

Tópicos em Gestão da Informação II

Aula 02 – Estatística da vida diária

Prof. Dalton Martins
dmartins@gmail.com

Gestão da Informação
Faculdade de Informação e Comunicação
Universidade Federal de Goiás



Documentário

- O prazer da estatística:
 - mostra exemplos fundamentais de uso da estatística na cotidiano de cidades e problemas humanos de uma forma simples e focada em suas aplicações;
 - <https://www.youtube.com/watch?v=xLr68J2yDJ8>

O que é estatística

- É, antes de qualquer coisa, um método científico que determina questões de pesquisa
 - Projeta estudos e experimentos;
 - Coleta, organiza, resume e analisa dados;
 - Interpreta resultados e esboça conclusões.

As estatísticas erram

- É fundamental para um gestor da informação despertar um certo nível de ceticismo perante as informações que lhe chegam a mão: isso está mesmo correto?
- A mídia apresenta muitos exemplos de estatísticas erradas e descobrir alguns desses problemas é um bom treino para se tornar mais confiante com a estatística e desarmar qualquer explosão de números!

Detectando erros, exageros e mentirinhas leves

- A primeira coisa a fazer é se perguntar se o resultado está de fato correto e procurar investigar;
- Algumas dicas:
 - Verifique se a soma está correta → as porcentagens somam mesmo 100%. A soma do número de pessoas em cada categoria resulta no número de pessoas entrevistadas;
 - Verifique duas vezes até os cálculos mais simples;
 - Sempre procure por um total para que você possa colocar os resultados em perspectiva adequada;
 - Ignore os resultados baseados em amostrar muito pequenas!

Exemplos



Fonte: Jornal Folha de S. Paulo de 30/08/2008.



Fonte: Jornal O Estado de S. Paulo de 30/08/2008.

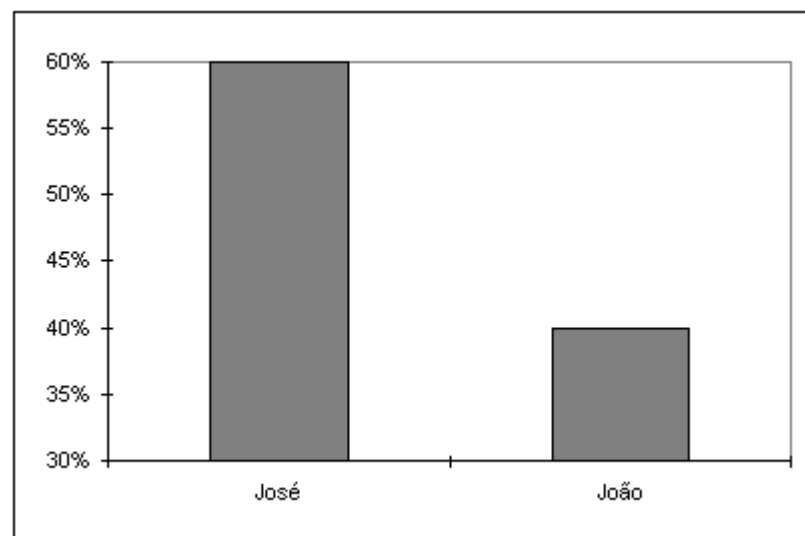
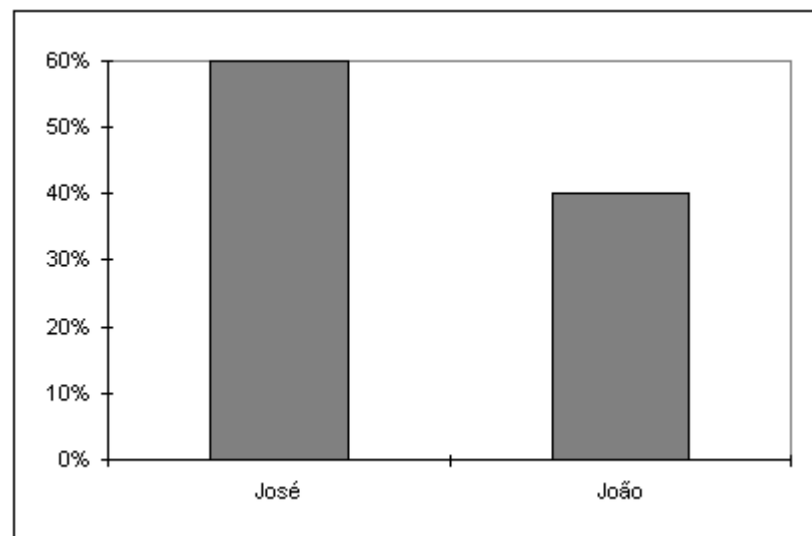
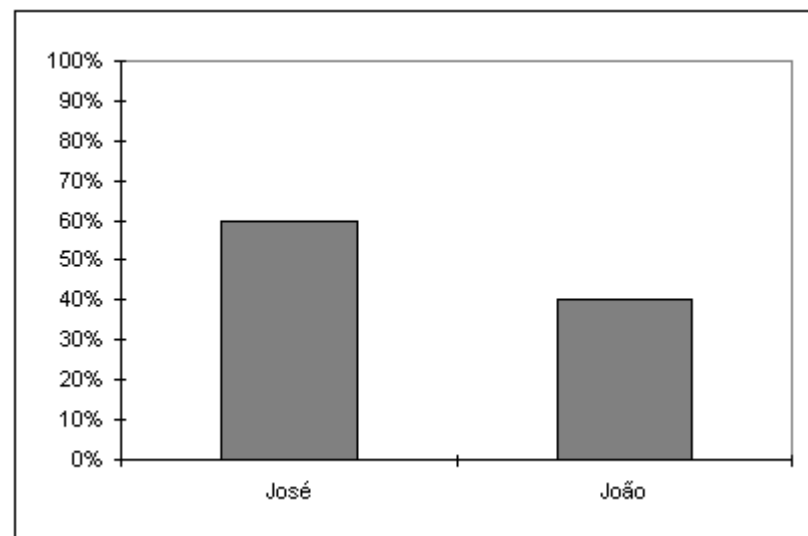
Obs: margem de erro de 2% nas pesquisas, sendo uma do Ibope e outra do DataFolha.

Exemplos

Tabela 1 - Intenção de voto em 15-11-1997

Candidato	Intenção de voto
José	60%
João	40%

Fonte: IBPE



Exemplos

cotidiano

Governo de SP revê estatísticas de crimes e 'descobre' 459 roubos

GIBA BERGAMIM JR.
DE SÃO PAULO

08/01/2014 ☉ 03h15

 Recomendar < 428

 Tweetar < 25

 +1 33

 OUVIR O TEXTO

 Mais opções

Após identificar erros nas estatísticas sobre crimes publicadas em 25 de dezembro, a Secretaria de Segurança Pública de São Paulo registrou 281 roubos a mais na capital paulista em novembro de 2013 e outros 178 no mesmo mês de 2012. Nos dois meses, são 459 novos crimes.

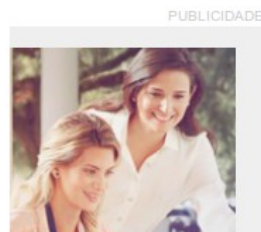
Os novos casos representam 2% dos 8,4 mil roubos de novembro de 2012 e 3% dos 10,2 mil registros de novembro de 2013 –sexto mês seguido em que houve crescimento.

Os números são residuais e alteram ligeiramente o percentual de aumento dos roubos em toda a cidade –em vez de 19,6%, o correto é 20%.

As incorreções foram detectadas após questionamentos, no dia seguinte à divulgação, feitos pela **Folha**, que apontou problemas como distritos policiais com apenas um caso –o de Sapopemba (zona leste)– ou alta surpreendente de até 15.400 % –o dos Jardins (zona oeste).

Na noite de anteontem, as informações foram atualizadas no site da pasta.

Foram corrigidas estatísticas de 5 dos 93 distritos policiais da cidade. A pedido da reportagem, a pasta está analisando dados de outras delegacias.



Revelando estatísticas enganosas

- As estatísticas que se referem à criminalidade são um grande exemplo de como a estatística é utilizada para mostrar os dois lados de uma história, em que apenas um é verdadeiro;
- Como dois políticos podem argumentar o aumento e a diminuição da criminalidade ao mesmo tempo para um mesmo local?

Exemplo

Número de crimes nos EUA

Ano	Crimes
1987	13.508.700
1988	13.923.100
1989	14.251.400
1990	14.475.600
1991	14.872.900
1992	14.438.200
1993	14.144.800
1994	13.989.500
1995	13.862.700
1996	13.493.900
1997	13.175.100

Entre 1987 e 1993 parece que o índice de criminalidade aumentou nos EUA.

Uma informação fundamental foi deixada de fora da história aqui. Outra coisa importante aumentou também no período → a população!

Exemplo

Número de crimes nos EUA

Ano	Crimes	População	Taxa de Crimes
1987	13.508.700	243.400.000	5.550,00
1988	13.923.100	245.807.000	5.664,24
1989	14.251.400	248.239.000	5.741,00
1990	14.475.600	248.710.000	5.820,27
1991	14.872.900	252.177.000	5.897,80
1992	14.438.200	255.082.000	5.660,22
1993	14.144.800	257.908.000	5.484,44
1994	13.989.500	260.341.000	5.373,53
1995	13.862.700	262.755.000	5.275,90
1996	13.493.900	265.284.000	5.086,59
1997	13.175.100	267.637.000	4.922,75

Entre 1987 e 1993 a taxa de crimes diminuiu e não aumentou!

O que é melhor para avaliar esse aspecto representado pelos dados anteriores: o número absoluto de eventos ou a taxa relativa ao número da população?

Razões, taxas e porcentagens

- Razão:
 - É uma fração que divide duas quantidades;
 - Exemplo: a razão de meninas por meninos é de 3 para 2;
- Taxa:
 - É uma razão que reflete uma quantidade por certa unidade;
 - Exemplo: seu carro faz 60 km por hora em média;
- Porcentagem:
 - É um número entre 0 e 100 que reflete a proporção do todo.
 - Para encontrar o aumento/diminuição de porcentagem entre dois valores em dois períodos diferentes, faça
 - A quantidade de “depois” menos a quantidade “antes” e divida o resultado pela quantidade de “antes”.

Exercício

Ano	Crimes	População	Taxa de Crimes
1987	13.508.700	243.400.000	5.550,00
1988	13.923.100	245.807.000	5.664,24
1989	14.251.400	248.239.000	5.741,00
1990	14.475.600	248.710.000	5.820,27
1991	14.872.900	252.177.000	5.897,80
1992	14.438.200	255.082.000	5.660,22
1993	14.144.800	257.908.000	5.484,44
1994	13.989.500	260.341.000	5.373,53
1995	13.862.700	262.755.000	5.275,90
1996	13.493.900	265.284.000	5.086,59
1997	13.175.100	267.637.000	4.922,75

1. Quanto variou a taxa de crimes entre 1987 e 1993?
2. E entre 1987 e 1997?

Sintetizando dados com estatísticas

- Todos os conjuntos de dados têm uma história e, se usadas de maneira apropriada, as estatísticas são a melhor maneira de contar essa história;
- As estatísticas são usadas para sintetizar algumas informações mais básicas dentro de um conjunto de dados:
 - Permite entendermos o que, de maneira geral, está acontecendo nos dados;
 - Somente depois vamos formular algumas hipóteses e procurar relações entre os itens medidos.

Tipos de dados que podem ser sumarizados

- Dados categóricos:
 - Coletam qualidades ou características a respeito de um indivíduo, como a cor dos olhos de uma pessoa, o sexo, o partido político ou a opinião sobre determinado assunto;
 - Os dados categorizados são frequentemente resumidos por meio das porcentagens de indivíduos que se enquadram em cada uma das categorias definidas;
 - Ainda podemos utilizar tabelas cruzadas, que resumem os dados utilizando duas variáveis categóricas na linha e coluna da tabela, tais como sexo e partido político para observarmos a distribuição dos indivíduos entre essas categorias;
- Dados numéricos:
 - São características mensuráveis, tais como altura, peso, idade ou renda financeira;
 - Pelo fato de terem significado numérico, podem ser resumidos de várias formas;
 - Determinadas características de dados numéricos podem ser descritas usando as estatísticas, tais como localização do centro, a maneira como os dados estão distribuídos e a localização de alguns marcos.

Dados categóricos

exemplo de tabela cruzada

Estudo sobre preferência de programa de aprendizado.

Escola	Período	Preferência de Aprendizado			Totais
		Individual	Grupo	Sala Aula	
1	Padrão	10	17	26	53
1	Integral	5	12	50	67
2	Padrão	21	17	26	64
2	Integral	16	12	36	64
3	Padrão	15	15	16	46
3	Integral	12	12	20	44

Fonte: Stokes et al. (2000)

Dados categóricos

exemplo de tabela cruzada

Banco do Nordeste do Brasil: taxa de crescimento real do saldo das operações de crédito por estado e atividade no Nordeste, média anual 2004-2011 (%)					
Estados	Total	Rural	Industrial	Comercial	Serviços
Alagoas	6,97	4,13	9,61	19,49	*
Bahia	10,89	4,78	17,27	39,61	44,49
Ceará	18,94	8,88	20,49	62,26	46,23
Maranhão	18,92	17,62	21,16	40,97	43,24
Paraíba	11,02	-0,55	12,83	46,76	51,23
Pernambuco	23,25	6,54	48,94	57,68	79,25
Piauí	1,54	-1,3	-1,79	29,32	35,97
Rio Grande do Norte	18,30	4,00	35,34	51,21	70,04
Sergipe	12,36	9,00	21,66	37,64	**
Nordeste	13,60	6,00	18,99	43,87	46,85

Fonte: Ministério do Planejamento/DEST. Elaboração Araújo (2013, p. 190); (*) Não possui saldo a partir de 2007; (**) não possui saldo a partir de 2008.

Dados numéricos

- Chegando ao centro:
 - O modo mais comum de sintetizar um conjunto de dados é descrever onde está o centro;
 - O centro de um conjunto de dados pode ser medido de diferentes formas e o método escolhido pode influenciar muito nas conclusões que podem ser tiradas;
 - As formas mais comuns de representar o centro de um conjunto de dados é usando a **média** e a **mediana**.

Média

- O que fazer para encontrar a média para um conjunto de dados:
 - Some todos os números do conjunto de dados;
 - Divida pelo número de números do conjunto de dados, n .
- Pela forma como a média é calculada os valores muito discrepantes do conjunto, sejam mais altos ou mais baixos tendem a “puxar” a média para cima ou para baixo, distorcendo o resultado.

Exemplo

Cargos	Salários
Presidente	R\$ 20.000,00
Diretor	R\$ 8.000,00
Engenheiro	R\$ 2.000,00
Estagiário	R\$ 800,00

Média: R\$ 7.700,00

Mediana

- A mediana de um conjunto de dados é o valor que se encontra exatamente no meio
- O que fazer para encontrar a mediana para um conjunto de dados:
 - Ordene os números do menor para o maior;
 - Se o conjunto de dados possuir um número ímpar de números, escolha o número que estiver exatamente no meio → este é a mediana;
 - Se o conjunto de dados possuir um número par de números, pegue os dois números que estiverem exatamente no meio e faça a média deles para encontrar a mediana.

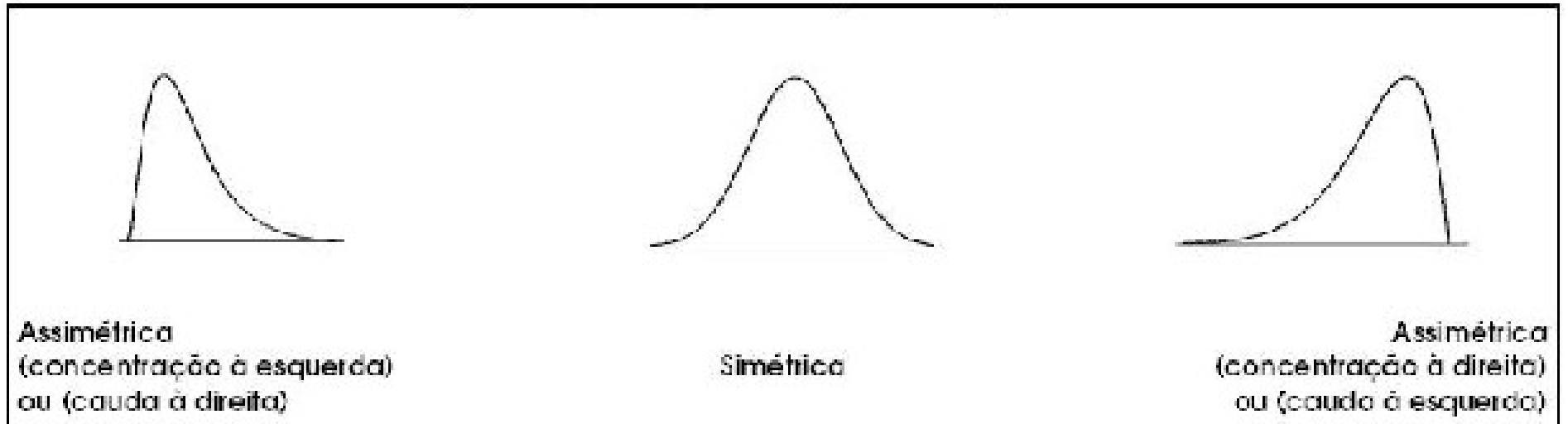
Exemplo

Cargos	Salários
Presidente	R\$ 20.000,00
Diretor	R\$ 8.000,00
Engenheiro	R\$ 2.000,00
Estagiário	R\$ 800,00

Mediana: R\$ 5.000,00

A mediana é menos influenciada por valores tão discrepantes nas extremidades.

Distorções



Se os valores forem discrepantes na extremidade inferior, **muito pequenos**, a curva será assimétrica para a **esquerda**.

Se os valores forem discrepantes na extremidade superior, **muito grandes**, a curva será assimétrica para a **direita**.

Distorções

- A média de um conjunto de dados é influenciada por valores discrepantes;
- Isso não acontece com a mediana;
- Se alguém lhe apresentar a média de um conjunto de dados, pergunte também pela mediana:
 - Isso te dará uma visão aproximada se há muita ou pouca distorção na distribuição dos dados.

Contabilizando variação

- A variabilidade é o que faz a estatística ser como é → sempre existe variação em um conjunto de dados;
- O desvio padrão representa de longe a medida mais utilizada para calcular variabilidade:
 - Representa a distância normal de qualquer ponto no conjunto de dados até o centro.
 - Ele é, aproximadamente, a distância média até o centro. No caso, o centro é a média.

Desvio padrão

- Como obter o desvio padrão:
 - Encontre a média do conjunto de dados;
 - Subtraia o valor da média para cada um dos números;
 - Eleve ao quadrado cada resultado obtido no passo anterior;
 - Some todos os resultados obtidos no passo anterior;
 - Divida a soma pela quantidade de números no conjunto de dados menos 1 ($n-1$);
 - Descubra a raiz quadrada do número resultante.

Desvio padrão

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

Interpretando o desvio padrão

- Um desvio padrão pequeno, basicamente, significa que os valores do conjunto de dados estão, na média, próximos do centro desse conjunto;
- Um desvio padrão grande significa que os valores do conjunto de dados estão, na média, mais afastados do centro.

Exemplo I

Cargos	Salários
Presidente	R\$ 20.000,00
Diretor	R\$ 8.000,00
Engenheiro	R\$ 2.000,00
Estagiário	R\$ 800,00

Desvio padrão: R\$ 8.784,08

Exemplo II

Cargos	Salários
Presidente	R\$ 3.000,00
Diretor	R\$ 2.000,00
Engenheiro	R\$ 1.500,00
Estagiário	R\$ 1.000,00

Desvio padrão: R\$ 853,91

A importância do desvio padrão

- O desvio padrão não aparece muito na mídia e isso é um grande problema;
- Se você descobrir onde fica o centro dos dados sem saber o quanto tais dados variam, você tem apenas uma parte da história;
- A variedade é o tempero da vida, assim, sem uma indicação de quanto diversificado e variado são os dados, você não saberá o verdadeiro tempo dos dados!
- Sem o desvio padrão não é possível comparar dois conjuntos de dados de maneira eficiente.

A importância do desvio padrão

- Só saber a média e a mediana não é o suficiente;
- Por exemplo, os seguintes conjuntos possuem a mesma média e mediana:
 - 199,200,201
 - 0,200,400
- Porém, sua variabilidade é bastante diferente:
 - O desvio padrão do primeiro conjunto é 1;
 - Do segundo é 200!!!

Exercícios

- Baixe o arquivo na página do curso chamado:
 - <https://gestaodainformacao.fic.ufg.br/pages/66915-dalton-lopes-martins>
 - Auxílio a Pesquisadores – FAPEG 2013.ods
- Faça:
 - A organização dos dados em formato moeda;
 - Ordene do maior para menor;
 - Encontre a somatória das colunas;
 - Encontre a média, mediana e desvio padrão para os dados;
 - Analise os resultados: o que eles dizem!