UNIDAD DIDÁCTICA 5-2

INFERENCIA SOBRE UNA POBLACIÓN NORMAL

1. Introducción

En esta parte del Unidad Didáctica 5 se introducen las técnicas básicas de inferencia sobre una población normal, sobre cuya media y desviación típica se desea obtener conclusiones a partir del análisis de los datos de una muestra extraída de la misma.

Para hacerlas más intuitivas, las cuestiones tratadas se introducen a partir del análisis de un problema concreto

Sobre dicho problema se recuerdan, en primer lugar, los conceptos de hipótesis nula, y de riesgos de primera y segunda especie que se vieron en el capítulo anterior y se introduce el concepto de intervalo de confianza, de gran importancia en la Inferencia Estadística.

Se expone también la forma de llevar a cabo los análisis mediante la utilización del paquete estadístico Statgraphics, ya manejado en unidades anteriores.

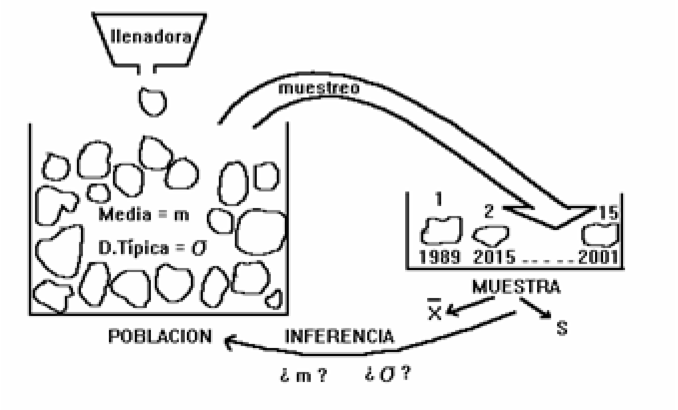
2. Un ejemplo

Una máquina llenadora de bolsas de malla de mandarinas, de las que se usan en los almacenes de confección de esta fruta, se regula para obtener un peso de 2000 gramos. Debido a una serie de causas de variabilidad (variación en el calibre de las frutas, imprecisión en las pesadas automáticas,...) es imposible obtener constantemente bolsas que pesen exactamente 2000 gramos; el peso obtenido es realmente una variable aleatoria, definida sobre la población de todas las bolsas que se confeccionan.

1989	2015	1962	2013	1983	1989	1992	201
1958	2023	1980	1977	1994	2017	2001	

¿Qué puede decirse sobre la media y sobre la varianza de la variable estudiada?

En especial, ¿puede admitirse que la máquina está bien regula¬da o, por el contrario, hay evidencia de que la media es diferente de 2000 y debe, por tanto, procederse a reajustar la máquina?



3. Análisis de los datos

3.1. Normalidad de los datos

La mayor parte de las técnicas de inferencia estadística clásicas sobre variables aleatorias continuas, asumen que las variables muestreadas se distribuyen normalmente. Antes, por tanto, de aplicar estas técnicas es aconsejable analizar hasta qué punto el modelo normal es adecuado, a la vista de los datos obtenidos, para analizar la pauta de variabilidad de la variable estudiada.

Nota importante: todo modelo matemático postulado para el estudio de un fenómeno real (sea la distribución normal, el péndulo matemático, o los rectángulos) implica necesariamente una simplificación y abstracción de ésta. (Carece por tanto de sentido la cuestión de si un determinado modelo es "verdadero" o "correcto" para aplicarlo a una realidad concreta! Sin embargo sí que es importante plantearse si la realidad estudiada es "lo suficientemente parecida" a la postulada por el modelo, como para que la aplicación de éste conduzca a resultados "válidos" o "útiles" en la práctica.

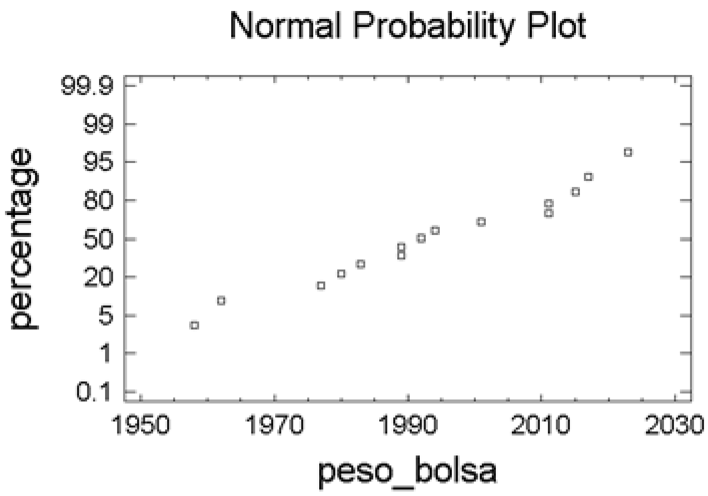
Autoevaluación: ¿Existen en el mundo muchos péndulos constituidos por un hilo inextensible y sin peso del que cuelga una masa puntual sin volumen que oscila sin rozamiento en el vacío?

¿Quiere ello decir que la fórmula del periodo del péndulo matemático no puede aplicarse nunca en la práctica? ¿Se podría aplicar dicha fórmula para calcular aproximadamente el periodo de un péndulo constituido por un trozo de papel colgado de una goma oscilando en medio de un vendaval?

Existen diversos tests estadísticos formales para estudiar la normalidad de unos datos (test Gi-dos, test de Kolmogorov, etcétera...). En nuestra opinión estos tests son de dudosa utilidad práctica, constituyendo "la respuesta correcta a una pregunta equivocada". En efecto, la pregunta a la que pretenden responder (¿proceden mis datos de una distribución normal teórica?) carece de sentido.

La pregunta realmente relevante sería: "¿es la pauta de variabilidad constatada en los datos lo suficientemente parecida a la postulada por el modelo matemático de la distribución normal, como para que pueda aplicarles los resultados teóricos deducidos para ésta?".

A esta pregunta, en vez de con test estadísticos formales de ajuste a distribuciones, se responde mejor a partir de una representación gráfica sencilla como un histograma o un gráfico en papel probabilístico normal. Estas representaciones pueden además detectar la presencia de datos anómalos (posiblemente la más grave y frecuente forma de no normalidad) y sugerir modelos alternativos en los casos en los que el normal aparezca como poco adecuado



Dado que quince datos es un número muy reducido para elaborar un histograma, parece preferible en el ejemplo limitarse a obtener una representación de los datos en papel probabilístico normal. Tal como se ve en la figura, dicha representación no muestra ningún indicio claro de no normalidad, por lo que se asumirá el modelo normal como adecuado para el análisis posterior.

3.2. Análisis descriptivo de la muestra

La primera fase del análisis de cualquier conjunto de datos debe siempre consistir en un estudio descriptivo de los mismos, tanto mediante gráficos (histograma, diagrama Box-Whisker, Plot Normal), como calculando los principales "estadísticos (media, desviación típica, cuartiles, coeficientes de asimetría y curtosis, etc...)

Este análisis puede realizarse, por ejemplo, mediante Statgraphics que proporciona la salida que se recoge a continuación

```
Summary Statistics for peso_bolsa

Count = 15
Average = 1993.6
Median = 1992.0
Variance = 391.971
Standard deviation = 19.7983
Minimum = 1958.0
Maximum = 2023.0
Range = 65.0
Lower quartile = 1980.0
Upper quartile = 2013.0
Skewness = -0.256502
Stnd. skewness = -0.405564
Kurtosis = -0.750953
Stnd. kurtosis = -0.593681
```

A la vista de los resultados la hipótesis de normalidad sigue apareciendo como aceptable, puesto que los valores de los coeficientes de asimetría y curtosis estandarizados son bastante cercanos a cero.

Sin embargo, un simple análisis descriptivo no es suficiente para tomar decisiones a partir de los datos

Autoevaluación: la media muestral ha resultado igual a 1993.6, y es por tanto diferente de 2000 ¿Quiere ello decir que la máquina llenadora se ha desajustado y que, por tanto, hay que detener la producción y proceder a reajustar la máquina?

Procediendo de esta forma sistemáticamente:¿mejoraremos mucho nuestro proceso?, ¿lo empeoraremos?

Si la máquina llenadora está bien ajustada, o sea si su media es m = 2000: ¿Es seguro que en una muestra al azar de 15 bolsas se obtendrá siempre exactamente $\bar{x}$=2000? ¿Es poco probable que pase eso? ¿Es prácticamente imposible que eso ocurra alguna vez?

Necesitamos por tanto algún procedimiento que nos permita obtener conclusiones sobre el valor de m en la población, a partir de la información que hemos obtenido en la muestra. Este tipo de procedimientos constituyen el objeto de la Inferencia Estadística.

3.3. Estudio de la hipótesis m=2000. Enfoque del problema

Dado que la máquina se ajustó inicialmente a 2000 grs. y que no se ha producido ningún hecho excepcional en la planta (la muestra se ha tomado para hacer un control rutinario), parece razonable asumir como hipótesis de salida que m es igual a 2000 a no ser que la información contenida en la muestra indique lo contrario

A esta hipótesis que se toma como base de partida y que, en cierto sentido, refleja nuestro conocimiento previo de la situación, se le denomina en Inferencia Estadística Hipótesis Nula H_0

El proceso de razonamiento seguido para estudiar si esta hipótesis nula es o no admisible, es intuitivamente el siguiente:

Si m = 2000 entonces $\bar{x}$ debería ser ≈2000. Por tanto, se admitirá que m=2000 si $\bar{x}$ es "cercana" a 2000 y se rechazará dicha hipótesis si $\bar{x}$ no es "cercana" a 2000.

Pero...¿qué debería entenderse por "cercana"?

La solución consiste en basarse en el estadístico tcalc $=\frac{\bar{x}-2000}{S/\sqrt{N}}$ (donde N=15 es el tamaño de muestra), que de acuerdo con lo visto en la unidad didáctica 5.1 Distribuciones en el muestreo, seguirá una distribución t de student con 14 grados de libertad (y será por tanto “cercana a cero”) si es cierta la hipótesis de que m es igual a 2000.

Por otra parte, si m es diferente de 2000, la tcalc tomará en general valores más lejanos de cero (positivos si m>2000 ó negativos si m<2000) que los esperables para una variable t de Student. Por tanto, una forma razonable de proceder será rechazar la hipótesis nula si la tcalc resulta demasiado grande (en valor absoluto) para ser una t de Student

Y ¿cómo podemos cuantificar hasta qué punto la tcalc es "demasiado grande" para ser una t de Student? calculando la probabilidad de que una t de Student tome valores tan grandes o más que la tcalc. A esta probabilidad, que puede obtenerse en las tablas de la t, se le denomina el p-value, y contra más baja sea mayor será la evidencia respecto a la falsedad de H_0.

Frecuentemente se opera en Estadística rechazando la hipótesis nula si el p-value es inferior a 0.05. Ello es equivalente, en nuestro ejemplo, a rechazar la hipótesis nula si el estadístico t resulta mayor (en valor absoluto) al valor correspondiente al punto $t_{14}(0.05)$ que, como puede consultarse en la tabla correspondiente, es 2.14.

Como en el ejemplo

$$|t_{calc}| = \left|\frac{\bar{x}-2000}{s/\sqrt{N}}\right| = \left|\frac{1993.6-2000}{19.8/\sqrt{15}}\right| = |-1.25| < 2.14$$

se deduce que la hipótesis nula m = 2000 es aceptable, o sea, que es compatible con los datos observados en la muestra.

Autoevaluación: contra más pequeño sea el p-value, más significativa se dice que es la diferencia entre m y 2000. Que una diferencia sea muy significativa estadísticamente, ¿quiere decir que es muy grande? Cuando en el ejemplo considerado se admite la hipótesis nula, ¿quiere decir que hemos demostrado estadísticamente que m es igual a 2000?

SOLUCIÓN:

Un error muy frecuente, es la confusión entre significación estadística e importancia práctica. El problema es de naturaleza semántica, y deriva de utilizar, para designar un concepto técnico estadístico concreto, un vocablo -“significativo”- que tiene un sentido diferente en el lenguaje habitual. Si la diferencia entre m y 2000 es “muy significativa estadísticamente”, la interpretación práctica correcta es que resulta casi seguro que dicha diferencia no es nula, y no necesariamente que la diferencia en cuestión sea muy importante.

Por otra parte, el que no se rechace una H_0, no significa que se haya demostrado que dicha hipótesis nula es cierta, sino sólo que la misma es compatible con los datos observados, como lo serían probablemente también muchas otras hipótesis alternativas.

En síntesis, la operativa general propuesta para analizar si es admisible la hipótesis H_0: $m = m_0$ es la siguiente:

Calcular

$$t_{calc} = \frac{\bar{x} - m_0}{s/\sqrt{N}}$$

y rechazar H0 si $|t_{calc}| > t_{N-1}(\alpha)$, donde $t_{N-1}(\alpha)$ es un valor, que se busca en tablas tal que

$P(|t_{N-1}| > t_{N-1}(\alpha)) = \alpha$, donde en nuestro ejemplo hemos operado con $\alpha = 0.05$.

Riesgos de 1ª y 2ª especie

α será por lo tanto la probabilidad que tenemos, operando de esta forma, de equivocarnos rechazando H0 cuando es cierta. A α se le denomina, como sabemos, Riesgo de 1ª especie.

(Nota: en general en Estadística la palabra "riesgo" es un término técnico que significa "probabilidad de cometer un error")

En general se utilizan valores de α=0.05 (como en el ejemplo) ó α=0.01 (si se requiere tener un riesgo de 1ª especie muy bajo).

Autoevaluación: ¿No sería mejor operar con riesgos de 1ª especie lo más bajos posibles? ¿Qué problema se presentaría en la práctica si queremos reducir al máximo la probabilidad de equivocarnos rechazando la H_0 cuando sea cierta?

Si se opera con α muy bajo (en el ejemplo: si sólo queremos reajustar la máquina si estamos muy seguros de que se ha desajustado), se aumenta a cambio el Riesgo de 2ª especie, que es, como sabemos, la probabilidad de aceptar H_0 cuando es falsa (no reajustar la máquina pese a haberse desajustado).

Se considera generalmente que α=0.05 es un compromiso razonable entre ambos riesgos.

Nota técnica importante: el ingeniero no debe, sin embargo, obsesionarse ciegamente en sus estudios con unos valores concretos de α (0.05 ó 0.01) La utilización de unos valores límites o “críticos” para el p-value que separan los resultados “significativos” de los “no significativos” no es más que un anacronismo, reflejo de épocas en las que el cálculo exacto de estos p-values se hallaba fuera del alcance del investigador, que sólo podía hacerse una idea al respecto comparando los valores por él obtenidos con los reflejados en unas tablas que generalmente se limitaban a estos dos niveles. Es el p-value, por tanto, lo que refleja el grado de evidencia de unos resultados contra la H_0 y, en consecuencia, lo que debería acompañar al análisis de dichos resultados, y no sólo la constatación de si resulta superior o inferior al 5%. Porque ¿qué diferencia hay, en la práctica, entre un p-value del 4.9% o del 5.1%?

3.4. Intervalo de confianza para m

Que la H_0 m=2000 sea admisible para los datos obtenidos, no quiere decir que hayamos demostrado que m es igual a 2000. De hecho, muchos otros valores hipotéticos de m también hubieran resultado admisibles con los datos obtenidos.

¿Cómo podemos, a partir de la muestra, construir un intervalo que tenga a priori una probabilidad elevada (1-α) de contener el valor desconocido m de la media poblacional?

Como

$$\frac{\bar{x} - m}{s/\sqrt{N}}$$

se distribuye como una t_{N-1}, y siendo $P(|t_{N-1}| > t_{N-1}(\alpha)) = \alpha$, se cumplirá que:

$$P\left(-t_{N-1}(\alpha) < \frac{\bar{x} - m}{s/\sqrt{N}} < +t_{N-1}(\alpha)\right) = 1 - \alpha$$

de donde se obtiene inmediatamente

$$P\left(\bar{x} - t_{N-1}(\alpha)\frac{s}{\sqrt{N}} < m < \bar{x} + t_{N-1}(\alpha)\frac{s}{\sqrt{N}}\right) = 1 - \alpha$$

Por tanto

$$\left[\bar{x} - t_{N-1}(\alpha)\frac{s}{\sqrt{N}}\,;\, \bar{x} + t_{N-1}(\alpha)\frac{s}{\sqrt{N}}\right]$$

constituye un intervalo de confianza con una probabilidad 1-α de contener a m. (A 1-α se le denomina también "nivel de confianza").

Autoevaluación: ¿qué interpretación práctica tiene la probabilidad 1-α asociada a un determinado intervalo de confianza?

SOLUCIÓN:

El aserto probabilístico asociado a un determinado intervalo de confianza, como por ejemplo

$$P\left(\bar{x} - t_{N-1}(\alpha)\frac{s}{\sqrt{N}} < m < \bar{x} + t_{N-1}(\alpha)\frac{s}{\sqrt{N}}\right) = 1 - \alpha$$

debe interpretarse como que existe una probabilidad 1- α de que los dos límites del intervalo (que son aleatorios al depender de los parámetros muestrales $\bar{x}$ y s) comprendan entre ellos al verdadero valor (desconocido) de la media poblacional m, y no como que existe una probabilidad 1- α de que la media esté entre los dos valores obtenidos en un intervalo concreto

La media m no es una variable aleatoria, porque no existe una población a la que vayan asociados los posibles valores de m, y en la que una proporción 1- α de los individuos tengan valores de m comprendidos en dicho intervalo. Lo que sí que existe es una población de posibles muestras, a cada una de las cuales le corresponde un determinado intervalo de confianza, y en la que una proporción 1- α de las muestras llevan asociados intervalos de confianza que cubren al verdadero valor de m.

Por tanto, cuando se dice (hablando con poca precisión) que existe una "probabilidad" 1- α de que un determinado intervalo contenga al verdadero valor de m, el término "probabilidad" no debe entenderse en el sentido técnico frecuencialista, sino en un sentido más coloquial como medida del grado de confianza que se puede tener en que la afirmación anterior es cierta.

En el ejemplo:

$$\left[1993.6 - 2.14\frac{19.8}{\sqrt{15}}\,;1993.6 + 2.14\frac{19.8}{\sqrt{15}}\right] = [1982.7\,;2004.5]$$

constituye un intervalo de confianza para la media m del proceso, para un nivel de confianza del 95%.

Nota: como es lógico, el intervalo contiene el valor m=2000, lo que era de esperar dado que la hipótesis m=2000 ha resultado aceptable.

3.5. Intervalo de confianza para σ

Siendo s^2 la varianza muestral obtenida en una muestra de tamaño N extraída de una población Normal de varianza σ^2, se demuestra que

$$(N-1)\frac{s^2}{\sigma^2}$$

sigue una distribución Gi-dos con N-1 grados de libertad.(La distribución Gi-dos fue estudiada en la unidad didáctica anterior).

A partir de este resultado es sencillo obtener un intervalo de confianza para σ, tal como se expone a continuación.

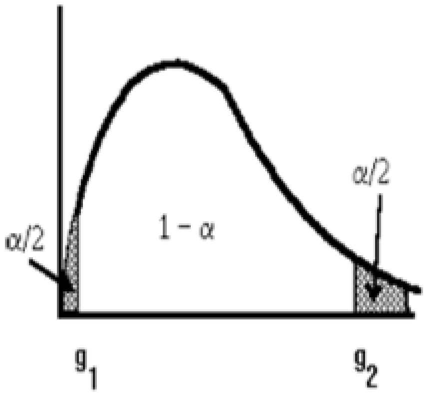
En efecto, como

$$(N-1)\frac{s^2}{\sigma^2}$$

sigue una distribución Gi-dos con (N-1) grados de libertad, de la tabla de la Gi-dos es posible obtener dos valores, g_1 y g_2, tales que

$P(g1<\chi^2 <g2) = 1- \alpha$

(Por ejemplo, $P(5.63< \chi2<26.1) = 0.95$)



Por tanto,

P(g1< $(N-1)\dfrac{s^2}{\sigma^2}$ <g2)

Es igual a 1- α, de donde se obtiene

$$P\left((N-1)\frac{s^2}{g_2} < \sigma^2 < (N-1)\frac{s^2}{g_1}\right) = 1-\alpha$$

constituirá, en consecuencia un intervalo de confianza para la varianza poblacional σ^2, y

$$\left[\sqrt{(N-1)\frac{s^2}{g_2}}\ ;\ \sqrt{(N-1)\frac{s^2}{g_1}}\right]$$

será un intervalo de confianza para la desviación típica poblacional σ.

En el ejemplo

$$\left[\sqrt{(15-1)\frac{19.8^2}{26.1}}\ ;\ \sqrt{(15-1)\frac{19.8^2}{5.63}}\right]$$

= [14.5 ; 31.2] será el intervalo de confianza, con un nivel de confianza del 95%, para σ.

4. Análisis mediante statgraphics

La obtención de intervalos de confianza para la media y la varianza de una población, así como el contraste de una hipótesis sobre el valor de la media, puede llevarse a cabo mediante la opción describe ... numeric data ... one-variable analysis del paquete Statgraphics.

La única entrada que hay que dar es el nombre de la variable en la que están guardados los valores muestrales. (En el ejemplo a dicha variable, que contiene los 15 valores observados para los pesos de las bolsas, se le ha denominado peso-bolsas)

El programa proporciona en primer lugar los estadísticos básicos descriptivos de la muestra. Mediante el botón derecho del ratón es posible solicitar un test de hipótesis para un determinado valor m_0 de la media y el riesgo α de 1ª especie que se desee (m_0 y α se fijan dentro de pane_options).

El resultado es:

```
Hypothesis Tests for peso_bolsa

Sample mean = 1993.6
Sample median = 1992.0


t-test
------
Null hypothesis: mean = 2000.0
Alternative: not equal

Computed t statistic = -1.25198
P-Value = 0.231089

Do not reject the null hypothesis for alpha = 0.05.
```

También mediante el botón derecho del ratón es posible solicitar intervalos de confianza, con un determinado nivel de confianza, para m y para □:

```
Confidence Intervals for peso_bolsa
-----------------------------------
95.0% confidence interval for mean: 1993.6 +/- 10.9639   [1982.64;2004.56]
95.0% confidence interval for standard deviation: [14.4948;31.2238]
```

Fuentes

Métodos Estadísticos en Ingeniería (Romero Villafranca, Rafael)