ESTADISTICA II: TEORIA Y PRACTICAS

Table of Contents

[2 REPASO 2](#_Toc138639177)

[2.1 CUANTILES Y PERCENTILES 2](#_Toc138639178)

[2.2 MUESTRAS ALEATORIAS 3](#_Toc138639179)

[2.3 LEY DE GRANDES NUMEROS 5](#_Toc138639180)

[2.4 ESTANDARIZACION 5](#_Toc138639181)

[2.5 TEOREMA CENTRAL DEL LIMITE 5](#_Toc138639182)

[2.6 INTERVALOS DE CONFIANZA 7](#_Toc138639183)

[3 ESTIMACIÓN DE PARÁMETROS 11](#_Toc138639184)

[3.1 MOMENTOS 13](#_Toc138639185)

[3.2 MAXIMA VEROSIMILITUD 18](#_Toc138639186)

[3.3 INFORMACION DE FISHER 24](#_Toc138639187)

[3.4 COTA DE CRAMER-RAO 29](#_Toc138639188)

[3.5 DISTRIBUCION ASINTOTICA DE EMV 33](#_Toc138639189)

[3.6 ERROR CUADRÁTICO MEDIO Y SESGO 34](#_Toc138639190)

[3.7 CONSISTENCIA 42](#_Toc138639191)

[3.8 INTERVALOS DE CONFIANZA EXACTOS 42](#_Toc138639192)

[4 TEST DE HIPÓTESIS 51](#_Toc138639193)

[4.1 PLANTEO DEL PROBLEMA 51](#_Toc138639194)

[4.2 TESTS COMUNES 58](#_Toc138639195)

[4.3 P-VALOR 64](#_Toc138639196)

[4.4 POTENCIA Y REGION DE RECHAZO 65](#_Toc138639197)

[4.5 DUALIDAD 69](#_Toc138639198)

[4.6 COCIENTE DE VEROSIMILITUD (TEST DE NEYMAN-PEARSON) 70](#_Toc138639199)

[4.7 COCIENTE DE VEROSIMILITUD GENERAL 72](#_Toc138639200)

[4.8 COMPARAR DOS MUESTRAS NORMALES 81](#_Toc138639201)

[4.9 TEST NO-PARAMETRICO (WILCOXON-MANN-WHITNEY) 84](#_Toc138639202)

[4.10 TEST CON DATOS CATEGORICOS 89](#_Toc138639203)

[4.11 HOMOGENEIDAD 91](#_Toc138639204)

[5 PRACTICAS 94](#_Toc138639205)

[5.1 GUIA 4 EMV-MOMENTOS 94](#_Toc138639206)

[5.2 GUIA 5 INTERVALOS DE CONFIANZA 108](#_Toc138639207)

[5.3 GUIA 6 TEST DE HIPÓTESIS 129](#_Toc138639208)

[5.4 GUIA 7 COMPARAR DOS MUESTRAS 143](#_Toc138639209)

[5.5 GUIA 8 DATOS CATEGORICOS 163](#_Toc138639210)

# REPASO

## CUANTILES Y PERCENTILES

El *p*-ésimo cuantil de X es el valor tal que . Por ejemplo, la mediana es

En general se escribe (usando la acumulada). Con una densidad continua, el cuantil es el valor tal que haya un area a la izquierda de y un area a la derecha de .

Por ejemplo, para una uniforme , el cuantil 0.6 sería:

A picture containing text, diagram, line, font

Description automatically generated

A la izquierda está la densidad y a la derecha la acumulada.

Para una normal, el cuantil 0.6, sería:

A picture containing diagram, plot, line, slope

Description automatically generated

Los cuantiles son una medida de la **ubicación** de una variable aleatoria.

Los cuantiles a veces se describen en términos de percentiles. Por ejemplo, el 60mo percentil es lo mismo que el cuantil 0.6, así que decir que sos más alto que el 60% de la población indica que la probabilidad de que seas más alto que una persona elegida al azar de la población es 60%.

## MUESTRAS ALEATORIAS

La estadística consiste en analizar **muestras** tomadas de **poblaciones** y la idea es que la muestra que agarramos para describir a la población sea una **muestra aleatoria**. (que sea representativa de la poblacion que se quiere describir).

Por ejemplo, si queres saber como esta distribuida la altura en una escuela de 1000 personas (la poblacion), no estaría bueno seleccionar tu muestra de los 40 alumnos que juegan al basquet; si seleccionas como muestra a los que juegan al basquet y con esa muestra queres describir la poblacion total no vas a llegar a algo cierto sobre la poblacion.

La muestra debería ser aleatoria en el sentido de que no seleccionas elementos de la población que tengan características específicas o particulares, sino que cada elemento de la población tiene la misma probabilidad de ser seleccionado para formar parte de la muestra.

De esta forma, se garantiza que la muestra sea representativa de la población completa, evitando sesgos en las conclusiones o inferencias que se obtengan de la muestra. Entonces, una muestra aleatoria quiere decir que cada individuo tiene la misma probabilidad de ser elegido para la muestra.

La idea también es lograr una muestra en la que cada medición es independiente, porque asi tenemos la mayor cantidad posible sobre la población. Esto es los mismo que decíamos antes: la seleccion de un elemento de la población para tenerlo en la muestra no afecta la selección de ningun otro elemento de la población.

De las muestras, se pueden obtener **resumenes**. Por ejemplo, la media de los datos, la varianza de los datos, etc...

Un resumen (como la media o varianza) es una función de la muestra porque toma los valores de la muestra y devuelve algun resultado *y*. A esto se le llama un **estadístico**. Un estadístico es una función de una muestra. Los resumenes, por ejemplo, de la muestra son estadisticos.

Si se calcula la media, por ejemplo, de una muestra de 10 números que sale de una distribución cualquiera, el valor que obtenga no será exactamente igual a la media de la población; por ahí un poco más alto o un poco más bajo. Si tomamos muestras de 10 elementos una y otra vez (calculando la media de cada una), encontramos que algunas medias muestrales se acercan mucho más a la media poblacional que otras. Algunos serían más altos que la media de la población y otros serían más bajos.

Ahora tomamos muestras de 10 números y les calcular la media una y otra vez, digamos unas 1000 veces, y luego a esas 1000 medias las ponemos en un grafico, como un histograma. Esta distribución de las medias es una muy buena aproximación a la **distribución muestral de la media.** La distribución muestral de la media es la distribución teórica a la que se aproxima el grafico mientras tomamos más muestras. A medida que el número de muestras se acerca al infinito, la distribución de frecuencias relativas se acerca a la distribución muestral (de la media).

La distribución muestral de la media para una muestra de tamaño 10 era solo un ejemplo; *hay una distribución muestral diferente para cada tamaño de muestra*. Una cosa es el **número de muestras** y otra el **tamaño de la muestra**. El número de muestras es la cantidad de muestras del mismo tamaño que se estan usando. La distribución muestral se la aproxima aumentando el numero de muestras, pero cada tamaño tiene su distribución.

Lo que toca ver es que pasa si el número o el tamaño de las muestras son grandes. El número de muestras tiene que ser grande para acercarnos a una distribución muestral graficamente. Pero lo interesante es cuando tenemos muestras de gran tamaño, aunque no sean muchas. En esos casos se llega a las conclusiones de la ley de grandes numeros y el teorema central del limite.

## LEY DE GRANDES NUMEROS

Intuitivamente, el promedio de varias mediciones tiende a ser mejor que una sola medición. Lo que dice la ley de grandes numeros es que el promedio de muchas muestras independientes es muy parecido a la media de la distribución subyacente.

En otras palabras, si tenemos muchas observaciones que salen de una población, el promedio de estas observaciones deberían ser parecidas a la media de la población.

Por ejemplo, si lanzás una moneda (donde las probabilidades de obtener cara o cruz son iguales, 0.5) 5 veces, la proporción de caras (por ejemplo) no necesariamente será 0.5. Sin embargo, si tirás la moneda muchas más veces, digamos 1,000 o 10,000 veces, la proporción de caras se acercará cada vez más a 0.5 del total. Esto no significa que obtendrás exactamente 500 caras en 1,000 lanzamientos o 5,000 caras en 10,000 lanzamientos, pero sí significa que la proporción de caras se acercará más a 0.5 cuanto más veces lances la moneda.

## ESTANDARIZACION

Si a una variable aleatoria se le resta su media y esa diferencia se divide por el desvío de esa variable, queda una variable aleatoria (Z) de esperanza 0 y desvío 1, **para cualquier distribución.**

## TEOREMA CENTRAL DEL LIMITE

Tomemos una muestra i.i.d. que salen de una distribución . La media es

Ahora, es una variable aleatoria. tiene su esperanza y su varianza .

Podemos saber la esperanza de y la varianza de con eso

Entonces la esperanza del promedio es igual a la esperanza de la distribucion de la que salen las muestras.

Luego,

Lo que dice el teorema central del límite es que:

Es decir, la distribución de converge en :

Resumiendo,

Con ., son variables aleatorias con esperanza y desvío . Definimos

Entonces,

Y también, porque es múltiplo de , se puede estandarizar

Además la varianza muestral es

Podemos usarla en vez de la varianza poblacional y

Si en vez de la varianza poblacional usamos la varianza muestral,Y

Y podemos demostrar que, se cumple que

Porque recordamos que la chi es la suma *n* normales estandar de .

## INTERVALOS DE CONFIANZA

Imaginemos que tenemos una fábrica de cajas de cereales. Todos los días, las máquinas llenan las cajas con cereales, pero no siempre la misma cantidad exacta, por ende, todas las cajas terminan teniendo un peso un poco diferente. Quisiéramos estimar cual es el promedio del peso de todas las cajas que se llenan un día (nuestra población). Tomamos una muestra de 100 cajas que se llenaron ese día y la media muestral de los pesos de esas 100 cajas nos da , mientras que el desvío estándar muestral nos da .

La media de la población no va a ser exactamente igual a , pero lo que podríamos hacer es construir un **intervalo de confianza** alrededor de 12.05 que probablemente contenga la media poblacional real . También quisiéramos calificar el nivel de seguridad que tenemos de que la media real está en ese intervalo.

Para ver cómo construir un intervalo de confianza, llamamos a la media poblacional verdadera y al desvío estándar verdadero. La muestra aleatoria son los 100 pesos distintos que obtuvimos de las 100 cajas que agarramos como muestras.

Teníamos como la media muestral. Como es la media de una muestra grande de 100, el teorema central del límite nos dice que esa media muestral viene de una distribución normal con media = (la media poblacional) y con desvío estándar . Es decir, las posibles medias de las muestras que podremos tomar de nuestra población salen de una distribución normal como la de la imagen, donde la media de esa normal es la media de la población que estamos estudiando.

A picture containing plot, line, diagram

Description automatically generated

En esta imagen vemos una distribución normal con media y tenemos una región en el centro, excluyendo dos colas (ya que el 95% de los datos de una normal están a 1.96 desvíos estándar de ambos lados). Esto significa que el 95% de los valores que tomemos de la normal van a estar en esa región. Uno de los posibles valores que puede tomar esa distribución de las medias es la esperanza que efectivamente encontramos, 12.05, marcada en el gráfico con . Pero la realidad es que esa es solo un posible lugar del que el numero 12.05 podría haber salido. El 95% de las distintas medias muestrales del mismo tamaño que hubiéramos tomado habrían caído en esa región. Debajo de la normal hay una recta centrada en que tiene exactamente la misma longitud que el 95% central de la normal. *Ese* es un **intervalo de confianza de nivel 95%**. Ese intervalo efectivamente cubre la media verdadera.

En esta otra imagen tenemos una situación diferente:

A picture containing plot, line, diagram, text

Description automatically generatedEn este caso, la media de nuestra muestra justo cayó fuera del intervalo de 95% de la normal (lo cual tiene una chance de suceder del 5%). El intervalo que construiríamos entonces *no*cubriría la media verdadera de la población. Ósea que, si queremos construir un intervalo de confianza al 95%, hay un 5% de posibilidades de que el intervalo que terminemos construyendo *no* contenga a la media poblacional.

El desvío estándar de las medias muestrales es desconocido (porque el desvío de la población es desconocido). En este caso, tenemos que el desvío muestral es . Usamos ese desvío para aproximar el poblacional y construimos un intervalo de confianza para la media muestral del peso de las cajas de cereales entre () o (). Estamos seguros al 95% de que la media verdadera está en ese intervalo.

Veamos como hallar un intervalo de confianza con cualquier nivel de confianza (). Aca abajo tenemos una normal que describe medias muestrales:A picture containing diagram, line, plot, slope

Description automatically generated

Lo que hacemos es definimos como el cuantil que deja debajo de cada cola. Por ejemplo, queremos un nivel de confianza del 95%, por ende el cuantil que deja 0.025 debajo de la curva es . Similarmente, 1.96 deja 0.025 debajo de la curva hacia la derecha. El área en el medio es (). Por ende, el intervalo () va a cubrir la media para una proporción de todas las muestras que podrían tomarse.

# ESTIMACIÓN DE PARÁMETROS

En **probabilidad**, teníamos una distribución (su función de densidad) y podíamos analizar las propiedades de las variables aleatorias que salían de esa distribución. Por ejemplo, sabiendo la función de la distribución sacábamos la esperanza de una variable aleatoria haciendo: . O, para sacar la varianza, hacíamos

En **estadística**, en cambio, tenemos valores de variables aleatorias que salen de alguna distribución, pero no sabemos cuál es esa distribución, o no sabemos alguno de sus parámetros.

Por ejemplo, la altura de las personas es una variable aleatoria pero no sabemos la función de densidad que genera esos valores. En ese caso, si decimos que salen de una distribución normal, entonces Pero aún así no sabemos completamente la distribución, porque son desconocidos; lo que tenemos es una muestra de datos que salen de esa distribución.

Los parámetros desconocidos los llamamos que salen del conjunto (el espacio paramétrico), ósea son todos los valores que podría tomar . Por ejemplo, si supiéramos que nuestros datos salen de una exponencial, tenemos un parámetro que no conocemos, pero sí sabemos es que obligatoriamente van a ser valores positivos y serían todos los reales positivos.

La **inferencia** estadística consiste en resolver el problema estadístico de tener datos, pero no exactamente la distribución que los genera. Teniendo una muestra aleatoria i.i.d. que sale de y cuyo parámetro no conocemos, los problemas estadísticos son:

* Estimar el parámetro: estimar el valor de la distribución y la calidad de la estimación.
* Estimar intervalos: con cierta confianza obtener un intervalo que contiene el parámetro
* Testear hipótesis: testear si una afirmación sobre el parámetro es o no es verdadera

Entonces, con i.i.d. una muestra aleatoria que sale de cuyo parámetro sale de y no es conocido; el valor de debería estimarse a partir de la muestra.

Un **estimador** del parámetro basado en la muestra aleatoria *es una función* de la muestra: . Esa funcion devuelve el valor estimado del parámetro para esos valores de la muestra. **Un estimador *es* una variable aleatoria.**

Ósea, si tenemos los valores como muestra, entonces al evaluarlos en el estimador vamos a obtener . Y si tenemos los valores , que son una muestra de la misma distribución y los evaluamos en el estimador, vamos a obtener *otro* valor de . Esto es importante. Distintas muestras van a dar distintos valores cuando los metamos en el estimador. Cada valor que devuelva es una **estimación** del parámetro. Si vemos la estimación que salió de los datos y la que salió de otros datos , estas dos estimaciones pueden ser iguales, parecidas o nada que ver una con la otra. Eso va a depender de cuantos datos haya y de que tan bueno sea el estimador

Nuestro objetivo es acercarnos lo más posible al parámetro real, entonces lo mejor sería que con distintas muestras nos dé más o menos la misma estimación del parámetro. Para que esto pase, se necesitan

* Muchas muestras, porque mientras más datos, más va a ser representativa de la población total.
* Un buen estimador (insesgado, eficiente)

Estimadores comunes:

* Momentos
* Máxima verosimilitud

otros:

* Estimación de Bayes con prior y posterior probs.
* Menores cuadrados
* QMLE
* Variables Instrumentales

## MOMENTOS

Este no es un buen estimador, pero es fácil de hacer. Si tenemos el momento **poblacional** es:

(y si es discreta, la suma en vez de la integral)

Y con muestras que salen de esa distribución , el momento **muestral**  es:

Entonces, es el **estimador** **de .**

Por la ley de grandes números, cuando *n* es muy grande.

Si hacemos , se puede despejar el parámetro y se obtiene una función que es el **estimador del parámetro**.

### DENSIDAD PUNTUAL

Supongamos que X es una variable aleatoria discreta con función de densidad:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| X | 0 | 1 | 2 | 3 |
|  |  |  |  |  |

Y el parámetro .

Además, tenemos la muestra: (3,0,2,1,3,2,1,0,2,1). Usamos el método de momentos:

* 1. La media poblacional teórica es
  2. La media muestral es:
  3. Resolvemos la ecuación despejando que surge de igualar ,

### LAPLACE

Con una muestra i.i.d con densidad

Si calculamos el momento teórico de primer orden (la esperanza), nos da:

Y si tratáramos de resolver , no llegaríamos a nada porque no contiene el parámetro desconocido. En cambio, si calculamos el segundo momento poblacional,

Observamos que la función que teníamos () es la distribución de Laplace con parámetros . Laplace es simétrica alrededor de su media (su parámetro ). Por ende, expresa lo mismo que (o que , pero en este caso no nos ayuda mucho tener la parte negativa). Ahora bien, lo que estamos haciendo al sacar el segundo momento es multiplicarla por , que también es simétrica alrededor del origen. Si multiplicamos una función simétrica alrededor de algún eje por otra que también es simétrica alrededor de ese mismo eje, el producto resultante es, también, simétrico alrededor del eje.

El sale de la integral porque es una constante y sabemos que el en el intervalo positivo que estamos integrando es siempre nada más que . También sacamos el que no depende de *x*.

Ahora, notamos que nos queda una función Gamma en la integral:

Sabiendo que , lo igualamos al momento muestral del mismo orden. El momento de segundo orden de una muestra:

Resolviendo obtenemos que el estimador de :

### UNIFORME

Con una muestra aleatoria de la distribución uniforme , con . EL primer momento teórico es:

El estimador va a ser

### NORMAL y

Usamos el método de momentos con una muestra aleatoria para estimar los parámetros y la normal.

Los momentos poblacionales teóricos son:

Ya sabemos la normal tiene de parámetro su esperanza.

Este momento lo podemos obtener sabiendo que para la normal y , entonces , por ende,

Ahora tenemos el sistema de ecuaciones para y

El estimador de va a ser la esperanza muestral.

Podemos reemplazarlo en la segunda ecuación y tenemos:

Para hacer esto, veamos primero el cuadrado de la suma

Si dividimos todo por , nos queda el cuadrado del promedio

Entonces, podemos escribir

Volviendo a la ecuación,

Podemos sumar y restar , tenemos:

Notamos que

Entonces,

Y ahí tenemos un cuadrado de un binomio

Es la varianza. Ahí tenemos nuestro estimador

Dejamos demostrado también que

## MAXIMA VEROSIMILITUD

Este estimador es más acertado que el de momentos y cuando *n* es grande es casi seguro que va a dar un estimador muy preciso.

Como asumimos que la muestra aleatoria es independiente, la distribución conjunta (la probabilidad de que se den esos valores de la muestra todos juntos) es el producto de las probabilidades individuales.

Esa función conjunta se llama función de verosimilitud

La idea es que la probabilidad de que justo se den todos esos valores de la muestra en conjunto va a darse en algún valor particular del parámetro, digamos , y para otros valores de , es bajo. Ósea queremos encontrar el valor de que maximiza la .

Además, es lo mismo tomar el log en cualquier base, porque es aplicar una transformación creciente y el máximo va a estar en el mismo lugar del dominio.

Y el máximo va a estar en

### DENSIDAD PUNTUAL

Supongamos que X es una variable aleatoria discreta con función de densidad:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| X | 0 | 1 | 2 | 3 |
|  |  |  |  |  |

Y el parámetro .

Además, tenemos la muestra: (3,0,2,1,3,2,1,0,2,1). La estimación de máxima verosimilitud de va a ser así:

Primero obtenemos la , que al asumir que son muestras independientes, definimos como el producto de la probabilidad de que ocurran esas muestras:

Reemplazamos con las probabilidades que conocemos:

Para que sea fácil de maximizar, tomamos la .

Con como una constante que no depende de así que no nos importa su valor. Esta función  es mucho más fácil de maximizar (derivar e igualar a 0)

Entonces nuestro (distinto del método de momentos)

### LAPLACE

Con una muestra i.i.d con densidad

La es

Igualamos la derivada a 0

Y entonces el estimador de nos queda

### NORMAL y

Estimemos ahora con la máxima verosimilitud una muestra normal. Queremos estimar los parámetros y .

Ahora, tenemos un vector de parámetros, . Escribimos la .

Ahora toca igualar el gradiente a 0:

Despejamos y obtenemos que

Como, estamos restando veces , entonces,

Y para ,

Entonces,

### PARETO

Distribución de Pareto,

Asumiendo que es dado, estimamos :

### UNIFORME

Con una muestra aleatoria distribuida normalmente en el intervalo con .

La densidad sería:

Y la función de verosimilitud sería

El estimador de máxima verosimilitud debe ser un valor de tal que Además debe ser el valor que maximiza entre todos los que cumplen . es una función decreciente en (es decir que a mayormenor va a ser el valor de ), entonces el estimador debería ser el menor valor posible deque sea mayor que cualquiera de los valores de la muestra. Para ser mayor que cualquiera de la muestra, debe ser el máximo entre los valores de la muestra y no va a ser mayor que ese valor. Por esto,

El problema es que, por el dominio de la función , sabemos que aun el mayor valor de la muestra debe ser menor que el parámetro, por lo que siempre va a subestimar el verdadero valor del parámetro.

### UNFORME (EXCEPCIÓN)

Con una muestra aleatoria distribuida normalmente en el intervalo con .

Supongamos que ahora escribimos la función:

En un caso como este no existe un estimador de máxima verosimilitud. Esto se debe a que si usamos la misma condición de antes , esto no es posible porque el parámetro debe ser siempre estrictamente mayor a los valores de .

## INFORMACION DE FISHER

En la estimación de parámetros, se quiere obtener información del parámetro de una distribución usando una muestra de datos que salen de esa distribución.

Para eso hallamos estimadores, ya sea con el método de momentos, con máxima verosimilitud o intervalos de confianza.

Intuitivamente, si un evento cualquiera tiene probabilidad baja, observar que ocurre ese evento nos da mucha información, porque es un evento raro. Entonces, para , si es el valor verdadero del parámetro, la función de verosimilitud (la probabilidad conjunta de ocurran la muestra) toma un valor muy grande (el máximo), y su derivada, es cercana a 0.

Asumamos que tenemos **una sola observación**,

Por regla de la cadena,

Lo que esperamos es que sea cercana a 0, no es un evento raro así que eso no nos da información. Ahora nos va a estar interesando que **no** sea 0. Si eso pasa (, y por ende (al cuadrado para que sea positiva) es grande, entonces la variable aleatoria nos da mucha información sobre .

Es decir, los puntos donde la no es 0 nos dicen que hay que ajustar para mejorar la verosimilitud. El gradiente de la nos va a ir diciendo, en esos puntos, que tan lejos estamos y como acercarnos al máximo.

Entonces usamos para que nos dé cuánta información sobre X tenemos.

Dicho aún de otra forma, expresa que tan sensible es a cambios en . Si es grande (ósea, tenemos mucha información de ), quiere decir que incluso un cambio chico en el valor de va a causar un cambio grande en la probabilidad conjunta de observar esos datos de la muestra. El hecho de que cambie tanto la verosimilitud al moverse un poco por los valores de indica que podemos hallar el valor exacto de más precisamente que si la medida de información que tenemos () fuese menor.

La es una variable aleatoria[[1]](#footnote-2), por ende, quisiéramos considerar qué pasa en el caso promedio, la esperanza. Es decir, el efecto promedio de cambios en sobre la verosimilitud.

La información de Fisher se puede obtener con tres fórmulas, la primera es esta:

1)

Esto sigue era lo que dijimos que buscábamos: el efecto promedio de cambios en el parámetro sobre la verosimilitud.

Ahora, analizamos en profundidad la fórmula y obtendremos las otras maneras de calcularla.

Por la definición de esperanza, si , entonces . Tenemos que:

Ahora, cambiando el orden de diferenciación:

Esto es porque la integral sobre la densidad () es 1, y la derivada de 1 es 0.

Y también, con la misma lógica de cambiar el orden de diferenciación:

Por ende, la esperanza de la derivada de la es 0.

De esta forma, llegamos a la segunda forma de calcular la :

2)

Además, se puede ver también que:

Por la regla de la cadena,

Así que:

3)

NOTA:

La información de Fisher de una muestra aleatoria de tamaño *n* es simplemente *n* veces la información de Fisher de una única observación.

### BERNOULLI

Supongamos que tenemos una variable aleatoria X con distribución Bernoulli cuyo parámetro es desconocido y . Vamos a determinar la información de Fisher de X.

La función de densidad puntual es:

Y x puede ser 0 o puede ser 1.

Entonces,

Y además,

, entonces la información de Fisher nos queda:

### NORMAL DESCONOCIDO y CONOCIDO

Supongamos que tenemos una con desconocido, pero dado. Hallar la información de Fisher de X.

Teníamos que

Y su era

Por ende,

Así, nos queda que

NOTA:

Si hacemos una transformación del parámetro, nos va a dar diferentes expresiones de la información de Fisher con diferente parametrización. En particular, digamos que X es una variable aleatoria con distribución , y su valor del parámetro es desconocido, pero debe pertenecer al espacio paramétrico . Si es la información de Fisher de X, supongamos que el parámetro es reemplazado por otro, y , siendo una función diferenciable. Ahora, si es la información de Fisher cuando el parámetro es , tenemos que

## COTA DE CRAMER-RAO

Nuevamente tomamos una muestra aleatoria proveniente de una distribución con parámetro desconocido. Con el número de información de Fisher, se puede obtener un límite de la varianza del estimador del parámetro.

Con un estimador del parámetro , que es un estadístico (una función de la muestra). La esperanza del valor de  es . Su varianza es un valor finito. Considerando variable aleatoria y sabiendo que , si observamos la covarianza entre la primera derivada de la y el estimador, recordamos que

o también

Entonces:

Como es la esperanza del estimador y el estimador es una variable aleatoria, entonces es un valor fijo y sale de la esperanza en es termino, quedando: , que es 0 porque y queda:

Esta esperanza es la integral con respecto a todas las variables

Porque se puede tratar como constante (porque tomamos los valores de la muestra como dados) y sale de la integral.

Recordamos la desigualdad de Cauchy-Schwartz:

o también

Como la información de Fisher era la varianza de la primera derivada de la .

O como acabamos de ver que

Ahora sí, nos queda que el límite inferior de la varianza de un estimador es

Y mientras mayor sea la información menor va a ser la varianza del estimador y va a aumentar la calidad del estimador. Si fuese insesgado el estimador, es decir que , su derivada sería 1 y quedaría:

Entonces, bajo ciertas condiciones, no hay otro estimador insesgado de parámetro basado en una muestra i.i.d. de tamaño que tenga varianza menor a ese límite inferior.

### NORMAL CONOCIDO y DESCONOCIDO

Supongamos una muestra que viene un de una distribución normal , con conocido y con varianza desconocida. Calcular el límite inferior de varianza par cualquier estimador y comparar con la varianza muestral .

Teniendo que la distribución es

Entonces,

Entonces,

Y también,

Y el número de información de Fisher:

, que es el parámetro que desconocemos

es un valor fijo, no es una variable aleatoria,

Y entonces

El límite inferior de Cramer Rao es (la inversa de la información de Fisher)

Ahora, la varianza muestral es

Y se sabe que

Y entonces por definición de la Chi, que si

Y por la desigualdad,

Concluimos que la varianza del estimador es por lo menos mayor a la cota inferior de Cramer Rao .

## DISTRIBUCION ASINTOTICA DE EMV

Ahora, usemos el EMV  de y asumamos que el veradero valor del parámetro es . Cuando n es grande, la distribución del estimador es aproximadamente normal con media y varianza . Esto es algo que solo se sostiene si *n* tiende a infinito, por ende, el EMV es asintóticamente insesgado.  
En particular con una muestra de elementos i.i.d, la distribución de tiende a una normal Es lo mismo que decir que la **distribución asintótica de** es

Se puede usar este dato para un intervalo de confianza. Como no sabemos , usamos el estimador para la información y podemos decir que también es aproximadamente normal estándar. Sabiendo que la normal estándar es simétrica en el 0,

Si despejamos, tenemos un intervalo de confianza para el valor de .

Es un intervalo de confianza aproximado de nivel

## ERROR CUADRÁTICO MEDIO Y SESGO

Al haber varios estimadores para tener un valor del parámetro desconocido (momentos, máxima verosimilitud, etc.), podríamos hacer todos los métodos y sacar estimadores distintos uno con cada método, a partir de las mismas muestras. Cuando los métodos dan estimadores distintos (a veces pueden dar el mismo), hay medidas que sirven para compararlos, como la eficiencia o el error cuadrático medio.

Si tenemos un estimador de un parámetro de una muestra aleatoria, entonces la desviación de del valor real del parámetro mide la calidad del estimador. Entonces . Y como es una variable aleatoria, entonces quisiéramos saber el error cuadrático medio esperado, por lo que el **error cuadrático medio** de un estimador es .

Por otro lado, el **sesgo** de un estimador es la diferencia entre el valor esperado del estimador y el valor real del parámetro.

Si esa diferencia es 0, el estimador es insesgado.

Podemos ver que el ECM se puede interpretar así:

Por propiedades de la esperanza, podemos re-expresar esto como:

Si ahora sumamos y restamos

Tenemos por un lado , que es la varianza del estimador

Y al lado tenemos el resultado del cuadrado de un binomio, en este caso el cuadrado del sesgo.

Entonces, el error cuadrático medio tiene dos componentes. Uno mide la variabilidad del estimador, su precisión. El otro, mide su sesgo. Es decir, un estimador debería tener baja variabilidad y debe estar cerca del valor real del parámetro. Una combinación baja de varianza y sesgo hacen a un buen estimador. Para hallar un estimador con buenas propiedades de ECM, entonces tiene que ser un estimador que controle por varianza y por sesgo.

Si el estimador es insesgado, entonces el error cuadrático es igual a la varianza

### LAPLACE

Con una muestra i.i.d con densidad

Habíamos obtenido que el estimador de máxima verosimilitud era

Necesitamos saber . Si al calcular esa esperanza obtenemos que , entonces el estimador es insesgado y su error cuadrático medio es su varianza.

Calculamos , y el razonamiento es similar al cálculo que hicimos cuando analizamos sus momentos.

Efectivamente,

es un estimador insesgado de .

Esto quiere decir que el error cuadrático medio va ser simplemente la varianza del estimador

Calculamos

ya lo habíamos calculado, era .

Ahora nos queda calcular

Entonces,

### NORMAL y

Recordemos que si una muestra aleatoria viene de una distribución normal con varianza , el estadístico es la varianza muestral:

Mostramos que por propiedades del chi cuadrado:

Y también:

Ahora, si tenemos una muestra aleatoria salida de la normal, sabíamos que el estimador de máxima verosimilitud de la esperanza era la media muestral, y ahora notamos que:

Esto quiere decir que   es insesgado y

Similarmente, es un estimador insesgado de y su ECM es:

Algo importante es que, aunque muchos estimadores insesgados son buenos vistos desde ECM, controlar el sesgo no garantiza que el ECM esté controlado. En general ocurre que hay un trade-off entre varianza y sesgo tal que un incremento en sesgo puede reducir la varianza en mayor medida y resultar en un ECM aún mejor.

### NORMAL y (SESGO-VARIANZA)

Otro estimador de la varianza de una población normal (además de ) es el estimador de máxima verosimilitud (o el de momentos, que era el mismo).

A la vez, podemos ver que, como es insesgado,

Lo que observamos es que es un estimador sesgado de . Veamos qué pasa con su error cuadrático medio:

Este término se descompone en:

Por ende,

Ahora, considerando el ejemplo anterior,

por un lado y

Observemos que se cumple:

Simplemente porque en el segundo término no estamos restando , que es positivo

Ahora, vemos que

Esa desigualdad debe ser verdad porque de un lado siempre estamos dividiendo por un número menor que del otro.

Lo que implica esto es que

Y

tiene menor ECM que a pesar de que es insesgado. Pero claro, no es insesgado. Lo que estamos haciendo es intercambiar mayor sesgo por menor varianza y obtenemos un menor ECM.

Sin embargo, esto no implica que no haya que usar como estimador. Lo que dice ese análisis es que, en promedio, va a dar valores cercanos a más seguido que . Pero es sesgado y, en promedio, va a subestimar el valor de más seguido que .

De cualquier manera, el ECM es una función del parámetro y no va a haber un “mejor” estimador en términos de ECM. A veces, el ECM de dos estimadores se van a cruzar, es decir, para algunos valores del parámetro uno de los dos es mejor, mientras que, para otros valores, el otro resulta mejor. Para hallar el problema del “mejor estimador” es limitar la clase de estimadores. Esto se hace limitándose únicamente a estimadores insesgados y eligiendo nada más aquel con menor varianza.

Una propiedad de estimadores insesgados es que, si A y B son estimadores insesgados, la combinación lineal de los dos para cualquier *a*, también será insesgado.

### NORMAL (MINIMO ECM)

Ahora supongamos una muestra aleatoria normal y queremos estimar la varianza, hallar valores de tal que tenga el mínimo ECM posible. (notamos que si tenemos y si tenemos .

Llamemos

Ahora llamemos y tendríamos

Y que

El ECM de , va a ser

Lo que queda es minimizar esa función.

Ahora,

Entre corchetes tenemos un polinomio, y vemos que se optimiza esa función cuando la derivada es igual a 0.

Entonces, si , entonces,

## CONSISTENCIA

Una propiedad deseable de estimadores es que sean **consistentes**. Si tenemos un numero grande de observaciones, esperamos tener mucha información sobre cualquier parámetro desconocido , y por ende esperaríamos construir un estimador con ECM muy bajo. Un estimador es consistente si

Lo que quiere decir que a medida que aumenta la cantidad de observaciones, el ECM desciende a 0.

## INTERVALOS DE CONFIANZA EXACTOS

Usar intervalos sirven para obtener una estimación del parámetro desconocido  sin usar una función de la muestra (un estadístico) como estimador. La idea es encontrar un intervalo dentro del cual haya una probabilidad alta de que esté el parámetro . La longitud de ese intervalo nos da una idea de que tan acertadamente se puede calcular .

A veces se puede encontrar la fórmula matemática de la distribución muestral de un estimador (por ejemplo, para un estimador como el promedio sabemos que, con una muestra grande, su distribución es ). Con esa distribución muestral también se podría obtener un intervalo de confianza para le parámetro, que se llama **intervalo de confianza exacto**.

Para encontrar ese intervalo hay que saber la distribución de la población. Con eso, hay que encontrar la distribución muestral del estimador (el estadístico que estas usando para estimar el parámetro). Sabiendo la distribución muestral del estadístico se puede construir el intervalo.

Tomamos una muestra aleatoria de una distribución poblacional con parámetro . Dado un valor , se construye un intervalo de confianza de nivel . El puede ser cualquier número entre 0 y 1 pero suele ser 0,01, 0,02 o 0,05.

* 1. Encontrar una variable que sea una función de los datos y del parámetro.Llamémosle
  2. La distribución de , llamémosle (entonces , debe *no* depender del parámetro.
  3. Sabiendo la distribución de usamos valores y tal que la probabilidad de que esté entre esos valores es . Esto se expresa así: . Con esto se va despejando tal que nos queden cotas inferiores y superiores en las que está el parámetro.

Esa función se llama un **pivote**. Es aquella que satisface ser función de los datos y parámetros, pero sus valores no salen de una distribución que dependa del parámetro buscado. Por ejemplo, si tenemos una muestra aleatoria de una distribución normal con desconocido, pero conocido. Podemos usar como pivote a:

Esta función cumple las propiedades que le pedimos a un pivote: es función de los datos y del parámetro que buscamos; y además tiene distribución normal estándar la cual que **no** depende de , el parámetro que buscamos.

El pivote no es un estadístico porque, aunque es una función de la muestra, tiene un parámetro desconocido, y un estadístico no puede tener un parámetro desconocido. Usamos el estadístico para estimar un parámetro, por lo que se debe poder computar simplemente a partir de los datos y no puede tener un parámetro desconocido. Mientras que haciendo un intervalo de confianza queremos manipular el pivote para llegar a un intervalo sobre el parámetro desconocido, así que el pivote debe contener al parámetro.

Cuando elegimos dos puntos al armar el intervalo, quisiéramos que sea lo más pequeño posible.

### NORMAL DESCONOCIDO y CONOCIDO

Supongamos una muestra aleatoria con distribución normal con desconocido, pero conocido. Hallemos un intervalo de confianza de nivel para .

Usamos  para estimar . Entonces, definimos el pivote así:

Sabemos que si es el promedio de muestras de una distribución normal, entonces

Planteamos el intervalo

Recordamos que es, para siendo la densidad de la normal estándar:

Además. Como la normal estándar es simétrica,

Es así que:

Entonces,

El intervalo de confianza de nivel es:

### NORMAL CONOCIDO y DESCONOCIDO

Tenemos una muestra aleatoria de una distribución con conocido y desconocido. Buscamos un intervalo de confianza de nivel para .

Usamos el estimador que habíamos encontrado:

Y definimos al pivote:

Este estimador tiene distribución chi cuadrado con n grados de libertad, . Llamamos a y a  los cuantiles y de la . Entonces,

Tomamos el inverso recíproco de ambos lados de la desigualdad con ambos lados positivos, entonces cambia la dirección de la desigualdad

Así que el intervalo de confianza de para es

No es un intervalo simétrico, porque el chi cuadrado no es simétrico. A partir de esos ejemplos vemos que el procedimiento para obtener un pivote y obtener intervalos de confianza exactos es:

* Encontrar un estimador de .
* Encontrar una conexión entre el estimador y el parámetro, en general serían funciones que involucran al parámetro y al estimador.
* Entre las posibilidades que encontramos, elegimos la que nos dé una distribución estándar como pivote.

### NORMAL y DESCONOCIDOS

Ahora tenemos una muestra aleatoria con y desconocidos. Buscamos un intervalo de confianza de nivel para ambos.

El EMV o de momentos era:

Definimos la función

Sabiendo que

Entonces es un pivote porque .

Llamamos y a los cuantiles y de la t-student con grados de libertad. Así,

La t-student es simétrica alrededor del 0 como la normal estándar, entonces,

Y la desigualdad nos queda.

Ahora despejamos para tener todo en términos del parámetro desconocido:

El intervalo de confianza nivel nos queda:

Ahora, con , sabíamos que

Si nuestro estimador era, podemos manipularlo un poco y obtener la función

Esta función sabemos que tiene distribución chi cuadrado con grados de libertad porque

Ahora podemos construir el intervalo de confianza de nivel . Llamamos a y a los cuantiles y de la .

Nos queda que el intervalo es

### EXPONENCIAL

Buscamos un intervalo de confianza para el parámetro de la exponencial. Quisiéramos un intervalo de confianza para .

Primero, demostramos que si sigue una distribución exponencial con parámetro entonces tiene distribución exponencial con parámetro ½, lo cual es una . La densidad de es . Podemos ver que la densidad de *Y* es

Y si , entonces , que es una exponencial de parámetro .

De esa manera, tenemos un pivote.

Cada sigue una distribución , por lo que la distribución de es . Llamamos y a los cuantiles y de la .

Entonces,

Nos queda el intervalo para :

En resumen, el proceso general para obtener un intervalo de confianza de la forma , siendo dos estadísticos y por ende variables aleatorias (porque son variables aleatorias). Entonces los intervalos de confianza son intervalos aleatorios.

Supongamos que nos interesa el intervalo de confianza para . En diferentes experimentos observamos varios valores para . Y por ende tenemos diferentes intervalos para diferentes experimentos. Si repetimos el proceso una y otra vez, vamos a tener que entre todos los intervalos de confianza que resulten, alrededor del ·100 % de los que encontremos van a contener el parámetro .

# TEST DE HIPÓTESIS

Teníamos dos formas de estimar parámetros: formas para llegar a un punto en particular (momentos, EMV) o llegando a un intervalo donde es probable que exista el estimador (intervalos de confianza). La idea del test de hipótesis es decidir entre dos afirmaciones diferentes sobre el parámetro.

## PLANTEO DEL PROBLEMA

En el problema de inferencia estadística, queríamos conocer el parámetro de la distribución de nuestra muestra. Arrancábamos sabiendo solamente que ese parámetro existe en el espacio paramétrico (que era el conjunto donde podía existir ; por ejemplo, si sabíamos que la distribución era exponencial, era los reales positivos). Podríamos ver que se puede separar en dos subconjuntos disjuntos (no se intersecan) que llamamos y . Entonces, si son disjuntos, su intersección es el conjunto vacío: y también, si dijimos que estas dos partes surgen de separar en dos el espacio paramétrico, entonces . Lo que queremos saber ahora es si el parámetro desconocido está en o en .

Llamamos a la hipótesis de que y a la hipótesis de que . Como son disjuntos y su unión es el espacio paramétrico, solamente una de las dos puede ser verdadera. Debemos decidir si aceptamos o rechazamos la o la . A se le llama **hipótesis nula** y se le llama **hipótesis alternativa.** A este tipo de problemas con solamente dos decisiones posibles se les llama **test de hipótesis.** La decisión de aceptar una o la otra la vamos a tomar basándonos en observaciones salidas de una distribución y los valores observados nos van a dar información sobre el valor de . El procedimiento para elegir si aceptar o rechazar lo llamamos **criterio de decisión**.

La situación en la investigación científica es que se tiene una teoría actualmente aceptada, , y alguien sale con una explicación alternativa . Para rechazar la teoría que está vigente actualmente la evidencia debe ser mucho más consistente con la nueva teoría. se le llama la hipótesis del investigador porque es la que tenemos ganas de validar, como investigadores que se les ocurrió otra explicación a la que existe. Algo “nulo” es algo sin valor, sin efecto; por eso es la hipótesis de la continuidad, de no hacer cambio a la opinión corriente, ninguna mejora, ninguna diferencia, etc.

Por ejemplo, supongamos que de todas las placas que se produjeron en un periodo están falladas. Alguien sugiere un cambio en el proceso de producción para reducir la tasa de placas falladas. Llamamos a la proporción de placas falladas. Si el método alternativo que se propuso, que es el que tiene la carga de la prueba, asegura un entonces la hipótesis alternativa es . Pero usualmente no decimos que una teoría es *verdadera*.

Ahora supongamos que tenemos una muestra aleatoria de una distribución con parámetro que pertenece a . Además, si tenemos y cuya unión es . Lo que queremos es testear las hipótesis:

Los conjuntos y podrían representar únicamente un par de valores posibles de . Si es el caso, esta es una **hipótesis simple**. En cambio, si alguno de los dos conjuntos contiene más de un valor posible de entonces es una **hipótesis compuesta**. Bajo una hipótesis simple, la distribución de la observación es completamente conocida. Bajo una hipótesis compuesta, solamente se sabe que la distribución de las observaciones pertenece a cierta clase.

Por ejemplo, tenemos una muestra aleatoria de una distribución discreta. La hipótesis nula dice que son observaciones que vienen de una distribución Poisson con media (parámetro) desconocida. La hipótesis alternativa podría afirmar que *no* vienen de una Poisson. Ninguna de las dos hipótesis especifica completamente la distribución de probabilidad (en no se especifica ). Esa es una hipótesis compuesta. Si, en cambio, la hipótesis nula propusiera un valor específico de , sería una hipótesis simple.

Por ejemplo, una hipótesis nula simple tiene la forma de

Con unidimensional, hay dos formas usuales de expresar la hipótesis compuesta. **Hipótesis nula de una cola** tiene la forma de

O de

Cuando la hipótesis nula es simple, la hipótesis alternativa es una **hipótesis de dos colas.**

Con problemas de test de hipótesis es inevitable cometer errores. Es decir, quizás aceptamos o rechazamos la hipótesis nula erróneamente. Entonces, se considera qué tipos de errores se podrían cometer. Para cada valor de , la decisión de rechazar es la decisión incorrecta. Si se rechaza una hipótesis nula que era verdadera (un false positive) se le llama **error de tipo 1**. La decisión de aceptar una hipótesis una hipótesis alternativa que era falsa se llama **error de tipo 2**. Claro que , pero no podrían ser ambas. Solo un tipo de error es posible.

El problema aparece cuando, por ejemplo, sos muy estricto (solo aceptas en muy pocos casos) para tratar de evitar el primer tipo de error, quizás terminas aceptando alternativas que eran falsas (error de tipo 2). Al revés lo mismo, si sos muy abierto a aceptar la hipótesis nula, probablemente la aceptes cuando no deberías (error de tipo 1). \_Básicamente, si rechazas muchas hipótesis, alguna de las que rechazaste van a haber sido en realidad verdaderas; pero, si aceptas mucho, vas a terminar aceptando algo falso.

La probabilidad de cometer un error de tipo 1 se llama **nivel de significación** si la hipótesis nula es simple y se denota con .

Si la hipótesis nula es compuesta, la probabilidad de error de tipo 1 va a depender de qué miembro en particular de es verdadero; en este caso se define como el máximo de estas probabilidades.

La probabilidad de cometer error de tipo 2, por otro lado, se denota con

La probabilidad de rechazar cuando es falsa se llama **potencia** del test y es .

En general quisiéramos que la probabilidad de cometer cualquiera de los dos errores sea baja, pero ambos objetivos van uno en contra del otro, como mencionamos antes. Es decir, si elegimos un criterio de decisión que logre que sea chico entonces va a ser grande. Un criterio de decisión que siempre acepta no importa qué datos tenga va a tener un . Pero al mismo tiempo eso va a llevar a que . Similarmente, si tenemos un criterio que lleva a , eso lleva a que . Lo que hay que llegar es a un balance entre los dos objetivos de y pequeños.\_\*

En la práctica los errores de tipo 1 son más graves que los de tipo 2 (hay que tener mucho cuidado para rechazar una teoría establecida y aceptada; tenemos que estar muy seguros de que la nuestra es una teoría mejor). La forma en la que se llega a un balance entre es controlar por el error de tipo 1. Para eso, elegimos un número entre 0 y 1, y exigimos que . Este test se le llama y es el **nivel de significación del test**. Este nivel podría ser 1%, 5% y sería aceptable en algún uso, pero otros usos quizás requieran otro nivel.

Consideramos un problema donde se testean hipótesis que tienen la forma:

Como ya dijimos, antes de decidir cuál de las dos hipótesis aceptar, podemos observar la muestra que salió de una distribución con el parámetro desconocido . Llamamos como el espacio muestral () del vector de -dimensiones . Es decir, es el conjunto de todos los posibles grupos de observaciones que podrían haberse obtenido de la población.

En problemas de este tipo, especificamos un criterio de decisión partiendo el espacio muestral en dos subconjuntos. Un subconjunto contiene valores de **X** para los que el criterio de decisión va a aceptar ; el otro subconjunto contiene valores de **X** para los que el criterio de decisión va a rechazar y, entonces, aceptar . El subconjunto para el que es rechazada se llama **región crítica** del test. El criterio de decisión lo determinamos especificando la región crítica del test y el complemento de la región crítica debería tener los posibles grupos de observaciones que harían que aceptemos .

Usualmente, buscamos una función de la muestra que devuelva ciertos valores y tomamos la decisión basándonos en el valor de la función. Esta función *no* debe contener parámetros desconocidos. Es decir, la función en la que nos basamos para tomar la decisión de aceptar o rechazar la hipótesis nula es un estadístico que se llama **estadístico de prueba**. Este estadístico lo llamamos . Bajo la hipótesis nula, la distribución estadística de prueba se llama **distribución nula**.

Usualmente, un test que usa el estadístico *T* va a rechazar la hipótesis nula si *T* devuelve valores en algún intervalo fijo, o fuera de un intervalo fijo. El conjunto de valores del estadístico de prueba que lleva a rechazar la hipótesis nula se llama **región de rechazo**. Y el conjunto de valores que lleva a una aceptación de la hipótesis nula es la **región de aceptación**.

### EJ: BINOMIAL

Un político dice que va a ganar por lo menos el 50% de los votos en su ciudad y, por ende, ganar. Un experto analiza las propuestas del político y opina que va a perder. Para testear su suposición, el experto selecciona 15 votantes al azar y encuentra una cantidad *Y* de esa muestra que va a votar al candidato.

Decimos que es la tasa de apoyo del candidato. El político afirma que , nuestra hipótesis nula; el experto estima que , la hipótesis alternativa. Tenemos, entonces,

Si tenemos que *Y* entre 15 votantes eligen al candidato. Vemos que pasa si el valor de *Y* es pequeño, casi 0. De ser correcto lo que afirma el candidato, que por lo menos 50% de la ciudad lo vota a él, no sería totalmente inconcebible que , pero la probabilidad debería ser muy chica. Es mucho más probable que observemos si la hipótesis alternativa fuese correcta. Entonces, si observamos , intuitivamente, rechazaríamos la hipótesis nula () en favor de la alternativa. Si observamos otros valores muy chicos de Y, la conclusión debería ser la misma.

En este ejemplo, tomamos una decisión basándonos en el valor de Y, entonces, Y es el estadístico de prueba. Si Y tiene distribución binomial, , ahora seleccionamos un que completamente determine la distribución de nuestro estadístico de prueba. En general, seleccionamos un valor de que esté en y sea el más cercano a . En este caso, eso sería 0.5. Entonces, Bin(15, *p*) es la distribución nula.\_

A partir de lo anterior surge que rechazamos si el valor observado de *Y* es muy chico. La región de rechazo toma la forma , con *c* como una constante. Esto nos diría que tan chico tiene que ser Y para rechazar la hipótesis.

Veamos qué pasa si calculamos la probabilidad de error de tipo 1 si seleccionamos una región como . Por definición,

Ahora al revés, veamos qué pasa si sabemos que en la realidad el candidato terminará sacando 30% y su predicción fue equivocada. Podemos calcular la probabilidad de que, guiándonos por la muestra de , concluyamos que (que saca más del 50%) es verdad cuando en realidad no lo es. Esa sería la probabilidad de error de tipo 2.

Porque si es verdadera .

Ahora, veamos qué pasa si ajustamos la región de rechazo para que sea y sigamos viendo qué pasa con .

Y

Con este ejemplo se ve que el test con región de rechazo un riesgo de cometer error de tipo 1 bajo () pero no protege mucho contra el error de tipo 2 (). Si la región de rechazo es otra, los menores a 5, el error de tipo 2 baja a 0.278 y el error de tipo 1 sube a 0.151.

### EJ: OTRA BINOMIAL

Supongamos que queremos testear el parámetro *p* de una binomial con intentos.

Usamos como estadístico de prueba el número de éxitos, que llamamos *X*. Entonces, la región de rechazo (que consiste en ciertos valores del estadístico de prueba *X*) la podríamos setear en valores relativamente improbables bajo pero que sí son probables bajo . La región precisa se determina dado un valor de .

Primero supongamos que la región de rechazo son los valores . Es decir, si el estadístico de prueba devuelve esos valores bajo la hipótesis nula, eso indica que quizás habría que rechazarla a la hipótesis nula. El nivel de significación lo vamos a saber calculando qué probabilidad hay de que el estadístico devuelva valores de rechazo si asumimos que la hipótesis nula es verdadera.

Si, en cambio, la región de rechazo fuese , el nivel de significación sería

En general el proceso es al revés: partimos de un valor de y a partir de eso calculamos la región de rechazo. Si queríamos un pero el valor verdadero de *p* es 0.6, la potencia del test es la probabilidad de que *X* sea mayor a 7 pero ahora con probabilidad 0.6.

## TESTS COMUNES

### : NORMAL CON CONOCIDO

Supongamos que tenemos la muestra aleatoria con media desconocido y varianza conocida.

Quisiéramos testear las hipótesis:

Si la hipótesis nula es correcta entonces esperamos que la media muestral se acerque a la media poblacional . Bajo la hipótesis nula sabemos que , pues es conocido. Entonces,

Si la hipótesis nula era correcta, entonces el estadístico debe estar cerca al 0, ya que para que eso pase el numerador debe ser 0 (la media de la muestra es el valor ). Tomaremos una decisión basándonos en *Z*, por ende, *Z* es nuestro estadístico de prueba y su distribución es la distribución nula, .

Ahora supongamos que queremos probar la hipótesis al nivel de significación , es decir, bajo la hipótesis nula de que, la probabilidad de rechazar es . Tomamos una decisión basándonos en la desviación de *Z* del 0, es decir, si el valor de , rechazamos la hipótesis nula, con *c* una constante. La probabilidad de rechazar es con . Sabemos calcular el valor de *c* porque es el cuantil de la normal estándar. El criterio de decisión al final es: si , rechazamos la hipótesis nula con nivel de significación ; de otra forma no rechazaríamos.

En ese desarrollo, teníamos una hipótesis alternativa de dos colas. Si la hipótesis alternativa es de una cola, por ejemplo

Entonces rechazamos la hipótesis nula si *Z* es mayor que un número *c*. Si el nivel de significación es , podemos calcular el valor de *c* haciendo con . Entonces, Rechazamos la hipótesis nula si .

El otro caso de una cola sería si

De manera parecida, rechazaríamos si el estadístico de prueba es menor a un número predeterminado negativo *c*. Si el nivel de significación es , podemos calcular el valor de haciendo con . Así, . Con esa hipótesis alternativa, rechazamos la hipótesis nula si el estadístico de prueba .

### NORMAL CON DESCONOCIDO

Supongamos que tenemos la muestra aleatoria con media desconocido y varianza desconocida.

Quisiéramos testear las hipótesis:

Como antes, si es correcta, esperamos que la media muestral se acerque a la media poblacional . Pero en este caso no podemos usar

como estadístico porque es desconocido.

Sin embargo, podemos usar

Con

Para reemplazar a . La distribución de T es una t de student con grados de libertad.

Con esto, *T* es el estadístico de prueba y la distribución nula es . Ahora, queremos testear la hipótesis a nivel de significación ; es decir, bajo la hipótesis nula , la probabilidad de rechazar la hipótesis nula es . Como antes, tomamos una decisión en base a la desviación de *T* del 0; es decir si el valor de entonces rechazamos la hipótesis nula. Por ende, la probabilidad de rechazo de la hipótesis nula es . El valor de *c* lo podemos saber porque es el cuantil . El criterio de decisión final es, si rechazaríamos la hipótesis nula con nivel de significación ; de otra manera no rechazaríamos.

Ahora consideremos el test:

Entonces rechazamos la hipótesis nula si *T* es *mayor* que un número *c*. Si el nivel de significación es , podemos calcular el valor de *c* haciendo con . Entonces, Rechazamos la hipótesis nula si .

El otro caso de una cola sería si:

De manera parecida, rechazaríamos si el estadístico de prueba es menor a un número predeterminado negativo *c*. Si el nivel de significación es , podemos calcular el valor de haciendo con . Así, . Con esa hipótesis alternativa, rechazamos la hipótesis nula si el estadístico de prueba .

### NORMAL CON CONOCIDO

Supongamos que tenemos la muestra aleatoria con media conocido y varianza desconocida.

Usamos el estimador de :

Si la hipótesis nula, el valor estimado debería estar cerca de . Con esa idea, rechazaríamos la hipótesis nula si

fuese demasiado grande o pequeño.

Ahora, recordamos que el estadístico

Tiene distribución , por lo que podríamos usarlo como estadístico de prueba. Supongamos que queremos testear la hipótesis a nivel de significación . A partir de esa intuición, la probabilidad de rechazar la hipótesis nula:

Siendo *a* y *b* dos constantes. Llamemos al cuantil de la ; y al cuantil . Si llamamos y, entonces el criterio de decisión sería, si , rechazamos la hipótesis nula; de no ser así, aceptamos.

Ahora bien, con un test de una cola:

Rechazamos la hipótesis nula si .

Y en el caso

Rechazamos la hipótesis nula si .

En resumen, el proceso general para resolver un problema de test de hipótesis es

* 1. Encontrar un estimador para el parámetro de interés.
  2. Encontrar una relación entre el parámetro y el valor dado cuando la hipótesis nula es verdadera. Esto es para tener una idea intuitiva de cómo va a ser el criterio de decisión.
  3. Con los posibles criterios de decisión, elegimos el que tenga una distribución estándar para usar como estadístico de prueba. Determinar la región de rechazo basada en ese análisis y expresar la probabilidad de error de tipo 1 en base a la distribución nula.

### NORMAL CON DESCONOCIDO

Supongamos que tenemos la muestra aleatoria con media desconocido y varianza desconocida.

PASO 1:

Usamos el estimador de :

PASO 2:

Si la hipótesis nula es correcta, esperaríamos que y no sean muy distintos. Hay varios estadísticos que nos servirían para ver eso. Ejemplo, , , etc.

PASO 3:

Ahora bien, recordamos que

así que en este caso es nuestro estadístico de prueba. Lo que hicimos fue elegir uno de los estadísticos del paso 2 y modificarlo para tener un estadístico con distribución conocida.

PASO 4

Lo que hacemos ahora es analizar el nivel de significación . Lo que quiere decir eso es que sea la probabilidad hay de rechazar la hipótesis nula si esa hipótesis nula es verdadera (se cumple ). Como tomamos una decisión basada en el valor de , entonces si o , rechazamos la hipótesis nula. La probabilidad de rechazo es .

PASO 5

Con esa ecuación, podemos resolver que . El criterio de decisión es, si, rechazamos la hipótesis nula. En otro caso, no.

Veamos cómo sería con test de una cola:

En este caso rechazamos si

En cambio, si es

Rechazamos si .

### NORMAL CON DESCONOCIDA

Supongamos una muestra aleatoria de una distribución normal con media desconocida y varianza desconocida . Buscamos el p-valor del test de hipótesis

Sabemos que la región de rechazo para esta región es , con

Y

Ahora, supongamos que dadas las observaciones calculamos el estadístico del test y obtenemos , el p-valor sería:

Con y el nivel de significación deseado , entonces si el p-valor es , rechazamos la hipótesis nula, en cambio, no la rechazamos.

## P-VALOR

Como ya dijimos, la probabilidad de error de tipo 1, , usualmente se le llama nivel de significancia, o el nivel del test. Sin embargo, en la práctica el valor que se usa de es un poco arbitrario. Por eso es posible que tengamos diferentes soluciones para que para . El uso de esos números es nada más por conveniencia, no por otra consideración. Entonces, se necesita una cantidad más informativa que decir “rechazamos la hipótesis nula a nivel 0.05”.

Usemos *W* como estadístico de prueba, el **p-valor**, es el menor nivel de significancia para el que los datos observados indiquen que la hipótesis nula se debe rechazar.

Mientras menor sea el p-valor, más convincente va a ser la evidencia de que la hipótesis nula se debe rechazar. Si se tiene un nivel en particular de , el p-valor se puede usar para implementar un test de nivel . Como el p-valor es el menor nivel para el cual rechazamos la hipótesis nula, si el valor deseado de es mayor o igual al p-valor, la hipótesis nula se debería rechazar con ese valor de *.* La hipótesis nula se debe rechazar para cualquier valor de hasta e incluyendo el p-valor. En cambio, si es menor que el p-valor, no se puede rechazar la hipótesis. El p-valor sirve para evaluar que tanto los datos observados están en conflicto con la hipótesis nula.

Si rechazamos la hipótesis nula en favor de la alternativa, par avalores chicos de un estadístico. es decir, la región de rechazo es , el p-valor asociado con el valor observado de W se calcula

Y si la región de rechazo tuviese la forma , el p-valor sería

Y para un test de dos colas, con una región de rechazo es ;

El p-valor es simplemente la cola de la distribución nula. Si el test es de una cola, el p-valor es la probabilidad en una cola; si es de dos, es la probabilidad en las dos.

## POTENCIA Y REGION DE RECHAZO

Un test de hipótesis puede dar error de tipo II si no se rechaza cuando había que rechazarla. La potencia de un test es la probabilidad de que eso no pase. Ósea de rechazar cuando corresponde rechazarla.

O, dicho de otra forma:

Para que un test sea útil la probabilidad de error de tipo I y de tipo II deberían ser bajos. Para reducir la probabilidad de error de tipo I, simplemente elegíamos un valor chico como nivel de significación. Luego, se calcula la potencia. Si la potencia es grande, la probabilidad de error de tipo II es baja.

Por ejemplo, supongamos que tenemos las siguientes hipótesis

Rechazar significaría que es mayor a 80. Asumamos que tenemos 50 observaciones y el

Para calcular la potencia siempre es necesario un valor alternativo de . La potencia va a ser diferente para distintos valores de . Si ese es cercano al de , la potencia va a ser grande.

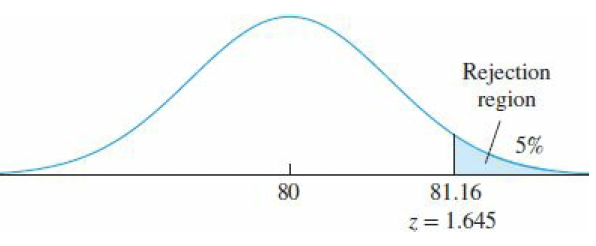
Los pasos para calcular la **potencia** son:

1. Calcular la región de rechazo

2. Calcular la probabilidad de que el estadístico de un resultado contenido en la región de rechazo si la hipótesis alternativa es correcta. Esa es la potencia.

Entonces primero calculamos la **región de rechazo:**

Ya sabemos que el estadístico tiene distribución . Habíamos asumido y . Bajo entonces, la distribución es . va a tomar valores en la normal de esos parámetros que va a tener esta pinta, *dada la hipótesis nula*:



Ahora, a hipótesis nula especifica que , por lo que valores demasiado grandes de empiezan a implicar que no sería cierta. El área del p-valor ca a ser el área (la probabilidad) a la derecha del valor observado de . Entonces, a nivel de confianza, buscamos en la tabla de la normal aquel que deje 5% de probabilidad hacia la derecha.

Ese conjunto serían los valores grandes más extremos que tienen posibilidad del 5% de salir. En este ejemplo, el valor a partir del cual queda 5% de la normal estandarizada hacia la derecha es 1.645, por ende, todos los valores del estadístico de prueba mayores a ese cuantil se rechazan. Esa es la región de rechazo.

Desarrollemos nuevamente:

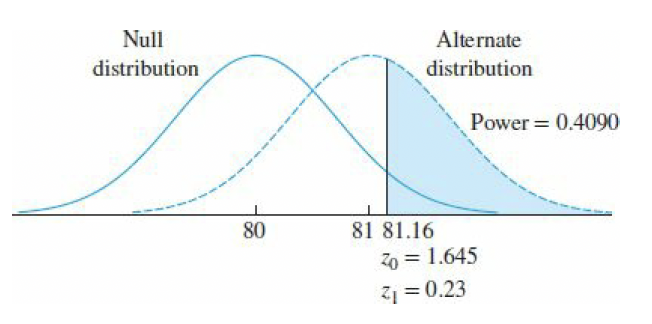
Lo que buscamos es esto

Aquellos valores (cuantiles) que dejan a la derecha 5%. Entonces, podemos darlo vuelta

Podríamos estandarizarla y buscar en la tabla:

El cuantil que deja 95% a la izquierda () es 1.645.

Ahora, calculamos la potencia. Esta es la probabilidad de que caiga en la región de rechazo si una alternativa es verdad, digamos . Ahora con esta hipótesis alternativa, la distribución de es . Acá vemos las dos distribuciones, bajo hipótesis nula y bajo la alternativa.



Se puede ver que bajo la hipótesis alternativa es mucho más probable tener valores que caigan en la región de rechazo.

Analicemos lo que teníamos antes:

Ese es el cuantil en la normal estándar. Podemos despejar para ver qué cuantil tenemos con la distribución nula del estimador.

Ahora, para obtener la potencia podríamos ver que pasa ese cuantil para la distribución de cuando . Estandarizamos:

Si buscamos en la tabla, ese cuantil deja 0.5912598 debajo de la curva. Pero la tabla muestra a la izquierda, y recordemos que rechazábamos a la derecha. Entonces, es la probabilidad de rechazar si . Es decir, la potencia es .

## DUALIDAD

Existe una dualidad entre los problemas de intervalos de confianza y test de hipótesis. Tomemos como ejemplo una muestra aleatoria normal con desconocido y conocido. Planteamos:

Para este test de hipótesis usábamos

Llamamos a al percentil de la normal estándar. Sabemos que aceptamos si

Que es lo mismo que

El proceso nos dice que aceptaríamos la hipótesis nula si se satisface esa desigualdad. Nos queda el intervalo

Que es el intervalo de confianza a nivel para el parámetro . Este proceso nos dice que podemos tener un intervalo de confianza de nivel a partir del resultado de un test de hipótesis con nivel de significación .

## COCIENTE DE VEROSIMILITUD (TEST DE NEYMAN-PEARSON)

Un test de hipótesis muy común es el cociente de verosimilitud. El test se basa en la función de verosimilitud y la intuición de que la función de verosimilitud tiene da ser mayor cerca del valor verdadero de .

### EXPONENCIAL

Consideramos el caso de hipótesis simples. Supongamos que es una muestra aleatoria de tamaño *n* de la distribución exponencial:

Aramos el test de hipótesis

Con . El nivel de significación del test es .

Si asumimos que es correcta, entonces la función de verosimilitud sería:

En cambio, si fuese correcta,

El cociente de verosimilitud lo definimos así:

Intuitivamente, si la evidencia sugiere , entonces va a ser grande y el cociente de verosimilitud va a ser pequeño. Así, rechazamos la hipótesis nula si el cociente de verosimilitud es lo suficiente mente pequeño, es decir con *k* como constante tal que bajo la hipótesis nula de .

Para hallar los resultados del criterio, expandimos la condición.

Ahora, como lo que tenemos es una exponencial de parámetro , entonces sabemos que

Entonces

Y llamamos teniendo en cuenta que . Ahora podemos ver la tabla de los valores del chi cuadrado con grados de libertad tal que la probabilidad

Resolvemos el valor de tal que . Encontrando el valor de *c*, buscamos *k* y definimos el test en términos del cociente de verosimilitud.

Por ejemplo, supongamos que y , y queremos hacer el test a nivel de significación con una muestra aleatoria de tamaño de una exponencial. Podemos ver la tabla del chi cuadrado con 10 grados de libertad y hallamos que el cuantil que deja 0.05 debajo de la curva es 3.94. Con ese dato podemos ver que .

Para hallar el criterio de rechazo en términos del cociente de verosimilitud, resolvemos *k*

Y despejando nos queda . Así que, volviendo al cociente original, rechazamos la hipótesis nula si

## COCIENTE DE VEROSIMILITUD GENERAL

Los cocientes de verosimilitud son útiles para comparar una hipótesis nula contra una alternativa compuesta.

Digamos que la hipótesis nula especifica que (podría ser un vector), se encuentra en un conjunto de valores posibles , es decir, . La alternativa propone que se encuentra en un conjunto que no se superpone con .

Ahora digamos que es el máximo de la función de verosimilitud para todo . Entonces . La representa la mejor explicación para los valores observadlos para todo . Similarmente, representa la mejor explicación para los datos para todo en todo el conjunto . Si pasara que , entonces la mejor explicación de los datos observados estaría en y *no* deberíamos rechazar . Sin embargo, si , entonces la mejor explicación de los datos se encontraría en y deberíamos considerar rechazar la hipótesis nula. El CVG está basado en el cociente .

Definimos el cociente así:

Un test de cociente de verosimilitud del tipo vs. usa como estadístico de prueba y la región de rechazo se determina por .

Notamos que . Un valor de cercano a 0 indica que la verosimilitud de la muestra es mucho menor bajo que bajo , lo que sugiere favorecer antes que . El valor de se elige tal que logre el valor deseado.

### : NORMAL CON DESCONOCIDO

Tenemos las hipótesis de una muestra aleatoria normal:

En esta situación nuestro parámetro es . Notamos que el conjunto es el conjunto , mientras que es el conjunto . Entonces nos queda que . El valor de es completamente desconocido. Ahora debemos hallar.

De la distribución normal sabemos que

Restringir al conjunto implica que , por lo que podemos encontrar si determinamos el valor de que maximiza sujeto a la restricción de . Entonces, el valor de que maximiza es la máxima verosimilitud de ese estimador con .

Entonces obtenemos reemplazando con y con en la función .

Ahora vamos a buscar . Sea el punto en el conjunto que maximiza la . Por el método de máxima verosimilitud tenemos que

Obtenemos reemplazando con . Eso nos da

Entonces, el cociente de verosimilitud nos queda

porque y entonces cuando rechazamos con una constante. Porque

La región de rechazo es entonces

Esto equivale a

Y ya que sabemos que , esa desigualdad es

Ahora, es un estadístico *t* y en este caso el test de cociente de verosimilitud es igual a a un t-test.

### : NORMAL CON DESCONOCIDO

Tenemos las hipótesis de una muestra aleatoria normal:

En esta situación nuestro parámetro es . Notamos que el conjunto es el conjunto , mientras que es el conjunto . Entonces nos queda que . El valor de es completamente desconocido. Ahora debemos hallar.

De la distribución normal sabemos que

Restringir al conjunto implica que , por lo que podemos encontrar si determinamos el valor de que maximiza sujeto a la restricción de . Entonces, el valor de que maximiza es la máxima verosimilitud de ese estimador con. El valor de   que maximiza la es . Entonces, se puede obtener reemplazando en y nos queda:

Ahora, para llamamos al punto en el conjunto que maximiza la función de verosimilitud . Por el método de máxima verosimilitud tenemos que:

lo obtenemos reemplazando en .

Y el CVG es

De nuevo sucede que porque . Entonces, cuando , entonces rechazamos la hipótesis nula. Es decir

Esta desigualdad se puede interpretar como función de , y se va a sostener la desigualdad si muy grande o muy chico. Es decir,

Lo mismo que decir que

Y acá vemos que tiene distribución . Esta situación termina siendo un test de .

NOTA:

El cociente de verosimilitud es una función de la muestra y se puede demostrar que, so tenemos parámetros libres bajo la hipótesis nula, y con *r* parámetros bajo la hipótesis alternativa, si *n* es grande, entonces  tiene una distribución con grados de libertad. En el primer ejemplo teníamos bajo la hipótesis nula, pero no se especifica , por ende, decimos que hay 1 parámetro libre . Bajo la hipótesis alternativa hay 2 parámetros libres: , entonces . Y en ese caso

### POISSON

Supongamos que se quieren comparar el número de quejas por cada uno de los dos turnos que entran por semana en una empresa. 100 observaciones independientes del número de quejas dan   para el turno 1 y   para el turno 2. Asumimos que el número de quejas por semana el en *i*-ésimo turno tiene distribución de Poisson con media para . Usamos el método de máxima verosimilitud para testear . Con nivel de significación .

La función de verosimilitud es l probabilidad conjunta de todos los y dados:

En nuestro ejemplo, y es desconocido. Bajo la hipótesis nula, tenemos una función de un solo parámetro:

Al maximizarla nos queda:

Tenemos que y . Con la función de máxima verosimilitud, vemos que se maximiza cuando y . se maximiza cuando la evaluamos en sus estimaciones de máxima verosimilitud.

Ahora, con los valores observados:

Entonces el valor observado de es

En este ejemplo el número de parámetros libres en es . Mientras que en . Por ende, la distribución aproximada de es . La región de rechazo a nivel contiene los valores de que exceden En este ejemplo, . . Es el valore que corta un área de 0.99 a la izquierda de . Dado que el calor observado de es mayor que rechazamos la hipótesis nula.

### MULTINOMIAL (DATOS CATEGORICOS)

Acá un problema de bondad de ajuste (el tipo de problemas que venimos haciendo, qué tan bien se ajustan los datos al modelo propuesto) para una multinomial. En el modelo, la hipótesis nula describe un vector de probabilidades *p* y especifica que con siendo un parámetro que puede o no ser conocido. Lo que queremos hacer es testear que tan bien se ajusta nuestro modelo contra otro donde las probabilidades conjuntas son libres. Si tenemos *m* variables,

La multinomial con *n* intentos era:

son las frecuencias observadas de cada . El numerador del cociente de verosimilitud va a ser:

Por la estimación de máxima verosimilitud, esta función conjunta se maximiza donde es evaluado en cada .

En , no tenemos restricción, entonces el denominador se maximiza por el estimador de máxima verosimilitud en:

Es decir, la probabilidad de cada observación es la frecuencia observada sobre el total de observaciones.

El CVG nos queda:

Como teníamos que

Entonces

Y

Llamamos y , queda

Bajo , las probabilidades son libres, solo deben sumar 1. Por ende . La distribución de chi cuadrado con grados de libertad.

## COMPARAR DOS MUESTRAS NORMALES

Supongamos que hay dos muestras independientes (cada elemento de ambas es independiente de todos los demás). En particular, tenemos la muestra y . Queremos ver si tienen en la misma distribución. Es decir, si y , testear si . Entonces:

Podríamos asumir que tanto F como G son normales

El problema ahora es ver si . Podríamos plantear:

En la hipótesis nula, lo que queremos ver es que es igual que decir

Teníamos estimadores de la media: , .

Entonces podemos usar como estadístico y sabemos que

Y entonces, de la misma forma que sabíamos que

Simplemente porque así estandarizábamos el promedio de una muestra: le restamos la media de la población (que en este caso también es normal) y esa diferencia la dividimos por la desviación estándar de la población.

Lo mismo podemos hacer con

Si no conocemos , un buen estimador es la media ponderada de ambas varianzas muestrales:

Expresado de otra forma:

Y dado que

Notamos que es un estimador consistente a medida que la cantidad de muestras va a infinito. Si reemplazamos en ,

Esto se puede demostrar:

Sabemos que la distribución chi-cuadrado es la suma de normales al cuadrado. Entonces si sumamos desvíos de la media (poblacional o muestral) al cuadrado y los dividimos por la varianza, esencialmente estamos sumando normales estándar

Y si teníamos que

Llamamos V a , entonces de otra forma:

Recordamos que si y entonces

Por ende, si teníamos

y

Entonces

Y como teníamos que

Entonces, llegamos a lo que queríamos demostrar:

## TEST NO-PARAMETRICO (WILCOXON-MANN-WHITNEY)

Supongamos que hay dos muestras independientes (cada elemento de ambas es independiente de todos los demás). En particular, tenemos la muestra y . Queremos ver si tienen en la misma distribución. Es decir, si y , testear si . Entonces:

Ahora no suponemos nada sobre las distribuciones. Solamente que son continuas y la única diferencia es un parámetro de posición. Es decir, es la misma corrida hacia un lado. Mas formalmente, .

Lo primero que hay que notar es que y y dos muestras de tamaños distintos y respectivamente. Como son distintos, alguna de las dos va a ser más chica, así que digamos que la más chica es . Ósea hay menos observaciones del tipo que . Nombramos a las medias poblacionales de ambas con y .

Lo primero que hacemos es ordenar todas las observaciones de las dos muestras. En total nos va a quedar una lista de observaciones salidas del grupo X y del de Y de tamaño . Una vez que ordenamos las observaciones, Tendríamos algo así:

Si antes teníamos cada una. Supongamos que las dos están de menor a mayor, entonces el subíndice 1, 2, 3 indica que el 1 es el número más chico, el 2 le sigue ordinalmente, y así. Le llamamos a ese número el **rango**.

X :

Y :

Al juntarlas nos va a quedar mezclado dependiendo de los tamaños verdaderos obviamente. El rejunte se debería ver una cosa así:

…

Nos quedaron todos los rangos mezclados, pero lo que hacemos es asignar nuevamente rangos a esta unión de observaciones, esta vez en el superíndice:

…

Nuestro estadístico, que llamaremos *W,* va a ser la suma de los índices de la muestra que tenía menos observaciones (habíamos dicho que esta era Y, con *m* observaciones).

…

Como dijimos que la única diferencia entre las distribuciones era la posición, entonces supongamos que < , los valores de deberían tender a ser menores que los de , lo que llevaría a que *W* sea menor (que algún número *c*). De la misma forma, si el caso fuera que > , entonces *W* sería mayor.

Cuando para ambas muestras se cumple que el tamaño de *m* y de *n* es mayor a 8, se puede demostrar que la distribución nula del estadístico *W* es aproximadamente normal con media y varianza . Con esos casos simplemente usamos la tabla normal para buscar el p-valor. queda que

Ahora el problema va a ser encontrar ese *c* contra el que vamos a comparar el valor del estadístico *W*.

Propongamos un ejemplo:

Si las combinamos

Pero ese es solo un ejemplo, supongamos que tenemos 4 valores que no sabemos que lugares ocupan en las observaciones y podrían estar en cualquier posición. Si asumimos que todavía no sabemos cuánto valen los valores de las muestras, pero podemos ver los posibles valores de *W*.

El estadístico entonces podría darnos .

Bajo las dos eran iguales, entonces no va a estar uno más corrido que el otro y las formas en las que se podría ordenar es . Para asignar rangos de las posibles combinaciones, hay .

|  |  |  |  |
| --- | --- | --- | --- |
| RANGOS | PROBS | W | PUNTUAL |
| {1, 2} | 1/10 | 3 | P(W=3) =1/10 |
| {1, 3} | 1/10 | 4 | P(W=4) =1/10 |
| {1, 4} | 1/10 | 5 | P(W=5) =1/5 |
| {1, 5} | 1/10 | 6 | P(W=6) =1/5 |
| {2, 3} | 1/10 | 5 | P(W=5) =1/5 |
| {2, 4} | 1/10 | 6 | P(W=6) =1/5 |
| {2, 5} | 1/10 | 7 | P(W=7) =1/5 |
| {3, 4} | 1/10 | 7 | P(W=7) =1/5 |
| {3, 5} | 1/10 | 8 | P(W=8) =1/10 |
| {4, 5} | 1/10 | 9 | P(W=9) =1/10 |

Generalizando ese ejemplo, bajo las distribuciones son exactamente iguales, entonces las probabilidades de los rangos son iguales. Habíamos dicho que *m < n* entonces son las posibles asignaciones de rango, todas con igual probabilidad:

Otro estadístico

Rechazamos la hipótesis nula si

a nivel .

En el ejemplo de antes, podemos ver la frecuencia de . Primero obtenemos las frecuencias de

|  |  |
| --- | --- |
| W |  |
| 3 |  |
| 4 |  |
| 5 |  |
| 6 |  |
| 5 |  |
| 6 |  |
| 7 |  |
| 7 |  |
| 8 |  |
| 9 |  |

Y las frecuencias de R va a ser puntual en los valores donde el otro no sea menor

Ahora podemos reemplazar con los valores que habíamos obtenido para las frecuencias de W.

Efectivamente, . Entonces rechazo si a nivel . Vara ver cómo funciona eso nos sirve plantear la misma densidad, pero acumulada:

Y quisiéramos rechazar si vemos valores muy extremos al evaluar nuestros datos en el estadístico, dado .

Entonces a nivel de confianza 0.2 rechazamos si por ejemplo.

## TEST CON DATOS CATEGORICOS

Un test Bernoulli tiene resultado 0 o 1, “fracaso” o “éxito”. Si se llevan a cabo varios intentos de Bernoulli y contamos cuantos éxitos hubo, podríamos testear la hipótesis de que la probabilidad de esa Bernoulli sea igual algún . (.

Para generalizar un intento Bernoulli tenemos un intento multinomial, que es un experimento que puede tener resultados (no solo 0 o 1, “fracaso” o “éxito”). Cuando era 0 o 1, había una sola probabilidad para el éxito y otra para el fracaso. Con un intento multinomial tenés *k* probabilidades para cada uno de los posibles *k* resultados. Por ejemplo, tirar un dado es multinomial con 6 posibles resultados : y el conjunto de probabilidades .

Digamos que tenemos una hipótesis nula de probabilidades multinomiales que nos dice que el conjunto de probabilidades es igual a un conjunto de valores . La hipótesis nula es .

Supongamos que tenemos un dado y queremos testear que es justo. Digamos que es la probabilidad de que salga el número *i*, la hipótesis nula va a ser que el dado es justo, entonces bajo la hipótesis nula, .

Tiramos el dado 600 veces y los anotamos en esta tabla:

|  |  |  |
| --- | --- | --- |
| CATEGORIA | OBSERVADOS | ESPERADOS |
| 1 | 115 | 100 |
| 2 | 97 | 100 |
| 3 | 91 | 100 |
| 4 | 101 | 100 |
| 5 | 110 | 100 |
| 6 | 86 | 100 |

Lo que dice esta tabla es que el numero 1 salió 115 de las 600 veces, y así. Bajo la hipótesis nula esperábamos que salga 1/6 de las veces del total 1/6\*600=100. La idea de este test de hipótesis es que si la hipótesis nula es verdadera esperaríamos que lo observado no se desvíe demasiado (se ajuste) a lo esperado. Entonces construimos un estadístico de prueba que mida la cercanía de los valores observados a los esperados (la bondad de ajuste). Este se llama estadístico chi cuadrado. Para definirlo, llamemos a el número de resultados (k=6 en el ejemplo del dado), y y los intentos observados y esperados, respectivamente, que llevan al resultado . El estadístico es

Entonces, mientras más grande sea , mayor va a ser la evidencia en contra de la hipótesis nula. Para sacar el p-valor del test, deberíamos saber la distribución bajo y, en general no se puede obtener esa distribución exactamente. Sin embargo, cuando los valores esperados son grandes, la distribución chi cuadrado para la que tenemos tabla se aproxima bastante. La distribución es , con tantos grados de libertad como categorías menos 1. Usar la distribución chi cuadrado sirve cuando todos los valores esperados son mayores a 5.

Entonces, para el ejemplo que teníamos del dado justo, podemos hacer

Ahora, para determinar el p-valor del estadístico, podemos usar la distribución chi cuadrado porque todos los valores esperados son mayores a 5. Buscamos en la tabla del chi cuadrado con 6 – 1 grados de libertad el cuantil 10%, que es 9.236. Gráficamente:

Entonces, podríamos decir que el p-valor es mayor a 0.10. Describimos qué tan bien una distribución multinomial particular se ajusta a nuestros datos, por eso se llama un test de **bondad de ajuste.**

## HOMOGENEIDAD

Otra forma de hacer un test de este tipo es, en vez de tener una hipótesis nula con ciertos valores especificados, tener una hipótesis nula que diga que al hacer varios intentos multinomiales cada intento tenga las mismas probabilidades.

Por ejemplo, si fabricamos clavos que queremos que tengan un cierto diámetro, nos pueden salir con más o con menos diámetro. Si tenemos 4 maquinas que hacen clavos, tomamos muestras de cada maquina y contamos la cantidad de clavos en cada categoría (menos diámetro, en rango aceptado, excedido).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | MUY POCO | OK | MUCHO | TOTAL |
| M 1 | 10 | 102 | 8 | 120 |
| M 2 | 34 | 161 | 5 | 200 |
| M 3 | 12 | 79 | 9 | 100 |
| M 4 | 10 | 60 | 10 | 80 |
| TOTAL | 66 | 402 | 32 | 500 |

Esta es una **tabla de contingencia.** Cada fila especifica una categoría de acuerdo con un criterio (maquinas) y cada columna una categoría respecto de otro criterio (diámetro). Cada intersección es una **celda**, por lo que hay 12 celdas. El numero en la celda es la cantidad de ensayos que resultaron ser de la categoría fila y la categoria columna . Ese es el valor **observado** de la celda .

La hipótesis nula es que la proporción de clavos es que la proporción de clavos con muy poco diámetro, diámetro aceptable y demasiado es la misma para todas las maquinas. En general, la hipótesis nula dice que no importa que columna elijamos, la probabilidad de los resultados asociados con las columnas es el mismo. La hipótesis nula entonces es:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | C 1 | C 2 | … | C J | TOTAL |
| F 1 |  |  | … |  |  |
| F 2 |  |  | … |  |  |
| … | … | … | … | … |  |
| F I |  |  | … |  |  |
| TOTAL |  |  | … |  |  |

Ahora, para armar el estadístico necesitamos computar el valor esperado para cada celda. La hipótesis nula decía que la probabilidad de que el resultado de un ensayo caiga en la columna es la misma para cada fila . La mejor estimación de esa probabilidad es la proporción de intentos cuyo resultado justo cae en la columna . Ósea, , entonces

El estadístico que usamos es

Bajo , la distribución es

# PRACTICAS

## GUIA 4 EMV-MOMENTOS

### POISSON

**Una de las primeras aplicaciones de la distribución Poisson fue realizada por Student en 1907, cuando estudiaba los errores cometidos al contar células de levadura o corpúsculos en la sangre con un hemacitómetro. En este estudio, las células de levadura se exterminaron y luego se mezclaron con agua y gelatina. Esta mezcla se esparció posteriormente sobre un vidrio y se dejó enfriar. Se utilizaron dos concentraciones distintas. El recuento se llevó a cabo sobre 400 cuadrados y los datos se resumen en la tabla que sigue:**

|  |  |  |
| --- | --- | --- |
| N CELULAS | CONCENTRACION 1 | CONCENTRACION 2 |
| 0 | 75 | 10 |
| 1 | 103 | 20 |
| 2 | 121 | 43 |
| 3 | 54 | 53 |
| 4 | 30 | 86 |
| 5 | 13 | 70 |
| 6 | 2 | 54 |
| 7 | 1 | 37 |
| 8 | 0 | 18 |
| 9 | 1 | 10 |
| 10 | 0 | 5 |
| 11 | 0 | 2 |
| 12 | 0 | 2 |

**Para cada uno de los dos conjuntos de datos, se debe estimar el parámetro λ mediante el método de los momentos y el de máxima verosimilitud**.

Definimos nuestras variables aleatorias

= "errores en primera concentración" ~

= "errores en primera concentración" ~

Por el método de **momentos**:

va a estar dado cuando

Entonces

va a estar dado cuando

Entonces

Por el método de **EMV**:

Planteamos la función de verosimilitud

Y para va a ser el mismo.

### GEOMETRICA

**Supongamos que X sigue una distribución geométrica, es decir, para k = 1, 2, .... y supongamos que se tiene una muestra independiente e idénticamente distribuida de tamaño n.**

**(a) Deberemos hallar el estimador de los momentos de p.**

**(b) Además, necesitaremos encontrar el estimador de máxima verosimilitud de p.**

### CONTINUA Y PUNTUAL

a) Momentos de

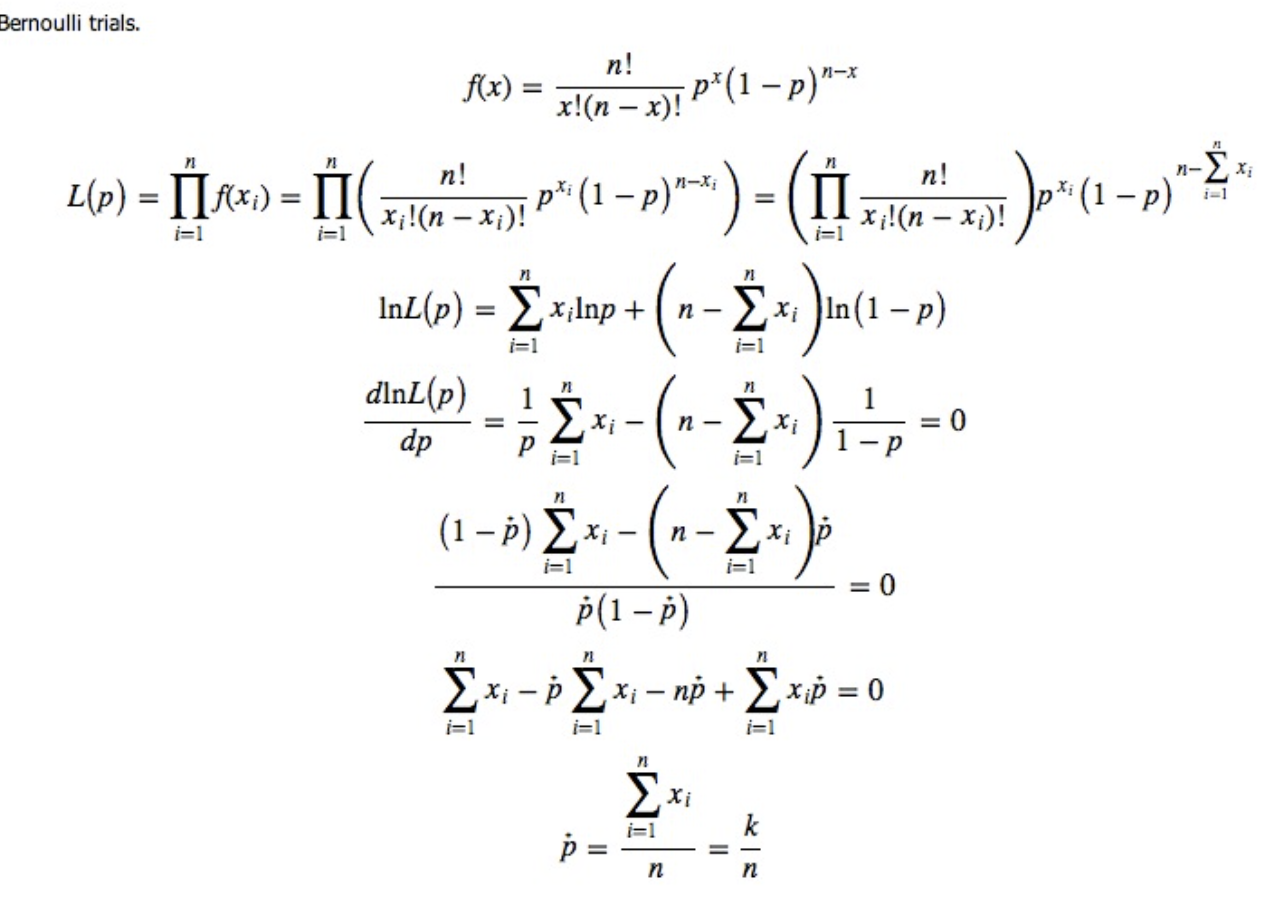
b)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *k* | 0 | 1 | 4 | 8 |
|  |  |  |  |  |

Y el parámetro .

* 1. La media poblacional teórica es
  2. La media muestral es el primer momento muestral:
  3. Resolvemos la ecuación despejando que surge de igualar ,

### BINOMIAL



**b) EMV de .**

La invarianza del EMV es la propiedad de que si tenés y tenés una función del parámetro, el EMV de , es . La varianza de una binomial es

Esto es cuna función del parámetro *p* . entonces

**c) Sesgo de**

El sesgo es

Porque

Porque

Entonces,

**D)**

Es consistente porque

### UNIFORME

**A) MOMENTOS**

No sirve el primer momento

Igualamos los momentos

**B)**

### BETA

Una muestra aleatoria en el intervalo

El EMV va a darse por

### HIPERGEOMETRICA

**En un esfuerzo para determinar el tamaño de una población de animales se capturaron 100 animales que fueron marcados y liberados. Un tiempo después, fueron capturados otros 50 animales de los cuales 20 estaban marcados. ¿Como se podría estimar el tamaño de la población? ¿Qué supuestos se deben hacer sobre el proceso de captura/recaptura?**

N es el tamaño estimado de la población, es el número de individuos marcados y liberados en la primera muestra (100 en este caso), es el total de individuos capturados en la segunda muestra (50 en este caso), es el número de individuos marcados recapturados en la segunda muestra (20 en este caso).

### LAPLACE

Con una muestra i.i.d con densidad

La es

Igualamos la derivada a 0

Y entonces el estimador de nos queda

### MULTINOMIAL

Hallar EMV cuando

Es hallar cuando

Es la misma densidad, porque ya la densidad del vector aleatorio es una probabilidad conjunta de la muestra (que es la definición de la )

### NORMAL

Con una muestra aleatoria normal:

A) Si es conocida ¿cuál es el EMV ?

B) Si es conocida ¿cuál es el EMV ?

C) Hallar los sesgos de los estimadores obtenidos en (a) y (b).

Dado que , porque es la definición de varianza (el desvío cuadrado de la media)

Los dos son insesgados.

D) [3.2.3](#_NORMAL_𝝁_y)

E) OK

F) [3.6.3](#_NORMAL_𝝁_y_1)

### EXPONENCIAL TRUNCADA

https://thescipub.com/pdf/jmssp.2008.284.288.pdf

A) Muestra aleatoria con densidad, con el método de momentos:

B) Muestra aleatoria con densidad, con el método de EMV:

Ahí tenemos un logaritmo de la indicadora. Ósea nos podría quedar o . Más opciones no hay. Como y y nuestro objetivo es que la sea lo más grande posible, entonces no queremos que ese término se vaya a , idealmente sería 0. Notamos que es 1 cuando , si algún valor de la muestra es mayor al parámetro, entonces la indicadora va a ser 0, que es lo que no queremos. Pero también necesitamos que sea tan grande como se pueda, porque el otro término va a influir también el valor final de la . Podemos concluir con ese razonamiento que el valor que maximiza la tan grande como se pueda, pero más chico que el más chico de la muestra (para que la indicadora no de 0). El único valor que cumple eso es el que sea el más chico de la muestra:

### UNIFORME

A) …

B) 3.2.5

C) La densidad de la distribución del máximo con una muestra aleatoria uniforme era, si M es el máximo de las muestras:

…

## GUIA 5 INTERVALOS DE CONFIANZA

### POISSON

**Hallar un intervalo de confianza de nivel aproximado (asintótico) 95% para cada estimador del ejercicio 1 de la Practica 4.**

Recordamos la distribución asintótica del EMV

Ahora,

Como no sabemos , usamos el estimador para la información y podemos decir que también es aproximadamente normal estándar. Sabiendo que la normal estándar es simétrica en el 0. Construimos el intervalo de confianza aproximado de nivel

Para :

Si despejamos, tenemos un intervalo de confianza para el valor de .

Para :

Si despejamos, tenemos un intervalo de confianza para el valor de .

### GEOMETRICA

**En un estudio ecológico acerca del comportamiento alimentario de los pájaros, se contó el número de escalas entre vuelos para distintos pájaros. Para los datos siguientes se supone que el n umero de escalas se puede modelar por una distribución geométrica.**

**(a) Estimar el parámetro *p* de la distribución geométrica.**

**nEscalas\_freq ={1:48, 2:31, 3:20, 4:9, 5:6, 6:5, 7:4, 8:2, 9:1, 10:1, 11:2, 12:1}**

**(b) Hallar un intervalo de confianza de nivel aproximado (asintótico) 95% para *p*.**

Derivadas de la loglik

La info de Fisher :

Fisher evaluada en el parámetro estimado con los datos

Buscamos el intervalo:

Si despejamos, tenemos un intervalo de confianza para el valor de .

### NORMAL

**Valores aleatorios de una normal:**

**5.3299 4.2537 3.1502 3.7032 1.6070 6.3923 3.1181 6.5941 3.5281 4.7433 0.1077 1.5977 5.4920 1.7220 4.1547 2.2799**

**Estimar media y varianza de la normal de la que salieron ()**

Usamos EMV:

(sum(5.3299, 4.2537, 3.1502, 3.7032, 1.6070, 6.3923, 3.1181, 6.5941, 3.5281, 4.7433, 0.1077, 1.5977, 5.4920, 1.7220, 4.1547, 2.2799))/(length(list(5.3299, 4.2537, 3.1502, 3.7032, 1.6070, 6.3923, 3.1181, 6.5941, 3.5281, 4.7433, 0.1077, 1.5977, 5.4920, 1.7220, 4.1547, 2.2799)))

Otra forma que era más fácil

mean(5.3299, 4.2537, 3.1502, 3.7032, 1.6070, 6.3923, 3.1181, 6.5941, 3.5281, 4.7433, 0.1077, 1.5977, 5.4920, 1.7220, 4.1547, 2.2799)

Ahora, la varianza:

jiji <- c(5.3299, 4.2537, 3.1502, 3.7032, 1.6070, 6.3923, 3.1181,

6.5941, 3.5281, 4.7433, 0.1077, 1.5977, 5.4920, 1.7220,

4.1547, 2.2799)

sum((jiji - mean(jiji))^2) / length(jiji)=

**B)** **Dar intervalos de confianza de nivel 90%, 95% y 99% para () :**

**Comparar las longitudes de los intervalos para . Hallar los intervalos de**

**confianza para  también usando la función t.test en R.**

**METODO EMV ASINTOTICO**

Derivadas de la :

Para mu:

Para sigma:

La info de Fisher :

…

… preguntar!

…

**METODO T-TEST**

Ahora tenemos una muestra aleatoria con y desconocidos. Buscamos un intervalo de confianza de nivel para ambos.

El EMV o de momentos era:

Definimos la función

Sabiendo que

Entonces es un pivote porque .

Llamamos y a los cuantiles y de la t-student con grados de libertad. Así,

La t-student es simétrica alrededor del 0 como la normal estándar, entonces,

Y la desigualdad nos queda.

Ahora despejamos para tener todo en términos del parámetro desconocido:

El intervalo de confianza nivel para nos queda:

**Nivel 90%**

[-qt(1-0.1/2, df=16-1)\*sqrt(var(jiji))/sqrt(16) + mean(jiji),

qt(1-0.1/2, df=16-1)\* sqrt(var(jiji))/sqrt(16) + mean(jiji)]

[2.800605, 4.421132]

**Nivel 95%**

[-qt(1-0.05/2, df=16-1)\* sqrt(var(jiji))/sqrt(16) + mean(jiji),

qt(1-0.05/2, df=16-1)\* sqrt(var(jiji))/sqrt(16) + mean(jiji)]

[2.625708, 4.596029]

**Nivel 99%**

[-qt(1-0.01/2, df=16-1)\* sqrt(var(jiji))/sqrt(16) + mean(jiji),

qt(1-0.01/2, df=16-1)\* sqrt(var(jiji))/sqrt(16) + mean(jiji)]

[2.248892, 4.972846]

Ahora, con , sabíamos que

Si nuestro estimador era, podemos manipularlo un poco y obtener la función

Esta función sabemos que tiene distribución chi cuadrado con grados de libertad porque

Ahora podemos construir el intervalo de confianza de nivel . Llamamos a y a los cuantiles y de la .

Nos queda que el intervalo es

var\_pop <- function(x) {

mean((x - mean(x))^2)

}

**Nivel 90%**

Ic <- c(16 \* var\_pop(jiji) / qchisq(1-0.1/2, df=16-1) , 16 \* var\_pop(jiji) / qchisq(0.1/2, df=16-1))

[2.051201 7.061256]

**NIVEL 95**

Ic <- c(16 \* var\_pop(jiji) / qchisq(1-0.05/2, df=16-1) , 16 \* var\_pop(jiji) / qchisq(0.05/2, df=16-1))

[1.865201 8.187521]

**NIVEL 99**

Ic <- c(16 \* var\_pop(jiji) / qchisq(1-0.01/2, df=16-1) , 16 \* var\_pop(jiji) / qchisq(0.01/2, df=16-1))

[1.563089 11.143735]

C)

sqrt(Ic)

99% [1.432201 2.657302]

95% [1.365724 2.861384]

99% [1.250236 3.338223]

D) Con el dato de podemos usar el estadístico

**Nivel 90%**

Ic <- c(-qnorm(1-0.1/2)\*2/sqrt(16) + mean(jiji), qnorm(1-0.1/2)\*2/sqrt(16) + mean(jiji))

[2.788442 4.433296]

**NIVEL 95**

Ic <- c(-qnorm(1-0.05/2)\*2/sqrt(16) + mean(jiji), qnorm(1-0.05/2)\*2/sqrt(16) + mean(jiji))

[2.630887 4.590851]

**NIVEL 99**

Ic <- c(-qnorm(1-0.01/2)\*2/sqrt(16) + mean(jiji), qnorm(1-0.01/2)\*2/sqrt(16) + mean(jiji))

[2.322954 4.898783]

E)

90% [2.788442 4.433296] -> 4.433296 -2.788442 =1.644854 ->1.644854/2=0.822427

Despejo n

N<-(2\*qnorm(1-0.1/2)/ 0.822427 \* 2)^2 = 64

95% [2.630887 4.590851] ->4.590851-2.630887=1.959964 ->1.959964/2=0.979982

N<-(2\*qnorm(1-0.05/2)/ 0.979982 \* 2)^2 = 64

99% [2.322954 4.898783]-> 4.898783-2.322954 ->2.575829/2=1.287915

N<-(2\*qnorm(1-0.01/2)/ 1.287915 \* 2)^2 = 64

F) Cuando se compara la tabla de distribución t con diferentes grados de libertad, se puede observar que a medida que el número de grados de libertad aumenta, la distribución t se aproxima cada vez más a la distribución normal estándar (z). Esto se debe a que la distribución t se vuelve más simétrica y se acerca a una forma de campana similar a la distribución normal a medida que los grados de libertad aumentan.

Cuando se compara una tabla de distribución t con un menor grado de libertad (n) con otra tabla con un mayor grado de libertad (m), se pueden notar algunas diferencias:

Colas más amplias: La distribución t con un menor grado de libertad tendrá colas más anchas en comparación con la distribución t con un mayor grado de libertad. Esto significa que hay una mayor probabilidad de obtener valores extremos en la cola de la distribución con menor grado de libertad.

Valores críticos más altos: Los valores críticos de la distribución t con menor grado de libertad serán más altos en comparación con la distribución t con mayor grado de libertad para el mismo nivel de confianza. Esto se debe a que con un menor grado de libertad, la distribución t es más dispersa y tiene una mayor variabilidad.

En resumen, al comparar una tabla de distribución t con un menor grado de libertad con otra tabla con un mayor grado de libertad, se puede observar que la distribución t con menor grado de libertad tiene colas más amplias y valores críticos más altos en comparación con la distribución t con mayor grado de libertad.

G) ni idea

### NORMAL

A)

B)

qt(0.05, df = n-1)

C)

### EXPONENCIAL

A) B) [3.8.4](#_EXPONENCIAL)

C)

D)

data <- c(25, 45, 50, 61, 39, 40, 45, 47, 38, 39, 27, 10, 59, 35, 66, 38, 32, 45, 21, 34)

ic <- c(qchisq(0.1/2, df=2\*20)/(2\*20\*mean(data)), qchisq(1-0.1/2, df=2\*20)/(2\*20\*mean(data)))

[0.01665157 0.03502417]

### EXPONENCIAL

A) EMV:

B)

Insesgado si

Es insesgado.

C)

### PARETO

A) Distribución de Pareto,

Asumiendo que es dado, estimamos :

C)

D)

c<-c(1.2721, 1.0545, 1.5742, 3.1482, 2.9438, 4.9067, 1.0073, 1.2991, 2.7042, 1.0774, 1.1509, 1.0235, 1.0166, 1.0938, 1.1320)

### GAMMA

EMV:

**B)**

### V o F

**Verdadero o falso: (si la respuesta es V dar una demostración, si la respuesta**

**es F dar un contraejemplo.)**

**(a) El EMV es insesgado.**

**Falso. Solo en el limite.**

**(b) El EMV es consistente.**

**Verdadero. Bajo condiciones generales, el EMV es consistente, lo que significa que converge en probabilidad al valor verdadero a medida que el tamaño de la muestra aumenta. Esta es una consecuencia del teorema de la convergencia de la verosimilitud (teorema de consistencia de los EMV).**

**(c) El EMV es asintóticamente insesgado.**

**Verdadero. Aunque el EMV puede no ser insesgado para un tamaño de muestra finito, suele ser asintóticamente insesgado. Esto significa que el sesgo tiende a cero a medida que el tamaño de la muestra tiende al infinito (consistencia)**

## GUIA 6 TEST DE HIPÓTESIS

### BERNOULLI

**Se arroja una moneda 8 veces en forma independiente para testear la hipótesis de que la probabilidad de obtener una cara es 1/2 versus la alternativa de que dicha probabilidad no es ½. El test rechaza si se observan 0, 1, 7 u 8 caras.**

**(a) ¿Cuál es el nivel de significación de este test?**

Llamemos a *X* = “suma de la cantidad de caras en 8 tiros”

El criterio de decisión está basado en la cantidad de caras después de 8 tiros, si son 0, 1, 7, 8 caras, rechazamos. Para contar la cantidad de éxitos en una serie de intentos independientes con distribución Bernoulli, usamos la función Binomial. Entonces, .

Planteamos el test de hipótesis:

El nivel de significación va a ser . La probabilidad de rechazar es lo mismo que decir la probabilidad de que salgan 0, 1, 7, 8 caras en 8 intentos.

Ahora, para sacar el nivel de significación vemos la probabilidad de rechazar si se cumple ():

**(b) Si en realidad la probabilidad de que un tiro resulte cara es 0.10, hallar la potencia del test.**

La potencia del test es la probabilidad de rechazar la hipótesis nula dado que efectivamente había que rechazarla ( era verdadera). Usamos que indica y calculamos

### VARIAS

**¿Cuál de las siguientes hipótesis es simple y cuál es compuesta?**

**(a) La distribución de X es Exponencial con parámetro 3.**

**Simple**

**(b) Un dado es insesgado, es decir, si X = resultado del dado, entonces E(X) = 3.5.**

**Simple**

**(c) La distribución de X es Normal con media 4 y varianza σ^2 > 1.**

**Compuesta**

**(d) La distribución de X es Gamma con parámetro .**

**Compuesta**

### V o F

**(a) Si el nivel de significación de un test decrece, debemos esperar que la potencia aumente.**

Falso. Cuando el nivel de significación (la probabilidad de cometer un error de Tipo I, es decir, rechazar incorrectamente la hipótesis nula) disminuye, normalmente la potencia del test (la probabilidad de rechazar correctamente la hipótesis nula cuando es falsa) también disminuye. Esto se debe a que se vuelve más difícil rechazar la hipótesis nula.

**(b) Si un test rechaza al nivel de significación α, la probabilidad de que la hipótesis nula sea cierta es α.**

Falso. El nivel de significación α es la probabilidad de rechazar la hipótesis nula dado que es cierta, no la probabilidad de que la hipótesis nula sea cierta.

**(c) Un error de tipo II es más grave que un error de tipo I.**

Depende. No es necesariamente verdadero ni falso; depende del contexto. En algunas situaciones, un error de tipo II (aceptar incorrectamente la hipótesis nula cuando es falsa) puede ser más grave que un error de tipo I, y viceversa.

**(d) El nivel de significación de un test estadístico es igual a la probabilidad de que la hipótesis nula sea cierta.**

Falso. El nivel de significación de un test es la probabilidad de rechazar incorrectamente la hipótesis nula, no la probabilidad de que la hipótesis nula sea cierta.

**(e) La probabilidad de que la hipótesis nula sea falsamente rechazada es igual a la potencia del test.**

Falso. La potencia de un test es la probabilidad de rechazar correctamente la hipótesis nula cuando es falsa. La probabilidad de rechazar falsamente la hipótesis nula es el nivel de significación del test.

**(f) La potencia de un test queda determinada por la distribución del estadístico bajo la hipótesis nula.**

Falso. La potencia de un test queda determinada por la distribución del estadístico bajo la hipótesis alternativa.

**(g) Un error de tipo I puede ocurrir cuando el estadístico del test cae en la región de rechazo del test.**

Verdadero. Un error de tipo I se produce cuando rechazamos la hipótesis nula cuando en realidad es cierta. Esto puede suceder si el estadístico del test cae en la región de rechazo.

**(h) El cociente de verosimilitud es una variable aleatoria.**

Verdadero. El cociente de verosimilitud, que se utiliza en pruebas de verosimilitud, es una función de los datos, que son variables aleatorias. Por lo tanto, el cociente de verosimilitud en sí mismo también es una variable aleatoria.

### NORMAL CONOCIDO

**Sea una muestra de una distribución normal con varianza igual a 25. Hallar la región de rechazo que da el test de Neyman-Pearson para un test de nivel α = 0.10 de H0: μ = 100 versus HA: μ = 125. ¿Cuál es la potencia del test? Repetir para α = 0.05.**

Funciones de verosimilitud bajo cada hipótesis

Ahora definimos

La región de rechazo va a ser , que es lo mismo que

Ahora si trato de llegar a un estadístico,

le resto 1800

divido y multiplico por 16 (*n*)

divido ambos lados por 16

multiplico ambos por -1 y cambio dirección de la desigualdad

Ahora empiezo a agregar cosas hasta tener el estadístico que es

resto mu

divido por

Y ahora es simplemente

Entonces

Es una región de rechazo a nivel 0.1

Potencia:

**B) Una aerolínea tiene dos precios en pasajes aéreos para realizar la ruta Buenos Aires - Córdoba. El precio mayorista tiene una media de 100 dólares con un desvío estándar de 5 dólares, mientras que el precio minorista tiene una media de 125 dólares con el mismo desvío estándar. El dueño de una agencia de viajes sospecha que no le están respetando el precio que le corresponde (mayorista) y considera hacer un reclamo. El promedio de los últimos 16 pasajes que adquirió fue de 121 dólares. Si quiere tener un 95% de seguridad en no hacer el reclamo cuando no corresponda, ¿qué decisión se tomará, en base a la muestra dada?**

Entonces, rechaza

### POISSON

[4.6.3](#_POISSON)

Muestra aleatoria Poisson

Cociente de verosimilitud UMP test nivel

El cociente de verosimilitud va a ser:

y la región de rechazo a cierto nivel tiene la forma

Teníamos que:

Ahora vemos si llegamos a algún estadístico conocido :

Nivel de significación :

Nos sirve el dato de porque *no* va a ser negativo entonce son hay que cambiar la dirección de la desigualdad

dado que

dado que

La suma de Poisson con algún parámetro es otra Poisson con los parámetros de cada uno sumado. como acá sumamos todas Poisson con el mismo parámetro entonces , dada la hipótesis nula

B)

NO RECHAZA

### EXPONENCIAL

**Una empresa vende dos tipos de l amparas. Las de calidad 1 tienen una duración con distribución Exp( = 0.1062) y las de calidad 2 con distribución Exp( =0.05). Un cliente ha realizado una compra de lámparas de la calidad 2 y quiere saber si realmente las lámparas recibidas son de esa calidad. Toma una muestra y mide sus tiempos de duración. Obtiene que . Quiere decidir si devuelve o no el envío y quiere tener un 95% de seguridad de devolver las lámparas cuando no son de la calidad pedida,**

**A) ¿Cuál es el test que debería usar? ¿Qué decisión adoptaría? Sugerencia: Recordar que independientes entonces .**

Nuestra hipótesis nula sería que en realidad son de calidad 2, ósea, que = 0.05 y la alternativa que sea de calidad 1.

Usando Neyman-Pearson, primero calculamos la función de verosimilitud, que para una exponencial () es el producto por tantas muestras de esa densidad.

Esa función, bajo cada hipótesis, nos da

El cociente de verosimilitud, entonces, es

Ese es nuestro cociente de verosimilitud, y lo que deberíamos ver es que, si es menor a un cierto punto crítico , debemos rechazar . Este punto *c* se determina tal que la probabilidad de rechazar la hipótesis nula cuando es verdadera (el error de tipo 1) es igual al nivel de significación que determinamos , porque estamos buscando la probabilidad de rechazar ( dado que la hipótesis nula es verdadera. Entonces .

Entonces, si teníamos que queríamos un nivel de significación = 0.05, se necesita un valor de . Si reemplazamos, esa fórmula nos queda:

pasamos la inversa de la constante al otro lado

tomamos el logaritmo de ambos lados

pasamos dividiendo

= multiplicamos ambos lados por

Y eso (|) sabemos que tiene distribución . Entonces,

Por eso, la región de rechazo es

Para cualquier nivel (no especificado)

Ahora buscamos que nos da una **potencia** de 95%.

El término , que sí tiene distribución chi cuadrado y entonces:

El termino tiene que ser igual al cuantil que deja 0.95 debajo de la curva de [

B)

### NORMAL

se rechaza

### V o F

A) si, el numerador es un subset del denominador

B) no. cuando el p valor es más chico que la significación que se fija entonces rechazas (ósea, hay más evidencia contra la hipótesis nula de la que por lo menos se fijó que debería haber con el nivel). el p valor era la mínima significación para la que se rechaza H0.

C) si . el p valor era la mínima significación para la que se rechaza H0.

D) no . idem c

E) no . idem c

### EXPONENCIAL

A)

hay que usar CVG:

Primero el EMV

Busco las derivadads de la loglik para obtener el estimador

evalúo la lik en h0 sobre la lik en el estimador EMV

Y rechazamos en

### NORMAL

[4.6.1.](#_𝝁:_NORMAL_CON)

[4.6.2.](#_,𝝈-𝟐.:_NORMAL_CON)

### EXPONENCIAL TRUNCADA

index

## GUIA 7 COMPARAR DOS MUESTRAS

### NORMAL

**Se utilizó una computadora para generar cuatro números aleatorios con distribución normal con media y varianza fija: 1.1650, 0.6268, 0.0751, 0.3516. Se generaron cinco números aleatorios más con la misma varianza pero posiblemente con una media diferente: 0.3035, 2.6961, 1.0591, 2.7971, 1.2641.**

**A) MEDIAS**

mean(1.1650, 0.6268, 0.0751, 0.3516) = 0.554625

mean(0.3035, 2.6961, 1.0591, 2.7971, 1.2641) = 1.62398

0.554625-1.62398= -1.069355

B) VARIANZA

en r:

var(1.1650, 0.6268, 0.0751, 0.3516) =0.2163101

var(0.3035, 2.6961, 1.0591, 2.7971, 1.2641) = 1.795

**o :**

((1.1650-0.554625)^2 + (0.6268-0.554625)^2+ (0.0751-0.554625)^2+ (0.3516-0.554625)^2)/(4-1)

0.2163101

((0.3035-1.62398)^2 + (2.6961-1.62398)^2 + (1.0591-1.62398)^2 + (2.7971-1.62398)^2+(1.2641-1.62398)^2)/(5-1)

1.179481

=0.6489303

=4.717924

=(0.6489303+4.717924)/(4+5-2)= 0.7666935

C) DESVIO

sqrt(0.7666935\*(1/4+1/5))

D) INTERVALO DE CONFIANZA

c(-qt(1-0.1/2, df=7)\*sqrt(0.7666935\*(1/4+1/5))+ (0.554625-1.62398),

-qt(0.1/2, df=7)\* sqrt(0.7666935\*(1/4+1/5))+ (0.554625-1.62398) )

[-2.18218744 0.04347744]

F) P VALOR

estadistico:

(0.554625-1.62398) / sqrt(0.7666935\*(1/4+1/5)) ~ t

-1.820559

P value = P(|T|>-1.820559)

P value = 1 - P(|T|<-1.820559)

P(T<-1.820559)

P(-T<-1.820559) = P(T>1.820559)

P VAL = P(T>1.820559) + P(T<-1.820559)

P VAL =( 1-P(T<1.820559)) + P(T<-1.820559)

P VAL= 0.0557386 + 0.0557386=0.1114772

G) nO

h)

### OZONO

**Se condujo un experimento para medir los efectos del ozono, un componente del smog. Se mantuvo un grupo de 22 ratas de 70 d  as de edad durante 7 d  as en un ambiente que conten  a ozono, y se anotaron las variaciones en sus pesos. Se mantuvo otro grupo de 23 ratas de la misma edad en un ambiente libre de ozono por un tiempo similar, y nuevamente se registraron las variaciones en sus pesos. Los datos (en gramos)  guran m as abajo,**

a <- c(1.0, 22.6, 38.5, 11.9, 20.7, 13.7, -7.9, 48.2, 33.2, 27.4, 27.7, 24.0, 54.4, 24.8, 60.4, 13.5, 33.1, 33.3, -7.8, 27.5, -18.1, 27.1, 17.1)

mean(a)= 16.48636

b <- c(-12.8, 30.2, 6.5, -8.2, -3.0, 28.0, 27.9, 22.1, 33.5, 31.4, 24.8, 9.9, -21.9, 17.8, 45.4, 13.5, 13.6, 23.8, 48.7, 9.1, 2.6, 19.8)

mean(b)= 22.88261

dif <-16.48636 - 22.88261= -6.39625

sample\_variance <- function(x) {

n <- length(x)

mean\_x <- mean(x)

variance <- sum((x - mean\_x)^2) / (n - 1)

return(variance)

}

> sample\_variance(a)

[1] 366.8842

> sample\_variance(b)

[1] 320.5431

= (22 - 1) \* 366.8842 = 7704.568

= (23 - 1) \* 320.5431 = 7051.948

Entonces,

pooled\_variance <- function(x, y) {

na <- length(x)

nb <- length(y)

sa <- sample\_variance(x)

sb <- sample\_variance(y)

pooled <- ((na-1)\*sa + (nb-1)\*sb )/ ((na) + nb - 2)

return(pooled)

}

344.2525

Bajo , , entonces

El estadistico tiene distribucion t\_(22+23-1)

Estadistico\_H0 = (16.48636-22.88261)/sqrt(344.2525\*(1/22+1/23))= -1.155995

p valor es:

P(|Estadistico\_H0| > -1.155995)

1 - P(|Estadistico\_H0| < -1.155995)

[ P(Estadistico\_H0 < -1.155995) ] + P(-Estadistico\_H0 < -1.155995)

[ P(Estadistico\_H0 < -1.155995) ] + P(Estadistico\_H0 > 1.155995)

P\_Valor = [ P(Estadistico\_H0 < -1.155995) ] + [ 1 - P(Estadistico\_H0 < 1.155995) ]

pt(-1.155995, df=22+23-1) + (1- pt(1.155995, df=22+23-1))

p.valor = 0.2539217

A nivel 0.05, no recahzamos

CODIGO COMPLETO:

a <- c(1.0, 22.6, 38.5, 11.9, 20.7, 13.7, -7.9, 48.2, 33.2, 27.4, 27.7, 24.0, 54.4, 24.8, 60.4, 13.5, 33.1, 33.3, -7.8, 27.5, -18.1, 27.1, 17.1)

b <- c(-12.8, 30.2, 6.5, -8.2, -3.0, 28.0, 27.9, 22.1, 33.5, 31.4, 24.8, 9.9, -21.9, 17.8, 45.4, 13.5, 13.6, 23.8, 48.7, 9.1, 2.6, 19.8)

n = length(a)

m=length(b)

sample\_variance <- function(x) {

n <- length(x)

mean\_x <- mean(x)

s = 0

for (i in x){

s = s + (i - mean\_x)^2

}

variance <- s / (n - 1)

return(variance)

}

pooled\_variance <- function(x, y) {

na <- length(x)

nb <- length(y)

sa <- sample\_variance(x)

sb <- sample\_variance(y)

pooled <- ((na-1)\*sa + (nb-1)\*sb )/ ((na) + nb - 2)

return(pooled)

}

sa = sample\_variance(a)

sb = sample\_variance(b)

print(sa)

print(sb)

sp = pooled\_variance(a,b)

print(sp)

Estadistico\_H0 = (16.48636-22.88261)/sqrt(344.2525\*(1/22+1/23))

p.valor = pt(Estadistico\_H0, df=n+m-1) + (1 - pt(-Estadistico\_H0, df=n+m-1))

print(p.valor)

t.test(a,b)

EL T.TEST de r da un p valor de 0.2534 que es mas o menos lo mismo que me dio a mi

El histograma muestra que mas o menos la media esta en 20 mas o menos para los dos asi que tiene sentid no rechazar. Masomenos normal es

A picture containing text, line, diagram, plot

Description automatically generated

> shapiro.test(a)

Shapiro-Wilk normality test

data: a

W = 0.96707, p-value = 0.6193

B)

c(-qt(1-0.05/2, df=n+m-1)\* 5.533113-6.39625), -qt(0.05/2, df=n+m-1)\* 5.533113-6.39625)

[-17.547507 4.755007]

Tal cual como da el t.test ¡!

### ACERO

**Se llevó a cabo un experimento para comparar la resistencia a la fractura del acero al níquel de alta pureza con el acero de pureza comercial del mismo tipo. Para m = 32 unidades, la resistencia promedio muestral fue de = 65.6 para el acero de alta pureza, mientras que = 59.8 para n = 38 unidades del acero comercial. Debido a que el acero de alta pureza es más costoso, su uso para cierta aplicación puede justificarse solo si su resistencia de fractura excede a la del acero de pureza comercial en más de 5.**

**a) Si se supone que σ = 1.1, pruebe las hipótesis pertinentes usando α = 0.01. Halle el p-valor.**

Bajo H0:

( (65.6-59.8) - 5) / (1.1 \* sqrt(1/32 + 1/38)) = 3.031203

OSEA, en la normal, viendo solamente la cola derecha, los valores que tienen probabilidad 0.01 estan despues del cuantil

Ósea si el estadístico hubiera dado un numero como o más grande, hubiera rechazado. Entonces rechazo la hipótesis nula

P-VALOR

Puedo rechazar la hipótesis nula porque

**b) Calcule la potencia para el test anterior cuando μx - μy = 6.**

Potencia =

Z= ((65.6 - 59.8) - 6) / (1.1 \* sqrt(1/32 + 1/38)) =

Ahora,

La potencia es entonces 1-pnorm(-1.462654 )= 0.928219

### ECM

**Se desea estimar la diferencia de medias de dos distribuciones normales independientes con igual varianza a través de una muestra con igual número de observaciones para cada distribución. Si fuera posible elegir, ¿que conduce a un menor error cuadrático medio del estimador: que el desvío estándar de las poblaciones se reduzca a la mitad o duplicar el tamaño de las muestras?**

CASO 1:

CASO 2: n\*2

CASO 1:

CASO 2:

DA LO MISMO.

### TAMAÑO DE MUESTRAS

**Se desean comparar dos muestras independientes para ver si sus medias poblacionales difieren. Supongamos que las observaciones en las dos muestras son normalmente distribuidas con la misma varianza conocida. Si hay disponibles un total de m sujetos para llevar a cabo el experimento, como debe repartirse este total entre las dos muestras si queremos:**

**a) tener el intervalo de confianza para lo más corto posible.**

lo estimamos con . El intervalo de confianza va a ser:

llamamos m-n la cantidad de sujetos en un grupo y n la cantidad en el otro grupo.

Ahora queremos minimizar. Notemos que el único número que no conocemos es *n*.

Recordamos que la normal es simétrica entonces

El mínimo valor de long que va a tomar esa función lo vamos a encontrar donde la derivada sea 0. De paso tomamos log para que sea más fácil.

**b) que el test sea lo más potente posible**

Para maximizar la potencia, necesitaríamos que

La potencia era la probabilidad de rechazar cuando hay que rechazar (la alternativa es verdadera). Rechazábamos cuando el estadístico de prueba era grande (cuando los valores bajo eran tan extremos que no era significativo aceptar). Es decir, al *observar* valores que provienen de dos distribuciones diferentes (en el caso ) si el test es bueno entonces debería poder rechazar en ese caso. De ser verdad el estadístico debería devolver esos valores extremos que hacen que se rechace la hipótesis nula. Es decir, esos valores son tan extremos que bajo no tiene sentido.

Entonces, quisiéramos maximizar , que es lo mismo que minimizar . Que tomas el logaritmo y es lo mismo que antes.

### TEST F DE VARIANZAS

Intervalo de confianza para

y son varianzas muestrales.

HIPOTESIS:

Llamemos

V debería ser más o menos 1 para aceptar y muy extrema para rechazar.

Bajo la hipótesis nula, se cumple que:

Sabíamos que

Entonces por definición de la F,

Por el criterio de decisión rechazaríamos si o , siendo constantes tal que , para nivel de significación Idealmente . Entonces el valor de c son los percentiles de la F, .

El intervalo de confianza quedaría

A nivel 0.05,

a <- c(1.0, 22.6, 38.5, 11.9, 20.7, 13.7, -7.9, 48.2, 33.2, 27.4, 27.7, 24.0, 54.4, 24.8, 60.4, 13.5, 33.1, 33.3, -7.8, 27.5, -18.1, 27.1, 17.1)

mean(a)= 16.48636

b <- c(-12.8, 30.2, 6.5, -8.2, -3.0, 28.0, 27.9, 22.1, 33.5, 31.4, 24.8, 9.9, -21.9, 17.8, 45.4, 13.5, 13.6, 23.8, 48.7, 9.1, 2.6, 19.8)

mean(b)= 22.88261

Bajo H\_0, entonces

366.8842/ 320.5431 = 1.144571

RR

q1 = qf(0.05/2, df1=length(a)-1, df2=length(b)-1)

0.4214492

q2 = qf(1-0.05/2, df1=length(a)-1, df2=length(b)-1)

2.393775

rechazo si V<0.4214492 o V>2.393775

V = 366.8842/320.5431 = 1.144571

Entonces **no** **rechazo** que sean la misma

p-value = 0.7598

### ABURRIMIENTO

**Los estudiantes universitarios hombres ¿se aburren m as que las mujeres? Esta pregunta se examino en el artículo `Boredom in Young and Adults Gender and Cultural Comparisons'. Los autores aplicaron una escala llamada Proneness de Aburrimiento a 61 estudiantes hombres y a 121 estudiantes mujeres de universidades de Estados Unidos. ¿La hip otesis de que la tasa de aburrimiento Proneness es m as alta para hombres que para mujeres >se encuentra sustentada por la información dada? Pruebe las hipótesis apropiadas usando un nivel de sonicación de 0.05.**

**DATOS:**

**H <- data.frame(n=61, mean=10.4, sd=4.83)**

**M <- data.frame(n=121, mean=9.26, sd=4.96)**

Entonces, planteo Hipotesis

Planteo el estadístico

Sp = ( (61-1)\*4.83 + (121-1)\*4.96 ) / (61 + 121 - 2)

4.916667

A nivel 0.05, rechazo los valores de cola izquierda si bajo que dejan

qt(0.05, df=81) = -1.663884

T = ( 10.4 - 9.26 ) / sqrt(4.916667\*1/61+1/121) = 3.824178

Entonces NO RECHAZO.

### RULEMANES

Se realiz o un estudio para comparar los rulemanes para la maquinaria fabricados a partir de diferentes compuestos (McCool 1979). Se testearon diez rulemanes de cada tipo. La tabla que sigue muestra el tiempo hasta que fallaron (en millones de ciclos):

t1 <- c(3.03, 5.53, 5.60, 9.30, 9.92, 12.51, 12.95, 15.21, 16.04, 16.84)

t2 <- c(3.19, 4.26, 4.47, 4.53, 4.67, 4.69, 12.78, 6.79, 9.37, 12.75)

A) teoría normal, a nivel 0.05 testear que son iguales

Sp^2 = 18.1015

Estadístico:

T = ( mean(t1)-mean(t2) ) / sqrt(18.1015\*1/10+1/10) = 2.852943)

qt(0.05/2, df=18) = -2.100922

qt(1-0.05/2, df=18) = 2.100922

T = ( mean(t1)-mean(t2) ) / sqrt(18.1015\*(1/10+1/10)) = 2.072309)

(r da lo mismo)

NO RECHAZO

p-valor:

P(|T|> 2.072309)=

1-pt(2.072309, df=18) + pt(-2.072309, df=18) = 0.05287564

A nivel 0.05 NO RECHAZO

B) WILCOXON

rangos\_men\_sum <- function(x, y) {

j <- sort(c(x, x))

if (length(x)<length(y) | length(x)==length(y)){

menor <- x

} else if (length(x)>length(y)) {

menor <- y

}

j <- sort(c(x, y))

s = 0

for (i in 1:length(j)) {

if (j[i] %in% menor){

s = s + i

}

}

return(s)

}

rangos\_men\_sum(t1, t2)

[1] 130

Rechazamos la hipótesis nula si

a nivel .

Wilcoxon rank sum exact test

data: t1 and t2

W = 75, p-value = 0.06301

alternative hypothesis: true location shift is not equal to 0

C) el de Mann Whitney, porque no son muy normales

D) Estimar ; la probabilidad de que un rulemán de tipo I dure masque uno de tipo II.

pi = sum(t1 > t2) / length(t1)

0.9

### MANN WHITNEY

[4.9](#_TEST_NO-PARAMETRICO_(WILCOXON-MANN-)

### CARPAS

liso <- c(25, 28, 16, 34, 38, 21, 29, 43, 32, 36)

estampado <- c(25, 28, 16, 34, 38, 21, 29, 43, 32, 36)

> rangos\_men\_sum(liso, estampado)

[1] 210

## GUIA 8 DATOS CATEGORICOS

### PARTICULAS

**La siguiente tabla da los datos correspondientes a la emisión de partículas alfas contadas en intervalos de un segundo de duración.**

**n |Observados**

**0 |5267**

**1 |4436**

**2 |1800**

**3 |534**

**4 |111**

**5+ |21**

**La primer columna indica la cantidad, n, de part  culas observadas en cada uno de los intervalos de un segundo de duraci on.(En 5267 intervalos no fue observada ninguna part  cula, en 4436 intervalos se observ o s olo una, etc.) Usar el test chi-cuadrado de bondad de ajuste para testear el ajuste de la distribuci on Poisson a los datos.**

1-Hipótesis:

2-Estimar el parámetro de la distribución de Poisson:

Primero necesitamos calcular la media de las observaciones. La media de una distribución de Poisson es igual a su parámetro. Para ello multiplicamos cada valor por su frecuencia y sumamos todos los productos. Dividimos la suma total por la suma de las frecuencias para obtener la media.

3-Calcular las frecuencias esperadas:

Usamos la fórmula de la probabilidad de Poisson, que es

Sustituimos λ por la media que calculamos y buscamos la probabilidad para cada número de emisiones de partículas. Multiplicamos cada probabilidad por el total de observaciones para obtener las frecuencias esperadas.

> for (i in 0:4) {

+ print(paste("P( X =", i, ")" ,dpois(i, 0.8371271)))

+ }

[1] "P( X = 0 ) 0.432952567872306"

[1] "P( X = 1 ) 0.362436327580497"

[1] "P( X = 2 ) 0.151702635921056"

[1] "P( X = 3 ) 0.0423314625569831"

[1] "P( X = 4 ) 0.00885920362227145"

> 1-ppois(4, 0.8371271)

"P( X >= 5 ) 0.0017178"

4-Calcular el estadístico chi-cuadrado:

Calculamos los valores esperados (el total de observaciones por la probabilidad de que suceda cada n si viniesen de la distribución:

= 0.432952567872306 \* 12169 =

5268.6

= 0.362436327580497 \* 12169 =

4410.488

= 0.151702635921056\* 12169 =

1846.069

= 0.0423314625569831\* 12169 =

515.1316

E\_4 = 0.00885920362227145 \* 12169 =

107.8076

= 0.0017178\* 12169 =

20.9039

Usamos que

Entonces

> esp <- c(5268.6 ,4410.488, 1846.069, 515.1316, 107.8076, 20.90391)

> obs <- c(5267, 4436, 1800, 534, 111, 21)

> s = 0

> for (i in 1:6) {

+ s = s + ( (obs[i]) \* log( ( obs[i] / esp[i] ), base=exp(1) ) )

+ }

> print(2\*s)

[1] 2.083784

> qchisq(1-0.05, df=5)

[1] 11.0705

Nos dio 2.083784 enronces no rechazamos que sea poisson

### TRAFICO

**La distribución Poisson ha sido usada por ingenieros para modelar el tráfico liviano, basado en el supuesto de que la tasa es aproximadamente constante y que el tránsito es liviano (es decir, que los autos individuales se mueven de modo independiente), la distribución del flujo de autos en un intervalo de tiempo dado o de área espacial fija debería ser aproximadamente Poisson (Gerlough y Schuhl 1955). La siguiente tabla muestra el número de giros a la derecha que se produjeron a lo largo de 300 intervalos de 3 minutos en una intersección específica. Ajustar una distribución Poisson y testear la bondad del ajuste usando el estadístico chi-cuadrado. Agrupar las últimas 4 celdas. Comentar el ajuste. Es útil saber que los 300 intervalos se tomaron a lo largo de varias horas del día y durante varios días de la semana.**

*n Frecuencia*

*0| 14*

*1 |30*

*2 |36*

*3 |68*

*4 |43*

*5 |43*

*6 |30*

*7 |14*

*8| 10*

*9 |6*

*10| 4*

*11 |1*

*12 |1*

*13+| 0*

freq <- c(14, 30, 36, 68, 43, 43, 30, 14, 10, 6, 4, 1, 1, 0)

A picture containing text, screenshot, font, document

Description automatically generated

a)

1. Las variables , (las muestras con las que estamos trabajando) y que son las que toma la función compuesta que define la , son aleatorias, por ende latambién es una variable aleatoria (es una función de variables aleatorias). [↑](#footnote-ref-2)