TALLER DE ESTADÍSTICA

7. MUESTRAS Y ESTIMACIONES. INFERENCIA ESTADÍSTICA.

MAURICIO CONTRERAS

MUESTRAS Y ESTIMACIONES EN LA ESO

Introducción

¿Cómo debe seleccionarse la muestra para que sea representativa de la población?. ¿Qué tamaño debe tener?. ¿Hasta qué punto es fiable la información obtenida en la muestra?. ¿Qué grado de error estaríamos dispuestos a admitir al extrapolar los datos de la muestra a toda la población?. Preguntas como éstas y otras parecidas son analizadas por la Inferencia Estadística (también llamada Estadística matemática).

El estudio de la Estadística inferencial es realmente muy difícil si no se utilizan recursos apropiados. En 3º y 4º de ESO y en Bachillerato se pretende que los estudiantes realicen actividades en las que, de una forma u otra, tengan que responder a las preguntas anteriores. Por ejemplo, la realización de encuestas reales en el centro sobre diversos temas (aficiones, uso del tiempo libre, etc) es una buena ocasión para tratar cuestiones relativas al muestreo y es, además, muy motivador.

La selección de la muestra se puede hacer de diferentes maneras (y es interesante recoger las propuestas de los alumnos para discutir si hay aleatoriedad o no) y con generadores muy variados (dados, ruletas, etc). La calculadora gráfica permite generar números aleatorios, seleccionar muestras, representar los datos, calcular medidas de centralización y dispersión, hacer simulaciones, obtener estimaciones de parámetros, etc. Posteriormente el estudiante deberá tomar una decisión a partir de la información obtenida. Esta sesión se dedicará a analizar las posibilidades de estos recursos en el aula.

1. Muestreo

- **<u>AFICIONES</u>**

a) ¿Cuáles son las aficiones de tus compañeros de centro?. ¿Cómo podrías saberlo?. ¿Es necesario preguntar a todos ellos?

Para recoger información de una población no es necesario obtener todos los datos, sino solamente los correspondientes a una parte de la población, a una <u>muestra</u>. Posteriormente, usaremos los datos de la muestra para inferir conclusiones sobre el comportamiento de la población. Surgen entonces algunas preguntas de interés:

- ✓ *¿Cuál debe ser el tamaño de la muestra para que sea representativa de la población?.*
- ✓ *¿Cómo debe seleccionarse la muestra para que la información no esté sesgada?.*
- ✓ *¿Hasta qué punto es fiable la información obtenida de la muestra?.*
- ✓ *¿Es válido predecir el comportamiento de la población basándose en los datos de la muestra?.*

b) Vamos a diseñar una encuesta para conocer las aficiones preferidas en tu centro. Piensa en cómo se puede diseñar la encuesta:

- ✓ Qué preguntas hacer.
- ✓ Cómo formular las preguntas para que no condicionen la respuesta.
- ✓ A cuántas personas hay que preguntar.
- ✓ A qué personas hay que preguntar.
- ✓ Cómo debe seleccionarse la muestra.

c) Con el modelo de encuesta diseñado, recoge información de tu centro sobre aficiones de tiempo libre. Construye tablas de frecuencias como las siguientes:

AFICIONES	PRIMERO	SEGUNDO	TERCERO	CUARTO	TOTAL
Cine					
Teatro					
TV					
Música					
Fútbol					
Baloncesto					
Atletismo					
Motociclismo					
Informática					
Excursiones					
TOTAL					

d) Representa gráficamente la información obtenida utilizando distintos diagramas:

- ✓ Dibuja, en unos mismos ejes, un diagrama de barras que muestre el número de aficionados a cada actividad para cada uno de los cursos.
- ✓ Dibuja, en unos mismos ejes, un diagrama de barras que muestre el número de aficionados a cada actividad para cuarto curso comparándolo con el total de encuestados.
- ✓ Haz lo mismo para comparar el total con los estudiantes de primero. Comenta las diferencias que observes.
- ✓ Dibuja un diagrama de sectores que muestre la información del total de encuestados.
- ✓ Dibuja un diagrama de sectores que muestre la información de cuarto curso y compáralo con el correspondiente al total de encuestados.

e) Analiza la información obtenida:

- ✓ ¿Qué proporción de estudiantes de primero hay en la muestra?. ¿Y de segundo?.
- ✓ ¿Qué proporción de encuestados son aficionados al cine?. ¿Y a la música?.

- ✓ ¿Qué proporción de estudiantes de tercer curso son aficionados al cine?. ¿Y al atletismo?.
- ✓ Si elegimos al azar un estudiante de tu centro, ¿cuál es la probabilidad de que sea aficionado a la Informática?. ¿Y al teatro?.
- ✓ Elegimos al azar un estudiante de segundo curso. ¿Qué probabilidad hay de que sea aficionado al fútbol?. ¿Y de que sea aficionado al baloncesto?.
- ✓ Elegimos al azar un estudiante de tu centro y resulta ser aficionado al motociclismo. ¿Hay muchas posibilidades de que sea de primero?. ¿Y de que sea de cuarto?.

Elegimos al azar un estudiante de tu centro. Designamos:

A = el estudiante elegido es de segundo curso.
B = el estudiante elegido es aficionado al cine.

Entonces el suceso que consiste en que el estudiante elegido es aficionado al cine sabiendo que es de segundo curso, se representa por B/A y se llama <u>suceso B condicionado por A</u>.

La probabilidad de este suceso, es decir, la probabilidad de que el estudiante elegido sea aficionado al cine sabiendo que es de segundo curso, se representa por p(B/A) y se llama <u>probabilidad condicionada</u>.

- **SONDEO ELECTORAL**

Se ha realizado una encuesta para conocer las intenciones de voto de los españoles por un determinado partido político A. En la ficha técnica del sondeo, leemos que el límite máximo de error es ± 2’8 %, es decir, ± 2’8 puntos de porcentaje, con una probabilidad del 95 %. En dicha encuesta se estima que el partido A obtendrá un porcentaje de votos del 33 %.

¿Entre qué valores mínimo y máximo puede fluctuar el porcentaje de votos del partido A, con una probabilidad del 95 % ?.

Si a es el porcentaje mínimo y b el máximo, se cumple que $a = 33 - 2'8$, $b = 33 + 2'8$. *El intervalo (a, b) se llama <u>intervalo de confianza con un nivel de confianza del 95%</u>. Se cumple que la probabilidad de que el porcentaje p de votos del partido A esté entre a y b es del 95%, o sea:* $p(a < p < b) = 0'95$.

- **ESTATURA MEDIA**

a) Para estimar la estatura media de los 934 estudiantes de un instituto, extraemos una muestra de 53 de ellos. La media de la muestra es 172’6 cm. Expresa este resultado sabiendo que en la ficha técnica se dice que el error máximo es de ± 1’8 cm, con una probabilidad de 0’90.

b) Si con el mismo estudio anterior admitimos que se cometa un error de ± 2’6 cm, el nivel de confianza, ¿será superior o inferior a 0’90?.

c) ¿Cómo podremos aumentar el nivel de confianza manteniendo la cota de error en ±1’8 cm?.

- **JUDÍAS**

Para contar el número de judías que hay en una bolsa procedemos así:

1) Sacamos un puñado de ellas, las señalamos, las contamos (187, por ejemplo) y las devolvemos a la bolsa.
2) Revolvemos largamente para que se mezclen y volvemos a extraer un buen montón, 411, de las cuales hay 44 señaladas.

¿Cuántas judías hay en la bolsa?.

- **SONDEO DE OPINIÓN**

En un sondeo de opinión entre los jóvenes valencianos de 15 a 24 años, una de las preguntas era: ¿Justificas que alguien acepte un soborno en su trabajo?. Respuesta: (1, nunca; 10, siempre). En una muestra de 2000 individuos, se obtuvo una puntuación de 2'63.

a) Expresa este resultado sabiendo que en la ficha técnica se dice que el error máximo es de ± 1'22 con un nivel de confianza del 95 %.

b) Si el error máximo fuera ± 0'6, ¿el nivel de confianza sería mayor o menor?.

- **COLONIA INFANTIL**

Una empresa de cosmética y perfumería desea conocer la aceptación de su nueva colonia infantil. Para ello decide preguntar a mil personas sobre la utilización y ventajas de dicha colonia.

a) ¿A qué mil personas debe elegir?.

   * A las mil primeras personas que entren en una boca de metro.
   * A las mil primeras personas que esperen a sus hijos a la puerta del colegio.
   * A las mil primeras personas que salgan del hipermercado.
   * A las mil primeras personas que abran la puerta a las diez de la mañana, en diversos barrios de la ciudad.

b) ¿Qué población queda excluida en cada una de las muestras anteriores?.

c) ¿A qué mil personas preguntarías tú?.

- **EQUIPO DE BALONCESTO**

Queremos seleccionar cinco alumnos de cada curso de ESO para realizar una competición de baloncesto. Carlos propone el siguiente procedimiento para seleccionar a los cinco alumnos de entre los 30 que componen el grupo de tercero de ESO C:

"Cojo 30 folios en blanco y en 5 de ellos pongo una marca; los doblo varias veces y hago que cada compañero coja uno. Los cinco de la señal formarán el equipo de baloncesto de la clase ".

a) ¿Crees que esta muestra elegida por Carlos es representativa de 3º de ESO C?. ¿Por qué?.

b) Inventa algún otro procedimiento para elegir a los 5 alumnos de tercero de ESO C.

- **ROPA DEPORTIVA**

Al director del Instituto le acaban de hacer una oferta de chandalls. Debe decir en diez minutos cuántos necesita de cada talla, para que graben en ellos el nombre del Centro.

Piensa varias posibilidades para elegir una muestra de 30 alumnos:

∗ Una clase de segundo de BUP.

∗ Las chicas de COU.

∗ Los tres primeros alumnos que encuentre de cada uno de los diez cursos del Instituto.

∗ Los 30 primeros alumnos que encuentre en el pasillo de la primera planta.

a) ¿Cuál crees que es la muestra más representativa de todos los alumnos?. ¿Por qué?.

b) ¿A qué parte de los alumnos excluyen las restantes muestras?.

c) Elige otra muestra que también represente a todos los alumnos del Centro.

- **EXTRAE MUESTRAS**

a) Utilizando la tabla de números aleatorios extrae una muestra de 12 individuos de una población de 720 habitantes. Explica detalladamente el procedimiento usado.

b) Con ayuda de la calculadora elige una muestra aleatoria de 15 personas, entre las 700 de un barrio determinado.

c) De una población de 1000 personas queremos extraer una muestra cuyo tamaño sea el 2'5 % de la población. Calcula el tamaño de la muestra y utiliza la tabla de números aleatorios para obtener esa muestra.

- **¿LISTA ALEATORIA?**

En un ordenador del Instituto apareció en pantalla la siguiente colección de números:

.4210367280 .3201243560 .0132789023. 9179689724 .3127581365 .0120646175

¿Se puede considerar esta colección como una tabla de números aleatorios?.

a) Haz un recuento del número de veces que aparece cada dígito, construye la tabla de frecuencias correspondiente. Compara la frecuencia relativa de cada cifra con su probabilidad teórica.

b) Aplica el test de poker: cuenta el número de clases del tipo aabcd que hay cuando dividimos la tabla en grupos de cinco dígitos. Compara la frecuencia relativa con la probabilidad teórica (0'5040).

- **<u>¿CIFRAS AL AZAR ?</u>**

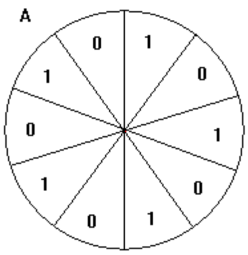
¿Son aleatorias las cifras del número π?.

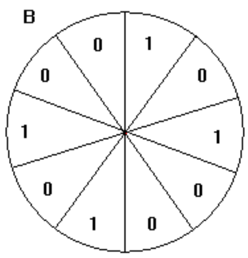
3’ 14159 26535 89793 23846 26433 83279 50288 41971 69399 37510 58209
74944 59230 78164 06286 20899 86280 34825 34211 70679 82148 08651
32823 06647 09384 46095 50582 23172 53594 08128 48111 74502 84102
70193 85211 05559 64462 29489 54930 38196 44288 10975 66593 34461
28475 48235 37867 83165 27120 19091 45648 56692 34603 48610 45432
66482 13393 60726 02491 41273 72458 70066 06315 58817 48815 20920
96282 92540 91715 36436 78925 90360 01133 05305 48820 46652 13841

a) Cuenta el número de veces que aparece cada dígito y compara la frecuencia relativa con la probabilidad teórica (0’1).

b) Aplica el test de poker. Cuenta el número de clases del tipo abcde, aabcd y aabbc. Compara las frecuencias relativas con las probabilidades teóricas. ¿Qué conclusiones obtienes?.

- **<u>¿CUÁL ES LA RULETA?</u>**





Hemos girado cada una de las ruletas 200 veces y hemos anotado los resultados en estas series:

Serie 1	1	0	0	1	0	0	1	1	0	1	0	0	0	1	0	0	1	0	1	0
	1	1	0	0	0	1	1	0	0	1	0	0	1	1	0	0	1	1	0	1
	1	0	1	0	0	0	0	1	0	1	0	0	1	1	0	1	0	0	1	0
	0	0	0	1	0	1	0	1	0	1	1	0	0	0	1	1	1	0	1	0
	1	0	0	1	0	1	0	1	0	1	0	1	0	0	1	1	1	0	1	0
	0	0	0	1	1	0	1	0	1	0	0	1	0	1	0	0	0	1	1	0
	0	1	1	0	1	0	1	1	0	0	0	1	1	1	0	1	0	1	0	1
	0	0	1	1	1	0	1	0	1	0	0	1	0	0	1	0	1	1	0	1
	0	0	0	1	0	1	1	1	0	1	0	1	0	0	1	0	1	1	0	1
	1	0	1	0	0	0	1	0	1	0	1	1	0	1	1	0	0	1	0	1

Serie 2	0	1	0	0	1	0	1	0	0	1	0	1	0	0	0	1	0	0	1	0
	1	1	0	1	0	1	0	1	0	1	1	0	0	0	1	0	1	0	0	0
	0	1	0	1	0	1	0	1	0	1	0	1	0	0	1	1	1	0	0	0
	1	0	0	1	1	0	1	0	1	0	0	1	0	1	0	0	0	1	0	0
	0	1	0	0	1	0	1	0	0	0	0	1	0	1	0	0	0	0	0	1
	1	0	0	1	0	0	1	0	0	1	0	0	0	1	0	0	1	0	1	0
	1	0	1	0	1	0	0	1	0	1	0	0	1	1	0	1	0	0	1	0
	1	0	0	0	1	0	0	0	0	1	0	0	1	1	0	1	0	1	0	0
	1	0	1	0	0	1	0	0	1	0	0	0	1	0	1	0	1	0	1	0
	0	0	1	0	0	0	0	1	0	0	1	0	0	1	1	0	0	1	0	0

Cada serie se ha obtenido girando una de las ruletas. ¿Cuál?. Explica.

Muestreo significa obtención de información a partir de muestras.

Población es el conjunto de datos o valores que se desea estudiar.

Una muestra es una parte del conjunto de datos estadísticos que se desea estudiar. Generalmente, el conjunto de datos es tan amplio que no se puede extraer la información directamente de todos ellos, sino que hay que seleccionar una muestra y limitar el estudio estadístico a los valores de la muestra.

Es posible obtener información bastante fiable de una población estudiando muestras obtenidas al azar. Esta información estará siempre afectada por un cierto grado de incertidumbre, pero el hecho de que las muestras sean extraídas al azar garantiza que las predicciones acerca de la población tengan alguna fiabilidad.

Debemos hacer la hipótesis de que las muestras aleatorias son representativas de la población de que proceden. Los elementos en una muestra obtenida al azar están en parecida proporción que en la población de la que se han obtenido.

Cuanto mayor es el tamaño de la muestra, mayor es la confianza que podemos tener en nuestra predicción.

- **¿CÓMO ES EL DADO?**

Un dado cúbico tiene todas sus caras marcadas con ceros y unos, pero no sabemos en cuántas caras hay 0 ni en cuántas hay 1.

Hemos lanzado 300 veces el dado y éstos son los resultados:

1 1 1 1 1 1 0 1 1 1 1 1 1 0 1 1 0 1 0 1 1 1 1 1 1 1 1 0 1 1
1 1 1 1 1 1 0 1 0 1 0 1 1 1 0 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1
1 1 1 1 1 1 1 1 1 1 1 0 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1
1 1 1 0 1 0 1 0 0 1 1 1 0 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 0 1
1 1 1 1 0 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1
1 1 1 1 1 1 0 1 1 1 1 1 1 0 0 1 1 1 1 0 0 1 0 1 0 1 1 1 0 1
0 1 1 1 1 1 0 1 1 0 1 1 1 1 1 1 1 0 0 1 1 0 1 0 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 0 0 1 1 1 0 1 1 0 1 0 1 1 1
1 1 1 1 1 1 1 1 0 0 1 0 1 1 1 1 1 1 1 0 1 1 1 1 0 1 0 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 0 0 1 1 1 1 1 1 1 1 1 1

Al tirar otra vez 300 veces el dado hemos obtenido la siguiente serie de ceros y unos:

1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 0 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1
1 1 1 1 1 1 1 0 1 1 1 1 0 1 1 1 1 1 1 0 1 1 1 1 0 0 1 0 1 1
1 1 1 0 1 1 1 1 1 1 0 1 0 1 0 1 0 1 1 1 1 0 1 1 1 0 0 0 0 1
1 1 1 1 1 0 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 0 1 1 1
1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 0 0 1 1 1 1 1 1 0 1 1 1
1 1 1 1 1 1 1 1 0 1 1 1 0 0 1 1 1 0 1 1 1 1 1 1 0 0 1 1 1 1
1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 0 1
1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 0 1 1 1
1 1 1 1 0 1 0 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 0 1 1 1 1 1 0

¿Cuántos ceros y cuántos unos crees que hay en el dado?.

- **<u>COMPOSICIÓN DE UNA BOTELLA</u>**

Una botella contiene 20 bolas de colores negro, rojo y verde. No sabemos cuántas de cada color, ni podemos verlo, porque la botella es opaca. Solo podemos ver, cuando la tumbamos, el color de la bola que queda junto al tapón, que es transparente.

A lo largo de varios dias hacemos 1000 veces la experiencia de agitar, inclinar la botella y anotar el color de la bola que se ve. Hemos obtenido estos resultados:

BOLA	NEGRA	ROJA	VERDE
FRECUENCIA	461	343	196

¿Cuál crees que puede ser la composición de la botella?

- **<u>EL ESTANQUE</u>**

Material: un sobre opaco con un número (desconocido para los alumnos) de fichas de colores.

Organización: grupos de cuatro.

En un estanque hay peces de distintas especies. Queremos saber qué proporción hay de peces de cada especie y cuántas especies diferentes hay, pero, como el agua está muy turbia, no podemos contarlos a simple vista.

Decidimos sacar un pez, anotar su especie (pero, para que no muera tenemos que devolverlo inmediatamente al agua) y repetir la misma operación varias veces.

¿Qué porcentaje de peces de cada especie hay?.

> *Se puede simular este problema utilizando un sobre con fichas de distintos colores (cada color representa una especie distinta de peces) y extrayendo diferentes muestras con reposición. Después se traslada la información de la muestra a la población de peces.*

- **<u>DADO OCTAÉDRICO</u>**

Tenemos un dado octaédrico (poliedro de 8 caras, triángulos equiláteros). Sus caras están numeradas con ceros y unos, pero no sabemos cuántos ceros ni cuántos unos hay. Al lanzarlo 300 veces, hemos obtenido los resultados:

1 1 1 1 1 0 1 0 1 1 1 0 1 1 1 1 0 1 1 1 0 1 1 0 1 1 1 1 0 1
1 0 0 1 0 0 0 0 1 0 0 1 1 1 1 0 0 1 0 0 1 0 0 0 1 1 0 0 1 1
1 0 0 0 0 0 0 0 0 1 1 0 1 1 0 1 1 0 1 1 0 1 0 0 1 1 0 0 1 1
1 0 1 1 1 1 1 1 1 0 0 1 0 1 0 1 1 1 0 1 1 1 1 0 1 0 1 1 1 1
1 1 1 1 1 1 0 1 1 0 1 0 0 1 1 1 1 1 1 1 0 1 1 0 0 0 1 1 0 1
0 1 1 1 0 0 0 1 1 1 0 1 1 1 1 1 1 0 1 1 1 0 0 0 1 1 0 0 0 0
1 1 1 1 1 1 0 1 1 0 1 1 0 1 1 1 1 1 1 1 0 0 0 1 1 0 1 0 1 1
0 0 1 0 1 0 1 1 1 1 0 1 0 0 1 1 1 1 1 1 1 0 0 0 1 1 1 0 1 1
1 1 1 1 0 1 1 1 1 0 1 1 1 0 1 1 1 0 1 1 0 0 1 0 0 1 0 1 0 0
1 0 1 1 0 1 0 1 1 1 1 1 0 0 1 0 0 0 0 1 0 0 1 1 1 0 0 0 1 0

¿Cuántos ceros y cuántos unos crees que hay en el dado?.

- **DADO CÚBICO**

Cada una de las caras de un dado cúbico se ha marcado con un 1 o con un 0, pero no sabemos cuántas de ellas se han marcado con 1 y cuántas con 0.

Al lanzar este dado 50 veces se obtuvo:

1 1 1 0 0 1 1 1 1 0 0 1 0 1 1 1 1 1 1 0 1 1 1 1 0 0 0 1 1 1 1 0 1 1 1 1 1 1 0 1 1 1 1 1 1 0 1 1 1 1

¿Cuántos ceros y cuántos unos podría tener el dado?.

- **¿CUÁNTOS PECES HAY EN EL ESTANQUE?**

En un estanque hay peces de una sola especie. Queremos saber cuántos hay, pero, como el agua está muy turbia, no podemos contarlos a simple vista. Decidimos sacar unos cuantos, marcarlos para distinguirlos de los otros, devolverlos al agua, sacar una segunda muestra en la que esperamos que haya peces marcados y sin marcar.

a) Con esta información, ¿podrías dar dos valores (máximo y mínimo) entre los cuales esté comprendido el número de peces del estanque?.

b) Supongamos que hemos marcado 10 peces, los hemos devuelto al agua y, en una segunda muestra hemos extraído 20 peces de los que hay 2 marcados. ¿Cuántos peces crees que habrá – aproximadamente – en el estanque?.

El problema puede ser simulado con una botella opaca que contenga bolas de colores. Puede utilizarse también una bolsa con bolas de la que se harán sucesivas extracciones con devolución.

Si se sustituye un número determinado de bolas por otras marcadas y se extraen muestras, el porcentaje de éstas en las muestras debe ser similar al de la botella.

Si al sustituir, por ejemplo, diez bolas de la botella por diez bolas azules y extraer cada grupo 200 bolas, mediante 20 muestras de diez bolas cada una, se obtiene en la clase una media de 19’67 bolas azules, el número total de bolas, N, debe verificar aproximadamente:

19’67 / 200 = 10 / N   de donde se puede determinar el tamaño de la población N.

- **JUDÍAS**

Para contar el número de judías que hay en una bolsa procedemos así:

I. Sacamos un puñado de ellas, las señalamos, las contamos (187, por ejemplo) y las devolvemos a la bolsa.

II. Revolvemos largamente para que se mezclen y volvemos a extraer un buen montón, 411, de las cuales hay 44 señaladas.

¿Cuántas judías hay en la bolsa?.

MUESTRAS Y ESTIMACIONES EN BACHILLERATO

Introducción

En Bachillerato se puede profundizar en el estudio de las técnicas de selección y procedimientos de muestreo, así como en el análisis de los datos obtenidos y la formulación de conjeturas. Se trata de seleccionar una muestra y utilizar los datos muestrales para realizar estimaciones sobre algún parámetro de la población. La calculadora gráfica es un poderoso instrumento que permite analizar fácilmente el comportamiento de los parámetros muestrales, así como obtener intervalos de confianza sobre los parámetros de estudio. En las siguientes actividades se muestran algunos ejemplos experimentados en 1º y 2º de Bachillerato en la asignatura de Matemáticas Aplicadas a las Ciencias Sociales.

1.– Tipos de muestreo

- **MUESTRAS ALEATORIAS**

Población *es el conjunto de individuos, cuyas características se pretenden estudiar.*

Muestra *es un subconjunto de la población.*

En Estadística se necesita obtener una muestra de ***n*** *elementos de una población de* ***N*** *individuos con el propósito de extraer conclusiones sobre la población a través de la muestra. Si la población es muy numerosa no tiene sentido obtener información de todos sus individuos, por razones de tiempo y dinero. Para recoger información acerca de la población se selecciona una muestra, es decir un subconjunto de la población y se efectúa con sus individuos una* ***encuesta****. Algunas preguntas de interés :*

∗ *¿**Cómo seleccionar la muestra** para que sea representativa de la población y no esté sesgada ?.*

∗ ***¿Cuál es el tamaño idóneo de la muestra ?****.*

Si la muestra es demasiado pequeña puede que la información obtenida no sea representativa de la población. Al aumentar el tamaño de la muestra se obtiene una mejor información, pero el tamaño no puede ser excesivo, por razones económicas.

∗ ***¿Es fiable la información obtenida en la muestra ?****.*

¿Hasta qué punto es representativa de la población la información contenida en la muestra?.

Estas cuestiones sobre tamaño y nivel de confianza de una muestra se estudian en ***INFERENCIA ESTADÍSTICA****.*

¿Cómo se selecciona una muestra ?

Para que la muestra sea representativa, debe ser una imagen miniaturizada de la población. Los caracteres interesantes en la muestra deben aparecer en la muestra con la misma proporción que en la población.

Para que esto ocurra y la información no presente sesgos, seleccionamos los individuos que componen la muestra al azar, mediante un sorteo. La muestra obtenida por este procedimiento se conoce con el nombre de **muestra aleatoria**. *En el caso de muestra aleatoria, todos los elementos de la población tienen la misma probabilidad* ***n / N*** *de formar parte de ella.*

Para obtener una muestra aleatoria se numeran los elementos de la población de forma que todos los números identificativos tengan la misma cantidad de dígitos. A continuación se eligen ***n*** *elementos con ayuda de la tabla de números aleatorios, para lo que basta leer números de la tabla de números aleatorios (o de la calculadora), rechazando aquellos que no correspondan a ninguno de los números identificativos de la población. La muestra estará formada por aquellos individuos de la población cuyos números de orden coincidan con los* ***n*** *números aleatorios seleccionados.*

***Ejemplo 1**.-* *Para extraer una muestra de 400 individuos de una población de tamaño 10000 numeramos sus elementos y escogemos 400 números diferentes de cuatro cifras de la tabla de números aleatorios (el 0000 será el 10000). Durante el proceso de selección de estos 400 números eliminaríamos los que aparezcan repetidos. A continuación realizaríamos una encuesta, preguntando a los 400 individuos que componen la muestra.*

***Ejemplo 2**.-* *Se desea confeccionar una apuesta de la lotería primitiva, en la que se señalan 6 números de 49. Para ello utilizamos la función randInt(1, 49, 6) de la calculadora gráfica TI–83. Así, pulsamos:* MATH ◄ [5] 1 , 49 , 6) ENTER

La apuesta estaría formada por los elementos de esta lista, siempre que no hayan repeticiones.

a) En una escuela hay 743 estudiantes. Se debe elegir 20 alumnos al azar. Explica el procedimiento más adecuado para efectuar la selección.

b) De una población de 1800 individuos queremos extraer una muestra cuyo tamaño sea el 1,5 % del tamaño de la población. Halla el tamaño de la muestra y explica el procedimiento de selección.

SOLUCIÓN:

a) *Se numeran los alumnos del 001 al 743 y se leen los números aleatorios en grupos de tres cifras. Se suprimen los números 000, 744, 745, ... , 999 y las repeticiones. Por ejemplo, empezando por el principio y en dirección horizontal obtenemos :*

593 915 803 052 098 827 188 702 482 848 041 909
657 490 464 290 659 956 776 364 772 040 461 527
062 966 214 391 801 896 839 915 114

b) *El tamaño de la muestra es 1,5% de 1800 = 1,5 x 1800 /100 = 1,5 x 18 = 27. Para extraer la muestra, utilizamos la función randInt(1, 1800, 27) de la calculadora gráfica TI–83. Para ello pulsamos:*

MATH ◄ [5] 1 , 1800 , 27) ENTER

La muestra está formada por los individuos de la población cuyos números de orden sean los de la lista obtenida, siempre que no hayan repeticiones.

- **TIPOS DE MUESTREO**

MUESTREO ALEATORIO SIMPLE

*El **muestreo aleatorio simple** es un procedimiento para seleccionar una muestra de una población que consiste en un sorteo en el que:*

a) *Todos los elementos de la población tienen las mismas posibilidades de ser elegidos, y*

b) *Los elementos de la muestra se eligen independientemente unos de otros, es decir, las posibilidades de cada elemento no dependen de cuáles son los otros elementos seleccionados.*

Podemos elegir los elementos de la muestra de uno en uno, o seleccionarlos todos al mismo tiempo.

Si el sorteo de los elementos se hace de uno en uno, es necesario que en cada etapa los elementos de la población que no han sido seleccionados anteriormente tengan las mismas probabilidades de ser elegidos en la siguiente etapa. Esto se puede conseguir de dos formas :

1) ***Muestreo aleatorio simple con reemplazamiento** : en cada etapa se devuelve a la población el elemento elegido de forma que pueda participar también en la siguiente etapa. Cada etapa es idéntica a la anterior y un mismo elemento puede ser elegido muchas veces. Se pueden obtener así muestras con elementos repetidos.*

2) ***Muestreo aleatorio simple sin reemplazamiento** : en cada etapa se separa el elemento seleccionado y no vuelve a participar en las siguientes etapas del sorteo. Cada etapa es diferente a la anterior porque la población a sortear va disminuyendo. En este caso, ya no se pueden producir repeticiones en la muestra.*

Estos dos procedimientos se diferencian si la población de la que extraemos la muestra es pequeña. En cambio, cuando es muy grande, pueden considerarse prácticamente iguales ya que las repeticiones son muy improbables.

En la práctica los dos procedimientos utilizan la tabla de números aleatorios o un generador aleatorio adecuado (ordenador, calculadora) para seleccionar los elementos que componen la muestra. En el caso (1) se admiten números repetidos y en el caso (2) se rechazan las repeticiones.

Si seleccionamos todos los elementos de la muestra al mismo tiempo, debemos buscar un procedimiento que asegure que todas las muestras del mismo tamaño tengan las mismas probabilidades de ser elegidas.

a) El centro Ximo Trinquet tiene un equipo de fútbol sala y un equipo de baloncesto. Los integrantes de cada uno de los equipos son:

   - Fútbol sala : Pepe, Juana, Ana, Javi, Ximo, Juanjo, Vicente, Marta y Daniel.
   - Baloncesto : Jordi, Antonio, Asun, Enrique, Mario, Ramón, Isabel y Maite.

   El programa deportivo de Canal 9 *Avall la bola* invita a tres estudiantes del equipo de fútbol sala y a dos del equipo de baloncesto a participar en uno de sus programas. Utiliza el muestreo aleatorio simple para seleccionar a los cinco estudiantes invitados. Explica detalladamente el procedimiento que sigues para realizar dicha selección.

MUESTREO ALEATORIO SISTEMÁTICO

Dividimos el tamaño de la población entre el tamaño de la muestra: $x=\frac{N}{n}$*. A continuación elegimos un número aleatorio de la tabla, A. Sumando y restando x a este número A, obtenemos los elementos de la muestra:*

$A-3x \quad A-2x \quad A-x \quad A \quad A+x \quad A+2x \quad A+3x$

Por ejemplo, para seleccionar una muestra de 400 individuos de una población de 10000 personas, dividimos el tamaño de la población entre el tamaño de la muestra: 10000 / 400 = 25. Elegimos un número aleatorio de la tabla que tenga cuatro cifras (el 0000 corresponde al 10000), por ejemplo, el 2427. Sumando y restando 25 a este número obtenemos los elementos de la muestra :

2352 2377 2402 2427 2452 2477 2502

MUESTREO ALEATORIO ESTRATIFICADO

Cuando la población está dividida en grupos que son significativos para los datos estadísticos que se están estudiando, es conveniente que la muestra refleje la composición de la población. Cada grupo de la población proporciona aleatoriamente una parte de la muestra (cada parte proporcional al tamaño del grupo de procedencia).

Así, si queremos extraer una muestra de tamaño 400 de una población de 10000 individuos en la que hay 6000 de estudios primarios, 3000 de estudios medios y 1000 de estudios superiores, elegimos al azar a, b y c personas de cada grupo tales que:

$$\frac{a}{6000}=\frac{b}{3000}=\frac{c}{1000}=\frac{400}{10000}$$

de manera que $a = 240 \quad b = 120 \quad c = 40$

MUESTREO ALEATORIO POR CONGLOMERADOS

Se eligen aleatoriamente unos grupos, cuyos elementos constituyen la muestra.

Así, podemos elegir fincas y formar la muestra con los habitantes de esas fincas (sin excluir a ninguno).

MUESTREO ALEATORIO POR ETAPAS

Se eligen aleatoriamente ciertos grupos y en cada uno se toman aleatoriamente ciertos elementos que componen la muestra.

Así, podemos elegir aleatoriamente calles; en ellas seleccionar fincas al azar y en éstas obtener también aleatoriamente individuos de la muestra.

b) En cierto barrio se quiere hacer un estudio para conocer mejor el tipo de actividades de ocio que gustan más a sus habitantes. Para ello, van a ser encuestados 100 individuos elegidos al azar.

1) Explica qué procedimiento de selección sería más adecuado utilizar: muestreo con o sin reemplazamiento. ¿Por qué?.

2) Como los gustos cambian con la edad y se sabe que en el barrio viven 2500 niños, 7000 adultos y 500 ancianos, posteriormente se decide elegir la muestra anterior utilizando muestreo estratificado.

2.1) Define los estratos.

2.2) Determina el tamaño muestral correspondiente a cada estrato.

- **BIBLIOTECA**

Una biblioteca pública está organizada en cinco secciones (en el cuadro adjunto se indica el número de libros existentes en cada sección).

Sección 1	Sección 2	Sección 3	Sección 4	Sección 5
500	860	1200	700	740

Con objeto de estimar el porcentaje de libros de edición española, se quiere seleccionar una muestra de un 5% del número total de libros, a través de muestreo aleatorio estratificado, considerando como estratos las secciones. Determina el número de libros que habría que seleccionar en cada sección si:

a) Seleccionamos el mismo número de libros de cada sección.

b) Utilizamos muestreo proporcional.

- **INSPECCIÓN FISCAL**

En un determinado país, el porcentaje de declaraciones fiscales que son correctas es del 60%, 40% y 80% según se trate de industriales, profesionales liberales o asalariados. Se sabe que del total de las declaraciones, el 10% son de industriales y el 20% de profesionales liberales. Se van a realizar 1500 inspecciones.

a) ¿Cuántos industriales, profesionales liberales y asalariados han de ser inspeccionados si se desea que la inspección sea proporcional a la probabilidad de declaración incorrecta en cada categoría socio–profesional?.

b) Compara esta distribución de las 1500 inspecciones con la que se tendría en el caso de hacerla proporcional al número de declaraciones de cada categoría.

MUESTRAS Y ESTIMACIONES ESTADÍSTICAS CON LA CALCULADORA GRÁFICA

Introducción

La calculadora gráfica es una herramienta potente que facilita enormemente los cálculos en Inferencia Estadística. La TI–83 dispone de los menús DISTR y TESTS para trabajar esta parte de la Estadística. El menú DISTR permite calcular probabilidades asociadas a diversos modelos probabilísticos (los más usuales, como la distribución normal, t de Student, ji–cuadrado, F, Poisson, etc). El menú TESTS contiene los contrastes de hipótesis más usuales (para una y dos muestras), así como intervalos de confianza y análisis de la varianza.

En Bachillerato se trata de usar esta herramienta para introducir los conceptos fundamentales relacionados con la Inferencia Estadística, centrándose especialmente en la estimación de parámetros. La enorme complejidad conceptual que supone el estudio de los tests de hipótesis hace recomendable que se analicen las ideas básicas y se utilice directamente la calculadora gráfica para comprobar hipótesis y resolver problemas. Evidentemente, no es necesario ni tiene mucho sentido estudiar los menús DISTR y TESTS en su totalidad, sino únicamente las opciones ligadas a las distribuciones más importantes (binomial, normal, t de Student, …)

1.– Distribuciones muestrales

A) DISTRIBUCIÓN MUESTRAL DE MEDIAS

El estudio de las propiedades de una población se efectúa a través de diversas muestras extraídas de la población. Los estadísticos (media, mediana, desviación típica, proporción,…) obtenidos en la muestra permiten decidir sobre los correspondientes parámetros en la población. Para ello necesitamos saber cómo se distribuyen dichos estadísticos en el conjunto de las posibles muestras.

Supongamos que en una población la variable aleatoria X tiene media μ y desviación típica σ. Extraemos una muestra de tamaño n y hallamos la media de la variable X en la muestra, $\overline{X}$. Repetimos el proceso con otras muestras de tamaño n, hallando la media, $\overline{X}$, en cada una de ellas. Entonces, se cumple que la media de todas las medias muestrales coincide con la media μ de la población. Además, la desviación típica de todas las medias muestrales es igual a $\frac{\sigma}{\sqrt{n}}$.

Si la distribución de la variable X en la población es normal, entonces la distribución de las medias muestrales también es normal. Es decir:

Si en una población la variable X es normal de media μ y desviación típica σ, entonces las medias muestrales $\overline{X}$ siguen una normal de la misma media μ y desviación típica $\frac{\sigma}{\sqrt{n}}$.

$$\boxed{\textbf{Si } X \approx N(\mu, \sigma) \quad \textbf{entonces} \quad \overline{X} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)}$$

Si la variable X en la población no sigue una distribución normal, pero se toman muestras de tamaño n > 30, entonces también se cumple que las medias muestrales siguen una normal de media μ y desviación típica $\frac{\sigma}{\sqrt{n}}$. Este resultado se conoce como **teorema central del límite**.

Si la desviación típica poblacional, σ, es desconocida, puede sustituirse por la desviación típica muestral, s, cumpliéndose, en ese caso, que: $\overline{X} \approx N\left(\mu,\ \frac{s}{\sqrt{n}}\right)$.

<u>Ejemplo</u>.- La estatura media de la población de cierto barrio es de 176 cm, con una desviación típica de 10 cm.

a) Calcula la media y la desviación típica de la distribución de las medias de las muestras de tamaño 36.

b) Halla la probabilidad de que una muestra de 36 personas tenga una estatura media de 176 cm o más.

a) La distribución de las medias muestrales es normal de media 176 cm y desviación típica $\frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{36}} = \frac{10}{6} \approx 1{,}67$ cm.

b) Si X es la estatura media de las muestras de tamaño 36, entonces se cumple X≈N(176, 1,67). Utilizando la calculadora gráfica: **normalcdf(176, 1E99, 176, 1.67) [ENTER]** da como resultado 0.499999999=0.5. Por tanto, la probabilidad de que la muestra tenga una estutura media de 176 cm o más es: $p(X \geq 176) = 0{,}5$

B) DISTRIBUCIÓN MUESTRAL DE PROPORCIONES

Cuando se trata de determinar la proporción o porcentaje de una población que posee un cierto atributo (vota al partido A / no vota al partido A, invierte en bolsa / no invierte en bolsa, éxito / fracaso), utilizamos el modelo de la distribución binomial. Así, si la probabilidad de éxito es P y la de fracaso Q=1–P, y tomamos muestras aleatorias de tamaño n≥30, entonces las proporciones muestrales siguen una distribución normal de media P y desviación típica $\sqrt{\frac{P \cdot Q}{n}}$. Es decir:

Si P es la proporción poblacional y n ≥ 30, entonces las proporciones muestrales siguen una distribución normal $N\left(P, \sqrt{\frac{P \cdot Q}{n}}\right)$.

En la práctica ocurre que las proporciones P y Q de la población son desconocidas. En estos casos se aproximan por las respectivas proporciones de una muestra, siempre que su tamaño sea n > 100.

Ejemplo.- En unas elecciones a alcalde, el 56% de los votantes optó por el candidato A, mientras que el 44% lo hizo por el candidato B.

a) Halla la distribución de proporciones de las muestras de tamaño 50 extraídas de la población.

b) Calcula la probabilidad de que en una muestra de 50 votantes haya, al menos, 30 favorables al candidato A.

a) La proporción de la población, para el candidato A, es p = 0,56; q = 0,44. La proporción de las muestras de tamaño 50 se distribuye según la curva normal $N\left(p, \sqrt{\frac{p \cdot q}{n}}\right) = N\left(0{,}56, \sqrt{\frac{0{,}56 \cdot 0{,}44}{50}}\right) = N(0{,}56; 0{,}07)$.

b) Treinta votantes a favor de A, entre 50, supone una proporción de $\bar{p} = \frac{30}{50} = 0{,}6$. Como $\bar{p} \approx N(0{,}56; 0{,}07)$, utilizando la calculadora gráfica: **normalcdf(0.6, 1E99, 0.56, 0.07) [ENTER]** da como resultado: 0.2838 ≈ 0.28. Entonces, la probabilidad pedida es: $p(\bar{p} \geq 0{,}6) = 0{,}28$.

ACTIVIDADES

- **RECIÉN NACIDOS**

En una ciudad, el peso de los recién nacidos se ha distribuido según la ley normal de media μ = 3100 gramos y desviación típica σ = 150 gramos.

Halla los parámetros de la distribución que siguen las medias de las muestras de tamaño 100.

- **ELECCIONES**

En las elecciones a decano de una facultad se presentaron dos candidatos: A y B. El resultado de la votación fue del 60% para A y 40% para B. Si antes de la votación se hizo una encuesta a 36 votantes, ¿cuál habría sido la probabilidad de acertar el ganador?. (Es decir, p(votar A) > 0,5).

- **OPOSICIONES**

Al acto de presentación de unas oposiciones asistió el 65% de los candidatos. Si se hubiesen tomado, elegidos al azar, 81 opositores, ¿cuál es la probabilidad de que se presenten menos de 55?.

- **LA PRESA**

El 40% de los ciudadanos de una comarca se opone a la construcción de una presa. Si se pregunta a 60 personas de esa comarca, ¿qué probabilidad hay de que ganen los que se oponen?.

- **PROGRAMAS CULTURALES**

Se sabe que el 60% de los adultos de un área geográfica asiste regularmente a programas culturales. Se obtiene una muestra aleatoria de 150 adultos. Halla la probabilidad de que la proporción muestral esté comprendida entre los valores 0,5 y 0,7.

2.–Estimación de parámetros

A) ESTIMACIÓN PUNTUAL

Generalmente no se suelen conocer exactamente las características de una población. Normalmente utilizamos muestras para describirlas, de manera que las características muestrales serán una estimación de las correspondientes características poblacionales.

Para describir una población compuesta por diversas categorías utilizamos las proporciones o frecuencias relativas de cada categoría. La proporción exacta de una categoría en la población, P, no es conocida y usamos la correspondiente proporción muestral, $\overline{P}$, como estimador.

Para describir una variable continua en la población es usual recurrir a la media y a la desviación típica. Normalmente la media, μ, y la desviación típica, σ, poblacionales son desconocidas y utilizamos la media muestral, $\overline{x}$, y la desviación típica muestral, s, como estimadores.

Ejemplo 1.- Un investigador mide la longitud total del tallo de 13 plantas de soja de una determinada especie a los 16 días de crecimiento, obteniendo los siguientes resultados:

20.2	**22.9**	**23.3**	**20.0**	**19.4**	**22.0**	**22.1**	**22.0**	**21.9**	**21.5**	**19.7**	**21.5**	**20.9**

¿Cuál es la longitud media del tallo de las plantas de soja de esa especie?. ¿Cuál es la desviación típica en esta clase de plantas?.

Evidentemente, no podemos saber con certeza cuál es la longitud media poblacional de esta especie de plantas, ni tampoco cuál es su desviación típica. Sin embargo, podemos dar como estimadores puntuales la media y la desviación típica muestral:

$\overline{x} = 21.3385 \approx 21.34$ cm. es una estimación puntual de μ.
s = 1.2190 ≈ 1.22 cm. es una estimación puntual de σ.

Ejemplo 2.- En una encuesta aleatoria de 265 personas de una población se encontraron 194 personas favorables a una determinada política. ¿Qué proporción de ciudadanos de la población son favorables a dicha política?.

Evidentemente, no podemos saber con certeza cuál es la proporción de individuos favorables en la población, pero podemos dar como estimación puntual la proporción muestral:

$\overline{P} = \frac{194}{265} = 0.732 \approx 73.2\,\%$ es una estimación puntual de P.

B) ESTIMACIÓN POR INTERVALOS DE CONFIANZA

Como es lógico, resulta muy arriesgado trasladar mecánicamente a la población los parámetros obtenidos en la muestra. Lo normal es que hayan desviaciones entre los parámetros muestrales y los poblacionales. Parece más acertado dar como estimación del parámetro un intervalo y no un único valor. La estimación por intervalos de confianza consiste en hacer razonamientos del siguiente tipo:

"No sabemos cuál es el valor buscado del parámetro w, pero la información contenida en la muestra indica que ese número está entre los valores a y b casi con seguridad".

Los extremos del intervalo [a, b] serán funciones de la muestra y se trata de determinarlos con un cierto nivel de seguridad o NIVEL DE CONFIANZA. El nivel de confianza mide el grado de seguridad que tenemos al afirmar que el valor del parámetro se encuentra en el intervalo [a, b]. Se expresa así:

$p(a \le w \le b) = 1 - \alpha$ = nivel de confianza

Donde α se llama nivel de error o nivel de significación.

Por ejemplo, determinar un intervalo de confianza con un nivel de significación del 5% es equivalente a obtener un intervalo con un nivel de confianza del 95%. Esto significa que si extraemos una muestra de la población y obtenemos un intervalo de confianza para el parámetro buscado y volvemos a repetir el proceso de extraer muestras y obtener los correspondientes intervalos de confianza, 95 de cada 100 de estos intervalos contendrán al verdadero valor de parámetro.

Para determinar un intervalo de confianza para un parámetro w necesitamos conocer la distribución muestral del correspondiente parámetro muestral $\overline{w}$. Por ejemplo, si $\overline{w}$ sigue una distribución normal de media w y desviación típica ES_w, una medida de la discrepancia entre el estimador $\overline{w}$ y el parámetro w es ES_w, que se llama ERROR TÍPICO DE MUESTREO.

Si $\overline{w} \approx N(w, ES_w)$ entonces, el intervalo $\left[\overline{w} - Z_{1-\frac{\alpha}{2}} \cdot ES_w,\ \overline{w} + Z_{1-\frac{\alpha}{2}} \cdot ES_w\right]$ es un intervalo de confianza para el parámetro w con un nivel de confianza $1-\alpha$, siendo $Z_{1-\frac{\alpha}{2}}$ el cuantil $1-\frac{\alpha}{2}$ de la distribución $N(w, ES_w)$.

En general, un intervalo de confianza para el parámetro w con un nivel de confianza $1-\alpha$, es un intervalo de la forma $\left[\overline{w} - k \cdot ES_w, \overline{w} + k \cdot ES_w\right]$, siendo $\overline{w}$ un estimador puntual de w y k el cuantil correspondiente de la distribución muestral que siga el estimador.

C) INTERVALO DE CONFIANZA PARA LA MEDIA

Si la población tiene media μ desconocida y desviación típica σ conocida, y extraemos una muestra de tamaño n con media $\overline{x}$ y desviación típica s, para n suficientemente grande se cumple que $\overline{x} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$. Por tanto, el error típico de muestreo es, en este caso, $ES_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$.

Entonces, aplicando lo visto en el apartado anterior, se cumple que:

El intervalo $\left[\overline{x} - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}},\ \overline{x} + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right]$ es un intervalo de confianza para la media μ con un nivel de confianza $1-\alpha$.

Lo habitual es que la desviación típica poblacional σ sea desconocida, en cuyo caso la media muestral $\overline{x}$ no sigue una distribución normal y entonces no se puede utilizar σ ni el cuantil $1-\frac{\alpha}{2}$ de la N(0, 1) para hallar el intervalo de confianza. Si σ es desconocida, la media muestral $\overline{x}$ sigue una distribución T de Student, que para valores de n grandes se puede aproximar por una distribución normal. En este caso se puede utilizar la desviación típica muestral s en lugar de σ, de forma que el intervalo de confianza para la media μ viene dado por:

$$\left[\overline{x} - Z_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}},\ \overline{x} + Z_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}\right]$$ para un nivel de significación α.

Ejemplo.- Se ha medido la longitud de 13 plantas de una especie de soja, obteniendo los siguientes resultados:

20.2	**22.9**	**23.3**	**20.0**	**19.4**	**22.0**	**22.1**	**22.0**	**21.9**	**21.5**	**19.7**	**21.5**	**20.9**

Halla un intervalo de confianza para la longitud media de esta especie de plantas, con un nivel de significación del 5%.

Para α=0.05, el nivel de confianza es 1−α=0.95. El cuantil correspondiente de la N(0, 1) es $Z_{1-\frac{\alpha}{2}} = Z_{0.975} = 1.96$, como puedes comprobar con la función **invNorm(0.975)** de la calculadora gráfica, pulsando [2nd] DISTR [3] 0.975) ENTER. Además, sabemos que la media y la desviación típica muestrales son: $\overline{x} = 21.3385 \approx 21.34$ cm. y s = 1.2190 ≈ 1.22 cm. Por lo tanto, el intervalo de confianza buscado es:

$$\left[21.34 - 1.96 \times \frac{1.22}{\sqrt{13}},\ 21.34 + 1.96 \times \frac{1.22}{\sqrt{13}}\right] = [20.6768,\ 22.0032] \approx [20.68,\ 22]$$

Tenemos una confianza del 95 % de que el intervalo [20.68, 22] contenga al verdadero valor de la media poblacional.

D) INTERVALO DE CONFIANZA PARA LA PROPORCIÓN

Si la proporción P con la que aparece una categoría en la población es desconocida y extraemos una muestra de tamaño n ≥ 30 obteniendo dicha categoría con una proporción muestral $\overline{P}$, entonces se cumple que $\overline{P} \approx N\left(P, \sqrt{\frac{P \cdot Q}{n}}\right)$, siendo Q=1−P. En este caso, el error típico de muestreo es $ES_{\overline{P}} = \sqrt{\frac{P \cdot Q}{n}}$. Por tanto, aplicando lo visto en el apartado anterior, se cumple:

El intervalo $\left[\overline{\mathbf{P}} - \mathbf{Z}_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\mathbf{P} \cdot \mathbf{Q}}{\mathbf{n}}},\ \overline{\mathbf{P}} + \mathbf{Z}_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\mathbf{P} \cdot \mathbf{Q}}{\mathbf{n}}}\right]$ **es un intervalo de confianza para la proporción P con un nivel de confianza 1−α.**

Lo habitual es que la proporción en la población, P, sea desconocida. Por tanto se desconocen P y Q. En este caso, se puede utilizar la proporción muestral $\overline{P}$ en lugar de P y $\overline{Q} = 1 - \overline{P}$ en lugar de Q. Así, **el intervalo de confianza para la proporción poblacional P con un nivel de confianza 1−α es:**

$$\left[\overline{P} - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\overline{P} \cdot \overline{Q}}{n}},\ \overline{P} + Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\overline{P} \cdot \overline{Q}}{n}}\right]$$

Ejemplo.- **En una encuesta aleatoria de 265 personas de una población se encontró que 194 de ellas se mostraron favorables a una determinada política. ¿Cuál es el intervalo de confianza del 95% para la proporción de la población total favorable a dicha política?.**

La proporción muestral es, en este caso, $\overline{P} = \frac{194}{265} = 0.732$. Por lo tanto, $\overline{Q} = 1 - \overline{P} = 1 - 0.732 = 0.268$. Para un nivel de confianza del 95% se cumple: 0.95=1−α → α=0.05 → $Z_{1-\frac{\alpha}{2}} = Z_{0.975} = 1.96$, como puedes comprobar con la función **invNorm(0.975)** de la calculadora gráfica. Entonces, el intervalo de confianza buscado es:

$$\left[0.732 - 1.96 \times \sqrt{\frac{0.732 \times 0.268}{265}},\ 0.732 + 1.96 \times \sqrt{\frac{0.732 \times 0.268}{265}}\right] = [0.6787,\ 0.7853]$$

Es decir, tenemos un 95 % de confianza de que la proporción de personas favorables a dicha política está comprendida entre el 67,9 % y el 78,5 %.

ACTIVIDADES

- **SELECTIVIDAD**

Una muestra aleatoria de 100 alumnos que se presenta a las pruebas de Selectividad, revela que la media de edad es de 18,1 años. Halla un intervalo de confianza de 90% para la edad media de todos los estudiantes que se presentan a las pruebas, sabiendo que la desviación típica de la población es de 0,4.

- **PRECIOS**

Se ha tomado una muestra de los precios de un mismo producto alimenticio en 16 comercios, elegidos al azar en un barrio de una ciudad, y se han encontrado los siguientes precios:

95	108	97	112	99	106	105	100	99	98	104	110	107	111	103	110

Suponiendo que los precios de este producto se distribuyen según una ley normal de varianza 25 y media desconocida:

a) ¿Cuál es la distribución de la media muestral?.

b) Determina el intervalo de confianza, al 95%, para la media poblacional.

- **ORDENADORES**

Se realizó una encuesta a 350 familias preguntando si poseían ordenador en casa, encontrándose que 75 de ellas lo poseían. Estima la proporción real de familias que disponen de ordenador con un nivel de confianza del 95%.

- **LECTORES DE PRENSA**

Tomada al azar una muestra de 500 personas en la Comunidad Valenciana, se encontró que 220 leían algún periódico habitualmente.

Calcula, con un nivel de confianza del 95 por ciento, el intervalo en el que se encontrará la verdadera proporción de lectores de periódicos y explica el proceso seguido para dicho cálculo.

3.– Intervalos de confianza con la calculadora gráfica

Podemos obtener intervalos de confianza con ayuda de la calculadora gráfica TI–83. Para ello usaremos el menú TESTS que aparece al pulsar la tecla STAT.

A) ESTIMACIÓN DE UNA PROPORCIÓN

Para obtener un intervalo de confianza para la proporción usaremos la función 1–PropZInt, a la que puedes acceder a través del menú TESTS. En la pantalla que aparece introduce el número de casos, x, el tamaño de la muestra, n, el coeficiente de confianza, C–Level, sitúa el cursor sobre Calculate y pulsa ENTER. En la siguiente pantalla se muestra el intervalo de confianza, la proporción y el tamaño muestral.

Ejemplo.- **En un sondeo electoral realizado a 273 personas de una población, se manifestaron 82 personas favorables a un determinado partido político. ¿Cuál es el intervalo de confianza del 95 % para la proporción de la población total que votará a dicho partido?.**

Pulsamos STAT ▸ [A] para activar la función 1–PropZInt. En la siguiente pantalla introducimos los valores: x = 82, n=273, C–Level=0.95. Situamos el cursor sobre Calculate y pulsamos ENTER. En la siguiente pantalla se muestra el intervalo de confianza, además de la proporción muestral y el tamaño de la muestra.

B) ESTIMACIÓN DE UNA MEDIA

Para obtener un intervalo de confianza para la media utilizaremos la función Zinterval del menú TESTS. Este menú se obtiene pulsando la tecla STAT. Al activar esta función aparece una pantalla que propone dos métodos de trabajo diferentes: Data y Stats. En el primero (Data) hay que almacenar todos los valores de la muestra en una lista y especificar el nombre de la lista que contiene los datos. En el segundo (Stats) basta dar un resumen de los estadísticos de la muestra, como la media muestral y n. En cada ocasión usaremos la parte del menú que nos interese. A continuación hay que indicar el nivel de confianza (C–Level) y finalmente, desplazar el cursor a la opción Calculate y pulsar ENTER. El resultado es una pantalla donde se indican el intervalo de confianza, la media, la desviación típica y el tamaño de la muestra.

Ejemplo.- Hemos pesado 28 corderos de una misma especie criados en idénticas condiciones ambientales, obteniendo los siguientes resultados (en kg):

4.3	**5.2**	**6.2**	**6.7**	**5.3**	**4.9**	**4.7**	**5.5**	**5.3**	**4.0**	**4.9**	**5.2**	**4.9**	**5.3**
5.4	**5.5**	**3.6**	**5.8**	**5.6**	**5.0**	**5.2**	**5.8**	**6.1**	**4.9**	**4.5**	**4.8**	**5.4**	**4.7**

Calcula un intervalo de confianza para la media μ de los pesos de los corderos de esa especie, con un nivel de confianza del 90%.

Introducimos los datos en la lista L_1 y a continuación pulsamos STAT ▸ [1] L_1 para seleccionar la función 1−Var Stats del menú CALC. De esta forma obtenemos los estadísticos muestrales, concretamente: $\overline{x} = 5.167857143$ y $S_X = 0.6543606274$. Podemos suponer que la desviación típica poblacional coincide con la muestral, osea $\sigma = S_x = 0.65$. Entonces, activando el menú Zinterval, situaremos en cursor en Data y pulsaremos ENTER; introduciremos como lista de datos L_1, con frecuencias iguales a 1 y un C–Level igual a 0.90. Situando el cursor sobre Calculate y pulsando ENTER, obtenemos el intervalo de confianza, junto con la media, desviación típica y tamaño muestral.

Haz este mismo ejercicio usando el menú Tinterval, en lugar de Zinterval y compara los resultados obtenidos. El menú Tinterval se basa en usar la distribución T de Student, en vez de la distribución normal.

ACTIVIDADES

- **GAFAS GRADUADAS**

En una determinada población se toma una muestra al azar de 256 personas. De esta muestra, el 20% de las personas lleva gafas graduadas y el 80% restante no. Calcula el intervalo de confianza aproximado para la proporción poblacional de las personas que llevan gafas graduadas para un nivel de confianza del 95%.

- **LIBROS CIENTÍFICOS**

Se desea hacer un estudio de mercado para conocer el precio medio de los libros científicos. Para ello, se elige una muestra aleatoria formada por 34 libros y se determina que la media muestral es de 34'90 euros con una desviación típica de 4'50 euros. Halla el intervalo de confianza para el precio medio de los libros científicos al nivel del 99%.

4.– Tamaño de una muestra

Uno de los problemas fundamentales de la Inferencia Estadística consiste en determinar el tamaño idóneo de la muestra para que ésta sea representativa de la población. Evidentemente, el tamaño muestral se obtiene teniendo en cuenta el error máximo admisible. Distinguimos para ello dos casos, según que se pretenda estimar una media o una proporción.

A) PARA UNA MEDIA

Cuando estimamos una media poblacional con un nivel de significación α, el tamaño idóneo de la muestra es

$$n = \left(\frac{Z_{1-\frac{\alpha}{2}} \cdot \sigma}{E} \right)^2 .$$

B) PARA UNA PROPORCIÓN

Cuando estimamos una proporción poblacional con un nivel de significación α, el tamaño idóneo de la muestra es

$$n = \left(\frac{Z_{1-\frac{\alpha}{2}}}{E} \right)^2 \cdot \overline{P} \cdot \overline{Q}.$$

Si utilizamos el límite máximo de error, obtenemos otra aproximación al tamaño de la muestra:

$$n = \left(\frac{Z_{1-\frac{\alpha}{2}}}{E} \right)^2$$

ACTIVIDADES

- **BOMBILLAS**

Un fabricante de bombillas sabe que la desviación típica de la duración de las bombillas es de 100 horas. Calcula el tamaño de la muestra que se ha de someter a prueba para tener una confianza del 95% de que el error de la duración media que se calcule sea menor que 10 horas.

- **CONTROL DE CALIDAD**

¿De qué tamaño conviene tomar la muestra de una línea de producción para tener una confianza del 95% de que la proporción estimada no difiere de la verdadera en más de un 5%?. Se sabe, por estudios previos, que la proporción de objetos defectuosos es del orden del 0,05.

- **ELECCIONES**

En una población de 8000000 de votantes, se sospecha que el 35 % de ellos votarán al partido A. ¿Cuál debe ser el tamaño de la muestra, que se desea encuestar, para que la proporción del número de personas en la muestra que vota al partido A, no se aparte de la proporción correspondiente en la población en más de dos centésimas, con una probabilidad del 95 % al menos?.

- **GRUPO SANGUÍNEO**

Queremos estimar la proporción de valencianos que tienen el grupo sanguíneo 0, con una precisión (o margen de error) de 0’02. ¿Cuál es el tamaño de la muestra que se debe utilizar, con una certeza del 95% ?.

COMPROBACIÓN EXPERIMENTAL DEL TEOREMA CENTRAL DEL LÍMITE

Introducción

El Teorema central del límite es el punto clave de la Inferencia Estadística y, probablemente, el teorema más importante de la Estadística Matemática. Sin embargo, su demostración es realmente difícil, incluso para los que llevan años de experiencia en Matemáticas. Afortunadamente, la calculadora gráfica es una herramienta que permite visualizar y comprobar el teorema central del límite, mediante simulaciones. Esto es especialmente importante, ya que, aunque una simulación no demuestra nada, si permite adquirir una intuición del teorema que, al fin y al cabo, es lo que se pretende en Bachillerato.

1. Distribución muestral de medias

Sea la X una variable aleatoria con media μ y desviación típica σ. Extraemos una muestra de tamaño n y hallamos la media de la variable X en la muestra, $\overline{X}$. Repetimos el proceso con otras muestras de tamaño n, hallando la media, $\overline{X}$, en cada una de ellas. Entonces, se cumple que la media de todas las medias muestrales coincide con la media μ de la población. Además, la desviación típica de todas las medias muestrales es igual a $\frac{\sigma}{\sqrt{n}}$.

Si la distribución de la variable X en la población es normal, entonces la distribución de las medias muestrales también es normal.

Si la variable X en la población no sigue una distribución normal, pero se toman muestras de tamaño $n > 30$, entonces también se cumple que las medias muestrales siguen una normal de media μ y desviación típica $\frac{\sigma}{\sqrt{n}}$. Este resultado se conoce como **teorema central del límite**.

Vamos a utilizar la calculadora gráfica para comprobar este teorema mediante simulaciones, tanto en el caso de que la variable de partida X sea normal, como si no lo es.

- **<u>MUESTREO DE UNA VARIABLE NORMAL</u>**

Si en una población la variable X es normal de media μ y desviación típica σ, entonces las medias muestrales $\overline{X}$ siguen una normal de la misma media μ y desviación típica $\frac{\sigma}{\sqrt{n}}$.

$$\boxed{\textbf{Si } X \approx N(\mu, \sigma) \quad \textbf{entonces} \quad \overline{X} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)}$$

Comprobaremos este resultado con la calculadora gráfica.

Vamos a simular 6 veces la extracción de una muestra de tamaño n=100 de una variable aleatoria normal X, de media 5 y desviación típica 1. Para ello generaremos seis listas (que almacenaremos en L_1, L_2, L_3, L_4, L_5 y L_6), las cuales contendrán las muestras generadas.

Posteriormente construiremos una nueva lista, M, con las medias de cada una de ellas. Finalmente, calcularemos la media y desviación típica de M y dibujaremos el histograma correspondiente, para comprobar el teorema central del límite.

- Pulsa [STAT] [ENTER] para iniciar el editor de listas estadísticas. Sitúa el cursor sobre el nombre de la lista L_1. Pulsa [CLEAR] para borrar dicha lista.

- Pulsa [MATH] [◄] 6 para seleccionar el comando 6: randNorm(. Pulsa 5 [,] 1 [,] 100 [ )] para escribir en la línea de edición la instrucción L_1 = randNorm(5, 1, 100).

- Pulsa [ENTER] y verás cómo se genera una lista con 100 valores de una variable normal de media 5 y desviación típica 1.

- Pulsa [2nd] [QUIT] para regresar a la pantalla principal. Pulsa [STAT] [►] [ENTER] [2nd] [L_1] para introducir el comando 1−Var Stats L_1. Pulsa [ENTER] y comprueba que la media $\overline{X}$ es, aproximadamente, 5 y que la desviación típica S_X es, aproximadamente, 1. Anota los valores de $\overline{X}$ y S_X obtenidos.

- Pulsa [2nd] [STATPLOT] [ENTER] para definir el Plot1 con las siguientes características:

Activado	On
Type	Histograma
Xlist	L_1
Freq	1

- Pulsa [ZOOM] 9 para seleccionar el comando 9: ZoomStat. En la pantalla gráfica se muestra el histograma, que, como puedes comprobar tiene aproximadamente la forma de la curva normal. Por tanto, la muestra L_1 es aproximadamente normal, de media 5 y desviación típica 1.

- Pulsa [STAT] [ENTER] para regresar al editor de listas estadísticas. Sitúa el cursor sobre el nombre de la lista L_2. Pulsa [CLEAR] para borrar dicha lista.

- En la línea de edición introduce el comando L_2 = randNorm(5, 1, 100) utilizando el mismo procedimiento que con la lista L_1. Pulsa [ENTER] y verás cómo se genera la lista L_2.

- Utiliza el mismo procedimiento para definir y generar las listas L3, L4, L5 y L6 como: $L_3 = L_4 = L_5 = L_6$ = randNorm(5, 1, 100).

- Pulsa [STAT] [►] [ENTER] [2nd] [L_2] para seleccionar el comando 1−Var Stats L_2. Pulsa [ENTER] y anota la media $\overline{X}$ y la desviación típica S_X obtenidas.

- Utilizando el comando 1−Var Stats del menú STAT CALC, obtén los parámetros estadísticos de las listas L_3, L_4, L_5 y L_6. En cada caso anota la media $\overline{X}$ y la desviación típica S_X y completa la siguiente tabla:

Muestra	L_1	L_2	L_3	L_4	L_5	L_6
Media						
Desviación típica						

Comprueba que todas las desviaciones típicas están próximas a 1.

- Vamos ahora a construir una nueva lista con las medias de todas las muestras obtenidas. Pulsa [STAT] [ENTER] para regresar al editor de listas estadísticas. Sitúa el cursor en la cabecera de una lista en blanco y pulsa [M] para nombrar la lista como M.
- Pulsa [ENTER] [ENTER] para definir el primer elemento de la lista. Introduce la media de la lista L_1 como primer valor. Introduce la media de L_2 como segundo elemento, la media de L_3 como tercer elemento, … y la media de L_6 como sexto elemento.
- Pulsa [STAT] [▸] [ENTER] [2nd] [LIST] y con las teclas de cursor [▾] [▴] y [ENTER] selecciona la lista M para escribir en la pantalla principal el comando 1−Var Stats ʟM.
- Pulsa [ENTER] y comprueba que la media sigue siendo, aproximadamente, 5 y la desviación típica es, aproximadamente, igual a $1/\sqrt{100} = 1/10$, tal como indica el teorema central del límite.
- Si ampliamos la lista M con los valores obtenidos en toda la clase (seis por estudiante), podemos construir el histograma correspondiente. Para ello hay que pulsar [2nd] [STATPLOT] [ENTER] para editar el Plot1 con las siguientes características:

Activado	On
Type	Histograma
Xlist	ʟM
Freq	1

El resultado es un histograma que tiene aproximadamente la forma de la curva normal de media 5 y desviación típica 1/10, tal como indica el teorema central del límite.

2. Muestreo de una variable binomial

Si la variable aleatoria X no sigue una distribución normal, pero se toman muestras de tamaño n > 30, entonces el teorema central del límite afirma que las medias muestrales siguen una normal de media μ y desviación típica $\frac{\sigma}{\sqrt{n}}$. Vamos a comprobar este resultado con la calculadora gráfica simulando una variable binomial B(10, 0'5), que tiene media $\overline{X} = n \cdot p = 10 \cdot 0'5 = 5$ y desviación típica $\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{10 \cdot 0'5 \cdot 0'5} = \sqrt{2'5} = 1'58113883 \approx 1'58$. Según el teorema central del límite, si extraemos muestras de tamaño n=100, las medias muestrales deben seguir una distribución normal de media 5 y desviación típica $\frac{\sigma}{\sqrt{n}} = \frac{1'58}{\sqrt{100}} = \frac{1'58}{10} = 0'158$.

- Pulsa [STAT] [ENTER] para iniciar el editor de listas estadísticas. Sitúa el cursor sobre el nombre de la lista L_1. Pulsa [CLEAR] para borrar dicha lista.
- Pulsa [MATH] [◂] 7 para seleccionar el comando 7: randBin(. Pulsa 10 [,] .5 [,] 100 [ )] para escribir en la línea de edición la instrucción L_1 = randBin(10, .5, 100).
- Pulsa [ENTER] y verás cómo se genera una lista con 100 valores de una variable binomial de parámetros 10 y 0'5.
- Pulsa [2nd] [QUIT] para regresar a la pantalla principal. Pulsa [STAT] [▸] [ENTER] [2nd] [L_1] para introducir el comando 1−Var Stats L_1. Pulsa [ENTER] y comprueba que la media $\overline{X}$ es, aproximadamente, 5 y que la desviación típica S_X es, aproximadamente, 0.158. Anota los valores de $\overline{X}$ y S_X obtenidos.

- Pulsa [STAT] [ENTER] para regresar al editor de listas estadísticas. Sitúa el cursor sobre el nombre de la lista L_2. Pulsa [CLEAR] para borrar dicha lista.

- En la línea de edición introduce el comando L_2 = randBin(10, .5, 100) utilizando el mismo procedimiento que con la lista L_1. Pulsa [ENTER] y verás cómo se genera la lista L_2.

- Utiliza el mismo procedimiento para definir y generar las listas L3, L4, L5 y L6 como: $L_3 = L_4 = L_5 = L_6$ = randBin(10, .5, 100).

- Utilizando el comando 1−Var Stats del menú STAT CALC, obtén los parámetros estadísticos de las listas L_2, L_3, L_4, L_5 y L_6. En cada caso anota la media $\overline{X}$ y la desviación típica S_X y completa la siguiente tabla:

Muestra	L_1	L_2	L_3	L_4	L_5	L_6
Media						
Desviación típica						

Comprueba que todas las desviaciones típicas están próximas a 1'58.

- Vamos ahora a construir una nueva lista con las medias de todas las muestras obtenidas. Pulsa [STAT] [ENTER] para regresar al editor de listas estadísticas. Sitúa el cursor en la cabecera de una lista en blanco y pulsa [N] para nombrar la lista como N.

- Pulsa [ENTER] [ENTER] para definir el primer elemento de la lista. Introduce la media de la lista L_1 como primer valor. Introduce la media de L_2 como segundo elemento, la media de L_3 como tercer elemento, … y la media de L_6 como sexto elemento.

- Pulsa [STAT] [▸] [ENTER] [2nd] [LIST] y con las teclas de cursor [▾] [▴] y [ENTER] selecciona la lista M para escribir en la pantalla principal el comando 1−Var Stats LM.

- Pulsa [ENTER] y comprueba que la media sigue siendo, aproximadamente, 5 y la desviación típica es, aproximadamente, igual a 0.158, tal como indica el teorema central del límite.

- Si ampliamos la lista M con los valores obtenidos en toda la clase (seis por estudiante), podemos construir el histograma correspondiente. Para ello hay que pulsar [2nd] [STATPLOT] [ENTER] para editar el Plot1 con las siguientes características:

Activado	On
Type	Histograma
Xlist	LN
Freq	1

El resultado es un histograma que tiene aproximadamente la forma de la curva normal de media 5 y desviación típica 0'158, tal como indica el teorema central del límite.

INFERENCIA ESTADÍSTICA EN ESO Y BACHILLERATO

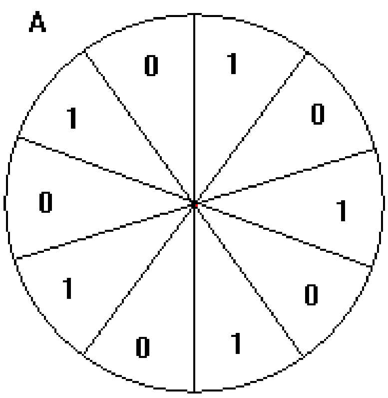
Introducción

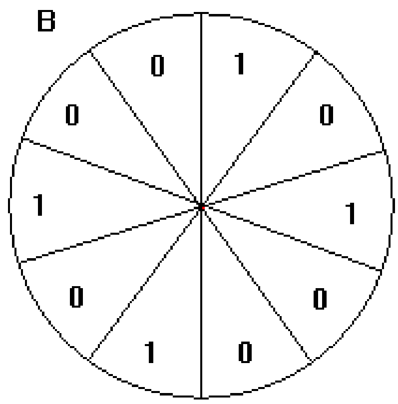
En el nuevo currículum de bachillerato para la asignatura de Matemáticas aplicadas a las Ciencias Sociales de 2º curso, la inferencia estadística ha quedado reducida al estudio de muestras y a la estimación estadística. No se incluyen, por tanto, los tests de hipótesis. Sin embargo, en 3º y 4º de ESO podemos tratar (de forma intuitiva y poco formal) algunas cuestiones relativas a contraste de hipótesis. Por ejemplo, a partir de una colección de 1 y 0 obtenida mediante una serie de lanzamientos de un dado cúbico con sus caras marcadas con 1 y 0, podemos preguntarnos cuántos unos y cuántos ceros tiene el dado y cómo están distribuidos entre sus caras. Un sencillo estudio estadístico basado en el análisis de los parámetros estadísticos usuales (media, moda, mediana, desviación típica, etc) permite formular una conjetura (que evidentemente no se puede contrastar en este nivel). Algunos modelos de calculadora gráfica disponen de un menú específico para realizar contrastes de hipótesis, lo que puede aprovecharse en Bachillerato, si bien se presentan grandes complicaciones conceptuales.

En esta sesión se analizarán algunos materiales que permiten introducir los tests de hipótesis a nivel intuitivo en ESO y Bachillerato.

1.– Inferencia estadística en la E.S.O.

- **<u>¿CUÁL ES LA RULETA?</u>**





Hemos girado cada una de las ruletas 200 veces y hemos anotado los resultados en estas series:

Serie 1	1	0	0	1	0	0	1	1	0	1	0	0	0	1	0	0	1	0	1	0
	1	1	0	0	0	1	1	0	0	1	0	0	1	1	0	0	1	1	0	1
	1	0	1	0	0	0	0	1	0	1	0	0	1	1	0	1	0	0	1	0
	0	0	0	1	0	1	0	1	0	1	1	0	0	0	1	1	1	0	1	0
	1	0	0	1	0	1	0	1	0	1	0	1	0	0	1	1	1	0	1	0
	0	0	0	1	1	0	1	0	1	0	0	1	0	1	0	0	0	1	1	0
	0	1	1	0	1	0	1	1	0	0	0	1	1	1	0	1	0	1	0	1
	0	0	1	1	1	0	1	0	1	0	0	1	0	0	1	0	1	1	0	1
	0	0	0	1	0	1	1	1	0	1	0	1	0	0	1	0	1	1	0	1
	1	0	1	0	0	0	1	0	1	0	1	1	0	1	1	0	0	1	0	1

Serie 2	0	1	0	0	1	0	1	0	0	1	0	1	0	0	0	1	0	0	1	0
	1	1	0	1	0	1	0	1	0	1	1	0	0	0	1	0	1	0	0	0
	0	1	0	1	0	1	0	1	0	1	0	1	0	0	1	1	1	0	0	0
	1	0	0	1	1	0	1	0	1	0	0	1	0	1	0	0	0	1	0	0
	0	1	0	0	1	0	1	0	0	0	0	1	0	1	0	0	0	0	0	1
	1	0	0	1	0	0	1	0	0	1	0	0	0	1	0	0	1	0	1	0
	1	0	1	0	1	0	0	1	0	1	0	0	1	1	0	1	0	0	1	0
	1	0	0	0	1	0	0	0	0	1	0	0	1	1	0	1	0	1	0	0
	1	0	1	0	0	1	0	0	1	0	0	0	1	0	1	0	1	0	1	0
	0	0	1	0	0	0	0	1	0	0	1	0	0	1	1	0	0	1	0	0

Cada serie se ha obtenido girando una de las ruletas. ¿Cuál?. Explica.

Muestreo significa obtención de información a partir de muestras.

Población es el conjunto de datos o valores que se desea estudiar.

Una muestra es una parte del conjunto de datos estadísticos que se desea estudiar. Generalmente, el conjunto de datos es tan amplio que no se puede extraer la información directamente de todos ellos, sino que hay que seleccionar una muestra y limitar el estudio estadístico a los valores de la muestra.

Es posible obtener información bastante fiable de una población estudiando muestras obtenidas al azar. Esta información estará siempre afectada por un cierto grado de incertidumbre, pero el hecho de que las muestras sean extraídas al azar garantiza que las predicciones acerca de la población tengan alguna fiabilidad.

Debemos hacer la hipótesis de que las muestras aleatorias son representativas de la población de que proceden. Los elementos en una muestra obtenida al azar están en parecida proporción que en la población de la que se han obtenido.

Cuanto mayor es el tamaño de la muestra, mayor es la confianza que podemos tener en nuestra predicción.

- **¿CÓMO ES EL DADO?**

Un dado cúbico tiene todas sus caras marcadas con ceros y unos, pero no sabemos en cuántas caras hay 0 ni en cuántas hay 1.

Hemos lanzado 300 veces el dado y éstos son los resultados:

1 1 1 1 1 1 0 1 1 1 1 1 1 0 1 1 0 1 0 1 1 1 1 1 1 1 1 0 1 1
1 1 1 1 1 1 0 1 0 1 0 1 1 1 0 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1
1 1 1 1 1 1 1 1 1 1 1 0 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1
1 1 1 0 1 0 1 0 0 1 1 1 0 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 1
1 1 1 1 0 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1
1 1 1 1 1 1 0 1 1 1 1 1 1 0 0 1 1 1 1 0 0 1 0 1 0 1 1 1 0 1
0 1 1 1 1 1 0 1 1 0 1 1 1 1 1 1 1 0 0 1 1 0 1 0 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 0 0 1 1 1 0 1 1 0 1 0 1 1 1
1 1 1 1 1 1 1 1 0 0 1 0 1 1 1 1 1 1 1 0 1 1 1 1 0 1 0 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 0 0 1 1 1 1 1 1 1 1 1 1 1

Al tirar otra vez 300 veces el dado hemos obtenido la siguiente serie de ceros y unos:

1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 0 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1
1 1 1 1 1 1 1 0 1 1 1 1 0 1 1 1 1 1 1 0 1 1 1 1 0 0 1 0 1 1
1 1 1 0 1 1 1 1 1 1 0 1 0 1 0 1 0 1 1 1 1 0 1 1 1 0 0 0 0 1
1 1 1 1 1 0 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 0 1 1 1
1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 0 0 1 1 1 1 1 1 0 1 1 1
1 1 1 1 1 1 1 1 0 1 1 1 0 0 1 1 1 0 1 1 1 1 1 1 0 0 1 1 1 1
1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 0 1
1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 0 1 1 1
1 1 1 1 0 1 0 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 0 1 1 1 1 1 0

¿Cuántos ceros y cuántos unos crees que hay en el dado?.

- **<u>DADO OCTAÉDRICO</u>**

Tenemos un dado octaédrico (poliedro de 8 caras, triángulos equiláteros). Sus caras están numeradas con ceros y unos, pero no sabemos cuántos ceros ni cuántos unos hay. Al lanzarlo 300 veces, hemos obtenido los resultados:

1 1 1 1 1 0 1 0 1 1 1 0 1 1 1 1 0 1 1 1 0 1 1 0 1 1 1 1 0 1
1 0 0 1 0 0 0 0 1 0 0 1 1 1 1 0 0 1 0 0 1 0 0 0 1 1 0 0 1 1
1 0 0 0 0 0 0 0 0 1 1 0 1 1 0 1 1 0 1 1 0 1 0 0 1 1 0 0 1 1
1 0 1 1 1 1 1 1 1 0 0 1 0 1 0 1 1 1 0 1 1 1 1 0 1 0 1 1 1 1
1 1 1 1 1 1 0 1 1 0 1 0 0 1 1 1 1 1 1 1 0 1 1 0 0 0 1 1 0 1
0 1 1 1 0 0 0 1 1 1 0 1 1 1 1 1 1 0 1 1 1 0 0 0 1 1 0 0 0 0
1 1 1 1 1 1 0 1 1 0 1 1 0 1 1 1 1 1 1 1 0 0 0 1 1 0 1 0 1 1
0 0 1 0 1 0 1 1 1 1 0 1 0 0 1 1 1 1 1 1 1 0 0 0 1 1 1 0 1 1
1 1 1 1 0 1 1 1 1 0 1 1 1 0 1 1 1 0 1 1 0 0 1 0 0 1 0 1 0 0
1 0 1 1 0 1 0 1 1 1 1 1 0 0 1 0 0 0 0 1 0 0 1 1 1 0 0 0 1 0

¿Cuántos ceros y cuántos unos crees que hay en el dado?.

- **<u>DADO CÚBICO</u>**

Cada una de las caras de un dado cúbico se ha marcado con un 1 o con un 0, pero no sabemos cuántas de ellas se han marcado con 1 y cuántas con 0.

Al lanzar este dado 50 veces se obtuvo:

1 1 1 0 0 1 1 1 1 0 0 1 0 1 1 1 1 1 1 0 1 1 1 1 0 0 0 1 1 1 1 0 1 1 1 1 1 1 0 1 1 1 1 1 1 0 1 1 1 1

¿Cuántos ceros y cuántos unos podría tener el dado?.

INFERENCIA ESTADÍSTICA EN BACHILLERATO

1.– Contraste de hipótesis en Bachillerato

- **EL FERIANTE**

Un feriante dispone de dos cajas exteriormente iguales, A y B, pero de contenidos diferentes: la primera contiene 3 bolas blancas y 3 negras; la segunda 4 blancas y 2 negras.

Por cada juego pagas 1 euro. Si eliges una caja y aciertas su contenido (proporción 3:3 o 2:1) ganas 80 céntimos de euro. El feriante permite elegir una de las cajas y extraer, con devolución, una bola 10 veces.

a) ¿Es muy ventajosa esta posibilidad respecto a la elección de las cajas por sorteo?.

b) Supongamos que hemos efectuado 10 extracciones de una de las cajas, con el resultado (b es "blanca" y n es "negra"): bbbnnbbbnb. ¿Podemos decir de qué caja se trata?.

Una vez extraída una muestra de una población, estamos interesados en averiguar si es o no cierta una hipótesis, H_1 relativa a toda la población.

Para ello la comparamos con otra H_0, que indica que las variaciones observadas en la muestras son debidas al azar.

La hipótesis ***H_0*** *se llama* ***hipótesis nula*** *y viene a decir que el resultado obtenido en la muestra se ha obtenido por casualidad (al azar).*

La hipótesis ***H_1*** *se llama* ***hipótesis alternativa*** *y afirma un contenido que explicaría los resultados obtenidos en la muestra.*

La comparación entre las dos hipótesis la haremos utilizando la información contenida en la muestra. Para ello calcularemos la probabilidad P de que, suponiendo cierta H_0 , se obtengan los resultados observados en la muestra.

Si esta probabilidad P es grande, no rechazamos H_0 , es decir, no rechazamos que los datos observados se hayan obtenido por casualidad.

Si P es pequeña rechazamos H_0 , es decir rechazamos que los datos se deban al azar y admitimos la hipótesis alternativa H_1 .

¿Cómo saber si P es grande o pequeña?. Comparándola con un valor α llamado nivel de significación. Los niveles de significación más utilizados son α=0'05=5% y α=0'01=1%.

Una vez elegido el nivel de significación α:

Si $P<\alpha$, rechazamos H_0 frente a H_1 con un nivel de significación α.

Si $P>\alpha$, no rechazamos H_0 frente a H_1 con un nivel de significación α

En las siguientes actividades la probabilidad P se puede calcular utilizando un modelo binomial de probabilidad.

- **¿ES INSESGADA UNA MONEDA?**

Llamamos "éxito" a la aparición de una cara al lanzar cierta moneda. La sometemos a un test con el fin de verificar la hipótesis H_0: "la moneda es insesgada, es decir que la probabilidad de éxito en una prueba es p=1/2".

a) La lanzamos 10 veces y obtenemos 9 éxitos. ¿Debemos rechazar H_0?.

b) Si al lanzar la moneda 20 veces, obtienes 18 cara, ¿rechazarás H_0 con un nivel de significación del 5%?.

- **ELECCIONES**

En un país se van a celebrar unas elecciones. ¿Qué porcentaje de votos obtendrá un determinado partido?. En un pueblo pequeño se han encuestado 10 electores, por sorteo, encontrando que 4 están a favor del partido en cuestión. ¿Qué podemos decir sobre la proporción, p, de votantes favorables que obtendrá en el pueblo?.

- **RATAS**

Analizando el color de la piel de 24 ratas agutí se ha observado la proporción 2:1 entre las de color amarillo y las grisáceas. ¿Crees que hay motivos para sospechar que hay un gen letal, es decir, que la proporción es realmente 2:1 y no 3:1 como ocurre normalmente?. ¿Ha sido casual la relación encontrada?.

- **LANZAMIENTO DE UN DADO**

Después de lanzar un dado 18 veces se han contabilizado seis "5". ¿Qué puedes decir acerca de la probabilidad, p, de obtener "5" en un lanzamiento, con un nivel de significación del 1%?.

2.– Contraste de hipótesis usando la curva normal

En las siguientes actividades la probabilidad P se puede calcular utilizando un modelo de probabilidad basado en la curva normal o la aproximación de la distribución binomial por la curva normal.

- **MEDICAMENTO**

Cierta enfermedad grave puede ser curada mediante un medicamento, A, con una probabilidad del 68%. Mediante un nuevo medicamento, B, se intenta superar dicho porcentaje de curaciones. Para comprobarlo se tratan 200 enfermos, de los que curan 150.

¿Podemos estar seguros de que el nuevo medicamento es mejor que el antiguo?. ¿No habrá sido casual el nuevo porcentaje, 75%, de curaciones?. En otras palabras, supongamos que el nuevo medicamento no es mejor que el antiguo. ¿Es muy probable o poco probable que, suponiendo de igual potencia curativa A y B, se obtengan 150 o más curaciones al tratar 200 enfermos?.

- **UNA MONEDA**

Deseamos verificar si una moneda es "buena". Elegimos un nivel de significación del 5%. Lanzamos la moneda 100 veces y obtenemos 60 veces cara. ¿Qué conclusión podemos sacar?. Repite los cálculos para el caso en que en 100 lanzamientos se obtengan 65 "caras". ¿Es insesgada la moneda?.

- **DEPORTE Y SALUD**

Repetidas estadísticas muestran que el porcentaje de accidentes entre los que practican el fútbol es del 22%. Durante cierto tiempo se han observado los accidentes entre 400 personas que practican la natación, resultando 36 accidentadas. ¿Se puede afirmar que la práctica del fútbol influye en el número de accidentes?.

- **CONTROL DE CALIDAD**

El 98'5% de los tornillos fabricados por una máquina cumplen las especificaciones de fabricación. A raíz de un accidente hay que volver a ajustar la máquina y, extraída una muestra aleatoria de 1500 tornillos de los fabricados después del ajuste se encuentra que sólo el 98% cumplen ahora las especificaciones. ¿Hay razones para sospechar que la máquina funciona peor?.

- **SONDEO ELECTORAL**

Supongamos que de una población de 8000000 de votantes se elige una muestra aleatoria de 2000 y que en esta muestra hay 700 personas, es decir el 35% de las interrogadas, que declaran su intención de votar por el partido C. ¿Qué podemos decir acerca del número de votos que recibirá el partido?.

- **TABAQUISMO**

Investigando los hábitos de consumo, un Instituto de Opinión Pública ha realizado en cierta región un encuesta en la que ha interrogado a 1500 personas; 870 de los interrogados han declarado ser fumadores. Si la encuesta cubría una población de 2030000 personas, ¿qué puedes decir acerca del número de fumadores en la población con un nivel de significación del 5%?.

- **GRANDES ALMACENES**

En una muestra aleatoria de tamaño 100 se ha encontrado que el 62% de los compradores en grandes almacenes compran en más de un departamento cada vez que lo hacen. Estima el porcentaje de la población con un nivel de significación del 5%. Repite el problema si el tamaño es 500, 1000 o 2000.

INFERENCIA ESTADÍSTICA CON LA CALCULADORA GRÁFICA

1.– Test de hipótesis

Se trata de poner en duda una hipótesis, H_0, llamada "hipótesis nula", frente a otra hipótesis, H_1, llamada "hipótesis alternativa". Las dos hipótesis hacen referencia a los valores de un parámetro en la población, y debemos decidir entre una y otra utilizando para ello la información contenida en la muestra.

La **hipótesis nula (H_0)** niega el contenido de la hipótesis alternativa, afirmando que el resultado obtenido en la muestra se debe al azar.

La **hipótesis alternativa (H_1)** afirma un contenido que se quiere probar contrastándolo con la hipótesis nula H_0.

Dependiendo del contenido de la hipótesis nula, tenemos un contraste de una proporción o de una media.

Ejemplos de contrastes de proporciones

1) Se lanza una moneda 10 veces, obteniendo 9 caras. ¿Se puede mantener el valor $p = \frac{1}{2}$ como probabilidad de cara?.
2) Una persona dice que puede distinguir entre dos tipos de cola. Después de hacerle 10 pruebas, acierta en 9 ocasiones. ¿La probabilidad de acertar en cada prueba es $p = \frac{1}{2}$?.

Ejemplos de contrastes de medias

1) La estatura media de los valencianos es de 170 cm. con desviación típica σ = 9 cm. Un muestreo realizado a 36 adultos da una media de 172 cm. ¿Es posible que la estatura media de los valencianos haya aumentado?.

2) El nivel de colesterol para una muestra de 144 personas mayores de 60 años tiene una media $\overline{x} = 235$ y desviación típica s = 45. ¿Se puede admitir que la media de colesterol de la población de mayores de 60 años es de 225 ?.

NIVEL DE SIGNIFICACIÓN Y NIVEL DE CONFIANZA

Se pretende rechazar o no rechazar la hipótesis nula H_0 frente a la hipótesis alternativa H_1.

Se acepta H_1 (se rechaza H_0) si, suponiendo cierta H_0, la probabilidad de que se obtenga el resultado observado en la muestra es pequeña.

Por el contrario, si esta probabilidad es grande, no se rechaza H_0. (Lo que no quiere decir que aceptemos H_0; simplemente indica que la hipótesis alternativa es menos fuerte que la hipótesis nula y, por tanto, no podemos rechazar H_0 si la comparamos con H_1).

Para decidir si esta probabilidad es grande o pequeña, se compara con un valor fijado de antemano, α, que se llama nivel de significación. Llamamos nivel de confianza a $1-\alpha$. Así:

Sea A el resultado observado en la muestra.

- **Se rechaza H_0 si $p(A / H_0) < \alpha$**
- **No se rechaza H_0 si $p(A / H_0) > \alpha$**

Al valor de la probabilidad $p(A / H_0)$ se le suele llamar p–valor asociado a los datos de la muestra. Por tanto:

- **Se rechaza H_0 si el p–valor es menor que el nivel de significación.**
- **No se rechaza H_0 si el p–valor es mayor que el nivel de significación.**

Se suelen utilizar los niveles de significación $\alpha = 0.05 = 5$ % y $\alpha = 0.01 = 1$ %, cuyos niveles de confianza asociados son $1-\alpha = 0.95 = 95$ % y $1-\alpha = 0.99 = 99$ %

REGIÓN CRÍTICA O DE RECHAZO

Es el conjunto de valores de la variable aleatoria X asociada a la muestra para los que se rechaza la hipótesis nula H_0. Contiene los valores de X obtenidos en la muestra y los posteriores.

Decimos que un valor de la variable aleatoria X asociada a la muestra es **significativo al nivel α** si pertenece a la región crítica correspondiente a dicho nivel de significación.

ERRORES EN EL CONTRASTE DE HIPÓTESIS

Podemos cometer dos tipos de errores al contrastar hipótesis:

Error de primera especie (TIPO I): La hipótesis nula H_0 es cierta y la rechazamos, dando por buena la hipótesis alternativa H_1.

Error de segunda especie (TIPO II): La hipótesis nula H_0 es falsa y no la rechazamos.

H_0	CIERTA	FALSA
NO SE RECHAZA	Acertamos	Error de segunda especie
SE RECHAZA	Error de primera especie	Acertamos

Es más grave un error de primera especie, ya que al aceptar la hipótesis alternativa H_1 (siendo H_0 cierta) estamos aceptando como bueno algo que no es mejor que H_0.

El nivel de significación α es, precisamente, la probabilidad de cometer un error de primera especie (ya que α es la probabilidad de rechazar la hipótesis nula H_0, siendo ésta cierta).

TEST BILATERAL Y TEST UNILATERAL

Un contraste sobre un parámetro w es **bilateral** cuando la hipótesis alternativa H_1 es del tipo $w \neq w_0$. Es decir, se trata de contrastar $w = w_0$ frente a $w \neq w_0$. La hipótesis alternativa H_1 se puede descomponer en dos desigualdades $w > w_0$ ó $w < w_0$. Por eso se llama test bilateral.

Si la hipótesis alternativa H_1 solamente contiene una desigualdad, se trata de un test **unilateral**. Así, contrastar H_0: $w = w_0$ frente a H_1: $w > w_0$ es un test unilateral. Otro test unilateral consiste en contrastar H_0: $w = w_0$ frente a H_1: $w < w_0$.

Ejemplo 1.- Se lanza una moneda 10 veces, obteniendo 9 caras. ¿La probabilidad de obtener cara es $p = \frac{1}{2}$?.

Aquí se contrasta H_0: $p = \frac{1}{2}$ frente a H_1: $\left(p > \frac{1}{2} \text{ ó } p < \frac{1}{2}\right)$. Se trata, pues, de un test bilateral.

Ejemplo 2.- Para curar cierta enfermedad se utiliza un medicamento A, con probabilidad $p = \frac{1}{2}$ de sanar. Se ha descubierto un nuevo medicamento B que se ha probado en 13 pacientes obteniendo 10 curaciones. ¿El nuevo medicamento B es mejor que el antiguo A ?.

Ahora se contrasta H_0: $p = \frac{1}{2}$ frente a $p > \frac{1}{2}$. Se trata, pues, de un test unilateral.

CONTRASTE DE UNA PROPORCIÓN

Test bilateral

Estudiamos la proporción P de individuos de una población que tienen un determinado atributo. Extraemos una muestra de tamaño n de la población y obtenemos una proporción de éxitos en la muestra igual a $\overline{P}$. A la vista de los resultados, ¿podemos afirmar que la proporción de éxitos en la población es P?. Establecemos las siguientes hipótesis:

H_0: La proporción de éxitos en la población es P
H_1: La proporción de éxitos en la población es $\neq$ P

¿Debemos rechazar H_0 con un nivel de significación α ?.

A partir de los datos muestrales construimos el intervalo de confianza para P con un nivel de significación α. Entonces pueden presentarse dos casos:

Si P pertenece al intervalo de confianza significa que la diferencia entre P y $\overline{P}$ es pequeña y por tanto, no hay grandes discrepancias entre la hipótesis nula y los datos observados. Es decir, la probabilidad de que se obtengan los datos observados en la muestra supuesta cierta la hipótesis nula debe ser grande. Lo que indica que el p–valor es grande (superior a α). Por tanto, no podemos rechazar H_0.

Si P no pertenece al intervalo de confianza significa que la diferencia entre P y $\overline{P}$ es grande y por tanto, hay grandes discrepancias entre la hipótesis nula y los datos observados. Es decir, la probabilidad de que se obtengan los datos observados en la muestra supuesta cierta la hipótesis nula debe ser pequeña. Lo que indica que el p–valor es pequeño (inferior a α). Por tanto, debemos rechazar H_0. En resumen:

$$\text{No rechazamos } H_0 \text{ si } P \in \left[\overline{P} - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{P \cdot Q}{n}},\ \overline{P} + Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{P \cdot Q}{n}}\right]$$

$$\text{Rechazamos } H_0 \text{ si } P \notin \left[\overline{P} - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{P \cdot Q}{n}},\ \overline{P} + Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{P \cdot Q}{n}}\right]$$

Este procedimiento de contraste es equivalente a:

$$\text{Rechazar } H_0 \quad \text{si } \left|\overline{P} - P\right| > Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{P \cdot Q}{n}}$$

$$\text{No rechazar } H_0 \quad \text{si } \left|\overline{P} - P\right| < Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{P \cdot Q}{n}}$$

Ejemplo.- Disponemos de una moneda cuyo aspecto no es simétrico. Lanzamos la moneda 1000 veces y obtenemos 550 veces "cruz". ¿Es una moneda regular?. Toma como nivel de significación 5% y después el 1%.

Se trata de un test bilateral, donde las hipótesis nula y alternativa son:

$$\begin{cases} H_0: \ p = \dfrac{1}{2} \\ H_1: \ p \neq \dfrac{1}{2} \end{cases} \qquad \text{La proporción muestral de cruces es } \overline{P} = \frac{550}{1000} = 0.55$$

a) Hallamos el intervalo de confianza para p con un nivel de significación α=0.05.

Utilizamos la función 1–PropZInt de la calculadora gráfica. Introducimos en x: 550, n:1000, C–Level: 0.95. Situamos el cursor en Calculate y pulsamos [ENTER]. En la siguiente pantalla obtenemos como intervalo de confianza: $[0.519, 0.581]$. Como p=$\frac{1}{2}$=0.5 no pertenece al intervalo de confianza, rechazamos H_0 con un nivel de significación del 5%.

b) Hallamos el intervalo de confianza para p con un nivel de significación α = 0.01.

Utilizamos la función 1–PropZInt de la calculadora gráfica. Introducimos en x: 550, n:1000, C–Level: 0.99. Situamos el cursor en Calculate y pulsamos [ENTER]. En la siguiente pantalla obtenemos como intervalo de confianza: [0.509,0.591]. Como p=$\frac{1}{2}$=0.5 no pertenece a este intervalo, rechazamos H_0 con un nivel de significación del 1%. Por lo tanto, con los niveles de significación elegidos debemos concluir que la moneda no es regular.

Test unilateral 1

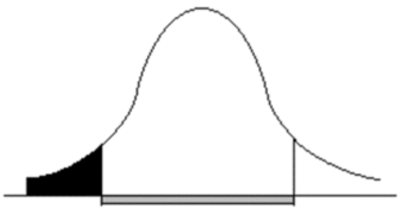
Estudiamos la proporción P de individuos de una población que tienen un determinado atributo. Extraemos una muestra de tamaño n de la población y obtenemos una proporción de éxitos en la muestra igual a $\overline{P}$. A la vista de los resultados obtenidos, establecemos el siguiente contraste de hipótesis:

$$\begin{cases} H_0 : \text{La proporción de éxitos en la población es } P \\ H_1 : \text{La proporción de éxitos en la población es } > P \end{cases}$$

¿Debemos rechazar H_0 frente a H_1 con un nivel de significación α?.

En este contraste se trata de calcular la probabilidad de que la proporción muestral $\overline{P}$ sea mayor que P siendo cierta la hipótesis nula H_0. Esta probabilidad es el p–valor que compararemos con el nivel de significación α. Si el p–valor es grande (mayor que α) no se rechaza H_0. Si el p–valor es pequeño (menor que α) se rechaza H_0 y se acepta H_1.

En este caso la región de rechazo solamente contiene una cola (a diferencia del contraste bilateral, en el que se incluyen las dos colas). Por tanto, el cuantil correspondiente de la distribución normal es $Z_{1-\alpha}$, porque el nivel de significación α se concentra únicamente en la cola de la izquierda.



Por tanto, si $P < \overline{P} - Z_{1-\alpha} \cdot \sqrt{\frac{P \cdot Q}{n}}$, entonces $\overline{P}$ toma valores mayores que P, es decir, los datos observados en la muestra son sensiblemente mayores que los supuestos en la hipótesis nula. Por tanto, aceptamos H_1: la proporción de éxitos en la población es > P, con un nivel de significación α.

En resumen:

Si $P < \overline{P} - Z_{1-\alpha} \cdot \sqrt{\frac{P \cdot Q}{n}}$, aceptamos H_1 (proporción poblacional > P).
En caso contrario, no rechazamos H_0.

Test unilateral 2

También podemos plantearnos el contraste:

$$\begin{cases} H_0 : \text{La proporción de éxitos en la población es } P \\ H_2 : \text{La proporción de éxitos en la población es } < P \end{cases}$$

Si hemos extraído una muestra obteniendo una proporción muestral de éxitos $\overline{P}$, ¿debemos rechazar la hipótesis H_0 frente a H_2 con un nivel de significación α ?

De forma análoga, calculamos la probabilidad de que $\overline{P}$ tome valores menores que P, suponiendo cierta la hipótesis nula. Esta probabilidad es el p–valor que comparamos con α. Si el p–valor es mayor que α, no rechazamos H_0 . Si el p–valor es menor que α, rechazamos H_0 y aceptamos H_2 .

Al igual que antes, la región de rechazo solamente contiene una cola, porque el nivel de significación se concentra en la cola de la derecha. El cuantil correspondiente de la distribución normal es $Z_{1-\alpha}$.

Por tanto, si $P > \overline{P} + Z_{1-\alpha} \cdot \sqrt{\dfrac{P \cdot Q}{n}}$, entonces $\overline{P}$ toma valores menores que P, es decir, los datos observados en la muestra son sensiblemente menores que los que se obtendrían si fuera cierta la hipótesis nula. Por tanto, aceptamos H_2: la proporción de éxitos en la población es < P, con un nivel de significación α.

En resumen:

Si $P > \overline{P} + Z_{1-\alpha} \cdot \sqrt{\dfrac{P \cdot Q}{n}}$, aceptamos H_2 (proporción poblacional<P). En caso contrario, no rechazamos H_0 .

CONTRASTE DE UNA PROPORCIÓN CON LA CALCULADORA GRÁFICA

Podemos resolver contrastes sobre proporciones, tanto bilaterales como unilaterales, utilizando la función 1–PropZTest de la calculadora gráfica TI–83. Esta función se encuentra en el menú TEST que se obtiene al pulsar la tecla [STAT].

Ejemplo.- Cuando se introdujo hace varios años una determinada política, el 67% de la gente votó a favor. Se piensa que actualmente hay un porcentaje mayor de votantes que está a favor de la misma política. Una muestra aleatoria de 265 votantes proporciona un porcentaje de 73.2 individuos que están de acuerdo con dicha política. Queremos examinar a través de un proceso de test si nuestra suposición es válida a partir de los obtenidos en la muestra.

Nos planteamos el contraste: $\begin{cases} H_0 : \text{La proporción poblacional es } P = 0.67 \\ H_1 : \text{La proporción poblacional es } > P = 0.67 \end{cases}$.

En este caso se cumple: $n = 265$, $\overline{P} = 0.732$ y $\overline{Q} = 1 - \overline{P} = 1 - 0.732 = 0.268$.

Puesto que $n \times \overline{P} = 265 \times 0.732 = 194 > 10$ y $n \times \overline{Q} = 265 \times 71 > 10$, podemos usar una distribución normal como aproximación a la binomial.

Pulsamos la tecla [STAT] y elegimos el menú TEST. En dicho menú pulsamos [5] para seleccionar la función 1–PropZTest.

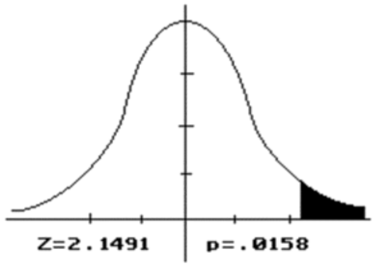
En la siguiente ventana, introduce como valor de P_0 = 0.67. Introduce como valor de X el número de éxitos obtenidos en la muestra, que es el 73.2% de 265, o sea 194. Introduce como valor de n el tamaño de la muestra, 265.

1–PropZTest
P_0 : .67
x : 194
n : 265
Prop $\neq P_0$ $<P_0$ $>P_0$
Calculate   Draw

A continuación sitúa el cursor sobre el tipo de test, en este caso $>P_0$, y pulsa ENTER. Sitúa el cursor en Calculate y pulsa ENTER.

En pantalla aparecen, por este orden, el tipo de test, el cuantil Z correspondiente de la normal, el p–valor, la proporción muestral $\hat{P}$ y el tamaño muestral n.

1–PropZTest
Prop>.67
Z=2.149060263
p=.0158147547
n=265



Si volvemos a la pantalla anterior y seleccionamos la opción Draw, al pulsar ENTER aparece una pantalla en la que se sombrea bajo la curva normal la región cuya área corresponde al p–valor de los datos.

Como el p–valor de los datos vale 0.0158 < 0.05, si fijamos un nivel de significación del 5%, debemos rechazar la hipótesis nula. Sin embargo, con un nivel de significación del 1%, no se podría rechazar la hipótesis nula, puesto que 0.0158>0.01.

ACTIVIDADES

- **CONTRA LA LEY**

Hace 10 años, el 52% de los ciudadanos estaban en contra de una ley. Recientemente, se ha elaborado una encuesta a 400 personas y 184 se mostraron contrarios a la ley. Con estos datos y con un nivel de significación del 0,01, ¿podemos afirmar que la proporción de contrarios a la ley ha disminuido?.

- **SONDEO ELECTORAL**

Un experto, basado en los anteriores comicios, sostiene que si se celebran elecciones generales en este momento tan sólo acudiría a votar el 48% de la población. No obstante, en un sondeo electoral realizado recientemente entre 1500 personas, 800 tienen intención de votar. ¿Supone esto, con un nivel de confianza del 99%, que el experto se equivoca y la intención de voto es mayor?.

CONTRASTE DE UNA MEDIA

Test bilateral

Estudiamos una población de media μ y desviación típica σ, conocidas. Extraemos una muestra de tamaño n, obteniendo una media muestral $\overline{x}$. A la vista de los resultados observados en la muestra, ¿podemos seguir afirmando que la media poblacional es igual a μ ?. Establecemos las siguientes hipótesis:

H_0: La media poblacional es μ
H_1: La media poblacional es $\neq \mu$

¿Debemos rechazar H_0 con un nivel de significación α ?.

Como σ es conocida, la distribución en el muestreo de la media muestral $\overline{x}$ es normal de media μ y desviación típica $\sigma/\sqrt{n}$.

A partir de los datos muestrales construimos el intervalo de confianza para μ con un nivel de significación α. Entonces pueden presentarse dos casos:

Si μ pertenece al intervalo de confianza, significa que la diferencia entre μ y $\overline{x}$ es pequeña y por tanto, no hay grandes discrepancias entre la hipótesis nula y los datos observados. Es decir, la probabilidad de que se obtengan los datos observados en la muestra, supuesta cierta la hipótesis nula, debe ser grande. Lo que indica que el p–valor es grande (superior a α). Por tanto, no podemos rechazar H_0.

Si μ no pertenece al intervalo de confianza, significa que la diferencia entre μ y $\overline{x}$ es grande y, por tanto, hay grandes discrepancias entre la hipótesis nula y los datos observados. Es decir, la probabilidad de que se obtengan los datos observados en la muestra, supuesta cierta la hipótesis nula, debe ser pequeña. Lo que indica que el p–valor es pequeño (inferior a α). Por tanto, debemos rechazar H_0. En resumen:

$$\text{No rechazamos } H_0 \text{ si } \mu \in \left[\overline{x} - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}},\ \overline{x} + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right]$$

$$\text{Rechazamos } H_0 \text{ si } \mu \notin \left[\overline{x} - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}},\ \overline{x} + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right]$$

Este procedimiento de contraste es equivalente a:

$$\text{Rechazar } H_0 \quad \text{si } \left|\bar{x} - \mu\right| > Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

$$\text{No rechazar } H_0 \quad \text{si } \left|\bar{x} - \mu\right| < Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

Ejemplo.- Hace algunos años, la media de estatura de los valencianos adultos era de 170 cm, con desviación típica σ = 9 cm. Pasado el tiempo, un muestreo realizado a 36 adultos da una media de 172 cm. ¿Podemos afirmar, con una confianza del 90 % que la estatura de los valencianos ha cambiado?.

Como la desviación típica poblacional es conocida, la media muestral sigue una distribución normal. Para un nivel de confianza $1-\alpha= 0.90$ el nivel de significación es $\alpha=0.10$. Por tanto, utilizaremos la función Zinterval de la calculadora gráfica. Para ello pulsamos [STAT] [◄] [7] Zinterval.

En la siguiente pantalla seleccionamos la opción Stats e introducimos los valores: $\sigma=9$, $\bar{x}=172$, n=36, C–Level=0.90. Situamos el cursor en la opción Calculate y pulsamos [ENTER]. En la siguiente pantalla se muestra el intervalo de confianza para la media: $[169.5,\ 174.5]$

Como μ =170 pertenece al intervalo de confianza [168.5, 174.5], no rechazamos la hipótesis nula H_0 . Con un nivel de confianza del 90 %, no podemos concluir que la estatura de los valencianos haya cambiado.

Nota importante

Si la desviación típica de la población, σ, es desconocida, entonces la distribución en el muestreo de la media muestral no es normal, sino que sigue una distribución T de Student. Si el tamaño de la muestra es suficientemente grande, se puede aproximar la distribucion T por una curva normal. Pero esto no es válido para muestras pequeñas.

Test unilateral 1

Estudiamos una población de media μ y desviación típica σ, conocida. Extraemos una muestra de tamaño n y obtenemos una media muestral $\bar{x}$. A la vista del resultado observado en la muestra, establecemos el siguiente contraste de hipótesis:

$$\begin{cases} H_0 : \text{La media en la población es } \mu \\ H_1 : \text{La media en la población es } >\mu \end{cases}$$

¿Debemos rechazar H_0 frente a H_1 con un nivel de significación α ?.

En este contraste se trata de calcular la probabilidad de que la media muestral $\bar{x}$ sea mayor que μ suponiendo cierta la hipótesis nula H_0. Esta probabilidad es el p–valor que compararemos con el nivel de significación α. Si el p–valor es grande (mayor que α) no se rechaza H_0. Si el p–valor es pequeño (menor que α) se rechaza H_0 y se acepta H_1.

Como la desviación típica poblacional σ es conocida, la media muestral sigue una distribución normal. En este caso la región de rechazo solamente contiene una cola (a diferencia del contraste bilateral, en el que se incluyen las dos colas). Por tanto, el cuantil correspondiente de la distribución normal es $Z_{1-\alpha}$, porque el nivel de significación α se concentra únicamente en la cola de la izquierda.

Por tanto, si $\mu < \bar{x} - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$, entonces $\bar{x}$ toma valores mayores que μ, es decir, los datos observados en la muestra son sensiblemente mayores que los supuestos en la hipótesis nula. Por tanto, aceptamos H_1: la media en la población es > μ, con un nivel de significación α.

En resumen:

Si $\mu < \bar{x} - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$, aceptamos H_1 (media poblacional > μ). En caso contrario, no rechazamos H_0 .

Test unilateral 2

También podemos plantearnos el contraste:

$$\begin{cases} H_0 : \text{La media en la población es } \mu \\ H_2 : \text{La media en la población es } < \mu \end{cases}$$

Si hemos extraído una muestra obteniendo una media muestral $\bar{x}$, ¿debemos rechazar la hipótesis H_0 frente a H_2 con un nivel de significación α ?

De forma análoga, calculamos la probabilidad de que $\bar{x}$ tome valores menores que μ, suponiendo cierta la hipótesis nula. Esta probabilidad es el p–valor que comparamos con α. Si el p–valor es mayor que α, no rechazamos H_0 . Si el p–valor es menor que α, rechazamos H_0 y aceptamos H_2 .

Al igual que antes, la distribución en el muestreo de la media muestral es normal y la región de rechazo solamente contiene una cola, porque el nivel de significación se concentra en la cola de la derecha. El cuantil correspondiente de la distribución normal es $Z_{1-\alpha}$.

Por tanto, si $\mu > \bar{x} + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$, entonces $\bar{x}$ toma valores menores que μ, es decir, los datos observados en la muestra son sensiblemente menores que los que se obtendrían si fuera cierta la hipótesis nula. Por tanto, aceptamos H_2: la media de la población es < μ, con un nivel de significación α.

En resumen:

Si $\mu > \overline{x} + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$, aceptamos H_2 (media poblacional<μ). En caso contrario, no rechazamos H_0 .

CONTRASTE DE UNA MEDIA CON LA CALCULADORA GRÁFICA

Podemos resolver contrastes sobre medias, tanto bilaterales como unilaterales, utilizando las funciones Z–Test y T–Test de la calculadora gráfica TI–83. Esta función se encuentra en el menú TEST que se obtiene al pulsar la tecla [STAT].

Ejemplo.- En una prueba atlética de velocidad celebrada el año pasado se obtuvo una marca media de 72 segundos y una desviación típica de 2.0 segundos. Recientemente se ha efectuado una modificación en la prueba. Para determinar el efecto de este cambio, se sometieron a prueba a diez atletas obteniéndose los siguientes tiempos:

76.2	**78.3**	**76.4**	**74.7**	**72.6**	**78.4**	**75.7**	**70.2**	**73.3**	**74.2**

Suponiendo que la desviación típica es la misma que antes de la modificación:

a) ¿Podemos concluir que ha cambiado el rendimiento medio de los atletas en esa prueba?.

b) ¿Podemos considerar que el rendimiento de los atletas ha aumentado?.

c) Supongamos que no puede asumirse que la desviación típica sea la misma que antes. ¿Se puede afirmar que ha cambiado el rendimiento de los atletas?.

a) En primer lugar introducimos los datos en una lista L_1. Se trata de estudiar el test

$$\begin{cases} H_0 : \text{La media en la población de atletas es } \mu = 72 \\ H_1 : \text{La media en la población de atletas es } \neq \mu = 72 \end{cases}$$

Como la desviación típica σ=2.0 es conocida, la distribución en el muestreo de la media muestral es normal. Utilizaremos, por tanto, la función Z–Test. Para ello pulsamos [STAT], seleccionamos el menú TEST y pulsamos [1].

Podemos utilizar dos procedimientos, según que indiquemos la lista donde están almacenados los datos de la muestra o que directamente introduzcamos la media de la muestra y el tamaño muestral.

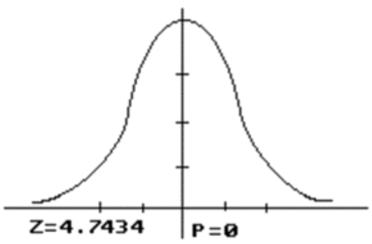
En el primer caso, seleccionamos la opción Data en la línea Inpt e introducimos en μ_0 el valor de la media poblacional (72), en σ el valor de la desviación típica poblacional (2), en List el nombre de la lista donde están los datos muestrales (L_1) y en Freq ponemos 1 (Si hubiese una lista con las frecuencias, introduciríamos el nombre de esa lista). En la línea μ : seleccionamos el tipo de test, en este caso $\neq\mu_0$.

Z–Test
Inpt: ***Data*** *Stats*
μ_0: 72
σ: 2
List: L_1
Freq: 1
μ : $\boldsymbol{\neq\mu_0}$ $<\mu_0$ $>\mu_0$
Calculate *Draw*

Si seleccionamos la opción Calculate, obtendremos los valores del cuantil de la normal (estadístico del test), el p–valor de los datos, la media, la desviación típica y el tamaño muestral.

Z–Test
$\mu \neq 72$
Z=4.74341649
p=2.1039906E–6
$\bar{x}$ =75
Sx=2.555169053
n=10

Si seleccionamos la opción Draw, se mostrará una pantalla en la que se sombrea bajo la curva normal la región cuya área corresponde al p–valor de los datos.



Un segundo procedimiento consiste en obtener la media muestral (y otros estadísticos muestrales) mediante el comando 1–Var Stat L_1. A continuación vamos a la pantalla Z–Test y en la línea Inpt: seleccionamos Stats. En μ_0 : introducimos la media poblacional (72). En σ: introducimos la desviación típica poblacional (2). En $\bar{x}$: ponemos el valor de la media muestral (75) y en n el tamaño muestral. En la línea μ: seleccionamos el tipo de test ($\neq\mu_0$).

Z–Test
Inpt: Data ***Stats***
μ_0: 72
σ: 2
$\bar{x}$: L_1
n: 1
μ : $\boldsymbol{\neq\mu_0}$ $<\mu_0$ $>\mu_0$
Calculate *Draw*

Si seleccionamos Calculate y pulsamos ENTER, obtenemos el estadístico test o cuantil de la normal, el p–valor, la media y el tamaño muestral. Y si seleccionamos Draw aparece una pantalla en la que se sombrea bajo la curva normal la región cuya área corresponde al p–valor de los datos.

Z–Test
$\mu \neq 72$
Z=4.74341649
p=2.1039906E–6
$\bar{x}$ =75
n=10

Con un p–valor de 0.000002 hay bastante evidencia de que la hipótesis nula debe ser rechazada. El rendimiento de los atletas ha cambiado.

b) Puesto que Z es positiva (4.74), podemos concluir que la media de la población es significativamente mayor que 72 segundos. El rendimiento general de los atletas ha aumentado. Esto lo podemos confirmar en el menú Z–Test, si seleccionamos como tipo de test $>\mu_0$.

c) Si suponemos que desconocemos el valor de la desviación típica poblacional σ, corresponde realizar el proceso mediante el uso del test T de Student. El contraste a plantear es:

$$\begin{cases} H_0: \text{La media en la población es } \mu = 72 \\ H_1: \text{La media en la población es } \neq \mu = 72 \end{cases}$$

Pulsamos [STAT] y seleccionamos el menú TEST. Pulsamos [2] para seleccionar el menú T–Test. Al igual que antes podemos seleccionar en la línea Inpt si queremos introducir una lista (Data) o directamente los estadísticos muestrales (Stats). Introducimos los valores tal como hemos hecho en apartados anteriores.

T–Test
Inpt: ***Data*** *Stats*
μ_0: 72
List: L_1
Freq: 1
μ : $\neq\mu_0$ $<\mu_0$ $>\mu_0$
Calculate Draw

Al seleccionar Calculate, obtenemos los valores del estadístico test T y el p–valor de los datos. Si seleccionamos Draw se mostrará la pantalla en la que se sombrea el área bajo la curva normal que representa el p–valor de los datos.

T–Test
$\mu \neq 72$
t=3.712800517
p=.0048240361
$\bar{x}$=75
Sx=2.555169053
n=10

Como el p–valor es pequeño (0.0048), debemos rechazar la hipótesis nula, es decir, el rendimiento de los atletas ha cambiado.

Ejemplo.- El peso de los adultos de cierta localidad se distribuye normalmente con media 65 kg y desviación típica 12 kg. Se elige una muestra, al azar, de 50 individuos de dicha localidad, resultando un peso medio de $\bar{x} = 70$ kg. Para una significación de α = 0.05, ¿puede decirse que los ciudadanos de esa localidad han aumentado de peso?.

En este caso tenemos: μ=65, σ=12, n=50, $\bar{x} = 70$. Establecemos el contraste:

$$\begin{cases} H_0 : \text{La media en la población es } \mu = 65 \\ H_1 : \text{La media en la población es } > \mu = 65 \end{cases}.$$

Para resolver el contraste hemos de ver si se cumple: $\mu < \bar{x} - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$.

Como la desviación típica poblacional σ es conocida, la distribución en el muestreo de la media muestral es normal. Con la calculadora gráfica observamos que el cuantil correspondiente de la distribución normal es: $Z_{1-\alpha} = Z_{0.95} = 1.64$. Entonces:

$$\bar{x} - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}} = 70 - 1.64 \times \frac{12}{\sqrt{50}} = 70 - 2.7831723 = 67.216828 \approx 67.22$$

Como μ=65<67.22, se rechaza H_0 con un nivel de significación del 5%. Es decir, con un nivel de confianza del 95% podemos afirmar que los individuos de esa localidad han aumentado de peso. Sin embargo este problema se puede resolver automáticamente con la calculadora gráfica, utilizando la técnica que hemos visto en el siguiente apartado.

En efecto, pulsamos [STAT] [◄] [1] Z–Test. En la siguiente pantalla seleccionamos la opción Stats e introducimos los valores: $\mu_0 = 65$, σ = 12, $\bar{x} = 70$, n = 50. Seleccionamos la opción $> \mu_0$, situamos el cursor sobre la opción Calculate y pulsamos [ENTER]. En la siguiente pantalla obtenemos un p–valor p=0.0016. Como el p–valor es menor que 0.05, rechazamos H_0 con un nivel de significación del 5%. Los individuos de la localidad han aumentado de peso.

ACTIVIDADES

- **SALARIOS**

El salario medio correspondiente a una muestra de 1600 personas de cierta población es de 935 euros. Se sabe que la desviación típica de los salarios en la población es de 200 euros. ¿Se puede afirmar, con un nivel de significación de 0,01, que el salario medio en dicha población es de 950 euros?.

- **JAQUECA**

Un laboratorio afirma que un calmante quita la jaqueca en 14 minutos en los casos corrientes. Con el fin de comprobar esta información, se eligen al azar 30 pacientes con jaqueca y se toma como variable en el experimento el tiempo que transcurre entre la administración del calmante y el momento en que desaparece la jaqueca. Los resultados obtenidos en esta muestra fueron: media 17 minutos y desviación típica 7 minutos. ¿Podemos admitir como cierta la afirmación del laboratorio a un nivel de confianza del 95%?.

- **TEST**

El estudio de un test de satisfacción de usuario que rellenan todos los demandantes de servicios de una gran empresa revela que la nota media que otorgan es de 5,70 puntos con una desviación típica de 0,5.

Posteriormente, se ha realizado un muestreo a 100 usuarios de la zona de influencia A, y a 49 usuarios de la zona B, obteniéndose puntuaciones medias respectivas de 5,6 y 5,85.

Con una confianza del 95%, ¿se puede afirmar que las diferencias entre las medias de cada muestra y de la población son debidas al azar, o se puede afirmar que son diferentes la nota media de la población y la de cada muestra?.

- **EMPLEO TEMPORAL**

Una encuesta realizada a 64 empleados de una fábrica, concluyó que el tiempo medio de duración de un empleo en la misma era de 6,5 años, con una desviación típica de 4. ¿Sirve esta información para aceptar, con un nivel de significación del 5%, que el tiempo medio de empleo en esa fábrica es menor o igual que 6 años?. Justifica adecuadamente la respuesta.

- **PROPAGANDA**

En los folletos de propaganda, una empresa asegura que las bombillas que fabrica tienen una duración media de 1600 horas. A fin de contrastar este dato, se tomó una muestra aleatoria de 100 bombillas, obteniéndose una duración media de 1570 horas, con una desviación típica de 120 horas. ¿Puede aceptarse la información de los folletos con un nivel de confianza del 95%?.

- **ESTUDIOS SUPERIORES**

Se ha llevado a cabo un estudio en diferentes países de la Unión Europea del porcentaje de la población que accede a la enseñanza superior.

En los países escogidos se han obtenido los valores siguientes (medidos en tanto por ciento):

23,5	35,0	29,5	31,0	23,0	33,5	27,0	28,0	30,5

Se supone que estos porcentajes siguen una distribución normal con desviación típica igual al 5 por ciento.

Se desea contrastar con un nivel de significación del 5% si los datos anteriores son compatibles con un valor medio del porcentaje de la población que cursa estudios superiores igual al 28 por ciento.

a) Plantea en el contraste cuáles son las hipótesis nula y la alternativa.

b) Determina la región crítica del contraste.

c) ¿Es posible aceptar la hipótesis con el nivel de significación indicado?.

- **JUVENTUD**

La Concejalía de Juventud de un Ayuntamiento maneja el dato de que la edad a la que los hijos se independizan de sus padres es una variable normal con media 29 años y desviación típica 3 años. Aunque la desviación típica no plantea dudas, sí se sospecha que la media ha descendido, sobre todo por la política de ayuda al empleo que ha llevado a cabo el Ayuntamiento. Así, de un estudio reciente sobre 100 jóvenes que se acaban de independizar, se ha obtenido una media de 28,1 años de edad.

a) Con un nivel de significación del 1%, ¿puede defenderse que la edad media no ha disminuido, frente a que sí lo ha hecho como parecen indicar los datos?. Plantea el contraste o test de hipótesis y resuélvelo.

b) Explica, en el contexto del problema, en qué consisten cada uno de los errores del tipo I y II.

INFERENCIA ESTADÍSTICA CON LA CALCULADORA GRÁFICA CLASSPAD 300 DE CASIO

Introducción

La calculadora ClassPad 300 permite obtener con facilidad estimaciones de parámetros, determinar intervalos de confianza, validar hipótesis, etc.

En las siguientes actividades estudiaremos algunas de las posibilidades de la ClassPad 300 para el estudio de la Inferencia Estadística en ESO y Bachillerato

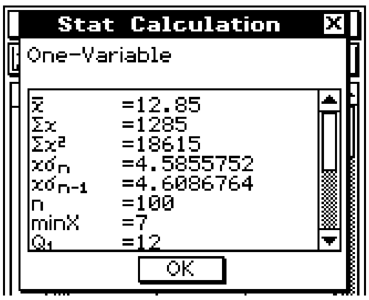
1. Inferencia estadística

1. INTERVALOS DE CONFIANZA

- **Intervalo de confianza para la media**
- El comando OneSampleZInt (situado en el teclado virtual [cat]) calcula el intervalo de confianza para la media poblacional cuando se conoce la desviación típica de la población. Para ello utiliza las fórmulas: $\left(\overline{x}-Z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}},\overline{x}+Z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}\right)$, siendo α el nivel de significación y $1-\alpha$ el nivel de confianza.
- Si se conoce la lista de datos, la sintaxis del comando es: **OneSampleZInt 1-α, σ, List, Frec**, siendo α el nivel de significación, σ la desviación típica, List el nombre de la lista de datos, Frec la lista que contiene las frecuencias de los datos.
- Si se conocen los parámetros estadísticos de la muestra, la sintaxis del comando es la siguiente: **OneSampleZInt 1-α, σ, $\overline{x}$, n**, siendo α el nivel de significación, σ la desviación típica, $\overline{x}$ la media y n el tamaño de la muestra.
- La siguiente tabla muestra las duraciones (en días) de 100 pastillas de jabón de una determinada marca. Halla un intervalo de confianza para la duración media de dichas pastillas con un nivel de significación α=0,05. Sigue los siguientes pasos:

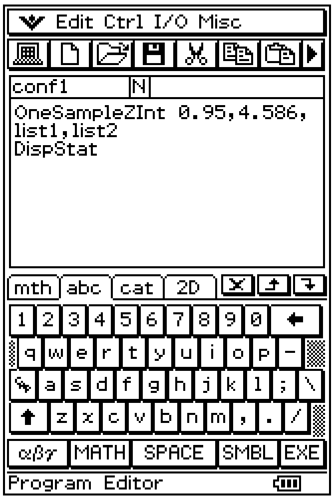
Duración (días)	7	12	17	22
Frecuencia	24	46	19	11

1) En el editor de listas de la aplicación Estadística, introduce en las listas list1 y list2 las duraciones y las frecuencias, respectivamente. Selecciona el comando Calc./ Una variable e introduce como frecuencias la lista list2. Toca el botón [Acep.] y observa el resultado.

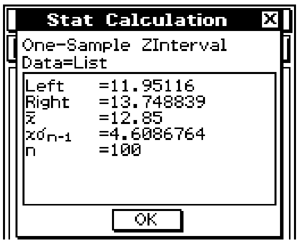


2) En el editor de programas de la aplicación Programas, selecciona el comando Edit / Archivo nuevo. En la siguiente ventana introduce como nombre del archivo conf1 y toca el botón [Acep.].

3) En la siguiente ventana, toca el botón del teclado virtual [cat] y selecciona Todo en la lista desplegable Forma. En el catálogo de comandos selecciona OneSampleZInt y toca el botón [INTRO] para introducir dicha función en la ventana de edición del programa. Con ayuda del teclado virtual [math] completa el comando OneSampleZInt 0.95, 4.586, list1, list2. Toca el botón [Ejec.]. En la siguiente línea del programa selecciona el comando E/S / Visualización / DispStat.



4) Selecciona el comando Edit / Guardar archivo. A continuación selecciona el comando /Cargador programa. En la lista desplegable Carpeta selecciona la carpeta donde está guardado el programa, en nuestro caso, la carpeta principal main. En la lista desplegable Nombre selecciona el nombre del programa, conf1. Haz clic en el botón [▸] o selecciona el comando Ejecutar / Ejecutar programa. Aparece una pantalla con el resultado del cálculo estadístico. En ella se indica que el intervalo de confianza del 95% es (11.95, 13.75).



- El peso medio de una muestra de 100 recién nacidos es 3200 g. Sabiendo que la desviación típica de los pesos de la población de recién nacidos es 150 gramos, halla el intervalo de confianza para la media poblacional con un nivel de significación de 0,05. Utiliza para ello un programa de nombre conf2 con los comandos OneSampleZInt 0.95, 150, 3200, 100 y DispStat.

- El comando OneSampleTInt (situado en el teclado virtual [cat]) calcula el intervalo de confianza para la media poblacional cuando se desconoce la desviación típica de la población y el tamaño de la muestra es pequeño. Para ello utiliza las fórmulas: $\left(\bar{x} - t_{n-1,\frac{\alpha}{2}} \cdot \frac{x\sigma_{n-1}}{\sqrt{n}}, \bar{x} + t_{n-1,\frac{\alpha}{2}} \cdot \frac{x\sigma_{n-1}}{\sqrt{n}}\right)$, siendo α el nivel de significación y 1-α el nivel de confianza.

- Si se conoce la lista de datos, la sintaxis del comando es: **OneSampleTInt 1-α, List, Frec**, siendo α el nivel de significación, List el nombre de la lista de datos, Frec la lista que contiene las frecuencias de los datos.

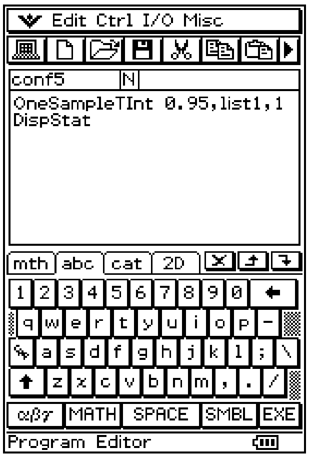
- Si se conocen los parámetros estadísticos de la muestra, la sintaxis del comando es la siguiente: **OneSampleTInt 1-α, $\bar{x}$, xσ_{n-1}, n**, siendo α el nivel de significación, xσ_{n-1} la desviación típica muestral, $\bar{x}$ la media y n el tamaño de la muestra.

- El gasto semanal de fotocopias, en céntimos de euro, para una muestra de 9 estudiantes es: 100, 150, 90, 70, 75, 105, 200, 120, 80. Halla un intervalo de confianza al 95% para la media de gasto semanal en fotocopias por estudiante. Sigue los siguientes pasos:

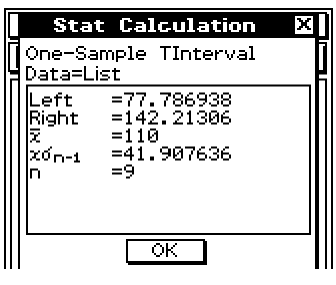
1) En el editor de listas de la aplicación Estadística, introduce en la lista list1 los gastos en fotocopias.

2) En el editor de programas de la aplicación Programas, selecciona el comando Edit / Archivo nuevo. En la siguiente ventana introduce como nombre del archivo conf5 y toca el botón [Acep.].

3) En la siguiente ventana, toca el botón del teclado virtual [cat] y selecciona Todo en la lista desplegable Forma. En el catálogo de comandos selecciona OneSampleTInt y toca el botón [INTRO] para introducir dicha función en la ventana de edición del programa. Con ayuda del teclado virtual [math] completa el comando OneSampleTInt 0.95, list1, 1. Toca el botón [Ejec.]. En la siguiente línea del programa selecciona el comando E/S / Visualización / DispStat.



4) Selecciona el comando Edit / Guardar archivo. A continuación selecciona el comando /Cargador programa. En la lista desplegable Carpeta selecciona la carpeta donde está guardado el programa, en nuestro caso, la carpeta principal main. En la lista desplegable Nombre selecciona el nombre del programa, conf5. Haz clic en el botón [▸] o selecciona el comando Ejecutar / Ejecutar programa. Aparece una pantalla con el resultado del cálculo estadístico. En ella se indica que el intervalo de confianza del 95% es (77.79, 142.21).



- En una prueba de 100 metros participa una muestra de 10 atletas con un tiempo medio de 11 segundos y una desviación típica muestral de 1.1547 segundos. Halla un intervalo de confianza para la media de tiempos en dicha prueba con un nivel de confianza del 95%. Utiliza para ello un programa de nombre conf6 con los comandos OneSampleTInt 0.95, 11, 1.1547, 10 y DispStat.

- **Intervalo de confianza para la proporción**

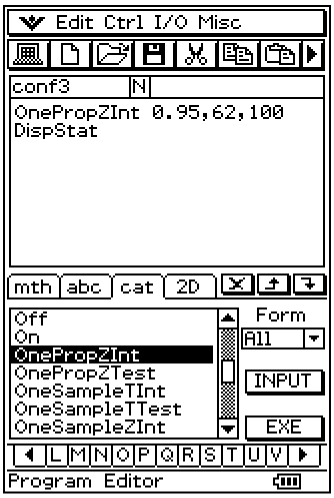
- El comando OnePropZInt (situado en el teclado virtual [cat]) calcula el intervalo de confianza para la proporción de éxitos en una población. Para ello utiliza las fórmulas: $\left(\frac{x}{n}-Z_{\frac{\alpha}{2}}\cdot\sqrt{\frac{1}{n}\cdot\frac{x}{n}\cdot\left(1-\frac{x}{n}\right)},\frac{x}{n}+Z_{\frac{\alpha}{2}}\cdot\sqrt{\frac{1}{n}\cdot\frac{x}{n}\cdot\left(1-\frac{x}{n}\right)}\right)$, siendo α el nivel de significación y $1-\alpha$ el nivel de confianza, x el dato y n el tamaño de la muestra.

- La sintaxis del comando es: **OnePropZInt 1-α, x, n**, siendo α el nivel de significación, x el dato y n el tamaño de la muestra.

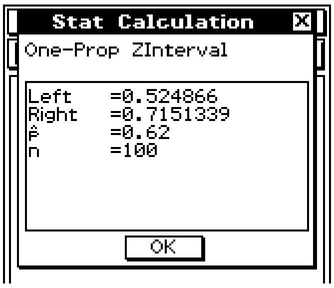
- Se ha lanzado 100 veces una moneda obteniéndose 62 caras. Halla un intervalo de confianza para la proporción de caras, con un nivel de confianza del 95%. Sigue los siguientes pasos:

1) En el editor de programas de la aplicación Programas, selecciona el comando Edit / Archivo nuevo. En la siguiente ventana introduce como nombre del archivo conf3 y toca el botón [Acep.].

2) En la siguiente ventana, toca el botón del teclado virtual [cat] y selecciona Todo en la lista desplegable Forma. En el catálogo de comandos selecciona OnePropZInt y toca el botón [INTRO] para introducir dicha función en la ventana de edición del programa. Con ayuda del teclado virtual [math] completa el comando OnePropZInt 0.95, 62, 100. Toca el botón [Ejec.]. En la siguiente línea del programa selecciona el comando E/S / Visualización / DispStat.



3) Selecciona el comando Edit / Guardar archivo. A continuación selecciona el comando /Cargador programa. En la lista desplegable Carpeta selecciona la carpeta donde está guardado el programa, en nuestro caso, la carpeta principal main. En la lista desplegable Nombre selecciona el nombre del programa, conf3. Haz clic en el botón [▶] o selecciona el comando Ejecutar / Ejecutar programa. Aparece una pantalla con el resultado del cálculo estadístico. En ella se indica que el intervalo de confianza del 95% es (0.525, 0.715).



- Tomando una muestra de 300 personas en una gran ciudad, se encontró que 104 de ellas leían el periódico regularmente. Halla, con un nivel de confianza del 90% un intervalo de confianza para la proporción de lectores de periódicos. Utiliza un programa conf4 con los comandos OnePropZInt 0.90, 104, 300 y DispStat.

2. TESTS DE HIPÓTESIS

- **Contraste de una media**

- El comando OneSampleZTest (situado en el teclado virtual [cat]) contrasta una hipótesis relativa a una media poblacional cuando la desviación típica de la población es conocida. Para una distribución normal se utiliza el estadístico: $Z = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$, siendo $\overline{x}$ la media de los datos de la muestra, μ_o la media poblacional supuesta, σ la desviación típica de la población y n el tamaño de la muestra.

- Si se conoce la lista de datos, la sintaxis del comando es: **OneSampleZTest "condición", μ_o, σ, List, Frec**, siendo List y Frec los nombres de las listas que contienen los datos y las frecuencias y siendo:

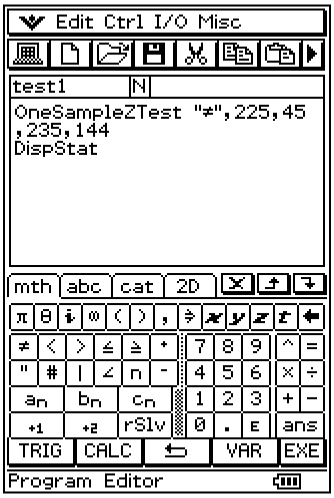
$$\text{Condición} = \begin{cases} \neq, & \text{si la prueba es de dos colas} \\ <, & \text{si la prueba es de cola inferior} \\ >, & \text{si la prueba es de cola superior} \end{cases}.$$

- Si se conocen los parámetros estadísticos de la muestra, la sintaxis del comando es la siguiente: **OneSampleZTest "condición", μ_o, σ, $\overline{x}$, n**, siendo σ la desviación típica poblacional, $\overline{x}$ la media y n el tamaño de la muestra.

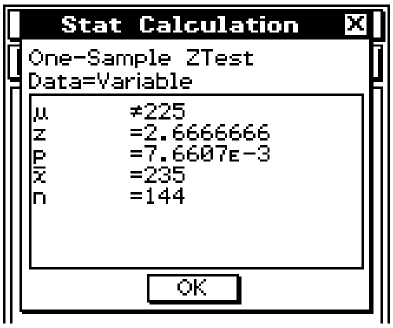
- El nivel de colesterol (en mg/dl) para una muestra de 144 personas mayores de 60 años tiene una media de 235, con una desviación típica de 45. ¿Se puede admitir que la media de colesterol de la población de mayores de 60 años es de 225, con un nivel de confianza del 95%?. Sigue estos pasos:

1) En el editor de programas de la aplicación Programas, selecciona el comando Edit / Archivo nuevo. En la siguiente ventana introduce como nombre del archivo test1 y toca el botón [Acep.].

2) En la siguiente ventana, toca el botón del teclado virtual [cat] y selecciona Todo en la lista desplegable Forma. En el catálogo de comandos selecciona OneSampleZTest y toca el botón [INTRO] para introducir dicha función en la ventana de edición del programa. Con ayuda del teclado virtual [math] completa el comando OneSampleZTest "≠", 225, 45, 235, 144. Toca el botón [Ejec.]. En la siguiente línea del programa selecciona el comando E/S / Visualización / DispStat.



3) Selecciona el comando Edit / Guardar archivo. A continuación selecciona el comando /Cargador programa. En la lista desplegable Carpeta selecciona la carpeta donde está guardado el programa, en nuestro caso, la carpeta principal main. En la lista desplegable Nombre selecciona el nombre del programa, test1. Haz clic en el botón [▸] o selecciona el comando Ejecutar / Ejecutar programa. Aparece una pantalla con el resultado del cálculo estadístico. Como el p-valor, $p=7.6\times10^{-3}$ es pequeño, podemos rechazar, con un nivel de significación del 5%, la hipótesis nula de que la media de colesterol es de 225 mg/dl.



- El peso de los adultos de la población valenciana se distribuye normalmente con media 65 kg y desviación típica 12 kg. Se elige una muestra, al azar, de 200 individuos, resultando un peso medio de 70 kg. Para un nivel de significación $\alpha=0.05$, ¿puede decirse que los valencianos pesan más que antes?. Utiliza un programa de nombre test3 con los comandos OneSampleZTest “>”, 65, 12, 70, 200 y DispStat.

- El comando OneSampleTTest (situado en el teclado virtual [cat]) contrasta una hipótesis relativa a una media poblacional cuando la desviación típica de la población es desconocida y la muestra es pequeña. Para una distribución t se utiliza el estadístico: $t=\dfrac{\bar{x}-\mu_0}{x\sigma_{n-1}/\sqrt{n}}$, siendo $\bar{x}$ la media de los datos de la muestra, μ_o la media poblacional supuesta, $x\sigma_{n-1}$ la desviación típica de la muestra y n el tamaño de la muestra.

- Si se conoce la lista de datos, la sintaxis del comando es: **OneSampleTTest “condición”, μ_o, List, Frec**, siendo List y Frec los nombres de las listas que contienen los datos y las frecuencias y siendo:

$$\text{Condición}=\begin{cases}\neq, & \text{si la prueba es de dos colas}\\ <, & \text{si la prueba es de cola inferior}\\ >, & \text{si la prueba es de cola superior}\end{cases}.$$

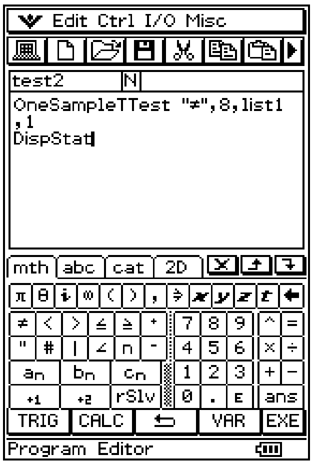
- Si se conocen los parámetros estadísticos de la muestra, la sintaxis del comando es la siguiente: **OneSampleTTest "condición", μ_0, $\bar{x}$, $x\sigma_{n-1}$, n**, siendo $x\sigma_{n-1}$ la desviación típica muestral, $\bar{x}$ la media y n el tamaño de la muestra.

- Según un estudio, el número medio de novelas leídas cada curso por los universitarios españoles es de 8. Se toma una muestra de diez estudiantes obteniéndose los datos siguientes de novelas leídas en el último curso: 14, 10, 5, 11, 0, 4, 7, 8, 13, 20. ¿Podemos admitir que ese valor medio es válido para la población de estudiantes muestreada?. Sigue los siguientes pasos:

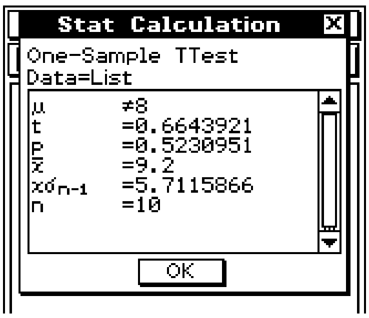
1) En el editor de listas de la aplicación Estadística, introduce los datos en la lista list1.

2) En el editor de programas de la aplicación Programas, selecciona el comando Edit / Archivo nuevo. En la siguiente ventana introduce como nombre del archivo test2 y toca el botón [Acep.].

3) En la siguiente ventana, toca el botón del teclado virtual [cat] y selecciona Todo en la lista desplegable Forma. En el catálogo de comandos selecciona OneSampleTTest y toca el botón [INTRO] para introducir dicha función en la ventana de edición del programa. Con ayuda del teclado virtual [math] completa el comando OneSampleTTest "≠", 8, list1, 1. Toca el botón [Ejec.]. En la siguiente línea del programa selecciona el comando E/S / Visualización / DispStat.



4) Selecciona el comando Edit / Guardar archivo. A continuación selecciona el comando /Cargador programa. En la lista desplegable Carpeta selecciona la carpeta donde está guardado el programa, en nuestro caso, la carpeta principal main. En la lista desplegable Nombre selecciona el nombre del programa, test2. Haz clic en el botón [▶] o selecciona el comando Ejecutar / Ejecutar programa. Aparece una pantalla con el resultado del cálculo estadístico. Como el p–valor, p=0.5230951, es mayor que el nivel de significación, no rechazamos la hipótesis nula de que la media sea 8.



- Un fabricante indica en el envoltorio que el género que contiene pesa 250 g. Se toma una muestra de 30 paquetes obteniendo un peso medio de 195 g, con una desviación típica de 10 g. ¿Hay evidencia de que por término medio los paquetes contienen menos género que lo indicado en la etiqueta?. Utiliza un programa de nombre test4 con los comandos OneSampleTTest “<”, 250, 195, 10, 30 y DispStat.

- **Contraste de una proporción**

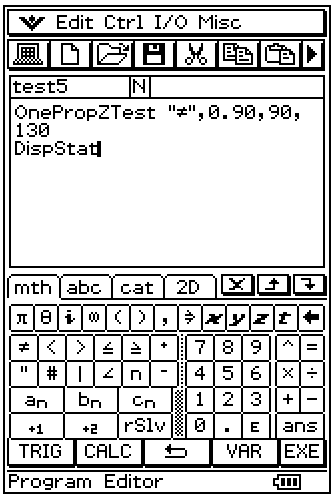
- El comando OnePropZTest (situado en el teclado virtual [cat]) contrasta si el número de éxitos alcanza un proporción fija. Para una distribución normal se utiliza el estadístico: $Z = \dfrac{\frac{x}{n} - p_o}{\sqrt{\dfrac{p_o \cdot (1 - p_o)}{n}}}$, siendo p_o la proporción esperada de la población y n el tamaño de la muestra.

- La sintaxis del comando es: OnePropZTest “cond”, p_o, x, n, siendo cond = {≠, <, >}, p_o la proporción esperada en la población (entre 0 y 1), x el número de éxitos obtenidos en la muestra, n el tamaño de la muestra.

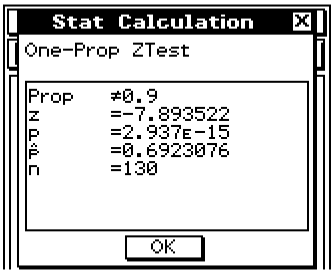
- Un medicamento es anunciado como eficaz al 90 por 100 para reducir las alergias en un período de 6 horas. Un hospital decide comprobarlo y suministra el medicamento a 130 pacientes obteniendo éxito en 90 de ellos. ¿Es cierta la eficacia que se afirma en el anuncio?. Sigue los siguientes pasos:

1) En el editor de programas de la aplicación Programas, selecciona el comando Edit / Archivo nuevo. En la siguiente ventana introduce como nombre del archivo test5 y toca el botón [Acep.].

2) En la siguiente ventana, toca el botón del teclado virtual [cat] y selecciona Todo en la lista desplegable Forma. En el catálogo de comandos selecciona OnePropZTest y toca el botón [INTRO] para introducir dicha función en la ventana de edición del programa. Con ayuda del teclado virtual [math] completa el comando OnePropZTest “≠”, 0.90, 90, 130. Toca el botón [Ejec.]. En la siguiente línea del programa selecciona el comando E/S / Visualización / DispStat.



3) Selecciona el comando Edit / Guardar archivo. A continuación selecciona el comando /Cargador programa. En la lista desplegable Carpeta selecciona la carpeta donde está guardado el programa, en nuestro caso, la carpeta principal main. En la lista desplegable Nombre selecciona el nombre del programa, test5. Haz clic en el botón [▶] o selecciona el comando Ejecutar / Ejecutar programa. Aparece una pantalla con el resultado del cálculo estadístico. Como el p–valor es menor que el nivel de significación, rechazamos la hipótesis nula de que la eficacia del medicamento sea la anunciada.



- Una asociación ecologista se opone a la construcción de una presa aduciendo que la mayor parte de los habitantes de la zona se oponen también a su construcción. Para comprobar tal opinión, se realiza un estudio preguntando a 400 ciudadanos. De ellos, están en contra de la presa 220. Para un nivel de confianza del 95%, ¿puede asegurarse que la mayoría de los habitantes de la zona se oponen a la construcción de la presa?. Utiliza un programa de nombre test6 con los comandos OnePropZTest “>”, 0.50, 220, 400 y DispStat.

- De cara a las próximas elecciones, se ha realizado una encuesta en una localidad de 10000 habitantes y, de una muestra de 500 personas, 320 han declarado que piensan votar al partido A. ¿Se puede afirmar que el porcentaje de votantes del partido A en esa población superará al 50%, con un nivel de significación del 5%?. Utiliza un programa de nombre test7 con los comandos OnePropZTest “>”, 0.50, 320, 500 y DispStat.

- Queremos decidir si una moneda está bien construida (no está sesgada). Para ello, la lanzamos 1000 veces, obteniendo 350 caras. Con un nivel de significación del 5%, ¿podemos afirmar que la moneda es correcta?. Utiliza un programa de nombre test8 con los comandos OnePropZTest “≠”, 0.50, 350, 1000 y DispStat.

2. Actividades

1) Se ha medido la longitud de 13 plantas de una especie de soja, obteniendo los siguientes resultados:

20,2	22,9	23,3	20,0	19,4	22,0	22,1	22,0	21,9	21,5	19,7	21,5	20,9

Halla un intervalo de confianza para la longitud media de esta especie de plantas, con un nivel de significación del 5%.

2) Una muestra aleatoria de 100 alumnos que se presenta a las pruebas de Selectividad, revela que la media de edad es de 18,1 años. Halla un intervalo de confianza de 90% para la edad media de todos los estudiantes que se presentan a las pruebas, sabiendo que la desviación típica de la población es de 0,4.

3) En un sondeo electoral realizado a 273 personas de una población, se manifestaron 82 personas favorables a un determinado partido político. ¿Cuál es el intervalo de confianza del 95% para la proporción de la población total que votará a dicho partido?.

4) Los gastos mensuales de las familias de un municipio se distribuyen normalmente. Si seleccionamos a 30 familias al azar y obtenemos como media 1500 euros de gastos y desviación típica 300 euros, halla un intervalo de confianza para la media de los gastos mensuales de las familias del municipio, con un nivel de confianza del 90%.

5) Se desea hacer un estudio de mercado para conocer el precio medio de los libros científicos. Para ello, se elige una muestra aleatoria formada por 34 libros y se determina que la media muestral es de 34,9 euros con una desviación típica de 4,5 euros. Halla el intervalo de confianza para el precio medio de los libros científicos con un nivel de confianza del 99%.

6) Cuando se introdujo hace varios años una determinada política, el 67% de la gente votó a favor. Se piensa que actualmente hay un porcentaje mayor de votantes que está a favor de la misma política. Una muestra aleatoria de 265 votantes proporciona un porcentaje de 73,2 individuos que están de acuerdo con dicha política. A partir de los datos de la muestra, ¿podemos admitir como válida nuestra suposición, con un nivel de confianza del 95%? ¿Y con un nivel de confianza del 99%?.

7) Hace 10 años, el 52% de los ciudadanos estaban en contra de una ley. Recientemente, se ha elaborado una encuesta a 400 personas y 184 se mostraron contrarios a la ley. Con estos datos y con un nivel de significación del 0,01, ¿podemos afirmar que la proporción de contrarios a la ley ha disminuido?.

8) Un experto, basado en los anteriores comicios, sostiene que si se celebran elecciones generales en este momento tan solo acudiría a votar el 48% de la población. No obstante, en un sondeo electoral realizado recientemente entre 1500 personas, 800 tienen intención de votar. ¿Supone esto, con un nivel de confianza del 99%, que el experto se equivoca y la intención de voto es mayor?.

9) Hace algunos años, la media de estatura de los valencianos adultos era de 170 cm, con desviación típica σ=9 cm. Pasado el tiempo, un muestreo realizado a 36 adultos da una media de 172 cm. ¿Podemos afirmar, con una confianza del 90% que la estatura de los valencianos ha cambiado?.

10) En una prueba atlética de velocidad celebrada el año pasado se obtuvo una marca media de 72 segundos y una desviación típica de 2 segundos. Recientemente se ha efectuado una modificación en la prueba. Para determinar el efecto de este cambio, se sometieron a prueba a diez atletas obteniéndose los siguientes tiempos:

76,2	78,3	76,4	74,7	72,6	78,4	75,7	70,2	73,3	74,2

Suponiendo que la desviación típica es la misma que antes de la modificación:

a) ¿Podemos concluir que ha cambiado el rendimiento medio de los atletas en esa prueba?.

b) ¿Podemos considerar que el rendimiento de los atletas ha aumentado?.

c) Si sabemos que la desviación típica no es la misma que antes de la modificación, ¿cuáles serían entonces las respuestas a los apartados (a) y (b)?.

11) Se sabe, por trabajos realizados por expertos, que la velocidad lectora media de los niños de 6 años es de 40 palabras por minuto, siendo la desviación típica de 12. Hemos tomado una muestra aleatoria de 49 niños de 6 años y les hemos medido su velocidad lectora, resultando una media de 42 palabras por minuto. ¿Podemos afirmar que nuestra media es compatible con la de los expertos a un nivel de confianza del 99%?.

12) Un estudio de un fabricante de televisores indica que la duración media de un televisor es de 10 años, con una desviación típica de 0,7 años. Suponiendo que la duración media sigue una distribución normal:

   a) Calcula la probabilidad de que un televisor dure más de 9 años.

   b) Calcula la probabilidad de que dure entre 9 y 11 años.

13) La probabilidad de que un esquiador debutante se caiga en la pista es 0,4. Si lo intenta 5 veces, calcula la probabilidad de que se caiga al menos 3 veces.

14) En una cierta prueba, el 35% de la población examinada obtuvo una nota superior a 6, el 25%, entre 4 y 6, y el 40% inferior a 4. Suponiendo que las notas siguen una distribución normal, halla la nota media y la desviación típica. ¿Qué porcentaje de la población tiene una nota que se diferencia de la media en menos de 2 unidades?

15) Se sabe que el peso de los recién nacidos en una determinada población sigue una distribución normal de media 3600 g y desviación típica 280 g. Se toma una muestra al azar de 196 de estos recién nacidos y se calcula la media. ¿Cuál es la probabilidad de que esta media esté entre 3580 g y 3620 g?.

16) En una muestra de 400 personas de una población hay 80 que tienen teléfono móvil. Calcula el intervalo de confianza aproximado para la proporción poblacional para un nivel de confianza del 95%.

17) Un laboratorio farmacéutico afirma que el número de horas que un medicamento de fabricación propia tarda en curar una determinada enfermedad sigue una variable normal con desviación típica igual a 8. Se trata un muestra de 100 enfermos a los que se les suministra el medicamento y se observa que la media de horas que tardan en curarse es igual a 32.

   a) Encuentra un intervalo de confianza, con un nivel de significación del 99% para la media del número de horas que tarda en curar el medicamento.

   b) Si el nivel de significación es igual a 0,05, ¿cuál es el tamaño de la muestra que habría que considerar para estimar el valor de la media con un error menor de 3 horas?.

18) Se sabe que 2 de cada 8 habitantes de una ciudad utiliza el transporte público para ir a su trabajo. Se hace una encuesta a 140 de esos ciudadanos.

   a) Halla el número esperado de ciudadanos que no van a su trabajo en transporte público.

   b) Halla la probabilidad de que el número de ciudadanos que van al trabajo en transporte público esté entre 30 y 45.

19) En una muestra de 600 personas de una ciudad se observa que 30 son inmigrantes.

   a) Halla un intervalo de confianza de nivel 0,95 para el porcentaje de inmigrantes en la ciudad.

   b) Si se quiere estimar el porcentaje de inmigrantes con un error máximo de 0,02, ¿cuál es el tamaño de la muestra que habría que considerar si se una un nivel de significación del 1%?.

20) El equipo directivo afirma que la media del recorrido que hacen los alumnos que asisten a un centro de bachillerato es, a lo sumo, igual a 2,5 km con una desviación típica igual a 0,5 km. Se toma una muestra de 81 alumnos y se obtiene para ellos un recorrido medio de 2,6 km.

   a) ¿Se puede aceptar con un nivel de significación igual a 0,05 la afirmación del equipo directivo?.

   b) ¿La respuesta del apartado anterior es la misma si el nivel de confianza es del 99%?

21) Cuando una máquina funciona correctamente, produce piezas cuya longitud sigue una ley normal de media 12 cm y desviación típica 1 cm. El encargado del control de calidad ha tomado una muestra de 25 piezas obteniendo una media de 11,5 cm.

   a) Contrasta la hipótesis de que la máquina está funcionando correctamente, con un nivel de significación igual a 0,05.

   b) Calcula el intervalo de confianza al nivel de 95% para la longitud media de las piezas que está produciendo la máquina.

22) La duración de las llamadas de teléfono, en una oficina comercial, sigue una distribución normal con desviación típica 10 segundos. Se hace una encuesta entre 50 llamadas y la media de duración obtenida en esa muestra es 35 segundos. Calcula un intervalo de confianza al 99% para la duración media de las llamadas.

23) Preguntadas 100 personas de cierta ciudad, elegidas al azar, si leen el periódico al menos una vez a la semana, solo 40 han contestado que sí. Halla un intervalo de confianza, con nivel de confianza del 99%, para la proporción de personas de esa ciudad que leen el periódico al menos una vez a la semana.

24) Al lanzar 5000 veces una moneda al aire salieron 3000 caras. ¿Se puede aceptar, con un nivel de significación del 0,05 que la moneda no está trucada?.

25) Con el fin de estimar la edad media de los habitantes de una gran ciudad, se tomó una muestra aleatoria de 300 habitantes, que arrojó una edad media de 35 años y una desviación típica de 7 años.

   a) Halla el intervalo del 95% de confianza en el que se encontrará la edad media de la población.

   b) ¿Qué nivel de confianza se debería usar para que el intervalo fuera $35 \pm 0,44$?