

DEFINICIONES DE EXÁMEN: ESTADISTICA 2:

Explica la diferencia entre muestreo probabilístico y no probabilístico. ¿Por qué utilizamos los métodos de muestreo probabilístico?

El muestreo probabilístico se basa en la selección de elementos muestrales utilizando criterios probabilísticos mientras que el muestreo no probabilístico utiliza criterios de facilidad, accesibilidad, conveniencia, ... Utilizamos el muestreo probabilístico porque nos permite conocer la probabilidad de cada muestra, con lo cual sabremos evaluar los resultados obtenidos y realizar inferencias para toda la población.

Define la propiedad de insesgadez de un estimador respecto a un parámetro tita y explica por qué es aconsejable utilizar estimadores insesgados.

Un estimador TE se dice insesgado con respecto a un parámetro tita si la media del estimador coincide con dicho parámetro desconocido, es decir, $E(\text{TE}) = \text{tita}$. Es aconsejable que suceda esta propiedad para asegurar que los posibles valores del estimador se distribuya alrededor del valor real que queremos aproximar. En caso contrario, podría ser que los posibles valores estuvieran mayoritariamente por debajo o por encima del verdadero valor del parámetro y entonces tendríamos problemas de infraestimación o sobreestimación.

¿Qué es un contraste de hipótesis? ¿Qué significa que tenga un nivel de significación del 5%?

Un contraste de hipótesis es una técnica estadística para tomar decisiones sobre una afirmación estadística realizada para la población. Si el nivel de significación es del 5% entonces estamos admitiendo que la probabilidad máxima de rechazar mi afirmación inicial sobre la población cuando realmente es cierta es del 5%, es decir, como mucho 5 de cada 100 veces que haga el contraste cambiaré de opinión inicial para equivocarme en la decisión.

Define función de distribución de una variable aleatoria continua. Relaciona este concepto con el de función de densidad.

La función de distribución de una variable aleatoria X en un punto x es la probabilidad acumulada hasta dicho valor x , es decir, la probabilidad de observar valores menores o igual a x : $F(x) = P(X \leq x)$. En variable aleatoria continuas dicha función de distribución es el área bajo la función de densidad entre menos infinito y el valor x . Por lo tanto, la función de densidad es la derivada de la función de distribución.

Define la propiedad de consistencia de un estimador respecto a un parámetro θ y explica por qué es aconsejable utilizar estimadores consistentes.

Un estimador se dice consistente respecto a un parámetro θ si los posibles valores del estimador están cada más cerca del verdadero valor de θ si el tamaño muestral tiende a infinito. Es aconsejable utilizar estimadores consistentes para asegurarnos de que el proceso de estimación mejora al disponer de más información, es decir, cuantos más datos (mayor tamaño muestral) dispongamos de la variable aleatoria utilizada.

¿Qué diferencia hay entre un contraste de hipótesis simples y uno de hipótesis compuestas? Pon un ejemplo de cada uno de ellos.

Un contraste de hipótesis simples se define cuando ambas hipótesis, nula y alternativa, son hipótesis simples, es decir, ambas definen una única distribución para la variable Y , por lo tanto, comparamos dos distribuciones de probabilidad. Mientras que un contraste de hipótesis compuestas define alguna de ellas o las dos familias de distribuciones de probabilidad. Sea X una variable aleatoria con distribución normal y desviación típica conocida $\sigma=2$. Entonces un contraste de hipótesis simples sería: ' H_0 ': $\mu=10$, ' H_1 ': $\mu=15$. Y un contraste de hipótesis compuestas sería: ' H_0 ': $\mu=10$, ' H_1 ': μ mayor o igual que 10. En el primer caso estamos comparando una Normal $(10,2)$ frente a una normal $(15,2)$. En el segundo caso estamos comparando una distribución normal $(10,2)$ frente a todas las normales $(\mu,2)$ con μ mayor o igual que 10.

¿Por qué es necesario conocer la distribución de un estadístico muestral?

El estadístico muestral es la forma de evaluar y comparar nuestros resultados empíricos con el modelo teórico propuesto, tanto cuando construimos intervalos de confianza como contrastes de hipótesis. Por lo tanto, tenemos que conocer su distribución para saber qué decisiones y con qué seguridad las tomamos en el problema.

Si queremos aumentar la precisión de un intervalo de confianza, ¿qué dos alternativas tenemos? ¿cuál es preferible?

La precisión de un intervalo de confianza depende del nivel de confianza y del tamaño muestral. Por lo tanto, tenemos que disminuir el nivel de confianza o aumentar el tamaño muestral. Preferimos, si es posible, aumentar el tamaño muestral porque de esa forma tendremos más precisión con la misma probabilidad de cubrimiento.

¿Qué es la potencia de un contraste? ¿Qué relación tiene con el error de tipo 2? ¿Qué valores debería tomar la potencia?

La potencia de un contraste es la probabilidad de rechazar la hipótesis nula cuando es cierto un valor concreto del parámetro. Por lo tanto, si el valor del parámetro pertenece a la hipótesis alternativa entonces la potencia será el complementario de la probabilidad del error de tipo 2, es decir, la probabilidad de tomar una decisión correcta al rechazar la nula. Es evidente que esta probabilidad debe ser lo más alta posible, es decir, un valor cercano a 1.

¿Por qué es necesario tomar una muestra en el estudio de una población? Enumera al menos tres razones para ello.

Los motivos de seleccionar una muestra es recopilar información sobre un fenómeno real cuando no tenemos acceso a toda la población porque es infinita, porque necesitamos conclusiones rápidas y económicas, porque el estudio necesario supone la destrucción del elemento o porque no conocemos todos los elementos de la población.

¿Por qué es deseable que un estimador sea eficiente?

Un estimador debe ser eficiente para asegurarnos que la estimación puntual proporcionada sea lo más fiable posible. Puesto que un estimador eficiente va a ser insesgado y su varianza mínima entonces estamos seguros que nuestra estimación va a estar alrededor y lo más cerca posible del parámetro desconocido.

¿Qué es un intervalo de confianza? ¿Qué significa que esté calculado al 95% de confianza?

Un intervalo de confianza es un intervalo de extremos aleatorios que va a contener en su interior el parámetro desconocido con una probabilidad fijada a priori. Si está calculado al 95%, significará que el 95% de las muestras y, por tanto, el 95% de los intervalos construidos de esa forma contendrán a dicho parámetro en su interior.

Define variable aleatoria y explica su clasificación en variable aleatoria discreta y variable aleatoria continua.

Una variable aleatoria es una aplicación que asigna un número real a cada suceso de un experimento aleatorio, es decir, una cuantificación numérica de los posibles resultados de un fenómeno aleatorio. Se dice que es variable aleatoria discreta si utiliza un conjunto finito o infinito numerable de números reales mientras que se dice variable aleatoria continua si toma valores en un conjunto infinito de la recta real.

¿Por qué seleccionamos una muestra aleatoria en lugar de realizar un estudio completo de la población? Justifica al menos tres motivos.

La selección de una muestra para su estudio se realiza principalmente por los siguientes motivos: la población es infinita o demasiado grande, el estudio implica la destrucción del elemento, las características poblacionales cambian si el estudio se prolonga demasiado tiempo y abaratar costes monetarios y temporales.

¿Qué es el nivel de confianza de un intervalo? ¿Qué significa que un intervalo tenga un nivel de confianza del 95%?

El nivel de confianza de un intervalo es la probabilidad que garantizamos para que dentro del intervalo se encuentre el verdadero valor del parámetro que queremos estimar. Si el intervalo está calculado con un nivel del 95%, significa que el 95% de los intervalos, es decir, el 95% de las muestras extraídas, contruidos contendrán en su interior al verdadero valor del parámetro.