Probabilidade e Estatística

Educação a Distância
Cruzeiro do Sul Educacional
Campus Virtual

Material teórico

Medidas de Dispersão ou Variação

Responsável pelo Conteúdo:
Profª Ms. Rosangela Maura C. Bonici

UNIDADE MEDIDAS DE DISPERSÃO OU VARIACÃO

- Introdução ao Conteúdo
- Cálculo da Variância e Desvio Padrão

Objetivo de Aprendizado

A proposta deste estudo é trabalhar com as medidas de dispersão.

Nela você irá aprender como calcular a variância e o desvio-padrão de dados brutos, de dados agrupados em distribuições de freqüência variável discreta e de dados agrupados em distribuições de freqüência variável contínua. Aprenderá também qual o significado do desvio-padrão e em quais situações práticas ele poderá ser empregado.

.

Atenção

Para um bom aproveitamento do curso, leia o material teórico atentamente antes de realizar as atividades. É importante também respeitar os prazos estabelecidos no cronograma.

Contextualização

Entenda o que é média e desvio padrão de uma prova

O entendimento natural que grande parte dos candidatos utiliza para avaliar seus desempenhos nas provas é: "acertei mais ou menos questões do que a média?". Claro que a premissa vigente é a de que os candidatos mais bem preparados superam, em número de acertos, a média das provas.

0 resultado de uma prova, normalmente, é conhecido através de informações como média e desvio padrão, bem como pela distribuição de frequências do número de acertos dos candidatos, expressos em forma gráfica. Este gráfico, denominado histograma, mostra, no eixo dos X (abscissas), o número de questões e, no eixo dos Y, o número de candidatos que acertaram o referido número de questões. O que significam essas informações?

Numa distribuição de frequências, há três medidas importantes: a moda, a mediana e a média. A primeira é o "pico", isto é, o ponto no eixo das abscissas de maior frequência. A mediana é o ponto, no eixo das abscissas, que divide as ocorrências em duas frações iguais, cada uma com 50% das frequências. A média é o ponto, no eixo das abscissas, que faria com que o gráfico ficasse equilibrado, não inclinando nem para a esquerda nem para a direita; em suma, a média é o ponto, no eixo das abscissas, situado na vertical que passa pelo centro de gravidade da figura.

O que se deseja em uma prova do Concurso Vestibular é um histograma formando uma curva simétrica, distribuídos entre 0 e 25 acertos, concentrando moda, mediana e média próximas a 15 acertos, exibindo uma distribuição balanceada de acertos, tanto à esquerda como à direita do centro de distribuição, de acordo com o gráfico apresentado na figura da página a seguir.

O gráfico mostra uma distribuição normal rigorosamente simétrica. No centro da distribuição, coincidem média, mediana e moda. Uma curva de distribuição normal (ou Curva de Gauss) tem como característica englobar 99,73% das ocorrências no intervalo compreendido entre a média e ± 3 desvios padrão, conforme detalhado no mesmo gráfico.

O desvio padrão de uma prova mede o grau de dispersão dos candidatos em relação à média, isto é, o quanto o conjunto de candidatos se distanciou da média, tanto além como aquém do centro de distribuição. Isso significa que os escores obtidos por 99,73% dos candidatos estarão compreendidos entre a média e ± 3 desvios padrão, ou seja, salvo raras exceções, todos os candidatos estarão neste intervalo.

Ao se elaborar uma prova, espera-se que o resultado da aplicação da mesma gere uma "curva de distribuição normal", isto é, essa prova deve gerar uma média de 15 acertos e os candidatos devem estar distribuídos simetricamente entre zero e 30 acertos (ou entre dois limites internos desse intervalo, equidistantes de 15).

A obtenção de uma distribuição simétrica com 100% das ocorrências entre 0 e 30 acertos, em uma prova de 30 questões e média de 15 acertos, pode ser possível quando se obtém um desvio padrão de 5 acertos. Nesse caso, se o histograma assumir formato semelhante ao da curva normal, todos os escores possíveis de serem obtidos pelos candidatos ficariam simetricamente contidos no intervalo entre a média e ± 3 desvios padrão, ou seja, entre O acertos (15 - (3 x 5)) e 30 acertos (15 + (3 x 5)).

Infelizmente, não é fácil obter uma curva de distribuição normal. O resultado obtido pelos candidatos em uma prova depende de muitos fatores, entre os quais podem ser destacados a preparação dos mesmos e o grau de dificuldade da prova. Por isso mesmo, é importante poder avaliar uma prova através do resultado obtido na sua aplicação.

A informação da média permite verificar o grau de facilidade da prova para a população que a realizou. Quanto menor a média (abaixo de 15 acertos), menor a facilidade dos candidatos com as questões da mesma. Quanto maior a média (acima dos 15 acertos), maior a facilidade dos candidatos com as questões propostas.

Uma prova, com histograma normal, com média de 15 acertos e desvio padrão de 4 acertos significa que 99,73% dos candidatos serão encontrados entre os escores 3 e 27 acertos, (15 - (3 x 4)) e (15 + (3 x 4)). Isso significa que os candidatos com escores 0, 1, 2, 28, 29 e 30 acertos serão 0,23% da população, portanto pouquíssimos, conforme mostrado na figura a seguir correspondente à distribuição com desvio 4.

Enquanto que numa prova, também com histograma normal, com média de 15 acertos, mas desvio padrão de 2 acertos significa que 99,73% dos candidatos serão encontrados entre os escores 9 e 21 acertos, (15 - (3 x 2)) e (15 + (3 x 2)). Isto é, também haverá poucos candidatos (0,23%) com escores de 0 até 8 e de 22 até 30 acertos, conforme mostrado na figura correspondente à distribuição com desvio 2.

Analisando as duas figuras, que seguem, é possível concluir que, quanto maior for o desvio padrão, mais aberta é a curva (maior dispersão), ou seja, maior variedade de escores obtidos pelos candidatos e melhores condições de discriminar a qualificação dos candidatos. Curvas muito fechadas (pequeno desvio padrão) significam menor dispersão, ou seja, grande concentração de escores e menor variedade dos mesmos (muitos empates). Em outras palavras, se houver muitos empates, a prova poderá não avaliar devidamente a preparação dos candidatos. Ao mesmo tempo, provas muito difíceis não diferenciam escores obtidos unicamente através de acerto casual.

Fonte: Vestibular da UFRGS 2005

Provas Comentadas - Processo de Avaliação

Publicado pela COPERSE - UFRGS

Universidade federal do Rio Grande do Sul. Disponível em:

<http://www.universitario.com.br/noticias/noticias_noticia.php?id_noticia=6560>. Acesso em: 10 Set. 2009

1 Introdução

As medidas de variação ou dispersão, avaliam a dispersão ou a variabilidade da sequência numérica em análise, são medidas que fornecem informações complementares à informação da média aritmética. As principais medidas de dispersão são: a variância e o desvio-padrão.

Usaremos as letras $\mathbf{s^2}$ para denotar a variância de uma amostra e **s** para denotar o seu desvio-padrão.

2 Cálculo da Variância e Desvio Padrão

Para calcular a variância e o desvio-padrão vamos analisar três casos:

i) Quando os dados ainda não foram agrupados em tabelas de frequência, ou seja, estão na forma de dados brutos ou rol;
ii) Quando os dados estão agrupados em distribuições de frequência variável discreta e
iii) Quando os dados estão agrupados em distribuições de frequência variável contínua.

2.1 Dados brutos ou rol

Para podermos calcular a variância e o desvio padrão de dados brutos vamos usar as fórmulas que seguem:

Fórmula para o Cálculo da Variância de Dados Brutos

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

s^2 = Variância
xi = Cada um dos valores assumidos pela variável
$\bar{x}$ = Média aritmética dos dados brutos
n = Total de elementos observados

Fórmula para o Cálculo do Desvio-Padrão de Dados Brutos

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

s = Desvio-padrão
s^2 = Variância

Vejamos um exemplo de utilização da variância e desvio-padrão. Calcule a variância e o desvio padrão das notas de três turmas de estudantes.

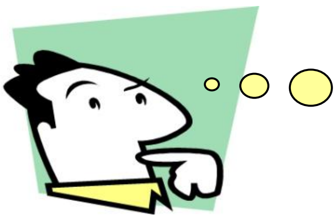
Quadro 1 – Notas de estudantes das Turmas A, B e C

Turma	Notas dos alunos								Média	Desvio-Padrão
A	4	5	5	6	6	7	7	8	6	**1,31**
B	1	2	4	6	6	9	10	10	6	**3,51**
C	**0**	**6**	**7**	**7**	**7**	**7,5**	**7,5**		**6**	**2,69**

Observe no quadro que a média e o desvio-padrão das notas já estão calculados. Vamos ver como isso foi feito.

O desvio-padrão da turma A foi calculado da seguinte forma:

1º) Determinar é a média aritmética das notas, pois a variância depende dela. Como são dados brutos vamos relembrar a fórmula para calculo da média



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

Usando as notas da turma A para fazer os cálculos temos:

$$\bar{x} = \frac{4+5+5+6+6+7+7+8}{8}$$

$$\bar{x} = \frac{48}{8} = 6$$

Concluímos que a média aritmética das notas vale **6.**

2º) Vamos calcular a variância das notas da turma A, para isso vamos usar a fórmula para o cálculo da variância de dados brutos que é

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

Vamos entender o que a fórmula está dizendo...

$\sum (x_i - \bar{x})^2$ (faça a diferença entre cada nota e a média aritmética e eleve ao quadrado, depois some cada uma dessas diferenças)

Depois divida o valor que encontrou pelo total de notas menos 1

Turma	Notas dos alunos							
A	**4**	**5**	**5**	**6**	**6**	**7**	**7**	**8**

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{(4-6)^2 + (5-6)^2 + (5-6)^2 + (6-6)^2 + (6-6)^2 + (7-6)^2 + (7-6)^2 + (8-6)^2}{8-1}$$

$$= \frac{4+1+1+0+0+1+1+4}{7} = \frac{12}{7} \cong 1{,}71$$

Temos que a variância das notas vale 1,71

3º) Vamos calcular o desvio-padrão das notas vamos usando a fórmula:

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

Substituindo a variância na fórmula e fazendo o cálculo temos:

$$s = \sqrt{1{,}71} = 1{,}31$$

Temos que o desvio-padrão vale 1,31.

Para calcular o desvio-padrão das turmas B e C foi procedido da mesma forma.

Considerações

Quadro 1 – Notas de estudantes das Turmas A, B e C

Turma	Notas dos alunos							Média	Desvio-Padrão
A	4	5	5	6	6	7	7	8	6	**1,31**
B	1	2	4	6	6	9	10	10	6	**3,51**
C	**0**	**6**	**7**	**7**	**7**	**7,5**	**7,5**		**6**	**2,69**

Observando o quadro1, podemos fazer as seguintes considerações:

As notas que geraram média 6 nas três turmas são bastante diferentes.

Os desvios-padrão são bem diferentes. O menor está na turma A, o intermediário na turma C e o maior na turma B.

O desvio-padrão nos mostra a variabilidade dos dados em relação à média. A grosso modo dizemos que o desvio-padrão nos mostra se a média aritmética sofreu pouca ou muita influência dos valores extremos (muito grandes ou muito pequenos). Nesse caso podemos afirmar que:

A turma A foi a menos influenciada por valores extremos

A turma C foi medianamente influenciada por valores extremos

A turma B foi a mais influenciada por valores extremos.

2.2 Distribuição de frequência variável discreta

Para calcular a variância e o desvio-padrão de uma distribuição de frequência variável discreta vamos usar as fórmulas a seguir:

Fórmula para o Cálculo da Variância da Distribuição de Frequência Variável Discreta

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

s^2= Variância
xi = Cada um dos valores assumidos pela variável
fi = freqüência absoluta
$\bar{x}$ = Média aritmética da variável discreta
$\sum fi - 1 =$ Soma do total de elementos observados menos 1

Fórmula para o cálculo do desvio-padrão

Distribuição de frequência variável discreta

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

s = Desvio-padrão
s^2= Variância

Vejamos um exemplo: O quadro 2 representa as notas de Matemática, calcule a variância e o desvio-padrão.

Quadro 2 – Notas de Matemática

Notas de Matemática (xi)	fi
2	3
3	5
4	8
5	4
Totais	20

As notas de Matemática estão agrupadas em uma distribuição de frequência variável discreta, para calcular a variância e o desvio-padrão temos que usar as fórmulas correspondentes.

1º Vamos calcular a variância usando a fórmula

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

Vamos entender o que ela significa

$\sum (x_i - \bar{x})^2 . f_i =$ devemos subtrair cada nota da média aritmética. Esse resultado deve ser elevado ao quadrado. Depois deve ser multiplicado pela respectiva frequência. Ao final fazer o somatório desses valores

$\sum fi - 1 =$somar o total de notas e subtrair 1

Primeiro, devemos calcular a média aritmética.
Para podermos,
depois usar a fórmula da variância

Lembra-se da fórmula da média aritmética ponderada? É ela que iremos usar!

$$\bar{X} = \frac{\sum xi . fi}{\sum fi}$$

Vamos usar a distribuição das notas de Matemática e abrir uma coluna para podermos multiplicar xi por fi e calcular a média

Notas de Matemática (xi)	fi	xi.fi
2	3	**2.3 = 6**
3	5	**3.5 = 15**
4	8	**4.8 = 32**
5	4	**5.4 = 20**
Totais	**20**	**73**

Calculando a média temos:

$$\overline{X} = \frac{\sum xi.fi}{\sum fi} = \frac{73}{20} = 3{,}65$$

A média aritmética das notas de Matemática é 3,65

Vamos calcular agora a Variância usando a fórmula. Para podermos fazer $\sum (x_i - \overline{x})^2 . f_i$, vamos abrir uma nova coluna na distribuição de frequência das notas de Matemática, para poder facilitar nossos cálculos

Notas de Matemática (xi)	**fi**	**$(xi - \overline{x})^2 . fi$**
2	**3**	**$(2 - 3{,}65)^2 . 3 = 8{,}17$**
3	**5**	**$(3 - 3{,}65)^2 . 5 = 2{,}11$**
4	**8**	**$(4 - 3{,}65)^2 . 8 = 0{,}98$**
5	**4**	**$(5 - 3{,}65)^2 . 4 = 7{,}29$**
Totais	**20**	**18,55**

Concluímos daí que $\sum (x_i - \overline{x})^2 . f_i$ vale 18,55, completando a resolução.

$$\sum fi - 1 = 20\text{-}1 = 19$$

Calculando temos $\frac{18{,}55}{19} = 0{,}98$

A variância das notas de Matemática vale 0,98

2º Vamos calcular o desvio-padrão usando a fórmula

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

s = Desvio-padrão
s^2= Variância

$s = \sqrt{0{,}98} \Rightarrow s = 0{,}99$ *(desvio-padrão)*

Considerações

Podemos concluir pelos cálculos que o desvio-padrão vale 0,99, o que nos demonstra uma variabilidade pequena nas notas de Matemática.

2.3 Cálculo da variância e do desvio-padrão da distribuição de frequência variável continua

Para calcular a variância e o desvio-padrão de variáveis continuas devemos proceder como para as variáveis discretas, tomando somente o cuidado de substituir o **xi pelos pontos médios de cada classe,** uma vez que a variável está agrupada com intervalos de classe.

Fórmula para o cálculo da variância da distribuição de frequência variável contínua

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

s^2= Variância
xi = Cada um dos valores assumidos pela variável
fi = freqüência absoluta
$\bar{x}$ = Média aritmética da variável discreta
$\sum fi - 1 =$ Soma do total de elementos observados menos 1

Fórmula para o cálculo do desvio-padrão

Distribuição de frequência variável contínua

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

s = Desvio-padrão
s^2= Variância

Vamos ver um exemplo: O quadro 3, representa um banco de horas de uma pequena empresa. Calcule a variância e o desvio-padrão.

Quadro 3 – Banco de horas dos empregados de uma empresa

Banco de horas (h)	fi
0 |- 4	1
4 |- 8	3
8 |- 12	5
12 |- 16	1
Total	10

1º) Para calcular a variância a primeira coisa que temos que conhecer é a média aritmética desse banco de horas, caso contrario, não tem como usar a fórmula da variância.

Lembra-se da fórmula da média aritmética ponderada?
É ela que iremos usar!

$$\overline{X} = \frac{\sum xi.fi}{\sum fi}$$

Na variável contínua para podermos calcular a média temos que fazer aparecer os **xi, calculando o ponto médio** entre cada uma das horas. Para isso vamos abrir uma coluna para distribuição para colocar o ponto médio e outra para podermos multiplicar xi por fi.

Banco de horas (h)	**fi**	**xi (ponto médio)**	**xi . fi**
0 |- 4	1	$\frac{0+4}{2} = 2$	2.1 = 2
4 |- 8	3	$\frac{4+8}{2} = 6$	3.6 =18
8 |- 12	5	$\frac{8+12}{2} = 10$	5.10 = 50
12 |- 16	1	$\frac{12+16}{2} = 14$	1.14 =14
Total	**10**	**-**	**84**

$\overline{X} = \frac{\sum xi.fi}{\sum fi} \Rightarrow \overline{x} = \frac{84}{10} = 8{,}4$ **Temos que a média do banco de horas é 8,4 h**

Agora sim, estamos em condições de calcular a variância

$$s^2 = \frac{\sum (x_i - \overline{x})^2 . fi}{\sum fi - 1}$$

Vamos usar a distribuição e abrir uma coluna para podermos calcular

$$\sum (x_i - \overline{x})^2 . fi$$

Banco de horas (h)	**fi**	**xi (ponto médio)**	**$(xi - \overline{x})^2$. fi**
0 |- 4	1	2	$(2 - 8{,}4)^2 . 1 = 40{,}96$
4 |- 8	3	6	$(6 - 8{,}4)^2 . 3 = 17{,}28$
8 |- 12	5	10	$(10 - 8{,}4)^2 . 5 = 12{,}80$
12 |- 16	1	14	$(14 - 8{,}4)^2 . 1 = 31{,}36$
Total	10	-	102,4

Temos que $\sum (x_i - \bar{x})^2 . fi = 102{,}4$ *e* $\sum fi - 1 = 10 - 1 = 9$

Aplicando os valores na fórmula vem: $\boxed{s^2 = \dfrac{\sum (x_i - \bar{x})^2 . fi}{\sum fi - 1}} \Rightarrow s^2 = \dfrac{102{,}4}{9} = 11{,}38$

Chegamos à conclusão de que a variância vale 11,38

2º Agora vamos calcular o desvio-padrão usando $s = \sqrt{s^2}$

Substituindo os valores temos: $s = \sqrt{11{,}38} = 3{,}37$

Considerações

Feitos os cálculos verificamos que a variância do banco de horas é 3,37, o que demonstra uma variabilidade média nas horas.

NOTA

Quanto maior o desvio-padrão maior a variação ou dispersão dos dados

Quanto menor o desvio-padrão, menor a variação ou dispersão dos dados

Finalizando

Finalizamos mais uma Unidade onde aprendemos a calcular a variância e calcular e interpretar o desvio-padrão.

Como vimos, o desvio-padrão fornece informações que complementam a informação da média aritmética, mostrando se a variação dos dados que geraram a média aritmética é pequena, média ou grande.

Só conseguimos identificar se um desvio-padrão é pequeno ou grande se tivermos dois conjuntos que tenham médias iguais para podermos comparar seus desvios-padrão.

Estou confiante e tenho certeza que vocês conseguiram acompanhar e que estão satisfeitos por terem conseguido vencer mais essa etapa.

Agradeço a todos, continuem se esforçando sempre e até a próxima!

Um forte abraço

Material Complementar

UOL educação. Disponível em

<http://educacao.uol.com.br/matematica/ult1705u77.jhtm>.

Infoescola. Disponível em**:** < http://www.infoescola.com/estatistica/variancia-e-desvio-padrao/>.

Vídeo sobre cálculo da media e do desvio-padrão. Disponível em:

<http://www.youtube.com/watch?v=8X9apoqlbgs>.

Site da ADVFN. Disponível em <http://wiki.advfn.com/pt/Desvio_Padr%C3%A3o>.

Referências

CRESPO A. A. Estatística Fácil, 11ª Ed. São Paulo: Saraiva, 1994.

DOWNING, D. **Estatística Aplicada**, 2ª Ed. São Paulo: Saraiva, 2002.

MORETTIN, L.G. **Estatística Básica,** 7ª Ed. São Paulo: Pearson, 2000.

NEUFELD, J.L.**Estatística Aplicada a Administração Usando o Excel**. São Paulo: Pearson, 2003.

SPIEGEL, M.R. **Estatística**, 3ª Ed. Coleção Schaum. São Paulo: Pearson, 1994.

SPIEGEL, M.R. **Probabilidade e Estatística**. Coleção Schaum. São Paulo: Pearson, 1977.

SILVA, E.M.,**Estatística Para os Cursos de**; Economia, Administração e Ciências Contábeis. 3ª Ed. São Paulo:Atlas, 1999.

Anotações

Educação a Distância
Cruzeiro do Sul Educacional
Campus Virtual

www.cruzeirodosulvirtual.com.br
Campus Liberdade
Rua Galvão Bueno, 868
CEP 01506-000
São Paulo SP Brasil
Tel: (55 11) 3385-3000

Universidade
Cruzeiro do Sul

Módulo
Centro
Universitário

UDF
Centro
Universitário

UNICID
Universidade
Cidade de S. Paulo