# ANOTACIONES DE LAS CLASES Y REPASO DE LAS UNIDADES

## Anotaciones de la primera clase (06/08)

**NOTA**: Si estoy estudiando la estatura promedio de los individuos de una población, el objetivo de estudio es la estatura y no los individuos.

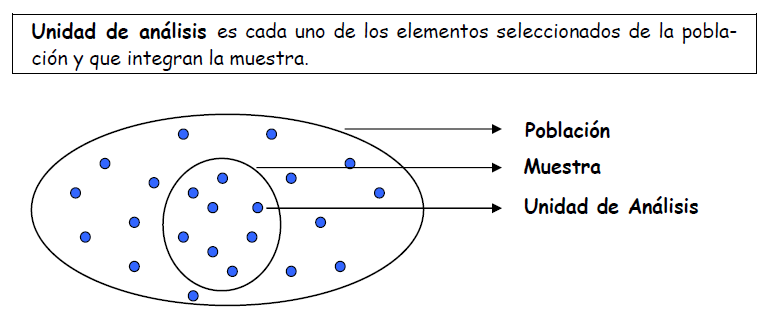
**NOTA**: Los datos en estadística son aleatorios. Los datos de las variables categóricas pertenecen a conjuntos mutuamente excluyentes. Los datos cuantitativos sí tienen carácter numérico. En ciertos casos se pueden establecer categorías estableciendo rangos con los valores posibles de los datos de las variables cuantitativas (por ejemplo una variable numérica es el peso de los individuos de una población y las variables categóricas que se establecen son: sobrepeso, peso normal, desnutrición, obesidad). Las variables numéricas son **discretas** o **continuas.**

**NOTA:** En ciertos casos una variable continua puede discretearse, es decir que es tratada como una variable discreta. Por ejemplo, se puede discretizar la variable edad y se puede discretizar la variable peso.

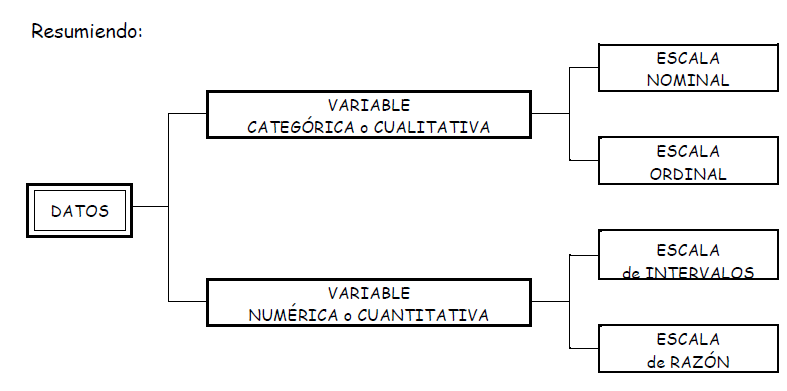
### DEFINICIONES



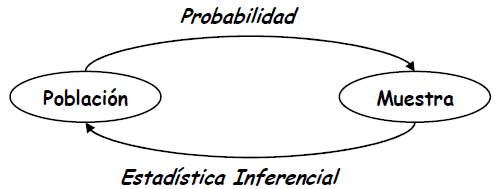




### Escalas de medición de los datos obtenidos según sus valores



### Relación entre la estadística y la probabilidad



## Anotaciones de la segunda clase (13/08)

**NOTA**: Hay tantas medias muéstrales como muestras puedan existir, estas medias no son las verdaderas y en el mejor de los casos son buenas aproximaciones de la media de la población, la cual es la verdadera. Lo mismo sucede al respecto de la variancia en la población y en la muestra

**NOTA**: Se usan las letras griegas para la población y se utilizan las letras latinas para las medidas en las muestras.

**NOTA**: *La puntuación es una medida individual* que es una medida de la distancia de una medida respecto de la media de la población en términos, o relativo al valor de la desviación media de la población. Es una medida de dispersión individual. En la puntuación el valor de la medida utilizado es cualquiera de los posibles de la variable

## Clase 3 (19/08):

**NOTA**: No se pueden prever los resultados de un experimento aleatorio. Los resultados de un experimento aleatorio son valores aleatorios.

**NOTA**: Observar que en el espacio muestral están todos los ene-adas ordenados (interesa el orden), es decir que importa el orden y por lo tanto pares ordenados con los mismos elementos pero en distinto orden son elementos distintos del espacio muestral.

**NOTA**: Recordar que en las variaciones sin repetición, todos los elementos de cada una de las variaciones son distintos entre sí. En este caso a lo sumo una variación puede tener tantos elementos como elementos en el conjunto a partir del cual se definen las variaciones. En cambio en las variaciones con repetición los datos en una misma variación si pueden ser iguales. En este caso la cantidad de elementos en la variación puede ser mayor que la cantidad de elementos en el grupo a partir del cual se definen las variaciones

Clase 09/09

## Distribución de Poisson

### La siguiente es la definición de proceso de Poisson

1) Se observan sucesos en un intervalo temporal o espacial con una tasa de ocurrencia λ. El valor promedio se obtiene de realizar el experimento varias veces en intervalos representativos y obtener el promedio

2) Los sucesos en intervalos disjuntos o no solapados son independientes.

3) La cantidad de sucesos en un intervalo es proporcional a la extensión del intervalo.

Se dice que la variable X, que es el número de éxitos en el intervalo, tiene una distribución de Poisson, y su función de masa de probabilidad es la que ya vimos en el libro.

**NOTA**: Es decir que el proceso de Poisson consiste en conteos de sucesos (éxitos) específicos en un intervalo espacial o temporal. Los sucesos en intervalos no solapados son independientes y la cantidad de sucesos que ocurran en un intervalo de extensión dada será proporcional a esta extensión.

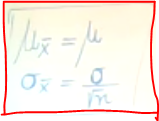
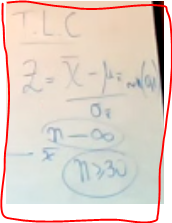
## CRITERIO DE DECIMALES AL UTILIZAR LA APLICACIÓN Y LAS TABLAS: SE USA LA CANTIDAD DE DECIMALES DE LA TABLA

## ANOTACIONES DE LA CLASE DEL 23/09/21

**NOTA**: La curva de la distribución normal tiene dos puntos de inflexión a un desvío estándar delante de la media y a un desvío estándar detrás de la media de la distribución. Son importantes las probabilidades estimadas en intervalos simétricos alrededor de la media expresada en términos de la desviación estándar.

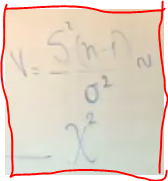
## ANOTACIONES DE LA CLASE 30/09/21

**NOTA**: Entonces, tenemos cierta variable aleatoria X (por ejemplo la estatura) de cierta población con determinados parámetros (media y dispersión estándar). Tomamos muestras independientes de la población, todas del mismo tamaño, relevamos la misma variable en las muestras y obtenemos los estadísticos correspondientes. Podemos definir entonces una nueva variable aleatoria (es aleatoria porque no se conocen a priori los valores que puede tomar la variable) que serán las medias a lo largo de diferentes muestras independientes de la población. Si la distribución de la variable en la población es normal, por la propiedad reproductiva de la distribución normal, la variable aleatoria definida que se denomina **media muestral** tendrá también distribución normal y la media de esta variable **media muestral (es una variable aleatoria muestral)** coincide con la media de la variable en la población. La desviación estándar de la **media muestral (se denomina error estándar y se pone un subíndice x raya)** es igual a la desviación estándar de la variable en la población dividida la raíz cuadrada de n. La media muestral es una distribución fundamental de muestreo

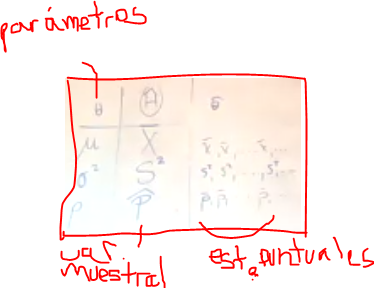


La imagen de la izquierda es el teorema del límite central. Cuando n tiende a infinito la distribución de la variable Z así definida tiende a ser normal estándar.

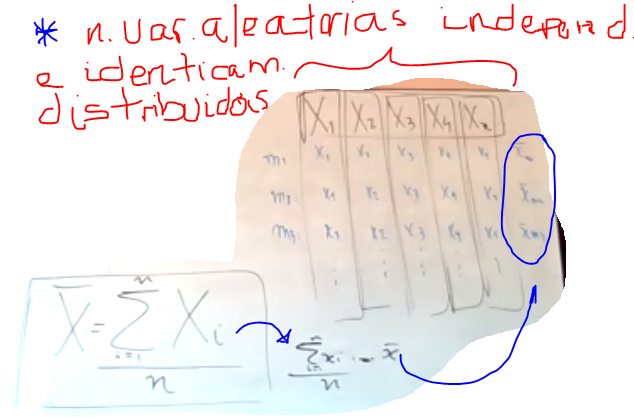
**NOTA**: Se define un nuevo estadístico a partir de las varianzas de la variable en las distintas muestras de tamaño n. Esta será una nueva distribución fundamental de muestreo. Esta nueva variable aleatoria tiene una distribución ji-cuadrada con n-1 grados de libertad (cuando la variable tiene distribución normal).



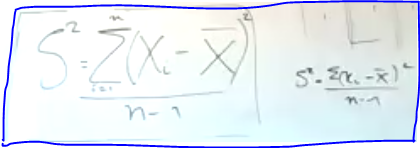
**NOTA**: Los parámetros son valores numéricos que describen parcial o totalmente a la población



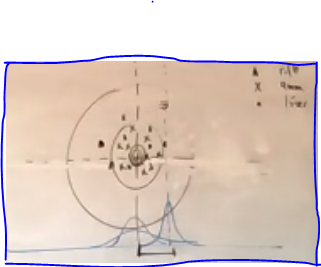
**NOTA**: Definición de una muestra aleatoria



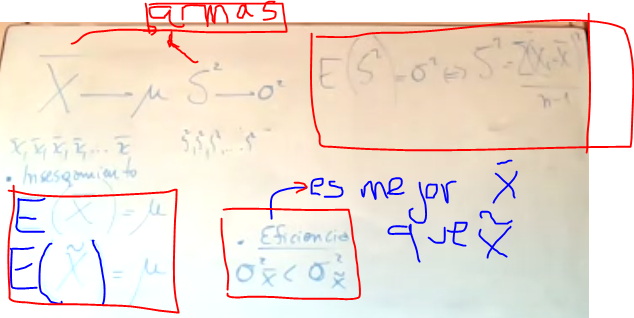
**NOTA**: Definición de varianza muestral, la que no tiene una distribución por sí misma y hay que afectarla por un factor para obtener la distribución ji-cuadrada. Esta variable aleatoria da cada una de las varianzas muestrales

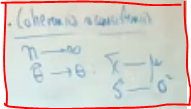


**NOTA**: Ejemplo con el disparo de armas. La curva centrada y con punta media es la del revolver. La más puntiaguda es el que tiene mira láser y la otra es el rifle



**NOTA**: Las siguientes se denominan estadísticos y estimadores cuando se utilizan para la estimación



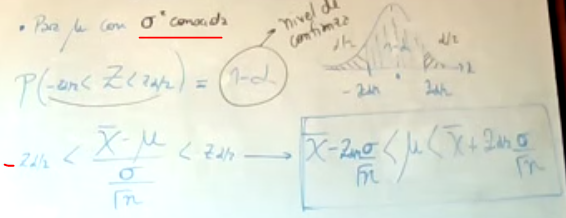


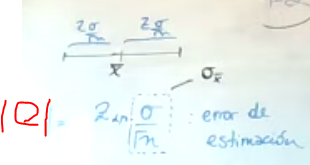
**NOTA**: Esto último dice que cuando el tamaño de las muestras tiende a infinito, los estadísticos tienden a los parámetros.

### Estimación por intervalos

#### Primer intervalo (conocemos la desviación estándar y conocemos la media en una muestra y queremos estimar la media de la población)

Basados en la información obtenida de una muestra por ejemplo la media, hacemos una estimación de la media de la población a partir de la definición de un intervalo simétrico alrededor de la media de la muestra





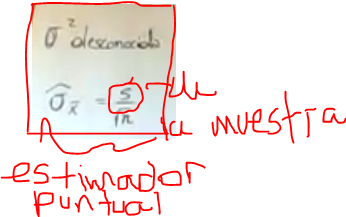
**NOTA**: e se denomina error de estimación

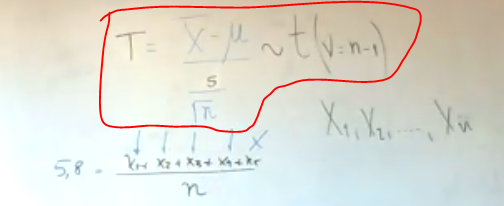
**NOTA**: Se observa que colocamos a la media en un intervalo simétrico alrededor de la media muestral (del estimador puntual)

**NOTA**: Para tener cierto nivel de confianza constante y teniendo una buena precisión hay que disminuir el valor del error estándar para disminuir el valor del radio del intervalo simétrico alrededor de la media muestral. Si no hacemos esto vamos perdiendo precisión porque arriesgamos poco. Para disminuir el error estándar hay que aumentar el tamaño de las muestras.

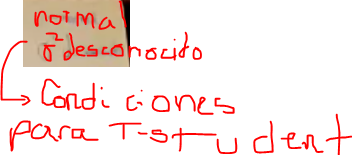
“Con un nivel de confianza tal, el intervalo contiene a la verdadera media de la variable”

#### Segundo intervalo (no conocemos la desviación estándar)





**NOTA:** Cuando cambiamos un parámetro por un estimador puntual perdemos un grado de libertad. En el ejemplo que se encuentra, para que la media tenga determinado valor podemos fijar solamente 4 de los valores y el 5 queda determinado por los demás (deja de ser independiente)



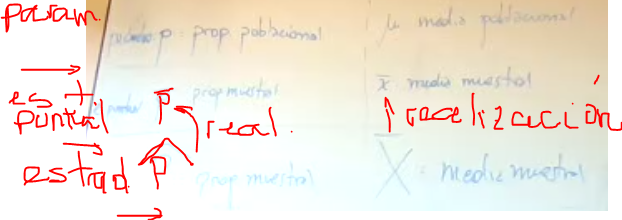
**NOTA**: También cuando el tamaño de la muestra es pequeño (menor a 30). Si n es lo suficientemente grande, la estimación puntual de la desviación se aproxima a la deviación estándar de la población por la propiedad de coherencia y por lo tanto se puede usar directamente una distribución normal estándar.

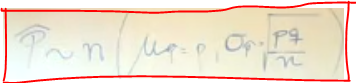
**NOTA**: La distribución t es simétrica parecida a la normal pero más achatada

#### Tercer intervalo (para estimar la varianza)

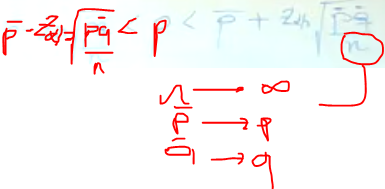
**NOTA**: La distribución de la ji-cuadrada se utiliza para la estimación de la varianza y no es simétrica y sesgada a derecha

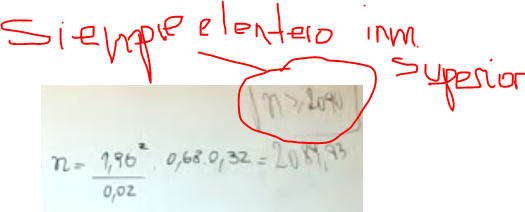
#### Cuarto intervalo (para estimar la proporción)





**NOTA**: La distribución fundamental de muestreo para la proporción muestral es siempre normal (porque tiene que ver con la distribución binomial y cuando n es muy grande tiende a una distribución normal). Los valores de p y q es la proporción de éxitos y fracasos verdaderos de la población

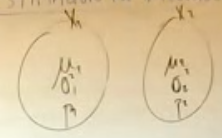




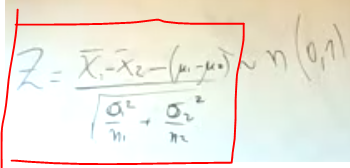
**NOTA**: Valores conservadores cuando no se conocen las proporciones en la muestra y se requiere el valor del tamaño de la muestra se hacen las siguientes suposiciones conservadoras para obtener el máximo valor para n

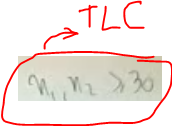


#### Ahora hacemos estimaciones por intervalos para varias poblaciones



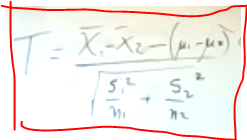
**NOTA**: Sirve para la comparación de los parámetros de las poblaciones en lugar de obtener los valores absolutos en cada población

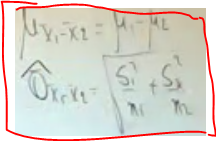


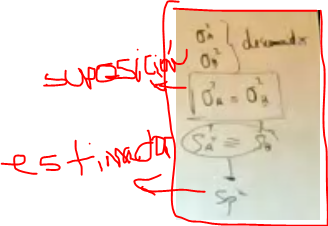


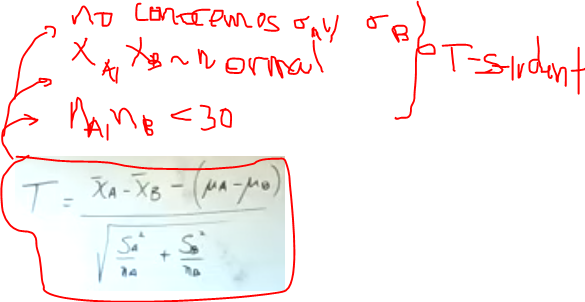
**NOTA**: Se cumple también el teorema del límite central

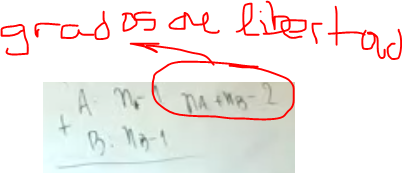
**NOTA**: Se cumple la propiedad reproductiva cuando ambas poblaciones tienen distribuciones normales



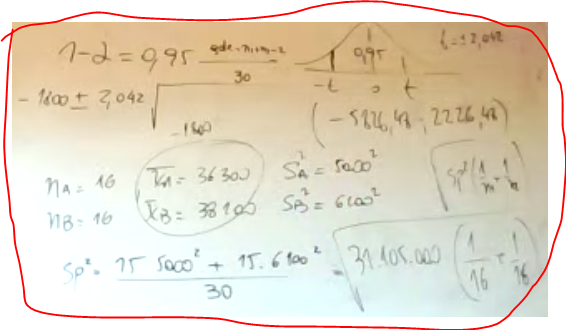


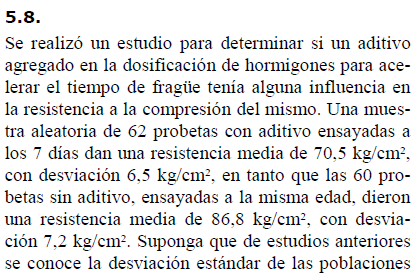
Ejercicio 5-27

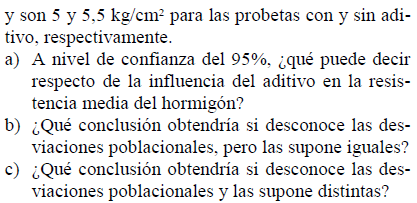




**Nota**: Con todo lo anterior se obtiene la distribución T



Tarea



Vamos a hacer un poco más de lo que pide el ejercicio para practicar lo que ya vimos. Vamos a estimar las medias de cada una de las poblaciones por medio del primer intervalo de estimación dado que conocemos las medias de las muestras y las desviaciones estándar de las poblaciones y el tamaño de las muestras es mayor a 30, de modo que se puede asumir, por el teorema del límite central, que la distribución Z es normal.

Entonces con un nivel de confianza del 95% estimamos las medias poblacionales

* Hormigón sin aditivo



Luego



Error de estimación

Intervalo de confianza . Se observa que también la precisión es buena

* Hormigón con aditivo

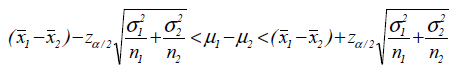
Error de estimación

Intervalo de confianza . Se observa que también la precisión es buena

En principio diríamos que el agregado del aditivo perjudica la resistencia del hormigón

* Ahora hacemos la estimación por intervalos en conjunto

Podemos suponer distribución normal dado el tamaño de las muestras.



Consideramos población 1 a la población de hormigón con aditivo

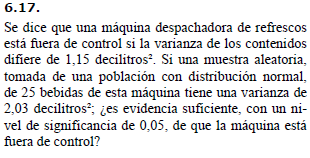
Error de estimación

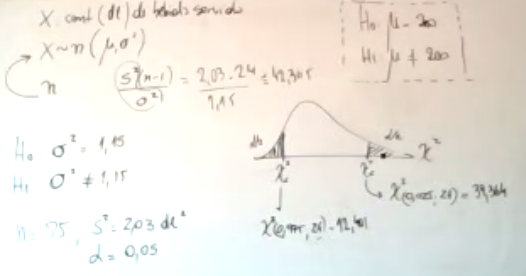
Intervalo de confianza

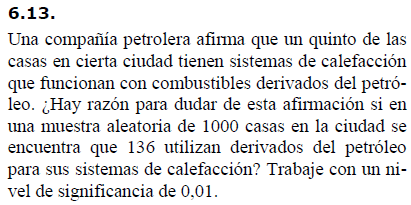
Se observa que ambos extremos del intervalo son negativos. Podemos decir primero, que con un nivel de confianza del 95% la diferencia de las verdaderas medias poblacionales está en el intervalo. Además se observa que los límites del intervalo son ambos negativos, indicando que con un nivel de confianza del 95%, la media de la población 1 es menor que la media de la población 2. Concluimos que el aditivo sí perjudica la resistencia del hormigón.

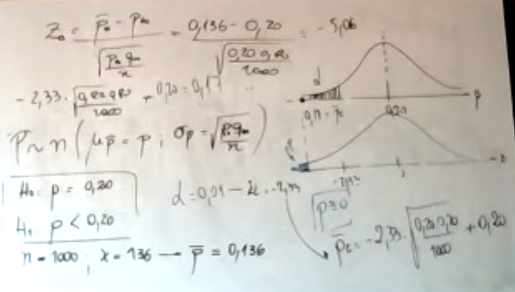
b) En este caso, tendremos que utilizar la fórmula con la distribución t dado que no hay fórmula para la distribución normal estándar cuando las desviaciones no son conocidas, pese a que el tamaño de las muestras es mayor que 30

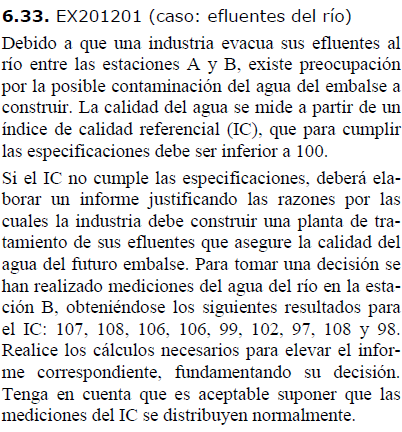
## Anotaciones de la clase del 14/10

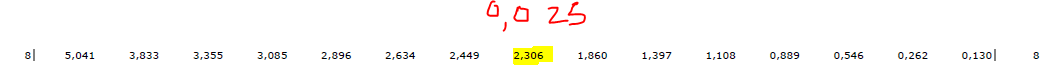


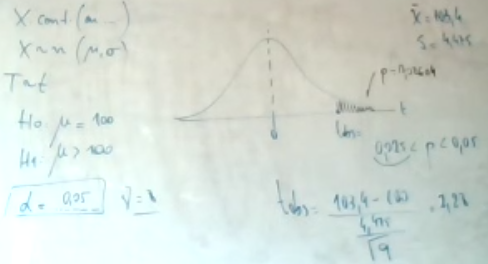




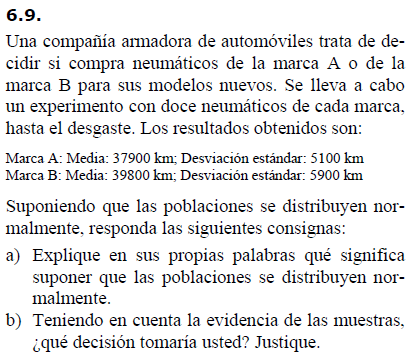




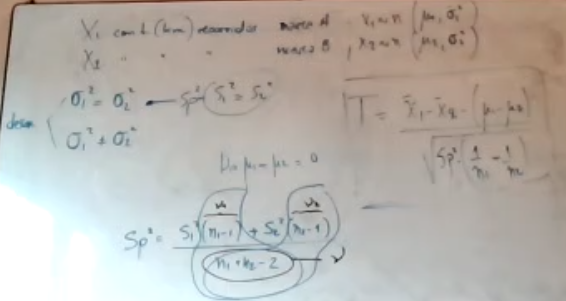




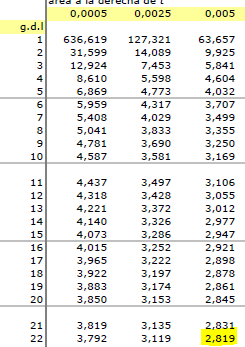
### Prueba de hipótesis para diferencia de medias



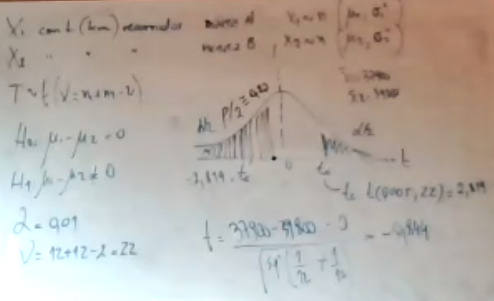
**NOTA**: En este caso son desconocidas las varianzas de las poblaciones de modo que hay que hacer una suposición justificada acerca de si son o no diferentes. Para verificar con cierto nivel de confianza si son o no significativamente distintas debemos hacer una prueba por intervalo del cociente de varianzas. Esto para tener bien definida la distribución T. Porque en caso de que las desviaciones estándar de la población podemos usar el estimador de varianza conjunta sp2, que era un promedio ponderado por los grados de libertad (de cada distribución) de las varianzas de las muestras



Hemos elegido un nivel de significancia del 1% en una prueba dedos colas. La hipotesis nula es que las medias de las poblaciones son iguales y la hipotesis alternativa es que son distintas y por eso es que deriva en una prueba de dos colas. El número de grados de libertas es igual a n1 +n2-2.



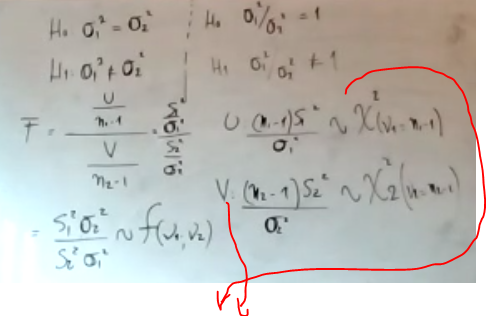
Este es el valor crítico



El valor de t observado está en la zona de aceptación, de modo que aceptamos la hipótesis nula.

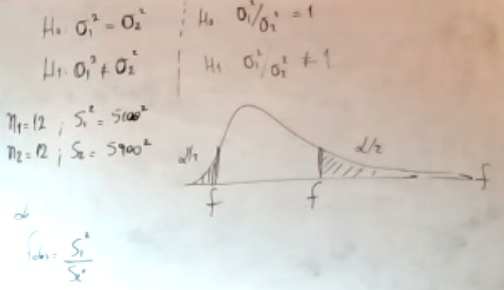


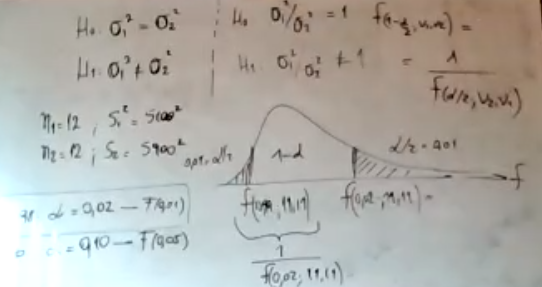
Vemos entonces que el valor de p/2 es aproximadamente igual a 0.2. Con lo que el valor de p es aproximadamente igual a 0.4. Con lo cual es riesgo real de cometer error de tipo 1 (rechazar mal) es del 40% con lo cual ni en pedo rechazamos

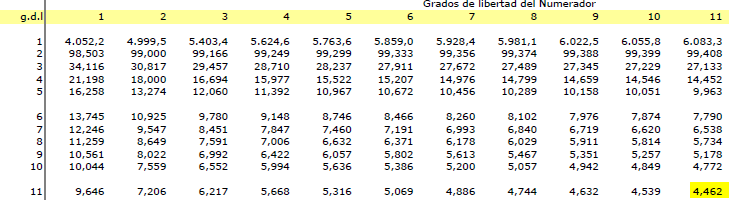


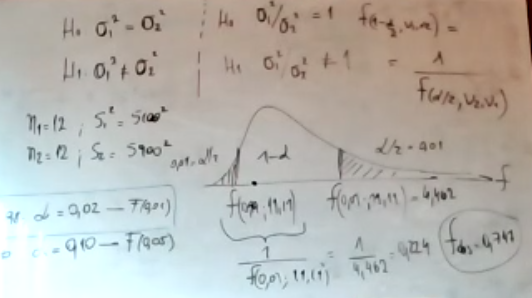
Esto siempre que las muestras provengan de una población con una distribución normal con sus respectivas varianzas y siempre que estas distribuciones normales sean independientes entre sí (bajo estas consideraciones originan las distribuciones ji-cuadrado con los grados de libertad que se indican).

**NOTA**: La ji-cuadrada así como la F de Fischer son distribuciones sesgadas

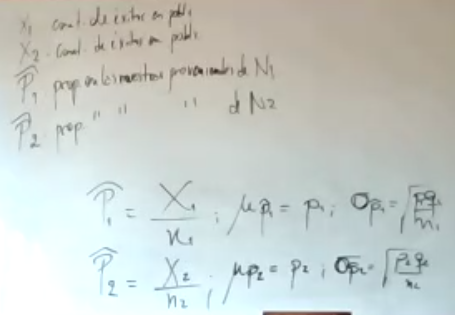


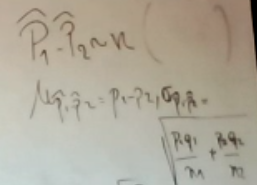




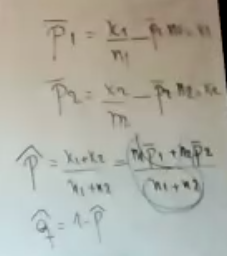
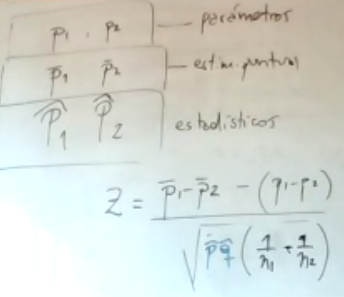


### Diferencia de proporciones

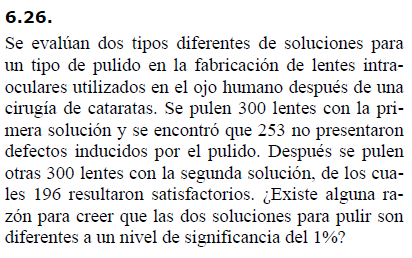


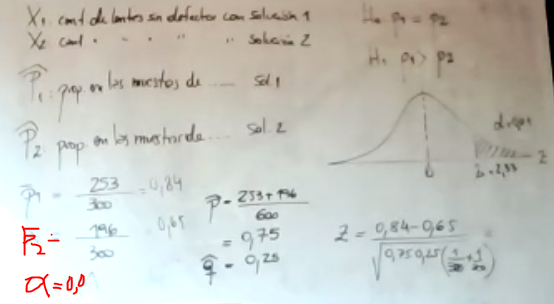


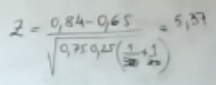
**NOTA**: La distribución de las proporciones en las muestras es normal para tamaños de muestra suficientemente grandes, tamaños tales que np y nq sea mayor que 5.



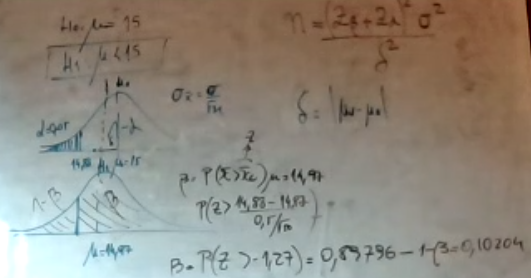
**NOTA**: Entonces tenemos un estimador de proporción conjunta. Vemos que igual es un promedio ponderado de las proporciones en las muestras respecto de los tamaños de las muestras. La diferencia de las proporciones muestrales es normal también







### Sensibilidad de una prueba de hipótesis



**NOTA**: La fórmula de la esquina superior derecha da el tamaño de una muestra necesaria para obtener determinada potencia de prueba. El delta es la diferencia entre la media hipotética y la media real

**NOTA**: La distribución ji-cuadrado exige la normalidad de la distribución y necesitamos una varianza conocida. Recordemos que la distribución ji-cuadrado es la que utilizamos para la prueba de hipótesis de la varianza. Cuando queremos hacer inferencias acerca de la varianza

