PARTE III

INFERÊNCIA ESTATÍSTICA

Introdução

A inferência estatística consiste num conjunto de métodos que servem para tomar decisões ou tirar conclusões (inferir) sobre uma população, com base numa amostra dessa população.

$$\textbf{Inferência Estat.}\begin{cases}\textbf{Estimação de Parâmetros}\begin{cases}\text{Estimação Pontual}\\ \text{Estimação Intervalar}\end{cases}\\ \\ \textbf{Testes de Hipóteses}\begin{cases}\text{Testes Paramétricos}\\ \text{Outros Testes}\end{cases}\end{cases}$$

Estimação de parâmetros

Suponhamos que temos, então uma amostra de dados e conhecemos a distribuição da qual os dados provêm, a menos de um ou mais parâmetros. Por exemplo:

- sabemos que os dados provêm de uma v.a. com distribuição normal $N(\mu, \sigma)$, mas que não conhecemos μ, ou desconhecemos mesmo μ e σ
- sabemos que os dados vêm de uma distribuição $Poi(\lambda)$, mas o valor médio λ é desconhecido.

Problemas em que o tipo da distribuição subjacente aos dados é especificado a menos de um conjunto de parâmetros desconhecidos e em que pretendemos *estimar* esses parâmetros são chamados problemas de estimação paramétrica ou de parâmetros.

Estatística

Suponhamos então que temos um conjunto de dados $x_1, \ldots, x_n$ obtidos aleatoriamente de uma população com uma lei de probabilidade conhecida (a menos de determinados parâmetros) e que pretendemos estimar um desses parâmetros, θ.

Problema

Como usar os dados para estimar o valor do parâmetro desconhecido θ?

Podemos sempre interpretar essa amostra $x_1, \ldots, x_n$ como uma realização de uma amostra aleatória $X_1, \ldots, X_n$ de X (X_i v.a.'s i.i.d. com X).

Definição

No âmbito da inferência estatística chamamos estatística a qualquer função das v.a.'s que constituem a amostra aleatória teórica $X_1, \ldots, X_n$, desde que tal função não inclua parâmetros desconhecidos.

Estimador e estimativa

Definição

- Qualquer estatística T, isto é, qualquer função $T = T(X_1, \ldots, X_n)$ da amostra aleatória que não envolva parâmetros desconhecidos usada para estimar um parâmetro θ, é chamada um estimador (pontual) de θ.
- O valor observado de T na amostra concreta $x_1, \ldots, x_n$, i.e. $\tilde{\theta} = T(x_1, \ldots, x_n)$ é chamado uma estimativa (pontual) de θ.

Estimador e estimativa

Assim:

- um estimador é uma v.a. (com uma certa distribuição)
- uma estimativa é um número (que, com uma "boa" escolha do estimador, deverá ser uma "aproximação razoável" para o valor do parâmetro que estamos a estimar).

Exemplo

Um estimador normalmente utilizado para estimar a média μ de uma distribuição normal $N(\mu, \sigma)$ é a v.a. média amostral

$$\overline{X} = \frac{X_1 + \cdots + X_n}{n}$$

e um valor

$$\overline{x} = \frac{x_1 + \cdots + x_n}{n}$$

obtido a partir da amostra observada $x_1, \ldots, x_n$ é uma estimativa de μ (correspondente ao estimador $\overline{X}$).

Classificação dos estimadores

Há vários critérios para classificar os estimadores.

Definição

Seja $T = T_n = T(X_1, \ldots, X_n)$ um estimador de um dado parâmetro θ. Dizemos que T_n é um estimador:

- centrado ou não enviesado de θ se verificar

$$E(T_n) = \theta.$$

- consistente se, para todo o $\epsilon > 0$,

$$\lim_{n\to\infty} P(\theta - \epsilon < T_n < \theta + \epsilon) = 1.$$

Dados dois estimadores centrados, diremos que é mais eficiente aquele que tiver menor variância, devendo esse ser o utilizado.

Exemplo

Seja $X_1, \ldots, X_n$ uma amostra aleatória do modelo $N(\mu, \sigma)$ e considere-se a v.a. média amostral $\overline{X}$ como estimador para μ. Como sabemos, tem-se

$$E(\overline{X}) = \mu,$$

ou seja, $\overline{X}$ é um estimador centrado para μ; além disso, pela LGN, temos, para qualquer $\epsilon > 0$

$$\lim_{n\to\infty} P(\mu - \epsilon < \overline{X} < \mu + \epsilon) = 1,$$

pelo que $\overline{X}$ é também um estimador consistente para μ.

Nota Pode provar-se que, neste caso, este é o "melhor" estimador centrado que existe para μ, no sentido em que não há outro estimador centrado para μ que tenha menor variância do que $\overline{X}$.

Estimadores para o valor médio e para a variância

Mais geralmente, têm-se os resultados contidos na seguinte proposição.

Proposição

Seja $X_1, \ldots, X_n$ uma amostra aleatória de dimensão n de uma v.a. X tal que $E(X) = \mu$ e $\mathrm{var}(X) = \sigma^2$ (finitos). Então:

- *A média amostral $\overline{X} = \dfrac{X_1 + \ldots + X_n}{n}$ é um estimador centrado e consistente para μ.*
- *A variância amostral $S^2 = \dfrac{1}{n-1} \sum_{k=1}^{n} (X_k - \overline{X})^2$ é um estimador centrado e consistente para σ^2.*

Estimativas de máxima verosimilhança

Existem vários métodos de obter estimadores, sendo um deles, o chamado método da máxima verosimilhança, o qual permite, em geral, obter estimadores com boas propriedades.

Vamos dar uma ideia deste método, no caso em que X é uma v.a. discreta. Suponhamos então que X é uma v.a. discreta e seja $x_1, \ldots, x_n$ uma amostra aleatória de dados proveniente da lei de X, em que θ é um parâmetro desconhecido. Para cada valor de θ, podemos calcular a probabilidade P_θ de obter exatamente a amostra $x_1, \ldots, x_n$. A um valor $\hat{\theta}$ que maximiza a probabilidade P_θ chamamos estimativa de máxima verosimilhança de θ. Na tabela seguinte indicamos as estimativas de máxima verosimilhança (m.v.) dos parâmetros de alguns modelos mais usuais.

Modelo	Parâmetro(s)	Estimativa(s) m.v.
$Bi(m,p)$	p	$\hat{p} = \frac{\overline{x}}{m}$
$Poi(\lambda)$	λ	$\hat{\lambda} = \overline{x}$
$U[a,b]$	$a,\ b$	$\hat{a} = \min\{x_k\},\ \hat{b} = \max\{x_k\}$
$N(\mu,\sigma)$	$\mu,\ \sigma$	$\hat{\mu} = \overline{x},\ \hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$
$Exp(\lambda)$	λ	$\hat{\lambda} = 1/\overline{x}$

Estimadores de m.v.

A cada estimativa de máxima verosimilhança corresponde um estimador de máxima verosimilhança. Por exemplo, sendo a estimativa de m.v. da média μ, no modelo normal $N(\mu, \sigma)$, dada pela média amostral $\overline{\mathbf{x}} = \dfrac{x_1 + \cdots + x_n}{n}$, o correspondente estimador de m.v. será a v.a. média amostral $\overline{X} = \dfrac{X_1 + \cdots + X_n}{n}$.

Nota: O estimador de máxima verosimilhança para a variância é dado por

$$\frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X})^2$$

o qual não coincide com a variância amostral

$$S^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X})^2.$$

A razão de preferirmos usar a v.a. S^2 (i.e. usarmos $n-1$ no denominador, em vez de n) como estimador para σ^2, tem a ver como o facto de este estimador ser centrado, o que não se passa com o estimador de máxima verosimilhança.

Estimação intervalar

Como acabámos de ver, no caso da estimação pontual para um certo parâmetro θ, é escolhido um estimador, i.e. uma estatística T com "boas propriedades" e calcula-se uma estimativa para θ com base no valor observado dessa estatística.

Na chamada estimação intervalar, a ideia é encontrar duas estatísticas T_1 e T_2 para as quais se tenha $T_1 \leq \theta \leq T_2$ com uma determinada probabilidade (grande, e fixada *a priori*). Mais precisamente, escolhido um determinado valor de α (pequeno, tipicamente $\alpha = 0.05, 0.01, 0.005$), encontram-se duas estatísticas T_1 e T_2 para as quais seja

$$P(T_1 \leq \theta \leq T_2) = 1 - \alpha.$$

Nesse caso, dizemos que o intervalo (aleatório) cujos extremos são essas estatísticas T_1 e T_2, i.e. o intervalo $[T_1, T_2]$, é um intervalo de $(1-\alpha)\times 100\%$ confiança (ou de nível de confiança $1-\alpha$) ou um intervalo com uma margem de erro de $\alpha \times 100\%$ (ou α) para θ.

Intervalo de confiança

Assim por exemplo, fixada uma probabilidade $1-\alpha = 0.95$, falaremos num intervalo de 95% de confiança para o parâmetro que estamos a estimar ou diremos que esse intervalo tem uma margem de erro de 5%.

Quando consideramos valores observados das estatísticas que definem os extremos dos intervalos aleatórios, encontramos então verdadeiros intervalos da reta real, os quais são estimativas intervalares para o parâmetro em causa.

A estes intervalos chamamos também intervalos de confiança (neste caso, deterministas, ou seja, não aleatórios) para esse parâmetro.

Intervalo de confiança para μ (população normal $N(\mu,\sigma)$, σ conhecido)

Problema

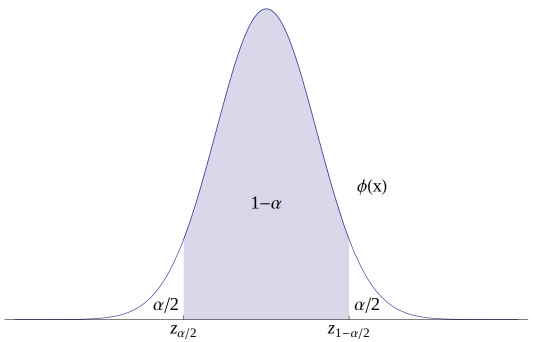
- Dada uma amostra aleatória $x_1, \ldots, x_n$ de uma v.a. com distribuição normal $N(\mu,\sigma)$, com μ desconhecido e σ conhecido,
- obter um intervalo de confiança para μ.

Vendo $x_1, \ldots, x_n$ como uma concretização de uma amostra aleatória $X_1, \ldots, X_n$, sabemos que

$$Z = \frac{\overline{X}-\mu}{\sigma}\sqrt{n} \frown N(0,1).$$

Temos

$$Z = \frac{\overline{X}-\mu}{\sigma}\sqrt{n} \frown N(0,1)$$



Assim, se $c = z_{1-\alpha/2}$ for o quantil de probabilidade $p = 1-\alpha/2$ da distribuição $N(0,1)$ (sendo, portanto, $-c$ o quantil de probabilidade $\alpha/2$), temos que

$$P\Big(-c \le \underbrace{\frac{\overline{X}-\mu}{\sigma}\sqrt{n}}_{Z} \le c\Big) = 1-\alpha.$$

De

$$P\Big(-c \le \frac{\overline{X}-\mu}{\sigma}\sqrt{n} \le c\Big) = 1-\alpha,$$

vem, explicitando μ,

$$P\Big(\overline{X} - c\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + c\frac{\sigma}{\sqrt{n}}\Big) = 1-\alpha.$$

Conclusão:

O intervalo aleatório

$$\Big[\overline{X} - c\frac{\sigma}{\sqrt{n}} \; , \; \overline{X} + c\frac{\sigma}{\sqrt{n}}\Big]$$

onde $c = z_{1-\alpha/2}$ é um intervalo de $(1-\alpha)\times 100\%$ de confiança para μ.

Neste caso, as duas estatísticas T_1 e T_2 de que falámos anteriormente são

$$\begin{aligned} T_1 &= T_1(X_1,\ldots,X_n) = \overline{X} - c\frac{\sigma}{\sqrt{n}}, \\ T_2 &= T_2(X_1,\ldots,X_n) = \overline{X} + c\frac{\sigma}{\sqrt{n}}, \quad c = z_{1-\alpha/2} \end{aligned},$$

A amplitude dos intervalos de confiança (dada por $2z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$):

- diminui quando o tamanho da amostra (n) aumenta
- aumenta quando a margem de erro (α) diminui (ou seja, quando a confiança aumenta): quando α diminui, o quantil $z_{1-\alpha/2}$ aumenta.

Exemplo

Suponhamos que a variável aleatória que representa o tempo de vida (em horas) de uma determinada bactéria tem distribuição normal $N(\mu,\sigma)$ com $\sigma = 0.18$. Foi estudada uma amostra aleatória de 8 bactérias desse tipo e o tempo médio de vida dessas bactérias foi de 1.63.
Determinemos um intervalo de 95% de confiança para o parâmetro μ dessa distribuição.
Temos $\overline{x} = 1.63$, $\sigma = 0.18$, $n = 8$ e $\alpha = 0.05$. Se determinarmos o quantil de probabilidade $p = 1 - \alpha/2 = 0.975$ da distribuição $N(0,1)$, tem-se que $c = z_{0.975} = 1.96$. Então, vem

$$c\frac{\sigma}{\sqrt{n}} = 1.96\frac{0.18}{\sqrt{8}} = 0.12,$$

pelo que um intervalo de 95% de confiança para μ é

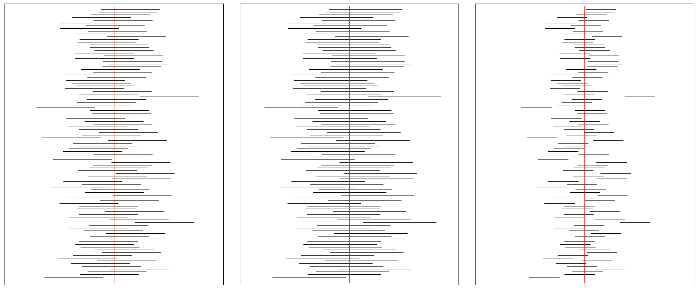
$$[1.63 - 0.12, 1.63 + 0.12] = [1.51, 1.75] .$$

Significado do intervalo de confiança

- *Confiança* não significa o mesmo que *probabilidade*.
  Ao dizermos que $[1.51, 1.75]$ é um intervalo de 95% de confiança para μ, **não** estamos a afirmar que a probabilidade de μ pertencer ao intervalo $[1.51, 1.75]$ é de 0.95.
- Esse intervalo $[1.51, 1.75]$ ou contém ou não contém μ (o qual é um valor fixo, embora desconhecido); dizer que $\mu \in [1.51, 1.75]$ é ou (100%) verdadeiro ou (100%) falso e por isso falar em probabilidade não faz qualquer sentido.
- Quando dizemos que esse intervalo é um intervalo de 95% de confiança para μ, isto deve ser interpretado no seguinte sentido:

O intervalo foi obtido por um processo que, em aproximadamente 95% dos casos, fornece intervalos que contêm μ.

IC's com diferentes graus de confiança



$1-\alpha = 0.9$ $\quad$ $1-\alpha = 0.95$ $\quad$ $1-\alpha = 0.8$

Amostra de modelo não normal (n grande)

O intervalo de confiança anteriormente obtido para o valor médio, supondo a variância conhecida, pode também, em consequência do Teorema Limite Central, ser utilizado para um modelo não normal, no caso de termos uma amostra de grande dimensão

Mais precisamente, neste caso apenas poderemos afirmar que, para n grande, se tem

$$P\left(\overline{X} - c\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + c\frac{\sigma}{\sqrt{n}}\right) \approx 1-\alpha, \quad c = z_{1-\alpha/2}.$$

Assim sendo, dizemos que o intervalo

$$\left[\overline{X} - c\frac{\sigma}{\sqrt{n}}, \overline{X} + c\frac{\sigma}{\sqrt{n}}\right], \ c = z_{1-\alpha/2}$$

é um intervalo de aproximadamente $(1-\alpha)\times 100\%$ de confiança para a média μ, ou que é um intervalo de $(1-\alpha)\times 100\%$ de confiança assintótico para μ.

Intervalo de confiança para p (população $Ber(p)$)

Como exemplo do que acabámos de referir, consideremos o seguinte:

Problema

- Dada uma amostra aleatória $x_1, \ldots, x_n$ proveniente da distribuição $Ber(p)$, com p desconhecido,
- obter um intervalo de confiança para p.

Neste caso, $\mu = p$ e como $\text{var}(X) = p(1-p)$, obtemos o intervalo assintótico

$$\left[\overline{X} - c\frac{\sqrt{p(1-p)}}{\sqrt{n}}, \overline{X} + c\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right], \ c = z_{1-\alpha/2},$$

onde $\overline{X}$ corresponde à proporção de sucessos em n experiências. Sendo n suficientemente grande, podemos (pela Lei dos Grandes Números) substituir, nos extremos do intervalo, p por $\overline{X}$ e obter o seguinte intervalo aleatório de aproximadamente $(1-\alpha)\times 100\%$ confiança para p:

$$\left[\overline{X} - c\frac{\sqrt{\overline{X}(1-\overline{X})}}{\sqrt{n}}, \overline{X} + c\frac{\sqrt{\overline{X}(1-\overline{X})}}{\sqrt{n}}\right], \ c = z_{1-\alpha/2},$$

Exemplo

Na véspera de uma eleição, foi efectuado um inquérito para saber das possibilidades de um determinado candidato. Foram inquiridos 1 002 eleitores, tendo-se 701 mostrado favoráveis à eleição do político em causa. Determinar um intervalo de (aproximadamente) 90% de confiança para a proporção de votantes no candidato.

Neste caso, $\overline{x} = \frac{701}{1\,002} = 0.6996$, $n = 1\,002$ e $c = z_{1-0.05} = z_{0.95} = 1.6449$; temos, então

$$\overline{x} - c\frac{\sqrt{\overline{x}(1-\overline{x})}}{\sqrt{n}} = 0.6754 \quad \text{e} \quad \overline{x} + c\frac{\sqrt{\overline{x}(1-\overline{x})}}{\sqrt{n}} = 0.7238$$

pelo que o intervalo de confiança será:

$$[0.6754, 0.7238].$$

A distribuição qui-quadrado

Antes de deduzirmos outros intervalos de confiança, torna-se necessário falar de outras distribuições contínuas, muito importantes: as distribuições qui-quadrado, t de Student e F de Fisher, as quais estão relacionadas com a distribuição normal.

Definição (Distribuição qui-quadrado)

Consideremos uma amostra aleatória $Z_1, Z_2, \ldots, Z_n$ da v.a. $Z \frown N(0,1)$, i.e. sejam Z_i v.a.'s i.i.d. tais que $Z_i \frown N(0,1)$. Chama-se distribuição qui-quadrado com n graus de liberdade à distribuição da variável aleatória

$$Q = Q_n = Z_1^2 + Z_2^2 + \cdots + Z_n^2.$$

Se Q tem uma distribuição qui-quadrado com n graus de liberdade, escrevemos $Q \frown \chi^2_n$.

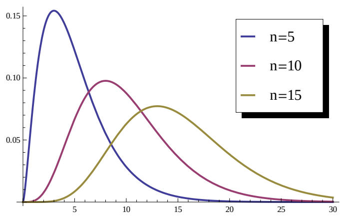
Distribuição χ^2_n

Se $Q \frown \chi^2_n$, então a sua função densidade de probabilidade f_Q é dada por

$$f_Q(x) = \begin{cases} 0, & \text{para } x \leq 0, \\ \frac{1}{2^{n/2}\Gamma(n/2)} e^{-x^2/2} x^{n/2-1}, & \text{para } x > 0, \end{cases}$$

onde Γ designa a chamada função gama.

Nota A função gama pode ser vista como uma extensão da função factorial; tem-se $\Gamma(n) = (n-1)!$, se $n \in \mathbb{N}$, e $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$, para $\alpha > 0$.



Função densidade de probabilidade da distribuição χ^2_n; $n = 5, 10, 15$.

Características teóricas da distribuição χ^2_n

Se $X \frown \chi^2_n$, tem-se

- $\mu_X = n$
- $\sigma^2_X = 2n$
- $\max\{n-2, 0\}$ é moda de X
- $\beta_1 = \sqrt{8/n}$

Atenção:

A distribuição χ^2_n não é simétrica. O quantil de probabilidade α não é simétrico do quantil de probabilidade $1-\alpha$.

É uma consequência imediata da definição que a soma de duas v.a.'s independentes com distribuição de qui-quadrado é ainda uma v.a. com distribuição de qui-quadrado. Mais especificamente, se $X \frown \chi^2_n$ e $Y \frown \chi^2_m$ são independentes, então $(X+Y) \frown \chi^2_{n+m}$.

Pelo TLC, uma vez que a distribuição de qui-quadrado é uma soma de n variáveis independentes[1] ela converge para uma distribuição normal quando $n \to \infty$.
Mais precisamente, se $X \frown \chi^2_n$, então, quando $n \to \infty$, $\frac{X-n}{\sqrt{2n}}$ converge (em distribuição) para a normal $Z \frown N(0,1)$, ou seja, para n grande, tem-se

$$\chi^2_n \overset{\text{aprox.}}{\frown} N(n, \sqrt{2n}).$$

No Mathematica, a função associada à distribuição qui-quadrado é a função ChiSquareDistribution.

[1] Sendo $Z_1, \ldots, Z_n$ independentes, pode provar-se que $Z_1^2, \ldots, Z_n^2$ também são independentes.

A distribuição t de Student

Definição (Distribuição t de Student)

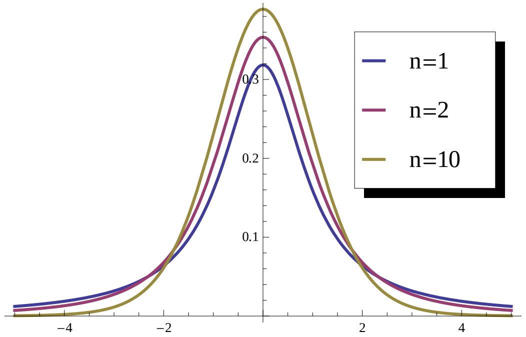
Se Z e Q_n são v.a.'s independentes tais que $Z \frown N(0,1)$ e $Q_n \frown \chi^2_n$, então dizemos que a v.a. $T = T_n$ definida por

$$T = T_n = \frac{Z}{\sqrt{Q_n}} \sqrt{n}$$

tem uma distribuição t de Student (por vezes apenas referida como distribuição t) com n graus de liberdade. Se T tem distribuição t de Student com n graus de liberdade, escrevemos simplesmente $T \frown t_n$.

A função densidade de probabilidade de $X \frown t_n$ é dada por

$$f_T(x) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}, \quad x \in \mathbb{R}.$$



Função densidade de probabilidade da distribuição $t_n; n = 1, 2, 10$.

Características teóricas da distribuição t de Student

Se $X \frown t_n$, então:

- Para $n > 1$, $\mu_X = 0$ (para $n = 1$, μ_X não existe)
- Para $n > 2$, $\sigma_X^2 = \frac{n}{n-2}$ (para $n \leq 2$, σ_X^2 não existe)
- Para $n > 3$, $\beta_1 = 0$ (para $n \leq 3$, β_1 não existe)

A distribuição t de Student é simétrica.

Pode provar-se que esta distribuição, para valores de n grandes, é aproximadamente $N(0, 1)$.

No Mathematica, a função associada à distribuição t de Student é a função StudentTDistribution.

A distribuição F de Fisher

Definição (Distribuição F de Fisher)

Se X e Y são duas v.a.'s independentes tais que $X \frown \chi_n^2$ e $Y \frown \chi_m^2$, então dizemos que a variável $V = V_{n,m}$ definida por

$$V = V_{n,m} = \frac{X/n}{Y/m}$$

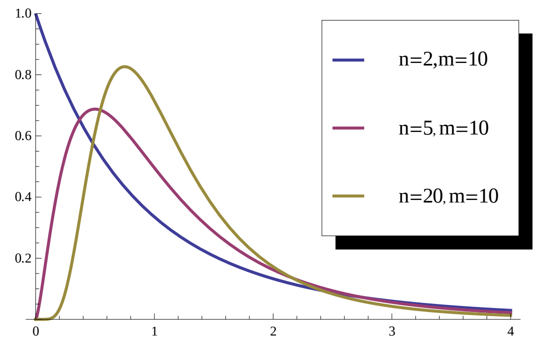
tem uma distribuição F de Fisher (ou apenas uma distribuição F, por vezes também referida como distribuição de Fisher-Snedecor) com n e m graus de liberdade. Nesse caso, escrevemos $V \frown F_{n,m}$.

Se $V \frown F_{n,m}$, então a sua função densidade de probabilidade é dada por

$$f_{n,m}(x) = \begin{cases} 0, & \text{para } x \leq 0, \\ c_{n,m} \dfrac{x^{(n-2)/2}}{(m+nx)^{(n+m)/2}}, & \text{para } x > 0, \end{cases}$$

onde

$$c_{n,m} = \frac{\Gamma(\frac{n+m}{2}) n^{n/2} m^{m/2}}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})}.$$



Função densidade de probabilidade da distribuição $F_{n,m}; n = 2, 5, 20; m = 10$.

Caraterísticas teóricas da distribuição F de Fisher

Se $X \frown F_{n,m}$, então:

- Para $m > 2$, $\mu_X = \frac{m}{m-2}$ (para $m \leq 2$, μ_X não existe)
- Para $m > 4$, $\sigma_X^2 = \frac{2m^2(m+n-2)}{n(m-2)^2(m-4)}$ (para $m \leq 4$, σ_X^2 não existe)

Não damos aqui as expressões do coeficiente de assimetria β_1 por ter uma expresão complicada.

Nota: A distribuição F não é simétrica.

No Mathematica, para trabalhar com a distribuição F, deve usar a função `FRatioDistribution`.

Notações para os quantis

Em tudo quanto se segue, o quantil de probabilidade p de cada uma das distribuições $Z \frown N(0,1)$, $U = \chi^2_m$, $T = t_m$ e $V = F_{n,m}$ será denotado, respetivamente, por z_p, $u_{m,p}$, $t_{m,p}$ e $v_{n,m,p}$.

Intervalo de confiança para μ (população normal $N(\mu, \sigma)$, σ desconhecido)

Consideremos agora o problema da determinação de um IC para o valor médio de uma distribuição $N(\mu, \sigma)$, no caso mais realista em que μ e σ são ambos desconhecidos. Neste caso, o intervalo

$$\left[\overline{X} - c\frac{\sigma}{\sqrt{n}}, \overline{X} + c\frac{\sigma}{\sqrt{n}}\right]$$

não é um estimador intervalar para μ, uma vez que os seus extremos envolvem o parâmetro desconhecido σ, ou seja, não são estatísticas.

Ideia: substituir a v.a. $Z = \frac{\overline{X}-\mu}{\sigma}\sqrt{n}$ usada na dedução do intervalo anterior, por

$$T = \frac{\overline{X} - \mu}{S}\sqrt{n},$$

onde S é a v.a. desvio padrão amostral, $S = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X})^2}$.

Pode provar-se que

$$T = \frac{\overline{X} - \mu}{S}\sqrt{n} \frown t_{n-1}$$

i.e. T tem uma distribuição t de Student com $n-1$ graus de liberdade. Então, se $c = t_{n-1,1-\alpha/2}$ for o quantil de probabilidade $1-\alpha/2$ de t_{n-1}, teremos, atendendo à simetria da distribuição t_{n-1},

$$P(-c \leq \underbrace{\frac{\overline{X} - \mu}{S}\sqrt{n}}_{T} \leq c) = 1 - \alpha$$

ou seja, teremos que

$$P\left(\overline{X} - c\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + c\frac{S}{\sqrt{n}}\right) = 1 - \alpha.$$

Assim:

$$\left[\overline{X} - c\frac{S}{\sqrt{n}}, \overline{X} + \frac{S}{\sqrt{n}}\right], \quad c = t_{n-1,1-\alpha/2}$$

é um intervalo de $(1-\alpha) \times 100\%$ confiança para o valor médio μ, no contexto de uma distribuição $N(\mu, \sigma)$, com σ desconhecido.

IC para o valor médio (população normal)

Uso do Mathematica

- No Mathematica, os IC para o valor médio (no caso de uma amostra aleatória proveniente de um modelo normal) podem ser obtidos com o comando `MeanCI`. Para usar este comando é, no entanto, necessário começar por "carregar"o pacote `HypothesisTesting`, o que deverá ser feito executando o comando: `Needs["HypothesisTesting`"]`.
- Por defeito, o IC é calculado assumindo que σ é desconhecido.
- Se quisermos determinar o IC correspondente ao caso da variância conhecida, devemos usar o comando `MeanCI` especificando `KnownVariance`$\rightarrow \sigma^2$, onde σ^2 é o valor conhecido para a variância da população.
- Por defeito, o IC calculado é um intervalo de 95% de confiança $(1-\alpha = 0.95)$. Se desejarmos um intervalo com outro nível de confiança, devemos especificar `ConfidenceLevel`$\rightarrow conf$ onde $conf$ é a confiança desejada; por exemplo, para um IC de 90% de confiança, devemos usar `ConfidenceLevel`$\rightarrow 0.9$.

Intervalo de confiança para σ^2 (população normal $N(\mu,\sigma)$)

Problema

- Dada uma amostra aleatória $x_1,\ldots,x_n$ proveniente da distribuição normal $N(\mu,\sigma)$, com σ desconhecido,
- obter um intervalo de confiança para a variância σ^2 (ou para o desvio padrão σ).

Pode provar-se que

$$U = (n-1)\frac{S^2}{\sigma^2} \frown \chi^2_{n-1}.$$

Então, sendo $c_1 = u_{n-1,1-\alpha_2}$ o quantil de probabilidade $1-\alpha/2$ da distribuição χ^2_{n-1} e $c_2 = u_{n-1,\alpha/2}$ o quantil de probabilidade $\alpha/2$ da mesma distribuição, podemos escrever

$$P\Big(c_2 \le \underbrace{(n-1)\frac{S^2}{\sigma^2}}_{U} \le c_1\Big) = 1-\alpha,$$

de onde se obtém

$$P\Big(\frac{n-1}{c_1}S^2 \le \sigma^2 \le \frac{n-1}{c_2}S^2\Big) = 1-\alpha.$$

Têm-se, assim, os seguintes IC (no contexto de uma distribuição normal $N(\mu,\sigma)$).

Intervalo de $(1-\alpha)\times 100\%$ de confiança para a variância σ^2

$$\left[\frac{n-1}{c_1}S^2\ ,\ \frac{n-1}{c_2}S^2\right],\quad c_1 = u_{n-1,1-\alpha/2},\ c_2 = u_{n-1,\alpha/2}$$

Intervalo de $(1-\alpha)\times 100\%$ de confiança para o desvio padrão σ

$$\left[\sqrt{\frac{n-1}{c_1}}S, \sqrt{\frac{n-1}{c_2}}S\right],\quad c_1 = u_{n-1;1-\alpha/2},\ c_2 = u_{n-1;\alpha/2}$$

IC para a variância (população normal)
Uso do Mathematica

- No Mathematica, os IC para a variância (no caso de uma amostra aleatória proveniente de um modelo normal) podem ser obtidos com o comando `VarianceCI`, o qual, tal como o comando `MeanCI`, necessita de carregamento prévio do pacote `HypothesisTesting`.
- Por defeito, o nível de confiança utilizado é $1-\alpha = 0.95$; outros níveis de confiança podem ser usados, desde que especifiquemos `ConfidenceLevel`$\rightarrow$ *conf*, onde *conf* é o valor de $1-\alpha$ desejado.

Intervalos unilaterais

Todos os intervalos de confiança referidos até agora são intervalos bilaterias, baseados no uso de duas estatísticas T_1 e T_2, para as quais se tem $P(T_1 \le \theta \le T_2) = 1-\alpha$. Podemos também considerar outro tipo de intervalos, ditos unilaterais, com um dos extremos infinito. Mais precisamente, se encontramos uma estatística T_1 tal que

$$P(T_1 \le \theta) = 1-\alpha,$$

então diremos que o intervalo $[T_1, +\infty[$ é um intervalo (unilateral direito) de $(1-\alpha)\%$ de confiança para θ.
De modo análogo, se tivermos uma estatística T_2 tal que

$$P(\theta \le T_2) = 1-\alpha,$$

diremos que o intervalo $]-\infty, T_2]$ é um intervalo (unilateral esquerdo) de $(1-\alpha)\times 100\%$ de confiança para θ.

Exemplo

Consideremos novamente o caso de uma amostra $X_1, \ldots, X_n$ proveniente de uma distribuição normal $N(\mu, \sigma)$, com σ conhecido. Como $Z = \frac{\overline{X}-\mu}{\sigma}\sqrt{n} \frown N(0,1)$, se $c = z_{1-\alpha}$ for o quantil de probabilidade $1-\alpha$ da normal reduzida, podemos afirmar que

$$P\Big(\frac{\overline{X}-\mu}{\sigma}\sqrt{n} \leq c\Big) = 1-\alpha$$

ou seja, que

$$P\Big(\overline{X} - c\frac{\sigma}{\sqrt{n}} \leq \mu\Big) = 1-\alpha.$$

Temos, então o seguinte **intervalo unilateral (direito) de $(1-\alpha)\times 100\%$ de confiança para μ**:

$$\Big[\overline{X} - c\frac{\sigma}{\sqrt{n}}, +\infty\Big[, \; c = z_{1-\alpha}.$$

Outros intervalos de confiança unilaterais correspondentes as casos bilaterias anteriormente estudados seriam obtidos de modo análogo.

Duas amostras

Os IC considerados até agora dizem respeito a um único valor médio, uma única proporção ou uma única variância. Em muitos casos há interesse em estudar situações que envolvem valores médios, proporções e variâncias relativos a duas amostras.

Quando temos amostragem envolvendo duas v.a.s podemos:

- Ter duas populações e duas amostras aleatórias $X_1, \ldots, X_n$ e $Y_1, \ldots, Y_m$ (com a mesma dimensão ou não), recolhidas independentemente uma da outra; exemplo: peso em populações masculina e feminina.
- Ter duas populações e duas correspondentes amostras aleatórias, intencionalmente relacionadas (logo, com a mesma dimensão), $X_1, \ldots, X_n$ e $Y_1, \ldots, Y_n$; exemplo: alturas de pares irmão/irmã.
- Ter uma população e duas amostras aleatórias $X_1, \ldots, X_n$ e $Y_1, \ldots, Y_n$ correspondentes a duas v.a.s que estão, eventualmente, relacionadas; exemplo: peso e altura numa população de homens ou peso antes e depois de um tratamento de emagrecimento numa certa população.

No primeiro caso temos duas amostras independentes e nos outros casos dizemos que temos amostras emparelhadas.

Amostras independentes de dois modelos normais

IC para a diferença dos valores médios

Passamos agora a estudar IC apropriados ao caso de duas amostras. Começamos com o caso de termos duas amostras independentes provenientes de modelos normais.

Suponhamos, então que temos duas amostras $X_1, \ldots, X_n$ e $Y_1, \ldots, Y_m$, independentes, com $X_i \frown N(\mu, \sigma)$ e $Y_i \frown N(\mu', \sigma')$ e que pretendemos obter IC para a diferença dos valores médios, $\mu - \mu'$.

Neste caso, há três hipóteses a considerar:

1. as variâncias σ^2 e σ'^2 são conhecidas;
2. as variâncias σ^2 e σ'^2 são desconhecidas, mas sabe-se que $\sigma = \sigma'$;
3. não temos qualquer informação sobre as variâncias σ^2 e σ'^2.

É possível deduzir IC de $(1-\alpha)\times 100\%$ de confiança para a diferença dos valores médios $\mu - \mu'$, para cada um dos casos acima mencionados. Vamos apenas indicar como se podem calcular esses IC, usando o Mathematica.

Amostras independentes de dois modelos normais

IC para a diferença dos valores médios

- No Mathematica, os IC para a diferença de valores médios , no caso de amostras independentes provenientes de modelos normais, podem ser obtidos com o comando `MeanDifferenceCI`, disponível no pacote `HypothesisTesting`.
- Por defeito, o IC é calculado assumindo que σ^2 e σ'^2 são quaisquer.
- Se quisermos determinar o IC correspondente ao caso das variâncias desconhecidas, mas iguais, devemos usar o comando `MeanDifferenceCI`, especificando `EqualVariances`$\rightarrow$ `True`.
- No caso de as variâncias serem conhecidas, devemos usar o comando `MeanDifferenceCI` especificando `KnownVariance`$\rightarrow \{\sigma^2, \sigma'^2\}$ onde σ^2 e σ'^2 são os valores conhecidos das variâncias.

IC para a diferença de valores médios

Amostras emparelhadas (de modelos normais)

No caso de termos duas amostras aleatórias emparelhadas $X_1, \ldots, X_n$ e $Y_1, \ldots, Y_n$ provenientes de modelos normais, $X \frown N(\mu, \sigma)$ e $Y \frown N(\mu', \sigma')$, pode provar-se que a v.a. diferença, $D = X - Y$, tem uma distribuição normal, com valor médio $\mu - \mu'$. No entanto, como as v.a.'s não são independentes, mesmo conhecendo as variâncias de X e de Y, não conhecemos a variância de D. Se usarmos o intervalo de confiança usual para uma população normal com variância desconhecida para a amostra das diferenças, $D_1 = X_1 - Y_1, \ldots, D_n = X_n - Y_n$, obtemos de imediato o seguinte intervalo de $(1-\alpha) \times 100\%$ de confiança para a diferença dos valores médios $\mu - \mu'$:

$$\left[\overline{D} - c\frac{S_D}{\sqrt{n}}, \overline{D} + c\frac{S_D}{\sqrt{n}}\right],$$

onde $S_D = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(D_i - \overline{D})^2}$ e $c = t_{n-1,1-\alpha/2}$.

No Mathematica, devemos usar o comando `MeanCI` aplicado à amostra das diferenças, sem especificação de variância.

Amostras independentes de dois modelos normais

IC para o quociente das variâncias

No caso de duas amostras independentes provenientes de modelos normais $X \sim (\mu, \sigma)$ e $Y \sim (\mu', \sigma')$, pode mostrar-se que a v.a. $V = \dfrac{S^2/\sigma^2}{S'^2/\sigma'^2}$ tem uma distribuição F de Fisher com $n-1$ e $m-1$ graus de liberdade. Com base nesta v.a., é fácil de obter o seguinte intervalo de $(1-\alpha) \times 100\%$ de confiança para o quociente das variâncias σ^2/σ'^2:

$$\left[\frac{1}{v_{n-1,m-1,1-\alpha/2}}\frac{S^2}{S'^2}, \frac{1}{v_{n-1,m-1,\alpha/2}}\frac{S^2}{S'^2}\right],$$

onde $v_{n-1,m-1,p}$ designa o quantil de probabilidade p da distribuição $F_{n-1,m-1}$.

No Mathematica, estes intervalos de confiança são obtidos como o comando `VarianceRatioCI`, disponível no pacote `HypothesisTesting`.

Testes estatísticos

Os chamados testes estatísticos constituem uma parte muito importante da inferência estatística.

Os primeiros testes que vamos estudar são testes paramétricos, i.e. são testes sobre valores de parâmetros de uma distribuição e baseiam-se nos IC que estudámos.

Tal como no caso da estimação de um parâmetro, partimos de uma amostra, $x_1, \ldots, x_n$, que admitimos ser uma realização de uma amostra aleatória $X_1, \ldots, X_n$ de uma determinada população, cuja distribuição envolve um parâmetro desconhecido θ.

Contudo, em vez estimarmos explicitamente o parâmetro desconhecido, o que pretendemos agora é usar essa amostra para testar uma hipótese (i.e. uma conjetura) particular sobre esse parâmetro desconhecido.
Se essa hipótese especificar completamente a distribuição subjacente aos dados, dizemos que se trata de uma hipótese simples; caso contrário, dizemos que estamos perante uma hipótese composta. Por exemplo, no caso de os dados virem de uma distribuição normal $N(\mu, \sigma)$ com σ conhecido, a hipótese $\mu = \mu_0$ é simples e a hipótese $\mu \leq \mu_0$ é composta.

Hipótese nula e hipótese alternativa

- Chamamos hipótese nula à hipótese que pretendemos testar e designamo-la por H_0.
- Em geral, é formulada uma hipótese alternativa, que designamos por H_1 (ou H_A).
- Nesse caso, dizemos que estamos perante um teste de hipóteses e que vamos testar a hipótese H_0 *versus* a hipótese H_1 (escrevemos, abreviadamente H_0 *vs* H_1). Quando referimos que H_1 é uma hipótese alternativa, queremos dizer que H_0 e H_1 devem ser incompatíveis, i.e., se se verificar H_0 não pode verificar-se H_1.

Exemplo

Suponhamos que temos uma moeda e que pretendemos testar se ela é ou não equilibrada. Estamos, assim, perante um modelo de Bernoulli $Ber(p)$, em que p é a probabilidade de sair cara. Neste caso, portanto, estamos interessados em testar a hipótese $H_0 : p = \frac{1}{2}$ *vs* $H_1 : p \neq \frac{1}{2}$.

Estatística de teste

- Chama-se teste de uma hipótese estatística H_0 a uma regra usada para decidir se os dados são ou não consistentes com essa hipótese e, portanto, se rejeitamos ou não essa hipótese.
- Essa regra vai ser baseada na utilização de uma estatística de teste, que é uma função $T = T(X_1, \ldots, X_n)$ da amostra aleatória $X_1, \ldots, X_n$.

  Em geral, essa estatística:
  - envolve um estimador do parâmetro que estamos a testar;
  - "mede" o afastamento dos dados em relação a H_0;
  - supondo a hipótese nula verdadeira, tem uma distribuição conhecida.

Exemplo

Suponhamos que estamos perante uma amostra proveniente de uma distribuição $N(\mu, \sigma)$, com μ desconhecido e σ conhecido, e que pretendemos testar

$$H_0 : \mu = \mu_0 \text{ vs } H_1 : \mu \neq \mu_0.$$

Fará sentido considerar como estatística de teste a v.a.
$T = T(X_1, \ldots, X_n) = \overline{X}$ ou, se for mais conveniente, a v.a. padronizda

$$T = T(X_1, \ldots, X_n) = \frac{\overline{X} - \mu_0}{\sigma}\sqrt{n}.$$

Note-se que esta v.a. envolve o estimador $\overline{X}$ para μ e mede a distância (expressa em desvios padrão) entre $\overline{X}$ e o seu valor esperado, quando H_0 é verdadeira, i.e. quando $\mu = \mu_0$; além disso, se a hipótese nula for verdadeira, essa v.a. tem uma distribuição conhecida (é normal $N(0,1)$).

Região crítica

Com a ajuda desta v.a. $T(X_1, \ldots, X_n)$, e dependendo de um valor α de probabilidade, por nós fixado (α pequeno), vamos definir uma região do plano, $\mathcal{C} = \mathcal{C}(\alpha)$, a chamada região crítica ou região de rejeição.
Com base na amostra aleatória concreta de dados $x_1, \ldots, x_n$, a regra de decisão consistirá em rejeitar H_0 se e só se o valor observado da estatística de teste, $T(x_1, \ldots, x_n)$, pertencer à região crítica.
Em resumo, o teste será da forma:

Rejeite-se H_0 se e só se $T \in \mathcal{C}$

Nota: $T \in \mathcal{C}$ deve ser entendido como: "o valor observado de T está em $\mathcal{C}$".

Erros de tipo I e tipo II

Ao tomar a decisão de rejeitar ou não a hipótese nula, podemos cometer dois tipos de erro:

1. rejeitarmos H_0, sendo H_0 verdadeira $\rightarrow$ erro de tipo I ou erro de primeira espécie
2. não rejeitarmos H_0, sendo H_1 verdadeira (H_0 falsa) $\rightarrow$ erro de tipo II ou erro de segunda espécie

Nível de significância

Em geral, consideramos mais grave cometer um erro de tipo I do que um erro do tipo II. Assim, especificamos um certo valor α pequeno, (tipicamente $\alpha = 0.05, 0.01$) e escolhemos um teste que garanta que a probabilidade de cometer um erro do tipo I seja igual a α, i.e.

$$P(\text{rejeitar } H_0 \,|\, H_0 \text{ verdadeira}) = \alpha.$$

A este valor de α chamamos nível de significância do teste. [2]
Denota-se por β a probabilidade de cometer um erro de segunda espécie, i.e.

$$\beta = P(\text{não rejeitar } H_0 \,|\, H_1 \text{verdadeira}).$$

O valor $1-\beta$ (i.e. a probabilidade de não cometer um erro do tipo II) é a chamada potência do teste (ao nível α).

[2] Esta definição pressupõe que a hipótese nula é simples; no caso de uma hipótese composta, por exemplo, do tipo $\theta \leq \theta_0$, a probabilidade de cometer um erro do tipo I varia em função dos valores que θ pode assumir quando H_0 é verdadeira; nesse caso, o nível de significância é definido como a "maior" dessas probabilidades .

Testes sobre o valor médio em população normal (σ conhecido)

Suponhamos que dispomos de uma amostra observada $x_1, \ldots, x_n$ correspondente a uma amostra aleatória $X_1, \ldots, X_n$ proveniente de uma distribuição $N(\mu, \sigma)$, com μ desconhecido e σ conhecido e que pretendemos testar a hipótese nula

$$H_0 : \mu = \mu_0$$

versus a hipótese alternativa

$$H_1 : \mu > \mu_0$$

onde μ_0 é uma certa constante especificada. Este tipo de hipótese alternativa é dita uma hipótese unilateral (em oposição a uma hipótese do tipo $H_1 : \mu \neq \mu_0$, dita bilateral).
Como $\overline{X}$ é um estimador natural de μ, faz sentido, atendendo à forma da hipótese alternativa H_1, procurar uma região de rejeição definida por uma expressão da forma

$$\overline{X} > k.$$

Se desejarmos um teste com nível de significância α, devemos determinar o valor de k para o qual se tenha um erro de tipo I igual α, isto é, k deverá ser tal que se tenha

$$P\Big(\overline{X} > k \,|\, H_0 \text{ verdadeira } \Big) = P\Big(\overline{X} > k \,|\, \mu = \mu_0\Big) = \alpha.$$

Seja $Z = \frac{\overline{X}-\mu_0}{\sigma}\sqrt{n}$. Temos

$$\overline{X} > k \iff Z > \underbrace{\frac{k-\mu_0}{\sigma}\sqrt{n}}_{c}$$

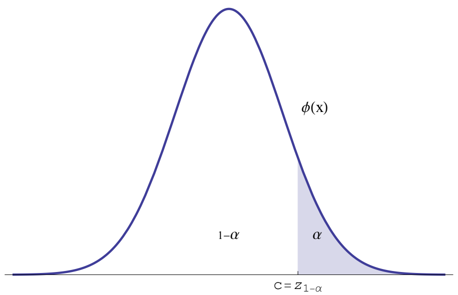
Logo, temos

$$P\Big(\overline{X} > k \,|\, \mu = \mu_0\Big) = \alpha \iff P\Big(Z > c \,|\, \mu = \mu_0\Big) = \alpha,$$

onde $c = \frac{k-\mu_0}{\sigma}\sqrt{n}$.

Neste caso, como σ é conhecido, se a hipótese nula se verificar, isto é, se $\mu = \mu_0$, sabemos que

$$Z = \frac{\overline{X}-\mu_0}{\sigma}\sqrt{n} \frown N(0,1).$$



Concluímos assim que $c = \frac{k-\mu_0}{\sigma}\sqrt{n}$ deverá ser o quantil de probabilidade $1-\alpha$ da distribuição normal $N(0,1)$ i.e. deverá ter-se $c = \frac{k-\mu_0}{\sigma}\sqrt{n} = z_{1-\alpha}$, ou seja, k deverá ser escolhido como

$$k = \mu_0 + c\frac{\sigma}{\sqrt{n}}, \quad c = z_{1-\alpha}.$$

Temos então que o teste (ao nível de significância α) consistirá no seguinte: rejeitaremos a hipótese nula H_0 $(\mu = \mu_0)$ se

$$\overline{X} > \mu_0 + c\frac{\sigma}{\sqrt{n}}$$

(não rejeitando H_0 se $\overline{X} \leq \mu_0 + c\frac{\sigma}{\sqrt{n}}$), ou, de modo equivalente, rejeitaremos H_0 se e só se (o valor observado da estatística de teste) $Z = \frac{\overline{X}-\mu_0}{\sigma}\sqrt{n}$ for superior a c, onde $c = z_{1-\alpha}$. Em resumo, o teste pode ser descrito do seguinte modo:

Rejeite-se H_0 se e só se

$$Z = \frac{\overline{X}-\mu_0}{\sigma}\sqrt{n} > c, \text{ onde } c = z_{1-\alpha}.$$

Teste e IC

Note-se que

$$\frac{\overline{x}-\mu_0}{\sigma}\sqrt{n} > c \iff \mu_0 < \overline{x} - c\frac{\sigma}{\sqrt{n}}$$
$$\iff \mu_0 \notin \left[\overline{x} - c\frac{\sigma}{\sqrt{n}}\,,\, +\infty\right[$$

Mas, $\left[\overline{x} - c\frac{\sigma}{\sqrt{n}}\,,\, +\infty\right[$, onde $c = z_{1-\alpha}$, é o intervalo de confiança unilateral direito (com margem de erro α) para μ, no caso σ conhecido, deduzido anteriormente.
Em conclusão: este teste equivale a tomar a decisão com base num IC unilateral direito para μ – rejeita-se H_0 se e só se μ_0 **não pertencer** a esse intervalo de confiança.

De modo análogo, se obtém o teste para o caso de outra hipótese unilateral

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu < \mu_0$$

ou o teste para o caso de uma hipótese bilateral

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu \neq \mu_0.$$

No primeiro caso, o teste corresponderá a tomar a decisão com base no intervalo de confiança unilateral esquerdo para μ, sendo

Rejeite-se H_0 se e só se

$$Z = \frac{\overline{X}-\mu_0}{\sigma}\sqrt{n} < -c, \quad c = z_{1-\alpha}.$$

No segundo, a decisão deve ser tomada com base no intervalo de confiança bilateral para μ e deverá ser

Rejeite-se H_0 se e só se

$$|Z| = \left|\frac{\overline{X}-\mu_0}{\sigma}\sqrt{n}\right| > c, \quad c = z_{1-\alpha/2}.$$

Exemplo

Sabe-se que, quando um sinal com valor μ é enviado de uma certa localidade A para uma localidade B, ele é recebido em B com uma distribuição $N(\mu, 2)$. Suponhamos que um determinado sinal foi enviado, independentemente e nas mesmas condições, 5 vezes, e que a média dos valores recebidos foi $\overline{x} = 9.5$. Vejamos se, ao nível de significância de 5%, podemos aceitar que o sinal enviado foi $\mu = 8$, ou se devemos rejeitar essa hipótese.
Neste caso, estamos perante um teste bilateral $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$, com $\mu_0 = 8$, no âmbito de um modelo normal com desvio padrão conhecido, $\sigma = 2$. A amostra selecionada tem dimensão $n = 5$ e o nível de significância que pretendemos é $\alpha = 0.5$. O valor observado da estatística de teste é

$$z = \frac{\overline{x}-\mu_0}{\sigma}\sqrt{n} = \frac{9.5-8}{2}\sqrt{5} = 1.68.$$

Por outro lado, temos $z_{1-\alpha/2} = z_{0.975} = 1.96$. Como $|z| = 1.68 < 1.96$, não rejeitamos a hipótese $\mu = 8$ ao nível de significância indicado.

Valor-p

Em alternativa à abordagem de fixar *a priori* o nível de significância do teste e rejeitar ou não H_0 em função do valor observado da estatística de teste, podemos começar por calcular o chamado valor-p (ou nível de significância do resultado), o qual é dado pela probabilidade de, admitindo que a hipótese H_0 é verdadeira, se obter um valor da estatística de teste "tão ou mais extremo" do que aquele que foi observado.

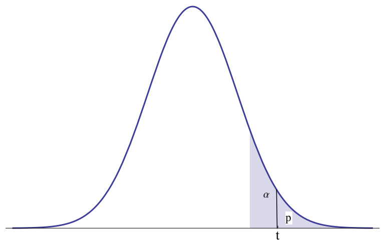
O que entendemos por "tão ou mais extremo" depende da hipótese alternativa que estamos a considerar; em geral, se a estatística de teste for T e t for o seu valor observado, tem-se :

- valor-$p = P(T \geq t)$, se a hipótese alternativa for do tipo $\theta > \theta_0$;
- valor-$p = P(T \leq t)$, se a hipótese alternativa for do tipo $\theta < \theta_0$;
- valor-$p = P(|T| \geq |t|)$, se a hipótese alternativa for do tipo $\theta \neq \theta_0$.

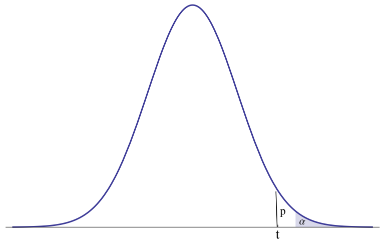
Quanto menor for o valor-p, mais "contrários" à hipótese nula são os dados. Calculado o valor-p, a conclusão do teste a um nível de significância α resulta da comparação desse valor com α, tendo-se:

- valor-$p < \alpha \Rightarrow$ rejeita-se H_0 ao nível α;
- valor-$p \geq \alpha \Rightarrow$ não se rejeita H_0 ao nível α.

Valor-p



Valor-p inferior a $\alpha \rightarrow$ rejeitamos H_0



Valor-p superior a $\alpha \rightarrow$ não rejeitamos H_0

Exemplo

Consideremos de novo o exemplo anterior da transmissão do sinal, e vejamos como seria a abordagem usando o valor-p.
Neste caso, como temos uma hipótese alternativa bilateral, tem-se

$$p = P\Big(|Z| \geq |z|\Big) = P\Big(|Z| \geq 1.68\Big) = 0.093.$$

Nota: O valor 0.093 acima pode obter-se facilmente usando o Mathematica, tendo em conta que, supondo a hipótese nula verdadeira, $Z \frown N(0,1)$.
Assim, não há evidência para rejeitar H_0 para qualquer nível de significância $\alpha \leq 0.093$; por exemplo, não rejeitamos H_0 ao nível de significância 5%, já que $0.093 > 0.05$; no entanto, rejeitaríamos H_0 ao nível de significância 10%, uma vez que $0.093 < 0.1$.

Procedimento usual

- Quando p é tal que $0.01 \leq p < 0.05$, é usual dizer-se que o resultado é significativo, havendo evidência para rejeitar H_0.
  (Em muitos pacotes de *software* estatístico, a um valor-p desta grandeza associa-se o símbolo $*$).
- Se $0.001 \leq p < 0.01$, dizemos que o resultado é muito significativo e que há grande evidência para rejeitar H_0.
  (Usa-se, geralmente, o símbolo $**$ como código para um valor deste tipo.).
- Se $p < 0.001$, dizemos que o resultado é altamente significativo e que há uma evidência muito grande para rejeitar H_0.
  (A um valor desta grandeza é usualmente atribuído o código $***$).
- Se $p \geq 0.05$, considera-se, em geral, que o resultado não é significativo e que não há evidência para rejeitar H_0.

Outro tipo de hipótese nula

Até aqui considerámos apenas o caso em que a hipótese nula é uma hipótese simples, da forma $H_0 : \mu = \mu_0$.
Se, por exemplo, pretendermos testar a hipótese

$$H_0 : \mu \leq \mu_0 \text{ vs } H_1 : \mu > \mu_0,$$

faz sentido considerar a mesma região crítica que no caso $H_0 : \mu = \mu_0$ *vs* $H_1 : \mu > \mu_0$.
De modo análogo, a região crítica para

$$H_0 : \mu \geq \mu_0 \text{ vs } H_1 : \mu < \mu_0,$$

é a mesma do que a correspondente às hipóteses $H_0 : \mu = \mu_0$ *vs* $H_1 : \mu < \mu_0$.

Testes sobre o valor médio em população normal (σ desconhecido)

Consideremos agora o caso em que a amostra provém de uma distribuição normal $N(\mu, \sigma)$, mas em que σ é desconhecido.
Suponhamos que pretendemos testar

$$H_0 : \mu = \mu_0 \text{ vs } H_1 : \mu > \mu_0.$$

Neste caso, em analogia com o que fizemos para o caso dos intervalos de confiança para o valor médio com σ desconhecido, a v.a. apropriada para usar como estatística de teste será

$$T = \frac{\overline{X} - \mu_0}{S}\sqrt{n},$$

a qual, como sabemos, tem uma distribuição t_{n-1} (t de Student com $n-1$ graus de liberdade).

Seguindo o processo descrito na secção anterior passo a passo, com as devidas adaptações, facilmente se conclui que o teste, ao nível de significância α, consistirá no seguinte:

Rejeite-se H_0 se e só se

$$T = \frac{\overline{X} - \mu_0}{S}\sqrt{n} \geq t_{n-1,1-\alpha},$$

onde $t_{n-1,1-\alpha}$ designa o quantil de probabilidade $1-\alpha$ da distribuição t_{n-1}.

De modo análogo se deduzem os testes adequados a outras hipóteses, neste caso em que σ é desconhecido.

Outros testes para modelo normal

Para todos os de intervalos de confiança referidos anteriormente podemos deduzir "correspondentes" testes de hipóteses, seguindo uma metodologia análoga à que descrevemos.
Note-se que, aos intervalos de confiança para a diferença de valores médios vão corresponder testes para a igualdade de valores médios e aos intervalos de confiança para o quociente de variâncias vão corresponder testes para a igualdade das variâncias, uma vez que

$$\mu = \mu' \iff \mu - \mu' = 0 \quad \text{e} \quad \sigma = \sigma' \iff \frac{\sigma}{\sigma'} = 1.$$

Nota: Um formulário com uma descrição sucinta de alguns dos testes referidos está disponível na plataforma.
Segue-se uma descrição de como usar os testes para o modelo normal, com o Mathematica.

Testes para o valor médio (modelo normal)
Uso do Mathematica

- Os comandos do Mathematica que vamos utilizar para testes relativos ao valor médio são: ZTest e TTest.
- Usamos ZTest (baseado no uso da estatística $Z = \frac{\overline{X}-\mu_0}{\sigma}\sqrt{n}$) quando conhecemos o valor da variância e TTest (baseado no uso da estatística $T = \frac{\overline{X}-\mu_0}{S}\sqrt{n}$), no caso contrário.
- Por defeito, os testes são bilaterais; para mudar o tipo de hipótse alterantiva, usa-se AlternativeHypothesis $\rightarrow$ "Less"ou AlternativeHypothesis$\rightarrow$ "Greater".
- Por defeito, ao aplicar o teste, este dá como resultado o valor-p. No entanto, é possível obter outro tipo de informação (por exemplo, podemos pedir o valor observado da estatística de teste ou a indicação de qual a decisão a tomar - rejeição ou não da hipótse nula, etc.).
- Por defeito, o nível de significância é $\alpha = 0.05$ (5%); o nível de significância do teste pode ser alterado, especificando SignificanceLevel $\rightarrow \alpha$, onde α é o valor desejado.

Testes para a igualdade das médias (modelo normal)
Uso do Mathematica

- Os comandos ZTest e TTest são também usados para testes relativos à igualdade de valores médios. Usa-se ZTest, se conhecermos as variâncias e TTest, caso contrário.
  Nota: Neste último caso, contariamente ao que acontece quando se usa o comando MeanDifferenceCI, não temos que distinguir o caso de variâncias desconhecidas, mas iguais, do caso de variâncias desconhecidas quaisquer - é feito um teste automático, a partir das amostras, para "concluir"qual das hipóteses a considerar e o teste é ajustado devidamente.
- Quando temos duas amostras independentes, os comandos são usados com $\{amostra_1, amostra_2\}$ (sendo cada uma das amostras dada como uma lista de números).
- No caso de duas amostras emparelhadas, usa-se o comando TTest com apenas uma lista: a que contém a amostra das diferenças; em alternativa, podem dar-se as duas amostras e usar o comando PairedTTest.

Testes sobre variâncias (modelo normal)
Uso do Mathematica

- Para testes sobre a variância (ou sobre igualdade de variâncias) referentes a amostras de população (ou populações) com distribuição normal, usaremos o comando FihserRatioTest.
- O teste de igualdade de variâncias é feito comparando o seu quociente (i.e. a sua razão) com o valor um (e não a sua diferença com o valor zero) e usa a estatística $V = \frac{S^2}{S'^2}$, a qual tem distribuição distribuição F de Fisher; daí o nome do comando usado para este tipo de teste.
- No caso de uma única amostra, o comando FihserRatioTest usa a estatística $U = (n-1)\frac{S^2}{\sigma_0^2}$ já referida, a qual tem uma distribuição de qui-quadrado com $n-1$ graus de liberdade.

Testes de ajustamento

A ideia de um teste de ajustamento é testar se uma determinada distribuição F_0 é ou não um modelo adequado para determinados dados. Assim, dado um conjunto de dados $x_1, \ldots, x_n$, que interpretamos como valores observados de uma amostra aleatória $X_1, \ldots, X_n$ de uma v.a. X, iremos testar

$$H_0: \ X \text{ segue a distribuição } F_0$$

vs

$$H_1: X \text{ não segue a distribuição } F_0$$

Podemos estar perante uma das seguintes situações:

1. F_0 está completamente especificada
2. F_0 envolve r parâmetros desconhecidos $\theta_1, \ldots, \theta_r$.

Exemplo

Efetuaram-se $1\,000$ lançamentos de 4 moedas e registou-se o número total de caras obtido. Os dados resultantes estão registados na tabela seguinte:

n. caras	0	1	2	3	4
n. lançamentos (4 moedas)	41	239	376	281	63

Sendo X a v.a. que representa o número da caras obtido no lançamento das 4 moedas, podemos, por exemplo, considerar como hipótese nula:

- $H_0 : X \frown Bi(4, 0.5)$; neste caso, F_0 estaria totalmente especificada
- $H_0 : X \frown Bi(4, p)$; neste caso, F_0 não estaria totalmente especificada, pois envolveria o parâmetro p desconhecido.

O teste de ajustamento de qui-quadrado

Este é um dos testes de ajustamento mais simples e adequa-se quer a variáveis qualitativas quer a variáveis quantitativas. Vamos, então, descrever o seu procedimento.

1. Classificam-se os n dados da amostra em k categorias ou classes $A_1, \ldots, A_k$, mutuamente exclusivas e exaustivas.
2. Para cada classe $A_j; j = 1, \ldots, k$, calculam-se:
   - a frequência (absoluta) observada dessa classe, o_j, i.e.

     $$o_j = \text{número de observações em } A_j$$

   - a frequência (absoluta) esperada dessa classe, e_j, i.e.

     $$e_j = np_j,$$

     onde p_j designa a probabilidade da classe A_j sob a validade da hipótese nula, i.e. $p_j = P(A_j |\ X \frown F_0)$.

     Nota: Quando F_0 envolve r parâmetros desconhecidos $\theta_1, \ldots, \theta_r$, p_j será a probabilidade de A_j sob a hipótese de X seguir uma distribuição F_0 com os valores dos parâmetros desconhecidos $\theta_1, \ldots, \theta_r$ substituídos pelas respetivas estimativas de máxima verosimilhança $\hat{\theta}_1, \ldots, \hat{\theta}_r$, i.e. $p_j = P(A_j \mid X \frown (F_0; \hat{\theta}_1, \ldots, \hat{\theta}_r))$.

3. Considerando o_j e e_j como valores observados de v.a.s O_j e E_j, respetivamente, pode mostrar-se que a v.a.

   $$Q_k = \sum_{j=1}^{k} \frac{(O_j - E_j)^2}{E_j}$$

   (que "mede" o afastamento das frequências observadas em relação às que seria de esperar se o modelo fosse válido) tem uma distribuição aproximada [3] de qui-quadrado com $k - r - 1$ graus de liberdade, i.e.

   $$Q_k \frown \chi^2_{k-r-1}.$$

   Nota: No caso em que F_0 é totalmente especificada, tem-se $r = 0$.
4. Considera-se, então, Q_k como estatística de teste, sendo a região de rejeição – para um teste com nível de significância α – da forma

   $$Q_k > u_{k-r-1,1-\alpha},$$

   onde $u_{k-r-1,1-\alpha}$ designa o quantil de probabilidade $1 - \alpha$ da distribuição χ^2_{k-r-1}.

[3] A aproximação é geralmente razoável se n for grande ($n \geq 30$) e $E_j \geq 5, \forall j$.

Exemplo

Consideremos novamente os dados do exemplo anterior (do lançamento de 4 moedas $1\,000$ vezes), correspondentes à seguinte tabela de frequências observadas

A_i	0	1	2	3	4
o_i	41	239	376	281	63

e testemos a hipótese H_0 : *os dados provêm de uma distribuição* $X \frown Bi(4, 0.5)$ *vs* a hipótese H_1 : *os dados não provêm dessa distribuição*. Sendo $X \frown Bi(4, 0.5)$, a f.m.p. de X e correspondentes frequências esperadas são:

A_i	0	1	2	3	4
p_i	0.0625	0.25	0.375	0.25	0.0625
e_i	62.5	250	375	250	62.5

Tem-se, então

$$q_k = \frac{(41-62.5)^2}{62.5} + \frac{(239-250)^2}{250} + \frac{(376-375)^2}{375} + \frac{(281-250)^2}{250} + \frac{(63-62.5)^2}{62.5} = 11.73.$$

Exemplo (cont.)

Neste caso, temos $k = 5$ classes e não há parâmetros desconhecidos na distribuição $F_0 = Bi(4, 0.5)$ (i.e. $r = 0$), pelo que vamos considerar o quantil de probabilidade de uma distribuição qui-quadrado com $k - 1 = 4$ graus de liberdade.
Escolhendo $\alpha = 0.05$, tem-se $u_{4,0.95} = 9.49$.Como o valor observado da estatística de teste, $q_k = 11.73$, é superior a 9.49, devemos rejeitar a hipótese de que os dados provêm de uma distribuição $Bi(4, 0.5)$ (ao nível de significância $\alpha = 0.05$).

Suponhamos, agora, que pretendemos testar H_0: *os dados provêm de uma distribuição* $Bi(4, p)$ *vs* H_1: *os dados não provêm de uma distribuição* $Bi(4, p)$ (com p desconhecido, a estimar).
Neste caso, vamos começar por calcular a estimativa m.v. para p, $\hat{p} = \dfrac{\overline{x}}{4}$:

$$\frac{\overline{x}}{4} = \frac{0 \times 41 + 1 \times 239 + 2 \times 376 + 3 \times 281 + 4 \times 63}{4 \times 1000} = 0.5215.$$

Exemplo (cont.)

A f.m.p. de $X \frown Bi(4, 0.5215)$ e as respetivas frequências esperadas são:

A_i	0	1	2	3	4
p_i	0.05243	0.2285	0.3736	0.2715	0.07397
e_i	52.43	228.5	373.6	271.5	73.97

Calculando o valor observado da estatística de teste (de modo análogo ao que fizemos atrás), vem $q_k = 4.95$. Como usámos uma estimativa m.v. para um parâmetro (i.e. $r = 1$), devemos procurar o quantil da distribuição qui-quadrado com $k - 1 - r = 5 - 1 - 1 = 3$ graus de liberdade. Para $\alpha = 0.05$, tem-se $u_{3,0.95} = 7.81$. Como $4.95 < 7.81$, não rejeitamos a hipótese nula de que os dados vêm de uma distribuição $Bi(4, p)$, sendo p estimado por $\hat{p} = 0.5125$.

Teste de ajustamento de qui-quadrado

Notas

- Os testes de qui-quadrado aplicam-se quer a variáveis qualitativas quer a variáveis quantitativas (agrupando os dados em classes);
- exigem amostras de dimensão grande ($n \geq 30$);
- pode ser necessário juntar uma ou mais classes das que foram inicialmente constituídas, por forma a garantir que $E_j \geq 5$;
- no caso de variáveis contínuas, é usual considerarem-se k classes equiprováveis, i.e., com $p_j = 1/k$, sendo $k \leq n/5$.

O Mathematica dispõe de um comando próprio para efetuar testes de ajustamento de qui-quadrado, `PearsonChiSquareTest`; no entanto, esse comando exige que a amostra seja dada "em bruto"(i.e. não podemos trabalhar diretamente com tabelas de frequências observadas) e a hipótese nula tem de envolver distribuições disponíveis no Mathematica.

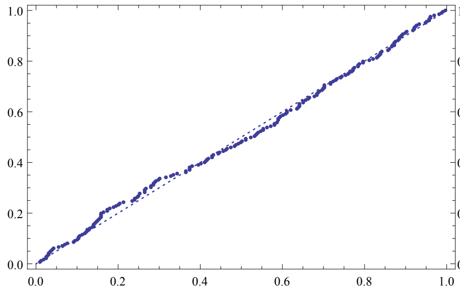
Gráfico de probabilidade normal

Como vimos, em muitas situações, é importante saber se os dados provêm de uma distribuição normal.
Um primeiro processo visual rápido para avaliar se os dados poderão vir de um modelo normal, usado normalmente antes de qualquer outro tipo de teste, consiste em esboçar o chamado gráfico de probabilidade normal.
Os pormenores de obtenção desse gráfico variam conforme o *software* utilizado, mas a ideia básica é comparar as frequências (relativas) acumuladas da amostra com os respetivos valores da função de distribuição teórica de uma distribuição normal. Se os dados se ajustarem a um modelo normal, os pontos do gráfico deverão dispor-se aproximadamente sobre uma reta.

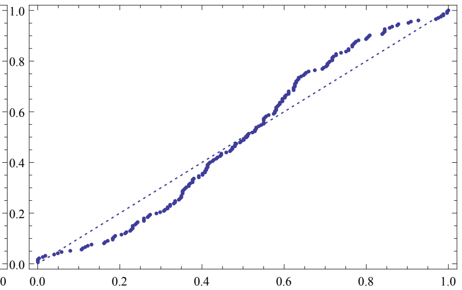
No Mathematica, um gráfico de probabilidade normal pode ser obtido com o comando `ProbabilityPlot`. Por defeito, a comparação é feita com os valores da função de distribuição da v.a. $N(\mu, \sigma)$ com os valores de μ, σ estimados a partir da amostra (substituídos por $\overline{x}$ e s).

Gráficos de probabilidade normal

Exemplos



Amostra de $X \frown N(0,1)$



Amostra de $X \frown t_3$

Teste de normalidade de Shapiro-Wilk

Existem muitos testes de normalidade, i.e. testes destinados a decidir se é razoável ou não assumir que os dados provêm de uma distribuição normal. (O próprio teste de ajustamento de qui-quadrado pode ser usado como um teste de normalidade.) De entre os diversos testes, um dos mais usados é o teste de Shapiro-Wilk. Este teste é baseado no uso da estatística

$$W = \frac{\left(\sum_{i=1}^{n} a_i X_{(i)}\right)^2}{(n-1)S^2}$$

em que as constantes a_i são dadas (variando com n). Tem-se $0 \leq W \leq 1$ e espera-se que $W \approx 1$ quando a hipótese nula (os dados provêm de um modelo normal) for verdadeira.

No Mathematica, o comando para efetuar um teste de Shapiro-Wilk é o comando `ShapiroWilkTest`.