3.1. Dada una muestra con tres datos se construye el estadístico

$$\psi = x_1 + \frac{1}{2}x_2 - \frac{1}{2}x_3$$

Estúdiese si se trata de un estimador fiel y consistente de la media de la población.

$$E\{\psi\} = E\{x_1 + \frac{1}{2}x_2 - \frac{1}{2}x_3\} = E\{x_1\} + \frac{1}{2}E\{x_2\} - \frac{1}{2}E\{x_3\}$$

Considerando muestreo aleatorio simple

$$E\{x_1\} = E\{x_2\} = E\{x_3\} = \mu$$

$$E\{\psi\} = \mu + \frac{1}{2}\mu - \frac{1}{2}\mu = \mu$$ es un estimador fiel

Calculemos ahora la varianza de ψ

$$\sigma^2(\psi) = E\{(\psi - E\{\psi\})^2\} = E\{(\psi-\mu)^2\} =$$

$$= E\{(x_1 + \frac{1}{2}x_2 - \frac{1}{2}x_3 - \mu)^2\} = E\{(x_1 - \mu + \frac{1}{2}x_2 - \frac{1}{2}\mu - \frac{1}{2}x_3 + \frac{1}{2}\mu)^2\} =$$

$$= E\{(x_1-\mu)^2\} + \frac{1}{4}E\{(x_2-\mu)^2\} + \frac{1}{4}E\{(x_3-\mu)^2\} +$$

$$+ \frac{2}{2}E\{(x_1-\mu)(x_2-\mu)\} + \frac{2}{2}E\{(x_1-\mu)(\mu-x_3)\} + \frac{2}{4}E\{(x_2-\mu)(\mu-x_3)\}$$

Considerando x_i como variables aleatorias independientes (no correlacionadas) entonces

$$E\{(x_1-\mu)(x_2-\mu)\} = E\{(x_1-\mu)(x_3-\mu)\} = E\{(x_2-\mu)(x_3-\mu)\} = 0$$

Por lo tanto

$$\sigma^2(\psi) = E\{(x_1-\mu)^2\} + \frac{1}{4}E\{(x_2-\mu)^2\} + \frac{1}{4}E\{(x_3-\mu)^2\} =$$

$$= \sigma^2 + \frac{1}{4}\sigma^2 + \frac{1}{4}\sigma^2 = \frac{3}{2}\sigma^2$$

Para estudiar la consistencia, necesitamos establecer una extensión general de ψ a muestras con tamaño n.
Consideremos n impar, entonces

$$\psi = x_1 + \frac{1}{n-1}x_2 + \frac{1}{n-1}x_3 + \frac{1}{n-1}x_4 - \dots + \frac{(-1)^i}{n-1}x_i + \dots + \frac{1}{n-1}x_n$$

Obviamente esta expresión incluye el caso $n=3$ anterior. Se verifica que

$$E\{\psi\} = \mu$$

Por otra parte

$$\sigma^2(\psi) = \sigma^2 + \underbrace{\frac{1}{(n-1)^2}\sigma^2 + \frac{1}{(n-1)^2}\sigma^2 + \dots + \frac{1}{(n-1)^2}\sigma^2}_{n-1}$$

$$\sigma^2(\psi) = \sigma^2 + \frac{1}{n-1}\sigma^2 = \frac{n-1+1}{n-1}\sigma^2 = \frac{n}{n-1}\sigma^2$$

$$\lim_{n\to\infty}\sigma^2(\psi) = \lim_{n\to\infty}\frac{n}{n-1}\sigma^2 = \sigma^2$$

Por lo tanto no es consistente!

3.2. Consideremos el estimador fiel de la varianza

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

Sabemos que $E\{s^2\} = \sigma^2$

Calculemos ahora $\sigma^2(s^2)$

$$\sigma^2(s^2) = E\{(s^2 - E\{s^2\})^2\} = E\{(s^2-\sigma^2)^2\} = E\{s^4\} - (\sigma^2)^2$$

$$\sigma^2(s^2) = E\{(s^2)^2\} - \sigma^4$$

- Como el cálculo general es algo extenso, podemos empezar por el caso de variables normales (gaussianas) en donde sabemos que

$$\frac{(n-1)s^2}{\sigma^2} = \chi^2_{n-1}$$

Recordemos que la χ^2 de Pearson con n grados de libertad cumple

$$E\{\chi^2_n\} = n$$

$$E\{(\chi^2_n - n)^2\} = 2n$$

es decir que $\sigma^2(\chi^2_{n-1}) = 2n$. Considerando la igualdad anterior tendremos que

$$\sigma^2\left[\frac{(n-1)s^2}{\sigma^2}\right] = \sigma^2(\chi^2_{n-1})$$

$$\frac{(n-1)^2}{\sigma^4}\sigma^2(s^2) = 2(n-1)$$

$$\boxed{\sigma^2(s^2) = \frac{2\sigma^4}{n-1}}$$ gaussiano

3.3 Una variable aleatoria verifica que la media de la población es nula $\mu=0$. Considerando una muestra de n datos, estudiar si el estimador t es un estimador fiel de la varianza ③

$$t=\frac{1}{n-1}\sum_{i=1}^{n}x_i^2$$

$$E\{t\}=\frac{1}{n-1}\sum_{i=1}^{n}E\{x_i^2\}$$

observemos que $E\{x_i\}=0$, por lo que

$$\sigma^2=E\{(x-E\{x\})^2\}=E\{x^2\}$$

de lo que deducimos que

$$E\{x_i^2\}=\sigma^2$$

$$E\{t\}=\frac{1}{n-1}\cdot n\sigma^2=\frac{n}{n-1}\sigma^2$$

Por lo tanto este estimador no es fiel.

Curiosamente sería fiel

$$t=\frac{1}{n}\sum_{i=1}^{n}x_i^2$$

3.4. Dadas dos variables aleatorias se toma una muestra de n pares de datos $\{(x_1,y_1);(x_2,y_2);\dots;(x_n,y_n)\}$

Consideramos

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i \qquad \bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i$$

$$E\{x\}=\mu \qquad E\{y\}=\nu$$

probar que $\quad cov(\bar{x},\bar{y})=\dfrac{cov(x,y)}{n}$

Observemos que

$$\operatorname{cov}(x,y) = E\{(x-\mu)(y-\nu)\} = E\{(x_i-\mu)(y_i-\nu)\}$$

Por lo que tendremos

$$\operatorname{cov}(\bar{x},\bar{y}) = E\{(\bar{x}-\mu)(\bar{y}-\nu)\} = E\left\{\left(\frac{1}{n}\sum_{i=1}^{n} x_i-\mu\right)\left(\frac{1}{n}\sum_{j=1}^{n} y_j-\nu\right)\right\} =$$

$$= E\left\{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)\right)\left(\frac{1}{n}\sum_{j=1}^{n}(y_j-\mu)\right)\right\} =$$

$$= \frac{1}{n^2} E\left\{\left(\sum_{i=1}^{n}(x_i-\mu)\right)\left(\sum_{j=1}^{n}(y_j-\nu)\right)\right\} =$$

$$= \frac{1}{n^2} E\left\{\sum_{i=1}^{n}(x_i-\mu)(y_i-\nu) + 2\sum_{i<j}(x_i-\mu)(y_j-\nu)\right\} =$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\underbrace{E\{(x_i-\mu)(y_i-\nu)\}}_{\operatorname{cov}(x,y)} + \frac{2}{n^2}\sum_{i<j}\underbrace{E\{(x_i-\mu)(y_j-\nu)\}}_{0} =$$

$$= \frac{1}{n^2}\cdot n\cdot \operatorname{cov}(x,y) = \frac{1}{n}\operatorname{cov}(x,y)$$

3.5. Consideremos una muestra de n datos proveniente de una población madre normal, calcular la varianza del estadístico

$$\lambda = \frac{\bar{x}-\mu}{s/\sqrt{n}}$$

Recordemos que se trata de una variable de t de Student con $n-1$ grados de libertad. Por otro lado

$$E\{\lambda\} = 0$$

ya que se trata de una distribución simétrica observamos que

$$\sigma^2(\lambda) = E\{(\lambda - E\{\lambda\})^2\} = E\{\lambda^2\}$$

Por otra parte λ sigue una distribución de t de Student con $n-1$ grados de libertad. Recordemos que si una variable t sigue una distribución de t de Student con n grados de libertad su densidad de probabilidad viene dada por

$$f(t) = \frac{\Gamma\left[\frac{1}{2}(n+1)\right]}{\Gamma\left[\frac{1}{2}n\right]\sqrt{\pi \cdot n}} \left[1 + t^2/n\right]^{-\frac{1}{2}(n+1)}$$

En este caso nos interesa averiguar

$$E\{t^2\} = \frac{\Gamma\left[\frac{1}{2}(n+1)\right]}{\Gamma\left[\frac{1}{2}n\right] \cdot \sqrt{\pi \cdot n}} \int_{-\infty}^{\infty} dt \cdot t^2 \cdot \left(1 + t^2/n\right)^{-\frac{1}{2}(n+1)} =$$

$$= \frac{\Gamma\left[\frac{1}{2}(n+1)\right]}{\Gamma\left[\frac{1}{2}n\right]\sqrt{\pi \cdot n}} \cdot 2 \int_{0}^{\infty} dt \cdot t^2 \cdot \left(1 + t^2/n\right)^{-\frac{1}{2}(n+1)} =$$

Consideremos el cambio de variable $z = t^2/n$; $t = \sqrt{z \cdot n}$

$dz = \frac{2t}{n} dt$

$$E\{t^2\} = \frac{\Gamma\left[\frac{1}{2}(n+1)\right]}{\Gamma\left[\frac{1}{2}n\right] \cdot \sqrt{\pi}\sqrt{n}} \cdot \int_0 \frac{2\, dt \cdot t}{n} \cdot n \cdot t \cdot \frac{1}{\left(1 + t^2/n\right)^{\frac{1}{2}(n+1)}} =$$

$$= \frac{\Gamma\left[\frac{1}{2}(n+1)\right]}{\Gamma\left[\frac{1}{2}n\right] \cdot \sqrt{\pi}\sqrt{n}} \cdot \int_0^{\infty} dz \cdot n\sqrt{z \cdot n} \quad \frac{1}{(1+z)^{\frac{1}{2}(n+1)}} =$$

$$= \frac{\Gamma\left[\frac{1}{2}(n+1)\right]}{\Gamma\left[\frac{1}{2}n\right]\sqrt{\pi}} \cdot n \cdot \int_0^{\infty} dz \; \frac{z^{\frac{1}{2}}}{(1+z)^{\frac{1}{2}(n+1)}}$$

Esta integral corresponde a la función Beta, definida como

$$B(x,y) = \int_0^\infty du \, \frac{u^{x-1}}{(1+u)^{x+y}} = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$$

Por lo tanto en nuestro caso tendremos que

$$\frac{1}{2} = x-1 \, ; \quad \frac{1}{2}(n+1) = x+y \, ;$$

$$x = 3/2 \, ; \quad y = \frac{1}{2}(n+1) - 3/2 = \frac{1}{2}n - 1$$

En consecuencia

$$E\{t^2\} = \frac{\Gamma\left[\frac{1}{2}(n+1)\right]}{\Gamma\left[\frac{1}{2}n\right]\sqrt{\pi}} \cdot n \cdot B\left(3/2, \tfrac{1}{2}n - 1\right) =$$

$$= \frac{\Gamma\left[\frac{1}{2}(n+1)\right]}{\Gamma\left[\frac{1}{2}n\right]\sqrt{\pi}} \cdot n \, \frac{\Gamma\left[3/2\right] \cdot \Gamma\left[\frac{1}{2}n - 1\right]}{\Gamma\left[\frac{1}{2}n - 1 + \frac{3}{2}\right]} =$$

$$= \frac{\cancel{\Gamma\left[\frac{1}{2}(n+1)\right]}}{\Gamma\left[\frac{1}{2}n\right] \cdot \sqrt{\pi}} \cdot n \cdot \frac{\Gamma\left[3/2\right]\Gamma\left[\frac{1}{2}n - 1\right]}{\cancel{\Gamma\left[\frac{1}{2}n + 1/2\right]}}$$

Por otra parte

$$\Gamma\left[\tfrac{1}{2}n\right] = \left(\tfrac{1}{2}n - 1\right)\Gamma\left[\tfrac{1}{2}n - 1\right]$$

$$\Gamma\left[3/2\right] = 1/2 \, \Gamma\left[1/2\right] = \tfrac{1}{2} \cdot \sqrt{\pi}$$

$$E\{t^2\} = \frac{n \cdot 1/2 \cdot \sqrt{\pi} \; \cancel{\Gamma\left[\frac{1}{2}n - 1\right]}}{\sqrt{\pi}\left(\frac{1}{2}n - 1\right)\cancel{\Gamma\left[\frac{1}{2}n - 1\right]}} = \frac{n}{n-2}$$

Como consecuencia de esta propiedad de la t-Student tendremos que para λ

$$\boxed{\sigma^2(\lambda) = \frac{n-1}{n-3}}$$

* Obsérvese que obviamente esta expresión sólo es válida para $n>3$.

3.6 Considerando una muestra de n datos de la forma $\{(x_1,y_1);(x_2,y_2);\ldots;(x_n,y_n)\}$ probar que

$$s(x,y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$$

es un estimador fiel de la covarianza de x, y;

Consideremos que

$$E\{s(x,y)\} = \frac{1}{n-1}\sum_{i=1}^{n}E\{(x_i-\bar{x})(y_i-\bar{y})\} =$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}E\{(x_i-\mu+\mu-\bar{x})(y_i-\nu+\nu-\bar{y})\} =$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\Big[E\{(x_i-\mu)(y_i-\nu)\}+E\{(x_i-\mu)(\nu-\bar{y})\}+$$
$$+E\{(\mu-\bar{x})(y_i-\nu)\}+E\{(\mu-\bar{x})(\nu-\bar{y})\}\Big]$$

Ahora bien

$$E\{(x_i-\mu)(\nu-\bar{y})\} = -E\{(x_i-\mu)\left(\frac{1}{n}\sum_{j=1}^{n}y_j-\nu\right)\} =$$

$$= -\frac{1}{n}\sum_{j=1}^{n}E\{(x_i-\mu)(y_j-\nu)\} = -\frac{1}{n}E\{(x_i-\mu)(y_i-\nu)\}$$

Idénticamente

$$E\{(\mu-\bar{x})(y_i-\nu)\} = -\frac{1}{n}\sum_{j=1}^{n}E\{(x_j-\mu)(y_i-\nu)\} =$$

$$= -\frac{1}{n}E\{(x_i-\mu)(y_i-\nu)\}$$

El término de los valores medios será

$$E\{(\bar{x}-\mu)(\bar{y}-\nu)\} = \operatorname{cov}(\bar{x},\bar{y}) = \frac{1}{n}\operatorname{cov}(x,y)$$

según hemos visto en un problema anterior.
Sustituyendo tendremos que

$$E\{S(x,y)\} = \frac{1}{n-1}\left[\sum_{i=1}^{n} E\{(x_i-\mu)(y_i-\nu)\} - \right.$$

$$- \frac{1}{n}\sum_{i=1}^{n} E\{(x_i-\mu)(y_i-\nu)\} - \frac{1}{n}\sum_{i=1}^{n} E\{(x_i-\mu)(y_i-\nu)\} +$$

$$\left. + \sum_{i=1}^{n} E\{(\bar{x}-\mu)(\bar{y}-\nu)\}\right] =$$

$$= \frac{1}{n-1}\left[n\cdot\operatorname{cov}(x,y) - n\cdot\frac{1}{n}\operatorname{cov}(x,y) - n\cdot\frac{1}{n}\cancel{\operatorname{cov}(x,y)} + n\,\cancel{\frac{1}{n}\operatorname{cov}(x,y)}\right] =$$

$$= \frac{1}{n-1}(n-1)\operatorname{cov}(x,y) = \operatorname{cov}(x,y)$$

Por lo tanto se trata de un estimador fiel.

37. Consideremos $\{x_1, x_2, \ldots, x_n\}$ son variables independientes con media μ y desviación típica σ provenientes de una misma distribución de probabilidad (incluso si no es normal tb. es válido). Construimos las variables

$$y_{j-1} = \frac{1}{\sqrt{j(j-1)}}\left[x_1 + x_2 + \ldots + x_{j-1} - (j-1)x_j\right]$$

$$2 \leq j \leq n$$

Veamos cuáles son sus propiedades.

- Por un lado

$$E\{y_{j-1}\} = \frac{1}{\sqrt{j(j-1)}} E\{[x_1 + x_2 + \dots + x_{j-1} - (j-1)x_j]\} =$$

$$= \frac{1}{\sqrt{j(j-1)}} [E\{x_1\} + E\{x_2\} + \dots + E\{x_{j-1}\} - (j-1)E\{x_j\}]$$

Como $E\{x_1\} = \dots = E\{x_j\} = \mu$

$$E\{y_{j-1}\} = \frac{1}{\sqrt{j(j-1)}} [\underbrace{\mu + \mu + \dots + \mu}_{j-1} - (j-1)\mu] = 0$$

- Por otra parte y respecto a la varianza se verificará que

$$\sigma^2(y_{j-1}) = E\{(y_{j-1} - E\{y_{j-1}\})^2\} = E\{y_{j-1}^2\}$$

$$y_{j-1} = \frac{1}{\sqrt{j(j-1)}} [(x_1-\mu) + (x_2-\mu) + \dots + (x_{j-1}-\mu) - (j-1)(x_j-\mu)]$$

$$y_{j-1}^2 = \frac{1}{j(j-1)} [(x_1-\mu)^2 + (x_2-\mu)^2 + \dots + (x_{j-1}-\mu)^2 + (j-1)^2(x_j-\mu)^2 + \\ + 2(x_1-\mu)(x_2-\mu) + 2(x_1-\mu)(x_3-\mu) + \dots \dots + \\ \cancel{+} 2(j-1)(x_{j-1}-\mu)(x_j-\mu)]$$

Por lo que

$$E\{y_{j-1}^2\} = \frac{1}{j(j-1)} [E\{(x_1-\mu)^2\} + E\{(x_2-\mu)^2\} + \dots + E\{(x_{j-1}-\mu)^2\} + (j-1)^2 E\{(x_j-\mu)^2\} + \\ + 2\underbrace{E\{(x_1-\mu)(x_2-\mu)\}}_{=0} + 2\underbrace{E\{(x_1-\mu)(x_3-\mu)\}}_{=0} + \dots \cancel{+}2(j-1)\underbrace{E\{(x_{j-1}-\mu)(x_j-\mu)\}}_{=0}]$$

$$= \frac{1}{j(j-1)} [\underbrace{\sigma^2 + \sigma^2 + \dots + \sigma^2}_{j-1} + (j-1)^2\sigma^2] = \frac{(j-1)[j-1+1]\sigma^2}{j(j-1)} = \sigma^2$$

En el caso de que las variables $\{x_1, x_2, .., x_n\}$ sean normales la variable y_{j-1} será normal ya que es suma de variables normales y esta distribución de probabilidad es estable.

3.8. Demostrar que las variables definidas en la transformación (3.40) cumplen que

$$\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i^2$$

recordemos que

$$y_1 = \frac{1}{\sqrt{1 \cdot 2}} (x_1 - x_2)$$

$$y_2 = \frac{1}{\sqrt{2 \cdot 3}} (x_1 + x_2 - 2x_3)$$

$$\vdots$$

$$y_{n-1} = \frac{1}{\sqrt{(n-1)n}} (x_1 + x_2 + \dots + x_{n-1} - (n-1)x_n)$$

$$y_n = \frac{1}{\sqrt{n}} (x_1 + \dots + x_n) = \sqrt{n} \cdot \bar{x}$$

Veamos que se cumple en los primeros casos

$n=2$

$$y_1 = \frac{1}{\sqrt{1 \cdot 2}} (x_1 - x_2)$$

$$y_2 = \frac{1}{\sqrt{2}} (x_1 + x_2)$$

$$y_1^2 + y_2^2 = \frac{1}{2} [x_1^2 + x_2^2 - 2x_1 x_2 + x_1^2 + x_2^2 + 2x_1 x_2] = x_1^2 + x_2^2$$

$n=3$ $\quad y_1 = \frac{1}{\sqrt{1\cdot 2}}(x_1 - x_2)$

$$y_2 = \frac{1}{\sqrt{2\cdot 3}}(x_1 + x_2 - 2x_3)$$

$$y_3 = \frac{1}{\sqrt{3}}(x_1 + x_2 + x_3)$$

$$y_1^2 = \frac{1}{2}(x_1^2 + x_2^2 - 2x_1x_2)$$

$$y_2^2 = \frac{1}{6}(x_1^2 + x_2^2 + 4x_3^2 + 2x_1x_2 - 4x_1x_3 - 4x_2x_3)$$

$$y_3^2 = \frac{1}{3}(x_1^2 + x_2^2 + x_3^2 + 2x_1x_2 + 2x_1x_3 + 2x_2x_3)$$

$$y_1^2 + y_2^2 + y_3^2 = \left(\frac{1}{2} + \frac{1}{6} + \frac{1}{3}\right)x_1^2 + \left(\frac{1}{2} + \frac{1}{6} + \frac{1}{3}\right)x_2^2 + \left(\frac{4}{6} + \frac{1}{3}\right)x_3^2 +$$

$$+\left(-1 + \frac{2}{6} + \frac{2}{3}\right)x_1x_2 + \left(-\frac{4}{6} + \frac{2}{3}\right)x_1x_3 + \left(-\frac{4}{6} + \frac{2}{3}\right)x_2x_3 =$$

$$= x_1^2 + x_2^2 + x_3^2$$

La demostración completa la haremos por inducción en el índice n. Supongamos que se cumple en el caso n veamos que esto implica que se cumple en el $n+1$

n $\quad \tilde{y}_1 = \frac{1}{\sqrt{1\cdot 2}}(x_1 - x_2)$

$$\vdots$$

$$\tilde{y}_{n-1} = \frac{1}{\sqrt{(n-1)n}}(x_1 + x_2 + \dots + x_{n-1} - (n-1)x_n)$$

$$\tilde{y}_n = \frac{1}{\sqrt{n}}(x_1 + x_2 + \dots + x_n)$$

Se verifica que $\quad x_1^2 + x_2^2 + \dots + x_n^2 = \tilde{y}_1^2 + \tilde{y}_2^2 + \dots + \tilde{y}_n^2$

Consideremos ahora el caso $n+1$.

$$\left.\begin{aligned} y_1 &= \frac{1}{\sqrt{1\cdot 2}}(x_1 - x_2) \\ &\vdots \\ y_n &= \frac{1}{\sqrt{n(n+1)}}(x_1 + \dots + x_n - n x_{n+1}) \\ y_{n+1} &= \frac{1}{\sqrt{n+1}}(x_1 + x_2 + \dots + x_{n+1}) \end{aligned}\right\}$$

Observemos que

$$y_1^2 + y_2^2 + \dots + y_n^2 + y_{n+1}^2 = y_1^2 + y_2^2 + \dots + y_{n-1}^2 + \tilde{y}_n^2 - \tilde{y}_n^2 + y_n^2 + y_{n+1}^2$$

Además $y_1 = \tilde{y}_1$; $y_2 = \tilde{y}_2$; $\dots$; $y_{n-1} = \tilde{y}_{n-1}$

$$y_1^2 + y_2^2 + \dots + y_n^2 + y_{n+1}^2 = \underbrace{\tilde{y}_1^2 + \tilde{y}_2^2 + \dots + \tilde{y}_{n-1}^2 + \tilde{y}_n^2}_{= \, x_1^2 + x_2^2 + \dots + x_n^2} - \tilde{y}_n^2 + y_n^2 + y_{n+1}^2$$

Por lo tanto

$$y_1^2 + y_2^2 + \dots + y_n^2 + y_{n+1}^2 = x_1^2 + x_2^2 + \dots + x_n^2 + \underbrace{y_n^2 - \tilde{y}_n^2 + y_{n+1}^2}$$

Analizaremos los últimos tres términos

$$y_n^2 = \frac{1}{n(n+1)}\left[x_1 + \dots + x_n - n x_{n+1}\right]^2$$

$$\tilde{y}_n^2 = \frac{1}{n}(x_1 + \dots + x_n)^2$$

$$y_{n+1}^2 = \frac{1}{n+1}(x_1 + \dots + x_n + x_{n+1})^2$$

De donde

$$\begin{aligned} y_n^2 - \tilde{y}_n^2 + y_{n+1}^2 = {} & \frac{1}{n(n+1)}\left[(x_1 + \dots + x_n)^2 - 2n(x_1 + \dots + x_n)x_{n+1} + n^2 x_{n+1}^2\right] + \\ & - \frac{1}{n}(x_1 + \dots + x_n)^2 + \\ & + \frac{1}{n+1}\left[(x_1 + \dots + x_n)^2 + 2x_{n+1}(x_1 + \dots + x_n) + x_{n+1}^2\right] \end{aligned}$$

De lo que obtenemos

$$y_n^2 - \tilde{y}_n^2 + y_{n+1}^2 =$$

$$= \frac{1}{n(n+1)} \Big[(x_1+\dots+x_n)^2 - 2n(x_1+\dots+x_n)x_{n+1} + n^2x_{n+1}^2 -$$

$$- (n+1)(x_1+\dots+x_n)^2 + n(x_1+\dots+x_n)^2 +$$

$$+ n\cdot x_{n+1}^2 + 2n\,x_{n+1}(x_1+\dots+x_n)\Big] =$$

$$= \frac{1}{n(n+1)}\left[n^2 x_{n+1}^2 + n\,x_{n+1}^2\right] = \frac{n(n+1)}{n(n+1)} x_{n+1}^2 = x_{n+1}^2$$

Volviendo a la primera expresión

$$y_1^2+y_2^2+\dots+y_n^2+y_{n+1}^2 = x_1^2+x_2^2+\dots+x_n^2+y_n^2-\tilde{y}_n^2+y_{n+1}^2 =$$

$$= x_1^2+x_2^2+\dots+x_n^2+x_{n+1}^2$$

3.9. En un experimento se ha tomado una serie de 10 datos obteniéndose los resultados

$$\{18, 21, 23, 19, 20, 21, 20, 19, 20, 17\} \quad n=10$$

Considerando que la distribución madre es normal indicar un intervalo de confianza al 95% para la media y la varianza.

i) como no nos indican la σ^2, debemos considerar la tipificación en una variable de tipo t-Student con $n-1$ grados de libertad

$$\frac{\bar{x}-\mu}{s/\sqrt{n}} \in t_{n-1}$$

$$\bar{x} - \frac{s}{\sqrt{n}} t_{n-1,\alpha/2} \leq \mu \leq \bar{x} + \frac{s}{\sqrt{n}} t_{n-1;\frac{\alpha}{2}}$$

- En el caso de muestra muestra tendremos que

$\bar{x} = 19.8$ $\quad s^2 = 2.844$ $\quad s = 1.686$

siendo $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$; $\quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$; $\quad s = \sqrt{s^2}$

- Puesto que el enunciado nos indica un nivel de confianza del 95%, $1-\alpha = 0.95$; $\alpha = 0.05$

$t_{\frac{\alpha}{2}, n-1} = t_{0,025;9} = 2.262$

El intervalo de confianza al 95% será

$$\bar{x} - \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2};n-1} \leq \mu \leq \bar{x} + \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2};n-1}$$

$$19.8 - \frac{1,686}{\sqrt{10}} \cdot 2.262 \leq \mu \leq 19.8 + \frac{1,686}{\sqrt{10}} \cdot 2.262$$

$$18.594 \leq \mu \leq 21.006$$

ii) En el caso de la varianza podemos construir un intervalo de confianza para σ^2 considerando que

$$\frac{(n-1)s^2}{\sigma^2} \in \chi^2_{n-1}$$

Considerando los percentiles de la χ^2 con n-1 grados de libertad

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2;n-1}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2};n-1}}$$

a un nivel de confianza $1-\alpha$

$$\text{prob}\left(\chi^2 \leq \chi^2_{\alpha/2;n-1}\right) = 1 - \alpha/2$$

considerando nuestro caso particular

$$\left.\begin{array}{l} \chi^2_{0,025;9} = 19.023 \\ \chi^2_{0,975;9} = 2.700 \end{array}\right\}$$

Por lo que el intervalo de confianza será

$$\frac{(10-1).\ 2.844}{19.023} \leq \sigma^2 \leq \frac{(10-1).\ 2.844}{2.700}$$

$$1.345 \leq \sigma^2 \leq 9.480$$

3.10 Se realiza una encuesta a 100 personas antes de unas elecciones para averiguar la intención de voto. De estos 100, 36 manifiestan su intención de votar por el partido A. Calcular con un nivel de confianza del 99% la proporción de votantes del partido A. ¿A cuántas personas habría que entrevistar para conocer esta proporción con un 2% de incertidumbre?

Si consideramos el intervalo de confianza para una distribución binomial tendremos que el estimador

$$\hat{p} = \frac{x}{n};$$ x nº de casos favorables, n " " " totales

$$\hat{p} = \frac{36}{100} = 0,36 \qquad \sigma(\hat{p}) \simeq \sqrt{\frac{p(1-p)}{n}} \simeq \sqrt{\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}}$$

Consideraremos el límite normal, esto es,

$$z \simeq \frac{\hat{p}-p}{\sqrt{\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}}} = \frac{\frac{x}{n}-p}{\sqrt{\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}}}$$

De estas hipótesis se deduce un intervalo de confianza basado en los percentiles de la normal

$$\frac{x}{n} - z_{\frac{\alpha}{2}} \sqrt{\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}} \leq p \leq \frac{x}{n} + z_{\frac{\alpha}{2}} \sqrt{\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}}$$

$$z_{\alpha/2} = z_{0.005} = +2.576$$

$$0.36 - 2.576 \sqrt{\frac{0.36\,(1-0.36)}{100}} \leq p \leq 0.36 + 2.576 \sqrt{\frac{0.36\,(1-0.36)}{100}}$$

$$0.2364 \leq p \leq 0.4836$$

Observemos que en este caso la semianchura del intervalo de confianza es

$$\Delta_n = z_{\frac{\alpha}{2}} \sqrt{\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}} = 0.1236$$

y la incertidumbre relativa sobre p es

$$u_r \simeq \frac{\Delta_n}{p} = \frac{0.1236}{0.36} = 0.34$$

es decir de un 34% ; si queremos disminuirla hasta un 2% considerando que la relación x/n se mantiene aproximadamente constante, tendremos

$$\frac{\Delta_n}{\Delta_{n'}} = \sqrt{\frac{n'}{n}} = \frac{0.34}{0.02} = 17$$

$$n' = (17)^2 \cdot n = (17)^2 \cdot 100 = 28900$$

3.11

se mide la actividad de una fuente radioactiva con un contador de pozo obteniéndose 3224 cuentas en un minuto. Obténgase con un nivel de confianza del 99% un intervalo para el valor medio de la tasa de desintegraciones por minuto.

Está claro que sólo tenemos una medida, por lo que la afirmación más evidente es

$$\mu \simeq 3224 \ \mathrm{min}^{-1} = \bar{x}$$

El contaje es poissoniano, por lo que esperamos que

$$\sigma^2 = \mu \simeq \bar{x}$$

con la aproximación anterior

$$\sigma^2 \simeq 3224 \ \mathrm{min}^{-2}$$

$$\sigma \simeq 56.78 \ \mathrm{min}^{-1}$$

sin embargo debido a que el valor medio del contaje es tan elevado podremos hacer la aproximación de que la distribución de probabilidad es aproximadamente gaussiana, y entonces el intervalo será

$$3224 - z_{\frac{\alpha}{2}} \cdot 56.78 \leq \mu \leq 3224 + z_{\frac{\alpha}{2}} \cdot 56.78$$

$$z_{\frac{\alpha}{2}} = z_{0.005} = 2.576$$

$$3224 - 2.576 \cdot 56.78 \leq \mu \leq 3224 + 2.576 \cdot 56.78$$

$$3078 \leq \mu \leq 3370$$

3.12 Para estudiar la intención de voto al partido PDX se entrevista a un conjunto de 1000 personas. De esta muestra 256 manifiestan su intención de voto favorable

i) encontrar un intervalo de confianza para la fracción p de votantes al partido PDX con un nivel de confianza del 95%

ii) a cuántas personas deberíamos entrevistar para que la semianchura de este intervalo fuera del 5%

(igual que 3.10)

$$\frac{X}{n} - z_{\frac{\alpha}{2}}\sqrt{\frac{\frac{X}{n}\left(1-\frac{X}{n}\right)}{n}} \le p \le \frac{X}{n} + z_{\frac{\alpha}{2}}\sqrt{\frac{\frac{X}{n}\left(1-\frac{X}{n}\right)}{n}}$$

$$z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$$

$$\hat{p} = 0.256$$

$$0.256 - 1.96\sqrt{\frac{0.256\,(1-0.256)}{1000}} \le p \le 0.256 + 1.96\sqrt{\frac{0.256\,(1-0.256)}{1000}}$$

$$0.229 \le p \le 0.283$$

En este caso la semianchura del intervalo es

$$\Delta_n = 0.027$$

$$\frac{\Delta_n}{\hat{p}} \simeq 0.106$$

Si deseamos bajar la anchura al 5%, tendremos que

$$\frac{\Delta_n}{\Delta_{n'}} = \frac{0.106}{0.05} = 2.12 = \sqrt{\frac{n'}{n}}$$

$$n' = (2.12)^2 \cdot 1000 = 4494$$

3.13 Un fabricante de pinturas afirma que con un bote de un litro se pueden pintar $20 m^2$. Se realiza un muestreo aleatorio simple y se obtiene

superficie (m^2) 18, 15, 21, 16, 19, 22

Considerando que la variable aleatoria "superficie" sigue una distribución modelo normal ¿qué podemos decir de la hipótesis del fabricante a un nivel de confianza del 90%?

En este caso consideraremos como hipótesis nula

$$H_0 : \mu_0 = 20$$

y como hipótesis alternativa

$$H_a : \mu_0 \neq 20$$

Como no tenemos más información sobre la distribución de probabilidad de la variable original, el estadístico de contraste es

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

. La región de aceptación comprenderá los valores

$$A = \{ t \,/\, -t_{\alpha/2;\, n-1} \leq t \leq t_{\alpha/2;\, n-1} \}$$

siendo $\alpha/2 = 0.05$ en este caso.

. Respecto a la muestra

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{6} [18 + 15 + 21 + 16 + 19 + 22] = 18.5$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{5} [(18-18.5)^2 + (15-18.5)^2 + (21-18.5)^2 + (16-18.5)^2 + (19-18.5)^2 + (22-18.5)^2]$$

$$s^2 = 7.5 \Rightarrow s = 2.74$$

En este caso obtenemos el valor del estadístico de contraste como

$$t = \frac{18.5 - 20}{2.74/\sqrt{6}} = -1.34$$

por otra parte $t_{0.05;5} = 2.015$

Esto significa que el valor medio de la muestra no indica una discrepancia significativa con la hipótesis nula al nivel de confianza del 90%.

3.14

Considerando una población de 10 000 hogares se realiza una encuesta de audiencia de un programa de TV a lo largo de distintos días de emisión, obteniéndose

5200, 6100, 4900, 4600, 4100, 4500, 3900, 6300, 4700, 4200

suponiendo que la distribución de probabilidad de audiencia es normal, realizar un test de hipótesis con un nivel de confianza del 90% sobre la afirmación de que la audiencia media es del 50%

De nuevo la afirmación para la hipótesis nula

$H_0: \mu = 0.5 = \mu_0$

$H_a: \mu \neq 0.5$

Consideraremos el estadístico t de Student

$$t = \frac{\bar{x} - \mu_0}{S/\sqrt{n}}$$

$$A \equiv \{t;\ -t_{\alpha/2;n-1} \leq t \leq t_{\alpha/2;n-1}\}$$

la muestra arroja las proporciones de audiencia

$$\{0.52,\ 0.61,\ 0.49,\ 0.46,\ 0.41,\ 0.45,\ 0.39,\ 0.63,\ 0.47,\ 0.42\}$$

$$\bar{x} = \frac{1}{10}[0.52+0.61+0.49+0.46+0.41+0.45+0.39+0.63+0.47+0.42] =$$

$$= 0.485$$

$$s^2 = \frac{1}{9}\sum_{i=1}^{10}(x_i-\bar{x})^2 = 0.0065 \quad ; \qquad s = 0.0809$$

De nuevo el estadístico de discrepancia será la t-Student

$$t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$$

Con la región de aceptación bivariante

$$A \equiv \{t \,/\, -t_{\alpha/2;n-1} \le t \le t_{\alpha/2;n-1}\}$$

El valor de la variable de discrepancia es

$$t = \frac{0.485-0.5}{0.081/\sqrt{10}} = -0.586$$

observemos que $t_{\alpha/2;n-1} = t_{0.05;9} = 1.83$

Por lo que aceptamos la hipótesis nula.

3.15 La dosis media que se desea impartir a un paciente en un tratamiento de radioterapia es de 2.00 Gy se realizan 10 medidas (en Gy) obteniéndose

1.98, 1.93, 1.98, 1.89, 1.92, 1.95, 1.94, 1.96, 1.93, 1.92

Establecer si al 99% de nivel de confianza la dosis es distinta de la deseada supondremos que podemos considerar que las variaciones en la dosis medida son el resultado de múltiples efectos pequeños debidos a diferentes magnitudes de influencia, es decir consideraremos el comportamiento gaussiano. De aquí nuestro estadístico será

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

con la hipótesis nula

$$H_0: \mu = 2.00 = \mu_0$$

$$H_a: \mu \neq 2.00$$

La región de aceptación

$$A = \{ t / -t_{\alpha/2; n-1} \leq t \leq t_{\alpha/2; n-1} \}$$

y la región crítica

$$C = \{ t / t \leq -t_{\alpha/2; n-1} \text{ ó } t_{\frac{\alpha}{2}; n-1} < t \}$$

Si consideramos la muestra de 10 dosis obtenidas

$$\bar{x} = 1{,}94 \qquad s = 0.028$$

$$t = \frac{1.94 - 2.00}{0.028/\sqrt{10}} = -6.776$$

$$t_{\frac{\alpha}{2}; n-1} = t_{0.005; 9} = 3.25$$

Por lo que observamos que el valor de t excede los límites de la región de aceptación. Por lo tanto rechazamos la hipótesis al 99% afirmando que (con este nivel de confianza) la dosis es diferente (menor) del valor prescrito.

3.16

Considérese que en un experimento de Bernoulli se lanza una moneda y se obtienen 10 caras y 20 cruces. Si se le asigna a este suceso una probabilidad binomial, úsese el método de máxima verosimilitud para obtener el valor de p (probabilidad de cara).

$$\text{prob}(x=r) = \binom{n}{r} p^r q^{n-r}$$

observemos que $n=30$ en este caso (el nº total de lanzamientos) y $q=1-p$. Considerando la probabilidad binomial, la probabilidad "a posteriori" del suceso es

$$\text{Prob} \equiv f(p) = \binom{30}{10} p^{10} (1-p)^{20}$$

Podemos considerar esta nuestra función de verosimilitud para p

$$\mathcal{L}(p) = \binom{30}{10} p^{10} \cdot (1-p)^{20}$$

Por ello, será máximamente verosímil el valor de p que verifique

$$\frac{d\mathcal{L}}{dp} = 0 = \frac{d}{dp}\left[\binom{30}{10} p^{10} (1-p)^{20}\right]$$

$$\frac{d\mathcal{L}}{dp} = \binom{30}{10}\left[10\, p^9 (1-p)^{20} + 20 \cdot p^{10} (1-p)^{19} (-1)\right] = 0$$

$$10\, p^9 (1-p)^{20} - 20\, p^{10} (1-p)^{19} = 0$$

$$10\,(1-p) - 20\,p = 0$$

$$10 = 30\,p \Rightarrow \boxed{p = 10/30}$$

Es importante observar que la estimación máximamente verosímil en este caso coincide con la asignación elemental del cociente de sucesos favorables sobre sucesos totales.

3.17

En un experimento para determinar la constante elástica de un resorte se han suspendido diferentes masas en un punto de la Tierra produciendo diferentes elongaciones (N.B. el experimento relaciona en este caso las masas inerciales o pesantes con la elongación en este lugar sin establecer explícitamente el valor de la aceleración de la gravedad). Las elongaciones se miden con un metro con $u_L = 0.002$ cm

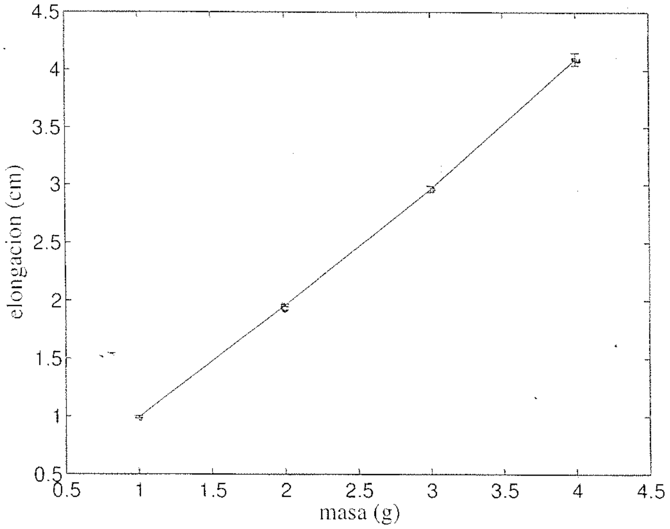
m (g)	l_1 (cm)	l_2 (cm)	l_3 (cm)	$\bar{l}$ (cm)	$u^A_{\bar{l}}$ (cm)	$u_{\bar{l}}$ (cm)
1.0	0.993	1.011	0.979	0.9943	0.0093	0.0095
2.0	1.954	1.944	1.978	1.9587	0.0101	0.0103
3.0	2.918	2.949	3.011	2.9593	0.0273	0.0274
4.0	4.143	4.157	3.987	4.0957	0.0545	0.0545

i) Para establecer la elongación para cada masa consideraremos

$$\bar{l} = \frac{1}{3}\sum_{i=1}^{3} l_i \qquad s^2(\bar{l}) = \frac{1}{n(n-1)}\sum_{i=1}^{3}(l_i-\bar{l})^2 \equiv u_A(\bar{l})$$

$$u(\bar{l}) = \sqrt{u_A^2(\bar{l}) + u_L^2}$$

u_L es una incertidumbre de tipo B



Obsérvese que las barras de incertidumbre son considerablemente pequeñas en el gráfico aunque desiguales.

ii) Consideraremos un ajuste lineal ponderado de modo que debemos minimizar

$$\chi^2(a,b) = \sum_{i=1}^{4} \left[\frac{y_i - a - b x_i}{\sigma_i}\right]^2$$

Consideraremos $\sigma_i \equiv u_i$ calculada para cada valor medio de elongación, esto es

$$\chi^2(a,b) = \frac{(0.9943 - a - b\cdot 1.0)^2}{(0.0095)^2} + \frac{(1.9587 - a - b\cdot 2.0)^2}{(0.0103)^2} +$$

$$+ \frac{(2.9593 - a - b\cdot 3.0)^2}{(0.0274)^2} + \frac{(4.0957 - a - b\cdot 4.0)^2}{(0.0545)^2}$$

Considerando las ecuaciones de minimización

$$\frac{\partial \chi^2}{\partial a} = 0; \quad \frac{\partial \chi^2}{\partial b} = 0;$$

$$\left.\begin{aligned} a\sum_{i=1}^{4}\frac{1}{\sigma_i^2} + b\sum_{i=1}^{4}\frac{x_i}{\sigma_i^2} &= \sum_{i=1}^{4}\frac{y_i}{\sigma_i^2} \\ a\sum_{i=1}^{4}\frac{x_i}{\sigma_i^2} + b\sum_{i=1}^{4}\frac{x_i^2}{\sigma_i^2} &= \sum_{i=1}^{4}\frac{x_i y_i}{\sigma_i^2} \end{aligned}\right\}$$

de donde $\Delta = \sum_{i=1}^{4}\frac{1}{\sigma_i^2}\sum_{i=1}^{4}\frac{x_i^2}{\sigma_i^2} - \left(\sum_{i=1}^{4}\frac{x_i}{\sigma_i^2}\right)^2$

$$a = \frac{1}{\Delta}\left[\sum_{i=1}^{4}\frac{y_i}{\sigma_i^2}\cdot\sum_{i=1}^{4}\frac{x_i^2}{\sigma_i^2} - \sum_{i=1}^{4}\frac{x_i y_i}{\sigma_i^2}\cdot\sum_{i=1}^{4}\frac{x_i}{\sigma_i^2}\right]$$

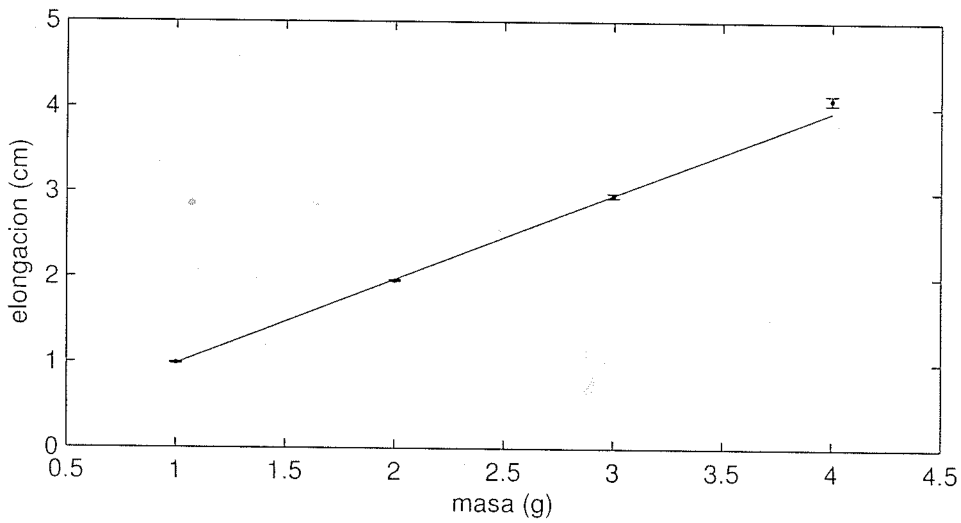
$$b = \frac{1}{\Delta}\left[\sum_{i=1}^{4}\frac{1}{\sigma_i^2}\cdot\sum_{i=1}^{4}\frac{x_i y_i}{\sigma_i^2} - \sum_{i=1}^{4}\frac{x_i}{\sigma_i^2}\sum_{i=1}^{4}\frac{y_i}{\sigma_i^2}\right]$$

$$\sigma_a^2 = \frac{1}{\Delta}\sum_{i=1}^{4}\frac{x_i^2}{\sigma_i^2} \;;\quad \sigma_b^2 = \frac{1}{\Delta}\sum_{i=1}^{4}\frac{1}{\sigma_i^2} \;;\quad \mathrm{cov}(a,b) = -\frac{1}{\Delta}\sum_{i=1}^{4}\frac{x_i}{\sigma_i^2}$$

De aquí obtenemos

$a = -0.002$ cm $\sigma_a = 0.017$ cm

$b = 0.988$ cm/g $\sigma_b = 0.010$ cm/g



ii) si debemos realizar un test del ajuste consideraremos un test de hipótesis de χ^2. De la forma

$$\chi^2 = \sum_{i=1}^{4} \frac{(y_i - \tilde{y}_i)^2}{\sigma_i^2} = \qquad \tilde{y}_i \equiv \text{ordenadas de la recta}$$

$$= \sum_{i=1}^{4} \frac{(y_i - a - bx_i)^2}{\sigma_i^2}$$

observemos que el nº de grados de libertad remanentes es $4-2=2$ ya que hemos usado la información experimental para ajustar a, b. De forma que

$$\chi^2 = 10.14$$

si buscamos ahora el percentil de la χ^2,

$$\chi^2_{\alpha;2} \begin{cases} \chi^2_{0.05;2} = 5.99 \\ \chi^2_{0.01;2} = 9.21 \end{cases}$$

En ambos casos los resultados parecen indicar cierta discrepancia con un modelo lineal. Rechazaríamos el ajuste considerando válidos los valores de las incertidumbres estimadas.

3.18 En un experimento se mide la densidad de una disolución a 298.15 K frente a la concentración de soluto (moles/l)

c $(\text{mol}\cdot l^{-1})$	0.000	0.005	0.010	0.015	0.020
ρ $(\text{kg}\cdot m^{-3})$	990.0	992.2	994.4	996.6	998.7

considerando que la incertidumbre relativa del densímetro es del 1% de la lectura, realizar un ajuste ponderado

$$\rho(c) = a + b\cdot c$$

Por tanto la tabla con las incertidumbres será

c (mol l^{-1})	0.000	0.005	0.010	0.015	0.020
ρ (kg m^{-3})	990.0	992.2	994.4	996.6	998.7
u_ρ (kg m^{-3})	9.90 (9.9)	9.92 (9.9)	9.94 (9.9)	9.97 (1.0)	9.99 (1.0)

hemos indicado excepcionalmente la incertidumbre en la densidad con 3 cifras significativas ya que en caso de usar dos todas las cifras serían casi idénticas.
El ajuste ponderado debe arrojar resultados casi idénticos al ajuste sin ponderar

$$a \sum_{i=1}^{5} \frac{1}{\sigma_i^2} + b \sum_{i=1}^{5} \frac{x_i}{\sigma_i^2} = \sum_{i=1}^{5} \frac{y_i}{\sigma_i^2}$$

$$a \sum_{i=1}^{5} \frac{x_i}{\sigma_i^2} + b \sum_{i=1}^{5} \frac{x_i^2}{\sigma_i^2} = \sum_{i=1}^{5} \frac{x_i y_i}{\sigma_i^2}$$

$$\Delta = \sum_{i=1}^{5} \frac{1}{\sigma_i^2} \sum_{i=1}^{5} \frac{x_i^2}{\sigma_i^2} - \left(\sum_{i=1}^{5} \frac{x_i}{\sigma_i^2} \right)^2$$

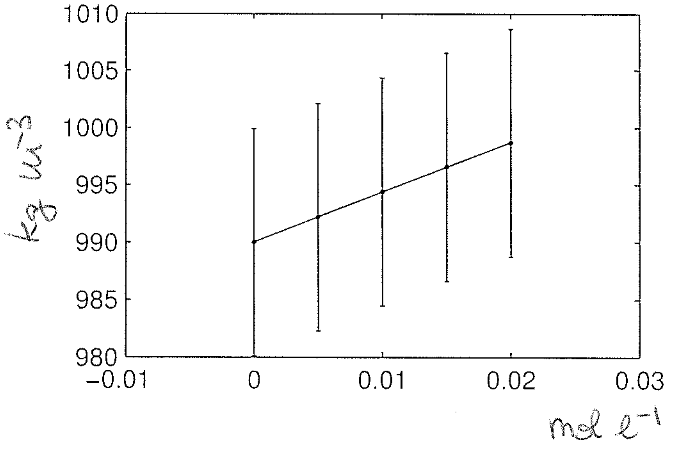
$$a = \frac{1}{\Delta} \left[\sum_{i=1}^{5} \frac{y_i}{\sigma_i^2} \cdot \sum_{i=1}^{5} \frac{x_i^2}{\sigma_i^2} - \sum_{i=1}^{5} \frac{x_i y_i}{\sigma_i^2} \cdot \sum_{i=1}^{5} \frac{x_i}{\sigma_i^2} \right]$$

$$b = \frac{1}{\Delta} \left[\sum_{i=1}^{5} \frac{1}{\sigma_i^2} \sum_{i=1}^{5} \frac{x_i y_i}{\sigma_i^2} - \sum_{i=1}^{5} \frac{x_i}{\sigma_i^2} \sum_{i=1}^{5} \frac{y_i}{\sigma_i^2} \right]$$

$$\sigma_a^2 = \frac{1}{\Delta} \sum_{i=1}^{5} \frac{x_i^2}{\sigma_i^2} \quad ; \quad \sigma_b^2 = \frac{1}{\Delta} \sum_{i=1}^{5} \frac{1}{\sigma_i^2}$$

$a = 990.0$ kg m^{-3} $\qquad \sigma_a = 7.7$ kg m^{-3}

$b = 436$ kg m^{-3} (mol l^{-1})$^{-1}$ $\qquad \sigma_b = 630$ kg m^{-3} (mol l^{-1})



observese que las barras de incertidumbre del 1% indicadas parecen excesivas a la hora de analizar los datos

- Si realizamos un ajuste sin pesos $\sigma_i = \sigma \ \forall i$ los valores de los parámetros obtenidos por mínimos cuadrados serán:

$$\tilde{a} = 990 \ \text{kg m}^{-3}$$

$$\tilde{b} = 436 \ \text{kg m}^{-3} \ (\text{mol l}^{-1})^{-1}$$

lo cual no difiere significativamente de lo obtenido anteriormente. Si usamos el estimador para la varianza de los datos del propio ajuste

$$s^2 = \frac{1}{n-2} \sum_{i=1}^{5} \left[y_i - \tilde{a} - \tilde{b} \cdot x_i\right]^2 = 0.0013 \ (\text{kg m}^{-3})^2$$

$$s \simeq 0.04 \ \text{kg m}^{-3}$$

que dista significativamente de lo anunciado en el problema.

- Respecto al coeficiente de correlación de dos variables aleatorias, recordemos que su expresión es

$$r(x,y) = \frac{\operatorname{cov}(x,y)}{s_x \cdot s_y}$$

$$cov(x,y) = \frac{1}{n-1}\sum_{i=1}^{n} (x_i-\bar{x})(y_i-\bar{y}) = 0.0273 \ (kg\ m^{-3})(mol\ l^{-1})$$

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i-\bar{x})^2 = 6.25\cdot 10^{-5} \ (mol\ l^{-1})^2 \qquad s_x = 0.0079\ mol\ l^{-1}$$

$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n} (y_i-\bar{y})^2 = 11.88 \ (kg\ m^{-3})^2 \qquad s_y = 3.45\ kg\ m^{-3}$$

por lo tanto $$r(x,y) = \frac{cov(x,y)}{s_x\ s_y} = 1.0$$

- lo que nos indica una excelente correlación de los datos.

- considerando la velocidad del sonido

$$v = (1500 \pm 100)\ m\cdot s^{-1}$$

obtendremos Ks con su incertidumbre, considerando

$$k_s = 1/\rho v^2$$

$$\left(\frac{\partial k_s}{\partial \rho}\right) = -1/\rho^2 v^2 \qquad \left(\frac{\partial k_s}{\partial v}\right) = -2\ 1/\rho v^3$$

$$u_{ks}^2 = \left(\frac{\partial k_s}{\partial \rho}\right)^2 u_\rho^2 + \left(\frac{\partial k_s}{\partial v}\right)^2 u_v^2 =$$

$$= \left(\frac{1}{\rho^2 v^2}\right)^2 u_\rho^2 + 4\left(\frac{1}{\rho v^3}\right)^2 u_v^2 = k_s^2\left[\left(\frac{u_\rho}{\rho}\right)^2 + 4\left(\frac{u_v}{v}\right)^2\right]$$

k_s ($m\ s^2\ kg^{-1}$)	$4.49\cdot 10^{-10}$	$4.48\cdot 10^{-10}$	$4.47\ 10^{-10}$	$4.46\cdot 10^{-10}$	$4.45\cdot 10^{-10}$
u_{ks} ($m\ s^2\ kg^{-1}$)	$0.60\cdot 10^{-10}$	$0.60\cdot 10^{-10}$	$0.60\cdot 10^{-10}$	$0.60\cdot 10^{-10}$	$0.59\cdot 10^{-10}$

3.19 Se quiere ajustar por el método de mínimos cuadrados un conjunto de datos

$$\{(x_1,y_1),(x_2,y_2)\ldots,(x_n,y_n)\}$$

a una recta $y=a$

¿Cuál es el valor de la constante a?

¿Cuál es la incertidumbre de a si cada dato y_i tiene una incertidumbre σ_i?

Para ajustar a una constante consideraremos la minimización de la función

$$\chi^2(a)=\sum_{i=1}^{n}\frac{(y_i-a)^2}{\sigma_i^2}$$

- La condición de mínimo es

$$\frac{d\chi^2}{da}=0 \Rightarrow -2\sum_{i=1}^{n}\frac{(y_i-a)}{\sigma_i^2}=0$$

$$a \sum_{i=1}^{n}\frac{1}{\sigma_i^2}=\sum_{i=1}^{n} y_i/\sigma_i^2$$

$$a=\sum_{i=1}^{n} y_i/\sigma_i^2 \Big/ \sum_{i=1}^{n} 1/\sigma_i^2$$

- La incertidumbre de a puede ser obtenida directamente del sistema, la matriz del sistema es

$$H=\left(\sum_{i=1}^{n} 1/\sigma_i^2\right)$$

$$\sigma_a^2 = H^{-1} = 1\Big/\sum_{i=1}^{n} 1/\sigma_i^2$$

3.20 Comparar las fórmulas del ajuste a una recta $y = a + bx$ para las incertidumbres de los parámetros a, b con las halladas en (3.146) y (3.148)

Recordemos que en 3.146 y 3.148 se obtuvo

$$\sigma_a^2 = \frac{\sum_{i=1}^{n} x_i^2/\sigma_i^2}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2} \sum_{i=1}^{n} \frac{x_i^2}{\sigma_i^2} - \left(\sum_{i=1}^{n} \frac{x_i}{\sigma_i^2}\right)^2}$$

$$\sigma_b^2 = \frac{\sum_{i=1}^{n} 1/\sigma_i^2}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2} \sum_{i=1}^{n} \frac{x_i^2}{\sigma_i^2} - \left(\sum_{i=1}^{n} \frac{x_i}{\sigma_i^2}\right)^2}$$

En el apartado 3.185 relacionamos las incertidumbres de los parámetros con la inversa de la matriz del sistema lineal

$$\underbrace{\begin{bmatrix} \sum_{i=1}^{n} 1/\sigma_i^2 & \sum_{i=1}^{n} x_i/\sigma_i^2 \\ \sum_{i=1}^{n} x_i/\sigma_i^2 & \sum_{i=1}^{n} x_i^2/\sigma_i^2 \end{bmatrix}}_{H} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i/\sigma_i^2 \\ \sum_{i=1}^{n} y_i x_i/\sigma_i^2 \end{bmatrix}$$

$$H^{-1} = \begin{pmatrix} \sigma_a^2 & \mathrm{Cov}(a,b) \\ \mathrm{Cov}(a,b) & \sigma_b^2 \end{pmatrix}$$

Ahora bien la inversa de la matriz H es

$$H^{-1} = \frac{1}{\det(H)} \begin{bmatrix} \sum_{i=1}^{n} x_i^2/\sigma_i^2 & -\sum_{i=1}^{n} x_i/\sigma_i^2 \\ -\sum_{i=1}^{n} x_i/\sigma_i^2 & \sum_{i=1}^{n} 1/\sigma_i^2 \end{bmatrix}$$

siendo $\det(H) = \sum_{i=1}^{n} 1/\sigma_i^2 \sum_{i=1}^{n} x_i^2/\sigma_i^2 - \left(\sum_{i=1}^{n} x_i/\sigma_i^2\right)^2$

Por comparación directa podemos comprobar la correspondencia de las varianzas.