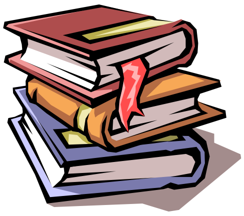
5: ESTIMACIÓN DE PARÁMETROS

Estamos en plena inferencia estadística, es decir, trabajaremos con la información muestral para estimar los valores de los parámetros poblacionales.



Actividad bibliográfica

1. Lea en las páginas 238 a 275, los apartados 9.1 al 9.5 inclusive, el apartado 9.7 y los apartados 9.9 al 9.12 inclusive del libro *Probabilidad y Estadística para Ingenieros* de Walpole, Myers y Myers.
   Tenga en cuenta las siguientes recomendaciones al estudiar este material:
   - Debe leer todo el capítulo con excepción de los apartados:
     - 9.6: Límites de tolerancia.
     - 9.8: Observaciones pareadas.
     - 9.13: Métodos bayesianos de estimación.
     - 9.14: Estimación de probabilidad máxima.
   - Para estudiar, comprender y resolver problemas de estimación debe considerar la distribución muestral del estadístico que permitirá construir un intervalo de confianza, para lo cual se debe reflexionar sobre los siguientes aspectos:
     - <u>Distribución de la población</u> de la cual proviene la muestra en estudio (normales, no normales o desconocidas).
     - Conocimiento de <u>otros parámetros poblacionales</u>. Por ejemplo, si se quiere estimar un intervalo de confianza para la media de la población, se debe tener en cuenta si se conoce o no la desviación estándar de la población en estudio.
     - <u>Tamaño de la muestra</u> seleccionada (muestras pequeñas o grandes).
   - El texto no hace hincapié en la forma en que se deben **interpretar** los intervalos de confianza, sólo se hace referencia en la página 257, en el

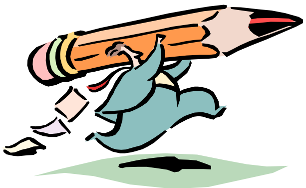
apartado *Interpretación del intervalo de confianza.* En el presente material encontrará ejercicios resueltos donde verá cómo debe hacer la interpretación adecuadamente. Además, encontrará una sección *Para pensar* que lo hará reflexionar al respecto.

- En distintos lugares del texto se mencionan las tablas A.3, A.4, etcétera. Usted ya sabe que debe trabajar con las tablas provistas por la cátedra y no con las que vienen en cada texto utilizado.
- Página 240: No es necesario que sepa la demostración dada en el ejemplo 9.1.
- Página 241: El primer párrafo de esta página: *Aunque S^2 es … se estima la varianza*, ¡es muy importante! Recuérdelo para entender otros temas.
- Página 244: Lea con atención el párrafo siguiente al cuadro.
- Página 246: Lea con atención los dos párrafos siguientes al Teorema 9.2.
- Páginas 247 y 248: Lea con atención los dos párrafos siguientes al cuadro de la página 247 y el primero de la página 248.
- Página 253: En el cuadro, en la expresión del límite inferior del intervalo, dentro de la raíz, la segunda varianza debe tener como subíndice 2 en lugar de 1.
- Página 254: Lea con atención el primer párrafo de la página.
- Página 254: Preste atención a la necesidad de *independencia* de las muestras, descripta en el segundo párrafo de la página.
- Página 254: Lea con atención el párrafo siguiente a la solución del ejemplo 9.6 (es el último párrafo de la página).
- Página 265: Preste atención a lo indicado en el párrafo posterior al cuadro.
- Página 266: Lea con mucha atención todos los conceptos volcados en esta página.
- Página 269: En el recuadro correspondiente al "Intervalo de confianza para p_1-p_2 de una muestra grande", en el último término de la desigualdad aparece + - $z_{\alpha/2}$, cuando lo correcto es + $z_{\alpha/2}$.
- Página 271: Preste atención a la población de la cual proviene la muestra.
- Página 273: En el último renglón aparece, en el argumento de la función probabilidad, la expresión $\frac{S_1^2}{S_1^2}$*, pero debe decir* $\frac{S_1^2}{S_2^2}$.

Estimación de parámetros

Para pensar

1. En la página 248 del libro de referencia para este capítulo, el ejemplo 9.4 concluye dando sólo un resultado, sin realizar interpretación de los valores obtenidos. Sabemos que no esperamos esto de un ingeniero, sino que deseamos que realice el cálculo para poder dar una interpretación y tomar decisiones en función de los resultados obtenidos, entonces le preguntamos: ¿Es correcto dar la siguiente conclusión: *Hay un 95% de confianza de que el contenido medio de los contenedores de ácido sulfúrico esté entre 9,74 y 10,26 litros*? ¿Por qué?



A trabajar solos...

Le proponemos que resuelva solo algunos ejercicios. Los desarrollos los encontrará al final del capítulo. Pero antes de resolverlos...

Recuerde que…

- En cada ejercicio debe:
  - Definir la o las variables en estudio.
  - Describir la distribución de la o las variables en estudio.
  - Plantear la solución del problema con la justificación correspondiente.
  - Efectuar los cálculos.
  - Interpretar los resultados.
- En las evaluaciones, a cada uno de estos ítems se les asignará un puntaje. Sumados darán el puntaje correspondiente a cada problema.
- Cuando calcule **grados de libertad**, que se indica con ν, si el valor es un número decimal, redondee al entero más próximo, según el redondeo matemático tradicional.
- Cuando calcule el **tamaño de muestra**, que se indica con **n**, si el valor es un número decimal, debe redondear *siempre* el entero inmediato superior.

Por ejemplo, si queda n = 45,000001, el valor a utilizar será n = 46; si queda n = 24,986, el valor a utilizar será n = 25.

- Hemos aprendido qué significa la notación z_α, t_α, etcétera, pero en la resolución de los ejercicios, a veces usaremos la notación z_c o t_c para indicar los valores críticos. Esta notación se usará cuando sea evidente a qué área corresponde.

¡Ahora sí, **a trabajar solos**!

1. Se estudia la peso de cierto producto que se vende en bolsas. Es aceptable suponer que los pesos están distribuidos de manera normal, y se conoce de experiencias previas que la desviación estándar de la población es de 2,6 g. Los resultados obtenidos de la muestra ensayada, en gramos, son los siguientes: 452, 450, 454, 450, 456, 450, 456, 450, 452, 450.

a) Estime mediante un intervalo de confianza del 95% el peso medio de las bolsas de la población.

b) ¿Qué podemos asegurar con una probabilidad de 0,95 sobre la medida del error máximo de la estimación?

c) ¿Qué tamaño debería tener la muestra para poder tener una confianza del 99% que el error máximo de la estimación sea de 1 gramo?

d) Grafique el error máximo probable de la media para un nivel de confianza del 95%, haciendo variar el tamaño de la muestra entre 10 y 155 unidades experimentales. Saque conclusiones.

2. Suponga que se quiere determinar la cantidad correcta de pintura que contienen los envases de 1 litro compradas a un conocido fabricante local. Por las especificaciones del productor se sabe que la desviación estándar de la cantidad de pintura es igual a 0,025 litros. Al seleccionar una muestra de 49 latas de las mismas características, el contenido promedio de pintura resultó ser igual a 0,996 litros.

a) ¿Es aceptable suponer que el contenido medio de pintura por lata tendría una distribución normal?

b) Estime el contenido real de la población de pintura en latas de un litro, mediante un intervalo de confianza del 99%.

c) Con base en estos resultados ¿es posible asegurar que el productor envasa menos de un litro en tales latas? ¿Por qué?

d) ¿Con qué grado de confianza podrá decirse que el contenido medio de pintura es de $0,996 \pm 0,005$?

3. Se llevan a cabo pruebas de resistencia a la tracción sobre dos tipos de cables. De experiencias previas se sabe que la desviación estándar de las resistencias para cada uno de los cables son conocidas y que es posible suponer

Estimación de parámetros

que las resistencias a la tracción de ambas poblaciones son normales e independientes. Las pruebas realizadas aportaron los siguientes datos:

Cable tipo	Tamaño de la muestra	Resistencia media muestral (en kg/mm^2)	Desviación estándar de la población (en kg/mm^2)
1	10	87,6	1,0
2	12	74,5	1,5

a) Construya un intervalo de confianza del 90% para la verdadera diferencia de la resistencia a la tracción de los cables. Saque conclusiones.

b) Si se adoptan tamaños de muestras iguales, determine el tamaño requerido de la muestra de modo que se tenga una confianza del 90% de que el error de la estimación sea menor que 0,50 kg.

4. Un fabricante de turbinas hidráulicas considera al tiempo que transcurre hasta que es necesario reparar la máquina por problemas de cavitación en el rotor de la misma como una variable aleatoria con media 5000 horas y desviación estándar de 40 horas. A este tiempo lo llama vida eficaz y acepta que la distribución de la vida eficaz es muy próxima a la normal.

El fabricante introduce una mejora en el proceso de fabricación del rotor que aumenta el tiempo de la vida eficaz promedio a 5050 horas y disminuye la desviación estándar a 30 horas.

Para comprobar lo que afirma el fabricante, se toma del proceso "antiguo" una muestra aleatoria de tamaño 16, de la que se obtiene una media muestral de 4998 horas y una desviación estándar de 41 horas; de la muestra aleatoria del proceso "mejorado" de tamaño 25, se obtiene una media de 5052 horas y una desviación estándar de 29 horas. Construya un intervalo de confianza del 95% para la verdadera diferencia de la vida eficaz, si los procesos "antiguo" y "mejorado" pueden considerarse como poblaciones independientes.

5. Se desea vender esferas para utilizar en un mecanismo de precisión. Las especificaciones del producto, en cuanto al diámetro de tales esferas, indican que no deben superar las 4,40 pulgadas. De 9 mediciones realizadas se tienen los siguientes datos:

Muestra	1	2	3	4	5	6	7	8	9
Diámetro (pulgadas)	4,32	4,39	4,40	4,44	4,34	4,36	4,37	4,42	4,38

Suponiendo normalidad:

a) Halle un intervalo de confianza del 99% para el diámetro medio de las esferas.

b) ¿Qué se podrá decir, con una confianza del 99%, sobre la magnitud posible del error si se utiliza a la media de la muestra de estas 9 observaciones como una estimación de la media de la población?

c) Sin variar la magnitud del error máximo probable, ¿cuál sería el tamaño de la muestra para un nivel de confianza del 90%?

d) Construya un intervalo con un nivel de confianza del 95% para la varianza poblacional.

6. Aceptando que la distribución del consumo anual de energía eléctrica del tipo residencial de los clientes de una empresa distribuidora del servicio es aproximadamente normal, y que la medición de 31 clientes seleccionados aleatoriamente dio los siguientes resultados:

Consumo medio anual, en miles de MWh/año, de clientes con tarifa residencial

490,9	435,3	580,6	510,1	600,4	495,1	520,2	501,5
483,2	520,9	460,1	500,8	490,7	490,7	499,6	498,3
510,6	473,4	474,9	490,1	410,2	488,6	500,5	505,4
491,0	421,8	490,6	439,9	506,4	490,2	475,5	

a) Estime mediante un intervalo de confianza del 95%, el consumo medio anual de los usuarios residenciales del servicio.

b) Estime mediante un intervalo de confianza del 99%, la varianza del consumo medio anual de los clientes con tarifa residencial.

7. Los resultados dados por una A.R.T. sobre el análisis del peso de los empleados que trabajan en las oficinas y de los empleados que trabajan en la planta de producción. Al tomar una muestra del peso de diez empleados que trabajan en las oficinas, se encontró que el peso promedio es 90 kg, con una desviación estándar muestral de 5 kg. Los resultados obtenidos en la muestra de quince empleados que trabajan en la planta de producción fueron 87 y 4 kg, para el peso promedio y la desviación estándar muestral, respectivamente. Suponiendo que el peso de los empleados está distribuido de manera normal, encuentre un intervalo de confianza del 95% para la diferencia entre las medias de las poblaciones de los dos grupos de empleados. Por otra parte, supóngase que las dos poblaciones normales tienen la misma desviación estándar. Saque conclusiones.

8. La pintura para señalamiento vial de rutas y autopistas se distribuye en dos colores: blanco y amarillo. El interés se centra en el tiempo de secado de la pintura. Se sospecha que la pintura de color amarillo se seca más rápidamente que la blanca. Se obtienen las siguientes mediciones del tiempo de secado de ambos tipos de pintura, en minutos:

Blanca	120	132	123	122	140	110	120	107		
Amarilla	126	124	116	125	109	130	125	117	129	130

a) Encuentre un intervalo de confianza del 99% para la diferencia entre los tiempos de secado promedio, suponiendo que las desviaciones estándar de éstos son iguales. Suponga que el tiempo de secado está distribuido de manera normal.

b) ¿Existe alguna evidencia que indique que la pintura amarilla se seca más rápidamente que la blanca?

9. Un industrial está interesado en la uniformidad de la máquina que utiliza para dosificar áridos. De manera específica, es deseable que la desviación estándar σ del proceso de dosificación sea menor de 0,5 kg; de otro modo existe un porcentaje mayor del deseable de dosificaciones incorrectas. Supóngase que la distribución del peso del árido dosificado es aproximadamente normal. Al tomar una muestra aleatoria de 20 dosificaciones realizadas, se obtiene una varianza muestral de 0,19 kg^2.

a) Determine el intervalo de confianza "superior" del 95% para la varianza poblacional.

b) Con un nivel de confianza del 95%, ¿apoyan los datos la condición de que la desviación estándar del proceso es menor que 0,5 kg?

10. Con el objeto de investigar el tiempo de secado de un nuevo tipo de pintura, se prepararon cuatro paneles de ensayo y se midió el tiempo de secado en los mismos, obteniéndose los siguientes resultados:

Panel	1	2	3	4
Tiempo	1h 54min	2h 2min	2h 5min	1h 55min

Con los datos de la muestra y suponiendo normalidad:

a) Construya un intervalo de confianza del 95% para la varianza del tiempo medio de secado real de la pintura.

b) Construya un intervalo de confianza del 95% para el tiempo medio de secado real de la pintura.

c) ¿Qué tamaño de muestra seleccionaría si, a nivel del 99%, es necesario limitar el error máximo probable al valor obtenido en el punto anterior?

11. Se debe efectuar el alumbrado en un predio municipal. Luego de analizar las propuestas de los distintos oferentes se vio la conveniencia de usar lámparas de dos fábricas del medio.

a) A los efectos de controlar la calidad de las lámparas, se realizaron pruebas de calidad, obteniéndose los siguientes resultados:

- Una muestra de 21 lámparas del fabricante A dio una vida media de 1400 horas y una desviación estándar de 120 horas.
- Una muestra de 25 lámparas del fabricante B dio una vida media de 1200 horas y una desviación estándar de 80 horas.

Se desea saber si es aceptable suponer que hay homogeneidad en la calidad de las lámparas empleadas en la obra, aunque hayan sido producidas por dos fábricas distintas, cada una con sus propios equipos, condiciones de elaboración y supervisión. Se acepta que el tiempo de vida de las lámparas de ambos proveedores está distribuido normalmente y son independientes.

b) Si con el mismo tamaño de muestras se hubieran obtenido las mismas vidas medias muestrales, pero con desviaciones estándares de 120 y 70 horas, respectivamente, ¿cuál sería la conclusión?

12. Una compañía de transporte debe decidir sobre la compra de neumáticos de tipo A o B. Para estimar la diferencia, se realiza una prueba empleando 16 neumáticos de cada marca. Los neumáticos se hacen trabajar hasta el desgaste total.

Los resultados son los mostrados en el cuadro siguiente:

Marca	Tamaño de la muestra	Recorrido medio (en km)	Desviación estándar (en km)
A	16	36.300	5.000
B	16	38.100	6.100

a) Construya un intervalo con un nivel de confianza del 90% para la diferencia de medias, suponiendo que las poblaciones tienen distribución normal y son independientes.

b) Determine un intervalo de confianza del 90%, para el cociente de las varianzas poblacionales. ¿Se justifica el supuesto de que las varianzas son iguales para calcular el intervalo de confianza para la diferencia de las medias?

13. El ingeniero a cargo del control de calidad de una planta debe detener el proceso de producción cuando se produce una rotura de los sacos en los que se envasa el producto que afecta una proporción superior a 0,008. Al realizar el control de la línea de bolsas de cemento portland de alta resistencia inicial observa que se rompieron 17 bolsas sobre un total de 1000, antes de salir de la planta.

a) ¿Debe detener el proceso de producción? Fundamente la respuesta construyendo un intervalo de confianza bilateral del 95% para la proporción real de sacos rotos.

b) ¿Qué tamaño debe tener la muestra si se desea, con nivel de confianza del 95%, que la proporción de defectuosas esté dentro del 0,004 de la proporción verdadera de defectuosas? Compare los resultados obtenidos con y sin estimación previa de la proporción real.

c) ¿A qué conclusión llega si construye un intervalo de confianza unilateral para los datos del punto a?

d) Suponga que se realizan ajustes en el proceso de producción y que en una segunda muestra aleatoria de 500 sacos envasados se rompen 4. Con base en los datos muestrales, construya un intervalo de confianza bilateral para la diferencia de proporciones correspondientes (antes y después de los cambios), para emitir conclusiones a nivel de confianza del 95%.

14. En una muestra aleatoria de 500 clientes de una empresa local se encontró que 340 se habían suscrito al servicio de correo electrónico y 160 al de correo electrónico e Internet.

a) Construya un intervalo de confianza del 95% para la proporción actual de clientes suscriptos al correo electrónico de esa empresa de servicios.

b) ¿Qué puede decir acerca del error cometido en la estimación, con un 95% de confianza?

c) A partir de los resultados obtenidos de la muestra preliminar ¿qué tan grande se requiere que sea la muestra, si se desea tener una confianza del 95% de que la estimación de p estará dentro de 0,02?

d) Asumiendo que no se dispone de una muestra preliminar para obtener información acerca de *p*, ¿qué tan grande se requiere que sea la muestra si se quiere tener una confianza al menos del 95% de que la estimación de *p* estará dentro de 0,02?

15. Se realiza un estudio para determinar la proporción de residentes en una ciudad y en sus suburbios que están a favor de la construcción de una planta de energía nuclear.

a) ¿Qué tan grande debe ser una muestra si se requiere una confianza al menos del 95% de que la estimación estará dentro del 0,04 de la proporción real de residentes de esa ciudad y sus suburbios que están a favor de la construcción de la planta de energía nuclear?

b) A falta de información previa, se tomó una muestra preliminar de tamaño 40 (observe que la muestra es mayor que treinta), resultando que 6 residentes estuvieron a favor de la construcción. Utilice esta estimación imperfecta para determinar en forma aproximada cuántas observaciones se necesitan para proporcionar el grado de precisión deseado en el punto anterior.

16. Se analiza la fracción de productos defectuosos producidos por dos líneas de producción. Una muestra aleatoria de 100 unidades provenientes de la línea 1 contiene 10 defectuosas, mientras que una muestra aleatoria de 120 unidades de la línea 2 tiene 25 que son defectuosas. Encuentre un intervalo de confianza del 99% para la diferencia en fracciones de productos defectuosos producidos por las dos líneas.

17. Se estudia la trabajabilidad de un hormigón plástico compactado mediante vibración normal. De acuerdo a las especificaciones del Reglamento, el mismo debe encuadrarse en un ámbito de consistencia A-2, de aspecto levemente cohesivo. Para ello se ajusta el proyecto del hormigón de modo de obtener un asentamiento del tronco cono de 8 cm, y de experiencias anteriores se sabe que la precisión del equipo de trabajo permite lograr una variabilidad de los resultados cuantificada por una desviación estándar de 1,5 cm. Se acepta también que el asentamiento medido está distribuido normalmente. Los resultados obtenidos de una muestra ensayada, medidos en cm, son los siguientes: 7,5; 8,0; 8,0; 7,5; 7,0; 8,5; 8,0; 9,0; 7,0; 8,0

a) Estime mediante un intervalo de confianza del 95% el asentamiento medio del hormigón estudiado.

b) ¿Qué podemos asegurar con una probabilidad de 0,95 sobre la medida del error máximo de la estimación?

c) ¿Qué tamaño debería tener la muestra para poder tener una confianza del 99% que el error máximo de la estimación sea de $\frac{1}{2}$ cm?

18. Se realizó un estudio para determinar si un aditivo agregado en la dosificación de hormigones para acelerar el tiempo de fragüe tenía alguna influencia en la resistencia a la compresión del mismo. Una muestra aleatoria de 62 probetas con aditivo ensayadas a los 7 días dan una resistencia media de 70,5 kg/cm^2, con desviación 6,5 kg/cm^2, en tanto que las 60 probetas sin aditivo, ensayadas a la misma edad, dieron una resistencia media de 86,8 kg/cm^2, con desviación 7,2 kg/cm^2. Suponga que de estudios anteriores se conocen las desviaciones poblacionales y son 5 y 5,5 kg/cm^2 para las probetas con y sin aditivo, respectivamente.

a) A nivel de confianza del 95%, ¿qué puede decir respecto de la influencia del aditivo en la resistencia del hormigón?

b) ¿Qué conclusión obtendría si desconoce las desviaciones poblacionales pero las supone iguales?

c) ¿Qué conclusión obtendría si desconoce las desviaciones poblacionales pero las supone distintas?

19. Se compara la resistencia de dos tipos de rosca para bulones utilizados en construcciones metálicas. Se prueban 20 piezas de las roscas tipo A y 18 de las del tipo B, en condiciones similares. Las piezas del tipo A dan una resistencia media de 78,3 kg, con desviación estándar 5,6 kg. Las del tipo B dan una resistencia media de 76,5 kg, con desviación estándar 6,3 kg. Se conoce de experiencias anteriores que las desviaciones poblacionales son 5 y 5,5 kg, para las roscas del tipo A y B, respectivamente, y que provienen de poblaciones normales.

a) ¿Hay suficientes razones para concluir a nivel del 95% que la resistencia media de las roscas del tipo A es más elevada que las del tipo B?

b) ¿Cuál es la conclusión, si no conoce los parámetros poblacionales?

20. El artículo "Mix Design for Optimal Strength Development of Fly Ash Concrete" *(Cement and Concrete Research, 1989, Vol. 19, Nº 4, pp. 634-640)* investiga la resistencia a la compresión del hormigón cuando se mezcla con ceniza muy fina (una mezcla de silicatos, alúmina, hierro, óxido de magnesio y otros componentes). Se sabe que la resistencia se comporta de manera aproximadamente normal.La resistencia a la compresión, en MPa, de nueve muestras de hormigón en condiciones secas, a la edad de 28 días, son las siguientes: 40,2; 30,4; 28,9; 30,5; 22,4; 25,8; 18,4; 14,2; 15,3

a) Encuentre un intervalo de confianza inferior del 99% para la resistencia a la compresión promedio.

b) Encuentre un intervalo de confianza bilateral del 95% para la resistencia a la compresión promedio.

21. Se investiga el diámetro de las barras de acero del tipo F-22 utilizadas en construcciones livianas de acero de nuestro medio, fabricadas en dos diferentes máquinas de extrusión. Para ello se toman dos muestras aleatorias.
De la primera máquina los resultados obtenidos de 15 mediciones dan un diámetro medio de 8,73 mm con varianza 0,35 mm^2, mientras que para la segunda máquina, de 18 mediciones el diámetro medio y la varianza son: 8,68 mm y 0,40 mm^2, respectivamente. Suponiendo que los diámetros de las poblaciones de barras producidos por ambas máquinas están distribuidos normalmente y que es posible aceptar que tienen la misma varianza, construya un intervalo de confianza bilateral del 95% para la diferencia en el diámetro promedio de la barra.

22. Un artículo publicado en Fire Technology investigó dos agentes dispersores de espuma que pueden emplearse en las boquillas de los equipos extinguidores de fuego. Al tomar una muestra aleatoria de cinco observaciones con una espuma que forma una película acuosa (AFFF), se obtuvo una media muestral de 4,7 y una desviación estándar de 0,6. Una muestra aleatoria de cinco observaciones con concentrados de tipo alcohólico (ATC) tuvo una media muestral de 6,9 y una desviación estándar de 0,8. Encuentre un intervalo de confianza del 95% para la diferencia en la dispersión de espuma promedio de estos dos agentes. ¿Puede obtenerse alguna conclusión sobre qué agente produce la mayor dispersión de espuma? Suponga que ambas poblaciones están bien representadas por distribuciones normales que tienen las mismas desviaciones estándar.

23. Se estudia la trabajabilidad de un hormigón plástico compactado mediante vibración normal. De acuerdo a las especificaciones del Reglamento, el mismo debe encuadrarse en un ámbito de consistencia A-2, de aspecto levemente cohesivo. Para ello se ajusta el proyecto del hormigón de modo de obtener un asentamiento del tronco cono de 8 cm, y de experiencias se acepta que el asentamiento medido está distribuido normalmente. Los resultados obtenidos de una muestra ensayada, medidos en cm, son los siguientes: 7,5; 8,0; 8,0; 7,5; 7,0; 8,5; 8,0; 9,0; 7,0; 8,0. Estime mediante un intervalo de confianza del 99%, la desviación estándar y la varianza del asentamiento del hormigón estudiado.

24. En un experimento se comparó el consumo de combustible para dos tipos de camiones con motor diesel, en condiciones similares. Doce vehículos del Tipo A y diez del Tipo B se probaron a 90 km/h. Si los doce camiones promediaron 16 km/l con desviación 1 km/l y los diez camiones consumieron 11 km/l con desviación 1,8 km/l.
Suponiendo normalidad:

a) Construya un intervalo de confianza del 98% para el cociente de las varianzas de los recorridos por litro de los vehículos tipo A y B. ¿Qué conclusión obtiene?

b) Determine un intervalo de confianza del 90% para la diferencia entre los rendimientos medios de cada tipo de camión, suponiendo que los rendimientos tienen varianzas iguales. ¿Qué conclusión obtiene?

25. En el laboratorio de ensayo se ha observado que las probetas de hormigón están siendo moldeadas deficientemente, hecho este que está afectando al resultado de ensayo obtenido.

a) Calcule un intervalo de confianza del 96% para la proporción de probetas de hormigón moldeadas defectuosamente, cuando se halla que en una muestra de 100 probetas, 8 son defectuosas.

b) ¿Qué tamaño debe tener la muestra si se desea, con nivel de confianza del 96%, que la proporción de defectuosas esté dentro del 0,05 de la proporción verdadera de defectuosas?

c) Grafique la variación del tamaño de la muestra, cuando el error varía entre 0,01 y 0,10, a nivel del 96%, utilizando 0,08 como estimación de la proporción verdadera.

26. Un artículo publicado en *Engineering Horizons* informa que 117 de 484 egresados de Ingeniería planeaban continuar estudiando para obtener un grado más avanzado. Si se considera esto como una muestra aleatoria de todos los graduados en 1990, encuentre un intervalo de confianza del 90% para la proporción de graduados que planean continuar con su educación.

27. En una muestra aleatoria de 1000 viviendas en una ciudad determinada, se encuentra que 228 de ellas utilizan calefacción en base a derivados del petróleo.

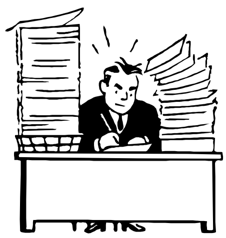
a) Encuentre el intervalo de confianza del 99% para la proporción de viviendas en esa ciudad que tiene ese tipo de calefacción.

b) ¿Qué tan grande debe ser la muestra si se requiere una confianza al menos del 99% de que la estimación estará dentro del 0,025 de la proporción real de viviendas en esa ciudad que tiene ese tipo de calefacción?

28. Según el Environment News (abril de 1975), "el análisis continuo de los niveles de plomo en el agua potable de varias comunidades de Boston reveló nive-

les elevados de plomo en los suministros de agua de Somerville, Brighton y Beacon Hill …". Los resultados preliminares de un estudio efectuado en 1974 indicaron que "el 20% de 248 hogares que se analizaron en estas comunidades reveló niveles elevados que exceden el estándar de la Agencia de Salud Pública de EE.UU. de 50 p.p.m". Al contrario, en Cambridge, que añade corrosivos al agua, "solamente el 5% de los 110 hogares analizados mostró niveles de plomo mayores que el estándar".
Obtenga un intervalo de confianza del 95% para la diferencia de proporciones de hogares que tienen niveles de plomo que exceden el estándar entre las comunidades de Somerville, Brighton y Beacon Hill, y la comunidad de Cambridge.

¡A repasar...!

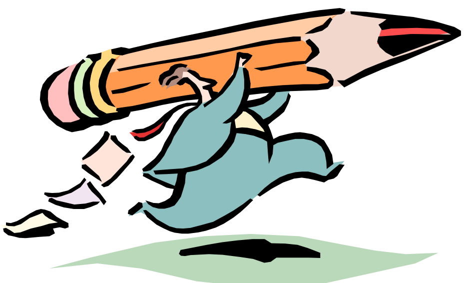


Este capítulo es muy importante para su futuro trabajo como ingeniero. Es preciso que lo domine antes de comenzar las aplicaciones prácticas y las autoevaluaciones.

Para autoevaluarse, responda las preguntas que están a continuación. Puede hacerlo con el material de estudio, pero asegurándose que "entiende" cada palabra, a tal punto que usted podría explicarle a un amigo, que no conoce el tema, de manera simple, los conceptos estudiados:

- ☑ **¿Qué trata la inferencia estadística?**
- ☑ **¿Cuáles son los métodos clásicos de estimación?**
- ☑ **¿Qué se espera de un buen estimador?**
- ☑ **¿Qué significa que un estimador sea insesgado?**
- ☑ **¿Qué significa que un estimador sea el más eficiente?**
- ☑ **¿Todos los estimadores que son más eficientes son insesgados?**
- ☑ **¿Qué es una estimación puntual?**
- ☑ **¿Qué es una estimación por intervalos?**
- ☑ **Respecto a los tipos de estimación analizados, ¿cuál es más precisa?, ¿cuál es más confiable?**
- ☑ **¿Qué es el error estándar de una estimación puntual?**

- ☑ **¿Cómo se calcula un intervalo de confianza para estimar la media poblacional? Analice cada una de las posibilidades y las condiciones que deben darse para utilizarlas correctamente.**
- ☑ **¿Cómo se calcula un intervalo de confianza para estimar la diferencia de medias poblacionales? Analice cada una de las posibilidades y las condiciones que deben darse para utilizarlas correctamente.**
- ☑ **¿Cómo se calcula un intervalo de confianza para estimar la proporción poblacional? Analice las condiciones que deben darse para utilizarlo correctamente.**
- ☑ **¿Cómo se calcula un intervalo de confianza para estimar la diferencia de proporciones poblacionales? Analice las condiciones que deben darse para utilizarlo correctamente.**
- ☑ **¿Cómo se calcula un intervalo de confianza para estimar la varianza poblacional? Analice las condiciones que deben darse para utilizarlo correctamente.**
- ☑ **¿Cómo se calcula un intervalo de confianza para estimar el cociente de varianzas poblacionales? Analice las condiciones que deben darse para utilizarlo correctamente.**
- ☑ **Analice las condiciones para calcular tamaños de muestra según la información disponible.**



Por favor, no avance al siguiente tema si tiene dudas o no recuerda las nociones aquí volcadas. Pero si se siente listo para continuar, es hora de empezar a trabajar con las **autoevaluaciones** y las **aplicaciones prácticas**...

Estimación de parámetros

Para pensar

1. En la página 248 del libro de referencia para este capítulo, el ejemplo 9.4 concluye dando sólo un resultado, sin realizar interpretación de los valores obtenidos. Sabemos que no esperamos esto de un ingeniero, sino que deseamos que realice el cálculo para poder dar una interpretación y tomar decisiones en función de los resultados obtenidos, entonces le preguntamos: ¿Es correcto dar la siguiente conclusión: *Hay un 95% de confianza de que el contenido medio de los contenedores de ácido sulfúrico esté entre 9,74 y 10,26 litros*? ¿Por qué?

   *Es **incorrecta** la conclusión dada. Pensemos un poquito...*
   Para obtener el intervalo tomamos una muestra aleatoria, calculamos la media y la varianza muestrales, y aplicando la expresión para el cálculo de intervalos de confianza para la media con varianza desconocida, tamaño de muestra pequeño y con una variable que se distribuye normalmente, obtenemos un resultado.
   Entonces pensemos un poquito más...
   Supongamos que el verdadero valor para la media poblacional fuera de 10,2 litros... ¿Usted tendría un 95% de confianza de que el valor de la media poblacional, que hemos supuesto de 10,2 litros, está entre 9,74 y 10,26 litros? Más claramente: ¿Hay un 95% de confianza de que el valor 10,2 esté entre 9,74 y 10,26?... ¡Nooooo! ¡Seguro que el valor 10,2 está entre 9,74 y 10,26! ¡Es matemática pura!
   Ahora supongamos que el verdadero valor para la media poblacional fuera de 9,7 litros... ¿Usted tendría un 95% de confianza de que el valor de la media poblacional, que hemos supuesto de 9,7 litros, está entre 9,74 y 10,26 litros? Más claramente: ¿Hay un 95% de confianza de que el valor 9,7 esté entre 9,74 y 10,26?... ¡Nooooo! ¡Es imposible que el valor 9,7 está entre 9,74 y 10,26! ¡Es matemática pura!
   Entonces... ¿Qué significa la expresión $9,74 < \mu < 10,26$?
   Vaya a la página 244 del libro y vea el esquema (Figura 9.3) dado para la interpretación de las estimaciones por intervalo de μ para muestras diferentes.
   Nosotros tomamos una muestra en particular, pero podríamos haber tomado otra que arrojara distintos valores para la media y la varianza muestrales, por lo que obtendríamos distintos intervalos que, como indica el esquema, podrían o no contener al verdadero valor de la media poblacional μ.
   Al trabajar con un 95% de confianza, estamos presumiendo que de todos los intervalos que podríamos calcular en función de todas la posibles muestras, el 95% de ellos contendrían realmente al valor de la media poblacional y habría un 5% que no lo contendría.

¡¡¡Y entonces... ¿Qué significa la expresión 9,74 < μ < 10,26?!!!

Según la evidencia muestral (de acuerdo a los resultados obtenidos en la muestra que seleccioné aleatoriamente), tenemos un 95% de confianza que el intervalo (9,74 ; 10,26) contenga al verdadero valor de la media poblacional μ. Es decir, tenemos un 95% de confianza de que la muestra seleccionada sea una de las que me permitían obtener un intervalo que contenga a la verdadera media poblacional.

Ejercicios resueltos

1. Se estudia la peso de cierto producto que se vende en bolsas. Es aceptable suponer que los pesos están distribuidos de manera normal, y se conoce de experiencias previas que la desviación estándar de la población es de 2,6 g. Los resultados obtenidos de la muestra ensayada, en gramos, son los siguientes: 452, 450, 454, 450, 456, 450, 456, 450, 452, 450.

a) Estime mediante un intervalo de confianza del 95% el peso medio de las bolsas de la población.

Si se selecciona una muestra de una población normal, o a falta de esto, si n es suficientemente grande, se puede establecer un intervalo de confianza de μ considerando la distribución muestral de $\overline{X}$, que según el teorema central del límite, es de esperarse que sea aproximadamente normal con media $\mu_{\overline{x}} = \mu$ y desviación estándar $\sigma_{\overline{x}} = \sigma / \sqrt{n}$.

Por lo tanto, se puede afirmar con una probabilidad de 1-α, que la variable normal estándar, $Z = \dfrac{\overline{x} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$ estará entre - $z_{\alpha/2}$ y $z_{\alpha/2}$.

A partir de esto podemos decir que:

Si $\overline{x}$ es la media de una muestra aleatoria de tamaño n de una población con varianza conocida σ^2, el intervalo de confianza del (1-α)100% para μ es:

$$\left(\overline{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \; ; \; \overline{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right)$$

donde $z_{\alpha/2}$ es el valor de z a la derecha del cual se tiene un área de $\alpha/2$.

En nuestro problema:
X:"Peso de las bolsas"

$X \sim$ Normal (μ ; σ = 2,6g)
$\overline{X} \sim$ Normal ($\mu_{\overline{x}} = \mu$; $\sigma_{\overline{x}} = \sigma/\sqrt{n}$) Por provenir la muestra de una distribución normal

Según los siguientes datos del problema:
$\overline{x}$ = 452 g
σ = 2,6 g
n = 10
$1-\alpha = 0{,}95 \Rightarrow z_{\alpha/2} = \pm 1{,}96$

Los límites de confianza del 95% son:

$$\overline{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 452\text{ g} \pm 1{,}96 \cdot \frac{2{,}6\text{ g}}{\sqrt{10}} = \mathbf{452\text{ g} \pm 1{,}6115\text{ g}}$$

Luego,

$$\mathbf{450{,}3885\text{ g} < \mu < 453{,}6115\text{ g} \qquad 1 - \alpha = 0{,}95}$$

🗣 *Se tiene un 95% de confianza, a partir de los datos de la muestra, que el intervalo (450,39 g ; 453,61 g) contiene a la verdadera media poblacional.*

b) ¿Qué podemos asegurar con una probabilidad de 0,95 sobre la medida del error máximo de la estimación?

> *Si se utiliza $\overline{x}$ como una estimación de μ, se puede tener una confianza del $(1-\alpha)100\%$ de que el error no excederá $z_{\alpha/2} \cdot \sigma/\sqrt{n}$.*

El máximo error de estimación con una confianza del $(1-\alpha).100\%$, según los siguientes datos del problema, es:

$$e = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 1{,}96 \cdot \frac{2{,}6\text{ g}}{\sqrt{10}} = \mathbf{1{,}6115\text{ g}}$$

🗣 *Se tiene un 95% de confianza de que el error máximo de la estimación es de 1,61 g, aproximadamente.*

c) ¿Qué tamaño debería tener la muestra para poder tener una confianza del 99% que el error máximo de la estimación sea de 1 gramo?

> *Si se utiliza $\overline{x}$ como una estimación de μ, se puede tener una confianza del $(1-\alpha)100\%$ de que el error no excederá una cantidad específica <u>e</u> cuando el tamaño de la muestra es*
>
> $$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{e}\right)^2$$

Estimación de parámetros

Según los siguientes datos del problema:

$\overline{x} = 452$ g

$\sigma = 2{,}6$ g

$e = 1$ g

$1-\alpha = 0{,}99 \Rightarrow z_{\alpha/2} = \pm\, 2{,}58$

Los límites de confianza del 99% son :

$\overline{x} \pm 2{,}58 \,.\, \dfrac{\sigma}{\sqrt{n}}$ siendo el error de la estimación $2{,}58 \,.\, \dfrac{\sigma}{\sqrt{n}}$.

Se desea que el error sea inferior a 1 g, o sea, $2{,}58 \,.\, \dfrac{2{,}6}{\sqrt{n}} < 1$

es decir, $\sqrt{n} > \dfrac{2{,}58 \,.\, 2{,}6}{1} = 6{,}708$ por lo que $n > 44{,}997$.

Luego, $\mathbf{n \geq 45}$

Así, pues, se puede tener una confianza del 99% de que el error de la estimación será menor que 1 gramo si se toma una muestra de tamaño 45 o mayor.

d) Grafique el error máximo probable de la media para un nivel de confianza del 95%, haciendo variar el tamaño de la muestra entre 10 y 155 unidades experimentales. Saque conclusiones.

$e = z_c \,.\, \dfrac{\sigma}{\sqrt{n}} = 1{,}96 \,.\, \dfrac{2{,}6}{\sqrt{n}}$

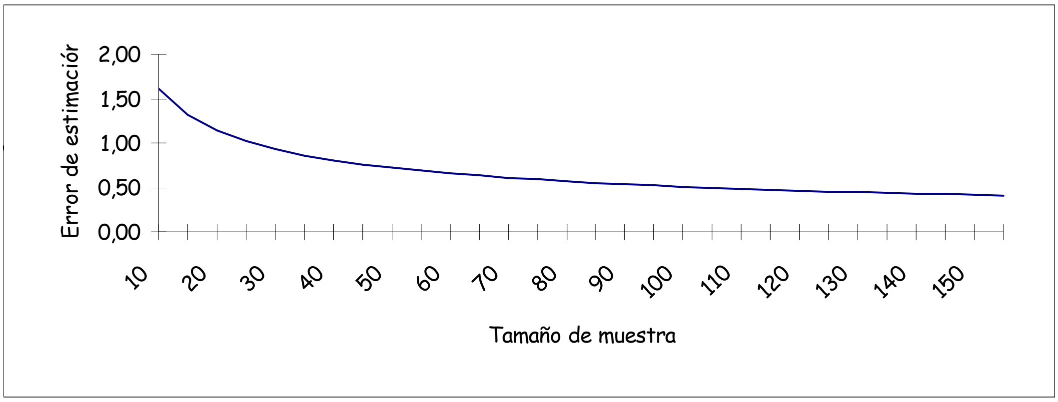
n	10	15	20	25	30	35
e	1,6115	1,3158	1,1395	1,0192	0,9304	0,8614

n	40	45	50	55	60	65
e	0,8057	0,7597	0,7207	0,6871	0,6579	0,6321

n	70	75	80	85	90	95
e	0,6091	0,5884	0,5698	0,5527	0,5372	0,5228

n	100	105	110	115	120	125
e	0,5096	0,4973	0,4859	0,4752	0,4652	0,4558

n	130	135	140	145	150	155
e	0,4469	0,4386	0,4307	0,4232	0,4161	0,4093



s
ión: Se observa que a medida que aumenta el tamaño de muestra, disminuye el error de estimación.

2. Suponga que se quiere determinar la cantidad correcta de pintura que contienen los envases de 1 litro compradas a un conocido fabricante local. Por las especificaciones del productor se sabe que la desviación estándar de la cantidad de pintura es igual a 0,025 litros. Al seleccionar una muestra de 49 latas de las mismas características, el contenido promedio de pintura resultó ser igual a 0,996 litros.

a) ¿Es aceptable suponer que el contenido medio de pintura por lata tendría una distribución normal?

Es aceptable suponer que el contenido medio de pintura por lata tendría una distribución normal porque la variable "media aritmética" se distribuye en forma normal y se conoce la varianza poblacional, además, se trabaja con una muestra grande y esto favorece las suposiciones de normalidad.

b) Estime el contenido real de la población de pintura en latas de un litro, mediante un intervalo de confianza del 99%.

X:"Cantidad de pintura en las latas"

$X \sim$ *Desconocida*

$\overline{X} \sim$ *Normal* $(\mu_{\overline{x}} = \mu\ ;\ \sigma_{\overline{x}} = \sigma/\sqrt{n}\ )$ *Por el teorema del límite central, ya que n es grande (n > 30)*

Según los siguientes datos del problema:

$\overline{x} = 0{,}996$ *litros*

$\sigma = 0{,}025$ *litros*

$n = 49$

$1-\alpha = 0{,}99 \Rightarrow z_{\alpha/2} = \pm\ 2{,}58$

Los límites de confianza del 99% son :

$$\overline{x} \pm z_c \cdot \frac{\sigma}{\sqrt{n}} = 0{,}996\ l \pm 2{,}58 \cdot \frac{0{,}025\ l}{\sqrt{49}} = \mathbf{0{,}996\ l \pm 0{,}003557\ l}$$

Luego,

0,9924 litros < μ < 0,9996 litros ***1 - α = 0,99***

🗣 *Se tiene un 99% de confianza, a partir de los datos de la muestra, de que el intervalo (0,9924 l ; 0,9996 l) contiene a la verdadera media poblacional.*

c) Con base en estos resultados ¿es posible asegurar que el productor envasa menos de un litro en tales latas? ¿Por qué?

Se puede tener un 99% de confianza, a partir de los datos de la muestra, que el productor envasa menos de un litro de pintura por lata, porque el valor 1 litro no está en el intervalo.

d) ¿Con qué grado de confianza podrá decirse que el contenido medio de pintura es de 0,996 $\pm$ 0,005?

$$z_c \cdot \frac{\sigma}{\sqrt{n}} = z_c \cdot \frac{0{,}025\ l}{\sqrt{49}} = 0{,}005 \text{ litros} \Rightarrow z_c = 1{,}4$$

siendo, $P(-1{,}4 \leq Z \leq 1{,}4) = \mathbf{0{,}83849}$

🗣 *Se tiene un 83,85% de confianza, a partir de los datos de la muestra, de que el intervalo (0, 991 l ; 1,001 l) contiene a la verdadera media poblacional.*

3. Se llevan a cabo pruebas de resistencia a la tracción sobre dos tipos de cables. De experiencias previas se sabe que la desviación estándar de las resistencias para cada uno de los cables son conocidas y que es posible suponer que las resistencias a la tracción de ambas poblaciones son normales e independientes. Las pruebas realizadas aportaron los siguientes datos:

Cable tipo	Tamaño de la muestra	Resistencia media muestral (en kg/mm^2)	Desviación estándar de la población (en kg/mm^2)
1	10	87,6	1,0
2	12	74,5	1,5

a) Construya un intervalo de confianza del 90% para la verdadera diferencia de la resistencia a la tracción de los cables. Saque conclusiones.

Si se tienen dos poblaciones con medias μ_1 y μ_2 y varianzas σ_1^2 y σ_2^2, respectivamente, el estimador puntual de la diferencia entre μ_1 y μ_2 lo da el estadístico $\overline{X}_1 - \overline{X}_2$. Por lo tanto, para obtener una estimación puntual de $\mu_1 - \mu_2$, se seleccionarán dos muestras aleatorias independientes, una de cada población, de tamaños n_1 y n_2, y se calculará la diferencia $\overline{x}_1 - \overline{x}_2$, de las medias muestrales. Por supuesto, se deben considerar las distribuciones muestrales de $\overline{X}_1 - \overline{X}_2$.

Se puede esperar que la distribución muestral de $\overline{X}_1 - \overline{X}_2$ está distribuida aproximadamente en forma normal con media $\mu_{\overline{x}_1-\overline{x}_2} = \mu_1 - \mu_2$ y desviación estándar $\sigma_{\overline{x}_1-\overline{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$.

Por lo tanto, se puede afirmar con una probabilidad de 1-α, que la variable normal estándar, $Z = \frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$ estará entre $-z_{\alpha/2}$ y $z_{\alpha/2}$.

A partir de esto podemos decir que:

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1-\alpha$$

$$P\left(-z_{\alpha/2} < \frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} < z_{\alpha/2}\right) = 1-\alpha$$

$$P\left((\overline{x}_1-\overline{x}_2) - z_{\alpha/2}\cdot\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} < \mu < (\overline{x}_1-\overline{x}_2) \pm z_{\alpha/2}\cdot\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right) = 1-\alpha$$

Si $\overline{x}_1$ y $\overline{x}_2$ son las medias de muestras aleatorias independientes de tamaño n_1 y n_2 de poblaciones con varianzas conocidas σ_1^2 y σ_2^2, respectivamente, un intervalo de confianza del (1-α)100% para μ_1-μ_2 es:

$$\left((\overline{x}_1-\overline{x}_2) - z_{\alpha/2}\cdot\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\ ;\ (\overline{x}_1-\overline{x}_2) + z_{\alpha/2}\cdot\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)$$

donde $z_{\alpha/2}$ es el valor de z a la derecha del cual se tiene un área de $\alpha/2$.

En nuestro problema:

X_1:"Resistencia a la tracción del cable 1"

X_2:"Resistencia a la tracción del cable 2"

$X_1 \sim$ *Normal* $(\mu_1\ ;\ \sigma_1 = 1{,}0\ kg/mm^2)$

$X_2 \sim$ *Normal* $(\mu_2\ ;\ \sigma_2 = 1{,}5\ kg/mm^2)$

$\overline{X}_1 - \overline{X}_2 \sim$ *Normal* $\left(\mu_{\overline{x}_1-\overline{x}_2} = \mu_1 - \mu_2\ ;\ \sigma_{\overline{x}_1-\overline{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)$ *Por provenir de muestras normales*

Según los siguientes datos del problema:

$\overline{x}_1 = 87{,}6\ kg/mm^2$	$\sigma_1 = 1{,}0\ kg/mm^2$	$n_1 = 10$
$\overline{x}_2 = 74{,}5\ kg/mm^2$	$\sigma_2 = 1{,}5\ kg/mm^2$	$n_2 = 12$

$1-\alpha = 0{,}90 \Rightarrow z_{\alpha/2} = \pm 1{,}645$

Los límites de confianza del 90% son:

$$(\bar{x}_1 - \bar{x}_2) \pm z_c \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = (87{,}6 - 74{,}5) \pm 1{,}645 \cdot \sqrt{\frac{1{,}0^2}{10} + \frac{1{,}5^2}{12}} =$$

$$= \mathbf{13{,}1 \pm 0{,}54\ kg/mm^2}$$

Luego,

$$\mathbf{12{,}56\ kg/mm^2 < \mu_1 - \mu_2 < 13{,}64\ kg/mm^2 \qquad 1 - \alpha = 0{,}90}$$

- *Se tiene un 90% de confianza, a partir de los datos de las muestras, que el intervalo (12,56 kg/mm^2 ; 3,64 kg/mm^2) contiene a la verdadera diferencia de medias poblacionales.*
- *Esto indica que hay una diferencia significativa entre las resistencias medias a la tracción de los dos tipos de cables, porque el valor 0 no pertenece al intervalo. Si el valor 0 está en el intervalo, nos indica que puede ocurrir que $\mu_1 = \mu_2$, es decir, no hay diferencia entre las medias poblacionales.*

b) Si se adoptan tamaños de muestras iguales, determine el tamaño requerido de la muestra de modo que se tenga una confianza del 90% de que el error de la estimación sea menor que 0,50 kg/mm^2.

Si los tamaños de las muestras son iguales, podemos escribir:

$$e = z_c \cdot \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}} = z_c \cdot \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{n}} = 1{,}645 \cdot \sqrt{\frac{1{,}0^2 + 1{,}5^2}{n}} < 0{,}50\ kg/mm^2$$

$$\Rightarrow n > \frac{z_c^2(\sigma_1^2 + \sigma_2^2)}{e^2} = \frac{1{,}645^2\left(1{,}0^2 + 1{,}5^2\right)}{0{,}50^2} = 35{,}18 \quad \Rightarrow \quad \mathbf{n \geq 36}$$

- *Con una confianza del 90% decimos que el error de estimación será menor que 0,50 kg/mm^2 si, tomando muestras de igual tamaño, éste es mayor o igual a 36.*

4. Un fabricante de turbinas hidráulicas considera al tiempo que transcurre hasta que es necesario reparar la máquina por problemas de cavitación en el rotor de la misma como una variable aleatoria con media 5000 horas y desviación estándar de 40 horas. A este tiempo lo llama vida eficaz y acepta que la distribución de la vida eficaz es muy próxima a la normal.

El fabricante introduce una mejora en el proceso de fabricación del rotor que aumenta el tiempo de la vida eficaz promedio a 5050 horas y disminuye la desviación estándar a 30 horas.

Para comprobar lo que afirma el fabricante, se toma del proceso "antiguo" una muestra aleatoria de tamaño 16, de la que se obtiene una media muestral de 4998 horas y una desviación estándar de 41 horas; de la muestra aleatoria del proceso "mejorado" de tamaño 25, se obtiene una media de 5052 horas y una desviación estándar de 29 horas. Construya un intervalo de confianza del 95% para la verdadera diferencia de la vida eficaz, si los procesos "antiguo" y "mejorado" pueden considerarse como poblaciones independientes.

X_A:"Tiempo de vida eficaz del rotor fabricado por el método antiguo"
X_M:"Tiempo de vida eficaz del rotor fabricado por el método mejorado"
$X_A \sim$ Normal (μ_A = 5000 h ; σ_A = 40 h)
$X_M \sim$ Normal (μ_M = 5050 h ; σ_M = 30 h)

$\overline{X}_A - \overline{X}_M \sim$ Normal ($\mu_{\overline{X}_A-\overline{X}_M} = \mu_A - \mu_M$; $\sigma_{\overline{X}_A-\overline{X}_M} = \sqrt{\frac{\sigma_A^2}{n_A}+\frac{\sigma_M^2}{n_M}}$) Por provenir de muestras normales

Según los siguientes datos del problema:
$\overline{x}_A$= 4998 h σ_A = 40 h n_A = 16
$\overline{x}_M$= 5052 h σ_M = 30 h n_M = 25
$1-\alpha = 0,95 \Rightarrow z_{\alpha/2} = \pm 1,96$

Los límites de confianza del 95% son :

$$(\overline{x}_A - \overline{x}_M) \pm z_c \cdot \sqrt{\frac{\sigma_A^2}{n_A}+\frac{\sigma_M^2}{n_M}} = (4998 - 5052) \pm 1,96 \cdot \sqrt{\frac{40^2}{16}+\frac{30^2}{25}} =$$

= - 54 ± 22,8573 h

Luego,

$-76,8573 \text{ h} < \mu_A - \mu_M < -31,1427 \text{ h}$ $1 - \alpha = 0,95$

- *Se tiene un 95% de confianza, a partir de los datos de las muestras, de que el intervalo (-76,8573 h ; -31,1427 h) contiene a la verdadera diferencia de medias poblacionales.*
- *Esto indica que hay una diferencia significativa entre los procesos "antiguo" y "mejorado", especialmente, diremos que el proceso mejorado tiene una vida eficaz mayor, ya que todos los valores del intervalo son negativos, lo cual indica que $\mu_A < \mu_M$.*

5. Se desea vender esferas para utilizar en un mecanismo de precisión. Las especificaciones del producto, en cuanto al diámetro de tales esferas, indican que no deben superar las 4,40 pulgadas. De 9 mediciones realizadas se tienen los siguientes datos:

Muestra	1	2	3	4	5	6	7	8	9
Diámetro (pulgadas)	4,32	4,39	4,40	4,44	4,34	4,36	4,37	4,42	4,38

Suponiendo normalidad:

a) Halle un intervalo de confianza del 99% para el diámetro medio de las esferas.

Con frecuencia, se intenta estimar la media de una población cuando se desconoce la varianza.

Recordemos que si se tiene una muestra aleatoria de una distribución normal, entonces la variable aleatoria $T = \dfrac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}}$ *tiene una distribución t de Student con n-1 grados de libertad.*

S es la desviación estándar muestral. En estas condiciones con σ *desconocida, T puede utilizarse para determinar un intervalo de confianza de* μ*. El procedimiento es el mismo que cuando se conocía* σ*, excepto que se sustituye* σ *por S y distribución normal estándar por distribución t, es decir, z por t.*

A partir de esto podemos decir que:

$$P(-t_{\alpha/2} < T < t_{\alpha/2}) = 1-\alpha$$

$$P\left(-t_{\alpha/2} < \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} < t_{\alpha/2}\right) = 1-\alpha$$

$$P\left(\overline{X} - t_{\alpha/2} \cdot \frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2} \cdot \frac{S}{\sqrt{n}}\right) = 1-\alpha$$

Si $\overline{x}$ *y s son la media y la desviación estándar de una muestra aleatoria de tamaño n de una población con varianza desconocida* σ^2*, el intervalo de confianza del* $(1-\alpha)100\%$ *para* μ *es:*

$$\left(\overline{x} - t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \; ; \; \overline{x} + t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}\right)$$

donde $t_{\alpha/2}$ *es el valor t con n-1 grados de libertad, lo que deja un área de* $\alpha/2$ *a la derecha.*

En nuestro problema:

X:"Diámetro de las esferas"

$X \sim Normal(\mu\,;\,\sigma)$

$\dfrac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \sim$ *t con n-1 grados de libertad*

Según los siguientes datos del problema:

$\overline{x} = 4{,}38$ *pulgadas*

$s = 0{,}03775$ *pulgadas*

$n = 9$

$1-\alpha = 0{,}99 \Rightarrow t_{0{,}005;8} = \pm\, 3{,}355$

Los límites de confianza del 99% son:

$$\overline{x} \pm t_c \,.\, \frac{s}{\sqrt{n}} = \overline{x} \pm t_{0,005,8} \,.\, \frac{s}{\sqrt{n}} = 4,38 \pm 3,355 \,.\, \frac{0,03775}{\sqrt{9}} =$$

$$= \mathbf{4,38 \pm 0,0422 \text{ pulgadas}}$$

Luego,

$$\mathbf{4,34 \text{ pulgadas} < \mu < 4,42 \text{ pulgadas}} \qquad \mathbf{1 - \alpha = 0,99}$$

Se tiene un 99% de confianza, a partir de los datos de la muestra, de que el intervalo (4,34 pulgadas ; 4,42 pulgadas) contiene a la verdadera media poblacional.

b) ¿Qué se podrá decir, con una confianza del 99%, sobre la magnitud posible del error si se utiliza a la media de la muestra de estas 9 observaciones como una estimación de la media de la población?

> *Si se utiliza $\overline{x}$ como una estimación de μ, se puede tener una confianza del $(1-\alpha)100\%$ de que el error no excederá $t_{\alpha/2} \,.\, \frac{s}{\sqrt{n}}$.*

El máximo error de estimación con una confianza del $(1-\alpha).100\%$, según los siguientes datos del problema, es:

$$e = t_c \,.\, \frac{s}{\sqrt{n}} = t_{0,005,8} \,.\, \frac{s}{\sqrt{n}} = 3,355 \,.\, \frac{0,03775}{\sqrt{9}} = \mathbf{0,0422 \text{ pulgadas}}$$

Se puede decir con un 99% de confianza, a partir de los datos de la muestra, que el error máximo probable será de 0,0422 pulgadas.

c) Sin variar la magnitud del error máximo probable, ¿cuál sería el tamaño de la muestra para un nivel de confianza del 90%?

> *Si se utiliza $\overline{x}$ como una estimación de μ, se puede tener una confianza del $(1-\alpha)100\%$ de que el error no excederá una cantidad específica e cuando el tamaño de la muestra es*
>
> $$n = \left(\frac{t_{\alpha/2} \,.\, s}{e}\right)^2$$

El tamaño de muestra, para tener un $(1-\alpha).100\%$ de confianza que el error es menor que e, en este caso, es:

$$e = t_c \,.\, \frac{s}{\sqrt{n}} = t_{0,05,8} \,.\, \frac{s}{\sqrt{n}} = 1,86 \,.\, \frac{0,03775}{\sqrt{n}} < 0,0422 \text{ pulgadas}$$

$$\Rightarrow \quad n \; > \; \frac{t_c^2 s^2}{e^2} = \frac{1,86^2 . 0,03775^2}{0,0422^2} = 2,77 \quad \Rightarrow \quad \mathbf{n \geq 3}$$

- *Con una confianza del 90% decimos que el error de estimación será menor que 0,0422 pulgadas si se toma una muestra de tamaño mayor o igual a 3.*

d) Construya un intervalo con un nivel de confianza del 95% para la varianza poblacional.

Si se toma una muestra de tamaño n de una población normal con varianza σ^2, y se calcula la varianza muestral s^2, se obtiene un valor del estadístico S^2. Esta varianza muestral calculada se utilizará como una estimación puntual de σ^2. Por lo que al estadístico S^2 se le llama estimador de σ^2.

Puede establecerse una estimación por intervalo de σ^2 utilizando el estadístico $\chi^2 = (n-1)\,.\,\frac{s^2}{\sigma^2}$ que tiene una distribución χ^2 con n-1 grados de libertad cuando las muestras se seleccionan a partir de una población normal.

$$P(\chi^2_{1-\alpha/2} < \chi^2 < \chi^2_{\alpha/2}) = 1-\alpha$$

$$P(\chi^2_{1-\alpha/2} < (n-1)\,.\,\frac{s^2}{\sigma^2} < \chi^2_{\alpha/2}) = 1-\alpha$$

$$P(\frac{s^2.(n-1)}{X^2_{\alpha/2;n-1}} < \sigma^2 < \frac{s^2.(n-1)}{X^2_{1-\alpha/2;n-1}}) = 1-\alpha$$

Si s^2 es la varianza de una muestra aleatoria de tamaño n de una población normal, un intervalo de confianza del $(1-\alpha)100\%$ para σ^2 es:

$$\left(\frac{s^2.(n-1)}{X^2_{\alpha/2;n-1}}\,;\,\frac{s^2.(n-1)}{X^2_{1-\alpha/2;n-1}}\right)$$

donde $\chi^2_{1-\alpha/2}$ y $\chi^2_{\alpha/2}$ son valores χ^2 con n-1 grados de libertad, con áreas $\alpha/2$ y $1-\alpha/2$, respectivamente, a la derecha.

En nuestro problema:

X:"Diámetro de las esferas"

$X \sim$ Normal $(\mu\,;\,\sigma)$

$(n-1)\,.\,\frac{s^2}{\sigma^2} \sim \chi^2$ *con n-1 grados de libertad*

Según los siguientes datos del problema:

$\overline{x}$ *= 4,38 pulgadas*

s = 0,03775 pulgadas

n = 9

Estimación de parámetros

$1-\alpha = 0{,}95 \quad \Rightarrow \quad \chi^2_{0,025;8} = 17{,}535 \quad y \quad \chi^2_{0,975;8} = 2{,}180$

El intervalo de confianza del 95% para la varianza poblacional es :

$$\frac{s^2.(n-1)}{\chi^2_{0,025;8}} < \sigma^2 < \frac{s^2.(n-1)}{\chi^2_{0,975;8}}$$

$$\frac{0{,}03775^2.(9-1)}{17{,}535} < \sigma^2 < \frac{0{,}03775^2.(9-1)}{2{,}180}$$

$$\mathbf{0{,}00065 < \sigma^2 < 0{,}00523 \qquad 1-\alpha = 0{,}95}$$

$$\mathbf{0{,}0255 < \sigma < 0{,}0723 \qquad 1-\alpha = 0{,}95}$$

- *Se tiene un 95% de confianza, a partir de los datos de la muestra, de que el intervalo (0,00065 pulgadas^2 ; 0,00523 pulgadas^2) contiene a la verdadera varianza poblacional.*

6. Aceptando que la distribución del consumo anual de energía eléctrica del tipo residencial de los clientes de una empresa distribuidora del servicio es aproximadamente normal, y que la medición de 31 clientes seleccionados aleatoriamente dio los siguientes resultados:

Consumo medio anual, en miles de MWh/año, de clientes con tarifa residencial

490,9	435,3	580,6	510,1	600,4	495,1	520,2	501,5
483,2	520,9	460,1	500,8	490,7	490,7	499,6	498,3
510,6	473,4	474,9	490,1	410,2	488,6	500,5	505,4
491,0	421,8	490,6	439,9	506,4	490,2	475,5	

a) Estime mediante un intervalo de confianza del 95%, el consumo medio anual de los usuarios residenciales del servicio.

X:"Consumo anual de los usuarios residenciales"

$X \sim Normal\ (\mu\ ;\ \sigma)$

$$\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} \sim \text{t con n-1 grados de libertad}$$

Aunque al ser n > 30, podemos aplicar el teorema del límite central y decir que:

$\overline{X} \sim Normal\ (\mu_{\overline{x}} = \mu\ ;\ \sigma_{\overline{x}} = s/\sqrt{n}\,)$ *(Ver último párrafo de la página 247 del libro de Walpole, Myers, Myers)*

Notas:

- *Como regla práctica si $n \geq 30$, aproximamos la distribución t a la distribución normal.*
- *En este ejercicio lo resolveremos de las dos maneras para observar si hay mucha o poca diferencia.*

Según los siguientes datos del problema:

$\overline{x} = 492$ *MWh/año*

Estimación de parámetros

s = 38 MWh/año
n = 31
$1-\alpha = 0{,}95 \Rightarrow t_{0,025;30} = \pm\, 2{,}042$

Los límites de confianza del 95%, usando la distribución t-Student, son:

$$\overline{x} \pm t_c \,.\, \frac{s}{\sqrt{n}} = \overline{x} \pm t_{0,025,30} \,.\, \frac{s}{\sqrt{n}} = 492 \pm 2{,}042 \,.\, \frac{38}{\sqrt{31}} =$$

$= \mathbf{492 \pm 13{,}94}$ ***MWh/año***

Luego,

$\mathbf{478{,}06}$ ***MWh/año*** $< \mu <$ $\mathbf{505{,}94}$ ***MWh/año*** $\qquad 1 - \alpha = 0{,}95$

🗣 *Se tiene un 95% de confianza, a partir de los datos de la muestra, de que el intervalo (478,06 MWh/año ; 505,94 MWh/año) contiene a la verdadera media poblacional del consumo anual de clientes con tarifa residencial.*

Según los siguientes datos del problema:
$\overline{x}$ *= 492 MWh/año*
s = 38 MWh/año
n = 31
$1-\alpha = 0{,}95 \quad \Rightarrow \quad z_{\text{área central} = 0,95} = \pm\, 1{,}96$

Los límites de confianza del 95%, usando la distribución normal, son:

$$\overline{x} \pm z_c \,.\, \frac{s}{\sqrt{n}} = \overline{x} \pm z_{\text{área central} = 0,95} \,.\, \frac{s}{\sqrt{n}} = 492 \pm 1{,}96 \,.\, \frac{38}{\sqrt{31}} =$$

$= \mathbf{492 \pm 13{,}38}$ ***MWh/año***

Luego,

$\mathbf{478{,}62}$ ***MWh/año*** $< \mu <$ $\mathbf{505{,}38}$ ***MWh/año*** $\qquad 1 - \alpha = 0{,}95$

🗣 *Se tiene un 95% de confianza, a partir de los datos de la muestra, de que el intervalo (478,62 MWh/año ; 505,38 MWh/año) contiene a la verdadera media poblacional del consumo anual de clientes con tarifa residencial.*

b) Estime mediante un intervalo de confianza del 99%, la varianza del consumo medio anual de los clientes con tarifa residencial.

X:"Consumo anual de los usuarios residenciales"

$X \sim \text{Normal}\,(\mu \,;\, \sigma)$

$$(n-1)\,.\, \frac{s^2}{\sigma^2} \sim \chi^2 \text{ con n-1 grados de libertad}$$

Estimación de parámetros

Según los siguientes datos del problema:
$s = 38$ *MWh/año*
$n = 31$
$1-\alpha = 0{,}99 \Rightarrow \chi^2_{0,005;30} = 53{,}672$ y $\chi^2_{0,995;30} = 13{,}787$

El intervalo de confianza del 99% para la varianza poblacional es:

$$\frac{s^2.(n-1)}{\chi^2_{0,005;30}} < \sigma^2 < \frac{s^2.(n-1)}{\chi^2_{0,995;30}}$$

$$\frac{38^2.(31-1)}{53{,}672} < \sigma^2 < \frac{38^2.(31-1)}{13{,}787}$$

$$\mathbf{807{,}12 < \sigma^2 < 3142{,}09} \qquad \mathbf{1 - \alpha = 0{,}99}$$

$$\mathbf{28{,}41 < \sigma < 56{,}05} \qquad \mathbf{1 - \alpha = 0{,}99}$$

- *Se tiene un 95% de confianza, a partir de los datos de la muestra, de que el intervalo (807,12 (MWh/año)^2 ; 3142,09 (MWh/año)^2) contiene a la verdadera varianza poblacional.*

7. Los resultados dados por una A.R.T. sobre el análisis del peso de los empleados que trabajan en las oficinas y de los empleados que trabajan en la planta de producción. Al tomar una muestra del peso de diez empleados que trabajan en las oficinas, se encontró que el peso promedio es 90 kg, con una desviación estándar muestral de 5 kg. Los resultados obtenidos en la muestra de quince empleados que trabajan en la planta de producción fueron 87 y 4 kg, para el peso promedio y la desviación estándar muestral, respectivamente. Suponiendo que el peso de los empleados está distribuido de manera normal, encuentre un intervalo de confianza del 95% para la diferencia entre las medias de las poblaciones de los dos grupos de empleados. Por otra parte, supóngase que las dos poblaciones normales tienen la misma desviación estándar. Saque conclusiones.

Si las varianzas poblacionales no se conocen y las distribuciones involucradas son aproximadamente normales, la distribución t debe considerarse como en el caso de una sola muestra.

Si no se está dispuesto a suponer que es normal, el trabajar con muestras grandes (mayores que 30), permitirá el uso de s_1 y s_2 en lugar de σ_1 y σ_2, respectivamente, en el supuesto de que $s_1 \cong \sigma_1$ y $s_2 \cong \sigma_2$. Una vez más, por supuesto, el intervalo de confianza es una aproximación.

En el caso que se desconozcan σ_1^2 y σ_2^2. Si $\sigma_1^2 = \sigma_2^2 = \sigma^2$, se obtiene una variable de la forma:

$$T = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_p.\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad \text{siendo} \qquad S_p^2 = \frac{(n_1-1)\,s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}$$

Estimación de parámetros

que tiene una distribución t con $n_1 + n_2 - 2$ grados de libertad y se puede afirmar con una probabilidad de $1-\alpha$, que la variable estará entre $-t_{\alpha/2}$ y $t_{\alpha/2}$. A partir de esto podemos decir que:

$$P(-t_{\alpha/2} < T < t_{\alpha/2}) = 1-\alpha$$

$$P\left(-t_{\alpha/2} < \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_p.\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} < t_{\alpha/2}\right) = 1-\alpha$$

$$P\left((\overline{x}_1 - \overline{x}_2) - t_{\alpha/2;\,\gamma}.S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} < \mu_1 - \mu_2 < (\overline{x}_1 - \overline{x}_2) + t_{\alpha/2;\,\gamma}.S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right) = 1-\alpha$$

> *Si $\overline{x}_1$ y $\overline{x}_2$ son las medias de muestras aleatorias de tamaño n_1 y n_2 de poblaciones con varianzas conocidas σ_1^2 y σ_2^2, respectivamente, un intervalo de confianza del $(1-\alpha)100\%$ para $\mu_1 - \mu_2$ es:*
>
> $$\left((\overline{x}_1 - \overline{x}_2) - t_{\alpha/2;\,\gamma}\,.\,S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\;;\;(\overline{x}_1 - \overline{x}_2) + t_{\alpha/2;\,\gamma}\,.\,S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)$$
>
> *donde $t_{\alpha/2;\,\gamma}$ es el valor de t a la derecha del cual se tiene un área de $\alpha/2$ con $\gamma = n_1+n_2-2$ grados de libertad.*

En nuestro problema:

X_1:"Peso de los empleados que trabajan en la oficina"

X_2:"Peso de los empleados que trabajan en la planta de producción"

$X_1 \sim$ *Normal* $(\mu_1 ; \sigma_1)$

$X_2 \sim$ *Normal* $(\mu_2 ; \sigma_2)$

$$\frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_p.\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim \text{t-Student con } n_1 + n_2 - 2 \text{ grados de libertad}$$

Según los siguientes datos del problema:

$\overline{x}_1 = 90$ kg $\qquad s_1 = 5$ kg $\qquad n_1 = 10$

$\overline{x}_2 = 87$ kg $\qquad s_2 = 4$ kg $\qquad n_2 = 15$

$1-\alpha = 0{,}95 \Rightarrow t_{0{,}025\,;\,10+15-2} = \pm\, 2{,}069$

Los límites de confianza del 95% son :

$$(\overline{x}_1 - \overline{x}_2) \pm t_{0{,}025\,;\,23}\,.S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \qquad \text{siendo } S_p^2 = \frac{(n_1-1)\,s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

$$(90-87) \pm 2{,}069.\ S_p\,.\sqrt{\frac{1}{10}+\frac{1}{15}} \qquad \text{siendo } S_p^2 = \frac{(10-1)\,5^2 + (15-1)\,4^2}{10+15-2} =$$

$= 19{,}5217\ kg^2$

$$(90 - 87) \pm 2{,}069 .\ 4{,}4183 .\ \sqrt{\frac{1}{10} + \frac{1}{15}} = \mathbf{3 \pm 3{,}732\ kg}$$

Luego,

$\mathbf{-\ 0{,}732\ kg < \mu_1 - \mu_2 < 6{,}732\ kg} \qquad \mathbf{1 - \alpha = 0{,}95}$

- *Se tiene un 95% de confianza, a partir de los datos de la muestra, de que el intervalo (- 0,732 kg ; 6,732 kg) contiene a la verdadera media poblacional.*
- *Esto indica que no hay una diferencia significativa entre los pesos de los empleados de distintas secciones. Al estar el cero en el intervalo, podemos decir que $\mu_1 - \mu_2$ puede tomar el valor cero, es decir, μ_1 puede ser igual a μ_2.*

8. La pintura para señalamiento vial de rutas y autopistas se distribuye en dos colores: blanco y amarillo. El interés se centra en el tiempo de secado de la pintura. Se sospecha que la pintura de color amarillo se seca más rápidamente que la blanca. Se obtienen las siguientes mediciones del tiempo de secado de ambos tipos de pintura, en minutos:

Blanca	120	132	123	122	140	110	120	107		
Amarilla	126	124	116	125	109	130	125	117	129	130

a) Encuentre un intervalo de confianza del 99% para la diferencia entre los tiempos de secado promedio, suponiendo que las desviaciones estándar de éstos son iguales. Suponga que el tiempo de secado está distribuido de manera normal.

X_B:"Tiempo de secado de la pintura blanca"

X_A:"Tiempo de secado de la pintura amarilla"

$X_B \sim Normal\ (\mu_B\ ;\ \sigma_B)$

$X_A \sim Normal\ (\mu_A\ ;\ \sigma_A)$

$$\frac{(\overline{X}_B - \overline{X}_A) - (\mu_B - \mu_A)}{S_p . \sqrt{\frac{1}{n_B} + \frac{1}{n_A}}} \sim \text{t-Student con } n_B + n_A - 2 \text{ grados de libertad}$$

Según los siguientes datos del problema:

$\overline{x}_B = 122\ min \qquad s_B = 11\ min \qquad n_B = 8$

$\overline{x}_A = 123\ min \qquad s_A = 7\ min \qquad n_A = 10$

$1 - \alpha = 0{,}99 \quad \Rightarrow \quad t_{0,005\ ;\ 8+10-2} = \pm\ 2{,}921$

Los límites de confianza del 99% son :

$$(\overline{x}_B - \overline{x}_A) \pm t_{0,005\ ;\ 16} .\ S_p \sqrt{\frac{1}{n_B} + \frac{1}{n_A}} \quad \text{siendo } S_p^2 = \frac{(n_B - 1)\ s_B^2 + (n_A - 1) s_A^2}{n_B + n_A - 2}$$

$$(122 - 123) \pm 2{,}921.\ S_p \cdot \sqrt{\frac{1}{8}+\frac{1}{10}} \quad \text{siendo } S_p^2 = \frac{(8-1)\,11^2 + (10-1)\,7^2}{8+10-2} = $$
$$= 8{,}97\ min^2$$

$$(122 - 123) \pm 2{,}921.\ 2{,}9954.\ \sqrt{\frac{1}{8}+\frac{1}{10}} = \mathbf{-1 \pm 4{,}15\ min}$$

Luego,

$$\mathbf{-\ 5{,}15\ min < \mu_B - \mu_A < 3{,}15\ min} \qquad \mathbf{1 - \alpha = 0{,}99}$$

- *Se tiene un 99% de confianza, a partir de los datos de las muestras, de que el intervalo (-5,15 min ; 3,15 min) contiene a la verdadera diferencia de medias poblacionales.*
- *Esto indica que no hay una diferencia significativa entre los tiempos de secado de cada una de las pinturas.*

b) ¿Existe alguna evidencia que indique que la pintura amarilla se seca más rápidamente que la blanca?

No existe evidencia que indique que la pintura amarilla seca más rápidamente porque el 0, es decir, el valor que indica que no hay diferencias está incluido en el intervalo, todas esta afirmaciones las podemos realizar con un 99% de confianza.

9. Un industrial está interesado en la uniformidad de la máquina que utiliza para dosificar áridos. De manera específica, es deseable que la desviación estándar σ del proceso de dosificación sea menor de 0,5 kg; de otro modo existe un porcentaje mayor del deseable de dosificaciones incorrectas. Supóngase que la distribución del peso del árido dosificado es aproximadamente normal. Al tomar una muestra aleatoria de 20 dosificaciones realizadas, se obtiene una varianza muestral de 0,19 kg^2.

a) Determine el intervalo de confianza "superior" del 95% para la varianza poblacional.

X:"Peso del árido dosificado"

$X \sim Normal\ (\mu\ ;\ \sigma)$

$(n-1)\,.\ \dfrac{s^2}{\sigma^2} \sim \chi^2$ *con n-1 grados de libertad*

Según los siguientes datos del problema:

$s^2 = 0{,}19\ kg^2$

$n = 20$

$1 - \alpha = 0{,}95 \quad \Rightarrow \quad \chi^2_{0,95;19} = 10{,}117$

El intervalo de confianza "superior" del 95% , está dado por

$$\sigma^2 < \frac{s^2.(n-1)}{\chi^2_{0,95;19}}$$

$$\sigma^2 < \frac{0,19.(20-1)}{10,117}$$

$\boldsymbol{\sigma^2 < 0,3568}$ $\qquad$ $\boldsymbol{1 - \alpha = 0,95}$

$\boldsymbol{\sigma < 0,5973}$ $\qquad$ $\boldsymbol{1 - \alpha = 0,95}$

Se tiene un 95% de confianza, a partir de los datos de la muestra, de que el intervalo (0 kg^2 ; 0,3568 kg^2) contiene a la varianza poblacional.

b) Con un nivel de confianza del 95%, ¿apoyan los datos la condición de que la desviación estándar del proceso es menor que 0,5 kg?

Con un nivel de confianza del 95%, a partir de los datos de la muestra, se puede decir que la desviación estándar poblacional del proceso es menor que 0,5 kg.

10. Con el objeto de investigar el tiempo de secado de un nuevo tipo de pintura, se prepararon cuatro paneles de ensayo y se midió el tiempo de secado en los mismos, obteniéndose los siguientes resultados:

Panel	1	2	3	4
Tiempo (en h y min)	1h 54min	2h 2min	2h 5min	1h 55min
Tiempo (en h)	*1,9 h*	*2,03... h*	*2,083... h*	*1,916... h*

Con los datos de la muestra y suponiendo normalidad:

a) Construya un intervalo de confianza del 95% para la varianza del tiempo medio de secado real de la pintura.

X:"Tiempo de secado de la pintura"

$X \sim Normal\ (\mu\ ;\ \sigma)$

$(n-1)\ .\ \dfrac{s^2}{\sigma^2} \sim \chi^2$ *con n-1 grados de libertad*

Según los siguientes datos del problema:

$s^2 = 0,00796\ h^2$

$n = 4$

$1 - \alpha = 0,95 \Rightarrow \chi^2_{0,025;3} = 9,348$ y $\chi^2_{0,975;3} = 0,216$

El intervalo de confianza del 95% , está dado por:

$$\frac{s^2.(n-1)}{\chi^2_{0,025;3}} < \sigma^2 < \frac{s^2.(n-1)}{\chi^2_{0,975;3}}$$

$$\frac{0,00796.(4-1)}{9,348} < \sigma^2 < \frac{0,00796.(4-1)}{0,216}$$

$\boldsymbol{0,0026 < \sigma^2 < 0,1106}$ $\qquad$ $\boldsymbol{1 - \alpha = 0,95}$

Estimación de parámetros

$$\mathbf{0{,}051 < \sigma < 0{,}3326} \qquad\qquad \mathbf{1 - \alpha = 0{,}95}$$

Se tiene un 95% de confianza, a partir de los datos de la muestra, que el intervalo (0,0026 h^2 ; 0,1106 h^2) contiene a la varianza poblacional.

b) Construya un intervalo de confianza del 95% para el tiempo medio de secado real de la pintura.

X:"Tiempo de secado de la pintura"

$X \sim Normal\ (\mu\ ;\ \sigma)$

$$\frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t \text{ con n-1 grados de libertad}$$

Según los siguientes datos del problema:

$\overline{x}$ = 1h 59 min = 1,983... h

s = 0,0892 h

n = 4

$1 - \alpha = 0{,}95 \Rightarrow t_{0{,}025;3} = \pm\, 3{,}182$

Los límites de confianza del 95% son :

$$\overline{x} \pm t_c . \frac{s}{\sqrt{n}} = \overline{x} \pm t_{0{,}025{,}3} . \frac{s}{\sqrt{n}} = 1{,}983 \pm 3{,}182 . \frac{0{,}0892}{\sqrt{4}} = \mathbf{1{,}983 \pm 0{,}1304\ h}$$

Luego,

$$\mathbf{-1{,}85\ h < \mu < 2{,}11\ h} \qquad\qquad \mathbf{1 - \alpha = 0{,}95}$$

Se tiene un 95% de confianza, a partir de los datos de la muestra, de que el intervalo (-1,85 h ; 2,11 h) contiene al tiempo medio de secado poblacional.

c) ¿Qué tamaño de muestra seleccionaría si, a nivel del 99%, es necesario limitar el error máximo probable al valor obtenido en el punto anterior?

$$e = t_c . \frac{s}{\sqrt{n}} = t_{0{,}005{,}3} . \frac{s}{\sqrt{n}} = 5{,}841 . \frac{0{,}0892}{\sqrt{n}} < 0{,}1304\ h$$

$$\Rightarrow \quad n > \frac{t_c^2 s^2}{e^2} = \frac{5{,}841^2 . 0{,}0892^2}{0{,}1304^2} = 15{,}98 \quad \Rightarrow \quad \mathbf{n \geq 16}$$

Con una confianza del 99% decimos que el error de estimación será menor que 0,1304 horas si se toma una muestra de tamaño mayor o igual a 16.

11. Se debe efectuar el alumbrado en un predio municipal. Luego de analizar las propuestas de los distintos oferentes se vio la conveniencia de usar lámparas de dos fábricas del medio.

a) A los efectos de controlar la calidad de las lámparas, se realizaron pruebas de calidad, obteniéndose los siguientes resultados:

- Una muestra de 21 lámparas del fabricante A dio una vida media de 1400 horas y una desviación estándar de 120 horas.
- Una muestra de 25 lámparas del fabricante B dio una vida media de 1200 horas y una desviación estándar de 100 horas.

Se desea saber si es aceptable suponer que hay homogeneidad en la calidad de las lámparas empleadas en la obra, aunque hayan sido producidas por dos fábricas distintas, cada una con sus propios equipos, condiciones de elaboración y supervisión. Se acepta que el tiempo de vida de las lámparas de ambos proveedores está distribuido normalmente y son independientes.

Una estimación puntual del cociente de dos varianzas poblacionales σ_1^2/σ_2^2 está dada por la razón s_1^2/s_2^2 de las varianzas muestrales. De aquí que el estadístico S_1^2/S_2^2 se llama estimador de σ_1^2/σ_2^2.

Si σ_1^2 y σ_2^2 son las varianzas de poblaciones normales, se puede establecer una estimación del intervalo de σ_1^2/σ_2^2 utilizando el estadístico:

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{\sigma_2^2}{\sigma_1^2} \cdot \frac{S_1^2}{S_2^2}$$

el cual tiene una distribución F con $n_1 -1$ y $n_2 -1$ grados de libertad, siendo n_1 y n_2 los tamaños de dos muestras aleatorias independientes que se seleccionan de dos poblaciones normales.

Por lo tanto, se puede afirmar con una probabilidad de $1 - \alpha$, que la variable F estará entre F_1 y F_2.

A partir de esto podemos decir que:

$$P(F_1 < F < F_2) = 1 - \alpha$$

$$P\left(F_{1;\, n_1 -1;\, n_2 -1} < \frac{\sigma_2^2}{\sigma_1^2} \cdot \frac{S_1^2}{S_2^2} < F_{2;\, n_1 -1;\, n_2 -1} \right) = 1 - \alpha$$

$$P\left(F_{1;\, n_1 -1;\, n_2 -1} \frac{S_2^2}{S_1^2} < \frac{\sigma_2^2}{\sigma_1^2} < F_{2;\, n_1 -1;\, n_2 -1} \cdot \frac{S_2^2}{S_1^2} \right) = 1 - \alpha$$

$$P\left(\frac{1}{F_2} \cdot \frac{S_1^2}{S_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{1}{F_1} \cdot \frac{S_1^2}{S_2^2} \right) = 1 - \alpha$$

> *Si s_1^2 y s_2^2 son las varianzas de muestras aleatorias independientes de tamaño n_1 y n_2, respectivamente, de poblaciones normales, entonces un intervalo de confianza del $(1 - \alpha)100\%$ para σ_1^2 / σ_2^2 es:*
>
> $$\left(\frac{1}{F_2} \cdot \frac{S_1^2}{S_2^2} \; ; \; \frac{1}{F_1} \cdot \frac{S_1^2}{S_2^2} \right)$$
>
> *donde F_1 es el valor de F con $n_1 -1$ y $n_2 -1$ grados de libertad a la derecha del*

cual se tiene un área de 1-α/2 y donde F_2 es el valor de F con n_1 -1 y n_2 -1 grados de libertad a la derecha del cual se tiene un área de α/2.

En nuestro problema:
X_A:"Tiempo de vida de las lámparas A"
X_B:"Tiempo de vida de las lámparas B"
$X_A \sim$ Normal (μ_A ; σ_A)
$X_B \sim$ Normal (μ_B ; σ_B)

$$\frac{S_A^2 / \sigma_A^2}{S_B^2 / \sigma_B^2} = \frac{\sigma_B^2}{\sigma_A^2} \cdot \frac{S_A^2}{S_B^2} \sim \text{F con } n_A - 1 \text{ y } n_B - 1 \text{ grados de libertad}$$

Según los siguientes datos del problema:

$\bar{x}_A$= 1400 h	s_A = 120 h	n_A = 21
$\bar{x}_B$= 1200 h	s_B = 100 h	n_B = 25

El intervalo de confianza para el cociente de varianzas poblacionales del (1 - α).100%, será:

$$\frac{1}{F_{0,01;\,20\,,24}} \cdot \frac{s_A^2}{s_B^2} < \frac{\sigma_A^2}{\sigma_B^2} < \frac{1}{F_{0,99;\,20\,,24}} \cdot \frac{s_A^2}{s_B^2}$$

Vemos que debemos calcular valores críticos de F, pero tenemos tablas que sólo tienen los valores críticos para 0,01 y 0,05, entonces debemos recurrir a una propiedad de la distribución F: $\mathbf{F_{1-\alpha;\nu_1,\nu_2} = 1/F_{\alpha;\nu_2,\nu_1}}$

Así, los valores críticos, que llamaremos, F_1 y F_2, serán:

✓ *Para el 98% de confianza*

$F_1 = F_{0,99;\,20\,;\,24} = 1 / F_{0,01;\,24\,;\,20} = 1 / 2,86 = 0,35$

$F_2 = F_{0,01;\,20\,;\,24} = 2,74$

Nota: 0,99 y 0,01 son las "colitas", es decir, el valor del área desde el valor crítico hasta +∞

✓ *Para el 90% de confianza*

$F_1 = F_{0,95;\,20\,;\,24} = 1 / F_{0,05;\,24\,;\,20} = 1 / 2,08 = 0,48$

$F_2 = F_{0,05;\,20\,;\,24} = 2,03$

Nota: 0,95 y 0,05 son las "colitas", es decir, indican el valor del área desde el valor crítico hasta +∞

En la fórmula vemos que necesitaremos los valores, $1/F_1$ y $1/F_2$, entonces, éstos serán:

✓ *Para el 98% de confianza*

$$\frac{1}{F_1} = \frac{1}{F_{0,99,20,24}} = \frac{1}{\frac{1}{F_{0,01,24,20}}} = F_{0,01;24;20} = 2,86$$

$$\frac{1}{F_2} = \frac{1}{F_{0,01,20,24}} = \frac{1}{2,74}$$

✓ Para el 90% de confianza

$$\frac{1}{F_1} = \frac{1}{F_{0,95,20,24}} = \frac{1}{\frac{1}{F_{0,05,100,60}}} = F_{0,05;24;20} = 2,08$$

$$\frac{1}{F_2} = \frac{1}{F_{0,05,20,24}} = \frac{1}{2,03}$$

El intervalo para estimar el cociente de varianzas, con un 98% de confianza, es:

$$\frac{1}{F_{0,01,20,24}} \cdot \frac{s_A^2}{s_B^2} < \frac{\sigma_A^2}{\sigma_B^2} < \frac{1}{F_{0,99,20,24}} \cdot \frac{s_A^2}{s_B^2}$$

$$\frac{1}{2,74} \cdot \frac{120^2}{100^2} < \frac{\sigma_A^2}{\sigma_B^2} < 2,86 \cdot \frac{120^2}{100^2}$$

$$\mathbf{0,5255 < \frac{\sigma_A^2}{\sigma_B^2} < 4,1184} \qquad \mathbf{1 - \alpha = 0,98}$$

Para las desviaciones típicas, al calcular la raíz cuadrada :

$$\mathbf{0,7249 < \frac{\sigma_A}{\sigma_B} < 2,0294} \qquad \mathbf{1 - \alpha = 0,98}$$

Se tiene un 98% de confianza, a partir de los datos de las muestras, de que hay homogeneidad entre las vidas medias de las lámparas A y B, ya que el 1 se encuentra en el intervalo, es decir, la suposición de que las varianzas son iguales está contemplada para el 98% de confianza.

El intervalo para estimar el cociente de varianzas, con un 90% de confianza, es:

$$\frac{1}{F_{0,05,20,24}} \cdot \frac{s_A^2}{s_B^2} < \frac{\sigma_A^2}{\sigma_B^2} < \frac{1}{F_{0,95,20,24}} \cdot \frac{s_A^2}{s_B^2}$$

$$\frac{1}{2,03} \cdot \frac{120^2}{100^2} < \frac{\sigma_A^2}{\sigma_B^2} < 2,08 \cdot \frac{120^2}{100^2}$$

$$\mathbf{0,7094 < \frac{\sigma_A^2}{\sigma_B^2} < 2,9952} \qquad \mathbf{1 - \alpha = 0,90}$$

Para las desviaciones típicas, al calcular la raíz cuadrada :

$$0,8422 < \frac{\sigma_A}{\sigma_B} < 1,7307 \qquad\qquad 1-\alpha = 0,90$$

Se tiene un 90% de confianza, a partir de los datos de las muestras, de que hay homogeneidad entre las vidas medias de las lámparas A y B, ya que el 1 se encuentra en el intervalo, es decir, la suposición de que las varianzas son iguales está contemplada para el 90% de confianza.

b) Si con el mismo tamaño de muestras se hubieran obtenido las mismas vidas medias muestrales, pero con desviaciones estándares de 120 y 70 horas, respectivamente, ¿cuál sería la conclusión?

X_A:"Tiempo de vida de las lámparas A"

X_B:"Tiempo de vida de las lámparas B"

$X_A \sim \text{Normal}\ (\mu_A ; \sigma_A)$

$X_B \sim \text{Normal}\ (\mu_B ; \sigma_B)$

$$\frac{S_A^2 / \sigma_A^2}{S_B^2 / \sigma_B^2} = \frac{\sigma_B^2}{\sigma_A^2} \cdot \frac{S_A^2}{S_B^2} \sim F \text{ con } n_A - 1 \text{ y } n_B - 1 \text{ grados de libertad}$$

Según los siguientes datos del problema:

$\bar{x}_A = 1400\ h$ $\quad s_A = 120\ h$ $\quad n_A = 21$

$\bar{x}_B = 1200\ h$ $\quad s_B = 75\ h$ $\quad n_B = 25$

$1 - \alpha = 0,98 \Rightarrow F_{0,01;20,24} = 2,74$ y $F_{0,99;20,24} = 1/F_{0,01;24,20} = 1/2,86$

El intervalo para estimar el cociente de varianzas, con un 98% de confianza, es:

$$\frac{1}{F_{0,01,20,24}} \cdot \frac{s_A^2}{s_B^2} < \frac{\sigma_A^2}{\sigma_B^2} < \frac{1}{F_{0,99,20,24}} \cdot \frac{s_A^2}{s_B^2}$$

$$\frac{1}{2,74} \cdot \frac{120^2}{70^2} < \frac{\sigma_A^2}{\sigma_B^2} < 2,86 \cdot \frac{120^2}{70^2}$$

$$1,0725 < \frac{\sigma_A^2}{\sigma_B^2} < 8,4049 \qquad\qquad 1-\alpha = 0,98$$

Para las desviaciones típicas, al calcular la raíz cuadrada :

$$1,0356 < \frac{\sigma_A}{\sigma_B} < 2,8991 \qquad\qquad 1-\alpha = 0,98$$

Se tiene un 98% de confianza, a partir de los datos de las muestras, de que no hay homogeneidad entre las vidas medias de las lámparas A y B, ya que el 1 no se encuentra en el intervalo.

Estimación de parámetros

El intervalo para estimar el cociente de varianzas, con un 98% de confianza, es:

$$\frac{1}{F_{0,05,20,24}} \cdot \frac{s_A^2}{s_B^2} < \frac{\sigma_A^2}{\sigma_B^2} < \frac{1}{F_{0,95,20,24}} \cdot \frac{s_A^2}{s_B^2}$$

$$\frac{1}{2,03} \cdot \frac{120^2}{70^2} < \frac{\sigma_A^2}{\sigma_B^2} < 2,08 \cdot \frac{120^2}{70^2}$$

$$\mathbf{1,4477 < \frac{\sigma_A^2}{\sigma_B^2} < 6,1127} \qquad \mathbf{1 - \alpha = 0,90}$$

Para las desviaciones típicas, al calcular la raíz cuadrada :

$$\mathbf{1,2032 < \frac{\sigma_A}{\sigma_B} < 2,4724} \qquad \mathbf{1 - \alpha = 0,90}$$

🗣 *Se tiene un 90% de confianza, a partir de los datos de las muestras, de que no hay homogeneidad entre las vidas medias de las lámparas A y B, ya que el 1 no se encuentra en el intervalo.*

12. Una compañía de transporte debe decidir sobre la compra de neumáticos de tipo A o B. Para estimar la diferencia, se realiza una prueba empleando 16 neumáticos de cada marca. Los neumáticos se hacen trabajar hasta el desgaste total.

Los resultados son los mostrados en el cuadro siguiente:

Marca	Tamaño de la muestra	Recorrido medio (en km)	Desviación estándar (en km)
A	16	36.300	5.000
B	16	38.100	6.100

a) Construya un intervalo con un nivel de confianza del 90% para la diferencia de medias, suponiendo que las poblaciones tienen distribución normal y son independientes.

El procedimiento visto para determinar los intervalos de confianza para $\mu_1 - \mu_2$ con varianzas desconocidas pero iguales $\sigma_1^2 = \sigma_2^2 = \sigma^2$ requiere la suposición de que las poblaciones son normales. Las desviaciones ligeras de la suposición de varianzas iguales o de normalidad, no alteran el grado de confianza del intervalo. Si las varianzas poblacionales son considerablemente diferentes, aún se obtienen resultados razonables cuando las poblaciones son normales, siempre que $n_1 = n_2$. Por lo tanto, en un experimento planeado, se debe hacer el esfuerzo de igualar el tamaño de las muestras.

Si se tienen dos poblaciones con medias μ_1 y μ_2 y varianzas σ_1^2 y σ_2^2, respectivamente, siendo estas varianzas desconocidas y distintas. El estadístico que se utiliza en este caso es:

$$T = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Estimación de parámetros

que tiene aproximadamente una distribución t con v grados de libertad,

donde $v = \dfrac{(s_1^2/n_1 + s_2^2/n_2)^2}{\dfrac{(s_1^2/n_1)^2}{n_1-1} + \dfrac{(s_2^2/n_2)^2}{n_2-1}}$

dado que v rara vez es un entero, se redondea al entero más cercano.
Por lo tanto, se puede afirmar con una probabilidad de $1 - \alpha$, *que la variable T estará entre* $-t_{\alpha/2}$ *y* $t_{\alpha/2}$.

A partir de esto podemos decir que:

$$P(-t_{\alpha/2;\,v} < T < t_{\alpha/2;\,v}) \approx 1 - \alpha$$

$$P\left(-t_{\alpha/2;\,v} < \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} < t_{\alpha/2;\,v}\right) \approx 1 - \alpha$$

$$P\left((\overline{x}_1 - \overline{x}_2) - t_{\alpha/2;\,v} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} < \mu < (\overline{x}_1 - \overline{x}_2) + t_{\alpha/2;\,v} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right) \approx 1 - \alpha$$

Si $\overline{x}_1$ *y* s_1^2, *y* $\overline{x}_2$ *y* s_2^2, *son las medias y varianzas de muestras pequeñas independientes de tamaño* n_1 *y* n_2, *respectivamente, de distribuciones aproximadamente normales con varianzas diferentes y desconocidas, un intervalo de confianza aproximado del* $(1 - \alpha)100\%$ *para* $\mu_1 - \mu_2$ *es:*

$$\left((\overline{x}_1 - \overline{x}_2) - t_{\alpha/2;\,v} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\ ;\ (\overline{x}_1 - \overline{x}_2) + t_{\alpha/2;\,v} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right)$$

donde $t_{\alpha/2}$ *es el valor de t a la derecha del cual se tiene un área de* $\alpha/2$.

En nuestro problema:

X_A: *"Recorrido del neumático A hasta su desgaste total"*

X_B: *" Recorrido del neumático B hasta su desgaste total"*

$X_A \sim$ *Normal* $(\mu_A ; \sigma_A)$

$X_B \sim$ *Normal* $(\mu_B ; \sigma_B)$

$$\frac{(\overline{X}_A - \overline{X}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}} \sim t \text{ con } v \text{ grados de libertad}$$

El intervalo de confianza del $(1 - \alpha).100\%$, *para la diferencia de medias poblacionales, con* ***varianzas poblacionales diferentes*** *y* ***desconocidas***, *es:*

$$(\overline{x}_A - \overline{x}_B) \pm t_{0,05\,;\,\nu} \cdot \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}} \qquad \text{siendo } \nu = \frac{(s_A^2/n_A + s_B^2/n_B)^2}{\frac{(s_A^2/n_A)^2}{n_A - 1} + \frac{(s_B^2/n_B)^2}{n_B - 1}}$$

Según los siguientes datos del problema:

$\overline{x}_A = 36300$ km $\quad s_A = 5000$ km $\quad n_A = 16$

$\overline{x}_B = 38100$ km $\quad s_B = 6100$ km $\quad n_B = 16$

$$\nu = \frac{(5000^2/16 + 6100^2/16)^2}{\frac{(5000^2/16)^2}{16-1} + \frac{(6100^2/16)^2}{16-1}} = 28{,}8872 \Rightarrow \nu = 29$$

$$1 - \alpha = 0{,}90 \Rightarrow t_{0,05\,;\,\nu = 29} = \pm 1{,}699$$

Luego, los límites de confianza del 90% son:

$$(\overline{x}_A - \overline{x}_B) \pm t_{0,05\,;\,29} \cdot \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$$

$$(36300 - 38100) \pm 1{,}699 \cdot \sqrt{\frac{5000^2}{16} + \frac{6100^2}{16}} = \mathbf{-1800 \pm 3350{,}1441}$$

Luego,

$$\mathbf{-5150{,}1441 \text{ km} < \mu_A - \mu_B < 1550{,}1441 \text{ km}} \qquad \mathbf{1 - \alpha = 0{,}90}$$

- *Se puede esperar con el 90% de confianza y en base a la evidencia muestral que el intervalo (- 5150,1441 km ; 1550,1441 km) contiene a la diferencia de las medias poblacionales.*
- *Esto indica que no hay una diferencia significativa entre los rendimientos de los neumáticos.*

b) Determine un intervalo de confianza del 90%, para el cociente de las varianzas poblacionales. ¿Se justifica el supuesto de que las varianzas son iguales para calcular el intervalo de confianza para la diferencia de las medias?

X_A: *"Recorrido del neumático A hasta su desgaste total"*

X_B: *"Recorrido del neumático B hasta su desgaste total"*

$X_A \sim \text{Normal}\,(\mu_A\,;\,\sigma_A)$

$X_B \sim \text{Normal}\,(\mu_B\,;\,\sigma_B)$

$$\frac{S_A^2/\sigma_A^2}{S_B^2/\sigma_B^2} = \frac{\sigma_B^2}{\sigma_A^2} \cdot \frac{S_A^2}{S_B^2} \sim F \text{ con } n_A - 1 \text{ y } n_B - 1 \text{ grados de libertad}$$

Según los siguientes datos del problema:

$\overline{x}_A = 36300$ km $\quad s_A = 5000$ km $\quad n_A = 16$

$\overline{x}_B$= 38100 km $\quad$ s_B = 6100 km $\quad$ n_B = 16

$1 - \alpha = 0{,}90 \quad \Rightarrow \quad F_{0,05;15,15} = 2{,}40 \quad y \quad F_{0,95;15,15} = 1/F_{0,05;15,15} = 1/2{,}40$

El intervalo de confianza para el 90% de confianza es:

$$\frac{1}{F_{0,05\,,15\,,15}} \cdot \frac{s_A^2}{s_B^2} < \frac{\sigma_A^2}{\sigma_B^2} < \frac{1}{F_{0,95\,,15\,,15}} \cdot \frac{s_A^2}{s_B^2}$$

$$\frac{1}{2{,}40} \cdot \frac{5000^2}{6100^2} < \frac{\sigma_A^2}{\sigma_B^2} < 2{,}40 \cdot \frac{5000^2}{6100^2}$$

$$\mathbf{0{,}2799 < \frac{\sigma_A^2}{\sigma_B^2} < 1{,}6125} \qquad \mathbf{1 - \alpha = 0{,}90}$$

Para las desviaciones típicas, al calcular la raíz cuadrada :

$$\mathbf{0{,}5291 < \frac{\sigma_A}{\sigma_B} < 1{,}2698} \qquad \mathbf{1 - \alpha = 0{,}90}$$

Se tiene un 90% de confianza, a partir de los datos de las muestras, de que no hay diferencias significativas entre las varianzas del desgaste de los neumáticos A y B, ya que el valor 1 se encuentra en el intervalo.

Por lo tanto, se puede usar la fórmula para varianzas desconocidas pero iguales, sin obtener grandes diferencias respecto al intervalo obtenido anteriormente para la diferencia de medias:

$\overline{x}_A$= 36300 km $\quad$ s_A = 5000 km $\quad$ n_A = 16

$\overline{x}_B$= 38100 km $\quad$ s_B = 6100 km $\quad$ n_B = 16

$1 - \alpha = 0{,}90 \quad \Rightarrow \quad t_{0,05\,;\,16+16\,-2} = \pm\, 1{,}697$

Los límites de confianza del 90% son :

$$(\overline{x}_A - \overline{x}_B) \pm t_{0,05\,;\,30} \,.S_p \sqrt{\frac{1}{n_A} + \frac{1}{n_B}} \qquad \text{siendo } S_p^2 = \frac{(n_A - 1)\, s_A^2 + (n_B - 1) s_B^2}{n_A + n_B - 2}$$

$$(36300 - 38100) \pm 1{,}697 .\, S_p \,. \sqrt{\frac{1}{16} + \frac{1}{16}}$$

$$\text{siendo } S_p^2 = \frac{(16 - 1)\, 5000^2 + (16 - 1)\, 6100^2}{16 + 16 - 2} = 31.105.000 \text{ km}^2$$

$$(36300 - 38100) \pm 1{,}697 .\, 5577{,}18567 . \sqrt{\frac{1}{16} + \frac{1}{16}} = \mathbf{-1800 \pm 3346{,}2004}$$

Luego,

$$\mathbf{-\,5146{,}2004 \text{ km} < \mu_A - \mu_B < 1546{,}2004 \text{ km}} \qquad \mathbf{1 - \alpha = 0{,}90}$$

- *Se espera con una confianza del 90% que el intervalo (- 5146,2004 km ; 1546,2004 km) contiene a la verdadera diferencia de las medias poblacionales.*
- *Esto indica que no hay una diferencia significativa entre los rendimientos de los neumáticos.*

13. El ingeniero a cargo del control de calidad de una planta debe detener el proceso de producción cuando se produce una rotura de los sacos en los que se envasa el producto que afecta una proporción superior a 0,008. Al realizar el control de la línea de bolsas de cemento portland de alta resistencia inicial observa que se rompieron 17 bolsas sobre un total de 1000, antes de salir de la planta.

a) ¿Debe detener el proceso de producción? Fundamente la respuesta construyendo un intervalo de confianza bilateral del 95% para la proporción real de sacos rotos.

Un estimador puntual de la proporción p de un experimento binomial está dado por el estadístico $\hat{P} = X/n$, donde X representa el número de éxitos en n intentos. Por lo tanto, la proporción muestral $\hat{p} = x/n$ se utiliza como la estimación puntual del parámetro p.

Si no se espera que la proporción desconocida p se acerque demasiado a cero o a uno, se puede establecer un intervalo de confianza para p considerando la distribución muestral de $\hat{P}$. A representar el fracaso en cada intento binomial por el valor 0 y el éxito por el valor 1, el número de éxitos, x, puede interpretarse como la suma de n valores que consta únicamente de ceros y unos, y $\hat{p}$ es justo la media muestral de estos n valores, de aquí que, de acuerdo al teorema central del límite, para una n lo bastante grande, $\hat{P}$ está distribuida aproximadamente en forma normal con media p y varianza p.q/n. Por lo tanto, se puede afirmar con una probabilidad de 1 - α, que la variable normal estándar,

$$Z = \frac{\hat{P} - p}{\sqrt{p.q/n}}$$

estará entre - $z_{\alpha/2}$ y $z_{\alpha/2}$.

A partir de esto podemos decir que:

$$P(- z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$$

$$P(- z_{\alpha/2} < \frac{\hat{P} - p}{\sqrt{p.q/n}} < z_{\alpha/2}) = 1 - \alpha$$

$$P(\hat{P} - z_{\alpha/2} . \sqrt{\frac{p.q}{n}} < p < \hat{P} + z_{\alpha/2} . \sqrt{\frac{p.q}{n}}) = 1 - \alpha$$

Cuando n es grande, se comete poco error al sustituir p por $\hat{p} = x/n$. Entonces se puede escribir:

$$P\left(\hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}.\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}.\hat{q}}{n}}\right) \approx 1 - \alpha$$

> *Si $\hat{p}$ es la proporción de éxitos en una muestra aleatoria de tamaño n, el intervalo de confianza del $(1 - \alpha)100\%$ para p es:*
>
> $$\left(\hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}.\hat{q}}{n}} \,;\, \hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}.\hat{q}}{n}}\right)$$
>
> *donde $z_{\alpha/2}$ es el valor de z a la derecha del cual se tiene un área de $\alpha/2$.*

En nuestro problema:
X:"Cantidad de sacos rotos"
$X \sim Binomial\ (n\ ;\ p)$
$\hat{p} \sim Normal\ (\mu = p\ ;\ \sigma^2 = p.q/n)$

Según los siguientes datos del problema:
$x = 17$
$n = 1000$
$\hat{p} = \dfrac{17}{1000} = 0{,}017$
$1 - \alpha = 0{,}95 \Rightarrow z_{0,025} = \pm 1{,}96$

El intervalo de confianza para la proporción poblacional para el 95%, es:

$$\hat{p} \pm z_c \cdot \sqrt{\frac{\hat{p}.\hat{q}}{n}} = 0{,}017 \pm 1{,}96 \cdot \sqrt{\frac{0{,}017.(1-0{,}017)}{1000}} = \mathbf{0{,}017 \pm 0{,}008}$$

Luego,
$\mathbf{0{,}009 < p < 0{,}025}$ $\qquad\qquad$ $\mathbf{1 - \alpha = 0{,}95}$

- *Se tiene un 95% de confianza, en base a la evidencia muestral, que el intervalo (0,009 ; 0,025) contiene a la proporción real de sacos rotos.*
- *Esto indica que **hay que detener el proceso** porque todos los valores del intervalo de confianza para la proporción de sacos rotos son superiores a 0,008.*

b) ¿Qué tamaño debe tener la muestra si se desea, con nivel de confianza del 95%, que la proporción de defectuosas esté dentro del 0,004 de la proporción verdadera de defectuosas? Compare los resultados obtenidos con y sin estimación previa de la proporción real.

Si se utiliza $\hat{p}$ como una estimación de p, se puede tener una confianza del $(1 - \alpha)100\%$ de que el error será menor que una cantidad especificada e cuando el tamaño de la muestra sea aproximadamente

$$n = \frac{z_c^2 . \hat{p}.\hat{q}}{e^2}$$

Entonces, el tamaño de muestra haciendo una estimación previa de la proporción es, para los datos del problema:

$$e = z_c . \sqrt{\frac{\hat{p}.\hat{q}}{n}} = 1{,}96 . \sqrt{\frac{0{,}017.(1 - 0{,}017)}{n}} < \mathbf{0{,}004} \Rightarrow$$

(para satisfacer la condición pedida, o sea, que la proporción esté dentro del 0,004 de la verdadera proporción, exigiría un tamaño de muestra)

$$\Rightarrow n > \frac{z_c^2 . \hat{p}.\hat{q}}{e^2} = \frac{1{,}96^2 . 0{,}017 . 0{,}983}{0{,}004^2} = 4012{,}31 \quad \Rightarrow \quad \mathbf{n \geq 4013}$$

El planteo anterior resulta engañoso en cuanto que se debe utilizar $\hat{p}$ para determinar el tamaño n de la muestra, y luego, al tomar una muestra del tamaño correspondiente, calculamos en ella, $\hat{p}$. Es decir, necesitamos n para calcular $\hat{p}$, y necesitamos $\hat{p}$ para calcular n...
¿Qué se hace en estos casos?
Si se puede hacer una estimación imperfecta de p sin tomar una muestra (por ejemplo, en base a la experiencia obtenida anteriormente), se podría utilizar este valor para $\hat{p}$ y entonces determinar n. Pero a falta de tal estimación, se podría tomar una muestra preliminar de tamaño n, en lo posible $n \geq 30$ para obtener una mejor estimación preliminar de p.

En ocasiones sería impráctico obtener una estimación de p para determinar el tamaño de la muestra con un grado especificado de confianza. Si esto pasa, se establece un límite superior para n al observar que $\hat{p}.\hat{q} = \hat{p}.(1 - \hat{p})$, lo cual debe ser cuando mucho igual que $\frac{1}{4}$, en virtud de que $\hat{p}$ debe tomar un valor entre 0 y 1. Este hecho puede verificarse completando el cuadrado.
De aquí que: $\hat{p}.\hat{q} = \hat{p}.(1 - \hat{p}) = -(\hat{p}^2 - \hat{p}) = \frac{1}{4} - (\hat{p}^2 - \hat{p} + \frac{1}{4}) = \frac{1}{4} - (\hat{p} - \frac{1}{2})^2$ lo cual es siempre menor que $\frac{1}{4}$, excepto cuando $\hat{p} = \frac{1}{2}$ y entonces $\hat{p}.\hat{q} = \frac{1}{4}$. Por lo tanto, si se sustituye n por $\hat{p} = \frac{1}{2}$ en la fórmula usada ante-

Estimación de parámetros

riormente, cuando, en realidad p difiere de ½, entonces n se hará más grande de lo necesario para el grado especificado de confianza y, en consecuencia, aumentará este último, luego la fórmula queda:

$$n = \frac{z_c^2}{4.e^2}$$

Entonces, el tamaño de muestra sin hacer una estimación previa de la proporción, para los datos de nuestro problema, es:

$$n > \frac{z_c^2}{4.e^2} = \frac{1{,}96^2}{4\,.\,0{,}004^2} = 60025 \quad \Rightarrow \quad \boldsymbol{n \geq 60025}$$

- *Al comparar los resultados obtenidos para cada situación puede verse que la información concerniente a p, que se obtuvo en una muestra preliminar, o tal vez por experiencia pasada, permite escoger una muestra más pequeña en tanto se conserve el grado de confianza requerido.*
  Un tamaño de muestra tan grande es el costo de la ignorancia, o sea, es el precio por tener información.

c) ¿A qué conclusión llega si construye un intervalo de confianza unilateral para los datos del punto a?

Para un nivel de confianza del 95%, un intervalo unilateral "inferior" es:

$$\hat{p} - z_c\,.\sqrt{\frac{\hat{p}.\hat{q}}{n}} = 0{,}017 - 1{,}645\,.\sqrt{\frac{0{,}017.(1-0{,}017)}{1000}} = \mathbf{0{,}017 - 0{,}00672}$$

Luego,

$\boldsymbol{p > 0{,}0102}$ $\qquad\qquad$ $\boldsymbol{1 - \alpha = 0{,}95}$

- *Se tiene un 95% de confianza que la proporción real de sacos rotos es superior a 0,0102.*
- *Esto indica que **hay que detener el proceso** porque todos los valores del intervalo de confianza superior para la proporción de sacos rotos son superiores a 0,008.*

d) Suponga que se realizan ajustes en el proceso de producción y que en una segunda muestra aleatoria de 500 sacos envasados se rompen 4. Con base en los datos muestrales, construya un intervalo de confianza bilateral para la diferencia de proporciones correspondientes (antes y después de los cambios), para emitir conclusiones a nivel de confianza del 95%.

Si se desea estimar la diferencia entre dos parámetros binomiales p_1 y p_2, es decir, la diferencia entre dos proporciones, se seleccionan muestras aleatorias independientes de tamaños n_1 y n_2 de las dos poblaciones binomiales con medias $n_1.p_1$ y $n_2.p_2$ y varianzas $n_1.p_1.q_1$ y $n_2.p_2.q_2$, res-

pectivamente, entonces se determinan los números x_1 y x_2, que indican las respectivas cantidades de éxitos en n_1 y n_2 intentos, para formar las proporciones $\hat{p}_1 = x_1 / n_1$ y $\hat{p}_2 = x_2 / n_2$.

Un estimador puntual para la diferencia de proporciones $p_1 - p_2$, está dado por el estadístico $\hat{P}_1 - \hat{P}_2$.

Se sabe que $\hat{P}_1$ y $\hat{P}_2$ están distribuidos en forma aproximadamente normal, con medias p_1 y p_2 y varianzas $p_1.q_1/n_1$ y $p_2.q_2/n_2$, respectivamente. Al seleccionar las muestras independientes de dos poblaciones, las variables $\hat{P}_1$ y $\hat{P}_2$ serán independientes, y entonces, por la propiedad reproductora de la distribución normal, se concluye que $\hat{P}_1 - \hat{P}_2$ está distribuida normalmente con media $p_1 - p_2$ y varianza $p_1.q_1/n_1 + p_2.q_2/n_2$, por lo que se puede asegurar que la variable

$$Z = \frac{\left(\hat{P}_1 - \hat{P}_2\right) - \left(p_1 - p_2\right)}{\sqrt{\dfrac{p_1 \cdot q_1}{n_1} + \dfrac{p_2 \cdot q_2}{n_2}}}$$

estará entre $-z_{\alpha/2}$ y $z_{\alpha/2}$.

A partir de esto, y tomando a $\hat{p}_1$ y $\hat{p}_2$, como estimaciones de p_1 y p_2, y con la condición de que $n_1\hat{p}_1$, $n_2\hat{p}_2$, $n_1\hat{q}_1$ y $n_2\hat{q}_2$ sean todas mayores o iguales que 5, podemos decir que:

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$$

$$P\left(-z_{\alpha/2} < \frac{\left(\hat{P}_1 - \hat{P}_2\right) - \left(p_1 - p_2\right)}{\sqrt{\dfrac{p_1 \cdot q_1}{n_1} + \dfrac{p_2 \cdot q_2}{n_2}}} < z_{\alpha/2}\right) = 1 - \alpha$$

$$P\left((\hat{P}_1 - \hat{P}_2) - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}_1 \cdot \hat{q}_1}{n_1} + \frac{\hat{p}_2 \cdot \hat{q}_2}{n_2}} < p_1 - p_2 < (\hat{P}_1 - \hat{P}_2) + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}_1 \cdot \hat{q}_1}{n_1} + \frac{\hat{p}_2 \cdot \hat{q}_2}{n_2}}\right) = 1 - \alpha$$

Si $\hat{p}_1$ y $\hat{p}_2$ son las proporciones de éxitos en muestras aleatorias de tamaño n_1 y n_2, respectivamente, un intervalo de confianza del $(1 - \alpha)100\%$ para $p_1 - p_2$ es:

$$\left((\hat{p}_1 - \hat{p}_2) - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}_1 \cdot \hat{q}_1}{n_1} + \frac{\hat{p}_2 \cdot \hat{q}_2}{n_2}} \; ; (\hat{p}_1 - \hat{p}_2) + - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}_1 \cdot \hat{q}_1}{n_1} + \frac{\hat{p}_2 \cdot \hat{q}_2}{n_2}} \right)$$

donde $z_{\alpha/2}$ es el valor de z a la derecha del cual se tiene un área de $\alpha/2$.

En nuestro problema:

X_1:"Cantidad de sacos rotos antes de los cambios en el proceso"

Estimación de parámetros

X_2:"Cantidad de sacos rotos después de los cambios en el proceso"
$X_1 \sim Binomial\ (n_1\ ;\ p_1)$
$X_2 \sim Binomial\ (n_2\ ;\ p_2)$
$\hat{P}_1 - \hat{P}_2 \sim Normal\ (\mu = p_1 - p_2\ ;\ \sigma^2 = p_1.q_1/n_1 + p_2.q_2/n_2)$

Según los siguientes datos del problema:

$x_1 = 17 \qquad n_1 = 1000 \qquad \hat{p}_1 = \dfrac{17}{1000} = 0{,}017$

$x_2 = 4 \qquad n_2 = 500 \qquad \hat{p}_2 = \dfrac{4}{500} = 0{,}008$

$1 - \alpha = 0{,}95 \Rightarrow z_{0,05} = \pm 1{,}96$

El intervalo de confianza para la proporción poblacional para el 95%, es:

$$(\hat{p}_1 - \hat{p}_2) \pm z_c . \sqrt{\frac{\hat{p}_1.\hat{q}_1}{n_1} + \frac{\hat{p}_2.\hat{q}_2}{n_2}}$$

$$(0{,}017 - 0{,}008) \pm 1{,}96.\sqrt{\frac{0{,}017\ .0{,}983}{1000} + \frac{0{,}008\ .0{,}992}{500}} = \mathbf{0{,}009 \pm 0{,}0112}$$

Luego,

$\mathbf{-\ 0{,}0022 < p_1 - p_2 < 0{,}0201} \qquad\qquad \mathbf{1 - \alpha = 0{,}95}$

🗣 *Se tiene un 95% de confianza, en base a la evidencia muestral, que el intervalo (-0,0022 ; 0,0201) contiene a la verdadera diferencia de proporciones de las poblaciones.*

14. En una muestra aleatoria de 500 clientes de una empresa local se encontró que 340 se habían suscrito al servicio de correo electrónico y 160 al de correo electrónico e Internet.

a) Construya un intervalo de confianza del 95% para la proporción actual de clientes suscriptos al correo electrónico de esa empresa de servicios.

X:"Cantidad de clientes suscriptos al correo electrónico"
$X \sim Binomial\ (n\ ;\ p)$
$\hat{P} \sim Normal\ (\mu = p\ ;\ \sigma^2 = p.q/n)$

Según los siguientes datos del problema:

$x = 340$
$n = 500$
$\hat{p} = \dfrac{340}{500} = 0{,}68$

$1 - \alpha = 0{,}95 \Rightarrow z_{0,05} = \pm 1{,}96$

Estimación de parámetros

El intervalo de confianza para la proporción poblacional para el 95%, es:

$$\hat{p} \pm z_c \cdot \sqrt{\frac{\hat{p}.\hat{q}}{n}} = 0,68 \pm 1,96 \cdot \sqrt{\frac{0,68.(1-0,68)}{500}} = \mathbf{0,68 \pm 0,041}$$

Luego,

$$\mathbf{0,6391 < p < 0,7209} \qquad \mathbf{1 - \alpha = 0,95}$$

- *Se tiene un 95% de confianza, en base a la evidencia muestral, que el intervalo (0,6391 ; 0,7209) contiene a la proporción real de clientes suscriptos al servicio de correo electrónico.*

b) ¿Qué puede decir acerca del error cometido en la estimación, con un 95% de confianza?

El error cometido en la estimación será inferior a 0,041 con una confianza del 95%.

c) A partir de los resultados obtenidos de la muestra preliminar ¿qué tan grande se requiere que sea la muestra, si se desea tener una confianza del 95% de que la estimación de p estará dentro de 0,02?

*Si **se utiliza** $\hat{p}$ como una estimación de p, se puede tener una confianza del (1 - α)100% de que el error será menor que una cantidad especificada e cuando el tamaño de la muestra sea aproximadamente*

$$n = \frac{z_c^2.\hat{p}.\hat{q}}{e^2}$$

Entonces, el tamaño de muestra haciendo una estimación previa de la proporción, de acuerdo a los datos del problema, es:

$$e = z_c \cdot \sqrt{\frac{\hat{p}.\hat{q}}{n}} = 1,96 \cdot \sqrt{\frac{0,68.(1-0,68)}{n}} < \mathbf{0,02} \quad \Rightarrow$$

$$\Rightarrow \quad n > \frac{z_c^2.\hat{p}.\hat{q}}{e^2} = \frac{1,96^2.0,68.0,32}{0,02^2} = 2089,83 \quad \Rightarrow \quad \mathbf{n \geq 2090}$$

d) Asumiendo que no se dispone de una muestra preliminar para obtener información acerca de p, ¿qué tan grande se requiere que sea la muestra si se quiere tener una confianza al menos del 95% de que la estimación de p estará dentro de 0,02?

*Si **no se utiliza** o no se conoce un valor de $\hat{p}$ como una estimación de p, se puede tener una confianza del (1 - α)100% de que el error será menor que una cantidad especificada e cuando el tamaño de la muestra sea aproximadamente*

$$n = \frac{z_c^2}{4.e^2}$$

Estimación de parámetros

Entonces, el tamaño de muestra sin hacer una estimación previa de la proporción, según los datos del problema, es:

$$n > \frac{z_c^2}{4.e^2} = \frac{1{,}96^2}{4\,.\,0{,}02^2} = 2401 \quad \Rightarrow \quad \mathbf{n \geq 2401}$$

15. Se realiza un estudio para determinar la proporción de residentes en una ciudad y en sus suburbios que están a favor de la construcción de una planta de energía nuclear.

a) ¿Qué tan grande debe ser una muestra si se requiere una confianza al menos del 95% de que la estimación estará dentro del 0,04 de la proporción real de residentes de esa ciudad y sus suburbios que están a favor de la construcción de la planta de energía nuclear?

X:"Cantidad de residentes que están a favor de la construcción de la planta de energía nuclear"

$X \sim$ *Binomial* $(n\,;\,p)$

$\hat{P} \sim$ *Normal* $(\mu = p\,;\,\sigma^2 = p.q/n)$

Según los siguientes datos del problema:

$e = 0{,}04$

$1 - \alpha = 0{,}95 \quad \Rightarrow \quad z_{0,05} = \pm\, 1{,}96$

Entonces, el tamaño de muestra sin hacer una estimación previa de la proporción, para los datos del problema, es:

$$n > \frac{z_c^2}{4.e^2} = \frac{1{,}96^2}{4\,.\,0{,}04^2} = 600{,}25 \quad \Rightarrow \quad \mathbf{n \geq 601}$$

b) A falta de información previa, se tomó una muestra preliminar de tamaño 40 (observe que la muestra es mayor que treinta), resultando que 6 residentes estuvieron a favor de la construcción. Utilice esta estimación imperfecta para determinar en forma aproximada cuántas observaciones se necesitan para proporcionar el grado de precisión deseado en el punto anterior.

Según los siguientes datos del problema:

$x = 6$

$n = 40$

$$\hat{p} = \frac{6}{40} = 0{,}15$$

Entonces, el tamaño de muestra haciendo una estimación previa de la proporción, es:

$$e = z_c\,.\sqrt{\frac{\hat{p}.\hat{q}}{n}} = 1{,}96\,.\sqrt{\frac{0{,}68.(1-0{,}68)}{n}} < \mathbf{0{,}04} \qquad \Rightarrow$$

Estimación de parámetros

$$\Rightarrow \quad n > \frac{z_c^2 . \hat{p} . \hat{q}}{e^2} = \frac{1{,}96^2 . 0{,}15 . 0{,}85}{0{,}04^2} = 306{,}13 \quad \Rightarrow \quad n \geq 307$$

16. Se analiza la fracción de productos defectuosos producidos por dos líneas de producción. Una muestra aleatoria de 100 unidades provenientes de la línea 1 contiene 10 defectuosas, mientras que una muestra aleatoria de 120 unidades de la línea 2 tiene 25 que son defectuosas. Encuentre un intervalo de confianza del 99% para la diferencia en fracciones de productos defectuosos producidos por las dos líneas.

X_1:"Cantidad de productos defectuosos provenientes de la línea 1"
X_2:"Cantidad de productos defectuosos provenientes de la línea 2"
$X_1 \sim$ Binomial $(n_1 ; p_1)$
$X_2 \sim$ Binomial $(n_2 ; p_2)$
$\hat{P}_1 - \hat{P}_2 \sim$ Normal $(\mu = p_1 - p_2 \; ; \sigma^2 = p_1.q_1/n_1 + p_2.q_2/n_2)$

Según los siguientes datos del problema:

$x_1 = 10 \qquad n_1 = 100 \qquad \hat{p}_1 = \frac{10}{100} = 0{,}1$

$x_2 = 25 \qquad n_2 = 120 \qquad \hat{p}_2 = \frac{25}{120} = 0{,}2083... \approx 0{,}21$

$1 - \alpha = 0{,}99 \quad \Rightarrow \quad z_{0,005} = \pm 2{,}58$

El intervalo de confianza para la proporción poblacional para el 99%, es:

$$(\hat{p}_1 - \hat{p}_2) \pm z_c . \sqrt{\frac{\hat{p}_1 . \hat{q}_1}{n_1} + \frac{\hat{p}_2 . \hat{q}_2}{n_2}}$$

$$(\hat{p}_1 - \hat{p}_2) \pm z_c . \sqrt{\frac{\hat{p}_1 . \hat{q}_1}{n_1} + \frac{\hat{p}_2 . \hat{q}_2}{n_2}}$$

$$(0{,}10 - 0{,}21) \pm 2{,}58 . \sqrt{\frac{0{,}10 \, . \, 0{,}90}{100} + \frac{0{,}21 \, . \, 0{,}79}{120}} = -\mathbf{0{,}11 \pm 0{,}1233}$$

Luego,

$$\mathbf{-0{,}2333 < p_1 - p_2 < 0{,}0133} \qquad \mathbf{1 - \alpha = 0{,}99}$$

- *Se tiene un 99% de confianza, en base a la evidencia muestral, que el intervalo (-0,2333 ; 0,0133) contiene a la verdadera diferencia de proporciones de las poblaciones.*
- *Como este intervalo incluye al cero, podemos decir que no hay diferencia significativa entre las proporciones de unidades defectuosas en cada línea de producción.*

17. Se estudia la trabajabilidad de un hormigón plástico compactado mediante vibración normal. De acuerdo a las especificaciones del Reglamento, el mismo debe encuadrarse en un ámbito de consistencia A-2, de aspecto levemente cohesivo. Para ello se ajusta el proyecto del hormigón de modo de obtener un asentamiento del tronco cono de 8 cm, y de experiencias anteriores se sabe que la precisión del equipo de trabajo permite lograr una variabilidad de los resultados cuantificada por una desviación estándar de 1,5 cm. Se acepta también que el asentamiento medido está distribuido normalmente. Los resultados obtenidos de una muestra ensayada, medidos en cm, son los siguientes: 7,5; 8,0; 8,0; 7,5; 7,0; 8,5; 8,0; 9,0; 7,0; 8,0.

a) Estime mediante un intervalo de confianza del 95% el asentamiento medio del hormigón estudiado.

X:"Asentamiento del hormigón plástico"

$X \sim$ *Normal* $(\mu ; \sigma)$

$\overline{X} \sim$ *Normal* $(\mu_{\overline{x}} = \mu \; ; \; \sigma_{\overline{x}} = \sigma / \sqrt{n})$ *Por provenir la muestra de una distribución normal*

Según los siguientes datos del problema:

$\overline{x} = 7{,}85$ cm

$\sigma = 1{,}5$ cm

$n = 10$

$1 - \alpha = 0{,}95 \Rightarrow z_{\alpha/2} = \pm 1{,}96$

Los límites de confianza del 95% son :

$$\overline{x} \pm z_c \cdot \frac{\sigma}{\sqrt{n}} = 7{,}85 \text{ cm} \pm 1{,}96 \cdot \frac{1{,}5 \text{ cm}}{\sqrt{10}} = \mathbf{7{,}85 \text{ cm} \pm 0{,}93 \text{ cm}}$$

Luego,

$\mathbf{6{,}92 \text{ cm} < \mu < 8{,}78 \text{ cm}}$ $\qquad\qquad$ $\mathbf{1 - \alpha = 0{,}95}$

🗣 *Se tiene un 95% de confianza, a partir de los datos de la muestra, que el intervalo (6,92 cm ; 8,78 cm) contiene a la media poblacional.*

b) ¿Qué podemos asegurar con una probabilidad de 0,95 sobre la medida del error máximo de la estimación?

$$e = z_c \cdot \frac{\sigma}{\sqrt{n}} = 1{,}96 \cdot \frac{1{,}5 \text{ cm}}{\sqrt{10}} < \mathbf{0{,}93 \text{ cm}}$$

🗣 *Se tiene un 95% de confianza de que el error máximo de la estimación sea de 0,93 cm .*

c) ¿Qué tamaño debería tener la muestra para poder tener una confianza del 99% que el error máximo de la estimación sea de ½ cm?

Según los siguientes datos del problema:

$\overline{x} = 7{,}85$ cm

$\sigma = 1{,}5$ cm

Estimación de parámetros

$e = \frac{1}{2}$ cm
$1 - \alpha = 0{,}99 \Rightarrow z_{\alpha/2} = \pm 2{,}58$

Los límites de confianza del 99% son : $\bar{x} \pm 2{,}58 \cdot \frac{\sigma}{\sqrt{n}}$ siendo el error de la estimación $2{,}58 \cdot \frac{\sigma}{\sqrt{n}}$.

Se desea que el error sea inferior a $\frac{1}{2}$ cm, o sea, $2{,}58 \cdot \frac{1{,}5}{\sqrt{n}} < 0{,}5$ cm

entonces, $n > \left(\frac{2{,}58 \cdot 1{,}5}{0{,}5}\right)^2 = 59{,}91$, por lo que, ***n ≥ 60***

- *Así, pues, se puede tener una confianza del 99% de que el error de la estimación será menor que $\frac{1}{2}$ cm si se toma una muestra de tamaño 60 o mayor.*

18. Se realizó un estudio para determinar si un aditivo agregado en la dosificación de hormigones para acelerar el tiempo de fragüe tenía alguna influencia en la resistencia a la compresión del mismo. Una muestra aleatoria de 62 probetas con aditivo ensayadas a los 7 días dan una resistencia media de 70,5 kg/cm^2, con desviación 6,5 kg/cm^2, en tanto que las 60 probetas sin aditivo, ensayadas a la misma edad, dieron una resistencia media de 86,8 kg/cm^2, con desviación 7,2 kg/cm^2. Suponga que de estudios anteriores se conocen las desviaciones poblacionales y son 5 y 5,5 kg/cm^2 para las probetas con y sin aditivo, respectivamente.

a) A nivel de confianza del 95%, ¿qué puede decir respecto de la influencia del aditivo en la resistencia del hormigón?

X_1:"Resistencia de los hormigones con aditivo"
X_2:"Resistencia de los hormigones sin aditivo"
$X_1 \sim$ Desconocida (con varianza poblacional conocida)
$X_2 \sim$ Desconocida (con varianza poblacional conocida)
$\bar{X}_1 - \bar{X}_2 \sim$ Normal ($\mu_{\bar{x}_1-\bar{x}_2} = \mu_1 - \mu_2$; $\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$) Por tener tamaños de muestras grandes

Según los siguientes datos del problema:

$\bar{x}_1 = 70{,}5$ kg/cm^2	$\sigma_1 = 5$ kg/cm^2	$n_1 = 62$
$\bar{x}_2 = 86{,}8$ kg/cm^2	$\sigma_2 = 5{,}5$ kg/cm^2	$n_2 = 60$

$1 - \alpha = 0{,}95 \Rightarrow z_{\alpha/2} = \pm 1{,}96$

Los límites de confianza del 95% son :

Estimación de parámetros

$$(\overline{x}_1 - \overline{x}_2) \pm z_c \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = (70{,}5 - 86{,}8) \pm 1{,}96 \sqrt{\frac{5^2}{62} + \frac{5{,}5^2}{60}} = \mathbf{-16{,}3 \pm 1{,}87\ kg/cm^2}$$

Luego,

$$\mathbf{-18{,}17\ kg/cm^2 < \mu_1 - \mu_2 < -14{,}43\ kg/cm^2} \qquad \mathbf{1 - \alpha = 0{,}95}$$

- *Se tiene un 95% de confianza, según la evidencia muestral, que el intervalo (-18,17 kg/cm^2 ; -14,43 kg/cm^2) contiene a la diferencia de las medias en la resistencia de los hormigones.*
- *Esto indica que* ***hay diferencia significativa*** *entre las resistencias medias de los hormigones usando o no aditivos para acelerar el tiempo de fragüe.*

b) ¿Qué conclusión obtendría si desconoce las desviaciones poblacionales pero las supone iguales?

X_1:"Resistencia de los hormigones con aditivo"
X_2:"Resistencia de los hormigones sin aditivo"
$X_1 \sim$ Desconocida
$X_2 \sim$ Desconocida

$\overline{X}_1 - \overline{X}_2 \sim$ *Normal* $\left(\mu_{\overline{X}_1-\overline{X}_2} = \mu_1 - \mu_2 \,;\, \hat{\sigma}_{\overline{X}_1-\overline{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right)$ *Por tener tamaños de muestras grandes*

Nota: Aunque no se conozcan las varianzas poblacionales no podemos utilizar la distribución t porque los tamaños de muestra son grandes y, además, no se sabe si las variables tienen distribución normal.

Según los siguientes datos del problema:

$\overline{x}_1 = 70{,}5\ kg/cm^2$	$s_1 = 6{,}5\ kg/cm^2$	$n_1 = 62$
$\overline{x}_2 = 86{,}8\ kg/cm^2$	$s_2 = 7{,}2\ kg/cm^2$	$n_2 = 60$

$1 - \alpha = 0{,}95 \;\Rightarrow\; z_{\alpha/2} = \pm 1{,}96$

Los límites de confianza del 95% son :

$$(\overline{x}_1 - \overline{x}_2) \pm z_c \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = (70{,}5 - 86{,}8) \pm 1{,}96 \sqrt{\frac{6{,}5^2}{62} + \frac{7{,}2^2}{60}} =$$

$$= \mathbf{-16{,}3 \pm 2{,}44\ kg/cm^2}$$

Luego,

$$\mathbf{-18{,}7366\ kg/cm^2 < \mu_1 - \mu_2 < -13{,}8634\ kg/cm^2} \qquad \mathbf{1 - \alpha = 0{,}95}$$

- *Se tiene un 95% de confianza, según la evidencia muestral, que el intervalo (-18,7325 kg/cm^2 ; -13,8675 kg/cm^2) contiene a la diferencia de las medias poblacionales.*
- *Esto indica que* ***hay una diferencia significativa*** *entre las resistencias medias de los hormigones usando o no aditivos para acele-*

rar el tiempo de fragüe. En particular se observa que el tiempo es significativamente menor si se agrega un aditivo.

c) ¿Qué conclusión obtendría si desconoce las desviaciones poblacionales pero las supone distintas?

Lo mismo que en el inciso b porque no se puede utilizar la distribución t por lo indicado anteriormente.

19. Se compara la resistencia de dos tipos de rosca para bulones utilizados en construcciones metálicas. Se prueban 20 piezas de las roscas tipo A y 18 de las del tipo B, en condiciones similares. Las piezas del tipo A dan una resistencia media de 78,3 kg, con desviación estándar 5,6 kg. Las del tipo B dan una resistencia media de 76,5 kg, con desviación estándar 6,3 kg. Se conoce de experiencias anteriores que las desviaciones poblacionales son 5 y 5,5 kg, para las roscas del tipo A y B, respectivamente, y que provienen de poblaciones normales.

a) ¿Hay suficientes razones para concluir a nivel del 95% que la resistencia media de las roscas del tipo A es más elevada que las del tipo B?

X_A:"Resistencia de las roscas tipo A"

X_B:"Resistencia de las roscas tipo B"

$X_A \sim Normal\ (\mu_A\ ;\ \sigma_A)$

$X_B \sim Normal\ (\mu_B\ ;\ \sigma_B)$

$\overline{X}_A - \overline{X}_B \sim Normal\ (\mu_{\overline{X}_A - \overline{X}_B} = \mu_A - \mu_B\ ;\ \sigma_{\overline{X}_A - \overline{X}_B} = \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}\ )$ *Por provenir de muestras normales*

Según los siguientes datos del problema:

$\overline{x}_A = 78,3\ kg$ $\quad \sigma_A = 5\ kg$ $\quad n_A = 20$

$\overline{x}_B = 76,5\ kg$ $\quad \sigma_B = 5,5\ kg$ $\quad n_B = 18$

$1-\alpha = 0,90 \quad \Rightarrow \quad z_{\alpha/2} = \pm\, 1,64$

Los límites de confianza del 90% son :

$$(\overline{x}_1 - \overline{x}_2) \pm z_c \cdot \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} = (78,3 - 76,5) \pm 1,64 \cdot \sqrt{\frac{5^2}{20} + \frac{5,5^2}{18}} =$$

$= \mathbf{1,8 \pm 2,8075\ kg}$

Luego,

$\mathbf{-\ 1,0075\ kg < \mu_A - \mu_B < 4,6075\ kg} \qquad \mathbf{1 - \alpha = 0,95}$

🗣 *Se tiene un 95% de confianza, según la evidencia muestral, que el intervalo (- 1,0075 kg ; 4,6075 kg) contiene a la diferencia de las medias poblacionales.*

🗣 *Esto indica que no hay una diferencia significativa entre las resistencias medias de los dos tipos de roscas.*

b) ¿Cuál es la conclusión, si no conoce los parámetros poblacionales?

No se conocen los parámetros poblacionales que son: la media y la varianza.

La media no se conoce porque es justamente la incógnita del problema, estamos estimando la media de la variable diferencia.

El desconocer la varianza, estando ante muestras pequeñas provenientes de poblaciones normales, supone el uso del estadístico t, aunque en distintas condiciones. Las condiciones son las siguientes:

- *No se conocen las varianzas poblacionales pero se suponen* ***iguales****:*

  X_A:"Resistencia de las roscas tipo A"

  X_B:"Resistencia de las roscas tipo B"

  $X_A \sim Normal\ (\mu_A\ ;\ \sigma_A)$

  $X_B \sim Normal\ (\mu_B\ ;\ \sigma_B)$

$$\frac{(\overline{X}_A - \overline{X}_B) - (\mu_A - \mu_B)}{S_p . \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \sim t \text{ con } n_A + n_B - 2 \text{ grados de libertad}$$

Según los datos del problema:

$\overline{x}_A = 78{,}3$ kg $s_A = 5{,}6$ kg $n_A = 20$

$\overline{x}_B = 76{,}5$ kg $s_B = 6{,}3$ kg $n_B = 18$

$1 - \alpha = 0{,}95 \Rightarrow t_{0{,}025\ ;\ 20+18-2} = \pm\ 2{,}0282$ (por interpolación)

Los límites de confianza del 95% son :

$$(\overline{x}_A - \overline{x}_B) \pm t_{0{,}025;36} . S_p \sqrt{\frac{1}{n_A} + \frac{1}{n_B}} \quad \text{siendo } S_p^2 = \frac{(n_A - 1)\, s_A^2 + (n_B - 1) s_B^2}{n_A + n_B - 2}$$

$$(78{,}3 - 76{,}5) \pm 2{,}0282.\ S_p . \sqrt{\frac{1}{20} + \frac{1}{18}}$$

$$\text{siendo } S_p^2 = \frac{(20 - 1)\ 5{,}6^2 + (18 - 1)\ 6{,}3^2}{20 + 18 - 2} = 35{,}2936 \text{ kg}^2$$

$$(78{,}3 - 76{,}5) \pm 2{,}0282.\ 5{,}9408 . \sqrt{\frac{1}{20} + \frac{1}{18}} = \mathbf{1{,}8 \pm 3{,}9147 \text{ kg}}$$

Luego,

$$\mathbf{-\ 2{,}1147 \text{ kg} < \mu_A - \mu_B < 5{,}7147 \text{ kg}} \qquad \mathbf{1 - \alpha = 0{,}95}$$

🗣 *Se tiene un 95% de confianza que la diferencia de las medias poblacionales se encuentren entre - 0,5878 kg y 4,1878 kg. Esto indica que no hay una diferencia significativa entre las resistencias medias de los dos tipos de roscas.*

- *No se conocen las varianzas poblacionales pero se suponen **distintas**:*
  X_A:"Resistencia de las roscas tipo A"
  X_B:"Resistencia de las roscas tipo B"
  $X_A \sim$ Normal $(\mu_A ; \sigma_A)$
  $X_B \sim$ Normal $(\mu_B ; \sigma_B)$

  $$\frac{(\overline{X}_A - \overline{X}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}} \sim t \text{ con } \nu \text{ grados de libertad}$$

  $$\text{siendo } \nu = \frac{(s_A^2/n_A + s_B^2/n_B)^2}{\frac{(s_A^2/n_A)^2}{n_A - 1} + \frac{(s_B^2/n_B)^2}{n_B - 1}}$$

  Según los datos del problema:
  $\overline{x}_A = 78{,}3$ kg $\quad s_A = 5{,}6$ kg $\quad n_A = 20$
  $\overline{x}_B = 76{,}5$ kg $\quad s_B = 6{,}3$ kg $\quad n_B = 18$

  $$\nu = \frac{(5{,}6^2/20 + 6{,}3^2/18)^2}{\frac{(5{,}6^2/20)^2}{20-1} + \frac{(6{,}3^2/18)^2}{18-1}} = 34{,}2692 \Rightarrow \nu = 34$$

  $1 - \alpha = 0{,}95 \Rightarrow t_{0,025\,;\,\nu=34} = \pm\, 2{,}032$ *(por interpolación)*

  Los límites de confianza del 95% son :

  $$(\overline{x}_A - \overline{x}_B) \pm t_{0,025\,;\,\nu=34} \cdot \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$$

  $$(78{,}3 - 76{,}5) \pm 2{,}032 . \sqrt{\frac{5{,}6^2}{20} + \frac{6{,}3^2}{18}} = \mathbf{1{,}8 \pm 3{,}9470 \text{ kg}}$$

  Luego,
  $$\mathbf{-\,2{,}1470 \text{ kg} < \mu_A - \mu_B < 5{,}7470 \text{ kg}} \qquad \mathbf{1 - \alpha = 0{,}95}$$

  - *Se tiene un 95% de confianza, en base a la evidencia muestral, que el intervalo (-2,1470 kg ; 5,7470 kg) contiene a la diferencia de las medias poblacionales.*
  - *Esto indica que no hay una diferencia significativa entre las resistencias medias de los dos tipos de roscas.*

20. El artículo "Mix Design for Optimal Strength Development of Fly Ash Concrete" (*Cement and Concrete Research, 1989, Vol. 19, Nº 4, pp. 634-640*) investiga la resistencia a la compresión del hormigón cuando se mezcla con ceniza muy fina (una mezcla de silicatos, alúmina, hierro, óxido de magnesio y otros componentes). Se sabe que la resistencia se comporta de manera

Estimación de parámetros

aproximadamente normal. La resistencia a la compresión, en MPa, de nueve muestras de hormigón en condiciones secas, a la edad de 28 días, son las siguientes: 40,2; 30,4; 28,9; 30,5; 22,4; 25,8; 18,4; 14,2; 15,3.

a) Encuentre un intervalo de confianza inferior del 99% para la resistencia a la compresión promedio.

X:"Resistencia a la compresión del hormigón cuando se mezcla con ceniza muy fina"

$X \sim Normal\ (\mu\ ;\ \sigma)$

$$\frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t \text{ con n-1 grados de libertad}$$

Según los siguientes datos del problema:

$\overline{x} = 25{,}12$ MPa

$s = 8{,}4203$ Mpa

$n = 9$

$1 - \alpha = 0{,}99 \Rightarrow t_{0{,}01;8} = +\ 2{,}896$

El límite del intervalo "inferior" para una confianza del 99% es :

$$\overline{x} - t_c \cdot \frac{s}{\sqrt{n}} = \overline{x} + t_{0{,}001{,}8} \cdot \frac{s}{\sqrt{n}} = 25{,}12 - 2{,}896 \cdot \frac{8{,}4203}{\sqrt{9}} =$$

$= \mathbf{25{,}12 - 8{,}1284}$ **MPa**

Luego,

$\mu >$ **16,9916 MPa** $\qquad\qquad$ $\mathbf{1 - \alpha = 0{,}99}$

🗣 *Se tiene un 99% de confianza, a partir de los datos de la muestra, que el intervalo terminal de límite inferior 16,99 Mpa, contiene a la media poblacional.*

b) Encuentre un intervalo de confianza bilateral del 95% para la resistencia a la compresión promedio.

X:"Resistencia a la compresión del hormigón cuando se mezcla con ceniza muy fina"

$X \sim Normal\ (\mu\ ;\ \sigma)$

$$\frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t \text{ con n-1 grados de libertad}$$

Según los siguientes datos del problema:

$\overline{x} = 25{,}12$ MPa

$s = 8{,}4203$ Mpa

$n = 9$

$1-\alpha = 0{,}95 \Rightarrow t_{0,025;8} = \pm 2{,}306$

Los límites de confianza del 95% son :

$$\bar{x} \pm t_c \cdot \frac{s}{\sqrt{n}} = \bar{x} \pm t_{0,025,8} \cdot \frac{s}{\sqrt{n}} = 25{,}12 \pm 2{,}306 \cdot \frac{8{,}4203}{\sqrt{9}} =$$

$$= \mathbf{25{,}12 \pm 6{,}4724\ MPa}$$

Luego,

$$\mathbf{18{,}6476\ MPa < \mu < 31{,}5924\ MPa \qquad\qquad 1-\alpha = 0{,}95}$$

🗣 *Se tiene un 95% de confianza, a partir de los datos de la muestra, que el intervalo (18,65 MPa ; 31,59 MPa) contiene a la media poblacional.*

21. Se investiga el diámetro de las barras de acero del tipo F-22 utilizadas en construcciones livianas de acero de nuestro medio, fabricadas en dos diferentes máquinas de extrusión. Para ello se toman dos muestras aleatorias.
De la primera máquina los resultados obtenidos de 15 mediciones dan un diámetro medio de 8,73 mm con varianza 0,35 mm^2, mientras que para la segunda máquina, de 18 mediciones el diámetro medio y la varianza son 8,68 mm y 0,40 mm^2, respectivamente. Suponiendo que los diámetros de las poblaciones de barras producidos por ambas máquinas están distribuidos normalmente y que es posible aceptar que tienen la misma varianza, construya un intervalo de confianza bilateral del 95% para la diferencia en el diámetro promedio de la barra.

X_1:"Diámetro de las barras hechas por la máquina 1"
X_2:"Diámetro de las barras hechas por la máquina 2"
$X_1 \sim Normal\ (\mu_1 ; \sigma_1)$
$X_2 \sim Normal\ (\mu_2 ; \sigma_2)$

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t \text{ con } n_1 + n_2 - 2 \text{ grados de libertad}$$

Según los datos del problema:

$\bar{x}_1 = 8{,}73\ mm \qquad s_1^2 = 0{,}35\ mm^2 \qquad n_1 = 15$
$\bar{x}_2 = 8{,}68\ mm \qquad s_2^2 = 0{,}40\ mm^2 \qquad n_2 = 18$
$1-\alpha = 0{,}95 \Rightarrow t_{0,025\,;\,15+18-2} = \pm 2{,}040$

Los límites de confianza del 95% son :

$$(\bar{x}_1 - \bar{x}_2) \pm t_{0,025\,;\,31} \cdot S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \qquad \text{siendo } S_p^2 = \frac{(n_1-1)\,s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

$$(8{,}73 - 8{,}68) \pm 2{,}04 \,.\, S_p \,.\, \sqrt{\frac{1}{15} + \frac{1}{18}}$$

$$\text{siendo } S_p^2 = \frac{(15-1)\,0{,}35 + (18-1)\,0{,}40}{15 + 18 - 2} = 0{,}3774$$

$$(8{,}73 - 8{,}68) \pm 2{,}04 .\ 0{,}6143 \,.\, \sqrt{\frac{1}{15} + \frac{1}{18}} = \mathbf{0{,}05 \pm 0{,}4381\ mm}$$

Luego,

$\mathbf{-0{,}3881\ mm < \mu_1 - \mu_2 < 0{,}4881\ mm}$ $\qquad\qquad$ $\mathbf{1 - \alpha = 0{,}95}$

- *Se tiene un 95% de confianza, en base a la evidencia muestral, que el intervalo (-0,3881 mm ; 0,4881 mm) contiene a la diferencia de las medias poblacionales.*
- *Esto indica que* ***no hay una diferencia significativa*** *entre los diámetros medios de las barras de acero hechos por las dos máquinas.*

22. Un artículo publicado en Fire Technology investigó dos agentes dispersores de espuma que pueden emplearse en las boquillas de los equipos extinguidores de fuego. Al tomar una muestra aleatoria de cinco observaciones con una espuma que forma una película acuosa (AFFF), se obtuvo una media muestral de 4,7 y una desviación estándar de 0,6. Una muestra aleatoria de cinco observaciones con concentrados de tipo alcohólico (ATC) tuvo una media muestral de 6,9 y una desviación estándar de 0,8. Encuentre un intervalo de confianza del 95% para la diferencia en la dispersión de espuma promedio de estos dos agentes. ¿Puede obtenerse alguna conclusión sobre qué agente produce la mayor dispersión de espuma? Suponga que ambas poblaciones están bien representadas por distribuciones normales que tienen las mismas desviaciones estándar.

X_1:"Dispersión de la espuma que forma una película acuosa (AFFF)"
X_2:" Dispersión de la espuma con concentrados de tipo alcohólico (ATC)"
$X_1 \sim Normal\ (\mu_1 ; \sigma_1)$
$X_2 \sim Normal\ (\mu_2 ; \sigma_2)$

$$\frac{\left(\overline{X}_1 - \overline{X}_2\right) - (\mu_1 - \mu_2)}{S_p . \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t \text{ con } n_1 + n_2 - 2 \text{ grados de libertad}$$

Según los datos del problema:

$\overline{x}_1 = 4{,}7 \qquad s_1 = 0{,}6 \qquad n_1 = 5$
$\overline{x}_2 = 6{,}9 \qquad s_2 = 0{,}8 \qquad n_2 = 5$
$1 - \alpha = 0{,}95 \Rightarrow t_{0{,}025\,;\,5+5-2} = \pm\, 2{,}306$

Estimación de parámetros

Los límites de confianza del 95% son :

$$(\overline{x}_1 - \overline{x}_2) \pm t_{0,025\,;\,8}\,.S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \qquad \text{siendo } S_p^2 = \frac{(n_1-1)\,s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$$

$$(4{,}7-6{,}9) \pm 2{,}306.\ S_p\,.\sqrt{\frac{1}{5}+\frac{1}{5}} \quad \text{siendo } S_p^2 = \frac{(5-1)\,0{,}6^2+(5-1)\,0{,}8^2}{5+5-2} = 0{,}5$$

$$(4{,}7-6{,}9) \pm 2{,}306.\ 0{,}7071\,.\sqrt{\frac{1}{5}+\frac{1}{5}} = \mathbf{-2{,}2 \pm 1{,}0313}$$

Luego,

$$\mathbf{-3{,}2313 < \mu_1 - \mu_2 < -1{,}1687} \qquad\qquad \mathbf{1 - \alpha = 0{,}95}$$

- *Se tiene un 95% de confianza, en base a la evidencia muestral, que el intervalo (-3,2313 ; -1,1687) contiene a la diferencia de las medias poblacionales.*
- *Esto indica que **hay una diferencia significativa** entre las dispersiones medias. En particular se observa que el agente dispersor ATC produce mayor dispersión de espuma.*

23. Se estudia la trabajabilidad de un hormigón plástico compactado mediante vibración normal. De acuerdo a las especificaciones del Reglamento, el mismo debe encuadrarse en un ámbito de consistencia A-2, de aspecto levemente cohesivo. Para ello se ajusta el proyecto del hormigón de modo de obtener un asentamiento del tronco cono de 8 cm, y de experiencias se acepta que el asentamiento medido está distribuido normalmente. Los resultados obtenidos de una muestra ensayada, medidos en cm, son los siguientes: 7,5; 8,0; 8,0; 7,5; 7,0; 8,5; 8,0; 9,0; 7,0; 8,0. Estime mediante un intervalo de confianza del 99%, la desviación estándar y la varianza del asentamiento del hormigón estudiado.

X:"Asentamiento del hormigón plástico"

$X \sim Normal\ (\mu\ ;\ \sigma)$

$(n-1)\,.\,\dfrac{s^2}{\sigma^2} \sim \chi^2$ *con n-1 grados de libertad*

Según los siguientes datos del problema:

$\overline{x} = 7{,}85$ cm

$s = 1{,}5$ cm

$n = 10$

$1 - \alpha = 0{,}95 \quad \Rightarrow \quad \chi^2_{0{,}025;9} = 19{,}023 \quad y \quad \chi^2_{0{,}975;9} = 2{,}700$

El intervalo de confianza del 95% , está dado por:

$$\frac{s^2.(n-1)}{\chi^2_{0,025;9}} < \sigma^2 < \frac{s^2.(n-1)}{\chi^2_{0,975;9}}$$

$$\frac{2,25.(10-1)}{19,023} < \sigma^2 < \frac{2,25.(10-1)}{2,700}$$

$$\mathbf{1,0645 < \sigma^2 < 7,500} \qquad \mathbf{1 - \alpha = 0,95}$$

$$\mathbf{1,0317 < \sigma < 2,7386} \qquad \mathbf{1 - \alpha = 0,95}$$

- *Se tiene un 95% de confianza, a partir de los datos de la muestra, que el intervalo (1,0645 cm^2 ; 7,5 cm^2) contiene a la varianza poblacional.*
- *Se tiene un 95% de confianza, a partir de los datos de la muestra, que el intervalo (1,0317 cm ; 2,7386 cm) contiene a la desviación estándar poblacional.*

24. En un experimento se comparó el consumo de combustible para dos tipos de camiones con motor diesel, en condiciones similares. Doce vehículos del tipo A y diez del tipo B se probaron a 90 km/h. Si los doce camiones promediaron 16 km/l con desviación 1 km/l y los diez camiones consumieron 11 km/l con desviación 1,8 km/l.

Suponiendo normalidad:

a) Construya un intervalo de confianza del 98% para el cociente de las varianzas de los recorridos por litro de los vehículos tipo A y B. ¿Qué conclusión obtiene?

X_A:"Rendimiento del combustible en camiones tipo A"

X_B:"Rendimiento del combustible en camiones tipo B"

$X_A \sim Normal\ (\mu_A\ ;\ \sigma_A)$

$X_B \sim Normal\ (\mu_B\ ;\ \sigma_B)$

$$\frac{S_A^2/\sigma_A^2}{S_B^2/\sigma_B^2} = \frac{\sigma_B^2}{\sigma_A^2}.\frac{S_A^2}{S_B^2} \sim F \text{ con } n_A - 1 \text{ y } n_B - 1 \text{ grados de libertad}$$

Según los siguientes datos del problema:

$\overline{x}_A = 16$ km/l $\qquad s_A = 1$ km/l $\qquad n_A = 12$

$\overline{x}_B = 11$ km/l $\qquad s_B = 1,8$ km/l $\qquad n_B = 10$

$1 - \alpha = 0,98 \Rightarrow F_{0,01;11,9} = 5,178$ y $F_{0,99;11,9} = 1/F_{0,01;9,11} = 1/4,632$

El intervalo de confianza para el 98% de confianza es:

$$\frac{1}{F_{0,01,11,9}}.\frac{s_A^2}{s_B^2} < \frac{\sigma_A^2}{\sigma_B^2} < \frac{1}{F_{0,99,11,9}}.\frac{s_A^2}{s_B^2}$$

$$\frac{1}{5,178}.\frac{1^2}{1,8^2} < \frac{\sigma_A^2}{\sigma_B^2} < 4,632.\frac{1^2}{1,8^2}$$

$$0,05961 < \frac{\sigma_A^2}{\sigma_B^2} < 1,4296 \qquad 1 - \alpha = 0,98$$

Para las desviaciones típicas, al calcular la raíz cuadrada :

$$0,2441 < \frac{\sigma_A}{\sigma_B} < 1,1957 \qquad 1 - \alpha = 0,98$$

- *Se tiene un 98% de confianza, a partir de los datos de las muestras, que el intervalo (0,05961 $(km/l)^2$; 1,4296 $(km/l)^2$) contiene al cociente de varianzas poblacionales.*
- *Esto indica que no hay diferencias significativas entre las varianzas de los consumos en los camiones tipo A y tipo B, ya que el 1 se encuentra en el intervalo.*

b) Determine un intervalo de confianza del 90% para la diferencia entre los rendimientos medios de cada tipo de camión, suponiendo que los rendimientos tienen varianzas iguales. ¿Qué conclusión obtiene?

X_A:"Rendimiento del combustible en camiones tipo A"
X_B:"Rendimiento del combustible en camiones tipo B"
$X_A \sim Normal\ (\mu_A\ ;\ \sigma_A)$
$X_B \sim Normal\ (\mu_B\ ;\ \sigma_B)$

$$\frac{\left(\overline{X}_A - \overline{X}_B\right) - \left(\mu_A - \mu_B\right)}{S_p . \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \sim t \text{ con } n_A + n_B - 2 \text{ grados de libertad}$$

Según los datos del problema:

$\overline{x}_A = 16$ km/l $\qquad s_A = 1$ km/l $\qquad n_A = 12$
$\overline{x}_B = 11$ km/l $\qquad s_B = 1,8$ km/l $\qquad n_B = 10$
$1 - \alpha = 0,90 \Rightarrow t_{0,05\ ;\ 12+10-2} = \pm 1,725$

Los límites de confianza del 90% son :

$$(\overline{x}_A - \overline{x}_B) \pm t_{0,05\ ;\ 20} . S_p \sqrt{\frac{1}{n_A} + \frac{1}{n_B}} \qquad \text{siendo } S_p^2 = \frac{(n_A - 1)\, s_A^2 + (n_B - 1) s_B^2}{n_A + n_B - 2}$$

$$(16 - 11) \pm 1,725.\ S_p . \sqrt{\frac{1}{12} + \frac{1}{10}}$$

$$\text{siendo } S_p^2 = \frac{(12 - 1)\, 1^2 + (10 - 1)\, 1,8^2}{12 + 10 - 2} = 2,008\ (km/l)^2$$

$$(16 - 11) \pm 1,725.\ 1,417 . \sqrt{\frac{1}{12} + \frac{1}{10}} = \mathbf{5 \pm 1,0466\ km/l}$$

Luego,

$$\mathbf{3,9534\ km/l < \mu_A - \mu_B < 6,0466\ km/l} \qquad \mathbf{1 - \alpha = 0,90}$$

Estimación de parámetros

- *Se tiene un 90% de confianza, según la evidencia muestral, que el intervalo (3,9534 km/l ; 6,0466 km/l) contiene a la diferencia de las medias poblacionales.*
- *Esto indica que hay una diferencia significativa entre los consumos medios de los dos tipos de vehículos, ya que el cero, que indicaría que no hay diferencia significativa, no está en el intervalo del 90% de confianza. Podemos decir que los camiones tipo A tienen mejor rendimiento que los tipo B.*

25. En el laboratorio de ensayo se ha observado que las probetas de hormigón están siendo moldeadas deficientemente, hecho este que está afectando al resultado de ensayo obtenido.

a) Calcule un intervalo de confianza del 96% para la proporción de probetas de hormigón moldeadas defectuosamente, cuando se halla que en una muestra de 100 probetas, 8 son defectuosas.

X:"Cantidad de probetas de hormigón moldeadas defectuosamente"

$X \sim Binomial\ (n\ ;\ p)$

$\hat{P} \sim Normal\ (\mu = p\ ;\ \sigma^2 = p.q/n)$

Según los siguientes datos del problema:

$x = 8$

$n = 100$

$\hat{p} = \dfrac{8}{100} = 0{,}08$

$1 - \alpha = 0{,}96 \quad \Rightarrow \quad z_{0,01} = \pm\, 2{,}05$

Los límites de confianza del 96% son:

$$\hat{p} \pm z_c \cdot \sqrt{\frac{\hat{p}.\hat{q}}{n}} = 0{,}08 \pm 2{,}05 \cdot \sqrt{\frac{0{,}08.(1-0{,}08)}{100}} = \mathbf{0{,}08 \pm 0{,}0556}$$

Luego,

$\mathbf{0{,}0244 < p < 0{,}1356} \qquad\qquad \mathbf{1 - \alpha = 0{,}96}$

- *Se tiene un 96% de confianza, según la evidencia muestral, que el intervalo (0,0244 ; 0,1356) contiene a la proporción real de probetas de hormigón moldeadas defectuosamente.*

b) ¿Qué tamaño debe tener la muestra si se desea, con nivel de confianza del 96%, que la proporción de defectuosas esté dentro del 0,05 de la proporción verdadera de defectuosas?

El tamaño de muestra, usando a las observaciones realizadas en la muestra como una estimación previa de la proporción, es:

Estimación de parámetros

$$e = z_c \cdot \sqrt{\frac{\hat{p}.\hat{q}}{n}} = 2{,}05.\sqrt{\frac{0{,}08.(1-0{,}08)}{n}} = \mathbf{0{,}05} \quad \Rightarrow$$

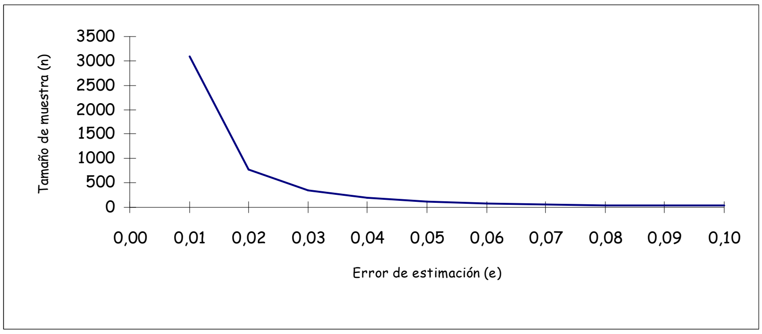
$$\Rightarrow \quad n > \frac{z_c^2.\hat{p}.\hat{q}}{e^2} = \frac{2{,}05^2.0{,}08.0{,}92}{0{,}05^2} = 123{,}72 \quad \Rightarrow \qquad \mathbf{n \geq 124}$$

- *Para tener un 96% de confianza que la proporción real de probetas defectuosas está dentro de un intervalo con un error inferior a 0,05, se debe tomar una muestra superior o igual a 124 probetas.*

c) Grafique la variación del tamaño de la muestra, cuando el error varía entre 0,01 y 0,10, a nivel del 96%, utilizando 0,08 como estimación de la proporción verdadera.

Usando la fórmula $n = \frac{z_c^2.\hat{p}.\hat{q}}{e^2} = \frac{2{,}05^2.0{,}08.0{,}92}{e^2}$

e	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	0,10
n	3093	773	344	193	124	86	63	48	38	31



- *Observamos que a medida que aumenta el error de estimación disminuye la muestra, lo cual es obvio, ya que una muestra pequeña da un mayor error de estimación que una muestra grande.*

26. Un artículo publicado en Engineering Horizons informa que 117 de 484 egresados de Ingeniería planeaban continuar estudiando para obtener un grado más avanzado. Si se considera esto como una muestra aleatoria de todos los graduados en 1990, encuentre un intervalo de confianza del 90% para la proporción de graduados que planean continuar con su educación.

X:"Cantidad de ingenieros que planean realizar estudios de posgrado"

$X \sim$ *Binomial* $(n ; p)$

$\hat{P} \sim$ *Normal* $(\mu = p ; \sigma^2 = p.q/n)$

Según los siguientes datos del problema:

$x = 117$

$n = 484$

$\hat{p} = \frac{117}{484} = 0{,}24$

$1 - \alpha = 0{,}90 \quad \Rightarrow \quad z_{0,05} = \pm 1{,}64$

Los límites de confianza del 90% son:

$$\hat{p} \pm z_c \cdot \sqrt{\frac{\hat{p}.\hat{q}}{n}} = 0{,}24 \pm 1{,}64 \cdot \sqrt{\frac{0{,}24.(1 - 0{,}24)}{484}} = \mathbf{0{,}24 \pm 0{,}0318}$$

Luego,

$\mathbf{0{,}2082 < p < 0{,}2718} \qquad \mathbf{1 - \alpha = 0{,}90}$

🗣 *Se tiene un 90% de confianza, según la evidencia muestral, que el intervalo (0,2082 ; 0,2718) contiene la proporción de egresados que planean seguir estudiando para obtener un grado más avanzado..*

27. En una muestra aleatoria de 1000 viviendas en una ciudad determinada, se encuentra que 228 de ellas utilizan calefacción en base a derivados del petróleo.

a) Encuentre el intervalo de confianza del 99% para la proporción de viviendas en esa ciudad que tiene ese tipo de calefacción.

X:"Cantidad de viviendas que utilizan calefacción en base a derivados del petróleo"

$X \sim Binomial\ (n\ ;\ p)$

$\hat{P} \sim Normal\ (\mu = p\ ;\ \sigma^2 = p.q/n)$

Según los siguientes datos del problema:

$x = 228$

$n = 1000$

$\hat{p} = \frac{228}{1000} = 0{,}228$

$1 - \alpha = 0{,}99 \quad \Rightarrow \quad z_{0,005} = \pm 2{,}58$

Los límites de confianza del 99% son:

$$\hat{p} \pm z_c \cdot \sqrt{\frac{\hat{p}.\hat{q}}{n}} = 0{,}228 \pm 2{,}58 \cdot \sqrt{\frac{0{,}228.(1 - 0{,}228)}{1000}} = \mathbf{0{,}228 \pm 0{,}0342}$$

Luego,

$\mathbf{0{,}1938 < p < 0{,}2622} \qquad \mathbf{1 - \alpha = 0{,}99}$

🗣 *Se tiene un 99% de confianza, en base a la evidencia muestral, que el intervalo (0,1938 ; 0,2622) contiene a la proporción real de viviendas que usan calefacción en base a derivados del petróleo.*

b) ¿Qué tan grande debe ser la muestra si se requiere una confianza al menos del 99% de que la estimación estará dentro del 0,025 de la proporción real de viviendas en esa ciudad que tiene ese tipo de calefacción?

El tamaño de muestra, haciendo una estimación previa de la proporción, es:

$$e = z_c \cdot \sqrt{\frac{\hat{p}.\hat{q}}{n}} = 2{,}58. \sqrt{\frac{0{,}228.(1-0{,}228)}{n}} = \mathbf{0{,}025} \quad \Rightarrow$$

$$\Rightarrow \quad n > \frac{z_c^2.\hat{p}.\hat{q}}{e^2} = \frac{2{,}58^2.0{,}228.0{,}772}{0{,}025^2} = 1874{,}61 \quad \Rightarrow \quad \mathbf{n \geq 1875}$$

🗣 *Para tener un 98% de confianza que la proporción real de viviendas que usan calefacción en base a derivados del petróleo está dentro de un intervalo con un error inferior a 0,025, se debe tomar una muestra superior a 1875 probetas.*

28. Según el Environment News (abril de 1975), "el análisis continuo de los niveles de plomo en el agua potable de varias comunidades de Boston reveló niveles elevados de plomo en los suministros de agua de Somerville, Brighton y Beacon Hill ...". Los resultados preliminares de un estudio efectuado en 1974 indicaron que "el 20% de 248 hogares que se analizaron en estas comunidades reveló niveles elevados que exceden el estándar de la Agencia de Salud Pública de EE.UU. de 50 p.p.m". Al contrario, en Cambridge, que añade corrosivos al agua, "solamente el 5% de los 110 hogares analizados mostró niveles de plomo mayores que el estándar".
Obtenga un intervalo de confianza del 95% para la diferencia de proporciones de hogares que tienen niveles de plomo que exceden el estándar entre las comunidades de Somerville, Brighton y Beacon Hill, y la comunidad de Cambridge.

X_1:"Cantidad de hogares que tienen niveles de plomo que exceden el estándar entre las comunidades de Somerville, Brighton y Beacon Hill"

X_2:"Cantidad de de hogares que tienen niveles de plomo que exceden el estándar en la comunidad de Cambridge"

$X_1 \sim$ *Binomial* $(n_1 ; p_1)$

$X_2 \sim$ *Binomial* $(n_2 ; p_2)$

$\hat{P}_1 - \hat{P}_2 \sim$ *Normal* $(\mu = p_1 - p_2 \; ; \sigma^2 = p_1.q_1/n_1 + p_2.q_2/n_2)$

Según los siguientes datos del problema:

$\hat{p}_1 = 0{,}20 \qquad n_1 = 248$

$\hat{p}_2 = 0{,}05 \qquad n_2 = 110$

$1 - \alpha = 0{,}95 \quad \Rightarrow \quad z_{0{,}025} = \pm 1{,}96$

Los límites de confianza del 95% son :

$$(\hat{p}_1 - \hat{p}_2) \pm z_c \,.\, \sqrt{\frac{\hat{p}_1 . \hat{q}_1}{n_1} + \frac{\hat{p}_2 . \hat{q}_2}{n_2}}$$

$$(0{,}20 - 0{,}05) \pm 1{,}96 \,.\, \sqrt{\frac{0{,}20 \,.\, 0{,}80}{248} + \frac{0{,}05 \,.\, 0{,}95}{110}} = \mathbf{0{,}15 \pm 0{,}0643}$$

Luego,

$$\mathbf{0{,}0857 < p_1 - p_2 < 0{,}2143 \qquad 1 - \alpha = 0{,}95}$$

- *Se tiene un 95% de confianza, según la evidencia muestral, que el intervalo (0,0857 ; 0,2143) contiene a la verdadera diferencia de proporciones de las poblaciones.*
- *Como este intervalo no incluye el cero, podemos decir que no hay diferencia entre las proporciones, debemos creer, con un nivel de confianza del 95% que el nivel de plomo en el agua de las ciudades de Somerville, Brighton y Beacon Hill es significativamente mayor que el nivel de plomo en el agua de Cambridge.*

Estimación de parámetros