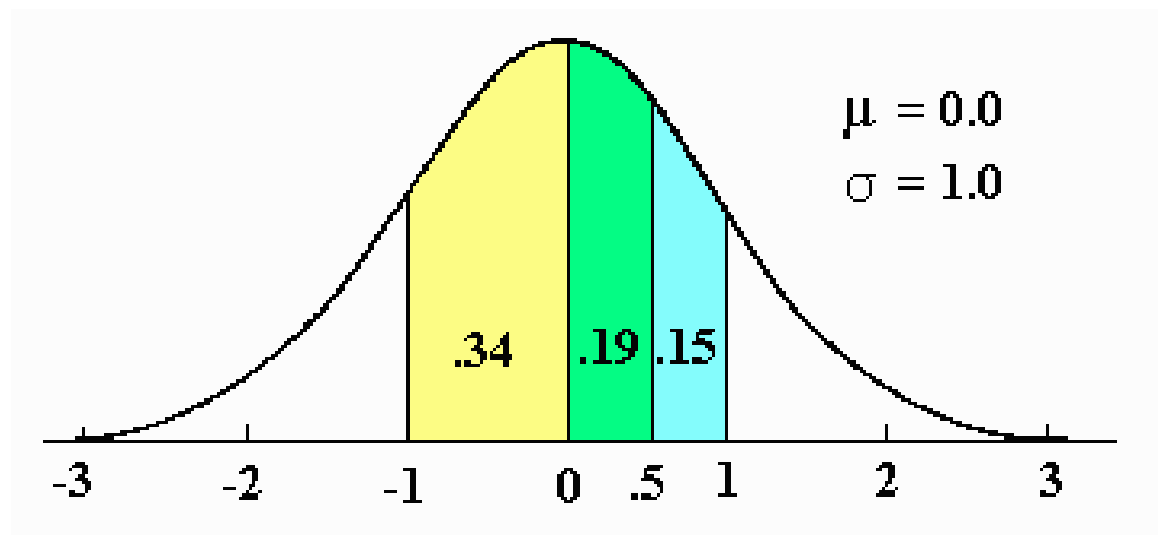


Distribuição Normal

Apontamentos para a disciplina de Estatística I





As distribuições Normais

- **Introdução: Curvas normais e distribuições normais**
- **A regra 68—95—99,7**
- **A distribuição normal padronizada (ou: padrão, estandardizada, reduzida)**
- **Cálculos com a distribuição normal**
 - **A tabela normal padronizada**
 - **Determinação de proporções normais**
 - **Determinação de um valor, dada uma proporção**
 - **Avaliação da normalidade**



As distribuições Normais

Apresentámos, até agora, uma ampla variedade de ferramentas **gráficas e numéricas** para descrever distribuições:

- instrumentos clássicos
 - Histogramas
 - polígonos de frequências
- instrumentos de análise exploratória de dados
 - diagramas de ramo-e-folhas
 - gráficos de extremos e quartis
- resumos numéricos
 - tendência central
 - variabilidade
 - assimetria e achatamento (curtose)



As distribuições Normais

Dispomos de uma estratégia clara para explorar dados de uma variável quantitativa individual, por exemplo:

- Inicie o trabalho com um gráfico, habitualmente recomenda-se um **ramo-e-folhas** ou um **histograma**
- Procure o padrão geral e os desvios acentuados a esse padrão, como por exemplo **valores discrepantes** (ouliers)
- Escolha um resumo numérico para dar uma breve descrição do centro e da dispersão

Agora vamos acrescentar mais uma etapa a esta estratégia:

- Por vezes o padrão geral de um grande número de observações é tão regular que podemos descrevê-lo por uma curva suave



As distribuições Normais

Curvas de densidade

Uma curva de densidade é uma curva que apresenta as seguintes propriedades:

- está sempre sobre (acima) o eixo horizontal
- tem uma área exactamente igual a 1 abaixo dela

Assim, uma curva de densidade descreve o padrão geral de uma distribuição. A área sob a curva e acima de qualquer intervalo de valores é a proporção de todas as observações que se enquadram naquele intervalo.



As distribuições Normais

A mediana e a média de uma curva de densidade

As medidas de centro e de dispersão discutidas anteriormente aplicam-se tanto a curvas de densidade como a conjuntos reais de observações.

A **mediana** de uma curva de densidade é o ponto de áreas iguais, o ponto que divide ao meio a área sob a curva.

A **média** de uma curva de densidade é o ponto de equilíbrio, no qual a curva se equilibraria se fosse feita de material sólido.

Para uma curva de densidade simétrica, a mediana e a média coincidem. Ambas estão no centro da curva.



As distribuições Normais

Distribuições normais

- Existe uma classe particularmente importante de curvas de densidade. Essas curvas de densidade são simétricas, têm um único pico e apresentam a forma de sino. Chamam-se *curvas normais* e descrevem **distribuições normais**.
- Todas as distribuições normais têm a mesma forma global
 - A curva de densidade exacta para uma distribuição normal particular é caracterizada pela sua média μ e seu desvio padrão σ .
 - A média está localizada no centro da curva simétrica, e coincide com a mediana (como vimos acima).



As distribuições Normais

- Alterando-se μ e deixando σ inalterado, deslocamos a curva normal ao longo do eixo horizontal, sem modificar a sua dispersão.
- o desvio padrão controla a dispersão de uma curva normal.
- O desvio padrão de uma curva normal pode ser determinado visualmente. Eis como proceder. Ao afastarmos-nos do centro μ numa ou noutra direcção, a curva passa de um decréscimo muito acentuado para um decréscimo cada vez menos acentuado. Os pontos onde ocorre essa mudança de curvatura (de convexa a côncava) estão situados à distância σ da média μ .



As distribuições Normais

A regra 68—95—99.7

Embora hajam muitas curvas normais, todas elas têm propriedades comuns. Em particular, todas as distribuições normais têm as propriedades descritas pela regra seguinte:

Na distribuição normal com média μ e desvio padrão σ :

- 68% das observações estão a menos de $\pm 1\sigma$ da média μ .
- 95% das observações estão a menos de $\pm 2\sigma$ de μ .
- 99.7% das observações estão a menos de $\pm 3\sigma$ de μ .

Na verdade os valores exactos são respectivamente:

68.27% ($\pm 1\sigma$); 95.45% ($\pm 2\sigma$), e 99.73% ($\pm 3\sigma$)



As distribuições Normais

- ❖ A Curva Normal é algo mais do que uma fórmula matemática. O seu principal interesse para os cientistas provém desta ser também um fenómeno natural, uma vez que é frequente encontrar variáveis com distribuições muito semelhantes à Normal.

Por exemplo:

Nas Ciências Naturais: Peso, altura, acuidade visual, força.

Na Psicologia: Quociente intelectual (QI), Extroversão, Raciocínio espacial (S).

- ❖ A representação gráfica da curva (ver próximo slide) permite apreciar a razão da sua universalidade.

As distribuições Normais

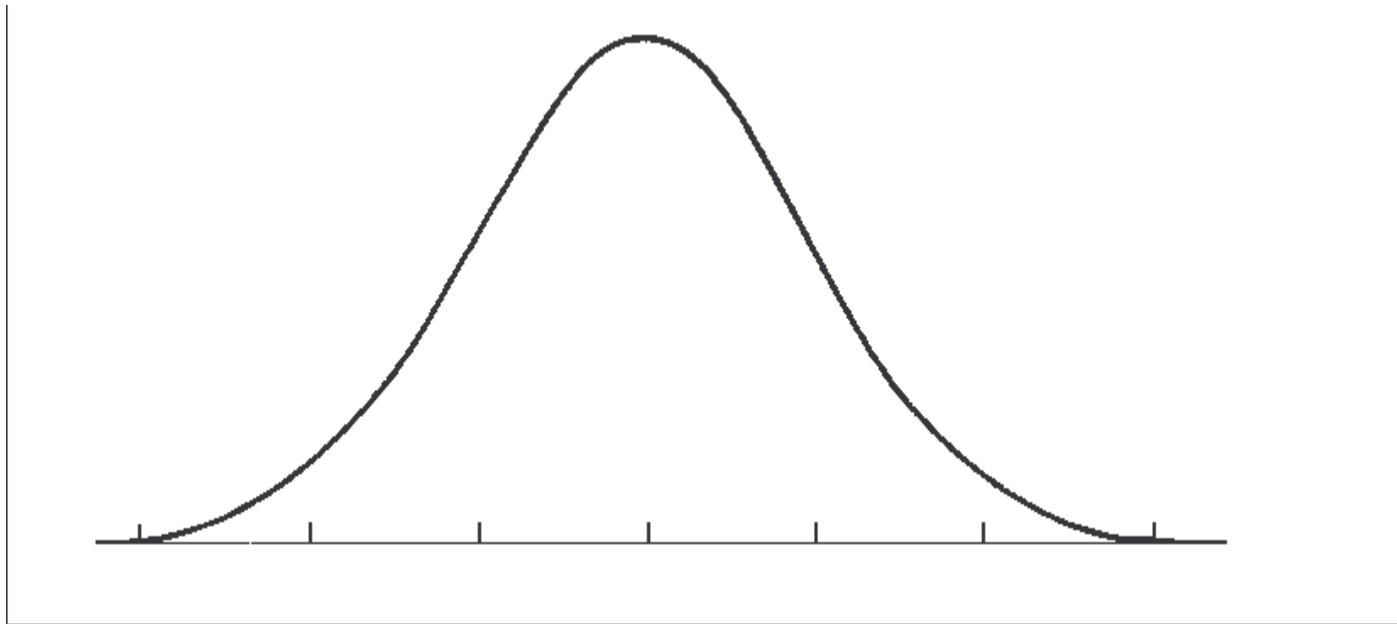


Figura 1 A distribuição normal ou curva normal

❖ Na maior parte das variáveis existe um valor central (a média) em torno do qual se situam a maioria dos indivíduos, e à medida que vamos tomando valores mais afastados da média observamos que estes são menos frequentes.



As distribuições Normais

- ❖ Matematicamente, uma variável aleatória distribui-se segundo um modelo normal, com parâmetros μ e σ , se a sua função de densidade de probabilidade para qualquer valor de X vem dada por:

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$

Onde $\pi = 3.1416\dots$ e $e = 2.718\dots$. Pode representar-se compactamente tudo isto da seguinte forma:

$$X \longrightarrow N(\mu, \sigma)$$



As distribuições Normais

❖ A Distribuição Normal Padronizada ou Reduzida

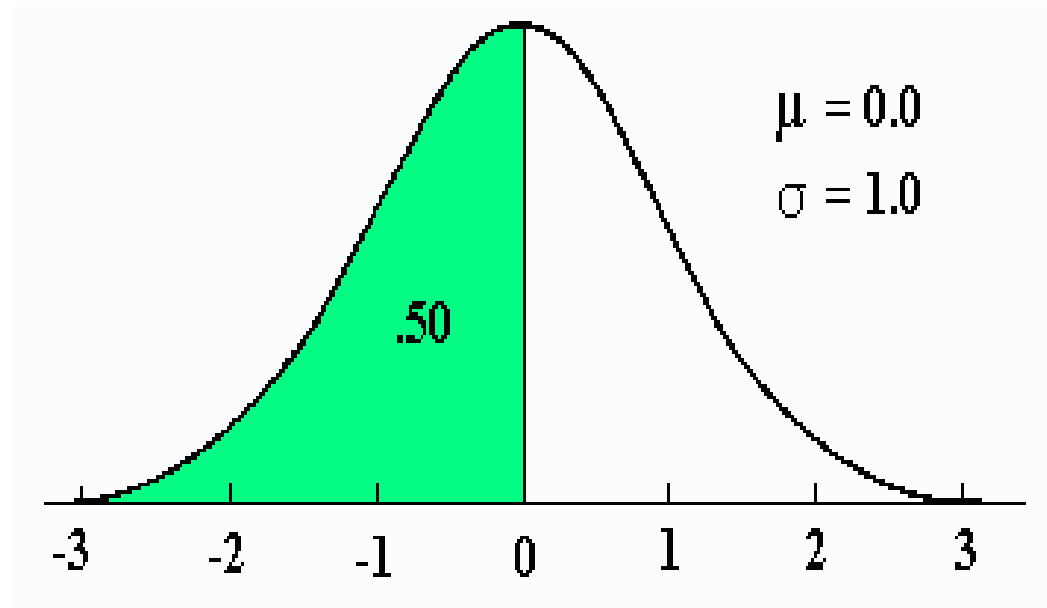
No expoente da fórmula anterior (descoberta por De Moivre) podemos reconhecer a fórmula empregue para obter pontuações típicas (ou estandardizadas), e portanto para variáveis tipificadas a função de densidade de probabilidade pode ser simplificada, dado que o desvio padrão é 1 e a média é 0.

Se uma variável X tem distribuição normal arbitrária $N(\mu, \sigma)$ com média μ e desvio padrão σ , então a variável reduzida

$$Z = \frac{X - \mu}{\sigma}$$

tem, igualmente, distribuição normal.

As distribuições Normais – A distribuição normal reduzida



$$Y = \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{z^2}{2}\right)}$$



As distribuições Normais

A maior parte do trabalho prático com variáveis aleatórias consiste em descobrir probabilidades associadas e valores.

Isso significaria integrar a função de densidade entre os valores de interesse.

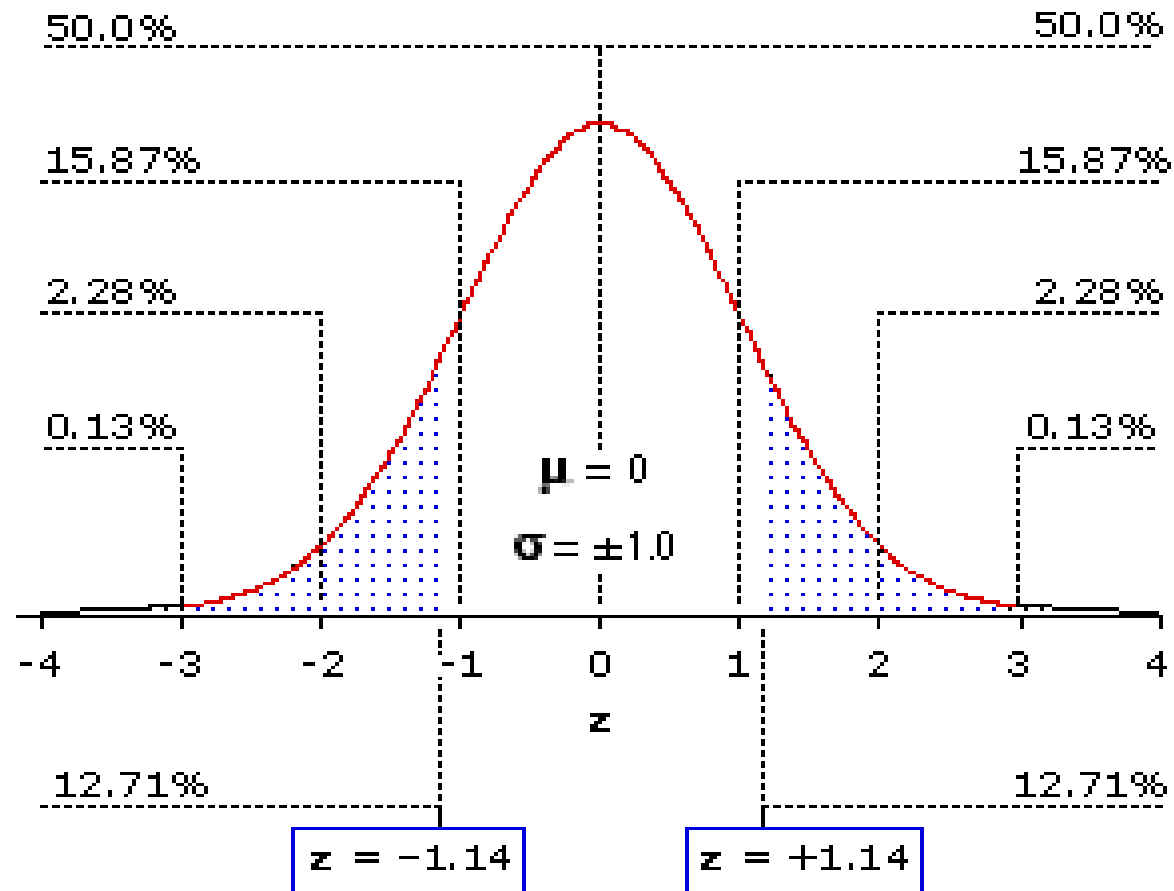
Para evitar ter que resolver este tipo de operações construíram-se tabelas apropriadas com as áreas já calculadas e cujo uso se baseia na aplicação no teorema de tipificação.

Segundo este teorema, a função de distribuição associada a um valor de uma variável aleatória, X , com distribuição normal, é a mesma que a função de distribuição tipificada, desse valor, na normal reduzida.

Por isso apenas foram construídas tabelas para a distribuição padronizada ou reduzida.

As distribuições Normais

A Distribuição Normal Reduzida





As distribuições Normais

De onde provêm os valores das áreas (percentagens) registados no gráfico precedente?

Por exemplo:

a) $P(z \geq +1.14) = 12.71\%$

b) $P(z \geq +2.00) = 2.28\%$

c) $P(z \leq -3.00) = 0.13\%$

Vamos utilizar uma das calculadoras web incluída nos sites:

<http://faculty.vassar.edu/lowry/VassarStats.html> (in Utilities --> Statistical Tables Calculator)

ou então:

http://davidmlane.com/hyperstat/normal_distribution.html



As distribuições Normais

Exemplos de problemas relacionados com a obtenção de probabilidades associadas a variáveis normais

✓ Suponha que a variável X segue uma distribuição $N(50,8)$, e queremos obter as seguintes probabilidades:

- a) Observar um valor quando muito (i.e., que no máximo) seja igual a 56
- b) Observar um valor que no mínimo seja igual a 52,8
- c) Observar um valor compreendido entre 40,8 e 48,3.



As distribuições Normais

Observar um valor quando muito (i.e., que no máximo) seja igual a 56?

Resolução:

- a) No primeiro caso trata-se de obter a probabilidade acumulada do valor 56, e para tanto basta estandardizar esse valor e procurar, na tabela da distribuição normal reduzida, a proporção da área que se encontra à sua esquerda. Ou seja:

$$P(X \leq 56) = P(z \leq (56-50)/8) = P(z \leq 0,75) = 0,7734.$$



As distribuições Normais

b) No segundo caso trata-se de obter a probabilidade que corresponde à área extrema da distribuição normal unitária acima de 52,8. Uma vez que a tabela da distribuição normal que utilizamos nos dá essa área directamente, basta que transformemos o valor de X numa pontuação tipificada (z). Assim:

$$P(X \geq 52,8) = P(z \geq (52,8 - 50)/8) = P(z \geq 0,35) = 0,3632.$$



As distribuições Normais

c) No terceiro caso trata-se de obter a área, da curva normal, limitada pelos valores 40,8 e 48,3. Trata-se de encontrar a diferença entre a probabilidade abaixo de 48,3 da que se encontra abaixo do valor 40,8 (Dada a forma da tabela que utilizamos, temos que ter em conta o conceito de simetria subjacente à distribuição normal). Assim:

$$\begin{aligned} P(40,8 \leq X \leq 48,3) &= \\ &= P((40,8-50)/8 \leq z \leq (48,3-50)/8) \\ &= P(-1,15 \leq z \leq -0,21) = 0,2917. \end{aligned}$$



As distribuições Normais

Exemplos de problemas relacionados com a obtenção de pontuações de uma variável normal com probabilidades concretas associadas [Neste caso usamos a fórmula: $X = z \sigma + \mu$].

✓ Suponha que, continuamos interessados em estudar a variável X , que segue uma distribuição $N(50,8)$, e queremos obter os valores de esta variável (X) para os quais se que cumprem as seguintes condições:

- a) Aquele para o qual a probabilidade de observar um valor quando muito igual a ele é 0,1736
- b) Aquele para o qual a probabilidade de observar um valor que no mínimo seja igual a ele seja 0,9207
- c) Os dois valores que incluam 50 % dos valores centrais.



As distribuições Normais

- a) Aquele para o qual a probabilidade de observar um valor quando muito igual a ele é 0,1736

Resolução:

- a) No primeiro caso trata-se de obter a probabilidade acumulada do valor (que deixa à sua esquerda) a área de 0,1736. Recorrendo ao procedimento de conversão na variável reduzida e socorrendo-se da tabela normal unitária, comprovará que se trata do valor $z = -0,94$. Agora basta reconverter este valor usando a média e o desvio padrão da distribuição:
- $$z_{0,1736} = -0,94 = ((X-50)/8), \text{ ou seja,}$$
- $$X = -0,94 * 8 + 50 = 42,48.$$



As distribuições Normais

b) No segundo caso trata-se de obter o valor de X que deixa uma área à sua direita de 0,9207. Esse valor pode ser obtido da tabela da distribuição unitária que utilizamos neste curso (Tabelas da Universidade de Vassar, EUA), subtraindo $1-0,9207=0,0793$. O valor que deixa à sua esquerda 0,0793 da área da distribuição Normal é $z = -1,41$. Então, convertendo para $N(50,8)$, teremos:

$$X = -1,41 * 8 + 50 = 38,72.$$



As distribuições Normais

- c) Trata-se de obter aquelas duas pontuações que deixam à sua esquerda e direita, respectivamente, áreas iguais a 0,25. Segundo a Tabela da distribuição normal reduzida essas pontuações teriam como valores típicos (ou reduzidos), respectivamente, -0,67 e 0,67. Reconvertendo esses valores, teremos:

$$z = -0,67 = ((X-50)/8), \text{ ou seja, } X = -0,67*8+50 \\ =44,64$$

e

$$z = 0,67 = ((X-50)/8), \text{ ou seja, } X = 0,67*8+50 = 55,36$$



As distribuições Normais

▪ Avaliação da normalidade

- **Como podemos julgar se os dados são aproximadamente normais?**
- **Diagramas de ramo-e-folhas e histogramas podem revelar as características distintamente não-normais de uma distribuição (outliers, assimetria acentuada, lacunas ou aglomerados)**
- **Poderá, ainda, usar a seguinte estratégia: marque os pontos média, média \pm desvio padrão, média ± 2 desvios padrão, no eixo dos X. O que nos dá a escala natural para a distribuição normal. Compara-se depois a contagem das observações em cada intervalo com a regra 68—95—99,7.**

NB: Conjuntos menores de dados raramente se adaptam à regra 68—95—99,7 de uma forma perfeita. Isto é verdadeiro mesmo para observações extraídas de uma população maior que tenha realmente uma distribuição normal!