

Apuntes Unidad 4

Intervalos de confianza

INTERVALOS DE CONFIANZA

Recordemos que cuando la media poblacional μ es desconocida no se puede determinar de forma precisa, pero si es posible estimarla a través de un intervalo $[a, b]$ que contiene a μ con cierta probabilidad. Para determinar un intervalo de confianza para la media de la población μ hacemos dos supuestos:

- La distribución de la población sigue una distribución aproximadamente normal.
- El valor de la desviación estándar de la población σ es conocido.

Por ejemplo, fijando la probabilidad de encontrar a μ en un 95%, tenemos que el intervalo de confianza viene dado por:

$$\left[\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}} \right]$$

Donde $\bar{X}$ corresponde a la media muestral, σ a la desviación estándar poblacional y n al tamaño de la muestra. Ahora bien, es importante recalcar que el intervalo obtenido es aleatorio, pues se genera a partir de la media muestral $\bar{X}$. El promedio muestral observado, denotado por $\bar{x}$, corresponde a una realización de la variable aleatoria $\bar{X}$. Manteniendo el nivel de confianza en 95% y reemplazando $\bar{X}$ por el promedio muestral observado $\bar{x}$, el intervalo fijo resultante se conoce como intervalo de confianza del 95% para μ . Este intervalo puede expresarse como:

$$\left[\bar{x} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96 \cdot \frac{\sigma}{\sqrt{n}} \right]$$

Finalmente, un intervalo de confianza de $(1 - \alpha) \cdot 100\%$ para una media poblacional μ , se puede expresar de la siguiente manera:

$$\left[\bar{x} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \right]$$

donde $\bar{x}$ corresponde al promedio muestral observado, n al tamaño de la muestra extraída

Curso: Probabilidad y estadística descriptiva

Unidad 4 : Inferencia estadística

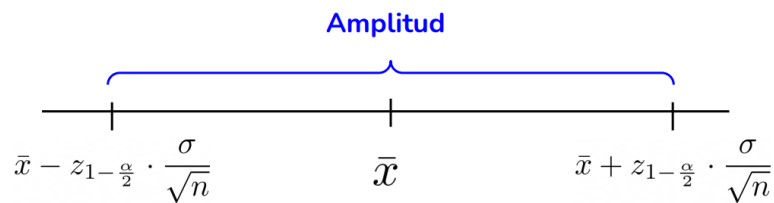
Tema: Intervalos de confianza.

Contenido: Intervalos de confianza

desde la población, σ a la desviación estándar de la población y $z_{1-\frac{\alpha}{2}}$ es el valor de z asociado a un nivel de confianza $(1 - \alpha) \cdot 100\%$.

AMPLITUD DE UN INTERVALO DE CONFIANZA

La amplitud de un intervalo de confianza corresponde a la diferencia entre el extremo superior y el extremo inferior del intervalo, como se muestra a continuación:



extremo superior – extremo inferior

$$\left(\bar{x} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \right) - \left(\bar{x} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \right)$$

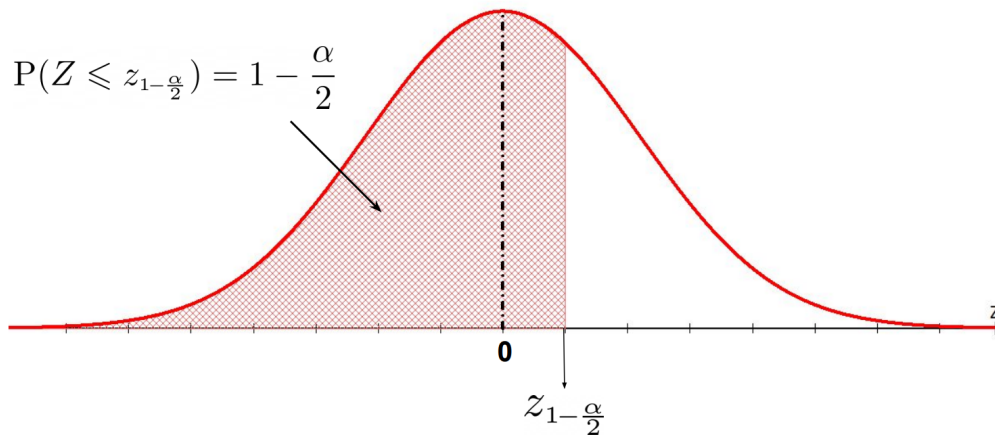
Considerando una población de interés con media μ desconocida, desviación estándar σ conocida y suponiendo que seleccionamos una muestra aleatoria de tamaño n de esa población y calculamos el promedio muestral $\bar{X}$, un intervalo que permite estimar μ con nivel de confianza $(1 - \alpha) \cdot 100\%$ es:

$$\left[\bar{X} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} , \bar{X} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \right]$$

La amplitud de este intervalo de confianza es igual a:

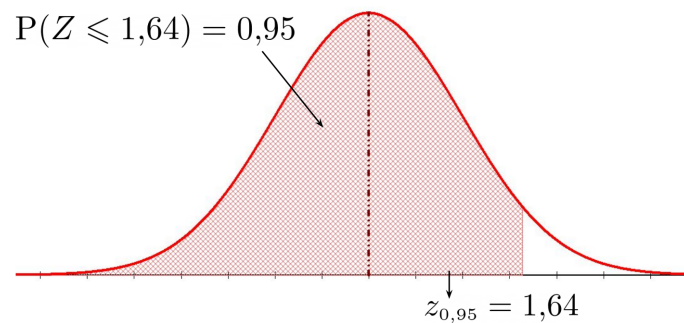
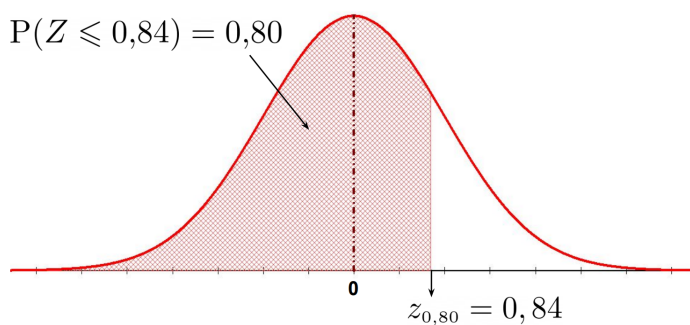
$$2 \cdot z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

Notemos que cuando los valores de σ y n permanecen constantes, esta amplitud solo cambia al modificar el valor del percentil $z_{1-\frac{\alpha}{2}}$. Este percentil representa el valor de la variable aleatoria Z que acumula un área bajo la curva normal estándar igual a $(1 - \frac{\alpha}{2})$, como se muestra en la siguiente figura:



Si cambiamos el nivel de confianza $(1 - \alpha) \cdot 100\%$, también cambiará el valor de $(1 - \frac{\alpha}{2})$, ya que ambos están relacionados por medio de α , y por tanto también se modificará el valor del percentil $z_{1-\frac{\alpha}{2}}$.

De hecho, cuando aumenta el nivel de confianza $(1 - \alpha) \cdot 100\%$, el valor del percentil $z_{1-\frac{\alpha}{2}}$ también aumenta, ya que el área en rojo debe ser mayor. Los siguientes gráficos ilustran esto:



Como la desviación estándar σ no cambia (es un parámetro de la población) y dado que la amplitud no depende del promedio muestral $\bar{X}$, al fijar el nivel de confianza $(1 - \alpha) \cdot 100\%$, la amplitud del intervalo de confianza es inversamente proporcional a $\sqrt{n}$. Esto se traduce en que, si n crece, la amplitud disminuye.

INTERPRETACIÓN DEL INTERVALO DE CONFIANZA

Consideremos una población donde se ha observado la variable de interés X y se busca estimar la media μ de esta variable. Para ello, podemos seleccionar una muestra de tamaño

Curso: Probabilidad y estadística descriptiva

Unidad 4 : Inferencia estadística

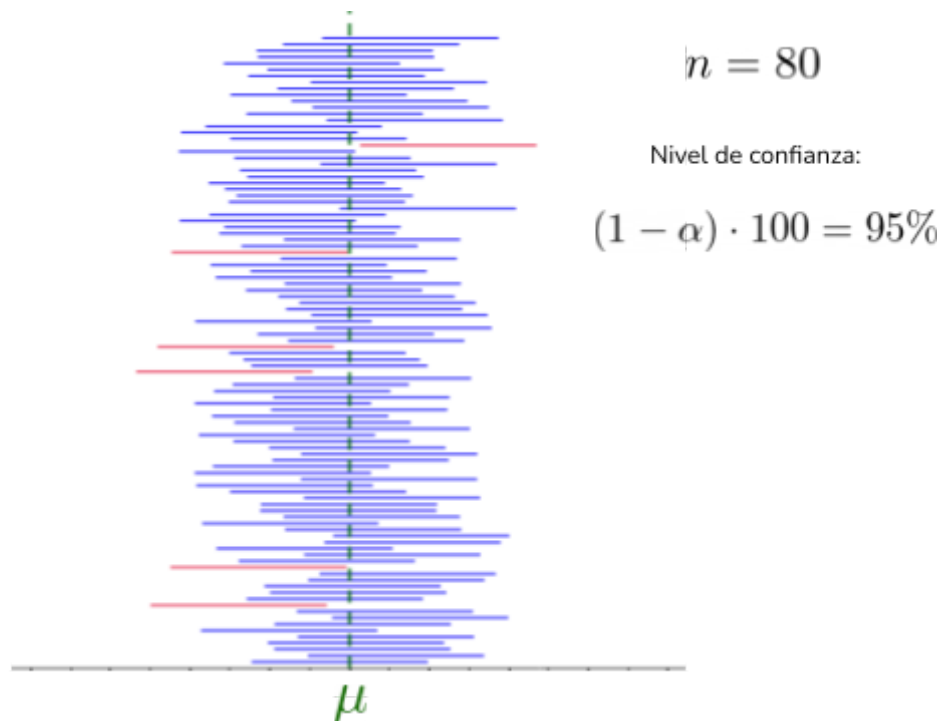
Tema: Intervalos de confianza.

Contenido: Intervalos de confianza

n de dicha población, calcular el promedio de dicha muestra $\bar{x}$, y utilizar este para construir un intervalo con nivel de confianza $(1 - \alpha) \cdot 100\%$.

Imagina que repetimos este proceso varias veces, manteniendo el mismo tamaño de la muestra y nivel de confianza, de tal forma que generemos muchos posibles intervalos. El nivel de confianza significa que **se espera que aproximadamente un $(1 - \alpha) \cdot 100\%$ de todos los intervalos contengan el verdadero valor de la media μ de la población.**

Lo anterior se ilustra en el siguiente gráfico que representa 100 intervalos de muestras de tamaño $n = 80$ y con nivel de confianza del 95% cada uno:



Como se puede observar, de los 100 intervalos de confianza, 94 de ellos contienen la media μ . Si consideramos muchos más intervalos de confianza, contruidos a partir de muestras de un tamaño razonable, deberíamos esperar que alrededor del 95% de ellos contenga a la media poblacional.

Curso: Probabilidad y estadística descriptiva

Unidad 4 : Inferencia estadística

Tema: Intervalos de confianza.

Contenido: Intervalos de confianza

ERROR DE ESTIMACIÓN

Esta expresión, utilizada para evaluar cuándo el valor de la media de la población μ pertenecía o no al intervalo de confianza, calculado a partir de una muestra aleatoria, se conoce como **error de estimación**, y se define como:

$$z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

Notemos que el error de estimación corresponde a la mitad de la amplitud del intervalo de confianza.

SÍNTESIS

- Imaginemos que queremos encontrar un intervalo $[a, b]$ que contenga el valor real de μ con una probabilidad del 95%. Esta probabilidad se suele denominar nivel de confianza, y puede ser distinta a 95%. Al intervalo que queremos encontrar lo denominamos intervalo de confianza.

Para determinar un intervalo de confianza para la media de la población μ hacemos dos supuestos:

- La distribución de la población sigue una distribución aproximadamente normal.
 - El valor de la desviación estándar σ de la población es conocido.
-
- Considerando la distribución del promedio muestral $\bar{X}$, un intervalo con un 95% de confianza nos indica que la probabilidad de que el verdadero valor del parámetro μ se encuentre contenido dentro del intervalo es igual a 0.95.
-
- Para estimar los valores de a y b , se debe estandarizar la variable $\bar{X}$. Este desarrollo nos permite obtener un intervalo $[a, b]$ al 95% de confianza, es decir, la probabilidad de que el siguiente intervalo contenga al verdadero valor de μ es del 95%.

$$\left[\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}} \right]$$

- Es importante aclarar que el intervalo obtenido es aleatorio, pues se genera a partir de la media muestral. El promedio muestral observado, denotado por $\bar{x}$ corresponde a una realización de la variable aleatoria $\bar{X}$. Si reemplazamos $\bar{X}$ por el promedio muestral observado $\bar{x}$, el intervalo fijo resultante se llama **intervalo de confianza de 95% para μ** . Este intervalo puede expresarse como:

$$\left[\bar{x} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96 \cdot \frac{\sigma}{\sqrt{n}} \right]$$

- Cuando estimamos intervalos de confianza, es usual querer estimarlos para distintos niveles de confianza. Por convención, el nivel de confianza se denota por $(1 - \alpha) \cdot 100\%$, y expresa la probabilidad de que el intervalo determinado contenga al valor real del parámetro estimado.

Un intervalo de confianza de $(1 - \alpha) \cdot 100\%$ para una media poblacional μ , se puede expresar de la siguiente manera:

$$\left[\bar{x} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \right]$$

donde $\bar{x}$ corresponde al promedio muestral observado, n al tamaño de la muestra extraída desde la población, σ a la desviación estándar de la población y $z_{1-\frac{\alpha}{2}}$ es el valor de z asociado a un nivel de confianza $(1 - \alpha) \cdot 100\%$.

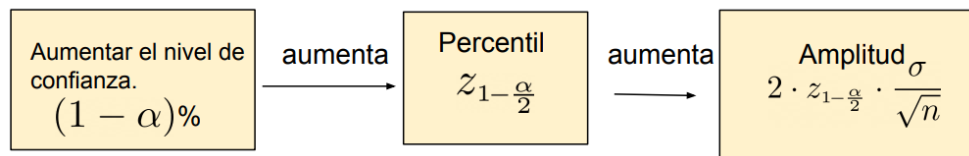
Considere el siguiente intervalo de confianza de $(1 - \alpha) \cdot 100\%$ para una media poblacional μ , con σ conocido:

$$\left[\bar{x} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} , \bar{x} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \right]$$

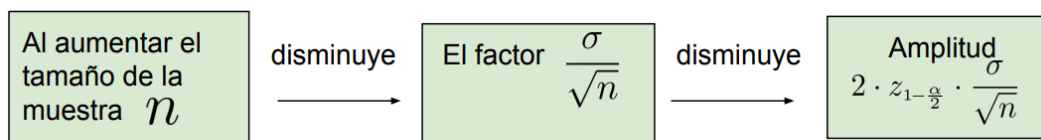
- La amplitud de este intervalo de confianza es igual a:

$$2 \cdot z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

- Si aumentamos el nivel de confianza, manteniendo todos los demás términos constantes, la amplitud del intervalo de confianza (IC) aumenta. Esto se ilustra en la siguiente tabla:



- Si aumentamos el tamaño de la muestra, manteniendo todos los demás términos constantes, la amplitud del intervalo de confianza (IC) disminuye.



- Considera una población donde se ha observado la variable de interés X con media μ desconocida. Si seleccionamos un gran número de muestras de tamaño n de esta población y a cada uno le calculamos su intervalo con un nivel de confianza $(1 - \alpha) \cdot 100\%$, entonces:

- El nivel de confianza significa que se espera que aproximadamente un $(1 - \alpha) \cdot 100\%$ de todos los intervalos de confianza contenga el verdadero valor de la media μ de la población.

- Un valor pertenecerá a un intervalo de confianza si es que se mantiene a una distancia máxima $z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ alrededor del promedio muestral. Esta distancia se denomina error de estimación.