**PREGUNTAS:**

**UNIDAD 5 - Estimación de parámetros**

**Recordamos:**

La **inferencia estadística** consta de los métodos mediante los cuales se hacen inferencias o generalizaciones acerca de una población.

Las inferencias se basan estrictamente en información obtenida de una muestra aleatoria seleccionada de la población, y el método bayesiano

La **inferencia estadística** se puede dividir en dos áreas principales: **estimación** y **pruebas de hipótesis**.

Cualquier inferencia deberá hacerse tomando en cuenta **no sólo el valor único**, sino **también la estructura teórica o la distribución de todos los valores x ¯** que se podrían observar a partir de las muestras de tamaño n.

Como resultado de lo anterior **surge el concepto de distribución muestral**, que **es la base del teorema del límite central.**

La distribución de probabilidad de un estadístico se denomina **distribución muestral.**

La **distribución muestral** de un estadístico **depende de la distribución de la población, del tamaño de las muestras y del método de selección de las muestras.**

**Cabe destacar que, a partir de las definiciones de las siguientes distribuciones, se puede afirmar que donde haya una t, F o χ2 la fuente era una muestra de una distribución normal.**

**Métodos de estimación clásicos**

**Si se emplea una estadísticapara estimar un parámetro desconocido θ,recibe el nombre de estimador de θ** y el valor específico del resultado de los datos muestrales, recibe el nombre de estimación puntual de θ

No se espera que un estimador logre estimar el parámetro de la población sin error. Lo que en realidad se espera es poder indicar con un cierto porcentaje de confianza que el valor real del parámetro no está muy alejado o en su defecto que se encuentra incluido en un cierto intervalo.

**Varianza de un estimador puntual**

Si 1 y 2 son dos estimadores insesgados del mismo parámetro de la población θ, deseamos elegir el estimador cuya distribución muestral tenga la menor varianza.

Si consideramos todos los posibles estimadores insesgados de algún parámetro θ, **al que tiene la menor varianza lo llamamos estimador más eficaz** de θ.

**Estimación por intervalo**

La exactitud de la estimación aumenta cuando las muestras son grandes; pero incluso así **no tenemos razones para esperar que una estimación puntual de una muestra dada sea exactamente igual al parámetro de la población.**

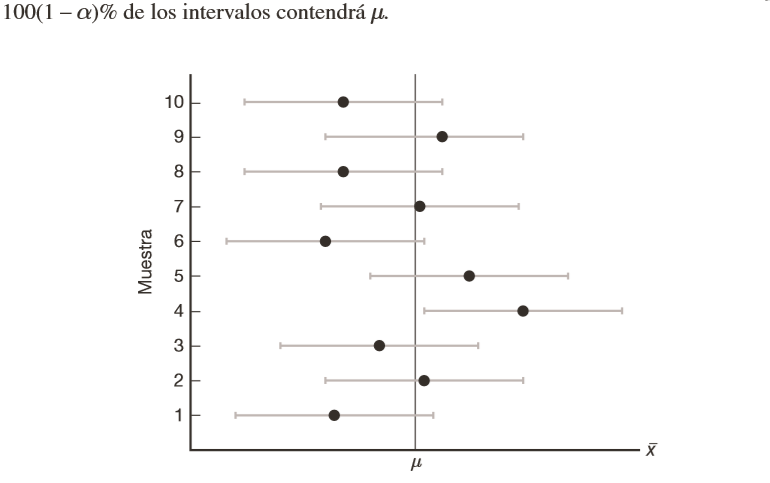
Hay muchas situaciones en que **es preferible determinar un intervalo** **dentro del cual esperaríamos encontrar el valor del parámetro**. Tal intervalo se conoce como estimación por intervalo.

Como muestras distintas suelen producir valores diferentes de y, por lo tanto, valores diferentes de L y θU, estos puntos extremos del intervalo son valores de las variables aleatorias correspondientes ΘLˆ y ΘUˆ. De la distribución muestral de seremos capaces de determinar Lˆ y Uˆ

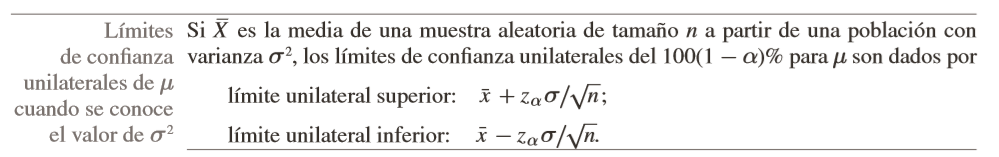
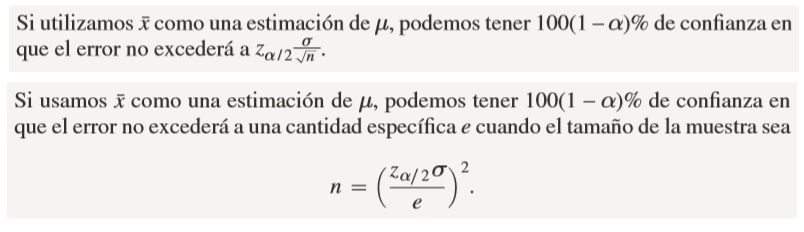
De manera que P(ΘLˆ< θ <ΘUˆ) sea igual a cualquier valor positivo de una fracción que queramos especificar.   
Si, por ejemplo, calculamos ΘLˆ y ΘUˆ, tales que **P (ΘLˆ< θ <ΘUˆ) = 1 − α**, para **0< α<1**, tenemos entonces **una probabilidad de 1 – α** de **seleccionar una muestra aleatoria que produzca un intervalo que contenga a θ**.   
**El intervalo** ΘLˆ< θ <ΘUˆ, que se calcula a partir de la muestra seleccionada, se llama entonces **intervalo de confianza** del 100(1 – α) %  
L**os extremos**, ΘLˆ y ΘUˆ, **se denominan límites de confianza inferior y superior**.  
**La fracción 1 – α se denomina coeficiente de confianza** o grado de confianza.  
Así, cuando α = 0.05, tenemos un intervalo de confianza del 95%, y cuando α = 0.01 obtenemos un intervalo de confianza más amplio del 99%.

El intervalo de confianza tan sólo sugiere que si se realiza el experimento y los datos se observan una y otra vez, aproximadamente 95% de tales intervalos contendrá el parámetro verdadero.

Los valores contenidos en el intervalo se deberían ver como **valores razonables**, dados los datos experimentales.



Ademas:



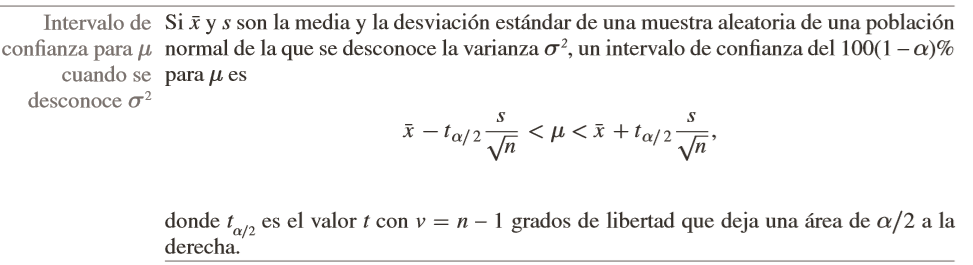
Interpretacion: En consecuencia, tenemos un 1-α% de confianza en que x promedio es menor que Limite superior

Cabe destaacar que el grado de confianza es exacto cuando las muestras se seleccionan de poblaciones normales.

**El caso en que se desconoce σ**

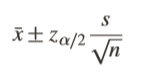
Si debemos tratar de estimar la media de una población **sin conocer la varianza** ysi tenemos una muestra aleatoria a partir de una **distribución normal**, entonces la variable aleatoria tiene una distribución t de Student con n – 1 grados de libertad.

T = (X – μ) / (S/ √n)



**Concepto de intervalo de confianza para una muestra grande**

Con frecuencia los estadísticos recomiendan que incluso cuando no sea posible suponer la normalidad, se desconozca σ y n ≥ 30, σ se puede reemplazar con s para poder utilizar el intervalo de confianza:



A menudo se hace referencia a esto como un intervalo de confianza para una muestra grande. La justificación para esto reside sólo en la presunción de que, con una muestra tan grande como 30 y una distribución de la población no muy sesgada, s estará muy cerca de la σ verdadera y, de esta manera, el teorema del límite central continuará siendo válido.

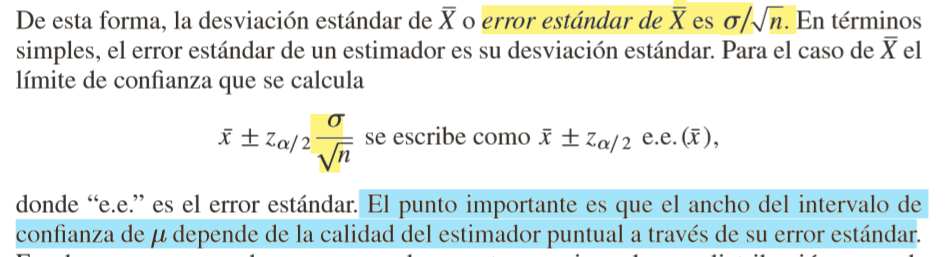
**Si usamos x¯ como una estimación de μ, podemos tener 100(1 – α) % de confianza en que el error no excederá a una cantidad específica e cuando el tamaño de la muestra sea**

**Error estándar de una estimación puntual**

Hicimos una distinción muy clara entre los objetivos de las estimaciones puntuales y las estimaciones del intervalo de confianza.   
Las **estimaciones puntuales** proporcionan un solo número que se extrae de un conjunto de datos experimentales  
Las segundas proporcionan un **intervalo razonable** para el **parámetro**, dados los datos experimentales; es decir, el 100(1 – α)% de tales intervalos que se calcula “cubren” el parámetro.

Estos dos métodos de estimación se relacionan entre sí. **El elemento en común es la distribución muestral** del estimador puntual.

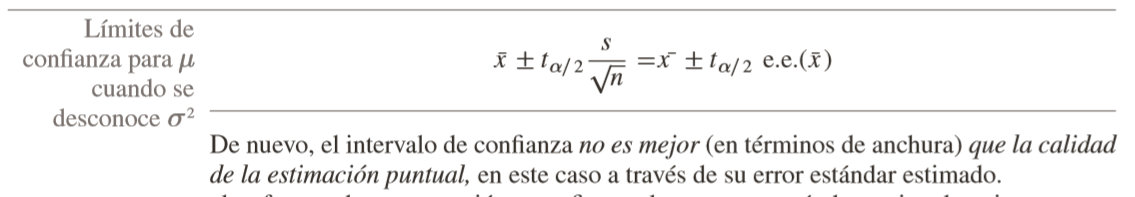
Indicamos antes que una medida de la **calidad** de un estimador insesgado es su **varianza**. La varianza de X es



El punto importante es que el ancho del intervalo de confianza de μ **depende de la calidad del estimador puntual** a través de su **error estándar**.

En el caso en que se **desconoce σ** y la muestra proviene de una **distribución normal**, **s** reemplaza a σ y se incluye el **error estándar estimado**

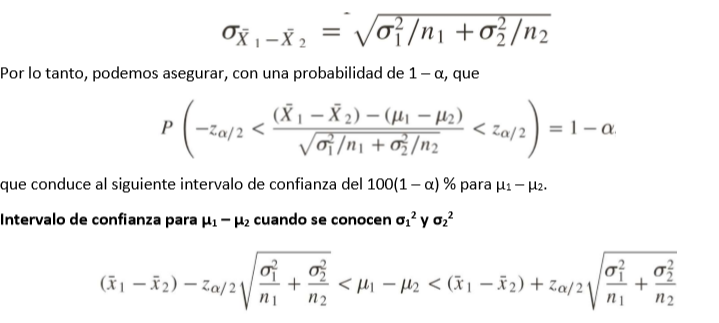
En este caso para hallar el intervalo utilizamos la distribución T



**Estimación de diferencia entre medias**

**Varianzas conocidas**

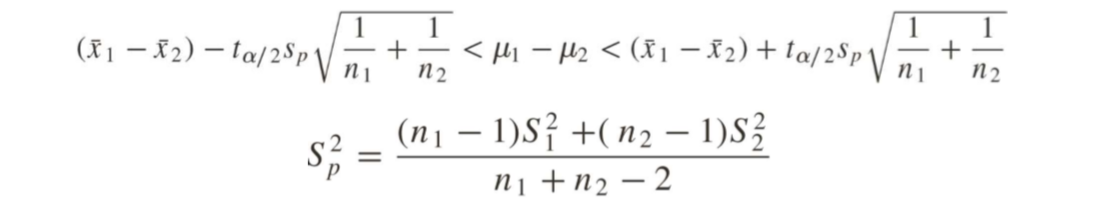
Si tenemos dos poblaciones con medias μ1 y μ2, y varianzas σ1 y σ2, respectivamente, el estadístico que da un estimador puntual de la diferencia entre μ1 y μ2 es X1 - X2. Por lo tanto, para obtener una estimación puntual de μ1 – μ2, se seleccionan dos muestras aleatorias independientes, una de cada población, de tamaños n1 y n2, y se calcula x1 - x2, la diferencia de las medias muestrales. Evidentemente, debemos considerar la distribución muestral de X1 - X2. Podemos esperar que la distribución muestral de X1 - X2 esté distribuida de forma aproximadamente normal con media μX1 - X2=μ1 - μ2 y desviación estándar.



Si las **varianzas no se conocen** y las dos distribuciones implicadas son **aproximadamente normales**, la distribución t resulta implicada como en el caso de una sola muestra.   
Si **no se está dispuesto a suponer normalidad, y tenemos muestras grandes** (digamos mayores que 30) **permitirán usar s1 y s2 en lugar de σ1 y σ2**, respectivamente, con el fundamento de que s1 ≈ σ1 y s2 ≈ σ2. De nuevo, por supuesto, el intervalo de confianza es aproximado.

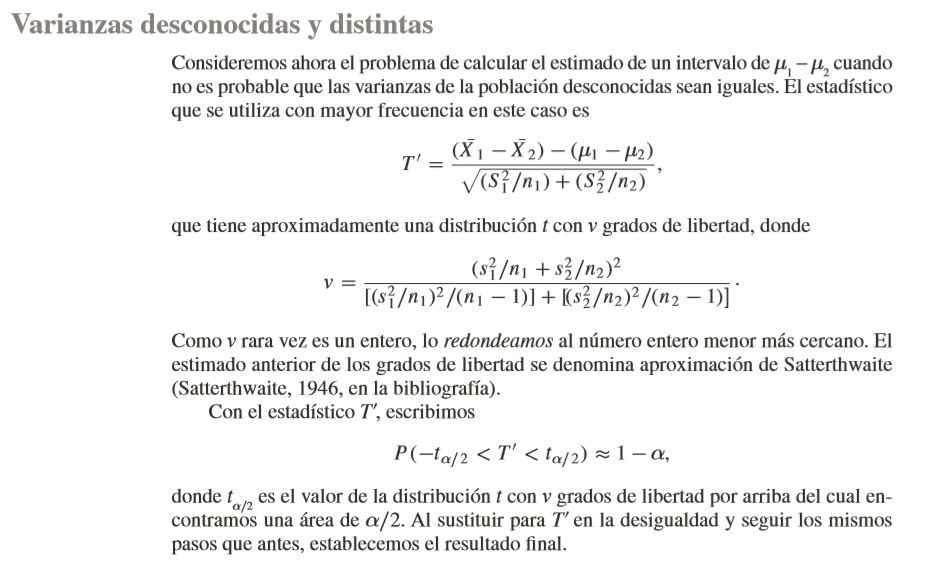
**Varianzas desconocidas pero iguales**

Si x1 y x2 son las medias de muestras aleatorias independientes con tamaños n1 y n2, respectivamente, tomadas de poblaciones **más o menos normales** con **varianzas iguales pero desconocidas**, un intervalo de confianza del 100(1 – α) % para μ1 – μ2 es dado por



Donde **Sp es la estimación agrupada de la desviación estándar** de la población y t α/2 es el valor t con v = n1 + n2 – 2 grados de libertad, que deja un área de α/2 a la derecha.

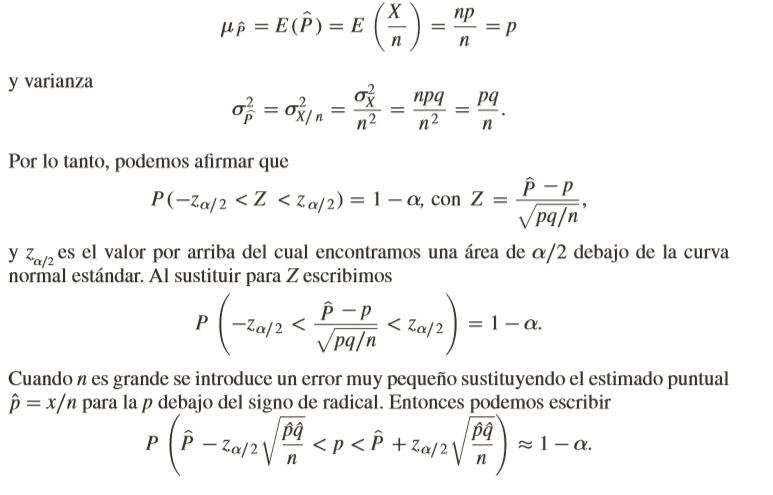
**Varianzas desconocidas y distintas**



**Estimación de una proporción**

**PARA UNA SOLA MUESTRA**

El estadístico , en donde x representa el número de éxitos en n ensayos, provee un estimador puntual de la proporción p en un **experimento binomial.** Por lo tanto, la proporción de la muestra , se utilizará como el estimador puntual del parámetro .   
**Si no se espera** que la proporción desconocida esté **demasiado cerca de 0 o de 1**, se puede establecer un intervalo de confianza para p considerando la distribución muestral de .   
Si en cada ensayo binomial asignamos el valor 0 a un fracaso y el valor 1 a un éxito, el número de éxitos, x, se puede interpretar como la suma de n valores que consta sólo de ceros y unos, y es sólo la media muestral de esos n valores.   
En consecuencia, por el **teorema del límite central**, para n suficiente mente grande está distribuida de forma casi normal con media

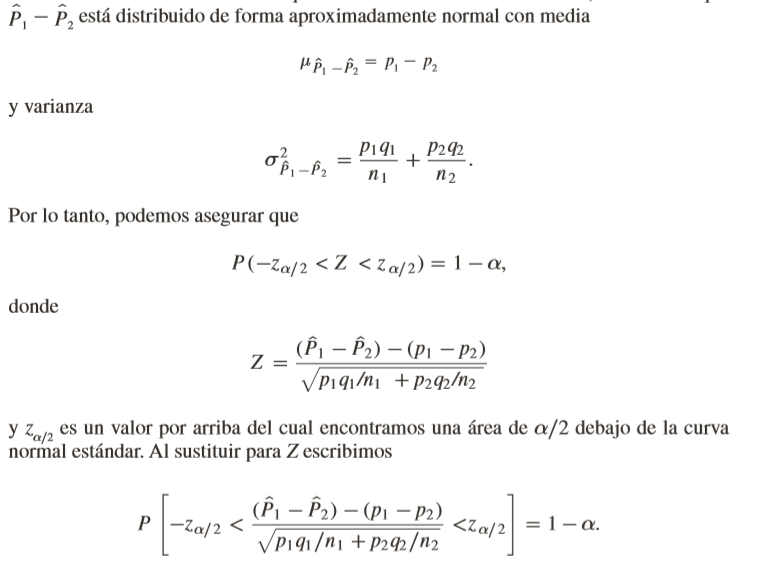


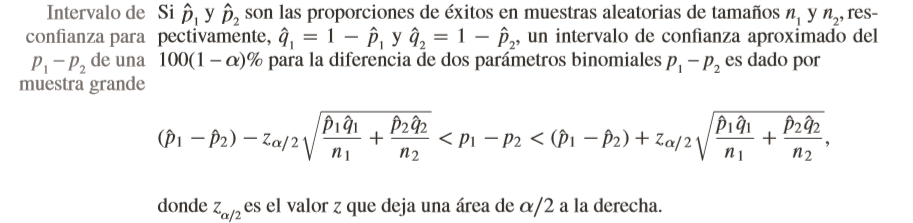
**CUIDADO**

Cuando **n es pequeña** y se cree que la proporción desconocida **p se acerca a 0 o a 1**, **el procedimiento del intervalo de confianza que se establece aquí no es confiable** y, por lo tanto, no se debería emplear. Para estar seguros se requiere que tanto n\* como n\* sean mayores que o iguales a 5.

**Dos muestras: estimación de la diferencia entre dos proporciones**

Se busca estimar la diferencia entre dos parámetros binomiales p1 y p2. Como:





**CUIDADO**

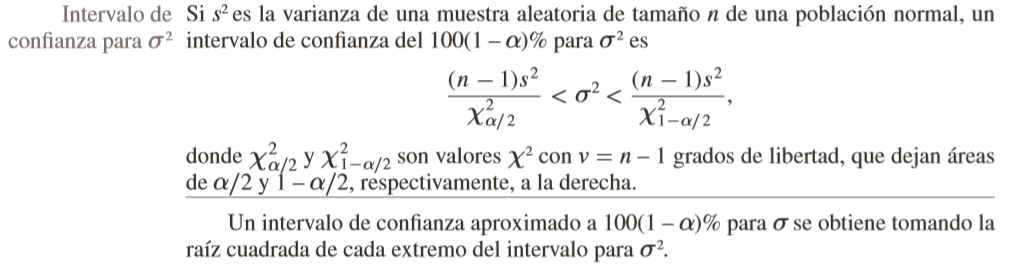
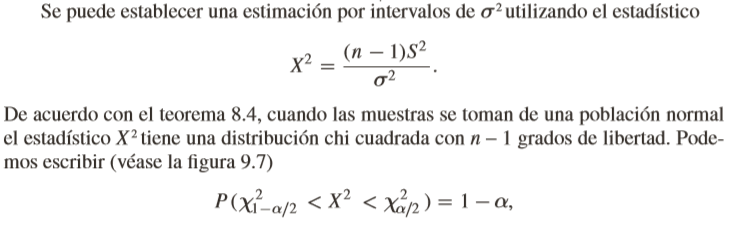
**R**emplazamos p1, p2, q1 y q2 bajo el signo de radical por sus estimaciones 1 = x1/n1],[ 2 = x2/n2] , [ 1 = 1 - 1 ]y [2 = 1 - 2]

SOLO SI (n1 1), (n11), (n2ˆ2) y (n11 ) **sean todas mayores que o iguales a 5.**

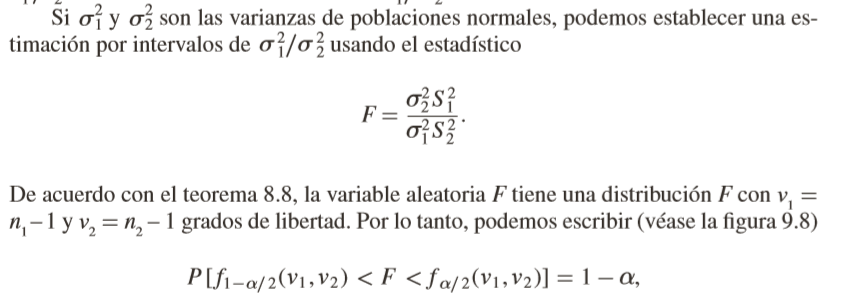
**ESTIMACION VARIANZA**

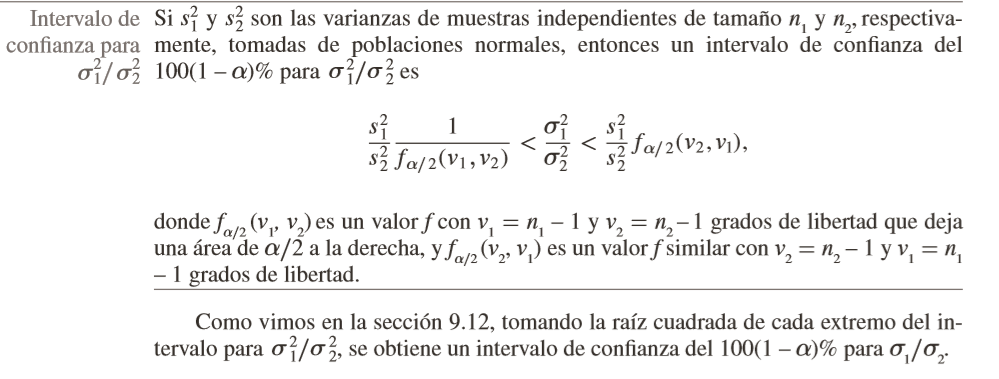
**Una sola muestra: estimación de la varianza**

Si extraemos una muestra de tamaño n de una población normal con varianza σ 2 y calculamos la varianza muestral sˆ2, obtenemos un valor del estadístico Sˆ2.



**Dos muestras: estimación de la proporción de dos varianzas**



****

**TENER EN CUENTA**

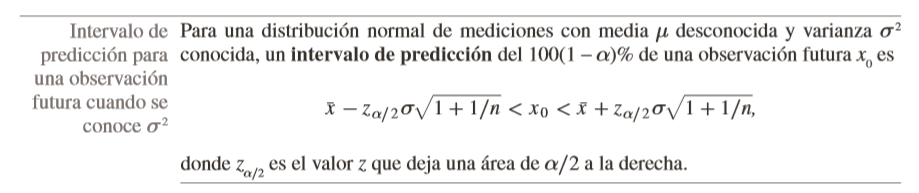
Cualquier aplicación de la **distribución t** es razonablemente insensible, es decir, **robusta**, a la suposición de normalidad.

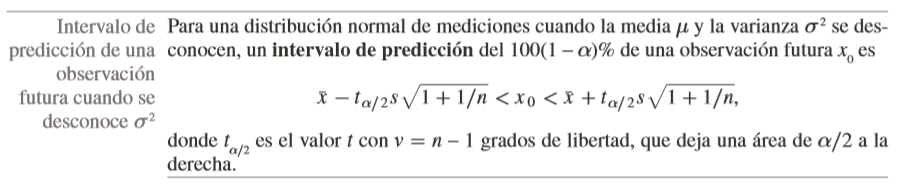
A diferencia del resultado de utilizar la distribución t, **la prueba χ2** para esta aplicación **no es robusta** para la suposición de normalidad (esto significa que cuando la distribución subyacente no es normal, la distribución muestral se aparta mucho de la distribución Ji cuadrada.)

**Intervalos de predicción**

Algunas veces, además de la media de la población, el experimentador podría estar interesado en predecir el valor posible de una **observación futura.**

Para predecir una nueva observación no basta con explicar la variación debida a la estimación de la media, también tendríamos que explicar la variación de una observación futura.





**Diferencia entre intervalos de confianza, intervalos de predicción e intervalos de tolerancia**

En el caso de los intervalos de confianza sólo se pone atención en el **parametro** de la población con el objetivo de asegurar su presencia.  
Se busca afirmar, por ejemplo, que Hay un 95% de confianza de que el verdadero contenido medio de los contenedores de ácido sulfúrico esté entre 9,74 y 10,26 litros

En el caso de **intervalos de tolerancia** nos interesa saber en dónde van a estar la mayoría de los valores por esta razón nos interesan los límites del intervalo de confianza  
**AUNQUE SEGÚN APUNTE NO ENTRAN**

En un caso totalmente diferente; Los **intervalos de predicción** se pueden aplicar cuando es importante **determinar un límite para un solo valor**. Aquí la media no es la cuestión, ni tampoco la ubicación de la mayoría de la población, lo que se requiere, más bien, **es la ubicación de una sola nueva observación.**

**¿Qué trata la inferencia estadística?   
¿Cuáles son los métodos clásicos de estimación?   
¿Qué se espera de un buen estimador?   
¿Qué significa que un estimador sea insesgado?   
¿Qué significa que un estimador sea el más eficiente?   
¿Todos los estimadores que son más eficientes son insesgados?   
¿Qué es una estimación puntual?   
¿Qué es una estimación por intervalos?   
Respecto a los tipos de estimación analizados, ¿cuál es más precisa?, ¿cuál es más confiable?   
¿Qué es el error estándar de una estimación puntual?   
¿Cómo se calcula un intervalo de confianza para estimar la media poblacional?   
Analice cada una de las posibilidades y las condiciones que deben darse para utilizarlas correctamente.   
¿Cómo se calcula un intervalo de confianza para estimar la diferencia de medias poblacionales?   
Analice cada una de las posibilidades y las condiciones que deben darse para utilizarlas correctamente.   
¿Cómo se calcula un intervalo de confianza para estimar la proporción poblacional?   
Analice las condiciones que deben darse para utilizarlo correctamente.   
¿Cómo se calcula un intervalo de confianza para estimar la diferencia de proporciones poblacionales?   
Analice las condiciones que deben darse para utilizarlo correctamente.   
¿Cómo se calcula un intervalo de confianza para estimar la varianza poblacional?   
Analice las condiciones que deben darse para utilizarlo correctamente.   
¿Cómo se calcula un intervalo de confianza para estimar el cociente de varianzas poblacionales?  
 Analice las condiciones que deben darse para utilizarlo correctamente.   
Analice las condiciones para calcular tamaños de muestra según la información disponible**.