Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Profesora: Lorena Correa A.
Ayudante: Felipe Calderón
Segundo Semestre 2003

GUIA 2 -METODOS ESTADISTICOS
ELM2400

Intervalos de Confianza

1. Se desea estimar la tasa de falla π de cierta máquina. Al observar 100 unidades durante cierto tiempo, se registró que 7 máquinas fallaron.

   (a) Construya intervalos de 90%, 92% y 96% de confianza para π. Use la aproximación Normal y luego construya intervalos exactos. Compare e interprete los resultados.

   (b) Si el ancho de un intervalo de 99% de confianza puede ser a lo más 0.10, ¿cuál es el tamaño de la muestra que se debe usar si π es desconocido?.

2. Un compuesto químico es testeado regularmente por su fabricante para determinar su contenido de azufre. Durante un período de observación se ha encontrado que el contenido medio es de 0.35% con una desviación estándar de 0.05%. Para un comprador de este compuesto es muy importante la cantidad de azufre que él contenga y por ello la mide en cuatro muestras que obtiene aleatoriamente de su compra. Si el promedio en dichas muestras excede el 0.5% rechaza la compra. ¿Puede confiar, en forma razonable, en que no rechazará la partida equivocadamente?.

3. Para controlar la calidad del concreto se utiliza una tarjeta de registro de una prueba de resistencia realizada a cubos de dimensiones especificadas y en base a ellas se determinó que ella tiene una distribución Normal de media $28N/mm^2$ y desviación estándar $6N/mm^2$. Se desea construir límites simétricos en torno a dicha media de tal forma de tener sólo una probabilidad de 1 en 5 de que el promedio de mediciones a cuatro cubos caiga fuera de dichos límites cuando el concreto mantiene los requerimientos del diseño. ¿Cuáles serían los límites si la probabilidad disminuye a 1 en 20?.

4. Para determinar la cantidad de cloro contenido en dos polímeros diferentes se realizaron 9 mediciones en el primero y 16 en el segundo, entregando promedios de 58.18 y 56.97 respectivamente. El método analítico utilizado es conocido de experiencias anteriores y se sabe que entrega resultados con una desviación estándar de 0.8. Encuentre un intervalo de 99 % de confianza para la verdadera diferencia entre los porcentajes de cloro de ambos polímeros.

5. En la construcción de un puente se necesita estudiar la posible aparición de pequeñísimas grietas. Se determina que la proporción de dm^2 construidos que tengan una de estas grietas no puede pasar de 1 por mil, y se necesita estar seguro de que se cumpla esta condición. Para ello se toma una muestra de N dm^2 en construcción que se observan para determinar la proporción de los que presentan alguna falla.

   (a) ¿Qué tipo de intervalo de una cola sería conveniente?. Derívelo.

   (b) Encuentre el intervalo de una cola, de 95% de confianza, sabiendo que se encontró fallas en 4 dm^2 de 5.000 observados.

6. Dos sucursales bancarias atienden un número de clientes por cada 5 minutos de acuerdo a distribuciones Poisson con parámetros λ_1 y λ_2 respectivamente. Se tomaron 2 muestras aleatorias de 120 períodos de 5 minutos en cada sucursal, encontrándose promedios muestrales de 9.5 y 10.1 clientes por período. Mediante intervalo de 99% de confianza para $(\lambda_1 - \lambda_2)$ indique si se puede afirmar que las tasas de atención son distintas.

7. La duración de una componente es aproximadamente Normal con una media proporcional a la cantidad de uno de los insumos utilizados, que no es una variable aleatoria, y varianza constante. Una muestra de tamaño $n = 8$ de las duraciones, en días, es 23, 31, 34, 46, 25, 49, 30. Los valores correspondientes a la cantidad de insumo utilizado, en gramos, son 18, 23, 22, 30, 15, 29, 16. Derive el estimador de máxima verosimilitud para la constante de proporcionalidad y obtenga un intervalo de confianza 0.95 para este parámtero suponiendo que la distribución de probabilidades del estimador es asintóticamente Normal.

8. El polvo detergente es comercializado en cajas que tienen un peso rotulado que se debe respetar. Con el objeto de estimar el peso μ y la desviación típica σ de las cajas de un lote producido en una jornada, se sacan 3 muestras independientes de tamaños n_1, n_2, n_3 respectivamente, y se obtienen $\bar{X}_i, \quad S_i^2, \quad i = 1, 2, 3$. Asuma normalidad de la población. Si

$$n_1 = 10, \quad n_2 = 15, \quad n_3 = 12$$

$$\bar{X}_1 = 151,5\ grs. \quad \bar{X}_2 = 152,0grs. \quad \bar{X}_3 = 150,5grs.$$

$$S_1^2 = 1,44grs^2 \quad S_2^2 = 1,21grs^2 \quad S_3^2 = 1,00grs^2$$

   (a) Construya un intervalo de 95% de confianza para μ.

   (b) Construya un intervalo de 95% de confianza para la desviación estándar σ.

9. Una muestra aleatoria, obtenida en 6 días, de la tasa de interés en el banco A es 2.34, 2.01, 2.65, 2.12, 2.76, 3.01. Para el banco B una muestra aleatoria de 6 días en las tasa de interés es 1.89, 2.23, 1.76, 2.34, 2.00, 2.81, 2.96. Estas tasas de interés se pueden asumir como provenientes de distribuciones Normales.

   (a) Muestre que en base al promedio de las tasas observadas del banco A, existen infinitos intervalos de confianza para la tasa de interés promedio. Explique porqué el intervalo simétrico es preferible. Explique porqué no se utiliza la mediana como base para construir un intervalo de confianza para la tasa de interés promedio aunque en muestras grandes también tiene distribución Normal.

   (b) Construya un intervalo de confianza 0.95 para el coeficiente de variación de la tasa de interés en el banco A. Interprete el significado de este intervalo.

   (c) ¿ Existe evidencia en los datos como para decir, con un nivel de confianza razonable, que el banco A tiene tasas más altas ?. ¿ Cómo puede cambiar la respuesta a esta pregunta si las observaciones corresponden a los mismos o distintos días ?. ¿ Cómo se puede validar el supuesto de que las varianzas son iguales ?.

   (d) Suponiendo que las observaciones corresponden a los mismos días, ¿ existe correlación entre las tasas de interés de estos dos bancos ?.

10. Si $Y_1, ..., Y_n$ es una muestra aleatoria de una población normal, explique cómo obtener un intervalo de confianza $(1 - \alpha)$ para la varianza poblacional σ^2 si la media poblacional μ es conocida. Explique en qué sentido el conocimiento de μ mejora la estimación por intervalos en comparación al caso μ desconocido.

11. En una muestra de 100 estudiantes egresados en 1992, el coeficiente de correlación entre la nota promedio en la universidad y el salario en el primer trabajo fue 0.45. En 1982 el coeficiente de correlación era 0.67. ¿ Diría usted, con un nivel razonable de confianza, que la correlación entre estas variables ha cambiado o aceptaría la hipótesis que las diferencias se deben sólo a errores muestrales ?. Explique claramente los supuestos necesarios para hacer el análisis.

12. El contenido efectivo de 15 paquetes de caramelos en gramos es 123, 131, 109, 108, 121, 120, 119, 131, 127, 119, 115, 118, 123, 121, 117.

   (a) Construya un intervalo de confianza 0.95 para la mediana de los pesos.

   (b) Construya un intervalo de confianza 0.50 para el percentil 0.20 de la distribución de los pesos. Interprete su resultado.

Test de Hipótesis

1. Sea $Y_1, ..., Y_n$ una muestra de la distribución Normal con media μ y varianza σ^2. La varianza de la distribución es conocida. Considere las hipótesis

$$H_0 : \mu \leq \mu_0, \qquad H_1 : \mu > \mu_0$$

   (a) Determine el valor de k en función de un nivel de significancia α^* predeterminado si se rechaza H_0 para $\overline{Y} > k$. Muestre que este valor de k puede obtenerse al definir la región de rechazo como

$$Z_c = \frac{\overline{Y} - \mu_0}{\sqrt{\sigma^2/n}} > Z_{1-\alpha^*}.$$

   (b) Mediante gráficos de la distribución Normal con medias μ_0 y $\mu > \mu_0$ represente los errores tipo I y tipo II.

   (c) Muestre que, para un nivel de significancia predeterminado, el error tipo II disminuye cuando aumenta n.

   (d) Calcule el valor de n necesario para que el error tipo II sea igual al nivel de significancia.

2. Para hacer un test de las hipótesis

$$H_0 : \lambda \geq 5, \qquad H_1 : \lambda < 5$$

se dispone de una muestra de tamaño 10 de la distribución Poisson con parámetro λ.

   (a) Calcule el error tipo I para los valores $\lambda = 5, 6, 7$ correspondientes a las regiones de rechazo de H_0 de la forma $\overline{Y} < k$, $k = 1, 2, 3$. Calcule el error tipo II para los valores $\lambda = 3, 4, 5$.

   (b) Determine el valor de k de manera que el nivel de significancia sea menor que 0.10.

   (c) Si el error tipo I máximo aceptado es 0.10, ¿ Porqué se debe tomar el valor de k que hace máximo este error bajo la condición que sea menor que 0.10?.

3. La variabilidad de un proceso en condiciones correctas, y medida en términos de la desviación estandar, es de 3 unidades. Se dispone de una muestra de tamaño 15, con los valores siguientes: 27, 17, 18, 30, 17, 22, 16, 23, 26, 20, 22, 16, 23, 21 y 17.

   (a) ¿ Se puede decir, con un nivel de significancia igual a 0.05, que el proceso está en condiciones correctas ?. Calcule el valor p del test.

   (b) ¿ Con que frecuencia se acepta que el proceso está funcionando en condiciones correctas si la desviación estandar efectiva es 4, 6, 8, 10?. Grafique la función potencia.

4. Una muestra de 10 piezas de acero del proveedor A ha dado una resistencia media a la tracción de 54.000 unidades con $s = 2.100$, mientras que otra muestra de 12 piezas del proveedor B ha resuelto en una media de 49.000 unidades y $s = 1.900$. Las piezas B son más baratas que las A y éstas últimas sólo serían rentables si tuviesen una resistencia media de al menos 2.000 unidades mayor que B sin tener mayor variabilidad. En caso contrario sería mejor comprar a B. ¿Qué decisión se tomaría?.

5. Se desea comparar dos tipos de ampolletas que difieren en el material de su filamento. Para ello se toma una muestra aleatoria de cada material y se mide el tiempo de duración de cada ampolleta. Suponga que para el primer tipo de ampolletas se toma una muestra de tamaño 30 obteniendose que el promedio de las observaciones es 10 unidades de tiempo con 12 observaciones superiores a 8. Para el segundo tipo de ampolletas se toma una muestra de tamaño 60 obteniendose una media igual a 15 y 33 observaciones superiores a 8. Suponiendo que las observaciones tienen distribución Exponencial,

   (a) ¿Puede concluirse que el material del segundo tipo de ampolletas es mejor que el de las primeras?. Use 5% de significancia.

   (b) Sean π_1 y π_2 las probabilidades de que una ampolleta dures más de 8 unidades de tiempo. Desarrolle un test para la hipótesis $\pi_1 < \pi_2$ determinando el valor p. ¿Puede concluir que el material del segundo tipo de ampolletas es mejor?.

6. Se han tomado datos de estaturas de 500 estudiantes de 18 y 19 años en una Universidad, con los resultados siguientes:

Frecuencia	6	17	51	119	149	96	48	12	2
Intervalo en cms.	150	155	160	165	170	175	180	185	190
	155	160	165	170	175	180	185	190	195

¿Puede aceptarse, con niveles razonables de confianza, que los datos provienen de una distribución Normal?. ¿ Cuál es la conclusión si la muestra consiste de los siguientes 8 datos solamente : 172, 170, 139, 192, 191, 175, 169, 168 ?.

7. La legislación impone a los aeropuertos ciertas normas con respecto al ruido emitido por los aviones en el despegue y aterrizaje. En los alrededores, el límite aceptado es de 80 decibeles. Más allá de este límite el aeropuerto debe pagar una multa. Los habitantes aseguran que el ruido en dicha zona sobrepasa el valor límite. El aeropuerto, que asegura que no es cierto, decide contratar un grupo de expertos, quienes para concluir asumen que la intensidad del ruido sigue una distribución $N(\mu,\ 49)$. Si se considera

$$H_0 : \mu = 80, \qquad H_1 : \mu = 78,$$

encuentre la región crítica de Neyman-Pearson para un nivel de significación del 5%. ¿Qué decisión tomarán lo expertos si se obtiene un promedio de 79.1 decibeles en una muestra de 100 aviones?. Encuentre el error tipo II y la potencia del test.

8. Use el lema de Neyman-Pearson para establecer la región crítica al hacer un test de

$$H_0 : \lambda = \lambda_0, \qquad H_1 : \lambda < \lambda_0$$

con un nivel de significancia igual a 0.05 y a partir de una muestra aleatoria $(Y_1, \ldots Y_n)$ de una población con función densidad

$$f(y) = \begin{cases} \lambda^2 y e^{-\lambda y} & , y > 0 \\ 0 & , \text{en otro caso} \end{cases}.$$

9. Se desea realizar una prueba de las hipótesis

$$H_0 : p = p_0, \qquad H_1 : p \neq p_0,$$

donde p es la proporción de elementos con una determinada característica. Para ello se observan n elementos, registrándose si cada uno posee la característica y el número total (de entre los n) que la presenta. Encuentre la regla de decisión mediante el método de razón de verosimilitud.

10. Muestras aleatorias de 250 personas en los grupos de edad 20–30, 30–40, 40–50, 50–60 años se clasificaron de acuerdo al número de horas de sueño diarias, con los siguientes resultados:

Edad	≤ 8	> 8
20–30	180	70
30–40	172	78
40–50	120	130
50–60	125	125

(a) Determine si las necesidades de sueño cambian según los grupos de edad.

(b) Si π_i es la proporción poblacional en el grupo i que tiene (≤ 8) horas de sueño diario, ¿ puede concluir que la proporción en el grupo de 30–40 años es mayor que la del grupo de 50–60 años ?.

(c) Compare las probabilidades π_i considerando todos los pares posibles.¿ En qué difiere su respuesta con la presentada en (a)?.

(d) Suponga que las variables edad y horas de sueño son medidas en forma contínua para una muestra de 120 personas. El coeficiente de correlación en la muestra es 0.41; ¿ existe evidencia como para afirmar que existe una correlación positiva entre la edad y las horas de sueño?.

11. Se sostiene que la vizcosidad del lubricante X es mayor que la del lubricante Y. Se obtienen 10 observaciones pareadas con los siguientes resultados:

X	120	130	128	170	190	210	230	250	270	290
Y	105	115	121	175	183	207	216	230	261	275

Suponiendo distribución Normal, ¿se puede decir, con un nivel de significancia del 5%, que no hay diferencias en las medias?. Si las observaciones son independientes, ¿ hay diferencias en las varianzas al 5%?.

12. Los habitantes de una ciudad reclaman que la distribución de su ingreso local difiere sustancialmente de la distribución nacional de los mismos. Con esta motivación se tomó una muestra aleatoria de 2000 ingresos familiares de dicha ciudad, que fueron clasificados y comparados con los correspondientes porcentajes nacionales. Los datos se muestran en la siguiente tabla. ¿Proveen los datos suficiente evidencia que indique que la distribución del ingreso de la ciudad difiere de la distribución nacional?.

Ingreso	Porcentajes Nacionales.	Ingreso en la ciudad. Frecuencias de clase.
más de 50.000	2	27
25.000-50.000	16	193
20.000-25.000	13	234
15.000-20.000	19	322
10.000-15.000	20	568
5.000 -10.000	19	482
menos de 5.000	11	174
Total	100	2000

13. Según el anuario del INE de España, el número de matrimonios mensuales realizados en España durante el año 1982 en miles de personas ha sido el siguiente

Meses	E	F	M	A	MY	JN	JL	A	S	O	N	D
Frecuencia	20	21	14	14	10	12	9	8	7	6	3	2

¿Es posible pensar que los datos distribuyen Exponencial?, ¿o tal vez Poisson?. Realice los tests correspondientes.

14. Una encuesta realizada a 120 trabajadores mostró que de las empresas grandes, que reunen el 50% de los trabajadores, un 65% del personal es hombre. En empresas medianas y pequeñas, que reunen al 30% y 20% de los trabajadores respectivamente, los porcentajes del personal de sexo masculino son sólo 55% y 51%. ¿ Existe evidencia en los datos como para acusar a las empresas grandes de discriminación sexual ?.