestra aleatoria obtenida de una población que se distribuye $\sim N(\mu_y, \sigma_y^2)$ con μ_y e σ_y desconocidas.
- Suponga $n = 30$, $\bar{y} = 20$ y $s_y = 4$
- Luego, $T = \frac{20-\mu_y}{\frac{4}{\sqrt{30}}}$. Dado que T proviene de una *t distribution*, entonces sabemos que para un $\alpha = 0.05$ tenemos que $t_{(29,1-\frac{\alpha}{2})} = 2.045$, y por lo tanto $P(-2.045 < T < 2.045) = 0.95$
- Del mismo modo, $P(20 - 2.045 \cdot \frac{4}{\sqrt{30}} < \mu_y < 20 + 2.045 \cdot \frac{4}{\sqrt{30}}) = P(18.507 < \mu_y < 21.494) = 0.95$

Cuan confiable es s_y como estimador de σ_y?

- Recuerda que si $z \sim N(0,1)$ y $x \sim \chi^2_{[n]}$, entonces $t = \frac{z}{\sqrt{\frac{x}{n}}} \sim t_{[n]}$.

  Notar que $t = \underbrace{\left(\frac{\overline{y} - \mu_y}{\sigma_y / \sqrt{n}}\right)}_{\sim N(0,1)} \Big/ \underbrace{\left(\frac{(n-1)S_y^2}{\sigma_y^2} \Big/ (n-1)\right)^{\frac{1}{2}}}_{\sim\sqrt{\chi^2_{n-1} / (n-1)}}$

- Luego, sabemos que $\frac{(n-1)S_y^2}{\sigma_y^2} \sim \chi^2_{n-1}$, y entonces podemos usar esto para definir un intervalo de confianza para σ_y^2.
- Definamos $\chi^2_{n-1_{(\alpha/2)}}$ y $\chi^2_{n-1_{(1-\alpha/2)}}$ como los valores críticos una distribución chi-cuadrado, cada uno de los cuales corresponde a las probabilidades $\alpha/2$ y $1-\alpha/2$ en cada extremo de la cola (ver tabla chi-cuadrado). Luego, $P\left(\chi^2_{n-1_{(1-\alpha/2)}} \leq \frac{(n-1)S_y^2}{\sigma_y^2} \leq \chi^2_{n-1_{(\alpha/2)}}\right) = 1-\alpha$. En efecto, el IC$1-\alpha$ de σ_y^2 es $\left[\frac{(n-1)S_y^2}{\chi^2_{n-1_{(\alpha/2)}}}, \frac{(n-1)S_y^2}{\chi^2_{n-1_{(1-\alpha/2)}}}\right]$. Para el IC $1-\alpha$ de σ, simplemente tomamos la SQRT de cada valor del IC.

Test de Normalidad

- Para derivar un intervalo de confianza asumimos que la distribución poblacional desde donde proviene la muestra es Normal. Es este un supuesto apropiado? Cómo saber si μ_y proviene de una normal?

qq-plots para testear Normalidad

- Dada una muestra aleatoria, rankea las observaciones de menor a mayor.
- Calcula la función de distribución acumuladad: para cada observación, cuál es la probabilidad de que se observe un valor menor o igual a ese.
- Graficar los cuantiles de la distribución acumulada contra los cuantiles que se obtendrían si esta muestra proviniera de una distribución Normal Estándar.
- Si los puntos se ubican aproximadamente alineados en una línea recta, entonces significa que la función de distribución poblacional de la cual proviene la muestra se parece a una Normal.

Ej: Usamos un qq-plot para testear normalidad en el precio de los autos. Se comportan los precios de los autos como una distribución Normal?

Test de Normalidad

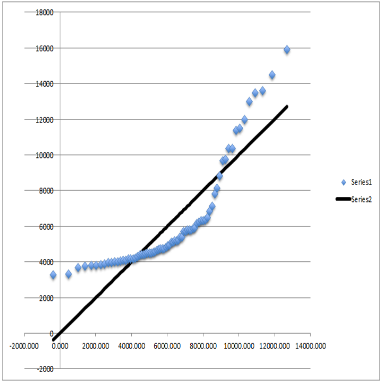


Figure: qq-plot

Qué pasa si fallamos el test de Normalidad?...

- Para calcular el intervalo de confianza de μ, necesitamos caracterizar la distribución del estadístico t, $(\overline{X}_n - \mu)/(S/\sqrt{n})$.
- Como vimos, construir una distribución exacta de este estadístico requiere de supuestos fuertes respecto a la distribución poblacional ($f(x) \sim \mathrm{Normal}$).
- Afortunadamente, existen aproximaciones útiles para la distribución de un estadístico-t, y dichas aproximaciones son independientes de la distribución poblacional. Sin embargo, son sólo validas para un n suficientemente grande. Estas aproximaciones son denominadas *distribuciones asintóticas* y pueden ser utilizadas para construir intervalos de confianza asintóticos.
- Comenzamos estudiando las propiedades en muestras grandes de los estimadores, las cuales son fundamentales para comprender la base de los intervalos de confianza asintóticos.

Outline

Propiedades de Estimadores en Muestras Grandes

- En estadística, evaluar cuál es el mejor estimador depende del comportamiento de la distribución muestral de cada estimador.
- Para algunos pocos estimadores, conocemos el comportamiento de sus distribuciones muestrales independiente del tamaño muestral (ej: conocemos $\bar{y}$ para $y \sim N(\mu_y, \sigma_y^2)$, con μ_y y σ_y^2 conocidos).
- Sin embargo, en la mayoría de los casos sólo es posible aproximar el *comportamiento limite* de la distribución muestral del estimador, y dicha aproximación será basada en el análisis de cómo la distribución muestral del estimador cambia conforme el tamaño muestral crece.
  - Ej: la distribución de la media muestral tiene a una normal a medida que aumenta el tamaño muestral, y esto es independiente de la distribución a partir de la cual provienen las observaciones.
- Cualquier proceso de estimación mejora conforme aumenta el tamaño muestral pues mas información siempre es mejor que menos.
- En efecto, las propiedades en muestras grandes sirven para evaluar la "calidad" de un estimador conforme aumenta el tamaño muestral (ej: podemos descartar estimador insesgados que no mejoran conforme aumenta n, como $\hat{\theta} = y_1$.

Convergencia en Probabilidad

Definition

Sea X_n una secuencia de v.a. indexadas por el tamaño muestral n. La v.a. X_n *converge en probabilidad* a una constante c si, para todo $\epsilon > 0$,

$$\lim_{n\to\infty} \Pr\{\,|X_n - c\,| > \epsilon\} = 0$$

Para ello decimos: $X_n \underset{p}{\rightarrow} c$, o alternativamente, $plim\, X_n = c$.

- Intuición: "Converge en probabilidad" significa que conforme n crece al infinito, toda la masa de la distribución empieza a concentrarse en un punto cercano a c.
- Ej: Supongamos que X_n toma dos valores, cero y n, con probabilidades $1 - \frac{1}{n}$ y $\frac{1}{n}$, respectivamente. Luego, $plim\, X_n = 0 \cdot plim\,(1 - \frac{1}{n}) + n \cdot plim\,(\frac{1}{n}) = 0 \cdot 1 + n \cdot 0 = 0$. En efecto, decimos que X_n converge en probabilidad a cero.

Consistencia

Definition

Si X_n tiene media μ_n y varianza σ_n^2 tal que los límites ordinarios de μ_n y σ_n^2 son c y 0, respectivamente, entonces X_n converge en probabilidad a c y por tanto $plim\, X_n = c$. Luego, $\hat{\theta}_n$ *es un estimador consistente* del parámetro poblacional θ (en que θ es un escalar, no una v.a.), si $\hat{\theta}_n$ converge en probabilidad a θ, i.e., $plim\, \hat{\theta}_n = \theta$.

- Ej: sea $\hat{\theta} = y_1$ una secuencia de v.a. y un estimador de μ_y. Cuán cerca está y_1 de μ_y para una muestra significativamente grande de tamaño n ? $E[y_1] = \mu$ y $Var[y_1] = \sigma_y^2$. Dado que $plim\, Var[y_1] = \sigma_y^2 \neq 0$, entonces $plim\, y_1 \neq \mu$, i.e., y_1 es un estimador *inconsistente* de μ_y. (Notar que $\hat{\theta}$ puede ser inconsistente por dos razones: si converge a una constante distinta de θ; o si no converge a una constante (sino a una v.a.).)
- NOTA: insesgamiento concierne a estimadores en muestras finitas, mientras que consistencia se refiere al comportamiento de la distribución muestral de un estimador conforme la muestra crece.

Consistencia

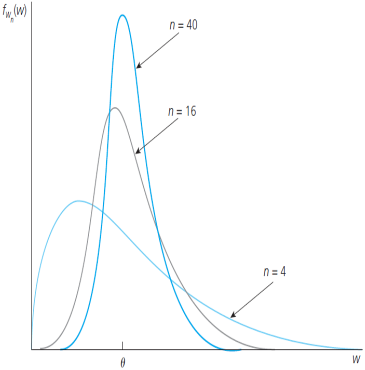


Figure: Consistency

Nota: cuando los estimadores son inconsistentes, todavía podemos definir sus límites de probabilidad, y eso será importante para saber cuan lejos se encuentra el límite de probabilidad de la distribución muestral, respecto de θ

Relación entre un estimador insesgado y uno consistente

Theorem

Si X_n es un estimador insesgado de θ (i.e. $E(X_n) = \theta$), y $Var(X_n) \to 0$ cuando $n \to \infty$, entonces plim $\overline{X}_n = \theta$, y en efecto X_n es un estimador consistente de θ.
Tarea: de un ejemplo de:
(i) Un estimador insesgado y consistente
(ii) Un estimador insesgado pero inconsistente
(iii) Un estimador sesgado e inconsistente
Demuestre en cada caso.

Ley de los Grandes Numeros (LLN)

Theorem

Ley de los Grandes Numeros: Sea $\{X_i\}_{i=1}^n$ *una muestra i.i.d, con* $E(X_i) = \mu < \infty$ *y* $Var(X_i) = \sigma^2 < \infty$*. Luego,* $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ *es un estimador consistente de* μ*.*

Proof.

Anteriormente demostramos que $Var(\overline{X}_n) = \frac{\sigma^2}{n}$, la cual converge a cero conforme $n \to \infty$. También demostramos que $\bar{X}_n$ es un estimador insesgado de μ. Por lo tanto, la proposición anterior implica que $plim\, \overline{X}_n = \mu$, y en efecto $\bar{X}_n$ es un estimador consistente de μ. □

La media muestral es un "buen" estimador de la media poblacional ya que conforme nuestra muestra crece al infinito, $\bar{X}$ se converge a μ.

Slutsky (1)

- Si $X_n$es una v.a. tal que:
  - $plim\, X_n = c$
  - $g(X_n)$ es una función continua y diferenciable
  - $g(X_n)$ no depende de n
- Entonces:

1. $plim\, g(X_n) = g(plim\, X_n)$
   - Ej: $plim\, X_n = \mu$ y $g(X_n) = \frac{1}{x_n}$, luego $plim\, g(X_n) = plim\, \frac{1}{X_n} = \frac{1}{plim\, X_n} = \frac{1}{\mu}$
   - Notese que esto no funciona para esperanzas....$E[X_n] = \mu$. No obstante, $E[\frac{1}{X_n}] =?$
2. Si X_ny $Y_n$son dos v.a. con límites de probabilidad θ y μ (i.e.,$plim\, X_n = \theta$ y $plim\, Y_n = \mu$), entonces:
   1. $plim\, (X_n \pm Y_n) = \theta \pm \mu$
   2. $plim\, (X_n \cdot Y_n) = \theta \cdot \mu$
   3. $plim\, (\frac{X_n}{Y_n}) = \frac{\theta}{\mu}$
   4. En general, $plim\, (g(X_n, Y_n)) = g(\theta, \mu)$

Convergencia en Distribución

Definition

X_n es una secuencia de v.a., con $X_n \sim f_n(\cdot)$ y X una v.a. con distribución $f(\cdot)$. Si para cada x tenemos que $f(x)$ es continua, $f_n(x) \xrightarrow[n\to\infty]{} f(x)$, entonces decimos que$X_n$ *converge en distribución* a X, y se escribe $X_n \underset{d}{\rightarrow} X$. (i.e., la distribución de X es la distribución límite de X_n).

- Ej: $t_{n-1} \underset{d}{\rightarrow} N(0,1)$ $\Rightarrow$conforme n crece al infinito, t_{n-1} converge a otra v.a. (una normal estándar). Entonces decimos que la media y varianza límite de t_{n-1} son 0 y 1, respectivamente.
- Convergencia en distribución se relaciona con un "histograma" de dos v.a.. Dos v.a. pueden tener la misma distribución pero sus realizaciones pueden ser completamente diferentes. Esto es distinto a la idea de convergencia en probabilidad, la cual establece que si $plim\, X_n = X$ entonces para valores grandes de n las realizaciones de las v.a. X_n y X deberían ser muy parecidas.

Slutsky (2)

- Si $g(X_n)$ es una v.a. tal que:
  - $X_n \underset{d}{\rightarrow} X$.
  - $g(X_n)$ es continua y diferenciable
  - $g(X_n)$ no depende de n
- Entonces:

1. $g(X_n) \underset{d}{\rightarrow} g(X)$.
   - Ej: sabemos que $t_{n-1} \underset{d}{\rightarrow} N(0,1)$. Luego $t_{n-1}^2 \underset{d}{\rightarrow} [N(0,1)]^2 = F_{[1,n]} = \chi^2_{[1]}$
2. Si $X_n \underset{d}{\rightarrow} X$ y *plim* $Y_n = \theta$, entonces $g(X_n, Y_n) \underset{d}{\rightarrow} g(X_n, \theta)$, y en efecto:
   - $X_n \cdot Y_n \underset{d}{\rightarrow} X_n \cdot \theta$
   - $X_n + Y_n \underset{d}{\rightarrow} X_n + \theta$
   - $\frac{X_n}{Y_n} \underset{d}{\rightarrow} \frac{X_n}{\theta}$

Teorema del Límite Central

- Si sabemos que *plim* $\bar{X}_n = \mu$, podemos entonces decir que $\bar{X}_n \underset{d}{\rightarrow} N(\cdot\,,\cdot\,)$?

Theorem

[Teorema del Limite Central, CLT] Sea $\{X_i\}_{i=1}^{n}$ *una muestra i.i.d, con* $E(X_i) = \mu < \infty$ *y* $Var(X_i) = \sigma^2 < \infty$. *Luego,*

$$\sqrt{n}(\bar{X}_n - \mu) \underset{d}{\rightarrow} N\left(0,\, \sigma^2\right)$$

Esto es equivalente a decir:

$$\frac{(\overline{X}_n - \mu)}{\sigma/\sqrt{n}} \underset{d}{\rightarrow} N\,(0,\, 1)$$

(demostración omitida)

Teorema del Límite Central

- Por LLN, la media muestral converge en probabilidad a μ conforme $n \to \infty$. En cambio, el CLT establece que conforme n crece, la distribución de la diferencia entre $\bar{X}_n$ y su límite μ, cuando es multiplicada por el factor $\sqrt{n}$ (esto es: $\sqrt{n}(\bar{X}_n - \mu)$), se aproxima a una distribución normal con media 0 y varianza σ^2
- IMPORTANTE: la distribución de $\sqrt{n}(\bar{X}_n - \mu)$ se aproxima a una normal independiente de la forma de la distribución de la cual provienen cada v.a. X_i . Dado que $\sqrt{n}(\overline{X}_n - \mu) \underset{d}{\to} N\left(0, \sigma^2\right)$, entonces $(\overline{X}_n - \mu) \underset{d}{\to} N\left(0, \frac{\sigma^2}{n}\right) \Rightarrow \overline{X}_n \underset{a}{\to} N\left(\mu, \frac{\sigma^2}{n}\right)$, i.e., $\bar{X}_n$ se distribuye *asintóticamente normal* con media μy varianza $\frac{\sigma^2}{n}$.
- La clave acá es que si tenemos una v.a.1 que converge a una v.a. 2 (es decir, no converge a una constante), y a priori desconocemos la distribución de la v.a.2, entonces podemos usar la transformación estabilizadora $\sqrt{n}(v.\bar{a}.2 - \mu)$ tal que la distribución límite de $\sqrt{n}(v.\bar{a}.2 - \mu)$ tiene una media y una varianza conocida. Luego podemos descomponer esta dist. y encontrar la dist. de la $v.\bar{a}.2$. Esta es la belleza del CLT!

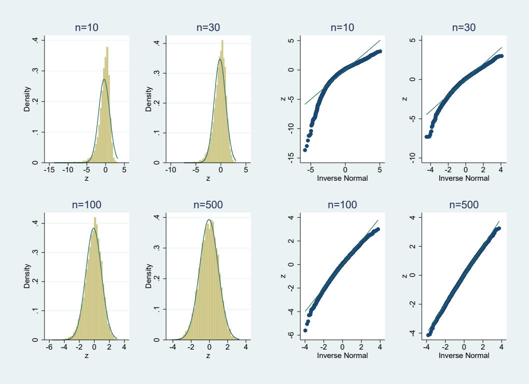
Qué es una "muestra grande"?

Ejercicio

1. Genere muchas (digamos $r = 1 \dots R$) muestras aleatoria de tamaño $n = (10; 30; 100; 500)$ a partir de una dit. exponencial con $\lambda = 0.5$ (y por tanto una media poblacional de $\frac{1}{0.5} = 2$), y (i) Calcule el promedio $\overline{y}_r$ para cada muestra r; (ii) Plotee un histograma de $\overline{y}_r$imponiendo una densidad normal con media μ_ny varianza σ_n^2. Deberías obtener 4 histogramas, una para cada tamaño muestral. (nota: para la inversa de la distribucion acumulada de una exponencial, use $-ln(1-p)/\lambda$).
2. Crear qqplots que confirmen el resultado del ejercicio anterior.
3. Para cada muestra $r = 1 \dots R$, crear un I.C. asintótico del 95% para la media poblacional y crear un indicador igual a 1 si el I.C contiene la media poblacional, y 0 si no. Para qué tamaños muestrales el CLT predice correctamente el área de los I.C. (use n=10,30,100)?

Ejercicio

Figure: Distribución Normal, Test de Normalidad, y CLT



Normalidad Asintótica y Eficiencia Asintótica

- La mayoría de los estimadores que utilizamos en estadística y econometría pueden ser escritos como funciones del promedio muestral, en cuyo caso podemos utilizar la LLN y el CLT para construir I.C. y hacer inferencia estadística.

Definition

Un estimador es asintóticamente eficiente si la varianza de la distribución asintótica de cualquier otro estimador consistente que se distribuye asintóticamente normal excede a $\frac{Var(\theta)}{n}$.

- Ej: $\bar{X}_n \underset{a}{\rightarrow} N(\mu, \frac{\sigma^2}{n})$, pero la $Median\, X_n \underset{a}{\rightarrow} N(\mu, (\frac{\Pi}{2}) \cdot \frac{\sigma^2}{n})$. Luego, decimos que la media es asintóticamente eficiente relativo a la mediana. (Notar que los estimadores consistentes tienen una varianza límite igual a cero, i.e., al evaluar la eficiencia asintótica no comparamos las varianzas límites entre los estimadores, sino las varianza de las distribuciones asintóticamente normales a las cuales cada uno de ellos converge. En otras palabras, estamos interesados en saber a qué velocidad va a converger el estimador.)

IC asintóticos: ejemplo

- El departamento de Tecnología e Información ha recolectado datos sobre la calidad del servicio telefónico que prestan los agentes de de call center de la compañía QuickCall.
  - La muestra consiste en llamadas atendidas por 5 agentes, cada hora, durante 28 semanas.
  - Cada registro de llamada tiene información sobre qué agente atendió la llamada, la fecha, hora/minutos, semana del año, duración de la llamada, y un indicador de si el problema del consumidor fue resuelto.

Table: Estadística Descriptiva

Nombre	# llamadas	Media x hr	Desv.Est	% Fracaso
Felipe	1,320	2.48	0.69	7.58%
Camila	1,320	3.54	0.53	7.05%
Javier	1,320	4.04	1.48	6.97%
Marta	1,320	2.99	1.00	7.27%
Violeta	1,320	4.56	1.82	7.88%
Total	6,600	3.52	1.41	7.48%

IC asintóticos: ejemplo I

Consideremos el caso del agente Felipe. Del un total de 1320 llamadas recibidas, la encuesta a los consumidores revela que el falló en resolver el problema en un 7.58% de ellas. Encuentre un intervalo de confianza α para la tasa de fracaso de este agente.

- $X_i \sim \mathrm{Bernoulli}(p)$, para $i = 1, 2, \cdots, n$. $X_i = 1$ si el agente falla en resolver el problema.
- Numero total de fracasos: $Y_n = \sum_{i=1}^{n} X_i \sim \mathrm{Bin}(n, p)$, con $E(Y_i) = np$ y $Var(Y_i) = np(1-p)$.
- Un estimador de p (la probabilidad de fracaso en una llamada), viene dada por:

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i = Y_n/n$$

$$SE(\hat{p}) = \sqrt{\widehat{p}(1-\hat{p})/n}$$

IC asintóticos: ejemplo II

- Sabemos de la Tabla 1 que: $\hat{p} = 7.58\%$ y $n = 1,320$. Usando el CLT, un IC al $\alpha\%$ se puede calcular como

$$7.58\% \pm z_\alpha \cdot 0.73\%.$$

- Luego, el IC al 95% es [6.15%,9.00%]