Distribuição Normal

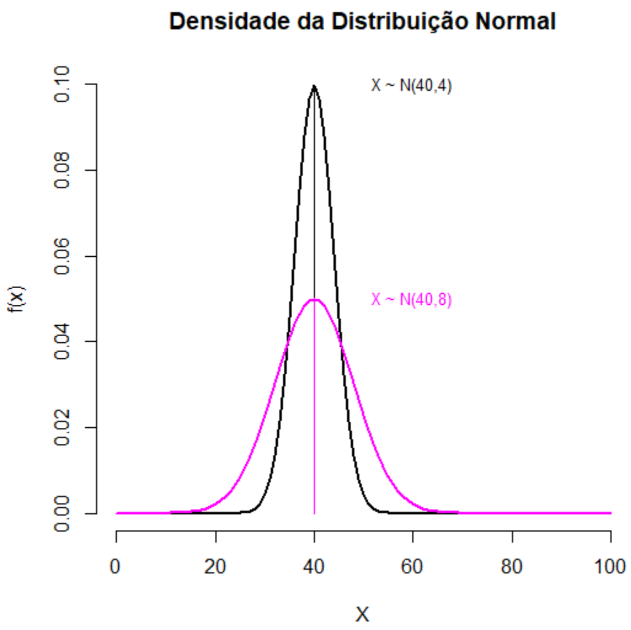
É totalmente definida por dois parâmetros: a média (μ) e a variância (σ^2).

Notação: seja X uma variável aleatória que segue uma distribuição normal com média μ e variância σ^2, $X \sim N(\mu, \sigma^2)$.

Normal Padrão (ou Gaussiana): é a distribuição normal com $\mu = 0$ e $\sigma^2 = 1$. A variável aleatória que segue uma distribuição normal padrão (ou gaussiana) é, usualmente, denotada por Z. Desta forma, $Z \sim N(0, 1)$.

Teorema: Se uma variável aleatória X é normalmente distribuída, com média μ e variância σ^2, isto é, $X \sim N(\mu, \sigma^2)$, então $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$.

OBS: com o teorema acima pode-se transformar qualquer variável aleatória normal em uma variável normal padrão. Vantagem: utilizar a tabela da normal (que serve apenas para a normal padrão) para qualquer variável normal.



Teorema: Sejam $X_1, X_2, \ldots, X_n$, n variáveis aleatórias independentes e normalmente distribuídas, sendo que $X_i \sim N(\mu_i, \sigma_i^2)$, para $i = 1, 2, \ldots, n$. Sejam $a_i \in \mathbb{R}$, números reais e Y uma combinação linear (ou soma com as X_i multiplicadas pelas a_i) dos X_i, isto é,

$Y = a_1X_1 + a_2X_2 + \cdots + a_nX_n$,

então a variável Y também é normalmente distribuída com média

$\mu_Y = a_1\mu_1 + a_2\mu_2 + \cdots + a_n\mu_n$

e variância

$\sigma_Y^2 = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2$.

Ou seja, $Y \sim N(\sum_{i=1}^{n} a_i\mu_i\,,\ \sum_{i=1}^{n} {a_i}^2{\sigma^2}_i)$.

Fazer os exemplos 79 até 81, páginas 204 até 208.

Teste de Hipótese

Definição: Uma hipótese estatística é uma afirmação ou conjectura acerca do(s) parâmetro(s) da distribuição de probabilidades de uma característica da população ou de uma variável aleatória. Em geral temos 2 hipóteses sendo testadas: H_0, hipótese nula que é a afirmação que fazemos; e H_1, hipótese alternativa que refuta a H_0.

Fazer os exemplos 82 até 84, páginas 223 até 224.

Definição: Um teste de uma hipótese estatística é um procedimento, ou regra de decisão, que nos possibilita decidir por rejeitar, ou não, a hipótese formulada, H_0, com base na informação obtida na amostra.

Fazer o exemplo 85, página 224.

Definição: A região crítica, RC, também chamada de região de rejeição, é o conjunto de valores assumidos pela variável aleatória ou estatística de teste para os quais a hipótese nula é rejeitada.

OBS: classificação do teste de hipótese quanto à região crítica:

(i) teste unilateral à esquerda, $RC = \{x \in \mathbb{R} | x < x_1\}$, onde x_1 é denominado valor crítico;

(ii) teste unilateral à direita, $RC = \{x \in \mathbb{R} | x > x_1\}$, onde x_1 é denominado valor crítico;

(iii) teste bilateral, $RC = \{x \in \mathbb{R} | x < x_1 \; ou \; x > x_2\}$, onde x_1 e x_2 são os valores críticos.

OBS: erros que podem ser cometidos em um teste de hipótese: erro tipo I e o erro tipo II.

Definição: O erro de tipo I consiste em rejeitarmos a hipótese nula, H_0, sendo essa verdadeira.

Definição: O erro de tipo II consiste em não rejeitarmos (aceitarmos) a hipótese nula, H_0, sendo essa falsa.

Decisão	Realidade	
	H_0 é verdadeira	H_0 é falsa
Rejeita H_0	Erro tipo I $P(erro\ tipo\ I) = \alpha$	Não há erro $P(acertar) = \alpha = 1 - \beta$
Não rejeita H_0	Não há erro $P(acertar) = \beta = 1 - \alpha$	Erro tipo II $P(erro\ tipo\ II) = \beta$

Definição: A probabilidade de cometermos o erro de tipo I é chamada de nível de significância, sendo denotada por α. Assim, $\alpha = P[erro\ tipo\ I] = P[rejeitar\ H_0 | H_0\ \text{é}\ verdadeira]$.

Definição: A probabilidade de cometermos o erro de tipo II é denotada por β. Assim, $\beta = P[erro\ tipo\ II] = P[não\ rejeitar\ H_0 | H_0\ \text{é}\ falsa]$.

OBS: o poder de um teste é a probabilidade de rejeitar H_0 quando ela é, de fato, falsa. É o caso de acertar com probabilidade $1 - \beta$.

Ver o exemplo na seção 1.2, página 226.

Procedimento (do teste de hipótese estatística):

1. Enunciar a hipótese H_0 a ser testada e a hipótese alternativa H_1;

2. Fixar o nível de significância do teste (α) e determinar a estatística de teste que será usada para testar a hipótese H_0;
3. Determinar a região crítica do teste (definindo se é unilateral a esquerda, à direita ou bilateral);
4. Determinar o valor da estatística de teste, utilizando os dados obtidos e a estatística de teste determinada em (2);
5. Conclusões (conclusão estatística e conclusão prática).

OBS: n é o tamanho da amostra.

Teste Z: testa se a média populacional, μ, de uma amostra aleatória $X_i, i = 1,2, \dots, n$, é igual a um valor fixado μ_0.

$H_0: \mu = \mu_0$

Onde $X \sim N(\mu, \sigma^2)$, sendo a variância, σ^2, *conhecida*. A estatística de teste é dada por

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

Isto é, Z segue uma normal padrão (gaussiana).

Fazer os exemplos 86 até 88, páginas 229 até 232.

Teste t para 1 média: testa se a média populacional, μ, de uma amostra aleatória $X_i, i = 1,2, \dots, n$, é igual a um valor fixado μ_0.

$H_0: \mu = \mu_0$

Onde $X \sim N(\mu, \sigma^2)$, sendo a variância, σ^2, *desconhecida*. A estatística de teste é dada por

$T = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \sim t(\nu)$.

Dizemos que T sege uma distribuição t de student com $\nu = n - 1$ graus de liberdade.

Fazer os exemplos 89 até 91, páginas 236 até 239.

Teste t para 2 médias: testa se as médias populacionais, μ_1 e μ_2, de 2 amostras aleatórias $X_{1i}, i = 1,2, \dots, n_1$, e $X_{2j}, j = 1,2, \dots, n_2$, são iguais.

$H_0: \mu_1 = \mu_2$

Onde $X_1 \sim N(\mu_1, {\sigma_1}^2)$ e $X_2 \sim N(\mu_2, {\sigma_2}^2)$, sendo as variâncias ${\sigma_1}^2 = {\sigma_2}^2 = \sigma^2$, *desconhecida*. A estatística de teste é dada por

$T = \frac{\bar{X_1} - \bar{X_2}}{\sqrt{S_c^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t(\nu)$.

Onde, $S_c^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$, é a variância estimada combinada das duas variâncias amostrais.

Dizemos que T sege uma distribuição t de student com $\nu = n_1 + n_2 - 2$ graus de liberdade.

Fazer os exemplos 92 até 94, páginas 242 até 246.

OBS: Tabela de Contingência é a tabela construída, em geral, para estudar a distribuição conjunta de duas variáveis qualitativas, mas também podem ser quantitativas. Considere duas variáveis X_1 e X_2 com k_1 e k_2 níveis, respectivamente.

X_1	X_2				
	Nível 1	Nível 2	...	Nível k_2	Total
Nível 1	n_{11}	n_{12}	...	n_{1k_2}	n_{1*}
Nível 2	n_{21}	n_{22}	...	n_{2k_2}	n_{2*}
...	...	...	...	...	...
Nível k_1	$n_{k_1 1}$	$n_{k_1 2}$	...	$n_{k_1 k_2}$	$n_{k_1 *}$
Total	n_{*1}	n_{*2}	...	n_{*k_2}	n

Onde,

n_{ij} representa o número de ocorrências do i-ésimo nível da variável X_1 e do j-ésimo nível da variável X_2;

$n_{i*} = \sum_{j=1}^{k_2} n_{ij}$;

$n_{*j} = \sum_{i=1}^{k_1} n_{ij}$;

$n = \sum_{j=1}^{k_2} n_{*j} = \sum_{i=1}^{k_1} n_{i*} = \sum_{j=1}^{k_2} \sum_{i=1}^{k_1} n_{ij}$.

Teste qui-quadrado para independência: testa se as variáveis X_1 e X_2 são independentes.

H_0: X_1 e X_2 são independentes

A estatística de teste é dada por

$\chi_{cal}^2 = \sum_{i=1}^{k_1} \sum_{j=1}^{k_2} \frac{\left(F_{o\,ij} - F_{e\,ij}\right)^2}{F_{e\,ij}} \sim \chi_\nu^2$, sendo $\nu = (k_1 - 1)(k_2 - 1)$, o grau de liberdade.

Sendo,

$F_{o\,ij}$ é a frequência observada;

$F_{e\,ij}$ é a frequência esperada e dada por $F_{e\,ij} = \frac{n_{i*} n_{*j}}{n}$.

Fazer os exemplos 95 até 97, páginas 250 até 259.

OBS: Tabelas utilizadas:

1. Tabela Normal Padrão, página 193;
2. Tabela t de Student, página 234;
3. Tabela qui-quadrada, página 249.