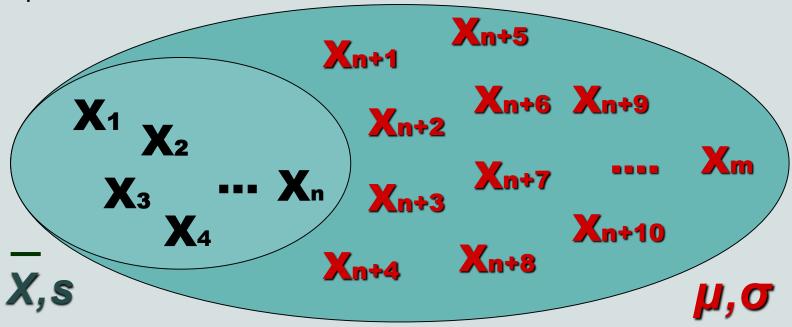
ESTADÍSTICA INFERENCIAL

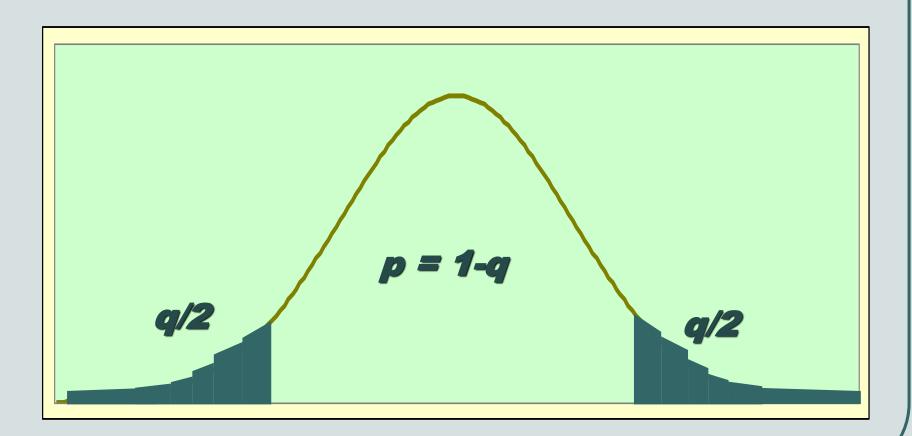
PRUEBA DE HIPÓTESIS

ESTADÍSTICA INFERENCIAL:

■ **Estadística Inferencial**: Se pretende hacer extensivos los análisis estadísticos, realizados con la muestra, a toda la población.



TEORÍA DE LAS PROBABILIDADES:



TEOREMA CENTRAL DEL LÍMITE:

- Sea X una variable continua, con promedio poblacional μ y desviación estándar poblacional σ, entonces:
 - X se <u>distribuye en forma normal</u>, cuando n es grande.
 - $\cdot \mu = \mu$
 - $\frac{\sigma}{x} = \frac{\sigma}{\sqrt{n}}$

ESTIMACIÓN PUNTUAL:

Inferencias:

- <u>Mientras más grande sea</u> <u>n</u>, <u>más se parece</u> <u>la distribución muestral a la distribución normal.</u>
 - El promedio de la muestra es una estimación puntual de μ.

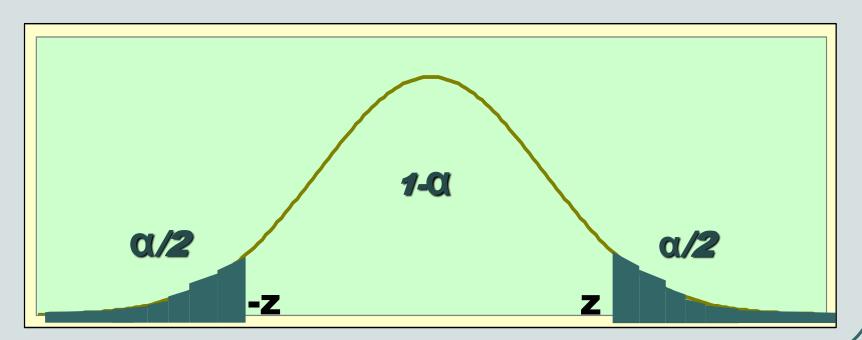
13/03/2019

5

ESTIMACIÓN POR INTERVALOS:

Así mismo, <u>µ está en el intervalo:</u>

$$< \overline{X} - z \cdot \sigma / \sqrt{n}$$
; $\overline{X} + z \cdot \sigma / \sqrt{n} >$, con p=1-\alpha.



ESTIMACIÓN POR INTERVALOS:

- □ Distribución t de Student:
 - Cuando no se conoce σ, μ está en el intervalo similar:
 - $<(\overline{x}-t.s/\sqrt{n};\overline{x}+t.s/\sqrt{n}>con p=1-\alpha.$
 - En él, <u>la distribución t-Student</u>
 <u>sustituye a Z</u> (empleando s en vez de σ).
 - t-Student es una familia de curvas, determinadas por los grados de libertad, por n-1.

Revisión de la Tarea extra-clase:

- Determine los intervalos de confianza de la distribución de frecuencias de la Actividad 2.1 (Ver <u>Hoja de Trabajo No.2</u>):
 - a) Primero, a partir de conocer el valor de **c**. (Actividad 3.5, incisos a y b)
 - b) Después, considerando que se desconoce **o**; por lo que se debe calcular previamente **s**. (Actividad 3.5, incisos a y b)

Problema:

- Comprobar una hipótesis con relación a un parámetro poblacional, a partir de una muestra representativa.
 - P/Ej.: Se conoce el promedio de las calificaciones en una asignatura de estudiantes de zonas urbanas y se quiere verificar (o rechazar) la hipótesis de que el promedio de estudiantes de zonas rurales difiere de aquel.
 - Es decir, se quiere probar H_{inv}: μ‡μ₀
- La exigencia supone una comparación; se necesita saber cuán grande es X-µ₀.

□ Procedimiento:

- Se necesita transformar la hipótesis investigativa en <u>hipótesis estadística</u>.
 - En el Ejemplo (Se conoce el promedio de las calificaciones en una asignatura de estudiantes de zonas urbanas y se quiere verificar la hipótesis de que el promedio de estudiantes de zonas rurales difiere de aquel) pueden identificarse dos hipótesis diferentes:
 - Una <u>nula</u>, **H₀**: μ=μ₀
 - Otra <u>alternativa</u>, H₁: μ‡μ₀

□ Procedimiento:

- Con vistas a avanzar hacia el objetivo, <u>se</u> debe trazar una estrategia de prueba:
 - Para tener la máxima confianza posible sobre
 H_{inv} se debe ser deliberadamente escéptico.
 - Se le dará <u>mayor credibilidad a la hipótesis nula,</u> H_0 : $\mu = \mu_0$, que a la alternativa, H_1 : $\mu = \mu_0$.
 - Sólo si la información proporcionada por la muestra <u>rechaza</u> la hipótesis nula, se considerará verdadera la hipótesis alternativa.

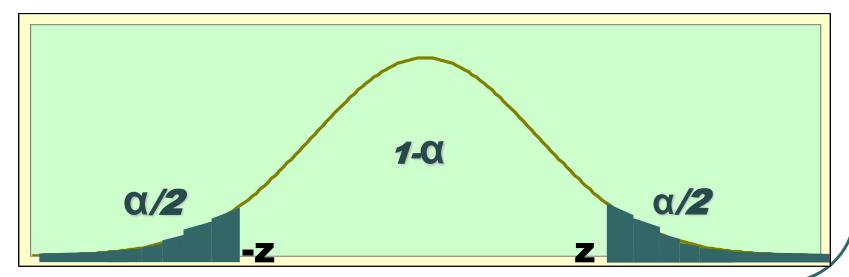
Procedimiento:

- Ahora lo que se requiere es de una <u>regla de</u> <u>decisión</u> (sobre la validez o no de H₀):
 - La misma gira en torno a saber en cuánto difiere \overline{X} de μ_0 ; es decir, cuán grande es \overline{X} - μ_0 .
- Pero se sabe que varía de muestra en muestra.
 - Luego, la información proporcionada por la muestra está sujeta a <u>cierto riesgo de cometer</u> <u>error al decidir</u>.

RIESGOS AL DECIDIR:	H ₀ es verdadera.	H ₀ es falsa.
Se rechaza H ₀	Se comete Error del Tipo I	No se comete error.
No se rechaza H ₀	No se comete error.	Se comete Error del Tipo II

Procedimiento:

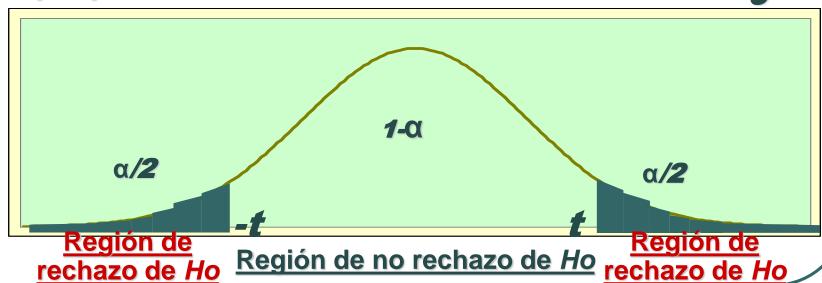
- No se puede lograr que las probabilidades α y β, de cometer error del tipo l y del tipo ll respectivamente, sean pequeñas a la vez.
- Luego, se le debe dar menor valor a α (.05 ó .01)



- □ Procedimiento:
 - Conocer <u>cuán grande</u> <u>es X-µ₀ está asociado</u> al intervalo de confianza para <u>µ</u>:
 - $<(\overline{X}-t. s/\sqrt{n}; \overline{X}+t. s/\sqrt{n}>con p=1-\alpha.$
 - Correspondiente a la distribución t-Student con n-1 grados de libertad.
 - Equivale a la expresión:

Procedimiento:

- La distribución t-Student constituye el estadígrafo de prueba.
- Con ella y α (.05 ó .01) se determinan dos regiones que permiten decidir sobre la validez o no de H₀.



Procedimiento general:

- 1. Planteamiento de las hipótesis (Hoy H1).
- 2. Determinación del <u>estadígrafo de prueba</u> (**t- Student**).
- 3. Formulación de la <u>regla de decisión</u> (especificación del valor de **a**: 0.05 ó 0.01)
- 4. Realización de los <u>cálculos</u> correspondientes, utilizando los estadígrafos muestrales y μ_0 .
- 5. <u>Decisión estadística</u>, verificando en qué región se encuentra el valor del estadígrafo de prueba.
- 6. Interpretación de los <u>resultados</u> (determinación del valor de **H**_{inv}).

Ejemplo 1:

El promedio de las calificaciones en una asignatura de estudiantes de zonas urbanas es de 7,5 puntos y se quiere verificar la hipótesis de que el promedio de estudiantes de zonas rurales difiere de aquel. Con una muestra de 30 estudiantes de zonas rurales se obtiene un promedio de 7,86 puntos y una desviación de 0,71.

1. Planteamiento de las <u>hipótesis</u>:

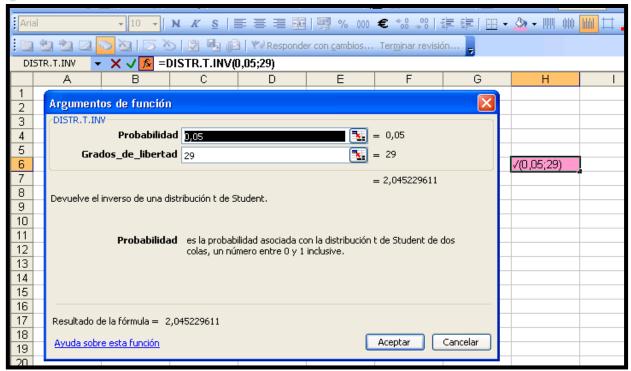
 $H_0: \mu = \mu_0, H_1: \mu \neq \mu_0$

- 2. Determinación del <u>estadígrafo de prueba:</u> t-Student, con n-1=30-1=29 grados de libertad.
- 3. Formulación de la regla de decisión:

Para t = 0.05 se tiene $t_{(29)} = 2.045$.

☐ Ejemplo 1:

• Se puede determinar el valor de **t** con Excel:



- □ Ejemplo 1:
 - 4. Realización de los <u>cálculos:</u> t = x μ₀ =2,778
 - 5. Decisión estadística:

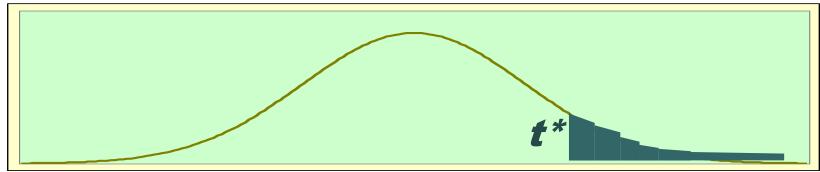
Puesto que 2,778 > 2.045, se rechaza H_o

6. Interpretación de los resultados:

Hay evidencias suficientes para afirmar, con un 95% de confianza, que se cumple **H**_{inv}; es decir, que existe diferencias entre las zonas rural y urbana.

- Consideraciones importantes:
 - H_{inv} se aprueba con una confianza de (1-U)100%;
 a U se le conoce como <u>nivel de significancia</u> de la prueba.
 - Al rechazar H_o se puede decir que el promedio muestral "difiere significativamente" de μ_0 .
 - Cuando no es posible rechazar H_o no puede decirse que ésta "se acepta", sólo que "<u>no hay</u> <u>evidencias suficientes para rechazarla</u>" o que "<u>no hay diferencias significativas</u>" (N.S.).

- Consideraciones importantes:
 - H_{inv} puede consistir en que <u>el promedio</u> <u>muestral resulta mayor (menor)</u> que μ₀. En esos casos se trabaja la distribución del estadígrafo de prueba <u>con una sola cola</u>.
 - Estudiar el <u>Ejemplo 2.3</u> de la <u>Hoja de Trabajo No.3</u>.



Región de no rechazo de Ho

Región de rechazo

Consideraciones importantes:

- Se han probado H_{inv} que involucran sólo variables continuas y en la que la hipótesis se refiere al parámetro μ₀. <u>Otras muchas</u> <u>situaciones pueden darse</u>.
 - Aún así, <u>el procedimiento general</u> de prueba de hipótesis <u>seguirá siendo válido</u>.
 - Aunque el planteamiento de <u>las hipótesis</u> <u>estadísticas sufrirán modificaciones</u>.
 - También se requerirán de <u>nuevos estadígrafos de</u> <u>prueba</u>.

☐ Ejemplo 2:

Un psicólogo opina que más del 30% de los niños del segundo ciclo tienen interés en conocer aspectos de educación sexual. Para probar su hipótesis elige una muestra de 150 niños y por medio de una entrevista detecta que 63 de ellos están interesados en el tema. Sométase a prueba la hipótesis con α=0,01.

1. Planteamiento de las hipótesis:

$$H_0$$
: p \leq 0,30, H_1 : p $>$ 0,30

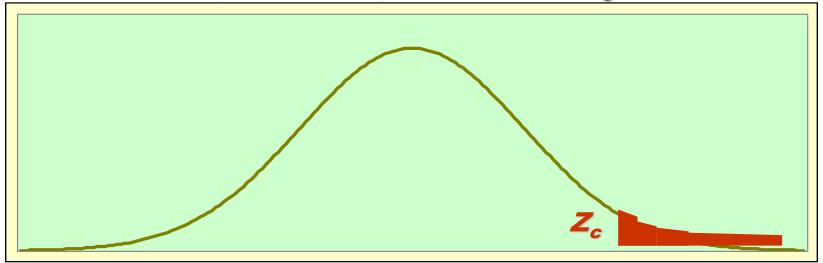
2. Determinación del estadígrafo de prueba:

$$z_c = (p-p_0)/\sqrt{p_0(1-p_0)/n}$$

$$n.p_0=150(0,30)>5 y n(1-p_0)=150(1-0,30)>5$$

- ☐ <u>Ejemplo 2</u>:
 - 3. Regla de decisión:

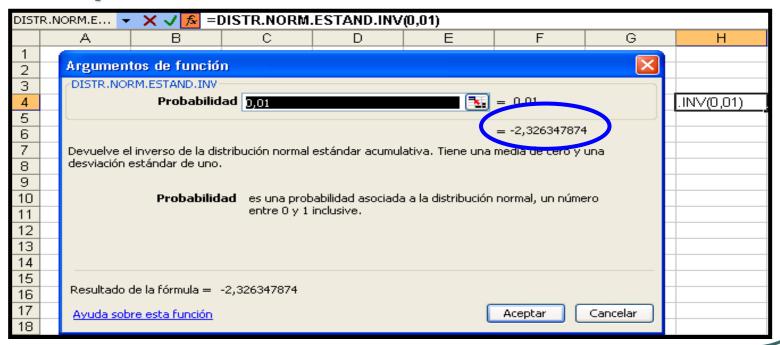
Como H_1 : p > 0,30; α quedará en la cola derecha; para $\alpha = 0,01$ se tiene $\mathbf{z}_c = 2,326$.



Región de no rechazo de Ho

Región de rechazo

- ☐ <u>Ejemplo 2</u>:
 - 3. Regla de decisión:
 - Se puede determinar el valor de **z** con Excel:



☐ <u>Ejemplo 2</u>:

4. Cálculos:

Con n=150, p=63/150=0,42 y p₀=0,30 se tiene: $z_c = (p-p_0)/\sqrt{p_0(1-p_0)/n} = 3,207$

5. Decisión estadística:

Como 3,207 > 2,326, se rechaza H_0 con α = 0,01.

6. Interpretación de los resultados:

Hay evidencias suficientes para considerar, con un 99% de confianza, que más del 30% de los niños del segundo ciclo se interesan por aspectos de la educación sexual.

☐ Ejemplo 3:

El director de una escuela considera que el nivel socio-económico de los estudiantes tiene una influencia en su rendimiento cognitivo. Para verificar su hipótesis aplicó una prueba de rendimiento y una encuesta, cuyos resultados se muestran en la siguiente tabla de contingencia.

¿Son <u>independientes</u> el nivel socioeconómico y el rendimiento cognitivo en la asignatura examinada con α=0,05?

☐ <u>Ejemplo 3</u>:

Frecuencias observadas:	Rendimiento cognitivo				
Nivel socio- económico	2	3	4	5	Total
Bajo	9	15	10	7	41
Medio	11	12	14	8	45
Alto	3	11	10	5	29
Total	23	38	34	20	115

□ Ejemplo 3:

1.Planteamiento de las hipótesis:

Ha: el nivel socio-económico y las calificaciones son independientes.

H₄: el nivel socio-económico y las calificaciones son independientes.

2. Determinación del estadígrafo de prueba:

¡Las proporciones se conservan! ¡Los resultados por niveles socio-económicos se comportan como en el total!

30

☐ <u>Ejemplo 3</u>:

Frecuencias esperadas:	Rendimiento cognitivo				
Nivel socio- económico	2	3	4		Total
Bajo	(23/115) 41	(38/115) 41	(34/115) 41	(20/115) 41	41
Medio	(23/115) 45	(38/115) 45	(34/115) 45	(20/115) 45	45
Alto	(23/115) 29	(38/115) 29	(34/115) 29	(20/115) 29	29
Total	23	38	34	20	115

☐ <u>Ejemplo 3</u>:

Frecuencias esperadas:	Rendimiento cognitivo				
Nivel socio- económico	2	3			
Bajo	23.41 /115	39 +1 /115	34.41 /115	20.41 /115 41	
Medio	23.45	38.45 /115	34.45 /115	20.45 /115 45	
Alto	2 /115	38.29 /115	34.29 /115	20.29 /115 29	
Total	23	38	34	20 115	

☐ Ejemplo 3:

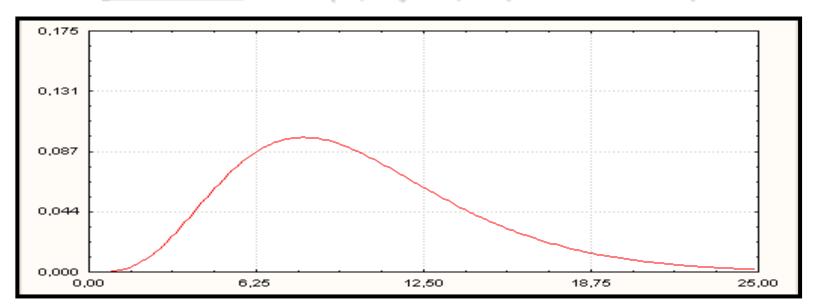
2.Determinación del <u>estadígrafo de</u> prueba:

¡Se necesita conocer cuán "grande" o "pequeña" es la diferencia entre la frecuencia observada y la frecuencia esperada correspondiente! ¡Se considerarán los cuadrados de las diferencias o; e; dado que algunas son negativas!

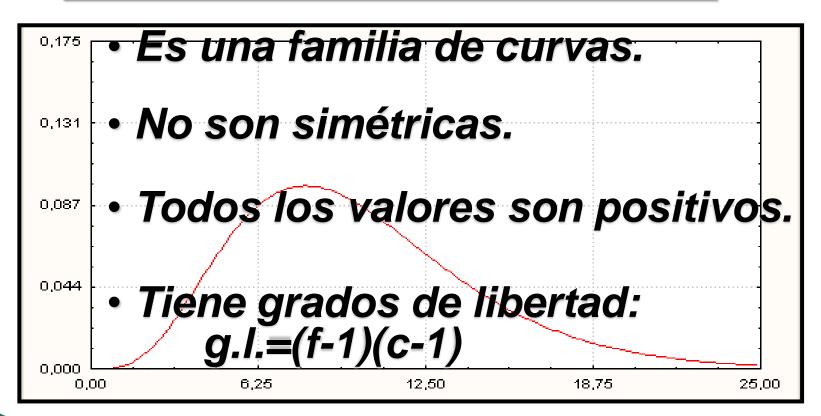
$$X^2 = \sum (o_i - e_i)^2 / e_i$$
 (Ji-cuadrada)

☐ Ejemplo 3:

2.Determinación del estadígrafo de pruebă? = \(\sum_{i} \cdot \eta_{i}\) \(\rac{1}{2} \cdot \eta_{i}\) (O_i-e_i)²/e_i (Ji-cuadrada)



Características de Ji-cuadrada:



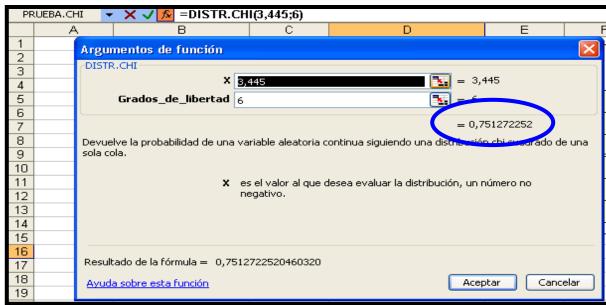
- ☐ Ejemplo 3:
 - 3. Formulación de la regla de decisión: Para α =0.05 se tiene X(6) =12,592; de modo que X_c debe ser mayor que ese valor.
 - 4. <u>Cálculos</u>:
 Sustituyendo en la fórmula,
 X_c=3.445.
 - 5. <u>Decisión estadística</u>: Como X_c =3,445 < 12,592=X(6) <u>no se</u>

☐ <u>Ejemplo 3</u>:

3. Regla de decisión:

 Se puede determinar directamente la probabilidad p con la <u>Distribución Chi</u>, en

Excel:



Tarea extra-clase:

- 1. En la *Hoja de Trabajo No.3*:
 - a) Estudie el Ejemplo 2.3.
 - b) Realice la <u>Actividad 2.1</u>, referida a una prueba de hipótesis de una sola cola.
- 2. Realice la prueba de independencia del *Ejemplo 3* para α=0.10