Estatística e Probabilidade

Farlei Roberto Mazzarioli
www.farlei.net

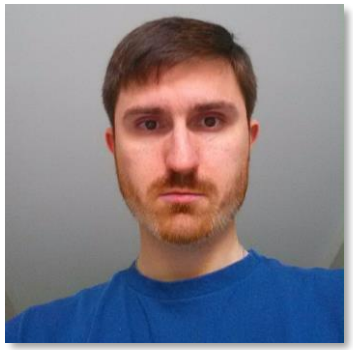
2021

Sumário

1. Conceitos básicos de estatística
2. Organização dos dados
3. Medidas de tendência central
4. Medidas de dispersão
5. Conceitos básicos de probabilidade
6. Distribuições binomial e de Poisson
7. Correlação linear e regressão linear
8. Distribuição normal
9. Estimativa de parâmetros
10. Teste de hipótese

Farlei Roberto Mazzarioli

Universidade Federal de São Carlos (UFSCar): Licenciatura Plena em Física; Especialização em Ensino de Filosofia.
Universidade Federal de Lavras (UFLA): Especialização em Informática em Educação.
Centro Universitário Claretiano: Licenciatura em Filosofia; Bacharel em Teologia.



Principais referências e recomendação para aprofundar os estudos:

“Estatística e Probabilidade” do autor Ricardo Vida. “Estatística” do autor José Carlos de Araújo Porto (Rochedo). “Estatística Descritiva” e “Estatística Indutiva” dos autores Mirtes Vitória Mariano; Christiane Mazur Lauricella; Alexandre Daliberto Frugoli.

Espaço reservado para vídeo do professor em apresentação online ou gravação de aula.

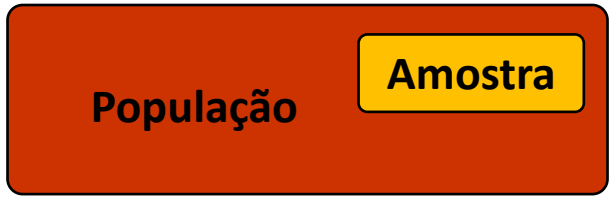
1. Conceitos básicos de estatística

A palavra “estatística” vem de *Status*, em latim “Estado”, e se referindo aos dados da administração do Estado, para saber número de soldados, estoque de alimentos, homens capazes de lutar e outros dados que permitam tomar decisões diante de um inimigo, ou seja, se negociar rendição, negociar paz, ir para a luta ou ir para a luta xingando a mãe do inimigo. A palavra *Statistics* surgiu em 1797 na Enciclopédia Britânica. Hoje “estatística” é um método científico que permite coletar dados, organizar dados, analisar dados e tomar decisões em condições de incertezas. Também existe a opção: liberdade ou φωδα-ση!

“Isso é Esparta!”, cena do filme *300* (2006).

População e amostra

População é o conjunto de elementos da questão de interesse e uma amostra representativa é o subconjunto que preserva as característica dessa população em escala menor. Normalmente não é economicamente viável coletar dados de toda a população e há condições em que isto seria desastroso, tal como um exame médico que precise retirar uma amostra de um órgão de um paciente. É preciso entender a natureza da amostra, ou seja, as variáveis estatísticas que são qualitativas (nominais ou ordinais) ou quantitativas (discretas ou contínuas).



Variáveis qualitativas:
Nominais (grandezas em paralelo): sexo, cor, raça, espécie, nação.
Ordinais (grandezas em série): níveis de escolaridade, faixa etária.

Variáveis quantitativas:
Discretas (valores inteiros): número de pessoas, número de defeitos.
Contínuas (valores fracionáveis): altura de pessoas, produção de café.

Técnicas de amostragem

Agora que você sabe o que é uma amostra e qual é a variável que será coletada, chega a vez de escolher a forma de coletar essa amostra, ou seja, a técnica de amostragem. A seguir temos os três tipos mais comuns que lhe permitirá formar a ideia do que é amostragem.

- **Amostragem aleatória:** os elementos de amostra são sorteados de forma simples dentro de todo o conjunto.

- **Amostragem sistemática:** os elementos já estão ordenados (linha de montagem, casas na rua, pessoas na fila) e se escolhe um intervalo em que se retirará as amostras (1 a cada 10, 1 a cada 200).

- **Amostragem estratificada:** os elementos estão divididos em estratos (subconjuntos) de forma que esta proporção deve ser levada em consideração, tal como sexo masculino e feminino, faixa de renda, nível de escolaridade, densidade populacional de um país.

Exemplo 1.1. Uma pesquisa é feita sobre as reprovações no curso de Ciência da Computação em uma determinada universidade, sorteando 10 alunos de cada sala de todos os períodos do curso e em todos os *campi* da universidade. Estes alunos deverão responder um questionário com 50 perguntas de alternativas que serão digitalizadas e organizadas em diversas tabelas por computador. Analise as afirmativas a seguir:

I. Seria mais simples e preciso fazer a amostragem aleatória dos alunos juntos, independente de sala e período.
II. A amostragem utilizada é estratificada aleatória, já que é aleatória em cada estrato que são as salas e períodos.
III. O histograma dos dados deve apresentar o número de alunos na coordenada y pela frequência na coordenada x.
IV. O número de alunos de cada sala e período, em que é estratificado, deve ser levado em consideração.

São verdadeiras apenas as afirmativas

a) I e II.
b) I e III.
c) II e III.
d) II e IV.
e) III e IV.

Exemplo 1.2. As variáveis estatísticas podem ser classificadas em qualitativas ou quantitativas, entre as qualitativas temos nominais que não definem ordem entre si e as ordinais que possuem ordem entre si, nas quantitativas temos as discretas que só admitem números inteiros e as contínuas que podem ser fracionadas. Analise as afirmativas a seguir:

I. São variáveis qualitativas nominais: sexo, cor, raça, religião, sabor de sorvete, gosto musical.
II. São variáveis qualitativas ordinais: grau de escolaridade, faixa etária, classe econômica.
III. São variáveis quantitativas discretas: número de pessoas, número de defeitos em um aparelho.
IV. São variáveis quantitativas contínuas: peso das pessoas, produção de minério, nota em provas.

São verdadeiras apenas as afirmativas

a) I, II, III e IV.
b) I, II e III.
c) I, II e IV.
d) II, III e IV.
e) II e III.

Exemplo 1.3. Uma pesquisa é feita com 4.000 alunos ao final do colegial de uma grande cidade, ela demonstra que 30% deles ficaram traumatizados com as aulas de Física, ou seja, carregam sequelas mensuráveis. Analise as afirmativas a seguir:

I. A população nesse caso são os alunos ao final do colegial.
II. A amostra corresponde aos 4.000 alunos da grande cidade.
III. A variável é ficar traumatizado com as aulas de Física.

É correto apenas o que se afirma em

a) I.
b) II.
c) III.
d) I e II.
e) I, II e III.

2. Organização dos dados

Como você vai organizar os dados coletados para poder interpretá-los? Os dados brutos devem ser dispostos em ordem crescente, tabelados na quantidade (ou intervalo de classe) em que se repetem, para então gerar um gráfico que permita uma visualização simples de todo o conjunto. Qual o nome de cada elemento disso?

Dados brutos: são os resultados da amostra recém coletados e registrados na sequência da coleta.

Rol: quando os dados brutos são organizados em ordem crescente se tornam o rol (lembre-se do “rol da fama”).

Tabela de frequência: depois esses dados são lançados em uma tabela em que é associado o valor e o número de vezes que ele aparece.

Histograma: chega o momento de “pôr no gráfico” os dados da tabela de frequência para que sejam melhor visualizados.

Como montar um histograma?

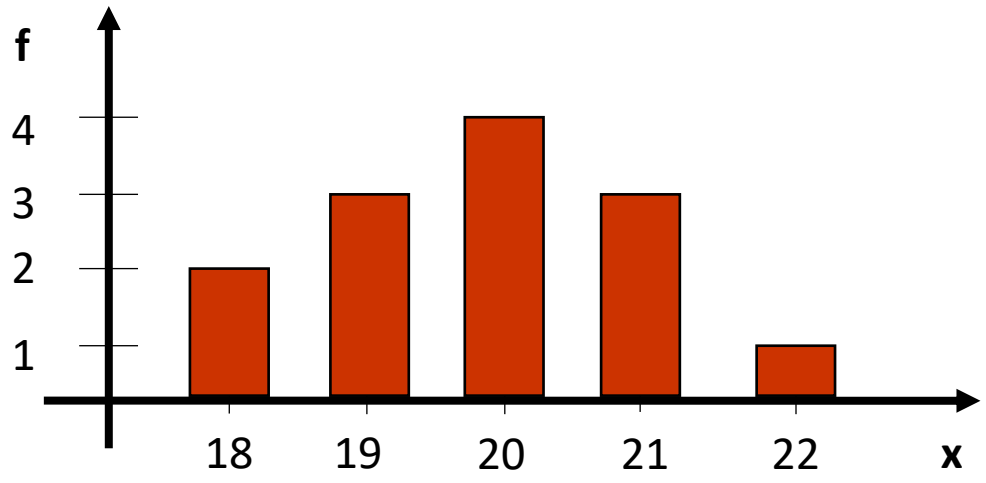
Considere a seguinte amostra da idade de alunos cujos dados brutos sejam 19, 18, 19, 20, 22, 21, 19, 20, 21, 18, 21, 20, 20. Colocando estes números em ordem crescente teremos o rol com que poderá ser montada a tabela e logo depois o histograma, todos a seguir.

Dados brutos: 19, 18, 19, 20, 22, 21, 19, 20, 21, 18, 21, 20, 20.
Rol: 18, 18, 19, 19, 19, 20, 20, 20, 20, 21, 21, 21, 22.

Tabela de frequência:

x	f
18	2
19	3
20	4
21	3
22	1
Σ	13

Histograma:



Distribuição de frequência

Esta tabela *muito chata* trabalha com a frequência relativa (f dividido por Σf) e a frequência acumulada (soma dela com as anteriores). Na tabela temos x = grandeza; f = frequência da grandeza; fr = frequência relativa; fr% = frequência relativa em porcentagem; fa = frequência acumulada; fra = frequência relativa acumulada; fra% = frequência relativa acumulada em porcentagem.

x	f	fr	fr%	fa	fra	fra%
18	2	0,154	15,4	2	0,154	15,4
19	3	0,231	23,1	5	0,385	38,5
20	4	0,308	30,8	9	0,692	69,2
21	3	0,231	23,1	12	0,923	92,3
22	1	0,077	7,7	13	1,000	100,0
Σ	13	1,000	100,0			

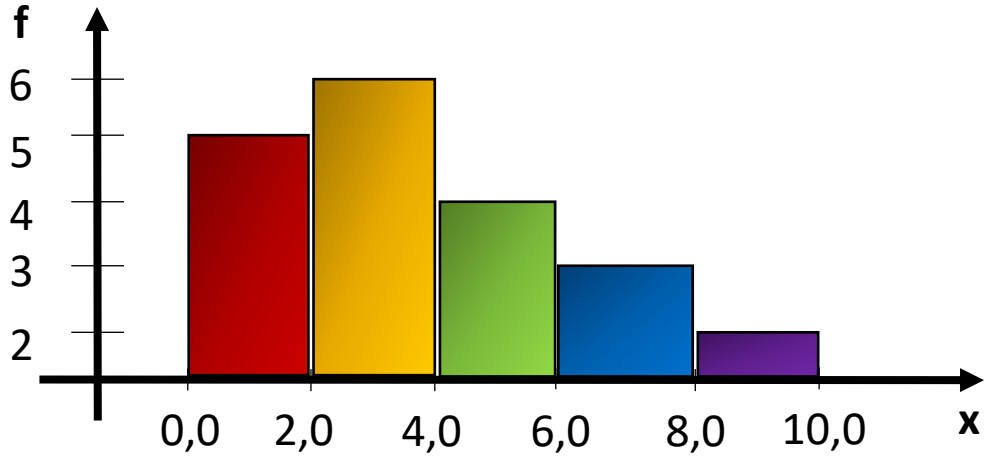
$$fr = \frac{f}{\Sigma f} \qquad fra = \frac{fa}{\Sigma f} \qquad fr\% = fr \cdot 100\% \qquad fra\% = fra \cdot 100\%$$

Intervalos de classe

Considere que a grandeza medida “x” seja dada em um intervalo de classe, como na tabela e no histograma a seguir. No exemplo, considere que as amostras de notas “x” de fechado em 0,0 até infinitamente perto de 2,0 com o símbolo “ l— ”. Dessa forma as barras do histograma são unidas já que os dados são contínuos.

Notas de 0 a 10 de uma turma de 20 alunos com um professor muito malvado.

x	f
0,0 l— 2,0	5
2,0 l— 4,0	6
4,0 l— 6,0	4
6,0 l— 8,0	3
8,0 l— 10,0	2
Σ	20



Como montar uma tabela dessas?

No slide anterior nós vimos as notas de uma turma de 20 alunos de um professor muito malvado e a sua tabela pronta e organizada em intervalos de classe. Para montar essa tabela é preciso dos valores das notas dispostos em rol, tendo o número de elementos calcula-se o número de classes, dos limites máximo e mínimo descobre-se a amplitude amostral e assim a amplitude de classe.

Rol: 0,0; 0,0; 0,5; 1,0; 1,0; 1,5; 2,0; 2,0; 2,5; 3,0; 3,5; 3,5; 4,0; 4,5; 5,0; 5,5; 6,5; 7,0; 7,5; 8,5; 9,0.

Número de elementos: n = 20

Número de classes: $n_c = \sqrt{n} = \sqrt{20} = 4{,}47 \cong 5$ (arredondar para cima) **n_c = 5**

Amplitude amostral: AA = limite máximo – limite mínimo
AA = 9,0 – 0,0 = 9,0

Amplitude de classe: $a = \frac{AA}{n_c} = \frac{9{,}0}{5} = 1{,}8 \cong 2$ (arredondar para cima) **a = 2**

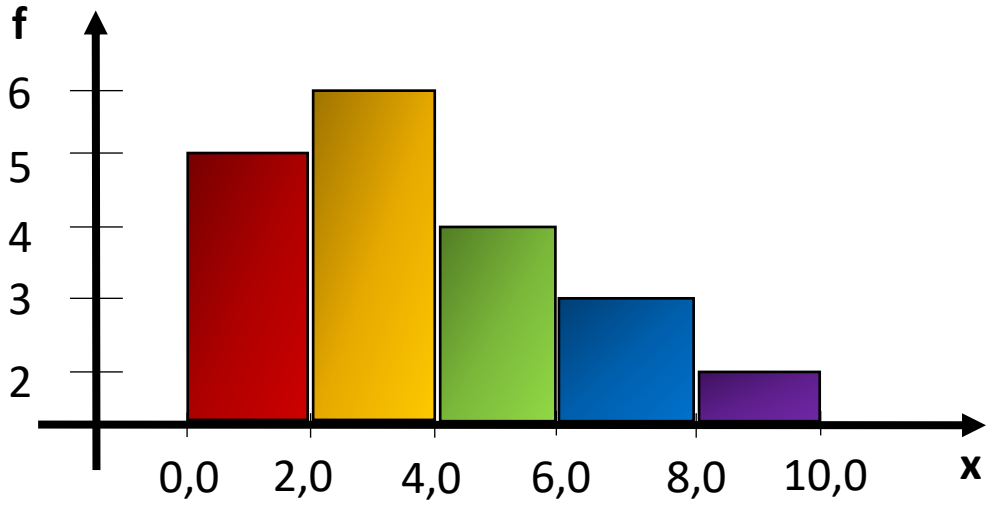
Confirmando os dados

O número de classes foi 5 e a amplitude de classe 2, então a tabela foi dividida nos 5 intervalos de 2 em 2 nesta forma: 0,0 l— 2,0; 2,0 l— 4,0; 4,0 l— 6,0; 6,0 l— 8,0; 8,0 l— 10,0.

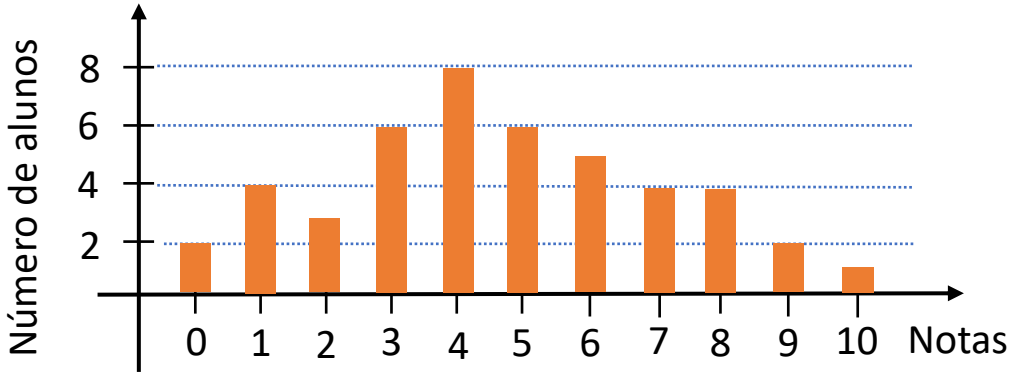
a = 2

n_c = 5

x	f
0,0 l— 2,0	5
2,0 l— 4,0	6
4,0 l— 6,0	4
6,0 l— 8,0	3
8,0 l— 10,0	2
Σ	20



Exemplo 2.1. Um histograma é um gráfico de barras que apresenta a variável em questão no eixo-x e a frequência no eixo-y, tal como a figura a seguir que apresenta as notas pelo número de alunos em uma sala de 45 alunos.



Com base no gráfico analise as afirmações a seguir:

I. Apenas um aluno tirou nota 10 e 51,1 % dos alunos tiraram nota 4 ou menor.
II. Apenas dois alunos tiram nota zero e 11 alunos tiraram nota 7 ou maior.
III. O grupo que possui notas entre 3 e 6 é composto de 22 alunos e é 55,5%.
IV. Se a nota para ficar feliz é tirar 7 ou mais, então 20 % deles estão felizes.

São verdadeiras apenas as afirmativas a) I e II. b) II e III. c) I e IV. d) II e IV. e) I e III.

I. Um aluno tirou nota 10 e (2 + 4 + 3 + 6 + 8)/45 = 0,511 = 51,1 %
II. Dois alunos tiraram nota 0 e para notas ≥ 7 temos 4 + 4 + 2 + 1 = 11.
III. 6 + 8 + 6 + 5 = 25 25/45 = 0,555 = 55,5 %
IV. Notas ≥ 7 são 4 + 4 + 2 + 1 = 11, logo 11/45 = 0,244 = 24,4 %

Exemplo 2.2. A organização dos dados estatísticos permite clara visualização destes dados, logo oferecendo fácil interpretação e maior assertividade na tomada de decisões com base neles.

Analise as afirmativas a seguir:

I. Um histograma é formado de retângulos com a base no intervalo de classe e frequência.
II. Definição de histograma é um gráfico cartesiano do rol em x pelo intervalo de classe em y.
III. Um histograma é péssimo instrumento para visualizar a moda da amostra.
IV. Uma tabela de frequências sem classe permite a fácil recuperação do rol dos seus dados.
V. Todo histograma é exclusivo para a determinação de variáveis qualitativas.

São verdadeiras apenas as afirmativas
a) I e II.
b) II e V.
c) I e IV.
d) I, II e III.
e) I, II e V.

3. Medidas de tendência central

As medidas de posição central são: média, moda e mediana. A média pode ser simples ou ponderada, a simples é a somatória de todos os elementos x dividido pelo número total de elementos n, enquanto na ponderada cada x é multiplicado pela frequência f. A moda é o valor ou elemento que aparece com maior frequência. A mediana é o valor que se encontra no meio do conjunto ordenado em ordem crescente, e se o número de elementos não for ímpar é só fazer a média entre os dois números beirando o meio do conjunto.

Média

simples

$$\bar{x} = \frac{\sum x}{n}$$

ponderada

$$\bar{x} = \frac{\sum x \cdot f}{\sum f}$$

Moda

Ex.: 5; 6; 6; 7; 7; 7 e 8.
A moda é 7.

Ex.: 4; 5; 6; 6; 7; 7; e 8.
As modas são 6 e 7.

Ex.: α λ π η π μ π λ
A moda é π

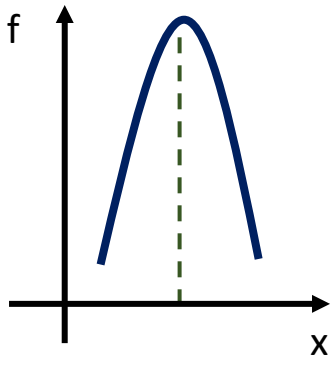
Mediana

Ex.: 5; 6; 6; 7; 7; 7 e 8.
A mediana é 7.

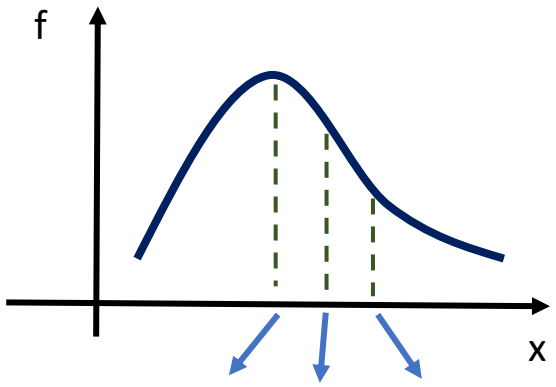
Ex.: 4; 5; 6; 6; 7; 7; 7 e 8.
A mediana é (6 + 7)/2 = 6,5.

Relação gráfica entre estes valores

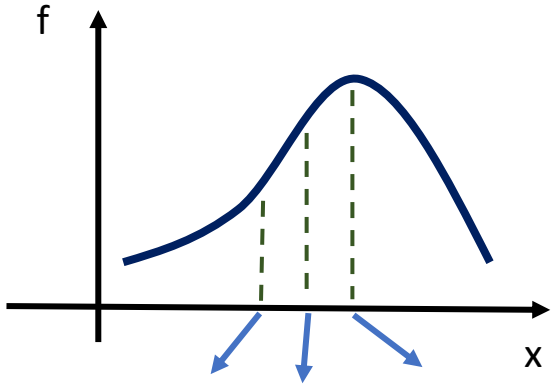
Quando fazemos um histograma, ou seja, gráfico (x, f), nós podemos ver a curva centralizada se média = moda = mediana, ou veremos essa curva deslocada para a esquerda se moda < mediana < média ou então a veremos deslocada para a direita se média < mediana < moda.



Média = Moda = Mediana



Moda Mediana Média



Média Mediana Moda

Cálculo da média sem classe

Considere como exemplo a idade em anos dos alunos: 18, 18, 19, 19, 19, 20, 20, 20, 20, 21, 21, 21, 22. Podemos ver que a quantidade total é n = 13. Podemos calcular a média de forma simples ao somar todos os valores ou usando a frequência com a média ponderada que economiza repetições dos números. Eu sei que isso é muito fácil, uma ofensa para o seu intelecto superior. Mas tenha paciência, em breve ficará mais difícil com choro e ranger de dentes como é o esperado para este curso terrível.

Média simples

$$\bar{x} = \frac{\sum x}{n} = \frac{18+18+19+19+19+20+20+20+20+21+21+21+22}{13} = \frac{258}{13} = 19{,}85$$

Média com frequência

$$\bar{x} = \frac{\sum x \cdot f}{\sum f} = \frac{18 \cdot 2 + 19 \cdot 3 + 20 \cdot 4 + 21 \cdot 3 + 22 \cdot 1}{2 + 3 + 4 + 3 + 1} = \frac{258}{13} = 19{,}85$$

Cálculo da média com classe

Considere aquelas notas de 20 alunos daquele professor malvado na tabela a seguir. O valor de “x” para calcular a média da classe é usando o ponto médio PM = (limite superior + limite inferior)/2.

x	f
0,0 ⊢ 2,0	5
2,0 ⊢ 4,0	6
4,0 ⊢ 6,0	4
6,0 ⊢ 8,0	3
8,0 ⊢ 10,0	2
Σ	20

x	PM	f	PM.f
0,0 ⊢ 2,0	1,0	5	5,0
2,0 ⊢ 4,0	3,0	6	18,0
4,0 ⊢ 6,0	5,0	4	20,0
6,0 ⊢ 8,0	7,0	3	21,0
8,0 ⊢ 10,0	9,0	2	18,0
Σ		20	82

$$\bar{x} = \frac{\sum PM \cdot f}{\sum f} = \frac{1{,}0 \cdot 5 + 3{,}0 \cdot 6 + 5{,}0 \cdot 4 + 7{,}0 \cdot 3 + 9{,}0 \cdot 2}{5 + 6 + 4 + 3 + 2} = \boxed{\frac{82}{20}} = 4{,}1$$

Atalho

Exemplo 3.1. As notas na Universidade Paulista (UNIP), a maior universidade do Brasil, correspondem a uma média simples entre duas provas e correspondentes práticas, se necessárias, gerando as notas NP1 e NP2, dessas se faz a média M = (NP1 + NP2)/2, que se for a partir de 7,0 o aluno estará aprovado. Como segunda chance existe um exame Ex que pode reconfigurar a média final MF na forma MF = (M + Ex)/2, que se for a partir de 5,0 o aluno será aprovado. O sistema arredonda automaticamente em até 0,25 a nota para aprovado na primeira média M e na média final MF. Analise as afirmações a seguir:

I. Um aluno que tire 0,0 nas duas provas (NP1 e NP2) pode ser aprovado se conseguir 9,5 no exame.
II. Um aluno pode ser aprovado se tirar 5,0 em todas as provas, ou seja, na NP1, na NP2 e no Exame.
III. Um aluno será aprovado se tirar 8,0 na NP1, 2,0 na NP2 e 4,0 no Exame.
IV. Um aluno será aprovado se tirar 4,0 na NP1, 2,0 na NP2 e 7,0 no Exame.

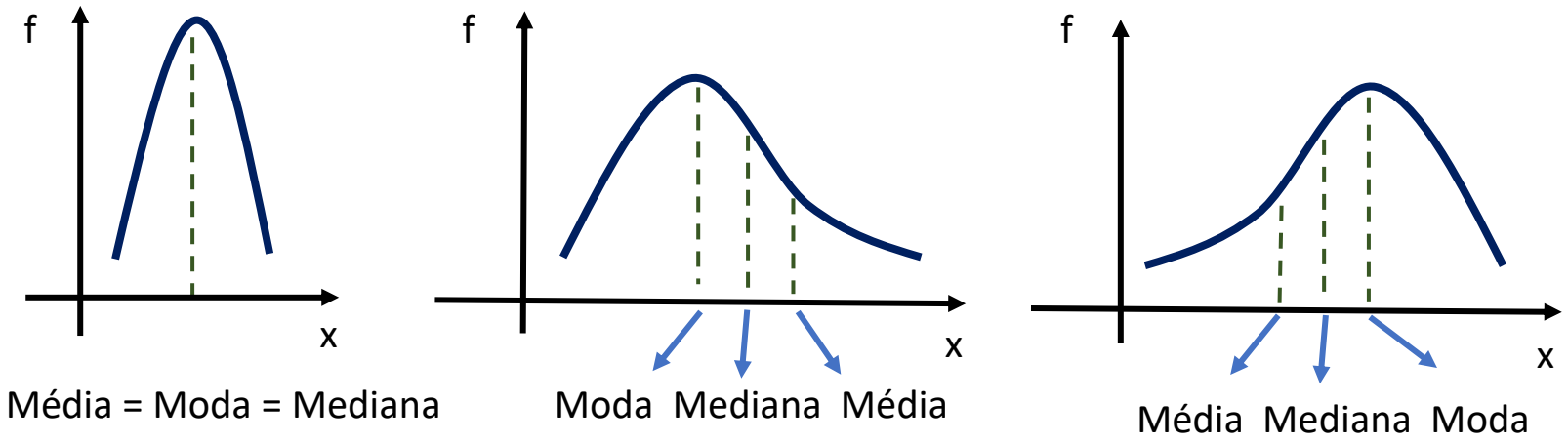
São verdadeiras apenas as afirmativas a) I e II. b) II e III. c) I e IV. d) I, II e III. e) I, II e IV.

I. M = (NP1 + NP2)/2 = (0,0 + 0,0)/2 = 0,0 MF = (M + Ex)/2 = (0,0 + 9,5)/2 = 4,75 → 5,0
II. M = (NP1 + NP2)/2 = (5,0 + 5,0)/2 = 5,0 MF = (M + Ex)/2 = (5,0 + 5,0)/2 = 5,0 → 5,0
III. M = (NP1 + NP2)/2 = (8,0 + 2,0)/2 = 5,0 MF = (M + Ex)/2 = (5,0 + 4,0)/2 = 4,5 → 4,5
IV. M = (NP1 + NP2)/2 = (4,0 + 2,0)/2 = 3,0 MF = (M + Ex)/2 = (3,0 + 7,0)/2 = 5,0 → 5,0

Exemplo 3.2. Na figura a seguir encontramos um gráfico da variável x no eixo-x e da sua respectiva frequência no eixo-y, sendo indicado neste as posições da média, moda e mediana, conforme distribuição dos dados. Analise as afirmações a seguir:

I. Moda é o valor ou elemento que aparece com maior frequência, enquanto a média simples é a somatória de todos os valores numéricos dos elementos dividido pelo número inteiro destes elementos.
II. Mediana é o valor que se encontra no meio do conjunto ordenado em ordem crescente, e se este número de elementos não for ímpar é só fazer a média entre os dois números beirando o meio do conjunto.
III. É evidente que o valor da média estará no intervalo entre o valor mínimo e o máximo do conjunto de elementos. A média pode coincidir com os valores da moda e da mediana ou não.

Está correto apenas o que se afirma em a) I. b) II. c) III. d) I e III. e) I, II e III.



4. Medidas de dispersão

As medidas de dispersão demonstram o quanto os dados estão dispersos e dão base numérica para tomada de decisões. A variância monstra o quanto o conjunto varia, assim uma variância zero significa que todos os valores de x são exatamente iguais. O desvio padrão é simplesmente calculado como a raiz da variância. O coeficiente de variação tem a vantagem de dar um valor em porcentagem. O escore é para cara valor de x, medindo o quanto este valor de x é raro ou comum no conjunto.

Variância

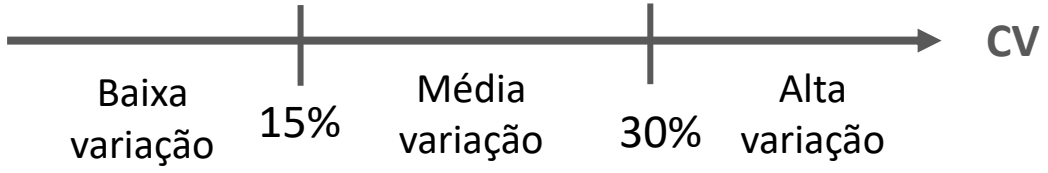
$$s^2 = \frac{\sum(x-\bar{x})^2}{n-1}$$

Desvio Padrão

$$s = \sqrt{s^2}$$

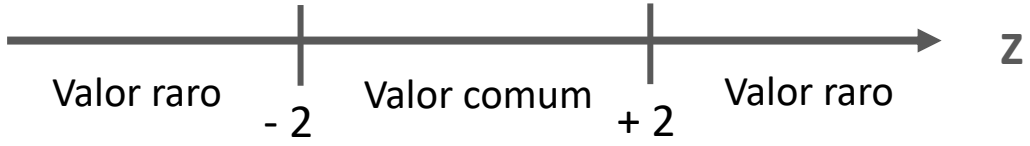
Coeficiente de variação

$$CV = \frac{s}{\bar{x}} \cdot 100\%$$



Escore

$$z = \frac{x-\bar{x}}{s}$$



Exemplo 4.1. Considere a idade em anos dos alunos: 18, 18,19, 19, 19, 20, 20, 20, 20, 21, 21, 21, 22. Calcule todas as medidas de dispersão que forem possíveis.

O cálculo das dispersões de um conjunto de dados será melhor efetuado com a montagem da tabela ao lado. Fique atento com a frequência nas contas, é fácil esquecer disso e errar tudo em uma tragédia grega.

x	f	$(x-\bar{x})^2$	$f \cdot (x-\bar{x})^2$
18	2	3,42	6,85
19	3	0,72	2,17
20	4	0,02	0,09
21	3	1,32	3,97
22	1	4,62	4,62
Σ	13	10,10	17,70

$$\bar{x} = \frac{\sum x}{n} = 19{,}85$$

$$s^2 = \frac{\sum (x-\bar{x})^2}{n-1} = \frac{17{,}70}{13-1} = 1{,}48$$

$$s = \sqrt{s^2} = \sqrt{1{,}48} = 1{,}21$$

$$CV = \frac{s}{\bar{x}} \cdot 100\% = \frac{1{,}21}{19{,}85} \cdot 100\% = 6{,}12\%$$

$$z = \frac{x-\bar{x}}{s}$$

x	z
18	-1,53
19	-0,70
20	+0,12
21	+0,95
22	+1,78

Exemplo 4.2. Considere que em uma turma de 20 alunos de um professor malvado as notas sejam descritas pela tabela malvada a seguir. Calcule todas as medidas de dispersão que forem possíveis.

Notas	Alunos
0,0 ⊢ 2,0	5
2,0 ⊢ 4,0	6
4,0 ⊢ 6,0	4
6,0 ⊢ 8,0	3
8,0 ⊢ 10,0	2
Σ	20

x	PM	f	$(x-\bar{x})^2$	$f \cdot (x-\bar{x})^2$	Escore
0,0 ⊢ 2,0	1,0	5	9,61	48,05	-1,18
2,0 ⊢ 4,0	3,0	6	1,21	7,26	-0,42
4,0 ⊢ 6,0	5,0	4	0,81	3,24	+0,34
6,0 ⊢ 8,0	7,0	3	8,41	25,23	+1,10
8,0 ⊢ 10,0	9,0	2	24,01	48,02	+1,86
Σ		20	44,05	131,18	

$$\bar{x} = \frac{\sum x}{n} = 4{,}1$$

O valor de x está sendo o PM.

$$s = \sqrt{s^2} = \sqrt{6{,}94} = 2{,}63$$

$$z = \frac{x - \bar{x}}{s}$$

$$s^2 = \frac{\sum (x-\bar{x})^2}{n-1} = \frac{131{,}18}{20-1} = 6{,}94$$

$$CV = \frac{s}{\bar{x}} \cdot 100\% = \frac{2{,}63}{4{,}1} \cdot 100\% = 64{,}24\%$$

Exemplo 4.3. Um professor muito cruel, ao lecionar a matéria Estatística e Probabilidade durante 5 anos, reprova a quantidade de aluno por ano: 10, 7, 5, 5 e 8. Sabe-se que a média é $\bar{x} = \Sigma x/n$ e a variância é $s^2 = [\Sigma(x - \bar{x})^2]/[n - 1]$, que o coeficiente de variação (CV) é dado por $(s/\bar{x}).100\%$, em que um valor menor que 15% é de baixa variação, de 15 a 30% de variação moderada e mais de 30% é alta variação; outro valor importante é o escore (z), definido como $(x - \bar{x})/s$, onde um valor entre -2 e +2 é comum, e raro se fora desse intervalo. Analise as afirmativas a seguir:

I. A média de reprovações por ano foi 7 com desvio padrão de 2,12 e todos os valores são comuns.
II. A variância é 5,1, o coeficiente de variação é 30,3 % e existem dois valores raros no conjunto.
III. A variância é 4,5, o coeficiente de variação é 30,3 % e o escore do valor 10 é de + 1,41.
IV. Todos os escores deram positivo e o coeficiente de variação deu baixa variação.

São verdadeiras apenas as afirmativas a) I, II e III. b) I e II. c) II e III. d) I e III. e) I e IV.

$\bar{x} = \Sigma x/n = (10 + 7 + 5 + 5 + 8)/5 = 35/5 = 7$

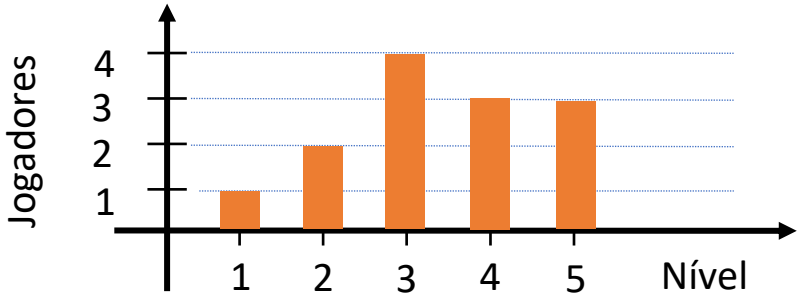
$$s^2 = \frac{\sum(x - \bar{x})^2}{n - 1} = \frac{(10 - 7)^2 + (7 - 7)^2 + (5 - 7)^2 + (5 - 7)^2 + (8 - 7)^2}{5 - 1} = 4{,}5$$

$s = \sqrt{(4{,}5)} = 2{,}12$ $\quad$ $CV = (s/\bar{x}).100\% = (2{,}12/7).100\% = 30{,}3\ \%$

$z = (x - \bar{x})/s \rightarrow z_1 = (10 - 7)/2{,}12 = +1{,}41 \quad z_2 = (7 - 7)/2{,}12 = 0$

$z_{3,4} = (5 - 7)/2{,}12 = -0{,}94 \quad z_5 = (8 - 7)/2{,}12 = +0{,}47$

Exemplo 4.4. Um jogo eletrônico em aplicativo de celular possui 5 níveis, este foi testado com 13 jogadores e ao final de 1 hora o nível alcançado por eles está representado no histograma a seguir. Determine o coeficiente de variação. Dados: $\bar{x} = \Sigma x/n = \Sigma(x_i.f_i)/n$; $s^2 = [\Sigma(x_i - \bar{x})^2.f_i]/[n - 1]$; $CV = (s/\bar{x}).100\%$.



Nível	Jogadores
1	1
2	2
3	4
4	3
5	3

x	f	$x_i.f_i$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2.f_i$
1	1	1	5,76	5,76
2	2	4	1,96	3,92
3	4	12	0,16	0,64
4	3	12	0,36	1,08
5	3	15	2,56	7,68
Σ	13	44		19,08

$\bar{x} = \Sigma x/n = \Sigma(x_i.f_i)/n = 44/13 = 3,4$

$s^2 = [\Sigma(x_i - \bar{x})^2.f_i]/[n-1] = 19,08/(13 - 1) = 1,59$

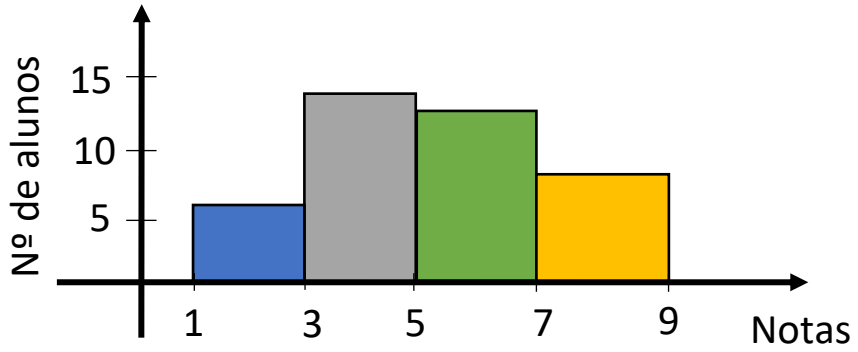
$s = \sqrt{s^2} = \sqrt{(1,59)} = 1,26$

CV = 47 %

$CV = (s/\bar{x}).100\% = (1,59/3,4).100\% = 47\ \%$

Exemplo 4.5. A tabela e seu histograma a seguir representam as notas de uma turma com 40 alunos. Dados: $\bar{x} = \Sigma x/n = \Sigma(x_i.f_i)/n$; $s^2 = [\Sigma(x_i - \bar{x})^2.f_i]/[n - 1]$; $CV = (s/\bar{x}).100\%$. Determine o coeficiente de variação CV.

Notas	Nº de alunos
1,0 ⊢-- 3,0	6
3,0 ⊢-- 5,0	14
5,0 ⊢-- 7,0	12
7,0 ⊢-- 9,0	8



x	f	PM	$PM_i.f_i$	$(PM_i - \bar{x})^2$	$(PM_i - \bar{x})^2.f_i$
1,0 ⊢-- 3,0	6	2,0	12	9,61	57,66
3,0 ⊢-- 5,0	14	4,0	56	1,21	16,94
5,0 ⊢-- 7,0	12	6,0	72	0,81	9,72
7,0 ⊢-- 9,0	8	8,0	64	8,41	67,28
Σ	40		204		151,60

$\bar{x} = \Sigma x/n = \Sigma(PM_i.f_i)/n = 204/40 = 5,1$

$s^2 = [\Sigma(PM_i - \bar{x})^2.f_i]/[n - 1] = 151,60/(40 - 1) = 3,89$

$s = \sqrt{s^2} = \sqrt{(3,89)} = 1,97$

$CV = (s/\bar{x}).100\% = (1,97/5,1).100\% = 38,7\ \%$ → $\boxed{CV = 38,7\ \%}$

5. Conceitos básicos de probabilidade

A palavra “probabilidade” vem do latim *probare* que significa “provar” no sentido de testar, tentar, provável, ou seja, a possibilidade de algo acontecer ou não no teste, na prova. Talvez você sinta espiritualmente isto como uma “provação” divina porque Deus te criou para ser grande, para ser o Rambo. O estudo da probabilidade leva em consideração o conjunto de todos os resultados possíveis de um experimento aleatório (espaço amostral) e o subconjunto de casos possíveis que se esteja estudando (eventos).

$$p = \frac{número\ de\ casos\ de\ estudo}{número\ de\ casos\ possíveis}$$

Um asteroide do nível que matou os dinossauros há 65 milhões de anos atrás costuma cair na Terra a cada 65 milhões de anos. O tempo desta prova é de 1:00 hora, logo qual a probabilidade de um asteroide desse nível cair na Terra durante a aplicação desta prova?

$$p = \frac{1\ hora}{65.000.000\ anos} = \frac{1\ h}{65.000.000.365.24\ h} = 1{,}75.10^{-12} = 0{,}000.000.000.175\ \%$$

Tipos de eventos

Vamos estudar três tipos de eventos: complementares, independentes e mutuamente exclusivos.

Eventos complementares: a probabilidade de um evento ocorrer (sucesso) será p e de não ocorrer (fracasso) será q de forma que elas se complementam em **p + q = 1**. Este "1" significa 100%.
Qual a probabilidade de não se obter o número 4 em um lançamento de um dado de 6 faces?
$p = 1/6$ $q = ?$ $p + q = 1 \rightarrow 1/6 + q = 1 \rightarrow q = 1 - 1/6 \rightarrow q = 5/6$

Eventos independentes: um evento não altera a probabilidade do outro ocorrer, logo a probabilidade do evento A, p(a), ocorrer simultaneamente com o evento B, p(b), será **p(a e b) = p(a).p(b)**.
Qual a probabilidade de se lançar dois dados de 6 faces e obter os resultados 5 e 6?
$p(a) = 1/6$ $p(b) = 1/6$ $p(a \text{ e } b) = p(a).p(b) = 1/6 \,.\, 1/6 = 1/36$

Eventos mutuamente exclusivos: a realização do evento A, p(a), exclui a possibilidade de ocorrer o evento B, p(b), logo a probabilidade de um ou outro será **p(a ou b) = p(a) + p(b)**.
Qual a probabilidade de se lançar um dado de 6 faces e se tirar 3 ou 5?
$p(a) = 1/6$ $p(b) = 1/6$ $p(a \text{ ou } b) = p(a) + p(b) = 1/6 + 1/6 = 2/6 = 1/3$

Exemplo 5.1. Em um laboratório farmacêutico, através de uma mutação, surge um vírus zumbi com 75% de chance de ser transmitido por mordida recebida de um zumbi. Analise as afirmativas a seguir:

I. Duas pessoas são mordidas, a probabilidade de nenhuma delas virar zumbi é 6,25%.
II. A probabilidade de alguém ser transformado em zumbi na quinta mordida é de 0,8%.
III. Um casal de namorados é mordido, a probabilidade de só ela virar zumbi é de 18,75%.
IV. Grupo de 10 pessoas, todas mordidas, probabilidade de ninguém virar zumbi é $9{,}53.10^{-12}$.

São verdadeiras apenas as afirmativas
a) I, II e III.
b) I e II.
c) II e III.
d) I e III.
e) I e IV.

I. p(a) = 0,25 x 0,25 = 0,0625 = 6,25 %
II. p(b) = 0,25 x 0,25 x 0,25 x 0,25 x 0,75 = 0,003 = 0,3 %
III. p(c) = 0,75 x 0,25 = 0,1875 = 18,75 %
IV. p(d) = 0,25 x 0,25 x (...) x 0,25 = $0{,}25^{10} = 9{,}53.10^{-7}$

p = 0,75 q = 0,25

Exemplo 5.2. Um curso universitário começa com 50 alunos no 1º ano, tem média de 35 alunos no início do 2º ano, 25 no início do 3º ano, 15 no 4º ano e se formam 5 em média por ano. Um veterano terrorista do 3º ano conta isso aos calouros, afirmando que apenas 1 a cada 10 deles vão se formar. Um calouro esperto ao saber disso, em reação, começa a fazer as contas das probabilidades desse mesmo veterano. Analise as afirmativas a seguir:

I. A probabilidade de um aluno, escolhido aleatoriamente, que irá passar do 3º para o 4º ano é de 60%.
II. A probabilidade de um aluno, escolhido aleatoriamente, que está no 3º ano e irá se formar é de 20%.
III. O esforço do aluno lhe permite fazer sua própria "sorte". A probabilidade mostra a tendência do grupo.
IV. Uma sala com 100% de alunos excelentes teria as mesmas probabilidades de 10% de alunos excelentes.

São verdadeiras apenas as afirmativas

a) I, II e III.
b) I e II.
c) II e III.
d) I e III.
e) I e IV.

p(a) = 15/25 = 0,6 = 60 %
p(b) = 5/25 = 0,2 = 20 %

Exemplo 5.3. Nos tempos da Guerra Fria um agente da KGB desconfia de um colega, então para extrair a sua confissão de traição faz uma roleta russa com um revólver com capacidade de 6 balas. Ele coloca uma bala e gira o tambor após cada disparo. Ignore a possibilidade de alguma munição falhar, de alguém sobreviver ao tiro na cabeça ou de existir compaixão no torturador russo. Analise as possibilidades a seguir:

I. São dados três tiros, a probabilidade de não morrer é de 52,9%.
II. São dados dois tiros, a probabilidade de morrer no segundo tiro é de 13,9%.
III. O carrasco atira 10 vezes, a probabilidade de sobreviver é 16,2%.
IV. O carrasco dá 5 tiros, a probabilidade de morrer no quinto tiro é 80,4%.

São verdadeiras apenas as afirmativas
a) I, II e III.
b) I e II.
c) II e III.
d) I e III.
e) I e IV.

I. $p(a) = (5/6) \times (5/6) \times (5/6) = 0{,}579 = 57{,}9\ \%$ $\quad p = 1/6 \quad q = 5/6$
II. $p(b) = (5/6) \times (1/6) = 0{,}139 = 13{,}9\ \%$
III. $p(c) = (5/6)^{10} = 0{,}162 = 16{,}2\ \%$
IV. $p(d) = (5/6)^{4} \times (1/6)^{1} = 0{,}0804 = 8{,}04\ \%$

Exemplo 5.4. Em um jogo de dois dados de seis lados é somado o resultado desses dois dados, podendo variar o resultado de 2 a 12. Ao montar uma tabela de possíveis resultados se obtém:

Resultado	2	3	4	5	6	7	8	9	10	11	12
Frequência	1	2	3	4	5	6	5	4	3	2	1

Os dados não são viciados, logo cada número possui a mesma chance de modo que são 36 possibilidades de combinações em cada lançamento. Determine:

a) a probabilidade de ser tirar um resultado menor ou igual a 5;

b) a probabilidade de se tirar o número 7;

c) a probabilidade de se tirar um número maior que 10.

Resolução

a) a probabilidade de ser tirar um resultado menor ou igual a 5;

Resultado	2	3	4	5	6	7	8	9	10	11	12
Frequência	1	2	3	4	5	6	5	4	3	2	1

$$p = \frac{1}{36} + \frac{2}{36} + \frac{3}{36} + \frac{4}{36} = \frac{1+2+3+4}{36} = \frac{10}{36} = 0{,}2778 = 27{,}78\%$$

b) a probabilidade de se tirar o número 7;

Resultado	2	3	4	5	6	7	8	9	10	11	12
Frequência	1	2	3	4	5	6	5	4	3	2	1

$$p = \frac{6}{36} = 0{,}1667 = 16{,}67\%$$

c) a probabilidade de se tirar um número maior que 10.

Resultado	2	3	4	5	6	7	8	9	10	11	12
Frequência	1	2	3	4	5	6	5	4	3	2	1

$$p = \frac{2}{36} + \frac{1}{36} = \frac{2+1}{36} = \frac{3}{36} = 0{,}0833 = 8{,}33\%$$

Você acha que já está difícil? Segundo o que eu estou vendo aqui as coisas vão piorar e depois piorar mais ainda. Isso é o normal da vida.

Mandy, Cartoon Network.

6. Distribuições binomial e de Poisson

Para problemas com apenas dois resultados possíveis e repetitivos trata-se de algo binomial. Caso a repetição seja em intervalos de espaço ou tempo trata-se de Poisson.

Distribuição Binomial

$$p(x) = \frac{n!}{x! \cdot (n-x)!} \cdot p^x \cdot (1-p)^{n-x}$$

Distribuição de Poisson

$$p(x) = \frac{\lambda^x \cdot e^{-\lambda}}{x!}$$

x = número de sucessos nas n repetições
n = número de repetições
p = probabilidade de sucesso em qualquer repetição
λ = número médio de sucesso em intervalo de espaço ou tempo

Fatorial n! = n.(n - 1).(n – 2). (...) .3.2.1 Número de Néper *e = 2,7183...*

Exemplo 6.1. Uma praça está repleta de pombos e estes defecam frequentemente sobre as pessoas. Cada pessoa que passa pela praça possui uma probabilidade de 20% de ser atingida pela matéria fecal destes amáveis pássaros. Um grupo de 10 pessoas passa pela praça e 5 delas são bombardeadas. Qual a probabilidade deste fétido evento? Dado: $p[x = r] = nCr.p^r.q^{n-r}$.

Organizando dados:

p = 0,2
n = 10
x = 5
p(x) = ?

$$\boldsymbol{p(x) = \frac{n!}{x! \cdot (n-x)!} \cdot p^x \cdot (1-p)^{n-x}}$$

$$p(5) = \frac{10!}{5! \cdot (10-5)!} \cdot 0{,}2^5 \cdot (1-0{,}2)^{10-5}$$

$$p(5) = 252 \cdot 0{,}2^5 \cdot 0{,}8^5$$

$$\boxed{p(5) = 0{,}0264 = 2{,}64\ \%}$$

A mesma conta, mas direto pela função combinação na calculadora:

$p[x = r] = nCr.p^r.q^{n-r}$ → $p[x = 5] = 10C5 \cdot 0{,}2^5 \cdot 0{,}8^{10-5} = 0{,}0264$

Exemplo 6.2. Uma esposa infiel costuma "pular a cerca" com média de cinco vezes por semana. Ignore o tamanho do chifre do marido. Qual é a probabilidade de que em uma semana esta esposa dê oito "puladas de cerca"? Dado: $p_x = \lambda^x.e^{-\lambda}/x!$

Organizando dados:

$\lambda = 5$
$x = 8$
$p(x) = ?$

$$\boldsymbol{p(x) = \frac{\lambda^x \cdot e^{-\lambda}}{x!}}$$

$$p(8) = \frac{5^8 \cdot e^{-5}}{8!}$$

$$\boxed{p(8) = 0{,}0653 = 6{,}53\ \%}$$

Exemplo 6.3. Uma prova de Estatística e Probabilidade é composta de 10 questões de múltipla escolha com 5 alternativas e apenas uma correta. Um aluno que não estudou marca aleatoriamente todas as questões da prova. Qual a probabilidade dele acertar 3 questões? Dado: $p[x = r] = nCr.p^{r}.q^{n-r}$.

Organizando dados:

p = 1/5 = 0,2
n = 10
x = 3
p(x) = ?

$$\boldsymbol{p(x) = \frac{n!}{x! \cdot (n-x)!} \cdot p^{x} \cdot (1-p)^{n-x}}$$

$$p(3) = \frac{10!}{3! \cdot (10-3)!} \cdot 0{,}2^{3} \cdot (1-0{,}2)^{10-3}$$

$$p(3) = 120 \cdot 0{,}2^{3} \cdot 0{,}8^{7}$$

$$\boxed{p(3) = 0{,}2013 = 20{,}13\ \%}$$

A mesma conta, mas direto pela função combinação na calculadora:

$p[x = r] = nCr.p^{r}.q^{n-r} \rightarrow p[x = 3] = 10C3 \cdot 0{,}2^{3} \cdot 0{,}8^{10-3} = 0{,}2013$

Exemplo 6.4. Um professor malvado costuma reprovar uma média de 40 alunos por semestre. Ignore o dilúvio de lágrimas. Qual a probabilidade de que em um semestre ele reprove exatamente 40 alunos? Dado: $p_x = \lambda^x . e^{-\lambda}/x!$

Organizando dados:

$\lambda = 40$
$x = 40$
$p(x) = ?$

$$\boldsymbol{p(x) = \frac{\lambda^x \cdot e^{-\lambda}}{x!}}$$

$$p(40) = \frac{40^{40} \cdot e^{-40}}{40!}$$

$$\boxed{p(40) = 0{,}0629 = 6{,}29\ \%}$$

Exemplo 6.5. Em um experimento escolar são plantados feijões com probabilidade de 85% de germinar. Cada grupo de alunos planta 5 grãos. Determine a probabilidade de um grupo obter a germinação de no máximo 2 grãos. Dado: $p[x = r] = nCr.p^r.q^{n-r}$.

$\mathbf{p[x = r] = nCr.p^r.q^{n-r}}$ $\quad$ $p = 0{,}85$ $\quad$ $q = 0{,}15$ $\quad$ $n = 5$ $\quad$ $x \leq 2 = \{0, 1, 2\}$

$p[x = 0] = 5C0 \times 0{,}85^0 \times 0{,}15^{5-0} = 0{,}0076\ \%$

$p[x = 1] = 5C1 \times 0{,}85^1 \times 0{,}15^{5-1} = 0{,}2152\ \%$

$p[x = 2] = 5C2 \times 0{,}85^2 \times 0{,}15^{5-2} = 2{,}4384\ \%$

$p[x \leq 2] = 0{,}0076 + 0{,}2152 + 2{,}4384 = 2{,}6612\ \%$

$\boxed{p[x \leq 2] = 2{,}6612\ \%}$

7. Correlação linear e regressão linear

A regressão linear permite obter a reta na forma **y = a.x + b** que melhor se aproxima dos pontos conhecidos e a correlação linear r determina a relação entre os números de x e de y de forma que **-1 ≤ r ≤ +1**. Os valores de a, b e r são dados pelas terríveis fórmulas a seguir:

Puro Osso, Cartoon Network.

$$a = \frac{n \cdot \sum x \cdot y - \sum x \cdot \sum y}{n \cdot \sum x^2 - (\sum x)^2}$$

$$b = \frac{\sum y}{n} - a.\frac{\sum x}{n}$$

$$r = \frac{n \cdot \sum x \cdot y - \sum x \cdot \sum y}{\sqrt{(n \cdot \sum x^2 - (\sum x)^2) \cdot (n \cdot \sum y^2 - (\sum y)^2)}}$$

Essa prova vai fazer um estrago. Se eu tivesse sentimentos até teria pena dessas pobres almas.

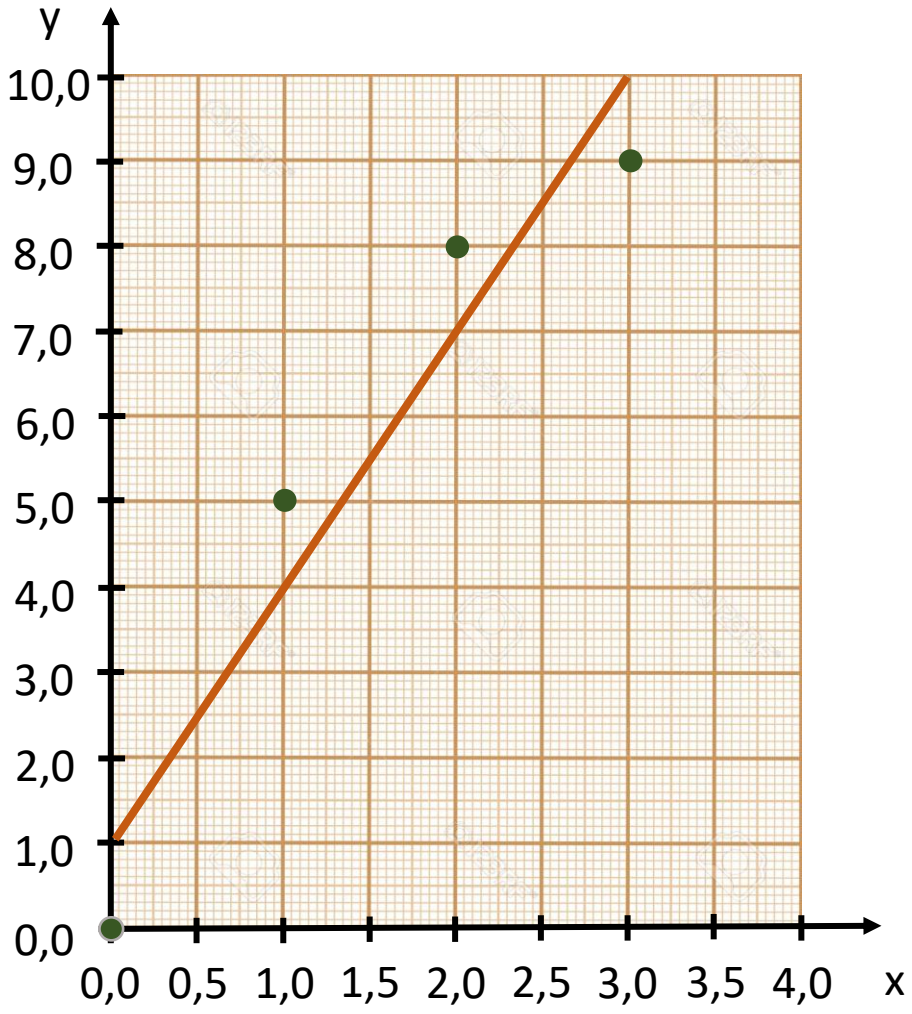
Exemplo 7.1. Uma função linear pode ser escrita na forma $y = a.x + b$ quando associamos variáveis nas coordenadas x e y de um gráfico, disso também resulta a correção linear r que varia de $-1 \leq r \leq +1$. Analise as afirmações a seguir:

I. A correlação é a relação entre variáveis x e y, sendo correlação linear se os pontos formarem uma reta.
II. A correlação será positiva se x crescendo y também crescer, ou se x decrescendo y também decrescer.
III. A correlação será negativa se x crescendo então y decrescer, ou se x decrescendo então y crescer.

Está correto apenas o que se afirma em

a) I e II.
b) I e III.
c) II e III.
d) I, II e III.
e) Nenhuma.

Exemplo 7.2. Considere os quatros pontos A(0;0), B(1;5), C(2;8) e D(3;9), construa um gráfico com eles e determine a reta y = a.x +b e o valor da correlação linear r.



Podemos perceber no gráfico que os pontos não se ligam em uma reta, entretanto uma reta foi desenhada no gráfico com o melhor equilíbrio visual entre os pontos. As equações no próximo slide apresentarão o resultado numérico em que esta reta y = a.x + b deverá ser a partir dos valores x e y dos pontos.

Ponto	x	y
A	0	0
B	1	5
C	2	8
D	3	9

Os pontos A(0;0), B(1;5), C(2;8) e D(3;9) são em número n = 4, eles precisam ser colocados na tabela a seguir em se que calcule antes x.y, x^2, y^2 e suas somatórias para minimizar sofrimento ao usar as equações.

$$\boldsymbol{a} = \frac{\boldsymbol{n} \cdot \sum \boldsymbol{x} \cdot \boldsymbol{y} - \sum \boldsymbol{x} \cdot \sum \boldsymbol{y}}{\boldsymbol{n} \cdot \sum \boldsymbol{x}^2 - (\sum \boldsymbol{x})^2} = \frac{4 \cdot 48 - 6 \cdot 22}{4 \cdot 14 - (6)^2} = 3$$

$$\boldsymbol{b} = \frac{\sum \boldsymbol{y}}{\boldsymbol{n}} - \boldsymbol{a}.\frac{\sum \boldsymbol{x}}{\boldsymbol{n}} = \frac{22}{4} - 3.\frac{6}{4} = 1$$

y = a.x + b → y = 3.x + 1

Ponto	x	y	x.y	x^2	y^2
A	0	0	0	0	0
B	1	5	5	1	25
C	2	8	16	4	64
D	3	9	27	9	81
Σ	6	22	48	14	170

$$\boldsymbol{r} = \frac{\boldsymbol{n} \cdot \sum \boldsymbol{x} \cdot \boldsymbol{y} - \sum \boldsymbol{x} \cdot \sum \boldsymbol{y}}{\sqrt{(\boldsymbol{n} \cdot \sum \boldsymbol{x}^2 - (\sum \boldsymbol{x})^2) \cdot (\boldsymbol{n} \cdot \sum \boldsymbol{y}^2 - (\sum \boldsymbol{y})^2)}}$$

$$r = \frac{4 \cdot 48 - 6 \cdot 22}{\sqrt{(4 \cdot 14 - (6)^2) \cdot (4 \cdot 170 - (22)^2)}} = +0{,}96$$

→ r = +0,96

8. Distribuição normal

Primeiro é preciso diferenciar média e desvio padrão, para amostra ($\bar{x}$, s) e para população (μ, σ). Existe uma relação entre essas médias e desvios padrões, visto que a média da população é a média das médias amostrais e o desvio padrão da população é $\sigma = s/\sqrt{n}$. Para amostras grandes (n ≥ 30) podemos considerar s ≈ σ.

Amostra

População

	Variância	Desvio Padrão	Escore
Amostral	$s^2 = \frac{\sum(x-\bar{x})^2}{n-1}$	$s = \sqrt{s^2}$	$z = \frac{x-\bar{x}}{s}$
Populacional	$\sigma^2 = \frac{\sum(x-\mu)^2}{n}$	$\sigma = \sqrt{\sigma^2}$	$z = \frac{x-\mu}{\sigma}$

Não é erro de digitação, um é “n – 1” e o outro é só “n”.

O valor da probabilidade

Quando x é uma variável contínua e a amostra com n ≥ 30 já é possível gerar um gráfico com uma curva em formato de sino, pois as densidades de probabilidade já se distribuem normalmente. A seguir as equações que foram invocadas das profundezas do inferno para a probabilidade assumir sua forma no mundo material.

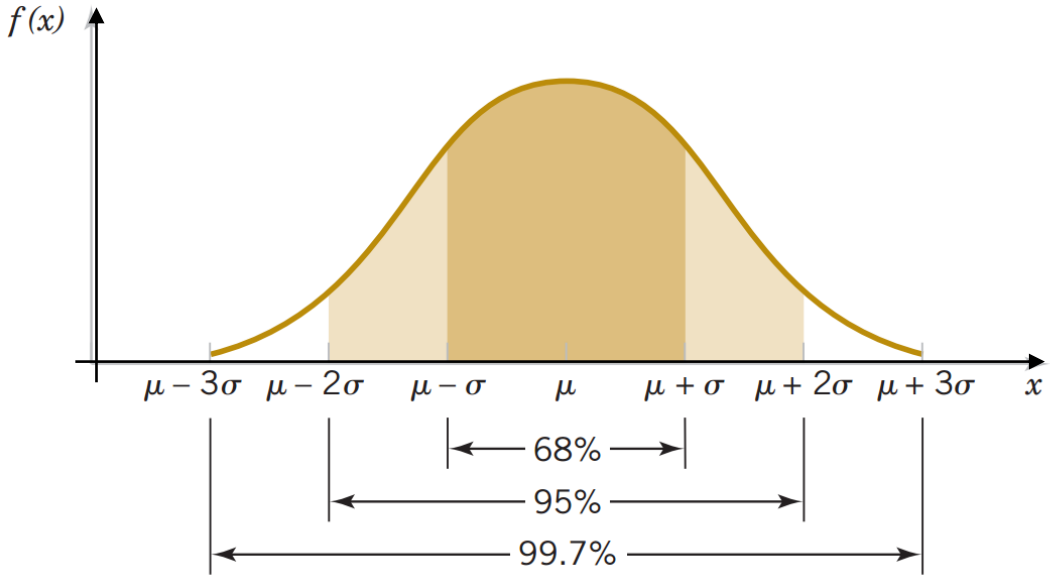
Função distribuição de probabilidades

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Cálculo da probabilidade entre “a” e “b”

$$p(a < x < b) = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

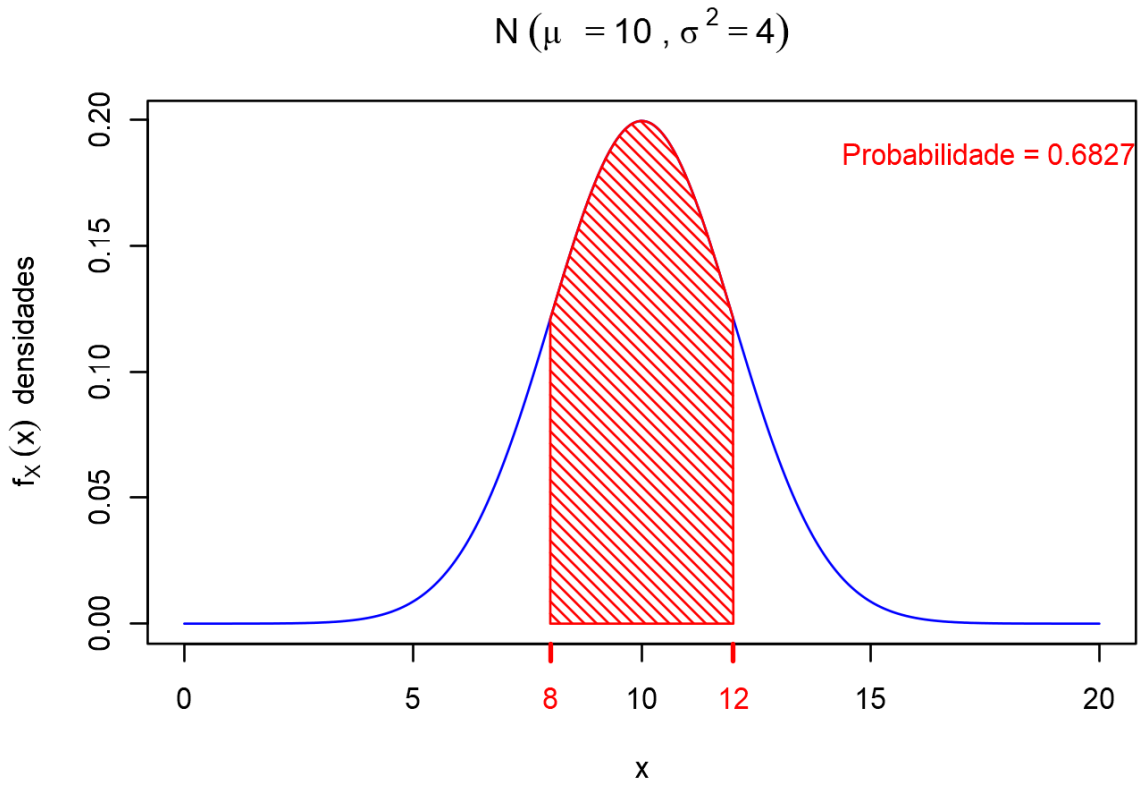
Da solução desta equação nasce o gráfico ao lado.



https://www.inf.ufsc.br/~andre.zibetti/probabilidade/normal.html

Vamos entender esse tipo de gráfico?

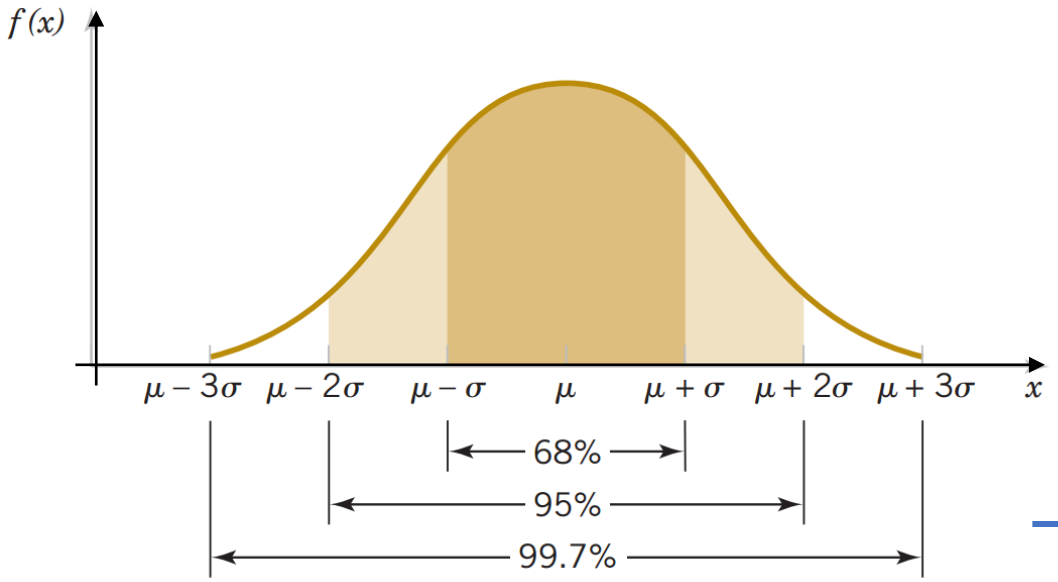
O gráfico a seguir foi gerado com média μ = 10 e variância σ² = 4. A probabilidade foi calculada com base naquela equação invocada do inferno para valores de x entre 8 e 12, logo temos p(8 < x < 12) = 0,6827.



https://www.inf.ufsc.br/~andre.zibetti/probabilidade/normal.html

Intervalo provável para x

O gráfico a seguir mostra que existe uma probabilidade de 68% de x estar no intervalo $\mu - \sigma < x < \mu + \sigma$, para qualquer valor de média μ e desvio padrão σ . Da mesma forma para outros intervalos.



$p(\mu - \sigma < x < \mu + \sigma) = 0{,}68 = 68\ \%$

$p(\mu - 2\sigma < x < \mu + 2\sigma) = 0{,}95 = 95\ \%$

$p(\mu - 3\sigma < x < \mu + 3\sigma) = 0{,}997 = 99{,}7\ \%$

https://www.inf.ufsc.br/~andre.zibetti/probabilidade/normal.html

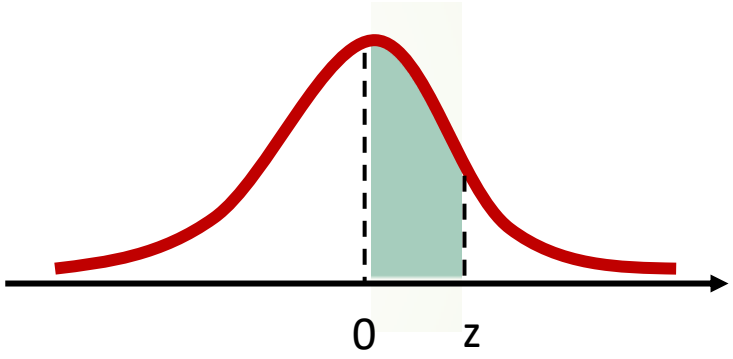
Tabela Normal

A tabela normal apresenta a probabilidade para o valor do escore z já calculado, ou seja, sabendo a média μ e o desvio padrão σ se calcula o escore z e na tabela se encontra a probabilidade p. Essa tabela foi criada sabendo que a área abaixo da curva é numericamente igual à probabilidade. Como usar essa tabela? Para z = 0,21 temos p = 0,0832 e para z = 0,32 temos p = 0,1255. Entendeu?

Segunda casa decimal de z

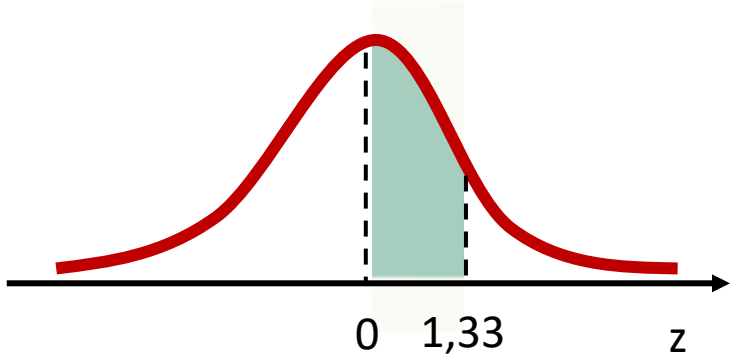
z	0,00	0,01	0,02	...
0,0	0,0000	0,0040	0,0080	...
0,1	0,0398	0,0438	0,0478	...
0,2	0,0793	0,0832	0,0871	...
0,3	0,1179	0,1217	0,1255	...
...	...	...	...	...

$$z = \frac{x - \mu}{\sigma}$$



Exemplo 8.1. Em uma turma boa de Estatística e Probabilidade a nota média é 6,5 com desvio padrão de 1,5. Sabe-se que o escore é $z = (x - \mu)/\sigma$ e é necessário usar a tabela normal de probabilidades. Qual a probabilidade de um aluno ter sua nota entre 6,5 e 8,5?

a) 21,04 %
b) 19,15 %
c) 40,82 %
d) 67,67 %
e) 74,92 %



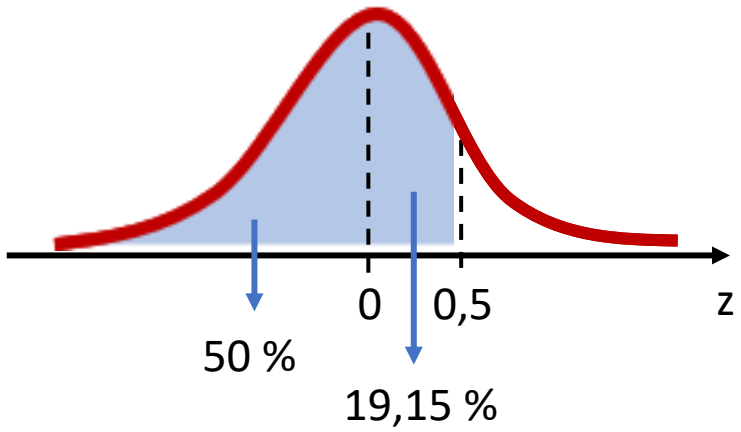
$z_{mín} = (x_{mín} - \mu)/\sigma = (6,5 - 6,5)/1,5 = 0,00 \rightarrow$ Tabela Normal $\rightarrow p_{mín} = 0,0000$

$z_{máx} = (x_{máx} - \mu)/\sigma = (8,5 - 6,5)/1,5 = 1,33 \rightarrow$ Tabela Normal $\rightarrow p_{máx} = 0,4082$

$p(6,5 < x < 8,5) = p(z_{mín} < z < z_{máx})$ = área do intervalo = 0,4082 – 0,0000 = 0,4082

Exemplo 8.2. O tempo de resposta na execução de um algoritmo é uma variável com distribuição normal com média 23 segundos e desvio padrão de 4 segundos. Sabe-se que o escore é $z = (x - \mu)/\sigma$ e é necessário usar a tabela normal de probabilidades. Qual é a probabilidade do tempo de resposta ser menor que 25 segundos?

a) 12,00 %
b) 21,04 %
c) 40,82 %
d) 69,15 %
e) 84,92 %

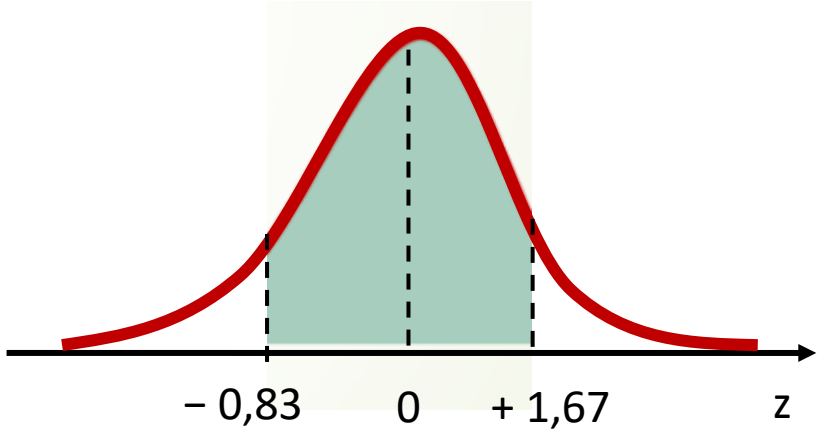


$z = (x - \mu)/\sigma = (25 - 23)/4 = 0{,}50 \rightarrow$ Tabela Normal $\rightarrow$ $p = 0{,}1915$

$p(x < 25) = p(z < 0{,}50) =$ área do intervalo $= 0{,}5000 + 0{,}1915 = 0{,}6915$

Exemplo 8.3. A porcentagem de absorção de água de um tipo de piso cerâmico possui distribuição normal com média 2,5 e desvio padrão 0,6. Sabe-se que o escore é $z = (x - \mu)/\sigma$ e é necessário usar a tabela normal de probabilidades. Qual a probabilidade de um piso escolhido aleatoriamente acusar absorção de água entre 2,0% e 3,5%?

a) 21,04 %
b) 19,15 %
c) 40,82 %
d) 67,67 %
e) 74,92 %



$z_{mín} = (x_{mín} - \mu)/\sigma = (2{,}0 - 2{,}5)/0{,}6 = -0{,}83 \rightarrow$ Tabela Normal $\rightarrow p_{mín} = 0{,}2967$

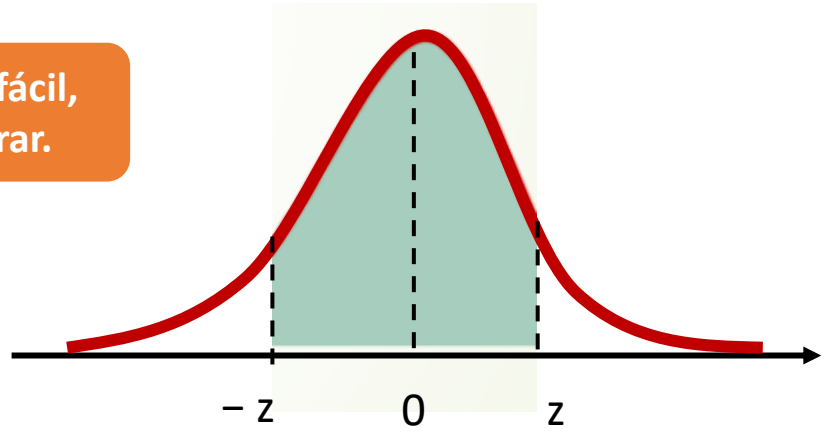
$z_{máx} = (x_{máx} - \mu)/\sigma = (3{,}5 - 2{,}5)/0{,}6 = +1{,}67 \rightarrow$ Tabela Normal $\rightarrow p_{máx} = 0{,}4515$

$p(2{,}0 < x < 3{,}5) = p(z_{mín} < z < z_{máx})$ = área do intervalo = 0,2967 + 0,4515 = 0,7492

9. Estimativa de parâmetros

Com base na média amostral podemos estimar a média populacional μ e ao usarmos a tabela normal teremos um intervalo de probabilidade dessa média populacional estimada estar, isto é o nível de confiança (c). Ou seja, o nível de confiança é a probabilidade p de μ estar entre – z e + z, logo é o dobro da tabela que dá de 0 até z.

Billy, Cartoon Network.



$$\bar{x} - z \cdot \sigma < \mu < \bar{x} + z \cdot \sigma \qquad \sigma = \frac{s}{\sqrt{n}}$$

Tabela t-Student

Quando o número de amostras for pequeno (n < 30) usa-se a tabela desenvolvida por Student para medir o valor **t**, que é a média padronizada para o conjunto. A tabela associa o Grau de Liberdade (GL = n – 1) ao nível de confiança (c). Como usar essa tabela? Exemplo 1: Para n = 5, GL = 5 – 1 = 4, e c de 95%, teremos t = 2,7764. Exemplo 2: Para n = 3, GL = 3 – 1 = 2, e c de 80%, teremos t = 1,8856.

Tabela de t por GL e confiança				
GL	50%	80%	90%	95%
1	1,0000	3,0777	6,3138	12,7062
2	0,8165	1,8856	2,9200	4,3027
3	0,7649	1,6377	2,3534	3,1824
4	0,7407	1,5332	2,1318	2,7764
...	...	...	...	...

Ao invés de usarmos o escore z, no seu lugar usaremos a letra t.

$$z = \frac{x - \mu}{\sigma} \rightarrow t = \frac{\bar{x} - \mu}{\sigma}$$

$$\bar{x} - z \cdot \sigma < \mu < \bar{x} + z \cdot \sigma$$

$$\downarrow$$

$$\bar{x} - t \cdot \sigma < \mu < \bar{x} + t \cdot \sigma$$

Distribuição Qui-Quadrado (X^2)

Um intervalo de confiança c para variância e desvio padrão é o qui-quadrado com limites à esquerda (L, Left) e à direita (R, Right). Assim pode-se entender o tamanho da variabilidade, ou chorar em desespero.

$$\alpha = \frac{1 \pm c}{2} \rightarrow \alpha_R = \frac{1+c}{2} \quad \alpha_L = \frac{1-c}{2} \rightarrow \text{Tabela} \rightarrow \frac{(n-1)s^2}{X_R^2} < \sigma^2 < \frac{(n-1)s^2}{X_L^2}$$

Tabela de X^2 por GL e alfa				
GL	0,99	0,95	0,90	0,10
1	0,002	0,004	0,016	2,706
2	0,020	0,103	0,211	4,605
3	0,115	0,352	0,584	6,251
4	0,297	0,711	1,064	7,779
5	0,554	1,145	1,610	9,236
...	...	...	...	...

Exemplo 9.1. Uma pesquisa com o preço de um remédio para tratamento de impotência sexual coletou uma amostra de 36 valores, com média de R$ 180,00 e desvio padrão de R$ 14,00. Determine a média do valor de toda a população usando um nível de confiança de 95%.

Organizando dados: n = 36 s = 14 $\bar{x}$ = 180 c = 95 %

$$\boldsymbol{\sigma} = \frac{\boldsymbol{s}}{\sqrt{\boldsymbol{n}}} = \frac{14}{\sqrt{36}} = R\$\ 2{,}33$$

Como a tabela mede de 0 a z e 95% corresponde de – z a + z, então usa-se 95/2 = 0,475.

c = 95 % →

z	0,06
1,9	0,4750

→ z = 1,96

$$\bar{\boldsymbol{x}} - \boldsymbol{z} \cdot \boldsymbol{\sigma} < \boldsymbol{\mu} < \bar{\boldsymbol{x}} + \boldsymbol{z} \cdot \boldsymbol{\sigma}$$

$$180 - 1{,}96 \cdot 2{,}33 < \mu < 180 + 1{,}96 \cdot 2{,}33$$

$$\boxed{R\$\ 175{,}43 < \mu < R\$\ 184{,}57}$$

Exemplo 9.2. Um remédio para tratamento de impotência sexual apresenta bons resultados para idades avançadas. Uma amostra de 16 usuários do remédio apresenta média de idade de 75 anos com desvio padrão de 10 anos. Determine o intervalo para a média populacional com nível de confiança de 95%.

Organizando os dados: n = 16 (n < 30, logo t-Student) $\bar{x} = 75$ s = 10 c = 95 %

GL = n – 1 = 16 – 1 = 15

$$\boldsymbol{\sigma} = \frac{\boldsymbol{s}}{\sqrt{\boldsymbol{n}}} = \frac{10}{\sqrt{16}} = 2{,}5$$

Tabela

c = 95 % →

GL	NC 95%
15	2,1315

→ t = 2,1315

$$\bar{\boldsymbol{x}} - \boldsymbol{t} \cdot \boldsymbol{\sigma} < \boldsymbol{\mu} < \bar{\boldsymbol{x}} + \boldsymbol{t} \cdot \boldsymbol{\sigma}$$

$$75 - 2{,}1315 \cdot 2{,}5 < \mu < 75 + 2{,}1315 \cdot 2{,}5$$

$$\boxed{69{,}67 \text{ anos} < \mu < 80{,}33 \text{ anos}}$$

Exemplo 9.3. O efeito colateral de um remédio contra impotência sexual causou a amputação do órgão específico em usuários com diabetes. Uma amostra de 20 usuários apresentou desvio padrão de 2 usuários. Determine o intervalo para a variância e o desvio padrão com nível de confiança de 90%.

Organizando os dados: n = 20 GL = n – 1 = 20 – 1 = 19 s = 2 c = 90 %

$$\alpha = \frac{1 \pm c}{2} = \frac{1 \pm 0{,}90}{2} \rightarrow \begin{cases} \alpha_1 = 0{,}95 \\ \alpha_2 = 0{,}05 \end{cases}$$

	alfa	
GL	0,95	0,05
19	30,144	10,117

$$\rightarrow \begin{cases} X_L^2 = 10{,}117 \\ X_R^2 = 30{,}144 \end{cases}$$

$$\frac{(\boldsymbol{n}-\mathbf{1})\boldsymbol{s}^2}{\boldsymbol{X}_R^2} < \boldsymbol{\sigma}^2 < \frac{(\boldsymbol{n}-\mathbf{1})\boldsymbol{s}^2}{\boldsymbol{X}_L^2}$$

$$\frac{(20-1)\cdot 2^2}{30{,}144} < \sigma^2 < \frac{(20-1)\cdot 2^2}{10{,}117}$$

$$\boxed{2{,}52 < \sigma^2 < 7{,}51}$$

$$\sqrt{2{,}52} < \sigma < \sqrt{7{,}51}$$

$$\boxed{1{,}58 < \sigma < 2{,}74}$$

Interpretando a resposta: a variância populacional varia de 2,52 a 7,51 e o desvio padrão populacional varia de 1,58 a 2,74 usuários do remédio.

10. Teste de hipótese

O teste de hipótese é um método mecânico para se fazer uma tomada de decisão em que uma hipótese será aceita ou rejeitada com base nos elementos amostrais. Definimos a hipótese a ser testada como "hipótese nula" $\mathbf{H_0}$ (nome ridículo...) expressa com igualdade (=) e a "hipótese alternativa" $\mathbf{H_a}$ expressa com desigualdade (≠, <, >) caso a hipótese inicial seja rejeitada.

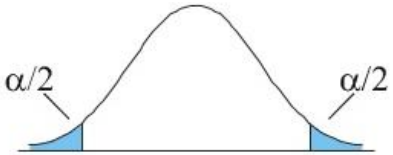
Decisão	H_0 é verdadeira	H_0 é falsa
Aceita-se H_0	Decisão correta	Erro tipo 2
Rejeita-se H_0	Erro tipo 1	Decisão correta

Considere a hipótese de que "estudar para Estatística e Probabilidade fará com que o aluno passe nessa matéria". Caso a hipótese seja verdadeira, se ele estudar passa (decisão correta) e se não estudar não passa (erro). Caso a hipótese seja falsa e não adiantar nem estudar, se ele estudar não passa (erro) e se não estudar também não passa economizando tempo e energia (decisão correta).

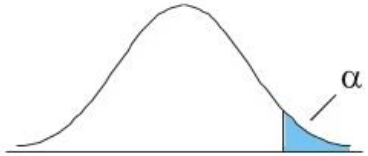
Como é uma hipótese estatística?

Verifica-se se a média populacional (μ) bate ou não com um valor inicialmente suposto para ela (μ_0) dentro de um nível de confiança aceitável, logo o teste é para verificar se este valor esperado está neste intervalo ou não.

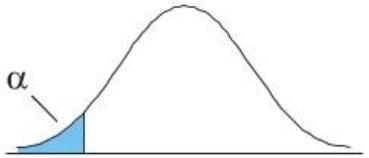
$H_0 : \mu = \mu_0$
$H_a : \mu \neq \mu_0$ Teste bilateral



$H_0 : \mu = \mu_0$
$H_a : \mu > \mu_0$ Teste unilateral à direita



$H_0 : \mu = \mu_0$
$H_a : \mu < \mu_0$ Teste unilateral à esquerda



https://pt.slideshare.net/rodrigomuribec/teste-de-hipteses-paramtricos

Como saber qual desses usar? Depende do problema em questão, você precisa entender o que está acontecendo para escolher qual deles usar.

Nível de significância (α)

Na estatística se coleta dados, organiza dados, analisa dados e se toma decisões em condições de incertezas com um nível de significância. O nível de confiança (c) e o nível de significância (α) são complementares (**c + α = 1**). Se o nível de confiança for de 95% (c = 0,95), o nível de significância de um erro será de 5% (α = 0,05).

Tabela de t_{cr} por GL e significância				
GL	**50%**	**20%**	**10%**	**5%**
1	1,0000	3,0777	6,3138	12,7062
2	0,8165	1,8856	2,9200	4,3027
3	0,7649	1,6377	2,3534	3,1824
4	0,7407	1,5332	2,1318	2,7764
...	...	...	...	...

Tabela de z_{cr} e significância					
α	**10%**	**5%**	**2,5%**	**1%**	**0,5%**
z_{cr}	1,280	1,645	1,960	2,333	2,575

Se a amostra for grande (n ≥ 30) teremos este intervalo em valores de $\mathbf{z_{cr}}$, se a amostra for pequena (n < 30) teremos este intervalo em valores de $\mathbf{t_{cr}}$.

Como fazer a tomada de decisão?

Após definir as hipóteses e o nível de significância é preciso usar a tabela para determinar o valor crítico para tomada de decisão (**z_{cr}** para n ≥ 30 ou **t_{cr}** para n < 30). Então calcula-se z ou t e verifica-se se está no intervalo desejado, tomando cuidado se o intervalo é bilateral, unilateral à esquerda ou unilateral à direita.

	Região de validade (n ≥ 30)	Região de validade (n < 30)
Bilateral	$-z_{cr} \cdot \frac{\alpha}{2} < z < +z_{cr} \cdot \frac{\alpha}{2}$	$-t_{cr} \cdot \frac{\alpha}{2} < t < +t_{cr} \cdot \frac{\alpha}{2}$
Unilateral à direita	$z < +z_{cr}$	$t < +t_{cr}$
Unilateral à esquerda	$z > -z_{cr}$	$t > -t_{cr}$

Valores estatísticos

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

Exemplo 10.1. Uma indústria produz pães de 500 gramas e para verificar se o produto está dentro das especificações ela retira uma amostra de 50 pães com média de 489,42 g e de desvio padrão de 2,55 g. O teste exige 1% de nível de significância. O produto atende as especificações?

Organizando dados: $\mu_0 = 500$ $\bar{x} = 489{,}42$ $n = 50$ $(n \geq 30) \rightarrow z$ $s = 2{,}55$ $\alpha = 1\% = 0{,}01$

$\alpha = 0{,}01 \rightarrow$

α	1%
z_{cr}	2,333

$\rightarrow z_{cr} = 2{,}333$

Definição do teste

$H_0 : \mu = \mu_0 \rightarrow$ Teste bilateral

$H_a : \mu \neq \mu_0$

$$-z_{cr} \cdot \frac{\alpha}{2} < z < +z_{cr} \cdot \frac{\alpha}{2}$$

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

$$-2{,}333 \cdot \frac{0{,}01}{2} < z < +2{,}333 \cdot \frac{0{,}01}{2}$$

$$z = \frac{489{,}42 - 500}{2{,}55/\sqrt{50}}$$

$$\boxed{-0{,}012 < z < +0{,}012}$$

$$\boxed{z = -29{,}34}$$

O valor z = -29,34 está fora do intervalo -0,012 < z < +0,012, portanto se descarta a hipótese H_0 e escolhe H_a na qual a produção não atende as especificações.

Exemplo 10.2. Um fabricante de lâmpadas afirma que a vida média de um tipo de lâmpada é acima de 750 horas com significância de 2,5%. Para verificar esta afirmação uma agência reguladora coleta uma amostra de 16 lâmpadas, obtém média de 735 h e desvio padrão de 20 h. A afirmação do fabricante é verdadeira?

Organizando dados: $\mu_0 = 750$ $\bar{x} = 735$ $n = 16$ $(n < 30) \rightarrow t$ $s = 20$ $\alpha = 2{,}5\ \% = 0{,}025$

$GL = n - 1 = 16 - 1 = 15$

Definição do teste

$H_0 : \mu = \mu_0 \rightarrow$ Teste unilateral à esquerda
$H_a : \mu < \mu_0$

$\alpha = 0{,}025 \rightarrow$

GL	α = 0,025
15	2,132

$\rightarrow t_{cr} = 2{,}132$

$$\boldsymbol{t > -t_{cr}}$$

$$t > -2{,}132$$

$$\boxed{t > -2{,}132}$$

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

$$t = \frac{735 - 750}{20/\sqrt{16}} \rightarrow \boxed{t = -3}$$

O valor t de -3 é menor que o valor crítico de -2,132, portanto a hipótese escolhida é será H_a, em que é falsa afirmação do fabricante.

Puro Osso, Cartoon Network.

Eu ainda não entendi o porquê de tanta crueldade.