Estadística, curso 2018-19

Ingeniería en tecnologías Industriales

Profesor: Adrián Serrano

Ejercicios del Bloque III

Ejercicios de aplicación inmediata de la teoría

Distribución en el muestreo de la media

a) Una población de un tipo de plantas tiene una talla media de 15 cm y desviación típica de 2.5 cm. Se toma al azar una muestra de 45 plantas. ¿Cuál es la probabilidad de que la media de las tallas de la muestra sea superior a 12.5? Qué hipótesis sobre la variable hay que añadir para responder a la misma pregunta si el tamaño muestral fuera de 10 plantas.

b) Consideremos una urna compuesta por 9 bolas numeradas del 1 al 9. Se extraen al azar y con reemplazamiento 40 bolas. Calcula la probabilidad de que la suma de las puntuaciones sea mayor que 195.

c) Un laboratorio fabrica comprimidos efervescentes en forma de disco cuyo diámetro X quiere controlar. La variable X tiene media 15 mm y desviación típica 4 mm. A intervalos fijos de tiempo se extraen muestras de 64 unidades de las que se miden los diámetros. Calcula la probabilidad de que la variable media muestral esté comprendida entre 13 y 16 mm.

Distribución en el muestreo de la proporción

a) El 4 % de las piezas que produce una máquina son defectuosas. Se toma una muestra aleatoria de 80 piezas. Calcula la probabilidad de que en la muestra existan menos de tres piezas defectuosas.

b) En unas elecciones generales el candidato A ha recibido el 45 % de los votos. Calcula la probabilidad de que en una muestra escogida al azar de 50 votantes más de la mitad haya votado a A.

Distribución en el muestreo de la diferencia de medias

a) En un aeropuerto operan dos copañías aéreas, A y B. El 60 % de los vuelos los opera la compañía A y el resto la compañía B. Los retrasos de los vuelos de la compañía A se distribuyen como una normal de media 18 minutos y desviación 4 minutos. Esta misma variable en la compañía B se distribuye como una normal de media 10 minutos y desviación típica de 3 minutos. Se toma una muestra de tamaño 100 al azar de vuelos, manteniendo la proporción de vuelos operados por cada compañía. Halla la probabilidad de que la diferencia de las medias muestrales de los retrasos de ambas compañías sea superior a 10 minutos.

Distribución en el muestreo de la desviación típica muestral

a) Si la altura de un grupo de población sigue una distribución normal N(176,12), calcular la probabilidad de que la desviación típica muestral sea menor de 10 para una muestra de tamaño 8.

Inferencia con una población

a) Las ausencias en días de un empleado de una empresa para un determinado año se aproxima por una normal de media μ desconocida y desviación típica de 2 días. Se pretende estimar μ usando la media de las ausencias en ese año de n trabajadores seleccionados de manera aleatoria en la empresa. ¿Qué tamaño muestral debería utilizarse para obtener un intervalo de confianza al 95 % con un error de $\pm 0{,}2$?. Si $\mu = 6{,}3$ y $n = 25$, ¿cuál es la probabilidad de que la media que resulta de la muestra se encuentre entre 6.1 y 6.5?.

b) Una de las entradas de cierta ciudad sufría constantemente retenciones de tráfico, de forma que el tiempo de espera en la cola formada por el semáforo allí instlado seguía una normal de media 10 minutos y desviación típica de 4 minutos. Con el fin de descongestionar ese punto y bajar la media de tiempo de espera, se habilitó una vía de acceso auxiliar. Transcurrida una semana se hizo un pequeño estudio sobre 36 vehículos y se obtuvo que tiempo medio de espera en el citado semáforo fue de 8.5 minutos. Las autoridades municipales mostraron su satisfacción y dijeron que la medida había funcionado, pero la opinión pública, sin embargo, defiende que la situación sigue igual.Suponiendo que la desviación típica se ha mantenido, contrasta la hipótesis defendida por la opinión pública frente a la de los responsables municipales a un nivel de significación del 5 %.

c) En un instituto de investigaciones dermatológicas se está investigando una afección cut ánea de tipo cancerígeno. Se eligen 140 ratas de una misma raza aleatoriamente y se les provoca el cáncer citados, a continuación se les frota con un medicamento. Se toma como variable rspuesta el número de horas que tarda el cáncer en desaparecer. Se obtiene una media muestral de 100 horas y un desviación típica muestral de 10 horas. Se puede aceptar la normalidad de la variable respuesta. Responde a las siguientes cuestiones referidas al enunciado anterior (los datos que se indican en cada apartado no se mantienen en los posteriores)

- c1) Calcula un intervalo de confianza al 90 % para la media de la variable respuesta.

- c2) Si se supiera que $\sigma = 9$ horas, obtén un intervalo de cofianza al 99 % para la media de la variable respuesta.

- c3) Si el tamaño muestral fuera $n = 10$ en lugar de 140, calcula el intervalo de confianza al 90 % de la media de la variable respuesta.

- c4) ¿Se puede aceptar con una significación del 5 % la hipótesis nula de que la media es menor que 105 horas?

- c5) Obtén un intervalo de confianza al 90 % para la varianza de la variable respuesta.

- c6) Contrasta la hipótesis $\sigma = 9$.

d) Se quiere estimar la media de la nómina mensual que reciben los directivos de las compañías multinacionales que operan en Europa.

- d1) Si la varianza de la nómina en la población es de 1000 ¿cuál es la varianza de la media muestral cuando el tamaño de la muestra es de 100?

- d2) Si en las condiciones del apartado anterior la media muestral fuera de 4008 , ¿se rechazaría con un nivel de significación del 5 % la hipótesis de que la nómina media es de 4000?

e) Dos estudiantes quieren contrastar si el consumo medio en teléfono móvil entre los estudiantes es como máximo de 10 Euros frente a si es mayor. El primero, en una muestra de 36 estudiantes, obtuvo una media de 10.4 euros con una desviación típica de 2 euros. El segundo, en una muestra de 49 estudiantes, obtuvo una media de 10.39 con una desviación de 2 Euros. ¿Qué decisión toma cada uno con un nivel de significación del 10 %?

f) En un hospital se observó que los pacientes abusaban del servicio de urgencias de manera que un 30 % de las consultas podías perfectamente haber esperado a concertar una cita con el médico de cabecera, porque no eran realmente urgentes. Puesto que esta situación ralentizaba el servicio, se realizó una campaña intersiva de concienciación. Transcurridos unos meses se ha recogido información de 60 consultas al servicio, de las cuales solo 15 no eran realmente urgencias.

- f1) Proporciona un intervalo de confianza al 95 % de la proporción poblacional de falsas urgencias.

- f2) Obtén el intervalo de confianza al 95.5 % de la proporción poblacional de falsas urgencias utilizando la fórmula del error máximo.

- f3) Hay personal del hospital que defiende que la campaña no ha mejorado la situación. Resuelve, con $\alpha = 0{,}01$, un contraste con dicha afirmación como hipótesis nula.

- f4) Si el contraste anterior llega a una conclusión equivocada, ¿qué tipo de error se habría cometido?

Estimación, intervalos de confianza para una población

1. **Control de calidad** Una empresa fabrica anillos de pistón para los motores de automóviles. El diámetro interno de los anillos es un valor crítico. El proceso de producción está diseñado para que el diámetro medio se comporte como una normal de media μ=74 milímetros y desviación σ=0.01 milímetros. A pesar de que el proceso está bien diseñado y tiene un mantenimiento adecuado, existen causas incontrolables que generan pequeñas distorsiones en los diámetros. Con objeto de controlar cuándo el proceso está realmente funcionando de forma anómala, cada hora se toma una muestra aleatoria de 5 anillos y se calcula la media y la desviación muestral de los diámetros. El número de muestras tomadas semanalmente es de 114, si el proceso está funcionando correctamente

   a) ¿qué proporción de las mismas se espera que tomen valores fuera del intervalo $\mu \pm 3\sigma/\sqrt{n}$?

   b) ¿Qué proporción de las muestras semanales se espera que permanezcan en el intervalo $(\mu,\ \mu + 3\sigma/\sqrt{n})$?¿y en $(\mu,\ \mu - 3\sigma/\sqrt{n})$?

   El proceso de producción se revisa si la proporción semanal de muestras supera los valores obtenidos en a) o en b).

   Si nos planteamos hacer un control similar sobre la cuasivarianza muestral de los diámetros de los anillos

   c) si el proceso funciona correctamente ¿qué intervalo propondrías para la cuasivarianza muestral de manera que la proporción esperada de muestras semanales que toman valores fuera de este intervalo sea la misma que la obtenida en a)?

2. **Duración de productos** El ingeniero de desarrollo de un fabricante de cubiertas está investigando la vida de las cubiertas para un nuevo compuesto de hule. Sabe que la vida útil sigue una distribución normal y que la desviación, $\sigma = 645km$, no cambiará con el nuevo compuesto pero que la media puede aumentar. Ha hecho 16 cubiertas y las ha probado hasta el fin de su vida útil en una prueba de carretera. La media muestral es de 60139 km. Al ingeniero le gustaría demostrar que la media de esta nueva cubierta excede 60000km.

   a) Construye un intervalo de confianza y establece conclusiones al 95 %.

   b) Si consideramos como hipótesis alternativa la sospecha del ingeniero, efectúa el contraste correspondiente a un nivel de significación del 5 %.

3. **Especificaciones de productos** El nivel de brillo de un cinescopio de televisión puede evaluarse midiendo la cantidad de corriente que se necesita para alcanzar un nivel de brillo particular. Un ingeniero ha diseñado un cinescopio que en su opinión necesitará 300 microamperes de corriente para producir el nivel de brillo deseado. Una muestra de 80 cinescopios da como resultado $\overline{x} = 317,2$ y $S = 15,7$.

   a) Construye un intervalo de confianza y establece conclusiones al 90 %.

   b) Si el tamaño muestras fuera de 20 cinescopios, repite el ejercicio anterior asumiendo las hipótesis que estimes necesarias.

   c) Efectúa el contraste correspondiente a un nivel de significación del 5 %, considerando como hipótesis alternativa la sospecha del ingeniero. Realiza el contraste para la situación propuesta en a) y en b).

4. **Inferencia con proporciones** Se estudia la fracción de circuitos integrados defectuosos producidos en un proceso de fotolitografía. Se prueba una muestra aleatoria de 300 circuitos, la cual da como resultado 13 circuitos defectuosos. Se quiere demostrar que la proporción de defectuosos es menor del 5 %.

   a) Construye un intervalo de confianza al 95 % y establece conclusiones.

   b) Efectúa el contraste de hipótesis correspondiente planteando como hipótesis alternativa la sospecha del enunciado. Asume $\alpha = 0{,}05$.

   c) Efectúa el contraste de hipótesis tal que la hipótesis nula plantea que la proporción de defectuosos es del 4 %.

5. **Inferencia con varianzas** Una marca de margarina dietética fue analizada para determinar el nivel de grasas poliinsaturadas (en porcentajes). Un muestra de seis tarrinas proporcionó los siguientes datos: 16.8, 17.2, 17.4, 16.9, 16.5, 17.1. Si asumimos normalidad de la variable en estudio

   a) Calcula un intervalo de confianza al 95 % para la media poblacional y otro para la varianza de la variable.

   b) Efectúa el contraste $H_0 : \sigma_0^2 = 1$ frente a $H_0 : \sigma_0^2 < 1$.

   c) Utilizando el mismo valor de la varianza muestral que se obtiene con los 6 datos del enunciado, repite el apartado b) suponiendo que el tamaño muestral es $n = 51$. Compara los resultados con los obtenidos en el apartado b) y analiza el efecto del tamaño muestral en las conclusiones.

Inferencia con dos poblaciones

1. **Comparación de medias**Un fabricante de componentes electrónicos está valorando utilizar un nuevo tipo de plástico. La resistencia a la ruptura del plástico es importante y se sabe que sigue una distribución normal. La desviación es la misma para el plástico 1 (nuevo) y el plástico 2 (clásico), $\sigma_1 = \sigma_2 = 1'0$ psi(libras/pulgada^2). A partir de la muestra aleatoria de 10 componentes fabricados con el plástico 1 y otra de 12 componentes fabricados con el clásico, se obtiene $\overline{x}_1 = 162'5$ y $\overline{x}_2 = 155'0$ respectivamente. La compañía no adoptará el plástico 1 a menos que su resistencia media a la ruptura exceda la del plástico 2 en al menos 10 psi. Con base en la información muestral, se quiere decidir si debería utilizarse el plástico 1:

   a) Construye un intervalo de confianza y establece conclusiones al 95 %.

   b) Efectúa el contraste con $\alpha = 0{,}05$ que plantea como hipótesis alternativa que la diferencia entre resistencias medias del plástico 1 y 2 no llega a ser 10.

2. **Comparación de medias**Se estudia la velocidad de combustión de dos cargas propulsoras sólidas diferentes usadas en el sistema de expulsión de la tripulación de un avión. Se sabe que la velocidad de combustión en ambas cargas propulsoras tienen distribución normal y aproximadamente la misma desviación estándar. Se prueban dos muestras aleatorias de $n_1 = 20$ y $n_2 = 20$; las medias y desviaciones muestrales de

la velocidad de combustión son $\overline{x}_1 = 18$ cm/s, $S_1 = 3,5$ cm/s y $\overline{x}_2 = 24$ cm/s, $S_2 = 3$ cm/s.

a) Construye un intervalo de confianza de 95 % para la diferencia de las medias $\mu_1 - \mu_2$. ¿Cuál es el significado práctico de este intervalo?

b) Efectúa el contraste de igualdad de medias a un nivel de significación del 5 %.

c) ¿Cómo se modificarían los apartados anteriores si los tamaños muestrales fueran $n_1 = 100$ y $n_2 = 105$ y no se asumiera normalidad?

3. **Comparación de medias**Dos proveedores fabrican un engranaje de plástico usado en una impresora láser. La resistencia a los impactos(psi=libra-fuerza/pulgada^2) de los engranajes es una característica muy importante. Se toma una muestra del proveedor A de 100 piezas que da como resultado una resistencia promedio de 290 y una desviación de 12. En una muestra de 160 piezas del proveedor B la media fue de 321 y la desviación de 22.

   Estudia, mediante un contraste de hipótesis, si los datos evidencian que la resistencia media del proveedor B supera en 20 a la de A.

4. **Comparación de proporciones** Se usan dos tipos diferentes de máquinas de moldeado por inyección para hacer piezas de plástico. Una pieza se considera defectuosa si presenta una merma excesiva o esta descolorida. Se seleccionan dos muestras cada una de tamaño 300, correspondientes a dos máquinas distintas, y se encuentran en una 15 piezas defectuosas y 8 en la otra. ¿Es razonable concluir que ambas máquinas producen el mismo porcentaje de piezas defectuosas? Haz el estudio mediante un intervalo al 99 % y, también, mediante contrastes de hipótesis.

5. **Tratamiento inferencial de dos poblaciones según los supuestos.** En el departamento de control de calid ad de una empresa, se quiere determinar si ha habido un descenso significativo de la calidad de su producto entre las producciones de dos semanas consecutivas a consecuencia de un incidente ocurrido durante el fin de semana. Deciden tomar una muestra de la prod ucción de cada semana, si la calidad de cada artículo se mide en una escala de 100, obtienen los resultados siguientes:

   Semana 1: 93, 86, 90, 90, 94, 91, 92, 96

   Semana 2: 93, 87, 97, 90, 88, 87, 84, 93

Asumimos la normalidad de la variable calidad en ambas semanas

b1) Suponiendo que las varianzas de la puntuación en las dos producciones son iguales, construye un intervalo de confianza para la diferencia de medias al nivel de 95Interpreta los resultados obtenidos.

b2) Suponiendo que las varianzas no son iguales, construye el intervalo pedido en b1)

b3) Proporciona un intervalo de confianza al 95 % para las medias poblacionales para cada semana por separado

b4) Si intervalos obtenidos en b3)se solapan, ¿podemos decir que con una confianza del 95 % las medias pueden coincidir en ambas semanas? Si las condiciones de b2) son ciertas, ¿se podría decir con el intervalo obtenido en b2) que con una significatividad del 5 % las medias son iguales?

Ejercicios sobre el concepto de inferencia

Responde a las siguientes cuestiones con ayuda de un gráfico

a) En un contraste se concluye que es significativo con $\alpha = 0{,}05$. ¿Qué error se puede estar comentiendo?

b) Un contraste concluye rechazando la hipótesis nula con $\alpha = 0{,}025$ pero la acepta si $\alpha = 0{,}01$. ¿El p-valor se encuentra entre 0.01 y 0.025?

c) ¿ Si un contraste es significativo con $\alpha = 0{,}05$ también lo es con $\alpha = 0{,}01$?

d) ¿El p-valor es la probabilidad de que cometamos un error de tipo I?

e) Una persona tiene unos hábitos muy regulares, de tal manera que todos los días que efectúa un paseo pasa por el punto A alrededor de las 16:30. La llegada exacta al punto A se puede modelar mediante una distribución uniforme de amplitud 10 minutos simétrica respecto de las 16:30. Efectuamos un contraste de manera que el error tipo I significa rechazar que la persona ha salido a pasear cuando realmente sí lo ha hecho. Si pronosticamos que la persona no ha salido a pasear cuando no ha pasado por A antes de las 16:34, ¿qué significatividad tiene el contraste?.