

ESTADÍSTICA II (Práctica Informática)

Uso de Tablas Dinámicas para el Cálculo de Intervalos de Confianza y Contraste de Hipótesis

Finalidad de la práctica:

1. Aprender la utilidad del uso de Tablas Dinámicas en la Inferencia Estadística
2. Calcular intervalos de confianza para la media, proporción y varianza poblacionales de una muestra dada.
3. Comprender e interpretar un contraste de hipótesis
4. Calcular el contraste para la media, la varianza y la proporción de una muestra dada

0. Uso de Tablas Dinámicas

Las **tablas dinámicas en Excel** nos ayudan a comparar grandes cantidades de datos e intercambiar fácilmente columnas por filas dentro de la misma tabla. Permiten realizar filtros por diferentes categorías que de otra manera necesitaríamos un tiempo considerable para construirlos. A partir de dichos filtros, es posible calcular diferentes estimadores (media muestral, cuasivarianza muestral, tamaño de la muestra, ...) para facilitar el cálculo de intervalos y test de hipótesis.

REPASO DE LA MATERIA TEÓRICA CORRESPONDIENTE

1. Cálculo del intervalo de confianza de una muestra dada

IC para la media cuando n es elevado

Tenemos una muestra de datos y queremos calcular el intervalo de confianza para la media o proporción real. Calculamos, en primer lugar, las estimaciones correspondientes a la media y la varianza real para ello es necesario obtener la media muestral (PROMEDIO) y la cuasivarianza muestral (VAR.S). A continuación, podemos calcular el radio del intervalo de confianza o error de estimación con la función INV.NORM.ESTAND y los extremos del intervalo se obtienen sumando y restando al estimador de la media o la proporción dicho error.

IC para la media de una NORMAL cuando n no es elevado

Tenemos una muestra de datos y queremos calcular el intervalo de confianza para la media real. Calculamos, en primer lugar, las estimaciones correspondientes a la media y la varianza real para ello es necesario obtener la media muestral (PROMEDIO) y la cuasivarianza muestral (VAR.S). A continuación, podemos obtener el percentil de la distribución t de Student con la función INV.T.2C y calculamos el error de muestreo multiplicando dicho percentil por la cuasidesviación típica y dividiendo por la raíz cuadrada del tamaño muestral. Los extremos del intervalo se obtienen sumando y restando al estimador de la media dicho error.

IC para la varianza o la desviación típica de una NORMAL

Tenemos una muestra de datos y queremos calcular el intervalo de confianza para la varianza real. Calculamos, en primer lugar, la estimación correspondiente a la varianza real para ello es necesario obtener la cuasivarianza muestral (VAR.S) y si es para la desviación típica, la cuasidesviación típica muestral (DESVEST.M). A continuación, podemos obtener los percentiles de la distribución Chi-cuadrado con la función INV.CHICUAD.CD (cola derecha) y INV.CHICUAD (cola izquierda). Por último, calculamos los extremos del intervalo dividiendo el producto de la cuasivarianza muestral por el tamaño muestral menos 1 entre cada percentil y, en el caso de la desviación típica, dividiendo por las raíces cuadradas de dichos percentiles.

2. Contraste de hipótesis

Un contraste de hipótesis es una regla de decisión para comprobar si los datos muestrales son compatibles o no con una hipótesis de trabajo planteada a priori. Para ello se construye un estadístico que mide la discrepancia entre los datos muestrales y la hipótesis planteada. Si dicha discrepancia es elevada se rechazará la hipótesis y, en caso contrario, seguiremos pensando que es cierta.

3. Cálculo del contraste de hipótesis para la media, la varianza y la proporción de una muestra dada

Tenemos una muestra de datos y queremos calcular el contraste de hipótesis para la media, varianza o proporción real. Calculamos, en primer lugar, las estimaciones correspondientes a la media, la varianza o proporción real para ello es necesario obtener la media muestral (PROMEDIO), la cuasivarianza muestral (VAR.S) o el cociente entre el número de éxitos (CONTAR.SI) y el tamaño muestral. A continuación, podemos calcular el estadístico de contraste y el percentil que genera la región crítica. Si el valor del estadístico pertenece a la región de aceptación entonces no podemos rechazar la hipótesis nula y seguimos creyendo en ella. Por el contrario, si el estadístico es un valor de la región crítica entonces tenemos evidencia para rechazar la hipótesis nula y crearemos en la alternativa.

Además, calcularemos el p-valor de cada contraste. Para ello, se obtiene la probabilidad de observar un valor del estadístico “peor”, en el sentido de la alternativa, que el proporcionado por la muestra. Se utilizará las funciones de distribución de las variables aleatorias de cada estadístico, bien sea normal (DISTR.NORM.ESTAND), t-Student (DISTR.T.N o DISTR.T.2C) o chi-cuadrado (DISTR.CHICUAD o DISTR.CHICUAD.CD). Interpretaremos el p-valor para

tomar la decisión de rechazar o no la hipótesis planteada. En caso de rechazar, explicaremos el sentido de nuestro rechazo y, por lo tanto, diremos la alternativa en la que creemos.

Las siguientes funciones están disponibles en EXCEL:

- **CONTAR.SI(rango valores; criterio)** cuenta el número de valores del rango que cumplen con el criterio dado.
- **DISTR.NORM.ESTAND(x)** proporciona la probabilidad de observar un valor menor o igual que x en una distribución normal estándar o tipificada, es decir, la probabilidad de la cola inferior de la distribución.
- **DISTR.T.N(x;g;acumulado)** proporciona el valor de la función de densidad (acumulado = 0) o de la función de distribución (acumulado=1) de una distribución t de Student con g grados de libertad en el punto x.
- **DISTR.T.CD (x;g)** proporciona la probabilidad, unilateral, de observar un valor mayor o igual a x en una distribución t de Student con g grados de libertad.
 - x: un valor entre 0 e infinito
 - Grados de libertad: parámetro de la distribución t de Student
- **DISTR.CHICUAD(x;g;acumulado)** proporciona el valor de la función de densidad (acumulado = 0) o de la función de distribución (acumulado=1) de una distribución chi-cuadrado con g grados de libertad en el punto x.
- **DISTR.CHICUAD.CD (x; grados de libertad)** devuelve la probabilidad de observar un valor mayor o igual que x en una distribución chi-cuadrado con grados de libertad, es decir, la probabilidad de la cola superior (derecha) de la distribución.
- **INV.CHICUAD (probabilidad; grados_de_libertad)** devuelve el percentil de la cola izquierda de la distribución chi-cuadrado con dichos grados de libertad.
 - Probabilidad: probabilidad fijada
 - Grados de libertad: Parámetro de la distribución Chi-cuadrado
- **INV.CHICUAD.CD (probabilidad; grados_de_libertad)** devuelve el percentil de la cola derecha de la distribución chi-cuadrado con dichos grados de libertad.
 - Probabilidad: probabilidad fijada
 - Grados de libertad: Parámetro de la distribución Chi-cuadrado
- **INV.NORM.ESTAND (probabilidad)** devuelve el inverso de la función de distribución de una $N(0,1)$, donde:
 - Probabilidad: probabilidad seleccionada
- **INV.T(probabilidad;grados de libertad)** devuelve el inverso de la distribución t de Student con grados de libertad que deja probabilidad en la cola izquierda de la distribución.
 - Probabilidad: probabilidad fijada
 - Grados de libertad: Parámetro de la distribución t-Student

- INV.T.2C (probabilidad; grados_de_libertad) devuelve el inverso de la distribución t de Student con grados de libertad que deja la mitad de probabilidad en cada cola de la distribución.
 - Probabilidad: probabilidad fijada
 - Grados de libertad: Parámetro de la distribución t-Student
- PROMEDIO(Rango): devuelve la media de los datos incluidos en el rango definido
- DESVEST.M (Rango): devuelve la cuasidesviación típica de los datos incluidos en el rango definido
- VAR.S (Rango): devuelve la cuasivarianza muestral de los datos incluidos en el rango definido

PROBLEMAS

Utilizando la hoja BANCO necesitamos calcular intervalos de confianza para varios parámetros de la población de clientes.

1. Calcula el intervalo de confianza para el salario medio de todos los clientes con una confianza del 95%.
2. Calcula el intervalo de confianza para la proporción de hombres que son clientes del banco con una confianza del 99%.
3. Calcula el intervalo de confianza para el salario medio de los hombres y de las mujeres, por separado, que son clientes del banco con una confianza del 98%.
4. Calcula el intervalo de confianza para la desviación típica del salario medio de los hombres y mujeres, por separado, que son clientes del banco con un nivel de confianza del 95%.
5. Calcula el intervalo de confianza del 95% para la proporción de clientes que tienen menos de 40 años

Utilizando la hoja BANCO plantear y resolver una serie de contrastes de hipótesis.

6. El año anterior el saldo medio en cuenta de un cliente era de 1342,45€, sin embargo se cree que ha disminuido debido a la crisis. Con un nivel de significación del 5%, ¿qué podemos afirmar sobre el saldo medio en cuenta de este año con respecto al anterior? Calcular el pvalor del contraste
7. Los medios de comunicación informan que se están produciendo mayores desigualdades económicas entre la población. Sabiendo que la desviación típica del saldo medio era de 385€ y con un nivel de significación del 3%, ¿se han producido mayores desigualdades económicas entre los clientes de la entidad bancaria? Calcular el pvalor del contraste
8. La tasa de clientes que tenían pagos pendientes ha sido en los últimos años del 2%. Con un nivel de significación del 5%, ¿ha aumentado dicha proporción entre los clientes de la entidad bancaria? Calcular el pvalor del contraste
9. Con un nivel de significación del 5%, ¿podemos afirmar que el saldo medio de las mujeres clientes del banco es igual a 1342,45€? Calcular el pvalor del contraste