REPASO

La siguiente unidad de este curso se trata de estimación puntual (ver programa). Para comprender estos contenidos es bueno recordar que la varianza de una población y de una muestra se mide de manera distinta. Es por esto que haremos un breve resumen para recordar sus diferencias y fórmulas.

Media

La media es una medida de localización o tendencia central de una variable. Su valor es lo que comúnmente conocemos como el promedio de una variable.

Para los datos de una **muestra**, la media se denota con $\bar{x}$, y para los datos de una población con μ, como vimos en clases pasadas. La media muestral es la sumatoria de las n observaciones, dividida por el tamaño de la muestra n:

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

Para la media de una **población**, μ viene dada por la sumatoria de las N observaciones, dividida por el tamaño de la población N:

$$\mu = \frac{\sum x_i}{\boldsymbol{N}}$$

Varianza

La varianza es una medida de dispersión o de variabilidad que está basada en la diferencia que existe entre las observaciones y la media. Esta diferencia, también conocida como desviación respecto de la media, se calcula para una **población** como $(x_i - \mu)$ y para una **muestra** como $(x_i - \bar{x})$. Para el cálculo de la varianza, primero es necesario elevar las desviaciones respecto a la media al cuadrado y luego calcular el promedio de estas desviaciones elevadas al cuadrado.

La varianza **poblacional** se denota como σ^2, y se expresa de la siguiente forma:

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

La varianza **muestral** se denota como s^2, su fórmula es la siguiente:

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

La varianza muestral se divide en $n - 1$ ya que es posible demostrar que al dividir las desviaciones con respecto a la media al cuadrado por $n - 1$, esta constituye un estimador no sesgado de la varianza poblacional.

La **desviación estándar** es la raíz cuadrada de la varianza y se denota σ para la población y s para la muestra.

ESTIMACIÓN PUNTUAL

Los estimadores puntuales son estadísticos que se obtienen desde una muestra y ayudan a estimar los parámetros poblacionales. Por ejemplo, podemos calcular el estadístico muestral $\bar{x}$ para obtener un estimador puntual de la media poblacional μ. También podemos hacer estimaciones puntuales a partir de la varianza muestral s^2 y la desviación estándar muestral s para hablar de la varianza poblacional σ^2 y de la desviación estándar poblacional σ, respectivamente. Otro ejemplo de estimador es la proporción muestral, calculada como $\bar{p} = \frac{x}{n}$, donde x es el número de observaciones de cierta característica que se quiere medir, dividido por el tamaño muestral n.

Distribución muestral

Como vimos en clases anteriores, la **media muestral** $\bar{x}$ es una variable aleatoria con una distribución de probabilidad que llamamos "distribución muestral de la media de las muestras". Esto lo resumiremos simplemente como: valor esperado de $\bar{x}$. La media de la variable aleatoria $\bar{x}$ es el valor esperado de $\bar{x}$. Sea $E(\bar{x})$ el valor esperado de $\bar{x}$ y μ la media de la población de la cual se seleccionó una muestra aleatoria simple. Si van a los libros incluidos en la bibliografía del curso, podrán ver que se puede demostrar que al emplear muestreo aleatorio simple, el valor esperado de $\bar{x}$ y μ son iguales:

$$E(\bar{x}) = \mu$$

Es importante tener en cuenta que si el valor esperado de un estimador puntual es igual al parámetro poblacional, el estimador puntual es **insesgado**. En este caso, $\bar{x}$ es un estimador insesgado de la media poblacional.

Para el caso de la **desviación estándar** de $\bar{x}$, denotada como $\sigma_{\bar{x}}$, se utilizan fórmulas distintas dependiendo del tipo de población sobre la cual se realizó el muestreo aleatorio simple. Esta población puede ser finita o infinita.

Para el caso de una **población finita**:

$$\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N-1}} * \left(\frac{\sigma}{\sqrt{n}}\right)$$

Y para el caso de una **población infinita**:

$$\sigma_{\bar{x}} = \left(\frac{\sigma}{\sqrt{n}}\right)$$

Siendo N el tamaño poblacional, n el tamaño muestral y σ sigma la desviación estándar poblacional.

En el caso de una población finita, se puede observar lo que se conoce como "**factor de corrección para una población finita**": $\sqrt{\frac{N-n}{N-1}}$. Cuando una población finita es grande y el tamaño muestral es pequeño, este factor de corrección es cercano a 1. Y cuando el factor de corrección es cercano a 1, la diferencia entre la desviación estándar muestral para muestras finitas o infinitas es insignificante. Por lo tanto, cuando una población finita es grande y el tamaño muestral es pequeño, podemos utilizar la fórmula para poblaciones infinitas y eso nos dará una buena aproximación para el cálculo.

La regla general plantea que la desviación estándar se calculará de la siguiente manera, siempre y cuando la población sea infinita o la población sea finita y el tamaño de la muestra sea menor o igual al 5% del tamaño de la población (es decir, cuando $\frac{n}{N} \leq 0{,}05$):

$$\sigma_{\bar{x}} = \left(\frac{\sigma}{\sqrt{n}}\right)$$

Distribución muestral de $\overline{x}$

Para identificar las características de la distribución muestral de $\bar{x}$ tenemos que identificar la forma de la distribución muestral. Consideraremos dos casos: una población con distribución normal y una población cuya distribución no es normal.

Como mencionamos en las clases anteriores, si la población se distribuye normalmente, la distribución de la muestra aleatoria será normal, independiente del tamaño muestral que tenga. Pero si la población no se distribuye normalmente, entonces debemos utilizar el teorema central del límite para determinar la forma de la distribución muestral de $\bar{x}$.

El teorema central del límite aplicado a la distribución muestral plantea que, cuando se seleccionan muestras aleatorias simple de tamaño n de una población, la distribución muestral de la media muestral $\bar{x}$ puede aproximarse mediante una distribución normal a medida que el tamaño de la muestra se hace grande. Diversas investigaciones han mostrado como la distribución muestral se aproximar a una distribución normal cuando el tamaño muestral es mayor o igual a 30.

Ejercicio 1

Los siguientes datos vienen de una muestra aleatoria simple.

5 8 10 7 10 14

¿Cuál es la estimación puntual de la media poblacional y de la desviación estándar?

Ejercicio 2

Imagínense hacemos una encuesta a estudiantes de la FEE acerca de si correr o no la solemne. De una muestra de 150 individuos, las respuestas fueron 75 Sí, 55 No y 20 sin opinión.

¿Cuál es la estimación puntual de la proporción de la población que responde Sí?

¿Cuál es la estimación puntual de la proporción de la población que responde No?

OTROS ESTIMADORES PUNTUALES

Hasta ahora hemos aprendido que las variables aleatorias se modelan a través de distribuciones de probabilidad. Además, sabemos que estas variables aleatorias tienen ciertos parámetros que las definen. Por ejemplo, no es lo mismo una distribución Poisson cuyo parámetro lambda λ es igual a 10 y otra Poisson cuyo parámetro lambda es igual a 3. Para

determinar el parámetro lambda, lo que hemos estado haciendo es definirlo según algún criterio. Para el caso de una distribución Poisson, sabemos que su promedio poblacional o esperanza es $\mathrm{E}(\mathrm{x}) = \lambda$, por lo tanto, si creemos que un evento se repite en promedio 20 veces por ciclo, definiremos $\lambda = 20$.

Ejercicio 3

Tomando en cuenta lo aprendido, ¿cómo modelarían las siguientes situaciones? ¿Cómo definirían la variable aleatoria? ¿Qué tipo de distribución utilizarían? ¿Cómo definirían el parámetro?

- Supongamos que tenemos una muestra aleatoria simple de un estudio que mide el efecto de un remedio. Este remedio puede haber sido exitoso o no. Se cree que en 23 de cada 100 personas el remedio funciona.
- El tiempo que pasa para que se produzca un auto en una industria. Según datos anteriores, en promedio, toman 10 horas en producirse cada auto.

Analicemos los ejercicios anteriores. En el primero caso sobre el remedio, se cree que la muestra tiene un parámetro determinado. En el segundo caso sobre la industria de autos, podemos definir el parámetro según datos históricos. Es decir, en ningún caso tenemos certeza absoluta de cómo modelar la variable aleatoria. La razón es sencilla: no tenemos la población completa. Piensen que si quieren probar el efecto de un remedio es imposible dárselo a todas las personas que padezcan una enfermedad. Piensen también que es muy difícil recolectar los datos de una industria completa. Quizás tengamos los datos de algunas fábricas en algún periodo, pero no tendremos los datos completos. Por lo tanto, si queremos modelar estas variables, tendremos que estimar el valor de estos parámetros.

Existen varios métodos para realizar estimaciones. Por ejemplo, puede que tengamos una creencia sobre el estimador, como en el ejercicio sobre el remedio. Otra posibilidad es que tengamos datos que nos den pistas sobre los parámetros, como en el ejercicio de la industria de autos. Según la información que tengamos o según cual queramos utilizar, usaremos estimadores de Bayes o estimadores de Máxima Verosimilitud.

ESTIMADORES DE BAYES

Los estimadores de Bayes unen creencias y datos. En la estimación se utiliza tanto una creencia sobre un parámetro como los datos que podamos extraer de una muestra. El nombre de estos estimadores viene del hecho que, para crear el parámetro, utilizan el Teorema de Bayes, el cual plantea lo siguiente:

$$\Pr(\mathrm{X}|\mathrm{Y}) = \frac{\Pr(\mathrm{Y}|\mathrm{X}) \cdot \Pr(\mathrm{X})}{\Pr(\mathrm{Y})}$$

Es decir, la probabilidad de X, dado que sucede Y, es igual a la probabilidad de Y, dado que sucedió X, multiplicado por la probabilidad de X, todo esto dividido por la probabilidad de que suceda Y.

Esto que suena muy complejo es más fácil de comprender con un ejemplo. Supongamos que X es una variable que representa las ventas de un mall, mientras que Y es una variable que indica cuánto tiempo falta para el próximo feriado. El Teorema de Bayes dice que podemos conocer cuál es la probabilidad de que las ventas sean altas cuando falta poco para un feriado, si conocemos cual es la probabilidad de que falte poco para un feriado cuando las ventas son altas y cuál es la probabilidad de las ventas.

Pongámonos en otro caso e imaginémonos que queremos saber cuál es la probabilidad de que una empresa saque al mercado un nuevo producto si la economía está en recesión. Según Bayes, podemos encontrar una respuesta si sabemos qué ha pasado con la economía cuando la empresa ha lanzado un producto. Podríamos decir que si la economía casi siempre ha estado mal cuando la empresa lanzó un producto, entonces la probabilidad de sacar un producto al mercado cuando la economía está mal será más alta. Lo mismo pasará si, por ejemplo, tengo la creencia de que la economía pasa por malos periodos con más frecuencia que la que pasa por buenos periodos (en cuyo caso Pr(X) sería alta).

Algo fundamental de comprender es que **para los bayesianos los parámetros también son variables aleatorias**. Esta es la diferencia más importante que veremos con los otros métodos de estimación. Con esta diferencia en consideración y manteniéndonos en la lógica bayesiana, supongamos que tenemos una variable aleatoria para determinar si las empresas cumplen o no con sus presupuestos anuales. Esta variable sería Bernoulli, con un parámetro p. En este caso, el proceso de la estimación de Bayes seguiría los siguientes pasos:

1. **Obtener una distribución a priori**: Antes de ver los datos se obtiene una distribución para el estimador. La creencia inicial que tengo sobre el estimador se plasma en una función de probabilidad. Por ejemplo, puedo creer que, en promedio, el 85% por ciento de las empresas cumplen con sus presupuestos. Por lo tanto, el parámetro p seguiría una distribución de probabilidad cuyo promedio sería 0,85. Esta distribución, por convención, se escribe como: $\xi(\mathrm{p})$, o sea, función xi de p.
2. **Observar los datos**: Una vez que ya modelamos la creencia que tenemos sobre el parámetro, procedemos a observar una muestra. Por ejemplo, recogemos datos de 100 empresas en Chile y analizamos si cumplen o no con sus presupuestos. Cada uno de los datos seguirá una distribución Bernoulli. Cada empresa es independiente de cómo se comporta la otra. Además, como aprendimos en cursos previos de estadística, la probabilidad de que dos eventos independientes ocurran a la vez, es igual a la multiplicación de la probabilidad de que cada evento ocurra. Por lo tanto, si tengo estas 100 empresas, cada una siendo una variable aleatoria, puedo crear una única función para la muestra multiplicando las f.d.p de cada observación. A esa función resultante, la llamaremos función de verosimilitud y la denotaremos por $\mathrm{f}(\mathbf{x}|\mathrm{p})$ (la **x** está en negrita porque representa a las 100 empresas, es decir no es un vector de una empresa).
3. **Obtener una distribución a posteriori**: Al comenzar este tercer paso ya tenemos definida una forma matemática para la creencia que tenemos sobre p. Además, tenemos datos sobre cómo se comportan las empresas en una muestra. Por lo tanto, lo que queda ahora es unir esta información. Para hacerlo, se usa el teorema de Bayes de la siguiente forma:

$$\xi(p|x) = \frac{f(x|p) \cdot \xi(p)}{f(x)}$$

Donde $\xi(p|x)$ es la distribución a posteriori de p condicional a una muestra x. Como podemos ver, es el mismo Teorema de Bayes que hemos explicado anteriormente. Lo que estamos haciendo es actualizar lo que creíamos de p usando los datos de la muestra, para así generar una nueva distribución de probabilidades para p.

4. **Obtener el estimador**: Después de realizados todos los pasos anteriores, contamos con una f.d.p para p que une creencias con datos. Supongamos que, por ejemplo, que los datos hayan arrojado un promedio muestral de 0.70, no de 0.85 como creíamos inicialmente. La función de distribución a posteriori va a intentar armonizar, en una sola ecuación estos dos hechos. Lo que tenemos una función de distribución de probabilidades y lo que finalmente queremos encontrar es una forma para estimar p. Para ello, se recurre al uso de funciones de pérdida o de costo, pero no ahondaremos en esto, dado que el resultado será, en condiciones típicas, siempre el mismo: el estimador de Bayes será el promedio de la distribución a posteriori. Por lo tanto, intuitivamente, podemos saber que el estimador de Bayes será más bajo que 0,85 y más alto que 0,70, ya que incorpora ambos promedios en una sola distribución.

STATA

Para aplicar lo aprendido sobre estimación puntual en Stata, usaremos el primer ejercicio, donde planteamos tener una muestra con seis observaciones: 5, 8, 10, 7, 10 y 14. Para introducir estas observaciones en el programa podemos simplemente abrir el Data Editor e ingresarlas manualmente, o bien, generar las observaciones con comandos:

```
. set obs 6
obs was 0, now 6
. generate var1 = 5 in 1
(5 missing values generated)
```

El primer comando <set obs #> abre en el editor el número de observaciones que tendrá la base, es por esto que aparece “obs was 0, now 6” estableciendo que el número de observaciones cambio de 0 a 6. El segundo comando <generate var1 = 5 in 1> genera la variable denominada var1 y reemplaza con un 5 en la observación 1, generando 5 valores perdidos, representados con un punto. Para la siguiente observación el comando sería <replace var1 = 8 in 2>, para poder reemplazar el valor perdido de la observación 2 en 8. Y así sucesivamente.

Al usar el comando summarize obtenemos el resumen de la variable generada var1, el número de observaciones, la media, la desviación estándar más la mínima y la máxima.

```
. sum var1
```

```
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        var1 |         6           9    3.098387          5         14
```

Esos mismos valores los podemos estimar nosotros. Por ejemplo, veamos a continuación cómo podemos calcular el valor de la desviación estándar sin usar el comando summarize. Para esto debemos recordar la fórmula de desviación estándar $s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$.

El primero paso es generar una segunda variable con la diferencia entre el valor de la observación y la media. Para esto, después del comando summarize de var1 se debe generar la variable var2 con lo necesitado. Este comando debe ser utilizado después de el comando summarize. Como vimos en clases anteriores, el comando summarize calcula una serie de estadísticos (pueden observarse todos si se especifica la opción <summarize, detail>), que quedan almacenados en la memoria temporal de Stata como variables locales y en este caso particular como la variable local <r(mean)>. Para ver estos estadísticos locales se usa el comando <ret list>.

```
. gen var2=var1-r(mean)
```

Luego de tener var2, que indica el diferencial con respecto a la media, podemos elevarla al cuadrado y generar una tercera variable, var3.

```
. gen var3=var2*var2
```

Finalmente podemos sumar todas las diferencias al cuadrado en una nueva variable, var4. Aquí utilizamos el comando <egen>, que permite generar nuevas variables utilizando alguna función. En este caso utilizamos la función sum, que realiza una sumatoria.

```
. egen var4=sum(var3)

. sum var4

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        var4 |         6          48           0         48         48
```

Finalmente se divide esta sumatoria por n-1 y se saca la raíz cuadrada mediante el comando display:

```
. di sqrt(48/5)

3.0983867
```