SUPER NOTA

NOMBRE DEL ALUMNO: PEREYRA CALVO CAROL DENISSE

TEMA: UNIDAD IV. DEMOGRAFIA

PARCIAL: CUARTO.

MATERIA: BIOESTADISTICA I.

MAESTRO: ICEL BERNARDO LEPE ARRIAGA

LICENCIATURA: ENFERMERÍA.

CUATRIMESTRE: CUARTO.

Frontera Comalapa, Chiapas a 23 de noviembre del 2024

Test para poblaciones normales

Una población

En la inferencia sobre una variable numérica en una población

El objetivo principal de los test de hipótesis

Es contrastar el valor de alguna medida de posición (media o mediana), de dispersión (varianza) o de algún otro parámetro poblacional.

Así, si se cuenta con información muestral sobre el número de horas diarias que un individuo está viendo la televisión, trataremos de ver si podemos aceptar que el promedio de horas en la población toma un determinado valor.

En principio, si hacemos inferencia sobre un resumen de la variable podemos considerar que dicha variable sigue una distribución y plantear el problema desde la óptica paramétrica.

En principio, si hacemos inferencia sobre un resumen de la variable podemos considerar que dicha variable sigue una distribución y plantear el problema desde la óptica paramétrica.

Esa distribución puede ser cualquiera de las muchas distribuciones tipo (binomial, Poisson, normal, uniforme, gamma,).

Por distintos motivos (variables, originales o transformadas, con distribución más o menos campaniforme, teorema del límite central, simplicidad en los procedimientos, ...), lo más habitual es la suposición de normalidad.

Entonces, suponiendo que la variable X sigue una distribución normal de media y desviación, plantearemos contrastes de hipótesis sobre dichos parámetros.

En la inferencia sobre dos variables numéricas en una población se trata de ver si existe relación lineal entre las mismas a partir de la información muestral.

Podemos tratar de analizar si existe relación entre la renta y el consumo de las familias y cuál es la intensidad y el sentido de la misma, concretándose la inferencia en el coeficiente de correlación.

Dos poblaciones

El objetivo fundamental es comparar la distribución de una variable cuantitativa X en las dos poblaciones (subpoblaciones)

Determinadas por las modalidades de una característica cualitativa dicotómica o, lo que es lo mismo, estudiar si la variable cuantitativa (variable respuesta) presenta diferencias significativas en cada uno de los dos niveles de la variable cualitativa (factor).

Esta comparación se realizará a partir de la información parcial que proporcionan dos muestras.

Si podemos aceptar normalidad, el objetivo general de comparar dos poblaciones se traduce en comparar las medias de la variable en cada una de ellas

Aunque suele tener únicamente un interés instrumental, también se pueden comparar las varianzas.

Para poder realizar estas comparaciones se utilizan dos muestras provenientes de individuos diferentes (estudiar las posibles diferencias salariales en función del sexo a partir de una muestra de hombres y una muestra de mujeres)

En este caso hablaremos de muestras independientes.

A veces se puede utilizar una muestra con los mismos individuos para las dos situaciones de la variable

A veces se puede utilizar una muestra con los mismos individuos para las dos situaciones de la variable en este caso hablaremos de muestras apareadas o relacionadas.

Siempre que se puedan considerar muestras relacionadas, este procedimiento proporciona en principio mejores inferencias.

El problema de comparar medias se puede generalizar a más de dos poblaciones, los llamados procedimientos ANOVA.

Poblaciones normales	Una población	Una variable	C. de H. sobre la media poblacional
			C. de H. sobre la varianza poblacional
		Dos variables	C. de H. sobre el coeficiente de correlación poblacional
	Dos poblaciones	Muestras independientes	C. de H. sobre la diferencia de medias poblacionales
			C. de H. sobre el cociente de varianzas poblacionales
		Muestras relacionadas	C. de H. sobre la diferencia de medias poblacionales

Este tipo de análisis, denominado genéricamente de correlación, conduce de forma inmediata al análisis de regresión: si existe relación trataremos de encontrar la función que mejor exprese esta relación.

Test para poblaciones binomiales y de Poisson

Nos encontramos con un modelo derivado de un proceso experimental puro, en el que se plantean las siguientes circunstancias

Se realiza un número n de pruebas (separadas o separables).

Cada prueba puede dar dos únicos resultados A y Ã.

La probabilidad de obtener un resultado A es p y la de obtener un resultado Ã es q, con q= 1-p, en todas las pruebas.

Esto implica que las pruebas se realizan exactamente en las mismas condiciones y son, por tanto, independientes en sus resultados.

Si se trata de extracciones, (muestreo), las extracciones deberán ser con devolución (reemplazamiento), o bien población grande (M.A.S).

El parámetro n será el número de extracciones (tamaño muestral) y el parámetro p la proporción de individuos de la población que poseen la característica en cuestión.

Se ha comentado que para que la probabilidad, de que en cada extracción obtengamos un individuo poseedor de la característica sea constante en todas las pruebas es necesario que las proporciones poblacionales no cambien tras cada extracción es decir se reemplace cada individuo extraído.

Sin embargo, si la población es muy grande, aunque no reemplacemos los individuos extraídos las variaciones en las proporciones de la población restante serán muy pequeñas y, aunque de hecho las probabilidades de, obtener un éxito varíen tras cada prueba, esta variación será muy pequeña y podremos considerar que son constantes.

En estadística, el test binomial es un test exacto de la significación estadística de las desviaciones de una teóricamente distribución esperada de las observaciones en dos categorías.

USO COMÚN

Un uso común del test binomial es en el caso donde la hipótesis nula es aquella en la que las dos categorías son igualmente probables de que ocurran (como el lanzamiento de una moneda).

Las tablas están ampliamente disponibles para dar la importancia observada en el número de observaciones en las categorías para este caso.

Sin embargo, como muestra el siguiente ejemplo, el test binomial no se limita a este caso.

Donde hay más de dos categorías, y un test exacto es requerido, el test multinomial, basado en la distribución multinomial, debe usarse en vez del test binomial.

MUESTRAS GRANDES

Un uso común del test binomial es en el caso donde la hipótesis nula es aquella en la que las dos categorías son igualmente probables de que ocurran (como el lanzamiento de una moneda).

Prueba χ^2 de Pearson y el G-test. Sin embargo, para pequeñas muestras estas aproximaciones se descomponen, y no hay alternativa para el test binomial.

La distribución de Poisson se aplica a varios fenómenos discretos de la naturaleza (esto es, aquellos fenómenos que ocurren 0, 1, 2, 3... veces durante un periodo definido de tiempo o en un área determinada) cuando la probabilidad de ocurrencia del fenómeno es constante en el tiempo o el espacio.

Ejemplos de estos eventos que pueden ser modelados por la distribución de Poisson incluyen:

- El número de autos que pasan a través de un cierto punto en una ruta (suficientemente distantes de los semáforos) durante un periodo definido de tiempo.
- El número de errores de ortografía que uno comete al escribir una única página.
- El número de llamadas telefónicas en una central telefónica por minuto.
- El número de servidores web accedidos por minuto.
- El número de animales muertos encontrados por unidad de longitud de ruta.
- El número de mutaciones de determinada cadena de ADN después de cierta cantidad de radiación.

Test basado en el estadístico Chi-cuadrado

Esta prueba puede utilizarse incluso con datos medibles en una escala nominal.

La hipótesis nula de la prueba Chi-cuadrado postula una distribución de probabilidad totalmente especificada como el modelo matemático de la población que ha generado la muestra.

Para realizar este contraste se disponen los datos en una tabla de frecuencias.

Para cada valor o intervalo de valores se indica la frecuencia absoluta observada o empírica (Oi).

A continuación, y suponiendo que la hipótesis nula es cierta, se calculan para cada valor o intervalo de valores la frecuencia absoluta que cabría esperar o frecuencia esperada (Ei=n·pi, donde n es el tamaño de la muestra y pi la probabilidad del i-ésimo valor o intervalo de valores según la hipótesis nula).

El estadístico de prueba se basa en las diferencias entre la Oi y Ei y se define como:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}.$$

Este estadístico tiene una distribución Chi-cuadrado con k-1 grados de libertad si n es suficientemente grande, es decir, si todas las frecuencias esperadas son mayores que 5. En la práctica se tolera un máximo del 20% de frecuencias inferiores a 5.

Si existe concordancia perfecta entre las frecuencias observadas y las esperadas el estadístico tomará un valor igual a 0

Por el contrario, si existe una gran discrepancia entre estas frecuencias el estadístico tomará un valor grande y, en consecuencia, se rechazará la hipótesis nula.

Así pues, la región crítica estará situada en el extremo superior de la distribución Chi-cuadrado con k-1 grados de libertad. Para realizar un contraste Chi-cuadrado la secuencia es:

 Analizar

 Pruebas no paramétricas

 Chi-cuadrado

En Valores esperados se debe especificar la distribución teórica activando una de las dos alternativas.

Por defecto está activada Todas las categorías iguales que recoge la hipótesis de que la distribución de la población es uniforme discreto.

La opción Valores requiere especificar uno a uno los valores esperados de las frecuencias relativas o absolutas correspondientes a cada categoría, introduciéndolos en el mismo orden en el que se han definido las categorías.

El recuadro Rango esperado presenta dos opciones: por defecto está activada Obtener de los datos que realiza el análisis para todas las categorías o valores de la variable

La otra alternativa, Usar rango especificado, realiza el análisis sólo para un determinado rango de valores cuyos límites Inferior y Superior se deben especificar en los recuadros de texto correspondientes.

El cuadro de diálogo al que se accede con el botón Opciones ofrece la posibilidad de calcular los Estadísticos Descriptivos y/o los Cuartiles, así como seleccionar la forma en que se desea tratar los valores perdidos.

Test de bondad de ajuste.

La bondad de ajuste de un modelo estadístico describe lo bien que se ajusta un conjunto de observaciones.

La bondad de ajuste de un modelo estadístico describe lo bien que se ajusta un conjunto de observaciones.

Las medidas de bondad en general resumen la discrepancia entre los valores observados y los valores esperados en el modelo de estudio.

Estas pruebas permiten verificar que la población de la cual proviene una muestra tiene una distribución especificada o supuesta.

Sea X: variable aleatoria poblacional f0 (x) la distribución (o densidad) de probabilidad especificada o supuesta para X.

Se desea probar la hipótesis: Ho: f(x) = f0 (x)

En contraste con la hipótesis alterna: Ha: f(x) no= f0 (x) (negación de Ho)

Prueba de bondad de ajuste chi cuadrado x2

El procedimiento de la prueba requiere una muestra aleatoria de tamaño n proveniente de la población cuya distribución de probabilidad es desconocida.

Estas n observaciones se pueden distribuir en k intervalos de clases y pueden ser representadas en histogramas.

La prueba se puede utilizar tanto para distribuciones discretas como para distribuciones continuas.

La prueba se puede sintetizar en los siguientes pasos.

1. Se colocan los n datos históricos (muéstrales) en una tabla de frecuencia.

2. Se propone una distribución de probabilidad una distribución de probabilidad de acuerdo con la tabla de frecuencia o con la curva que muestre un histograma o polígono de frecuencia.

3. Con la distribución propuesta, se calcula la frecuencia esperada para cada uno de los intervalos (FEi) de la siguiente manera:

- Si la variable es continua se utiliza de modelo matemático de la distribución propuesta y se evalúan todas las categorías y luego se multiplica por el número total de datos.
- Si la variable es continua se halla mediante la integración de la distribución propuesta y luego se multiplica por el número total de datos.

4. Se calcula el estadístico de prueba

5. Si el estimador C es menor o igual al valor correspondiente X2 con m-k-1 grados de libertad (K= números de parámetros estimados de la distribución propuesta estimada por los estadísticos muéstrales) y a un nivel de confiabilidad de 1-α, entonces no se puede rechazar la hipótesis de que los datos siguen la distribución que se propuso.