Estadistica II
Tema 1. Inferencia sobre una población

Curso 2009/10

Tema 1. Inferencia sobre una población

Contenidos

Tema 1. Inferencia sobre una población

Objetivos de aprendizaje

- Saber estimar valores de medias, varianzas y proporciones para una población a partir de muestras aleatorias simples
- Saber construir intervalos de confianza para la media de una población
  - En el caso de distribuciones normales
  - En el caso general a partir de muestras grandes
- Saber construir intervalos de confianza para proporciones con muestras grandes
- Saber construir intervalos de confianza para la varianza de una población normal

Tema 1. Inferencia sobre una población

Referencias en la bibliografía

- Meyer, P. "Probabilidad y aplicaciones estadísticas" (1992)
    - Capítulo 14
- Newbold, P. "Estadística para los negocios y la economía" (1997)
    - Capítulos 7 y 8 (hasta 8.6)

Inferencia

Definiciones

- Inferencia: proceso de obtención de información sobre valores desconocidos de la población a partir de valores muestrales
- Parámetro: valor desconocido de la población que queremos aproximar a partir de valores de una muestra
- Estadístico: una función de la información contenida en la muestra
- Estimador: variable aleatoria que depende de la información de la muestra y cuyos valores aproximan el valor del parámetro de interés
- Estimación: un valor concreto del estimador para una muestra dada

Inferencia

Ejemplo

Estimar el gasto familiar medio anual en alimentación en una región a partir de una muestra de 200 familias

- El parámetro de interés sería el valor promedio de dicho gasto en la región
- Un estadístico relevante en este caso sería la suma de los gastos de todas las familias en la muestra
- El estimador más razonable será el promedio del gasto familiar en la muestra
- Si para una muestra concreta el promedio de gasto en alimentación es de 3.500 euros, la estimación del gasto medio anual en la región será de 3.500 euros.

Estimación puntual

- Parámetros poblacionales de interés:
  - media o varianza de una población, o la proporción de la población que posee un determinado atributo
- Selección de estimadores:
  - Intuitivamente: por ejemplo, valores equivalentes de la muestra
  - Alternativamente, elección basada en propiedades de los estimadores

Propiedades de estimadores puntuales

- Sesgo: diferencia entre la media de un estimador y el valor del parámetro
  - Si el parámetro es μ y el estimador es $\hat{\mu}$, su sesgo es

$$\mathrm{Sesgo}[\hat{\mu}] = \mathrm{E}[\hat{\mu}] - \mu$$

- Estimadores insesgados: los que tienen sesgo igual a cero
  - Si el parámetro es la media de la población μ, la media muestral $\bar{X}$ tiene sesgo cero

Estimación puntual

Propiedades de estimadores puntuales

- Eficiencia: valor de la varianza del estimador
  - Medida relacionada con la precisión del estimador
  - Un estimador es más eficiente que otro si su varianza es menor
  - Eficiencia relativa de dos estimadores $\hat{\theta}_1$ y $\hat{\theta}_2$ de un mismo parámetro,

$$\text{Eficiencia relativa} = \frac{\text{Var}[\hat{\theta}_1]}{\text{Var}[\hat{\theta}_2]}$$

Comparación de estimadores

- Mejor estimador: estimador insesgado de mínima varianza
  - No siempre es conocido
- Criterio de selección: error cuadrático medio
  - Combinación de los dos criterios anteriores
  - El error cuadrático medio (ECM) de un estimador $\hat{\theta}$ se define como

$$\text{ECM}[\hat{\theta}] = \text{E}[(\hat{\theta} - \theta)^2] = \text{Var}[\hat{\theta}] + (\text{Sesgo}[\hat{\theta}])^2$$

Estimación puntual

Selección de estimadores

- Estimadores insesgados de mínima varianza
  - La media muestral para una muestra de observaciones normales
  - La varianza muestral para una muestra de observaciones normales
  - La proporción muestral para una muestra de observaciones binomiales
- Si no se conoce un estimador con buenas propiedades
  - Procedimientos generales para definir estimadores con propiedades razonables
    - Método de máxima verosimilitud
    - Método de los momentos

Estimación puntual

Ejercicio 1.1

- A partir de los valores de ventas de un producto en ocho días consecutivos,

$$8 \quad 6 \quad 11 \quad 9 \quad 8 \quad 10 \quad 5 \quad 7$$

  - obtén estimaciones puntuales para los siguientes parámetros de la población: media, varianza, desviación típica, proporción de días con ventas superiores a 7 unidades
  - si las ventas en otro periodo de seis días han sido

    $$9 \quad 8 \quad 9 \quad 10 \quad 7 \quad 10$$

    calcula una estimación para la diferencia de medias de unidades vendidas en los dos periodos

Resultados

$$\begin{aligned}
\bar{x} &= (8+6+11+9+8+10+5+7)/8 = 8 \\
s^2 &= \tfrac{1}{n-1}\textstyle\sum_i (x_i - \bar{x})^2 = 4, \quad s = \sqrt{s^2} = 2 \\
\hat{p} &= (1+0+1+1+1+1+0+0)/8 = 0{,}625 \\
\bar{x} - \bar{y} &= 8 - (9+8+9+10+7+10)/6 = -0{,}833
\end{aligned}$$

Estimación puntual

Ejercicio 1.1

- A partir de los valores de ventas de un producto en ocho días consecutivos,

$$8 \quad 6 \quad 11 \quad 9 \quad 8 \quad 10 \quad 5 \quad 7$$

  - obtén estimaciones puntuales para los siguientes parámetros de la población: media, varianza, desviación típica, proporción de días con ventas superiores a 7 unidades
  - si las ventas en otro periodo de seis días han sido

    $$9 \quad 8 \quad 9 \quad 10 \quad 7 \quad 10$$

    calcula una estimación para la diferencia de medias de unidades vendidas en los dos periodos

Resultados

$$\begin{aligned}
\bar{x} &= (8+6+11+9+8+10+5+7)/8 = 8 \\
s^2 &= \tfrac{1}{n-1}\textstyle\sum_i (x_i - \bar{x})^2 = 4, \quad s = \sqrt{s^2} = 2 \\
\hat{p} &= (1+0+1+1+1+1+0+0)/8 = 0{,}625 \\
\bar{x} - \bar{y} &= 8 - (9+8+9+10+7+10)/6 = -0{,}833
\end{aligned}$$

Estimación por intervalos

Motivación

- En muchos casos prácticos la información de una estimación puntual no es suficiente
  - Importante conocer también información sobre la magnitud del error
  - Por ejemplo, una estimación de crecimiento anual del 0,5 % tiene una interpretación muy diferente si el valor correcto puede estar entre el 0,3 % y el 0,7 %, o bien si puede estar entre el -1,5 % y el 3,5 %
  - En estos casos nos interesa conocer también información sobre la fiabilidad del estimador puntual
- La manera más habitual de proporcionar esa información es calcular un estimador por intervalos
  - Intervalo de confianza: rango de valores que contienen la cantidad que se estima con alta probabilidad

Estimación por intervalos

Concepto

- Estimador por intervalos
  - Regla basada en información muestral
  - Para calcular un intervalo que contenga el valor del parámetro
  - Con probabilidad alta
- Para un parámetro θ, y dado un valor $1-\alpha$, el nivel de confianza, un estimador por intervalos se define como dos variables aleatorias $\hat{\theta}_A$ y $\hat{\theta}_B$ que cumplen
$$P(\hat{\theta}_A \leq \theta \leq \hat{\theta}_B) = 1-\alpha$$
- Para dos valores de dichas variables aleatorias, a y b, tenemos un intervalo $[a, b]$ que denominamos intervalo de confianza al $100(1-\alpha)\,\%$ para θ
  - A $1-\alpha$ se le llama el nivel de confianza del intervalo
  - Si generamos muchos valores a y b utilizando la regla, se cumple $\theta \in [a, b]$ el $100(1-\alpha)\,\%$ de las veces (pero no siempre)

Cálculo de intervalos de confianza

Comentarios generales

- El intervalo está asociado a una probabilidad dada, el nivel de confianza
- De la definición de los valores del estimador por intervalos, $\hat{\theta}_A$ y $\hat{\theta}_B$, en

  $$P(\hat{\theta}_A \leq \theta \leq \hat{\theta}_B) = 1 - \alpha$$

  dichos valores se pueden obtener como cuantiles de la distribución del estimador, $\hat{\theta}$
- Para ello necesitamos conocer una distribución para alguna cantidad que relacione θ y $\hat{\theta}$, de la que calcular estos cuantiles
- Esta distribución es la base del cálculo de intervalos de confianza. Depende de
  - El parámetro que queramos estimar (media, varianza)
  - La distribución de la población
  - La información de que dispongamos (si conocemos otros parámetros)
- Estudiaremos en este tema diferentes casos particulares (distintos parámetros, distribuciones)

Media de una población normal con varianza conocida

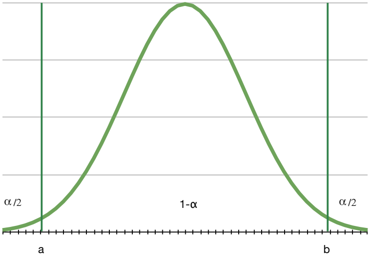
Hipótesis y objetivo

- Empezaremos considerando un caso especialmente sencillo, aunque poco realista
- Suponemos que
  - disponemos de una muestra aleatoria simple de n observaciones
  - la población sigue una distribución normal
  - conocemos la varianza poblacional σ^2
- Objetivo: construir un intervalo de confianza para la media poblacional μ (desconocida)
  - Para un nivel de confianza $1 - \alpha$ que escogemos o nos dan

Media de una población normal con varianza conocida

Procedimiento

- Sea $X_1, \dots, X_n$ la muestra aleatoria simple y $\bar{X}$ su media muestral, nuestro estimador puntual
- Nuestro primer paso es disponer de información sobre la distribución de una variable que relacione μ y $\bar{X}$
- Para esa distribución obtenemos dos valores a y b que definen un intervalo con la probabilidad deseada
- A partir de ellos definimos el intervalo para μ



Media de una población normal con varianza conocida

Procedimiento

- Para el caso que estamos considerando, la distribución de la media muestral cumple

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

  - Distribución conocida
  - Relaciona $\bar{X}$ y μ

- Construimos un intervalo que contiene la probabilidad deseada para una distribución normal estándar, buscando un $z_{\alpha/2}$ que cumpla

$$P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha$$

  - $z_{\alpha/2}$ es un valor tal que una distribución normal estándar toma valores mayores con probabilidad $\alpha/2$

Media de una población normal con varianza conocida

Procedimiento

- El intervalo siguiente tiene la probabilidad deseada

$$-z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}$$

- Sustituyendo valores muestrales y despejando en las desigualdades el valor de μ obtenemos el intervalo de confianza deseado

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Cálculo de intervalos de confianza

Ejercicio 1.2

Un proceso de envasado de un alimento produce bricks con un peso que sigue una distribución normal con desviación típica igual a 55 gr. Una muestra aleatoria simple de 50 bricks tiene un peso medio de 980 gr. Calcula un intervalo de confianza al 99 % para el peso medio de todos los bricks.

Resultados

$$\frac{\bar{X}-\mu}{55/\sqrt{50}} \sim N(0,1)$$

$$\alpha = 1-0{,}99 = 0{,}01, \qquad z_{\alpha/2} = z_{0{,}005} = 2{,}576$$

$$-2{,}576 \leq \frac{980-\mu}{55/\sqrt{50}} \leq 2{,}576$$

$$959{,}96 \leq \mu \leq 1000{,}04$$

Cálculo de intervalos de confianza

Ejercicio 1.2

Un proceso de envasado de un alimento produce bricks con un peso que sigue una distribución normal con desviación típica igual a 55 gr. Una muestra aleatoria simple de 50 bricks tiene un peso medio de 980 gr. Calcula un intervalo de confianza al 99 % para el peso medio de todos los bricks.

Resultados

$$\begin{aligned}
\frac{\bar{X}-\mu}{55/\sqrt{50}} &\sim N(0,1) \\
\alpha &= 1-0{,}99 = 0{,}01, \qquad z_{\alpha/2} = z_{0{,}005} = 2{,}576 \\
& -2{,}576 \leq \frac{980-\mu}{55/\sqrt{50}} \leq 2{,}576 \\
& 959{,}96 \leq \mu \leq 1000{,}04
\end{aligned}$$

Cálculo de intervalos de confianza

Procedimiento general

Pasos a seguir:

1. Identificar la variable con distribución conocida y la distribución sobre la que construiremos el intervalo de confianza
2. Buscar los percentiles de esa distribución que cubran el nivel de confianza elegido
3. Construir el intervalo para la variable con distribución conocida
4. Sustituir los valores muestrales
5. Despejar el valor del parámetro en este intervalo para construir otro intervalo específico para dicho parámetro

Cálculo de intervalos de confianza

Propiedades del intervalo

- El tamaño del intervalo de confianza es una medida de la precisión en la estimación
- En el caso que hemos estudiado dicho tamaño viene dado por

$$\frac{2z_{\alpha/2}\sigma}{\sqrt{n}}$$

- Por tanto, la precisión depende de:
  - La desviación típica de la población, cuanto mayor sea menos fiable será la estimación
  - El tamaño de la muestra, cuanto mayor sea mas fiable será la estimación
  - El nivel de confianza, cuanto mayor sea mayor será el tamaño del intervalo

Cálculo de intervalos de confianza

Ejercicio 1.3

Para el enunciado del ejercicio 1.2, calcula como cambia el intervalo si

- cambia el tamaño de muestra a 100 (manteniendo la media muestral)
- cambia el nivel de confianza al 95 %

Resultados

$$-2{,}576 \leq \frac{980-\mu}{55/\sqrt{100}} \leq 2{,}576$$

$$965{,}83 \leq \mu \leq 994{,}17$$

$$\alpha = 1 - 0{,}95 = 0{,}05, \qquad z_{\alpha/2} = z_{0{,}025} = 1{,}96$$

$$-1{,}96 \leq \frac{980-\mu}{55/\sqrt{50}} \leq 1{,}96$$

$$964{,}75 \leq \mu \leq 995{,}25$$

Cálculo de intervalos de confianza

Ejercicio 1.3

Para el enunciado del ejercicio 1.2, calcula como cambia el intervalo si

- cambia el tamaño de muestra a 100 (manteniendo la media muestral)
- cambia el nivel de confianza al 95 %

Resultados

$$
\begin{aligned}
& -2{,}576 \leq \frac{980 - \mu}{55/\sqrt{100}} \leq 2{,}576 \\
& 965{,}83 \leq \mu \leq 994{,}17 \\
\alpha &= 1 - 0{,}95 = 0{,}05, \qquad z_{\alpha/2} = z_{0{,}025} = 1{,}96 \\
& -1{,}96 \leq \frac{980 - \mu}{55/\sqrt{50}} \leq 1{,}96 \\
& 964{,}75 \leq \mu \leq 995{,}25
\end{aligned}
$$

Media de una población con muestras grandes

Motivación

- En muchos casos prácticos no sabemos si la distribución es normal y no conocemos su desviación típica
- Si el tamaño de la muestra es grande, el teorema central del límite nos permite construir intervalos aproximados

Hipótesis y objetivo

- Suponemos que
  - disponemos de una muestra aleatoria simple de n observaciones
  - el tamaño de la muestra es suficientemente grande
- Objetivo: construir un intervalo de confianza aproximado para la media poblacional μ (desconocida)
  - Para un nivel de confianza $1-\alpha$ que escogemos o nos dan

Media de una población con muestras grandes

Procedimiento

- Para el caso que estamos considerando, el teorema central del límite nos dice que para n elevado aproximadamente se cumple

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim N(0,1)$$

  donde S denota la desviación típica muestral
  - La misma distribución del caso anterior
  - Relaciona $\bar{X}$ y μ
- Construimos, como antes, un intervalo que contiene la probabilidad deseada para una distribución normal estándar, buscando un $z_{\alpha/2}$ que cumpla

$$P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha$$

Media de una población con muestras grandes

Procedimiento

- El intervalo siguiente tiene la probabilidad deseada

$$-z_{\alpha/2} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq z_{\alpha/2}$$

- Sustituyendo valores muestrales y despejando en las desigualdades el valor de μ obtenemos el intervalo de confianza deseado

$$\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}$$

Media de una población con muestras grandes

Ejercicio 1.4

Se ha realizado una encuesta a 60 personas en la que se pedía a los encuestados que valorasen de 0 a 5 la calidad de un servicio. La valoración media en la muestra fue de 2.8 puntos y la desviación típica muestral fue de 0.7 puntos. Calcula un intervalo de confianza al 90 % para la valoración media en la población

Resultados

$$\begin{aligned}
\alpha &= 1-0{,}9=0{,}1, \qquad z_{\alpha/2}=z_{0{,}05}=1{,}645\\
&-1{,}645 \leq \frac{2{,}8-\mu}{0{,}7/\sqrt{60}} \leq 1{,}645\\
&2{,}65 \leq \mu \leq 2{,}95
\end{aligned}$$

Media de una población con muestras grandes

Ejercicio 1.4

Se ha realizado una encuesta a 60 personas en la que se pedía a los encuestados que valorasen de 0 a 5 la calidad de un servicio. La valoración media en la muestra fue de 2.8 puntos y la desviación típica muestral fue de 0.7 puntos. Calcula un intervalo de confianza al 90 % para la valoración media en la población

Resultados

$$
\begin{aligned}
\alpha &= 1-0{,}9=0{,}1, \qquad z_{\alpha/2}=z_{0{,}05}=1{,}645\\
&-1{,}645 \leq \frac{2{,}8-\mu}{0{,}7/\sqrt{60}} \leq 1{,}645\\
&2{,}65 \leq \mu \leq 2{,}95
\end{aligned}
$$

Proporciones con muestras grandes

Motivación

- Queremos estimar la proporción de una población que cumple una cierta condición, partiendo de datos de una muestra
- La estimación de proporciones es un caso particular del caso anterior con datos no normales
- Nuestro estimador en este caso será la proporción muestral
- Si X_i representa si un miembro de una muestra aleatoria simple de tamaño n cumple o no la propiedad de interés, y la probabilidad de cumplimiento es p, entonces X_i sigue una distribución Bernoulli
- Deseamos estimar p, la proporción en la población que cumple la condición
- Partiendo de la proporción en la muestra que cumple la condición $\hat{p} = \sum_i X_i / n$
  - $\hat{p}$ es una media muestral

Proporciones con muestras grandes

Hipótesis y objetivo

- Suponemos que
  - disponemos de una muestra aleatoria simple de n observaciones que toman valores 0 ó 1
  - el tamaño de la muestra es suficientemente grande
- Objetivo: construir un intervalo de confianza aproximado para la proporción poblacional p (desconocida)
  - Para un nivel de confianza $1-\alpha$ que escogemos o nos dan

Proporciones con muestras grandes

Procedimiento

- En nuestro caso, por analogía con el caso anterior, $\mu = p$, $\sigma^2 = p(1-p)$
- El teorema central del límite nos dice que para n elevado se cumple aproximadamente
$$\frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \sim N(0,1)$$
donde $\hat{p}$ denota la proporción en la muestra
- Aproximamos $p(1-p)$ con el valor muestral correspondiente $\hat{p}(1-\hat{p})$
    - Se sigue teniendo aproximadamente la misma distribución
- Construimos, como antes, un intervalo que contiene la probabilidad deseada para una distribución normal estándar, buscando un $z_{\alpha/2}$ que cumpla
$$P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1-\alpha$$

Proporciones con muestras grandes

Procedimiento

- El intervalo siguiente tiene la probabilidad deseada

$$-z_{\alpha/2} \leq \frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})/n}} \leq z_{\alpha/2}$$

- Sustituyendo valores muestrales y despejando en las desigualdades el valor de p obtenemos el intervalo de confianza deseado

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Proporciones con muestras grandes

Ejercicio 1.5

En una muestra de 200 pacientes se han observado complicaciones importantes asociadas a una cierta enfermedad en 38 de los mismos. Se pide que calcules un intervalo de confianza al 99 % para la proporción de pacientes en la población que pueden tener complicaciones importantes con la enfermedad

Resultados

$$\hat{p} = 38/200 = 0{,}19$$

$$\alpha = 1 - 0{,}99 = 0{,}01, \qquad z_{\alpha/2} = z_{0{,}005} = 2{,}576$$

$$-2{,}576 \leq \frac{0{,}19 - p}{\sqrt{0{,}19(1-0{,}19)/200}} \leq 2{,}576$$

$$0{,}119 \leq \mu \leq 0{,}261$$

Proporciones con muestras grandes

Ejercicio 1.5

En una muestra de 200 pacientes se han observado complicaciones importantes asociadas a una cierta enfermedad en 38 de los mismos. Se pide que calcules un intervalo de confianza al 99 % para la proporción de pacientes en la población que pueden tener complicaciones importantes con la enfermedad

Resultados

$$\begin{aligned}
\hat{p} &= 38/200 = 0{,}19 \\
\alpha &= 1 - 0{,}99 = 0{,}01, \qquad z_{\alpha/2} = z_{0{,}005} = 2{,}576 \\
& -2{,}576 \leq \frac{0{,}19 - p}{\sqrt{0{,}19(1-0{,}19)/200}} \leq 2{,}576 \\
& 0{,}119 \leq \mu \leq 0{,}261
\end{aligned}$$

Media de una población normal con varianza desconocida

Motivación

- Queremos estimar la media de la población
- Sabiendo que la distribución de la población es normal
- Pero no conocemos la varianza de la población
- Si la muestra es pequeña los resultados anteriores no servirían
- Pero en este caso conocemos la distribución de la media muestral para cualquier tamaño de muestra

Hipótesis y objetivo

- Suponemos que
  - disponemos de una muestra aleatoria simple de n observaciones
  - la población sigue una distribución normal
- Objetivo: construir un intervalo de confianza para la media poblacional μ (desconocida)
  - Para un nivel de confianza $1 - \alpha$ que escogemos o nos dan

Media de una población normal con varianza desconocida

Procedimiento

- Para este caso, el resultado básico es que

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

  donde S denota la desviación típica muestral y t_{n-1} denota la distribución t de Student con $n-1$ grados de libertad
  - Es una distribución simétrica parecida a la normal (converge a ella para n elevado)
- Construimos un intervalo que contiene la probabilidad deseada pero ahora lo hacemos a partir de una distribución t de Student con $n-1$ grados de libertad, buscando un valor $t_{n-1,\alpha/2}$ que cumpla

$$P(-t_{n-1,\alpha/2} \leq T_{n-1} \leq t_{n-1,\alpha/2}) = 1 - \alpha$$

Media de una población normal con varianza desconocida

Procedimiento

- El intervalo siguiente tiene la probabilidad deseada

$$-t_{n-1,\alpha/2} \leq \frac{\bar{X}-\mu}{S/\sqrt{n}} \leq t_{n-1,\alpha/2}$$

- Sustituyendo valores muestrales y despejando en las desigualdades el valor de μ obtenemos el intervalo de confianza deseado

$$\bar{x} - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}} \leq \mu < \bar{x} + t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}$$

Media de una población normal con varianza desconocida

Ejercicio 1.6

Has medido la duración de una muestra de 20 bombillas de bajo consumo, y para dicha muestra has obtenido una duración media de 4520 horas, con una desviación típica muestral de 750 horas. Si la duración de estas bombillas se supone que sigue una distribución normal, calcula un intervalo de confianza al 95 % para la duración media (en la población) de las bombillas

Resultados

$$\alpha = 1 - 0{,}95 = 0{,}05, \qquad t_{n-1,\alpha/2} = t_{19,0{,}025} = 2{,}093$$

$$-2{,}093 \leq \frac{4520 - \mu}{750/\sqrt{20}} \leq 2{,}093$$

$$4169{,}0 \leq \mu \leq 4871{,}0$$

Media de una población normal con varianza desconocida

Ejercicio 1.6

Has medido la duración de una muestra de 20 bombillas de bajo consumo, y para dicha muestra has obtenido una duración media de 4520 horas, con una desviación típica muestral de 750 horas. Si la duración de estas bombillas se supone que sigue una distribución normal, calcula un intervalo de confianza al 95 % para la duración media (en la población) de las bombillas

Resultados

$$\begin{aligned} \alpha &= 1 - 0{,}95 = 0{,}05, \qquad t_{n-1,\alpha/2} = t_{19,0{,}025} = 2{,}093 \\ & -2{,}093 \leq \frac{4520 - \mu}{750/\sqrt{20}} \leq 2{,}093 \\ & 4169{,}0 \leq \mu \leq 4871{,}0 \end{aligned}$$

Varianza de una población normal

Motivación

- Hasta ahora solo hemos considerado intervalos de confianza para la media
- En algunos casos interesa también conocer un intervalo para la varianza
- No se conocen las distribuciones relevantes mas que para algunos casos
- Consideraremos únicamente el caso de datos normales

Hipótesis y objetivo

- Suponemos que
  - disponemos de una muestra aleatoria simple de n observaciones
  - la población sigue una distribución normal
- Objetivo: construir un intervalo de confianza para la varianza poblacional σ^2 (desconocida)
  - Para un nivel de confianza $1 - \alpha$ que escogemos o nos dan

Varianza de una población normal

Procedimiento

- Para este caso, el resultado básico es que

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

  donde S denota la desviación típica muestral y χ^2_{n-1} denota la distribución chi-cuadrado con $n-1$ grados de libertad
  - Distribución asimétrica que solo toma valores positivos
- Al igual que en los casos anteriores, el primer paso es construir un intervalo que contiene la probabilidad deseada
  - Como la distribución chi-cuadrado es asimétrica, necesitamos dos valores para definir el intervalo, $\chi^2_{n-1,1-\alpha/2}$ y $\chi^2_{n-1,\alpha/2}$

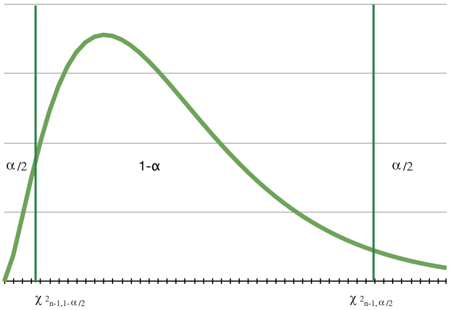
Varianza de una población normal
Procedimiento

- Seleccionamos los valores como

$$P(\chi^2_{n-1} \geq \chi^2_{n-1,1-\alpha/2}) = 1 - \alpha/2, \qquad P(\chi^2_{n-1} \geq \chi^2_{n-1,\alpha/2}) = \alpha/2$$

- Estos valores cumplen

$$P(\chi^2_{n-1,1-\alpha/2} \leq \chi^2_{n-1} \leq \chi^2_{n-1,\alpha/2}) = 1 - \alpha$$



Varianza de una población normal

Procedimiento

- El intervalo siguiente tiene la probabilidad deseada

$$\chi^2_{n-1,1-\alpha/2} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{n-1,\alpha/2}$$

- Sustituyendo valores muestrales y despejando en las desigualdades el valor de σ^2 obtenemos el intervalo

$$\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}$$

- Para la desviación típica el intervalo de confianza correspondiente será

$$\sqrt{\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}} \leq \sigma \leq \sqrt{\frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}}$$

Varianza de una población normal

Ejercicio 1.7

Para el ejercicio 1.6, se pide que calcules un intervalo de confianza al 95 % para la desviación típica de la duración de las bombillas

Resultados

$$\begin{aligned}
\alpha &= 1 - 0{,}95 = 0{,}05 \\
\chi^2_{n-1,1-\alpha/2} &= \chi^2_{19,0{,}975} = 8{,}907 \\
\chi^2_{n-1,\alpha/2} &= \chi^2_{19,0{,}025} = 32{,}852 \\
& 8{,}907 \leq \frac{19 \times 750^2}{\sigma^2} \leq 32{,}852 \\
& 325323 \leq \sigma^2 \leq 1199899 \\
& 570{,}37 \leq \sigma \leq 1095{,}40
\end{aligned}$$

Ejercicio 1.7

Para el ejercicio 1.6, se pide que calcules un intervalo de confianza al 95 % para la desviación típica de la duración de las bombillas

Resultados

$$\begin{aligned}
\alpha &= 1 - 0{,}95 = 0{,}05 \\
\chi^2_{n-1,1-\alpha/2} &= \chi^2_{19,0{,}975} = 8{,}907 \\
\chi^2_{n-1,\alpha/2} &= \chi^2_{19,0{,}025} = 32{,}852 \\
& 8{,}907 \leq \frac{19 \times 750^2}{\sigma^2} \leq 32{,}852 \\
& 325323 \leq \sigma^2 \leq 1199899 \\
& 570{,}37 \leq \sigma \leq 1095{,}40
\end{aligned}$$

Intervalos de confianza

Resumen para una población

- Para una muestra aleatoria simple

Parámetro	Hipótesis	Distribución	Intervalo
Media	Datos normales Var. conocida	$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$\mu \in \left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$
	D. no normales Muestra grande	$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim N(0,1)$	$\mu \in \left[\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}\right]$
	Proporciones Muestra grande	$\frac{\hat{P}-p}{\sqrt{\hat{P}(1-\hat{P})/n}} \sim N(0,1)$	$p \in \left[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$
	Datos normales Var. desconocida	$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$	$\mu \in \left[\bar{x} - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}, \bar{x} + t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}\right]$
Varianza	Datos normales	$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$	$\sigma^2 \in \left[\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}\right]$