TALLER 2
INFERENCIA ESTADÍSTICA

Profesor: Giovany Babativa

Ejercicios de entrenamiento.
La teoría estadística cuyas inferencias para los resultados son válidos solo cuando la distribución de la cual provienen los datos puede ser indexada por un conjunto finito de parámetros, se denomina estadística paramétrica. Cuando dicha distribución corresponde a la normal se desprenden las distribuciones chi-cuadrado, t-student y F, entonces para todos lo numerales suponga que la m.a. $X_1, \ldots, X_n$ es IID con una distribución normal.

1. De un lote de 3500 pilas se probaron 98 al azar. La vida promedio en esa muestra resultó ser de 3.5 horas con una desviación estándar de 0.9 horas. Construya un intervalo de confianza del 99% para la vida media del lote de pilas

2. Una muestra aleatoria de empleados de un grupo numeroso perteneciente a una empresa, entregó las siguientes calificaciones en un examen de aptitud: 63; 72; 56; 65; 66; 74; 57; 59; 63. Construya un intervalo de confianza del 95% para estimar la calificación promedio de todos los trabajadores de la empresa, suponiendo normalidad en la población.

3. Se desea medir la diferencia en ventas entre dos tipos de empleados en la actividad de seguros, unos con titulo profesional y otros de personas con estudios medios. Se toma una muestra de 41 empleados entre los primeros y la media de las ventas resulta ser 32($ miles), en tanto que la media de una muestra de 30 empleados con sólo estudios medios es de 25. Se encontró también que la varianza en la primera muestra es de 48 y en la segunda de 56. Determine el intervalo de confianza del 95% para estimar la diferencia en las ventas medias de los dos tipos de vendedores. (suponer normalidad).

4. Una pequeña empresa compró un lote grande de piezas electrónicas a una firma. En una muestra aleatoria de 50 piezas se comprobó que 5 eran defectuosas. Estime la proporción de piezas defectuosas de todo el lote, empleando un intervalo de 95% de confianza.

5. En dos ciudades se tomaron muestras de automóviles, cada una de 100 automóviles. En una ciudad 72 automóviles pasaron con éxito la prueba de seguridad, en la otra solamente lo hicieron 66. Construya un intervalo de 95% de confianza para estimar la diferencia de proporciones de autos seguros en las dos ciudades.

6. Se desea hacer una encuesta para estimar el porcentaje de personas mayores de edad, de una ciudad, que están a favor de cierto proyecto de ley. ¿Cuál debe ser el tamaño de la muestra, si se desea un nivel de confianza del 95% y un error máximo de estimación de 4%? (por estudios anteriores se sabe que la desviación estándar es de 10).

7. Se desea estimar el gasto medio en movilización de los alumnos de cierta Universidad que tiene 5000 alumnos. ¿Cuál es el tamaño de muestra necesario, si se desea un nivel de confianza del 95%, un error máximo de $\pm$\$2.000, y se sabe por estudios anteriores que la desviación estándar es de \$15.000?

8. Una muestra de 100 estudiantes elegida al azar dentro de la población de estudiantes de la Universidad Nacional, indica que el 55% de ellos no está de acuerdo con el estatuto estudiantil. Halle un intervalo de confianza del 95% para **estimar** la proporción de estudiantes que no están de acuerdo con el estatuto.

9. Una empresa fabricante de gaseosas está interesada en ajustar el comportamiento del dispensador liquido de manera tal que se pueda verter un promedio μ de onzas de líquido en cada botella. Se observó que la cantidad de líquido que dispensa la máquina tiene una distribución normal con $\sigma = 1.0$ onzas. Se elige aleatoriamente una muestra de $n = 9$ gaseosas llenas en un día determinado, y se mide el contenido de cada una.

   a. Si la media real es de 10 onzas, cual es la probabilidad de que una botella contenga menos de 9.5 onzas.

   b. Calcule la probabilidad de que la diferencia entre la media muestral y la media real μ sea menor que 0.3 onzas

   c. Cuántas observaciones se deben realizar en la muestra para que la diferencia entre la media muestral y la real μ no supere las 0.3 onzas con una probabilidad de 0.95.

10. Suponga que obtenemos una observación Y de una distribución exponencial con media θ. Utilice Y para construir un Intervalo de Confianza para con un coeficiente de confianza de 0.9. Utilice Y/θ como cantidad pivotal.

11. Un fabricante de fibras sintéticas desea estimar la tensión de ruptura media de una fibra. Diseña un experimento en el que se observan las tensiones de ruptura, en libras, de 36 hilos del proceso seleccionado aleatoriamente. Suponga que la tensión de ruptura de una fibra se encuentra modelada por la distribución normal con desviación estándar de 0.45 y que la media muestral fue 20,38 libras. Construir un Intervalo de Confianza estimado del 98% para el valor real de la tensión de ruptura promedio de la fibra.

12. Dos universidades financiadas por el gobierno tienen métodos distintos para inscribir a sus alumnos a principios de cada semestre. Las dos desean comparar el tiempo promedio que les toma a los estudiantes completar el trámite de inscripción. En cada universidad se anotaron los tiempos de inscripción para 100 alumnos seleccionados al azar. Las medias y las desviaciones estándares muestrales son las siguientes:

$$\overline{X}_1 = 50.2 \qquad \overline{X}_2 = 52.9$$
$$S_1 = 4.8 \qquad S_2 = 5.4$$

    Si se supone que el muestreo se llevo acabo sobre dos poblaciones distribuidas normales e independientes, obtener los intervalos de confianza estimados del 90, 95 y 99% para la diferencia entre las medidas del tiempo de inscripción para las dos universidades. Con base en esta evidencia, ¿se estaría inclinando a concluir que existe una diferencia real entre los tiempos medios para cada universidad?

13. En base al ejercicio 12, construir un intervalo de confianza estimado del 99% para el cociente de las varianzas de las dos poblaciones. Con base en este resultado, ¿es razonable la suposición de que las varianzas son iguales?

14. Una fábrica trabaja con m tipos diferentes de máquinas. Se cuenta con i máquinas de tipo i, $i = 1, \dots, m$, el costo de reparación semanal de cada máquina de tipo i se distribuye normal con media μ y varianza $i\sigma^2$, $i = 1, \dots, m$. El costo total de reparación semanal esperado por la fábrica es $\mu = \sum_{i=1}^{m} i\mu_i$. Se tomaron muestras aleatorias independientes, una muestra aleatoria $X_{11}, X_{12}, \dots, X_{1n_1}$ para la máquina de tipo 1; otra muestra aleatoria $X_{21}, X_{22}, \dots, X_{2n_2}$ para la máquina de tipo 2 y así sucesivamente para todos los tipos de máquina.

a. ¿Qué estimador emplearía con el objeto de estimar el costo de reparación semanal esperado para la fábrica?

b. Construya una cantidad pivotal que se distribuya normal estándar a partir de la estadística encontrada en el numeral anterior, que sirva para hacer pruebas de hipótesis cuando la varianza es conocida.

15. La magnitud del voltaje en el contador de energía de una trilladora es una variable aleatoria X. Se implementa un sistema de corrección con el que se espera que la magnitud del voltaje cambie. La magnitud del voltaje cuando se implementa el sistema de corrección es una variable aleatoria Y. Se supone que $X \sim N(\mu_X, \sigma^2)$ y $Y \sim N(\mu_Y, 2\sigma^2)$. Si se toman dos muestras independientes entre si, una muestra de X de tamaño 9 donde se obtiene una magnitud de voltaje promedio de 119 voltios y varianza muestral de 25 voltios^2, y otra muestra de Y de tamaño 8 donde se obtiene una magnitud de voltaje promedio de 123 voltios y varianza de 47 voltios^2, y asumiendo que las poblaciones X y Y son independientes:

a. ¿Existe suficiente evidencia estadística para afirmar que la varianza de la población Y es el doble de la de la población X?.

b. ¿Existe suficiente evidencia estadística para afirmar que hay diferencia entre las magnitudes de los voltajes medios antes y después de la conexión del sistema de corrección?. (Utilice un nivel de confianza del 95%)

c. Construya un intervalo de confianza al 95% para estimar σ^2.

d. Suponga que se plantean las siguientes hipótesis nula y alterna, $H_0 : \sigma^2 = \sigma_0^2$ vs $H_1 : \sigma^2 \neq \sigma_0^2$. ¿Para qué posibles valores de σ_0^2 no se rechaza la hipótesis nula?

16. Una empresa ha estado experimentando con dos disposiciones físicas distintas de su línea de ensamble. Se ha determinado que ambas disposiciones producen aproximadamente el mismo número de unidades terminadas al día. A fin de obtener una disposición que permita un mayor control del proceso, se sugiere que se adopte de manera permanente la disposición que exhiba la varianza más pequeña en el número de unidades terminadas. Dos muestras aleatorias independientes producen los resultados presentados en la tabla 1:

Linea de ensamble 1	Linea de ensamble 2
$n_1 = 21$ dias	$n_2 = 25$ dias
$s_1^2 = 1.432$	$s_2^2 = 3.761$

Tabla 1: Datos del ejercicio 16

Establezca un intervalo de confianza del 95% para $\frac{\sigma_1}{\sigma_2}$, la razón de las varianzas del número de unidades terminadas para las dos disposiciones de línea de ensamble, suponiendo que las distribuciones de los números de unidades terminadas para las dos líneas de ensamble son normales. Con base en el resultado anterior, ¿cual de las dos disposiciones recomendaría usted?

17. Se recibe un lote muy grande de artículos proveniente de un fabricante que asegura que el porcentaje de artículos defectuosos en la producción es del 1%. Al seleccionar una muestra aleatoria de 200 artículos y después de inspeccionarlos, se descubren 8 defectuosos. Obtener los Intervalo de confianza aproximados del 90% y 95% para la verdadera proporción de artículos defectuosos en un proceso de manufactura del fabricante. ¿Qué se puede concluir con respecto a la afirmación del fabricante?.

18. Los siguientes son los tiempos de secado (minutos) de hojas cubiertas de poliuretano bajo dos condiciones ambientales diferentes:

Condición 1	55.6	56.1	61.8	55.9	51.4	59.9	54.3	62.8	58.5	55.8
	58.3	60.2	54.2	50.1	57.1	57.5	63.6	59.3	60.9	61.8
Condición 2	55.1	43.5	51.2	46.2	56.7	52.5	53.5	60.5	52.1	47.0
	53.0	53.8	51.6	53.6	42.9	52.0	55.1	57.1	62.8	54.8

a. Halle un intervalo de 98% confianza para la diferencia entre las medias de los tiempos de secado bajo las dos condiciones ambientales. Suponga que las muestras son independientes entre si y provienen de poblaciones normales.

b. Un intervalo de confianza para el tiempo promedio de secado bajo la condición ambiental 1 es (56.041, 57.755) ¿Cuál es el nivel de confianza de este intervalo?

c. Halle un intervalo de 95% de confianza para la proporción de hojas cubiertas de poliuretano con tiempos de secado mayores que 60. No discrimine por condición ambiental.

d. Halle un intervalo de 95% de confianza para la varianza de tiempos de secado de las hojas cubiertas de poliuretano bajo la condición ambiental 2.

19. El número de accidentes del tránsito mortales en una ciudad es, en promedio, de 12 mensuales. Tras una campaña de señalización y educación se contabilizaron en 6 meses sucesivos: 8; 11; 9; 7; 10; 9 accidentes mortales. ¿Se puede decir a un nivel de significación del 5% que fue efectiva la campaña?

20. Se sabe que el peso promedio de mujeres entre 30 y 40 años en cierta región, ha sido históricamente de 53 kilos, con una desviación estándar de 5. En un estudio realizado en 16 mujeres de tales edades en esa región y que entregó una media de 50 kilos con una desviación estándar de 4.9 kilos.

    a. ¿Qué conclusión se puede sacar al nivel de significación del 5%, respecto del peso promedio?

    b. ¿Qué conclusión se puede sacar al nivel de significación del 1% respecto del peso promedio?

21. Shell de Colombia S.A. produce aceite y lo vende en presentaciones de balde. De acuerdo con la legislación colombiana el contenido neto del aceite en cada balde no puede ser menor de 5 galones o su equivalente en gramos, es decir 15 kilogramos.

    La superintendencia de industria y comercio con el propósito de verificar el cumplimiento de este requisito visitó la fábrica y midió el peso de una muestra de 15 baldes con aceite (producto empacado) y el peso de 17 baldes vacíos, y obtuvo los siguientes resultados:

	Promedio	desviación estándar
Baldes Vacíos	2.513 Kg	0.516 Kg
Empacado	18.101 Kg	1.316 Kg

    Asumiendo que:

    - El peso de los baldes vacío y el producto empacado se distribuye normal.
    - La varianza del peso del producto empacado es tres veces la varianza del peso de los baldes vacíos.

    ¿Existe evidencia estadística para afirmar que el contenido neto que entrega Shell en sus productos no cumple con las especificaciones dadas por la legislación colombiana? Utilice una confiabilidad del 95%.

22. En un programa de control de enfermedades crónicas, la hipertensión está incluida como la primera patología a controlar. 15 pacientes hipertensos son sometidos al programa y controlados en su presión antes y después de 6 meses de tratamiento. Los datos son los siguientes:

    ¿Se puede decir a un nivel de significación del 5% que el tratamiento es efectivo?

Inicio	180	200	160	170	180	190	190	180
	190	160	170	190	200	210	220	
Fin	140	170	160	140	130	150	140	150
	190	170	120	160	170	160	150	

23. El fabricante de cierta marca de cigarrillos sostiene que sus cigarrillos contienen en promedio 18 miligramos de nicotina por cigarrillo. Un organismo de control examina una muestra de 100 cigarrillos encontrando un contenido medio de 19.2 miligramos por cigarrillo, con una desviación estándar de 2 miligramos. ¿Puede el organismo concluir a un nivel del 5% de significación que el fabricante subestima el contenido medio de nicotina de sus cigarrillos?

24. El gerente de un reconocido restaurante de los alrededores de esta prestigiosa universidad realizó un estudio para conocer el comportamiento de sus ventas diarias (de lunes a viernes) durante el semestre académico. El resultado del estudio es que las ventas se distribuyen normal con promedio $\mu = \$350000$ y $\sigma = \$70000$. El gerente del restaurante, buscando una estrategia más agresiva de ventas, contrata dos espectaculares meseras para atender óptimamente al cliente. Después de dos meses, se toma una muestra de las ventas diarias durante cada uno de 25 días. Se obtiene un $\overline{X} = \$380000$ y $S = \$70000$. El dueño quiere saber si existe suficiente evidencia para afirmar que sus ventas diarias se incrementaron, para lo cual contrata un estudiante destacado del curso de Estadística II. Este le explica al gerente toda la teoría de intervalos de confianza que aprendió en las clase y le propone realizar tres I.C. el primero al 90%, el segundo al 95% y el tercero al 99%:

    a. Determine los intervalos adecuados para esta situación.

    b. Existe suficiente evidencia estadística para afirmar que luego de la contratación de las meseras hubo cambio en la varianza de las ventas. Utilice un nivel de significancia de 0.05.

    c. Si el α máximo tolerable por el gerente del restaurante es 0.01, ¿Qué intervalo escogería y por qué?. Concluya.

25. Una prestigiosa empresa diseñadora de automóviles sabe que el consumo promedio de combustible de su mejor vehículo es de 300 km/gal y su desviación estándar es de 10 km/gal. El director ha decidido explorar si un nuevo sistema de inyección electrónica inducida por pulsos electromagnéticos puede reducir el consumo de combustible. Luego de la implementación del sistema a 30 automóviles el consumo promedio fue de 282 km/gal. Use $\alpha = 0.05$.

    a. ¿Se puede afirmar que el nuevo sistema reduce el consumo de combustible?

    b. ¿Cual debe ser el mínimo consumo promedio muestral para que la probabilidad de rechazar la hipótesis nula dado que es cierta sea de 0.025?

26. Una empresa cuenta con 5 máquinas tipo A y 4 máquinas tipo B. El costo de reparación de cada máquina es como se muestra a continuación:

$$\text{Máquina A} \rightarrow X \sim N(\mu_A, \sigma^2)$$
$$\text{Máquina B} \rightarrow Y \sim N(\mu_B, 2\sigma^2)$$

    Construya un Intervalo de confianza al 95% para estimar el costo total que representa para la empresa la reparación de las 9 máquinas.