UNIVERSIDADE ESTADUAL DE CAMPINAS - UNICAMP
INSTITUTO DE FILOSOFIA E CIÊNCIAS HUMANAS - IFCH
DEPARTAMENTO DE ECONOMIA E PLANEJAMENTO ECONÔMICO - DEPE
CENTRO TÉCNICO ECONÔMICO DE ASSESSORIA EMPRESARIAL - CTAE

ELEMENTOS DE ESTATÍSTICA GERAL

Material para uso exclusivo nos Cursos do CTAE adaptado pelo Professor Luiz Antonio Teixeira Vasconcelos.

G^1 .09-07.75-100/

1975

ÍNDICE

I . As Funções da Estatística

Do ponto de vista de sua utilização a Estatística compreende duas funções muito amplas. A primeira delas é a Descrição, o resumo da informação de tal modo que se possa manuseá-la mais facilmente para analisá-la melhor. A segunda é a Indução que consiste em formular generalizações a respeito de uma população com base em amostras extraídas da mesma.

A Estatística Descritiva: muitas vezes, na investigação social, econômica, política ou no caso específico de um trabalho de consultoria empresarial, podemos nos encontrar numa situação de dispor de uma massa tão grande de dados que resulte difícil absorver a informação inteira contida nessa massa.

De alguma forma, portanto, a informação deve reduzir-se até um ponto no qual possamos ver claramente o que há nela. Isso pode ser conseguido através do cálculo de certas medidas tais como: Porcentagens, Médias, Desvios-Padrão, Coeficientes de Correlação, etc.

É necessário ter claro, porém, que ao reduzir um grande conjunto de dados em algumas medidas, perdemos necessariamente, certas informações, e podemos por isso mesmo, obter resultados enganadores. Nesse sentido, é conveniente que indiquemos com clareza as limitações de toda medida resumida e tenhamos muita precaução na interpretação delas.

A Estatística Indutiva: E muitas situações práticas, o investigador se vê obrigado, por várias razões, a generalizar com base numa informação limitada. Talvez a função mais importante da estatística seja a indução, entendida como o procedimento de inferir propriedades de uma população baseada em amostras de resultados conhecidos provenientes dessa população.

A indução estatística se baseia diretamente na teoria da probabilidade (que é um ramo da matemática) e implica na utilização de uma base racional muito mais complexa do que a descrição estatística.

Na discussão das funções da estatística cabem algumas advertências:

1. A estatística não deve ser encarada de maneira alguma como um método através do qual se possa provar tudo o que se queira. Na sua utilização as “regras do jogo” devem ser estabelecidas de tal forma que as interpretações nunca escapem dos limites dos dados disponíveis.
2. Os métodos estatísticos não se opõem à análise qualitativa dos casos particulares, ao contrário, ambos os métodos se complementam.
3. A análise estatística, por mais elaborada que seja, muito dificilmente poderá compensar as falhas de um projeto mal concebido ou de um instrumento de coleta de dados deficiente.

Na atividade de consultoria a utilização da estatística, como instrumento auxiliar de análise é muito freqüente. A elaboração de um diagnóstico empresarial implica em determinados levantamentos de informações que vão ser utilizadas para descrever a situação de setores da empresa, relacionar variáveis de um ou mais setores, e, além disso, podem ser usadas no momento de se fazer previsões sobre o comportamento da empresa com base na situação atual.

No caso das atividades de consultoria nas pequenas indústrias, é necessário ter clareza que a utilização dos métodos estatísticos para a análise empresarial, será muitas vezes restringida pela costumeira ausência de controles que forneçam as informações necessárias.

Além disso, determinados procedimentos estatísticos utilizados pelo consultor, escapam do nível geral de compreensão do empresário, portanto, no momento de serem justificados tais procedimentos, o consultor deve se preocupar em transmitir ao empresário, claramente, o significado econômico dos resultados obtidos.

II - A Estatística Descritiva

Antes de discutir as principais medidas empre4gadas para resumir um conjunto de dados, vamos enunciar alguns critérios de classificação e medição.

- Escala Nominal - é o nível mais simples de medição. Consiste na classificação dos indivíduos (observações) segundo determinadas características. Damos a cada categoria um nome conveniente com o objetivo de distinguí-las. A exigência dessa classificação em categorias é que elas englobem todos os casos do universo a que pertencem. Ex.: classificação de uma população segundo a cor da pele, a religião, etc.

- Escala Ordinal - nível de classificação que engloba a escala nominal e no qual agrupamos os indivíduos, não só em função de determinadas características, como também, segundo o grau que possuem essas características. Ex: classificação das famílias segundo o seu respectivo nível sócio-econômico em: “superior”, “média superior”, “média inferior”, “inferior”, etc.

- Escala de Intervalo - Esse nível é empregado em situações nas quais não só podemos classificar uma população, segundo o grau em que possui determinadas características, como também podemos indicar a distância exata entre esses graus. Ex.: classificação de uma população segundo a magnitude da renda anual do indivíduos em: entre $ l.000 e $ 1.500, $ 1.500 e % 2.000, etc.

II.1 - Proporções, Porcentagens e Razões

Seja o seguinte Quadro dos Produtos de uma Empresa:

Produto	Quantidade	Valor do Faturamento ($ 1000)
A	600	800
B	200	1.000
C	150	100
D	50	100
Total	**1000**	**2.000**

Para nos utilizarmos proporções para resumir os dados acima, temos que presumir que o método de classificação dos produtos, segundo a quantidade e o faturamento, foi tal que as categorias (A, B, C e D) são mutuamente exclusivas e exaustivas.

As proporções de cada categoria serão as seguintes:

Produto	**Proporção de Quantidade**	**Proporção do Valor do Faturamento**
A	0,60	0,40
B	0,20	0,50
C	0,15	0,05
D	0,05	0,05

:

A forma geral é:

$$\text{Proporção da Categoria C} = \frac{\text{n. de casos da categoria C}}{\text{n. total de casos}}$$

As porcentagens de cada categoria serão:

Produto	**Porcentagem-Quantidade**	**Porcentagem do Valor de Faturamento**
A	60%	40%
B	20%	50%
C	15%	5%
D	5%	5%

A forma geral é: Porcentagem da Categoria C = Proporção de C x 100

A razão de um número N_1 com respeito a outro número N_2 se define como N_1 dividido por N_2.

No quadro dado podemos calcular, por exemplo, a razão entre as categorias B e A em termos de quantidades vendidas.

Razão de B com respeito a A = 200 : 600 = 1 : 3, Isso significa que se tomarmos o total das vendas dos produtos A e B no período a que se refere o quadro, para cada 4 produtos das categorias A e B vendidos, 1 foi da categoria B e 3 foram da categoria A.

II. 2 Distribuição de Freqüência

Suponhamos que os dados apresentados no quadro 1 representem a renda anual de 100 pessoas de uma certa localidade, Os dados brutos apresentados nesta forma não servem praticamente em nada no sentido de propiciar alguma informação a respeito do que está sucedendo em termos de renda na localidade.

Quadro – 1

Renda anual de 100 indivíduos de uma certa localidade (em Cr$)

6360	8080	7260	3600	6000	3780	6120	3900	7200	5760
7320	11760	7800	3780	7320	6240	12600	8040	8160	9240
7440	14400	9960	7260	9840	6720	7020	7640	9120	12500
8580	7240	4680	7500	7560	8760	6900	7680	4920	5640
8400	13200	7500	9720	6600	11520	7920	6960	7080	9360
8700	4080	7380	7440	11400	9600	7840	9000	6200	13440
7620	4020	4200	6300	6420	14000	4800	7560	5520	6240
8520	15600	6360	6180	7640	7940	8880	5040	4260	7320
7740	6480	8040	9780	5160	6540	7400	6780	8160	8280
12360	7920	6060	9600	9000	5280	6660	5400	9480	7660

Uma observação superficial no Quadro 1 nos fornece a informação de que a maioria das pessoas recebe entre $ 6.000 e $ 9.000. Mas apenas com essa informação é difícil concluir alguma coisa sobre a distribuição como um todo.

Vamos agrupar os dados em categorias ou classes que expressem certos intervalos de variação de rendas, e nos utilizaremos desses grupos para dar uma visão conjunta da distribuição total.

Para isso devemos decidir quantas categorias vamos utilizar e quais são os seus limites. Não existe, porém, uma regra geral para esse procedimento e a decisão deverá depender dos objetivos que queremos atingir através da classificação.

No quadro 1, podemos observar que a renda varia de $ 3.600 até $ 15.600, portanto, a amplitude total da variação dos dados é de 12.000.

Suponhamos que nosso objetivo inicial é separar os indivíduos em grupos que tenham rendas anuais que diferem em $ 1.200 (ou rendas mensais que diferem em $100).

Devemos então, nesse caso, utilizar 10 categorias que agrupem as rendas em intervalos com amplitude de $ 1.200.

Quadro - 2

Distribuição da freqüência com os dados agrupados em intervalos de $ 1.220

Classes	Intervalos	Freqüência (f)	Freqüência Re4lativa - % - (f? f) x 100
1	3600 - 4800	9	9%
2	4800 - 6000	9	9%
3	6000 - 7200	21	21%
4	7200 - 8400	32	32%
5	8400 - 9600	12	12%
6	9600 - 10800	6	6%
7	10800 - 2000	3	3%
8	12000 - 3200	3	3%
9	13200 - 14400	3	3%
10	14400 - 5600	2	2%
		∑ f = 100	∑ f/ ∑ t x 100) = 100%

Obs. : os intervalos devem ser interpretados como: do limite inferior inclusivo, até o limite superior exclusivo.

Algumas vezes pode ser conveniente ao invés de indicar o número de casos em cada intervalo, indicar o número de casos que são maiores ou menores que um determinado valor. No nosso exemplo do Quadro – 2, podemos observar que há 18 indivíduos que ganham menos do que $ 6.000 por ano e que há 5 indivíduos cuja renda anual é maior do que 13.200.

Vamos apresentar então os dados na forma de freqüência acumulada (F) e conforme vimos podemos acumular as freqüências tanto “para cima” como “para baixo”, bastando para isso, observar quantos casos estão acima ou abaixo de um determinado valor.

Quadro - 3

Distribuição de Freqüência Acumulada

Acumulação para cima		Acumulação para baixo	
No. de casos abaixo de	**Freqüência acumulada**	**No. de casos acima de**	**Freqüência acumulada**
3600	0	3600	100
4800	9	4800	91
6000	18	6000	82
7200	39	7200	61
8400	71	8400	29
9600	83	9600	17
10800	89	10800	11
12000	92	12000	8
13200	95	13200	5
14400	98	14400	2
15600	100	15600	0

Por exemplo, podemos dizer que: “39 indivíduos têm renda anual abaixo de $7.200” ou “29 indivíduos têm renda acima de $ 8.400“.

Observações a respeito do agrupamento de dados em distribuições de freqüências:

- Ao resumir dados agrupando-os em classes de intervalos, perdemos, inevitavelmente, alguma informação importante. Mesmo assim, só dessa forma é que conseguimos uma visão mais clara do comportamento geral dos dados, visão essa, totalmente obscurecida quando tratamos com dados brutos.

- Um procedimento sempre útil no momento de analisarmos quantas classes devemos nos utilizar para agrupar um conjunto de dados, é calcular a diferença entre o maior e o menor valor deles. Em seguida temos que calcular a amplitude das classes. Isso deve ser feito levando em conta que o valor escolhido não pode ser altamente significativo no conjunto dos dados. Por exemplo: no agrupamento dos dados do

Quadro–1 (no Quadro 2) foi considerada que a diferença de $ 100 na renda mensal de dois indivíduos não é altamente significativa.

II.3 - Medidas de Tendência Central

Além das medidas até agora estudadas, que são utilizadas para resumir um determinado conjunto de dados, existem as medidas de tendência central ou de "tipismo" que descrevem os dados através de cálculo de valores chamados "médios". Em outras palavras, as medidas de tendência central, resumem (ou tipificam) através de um ou alguns números o comp0ortamento de uma seqüência grande de dados.

Existem diversas medidas de tendência central, que, em determinadas situações, podem fornecer resultados bastante distintos.

II.3.1 - A Média Aritmética

Num conjunto de dados com um número total de N observações, a média aritmética é definida como a soma de todas as observações dividida pelo número total delas, ou seja:

$$\bar{X} = \frac{X1 + X2 + \ldots Xn}{N} = \bar{X} = \frac{\tilde{\sum}_{i=1} Xi}{N}$$

Na seqüência desse texto sempre que possível notaremos o somatório $\sum_{I=1}$ simplesmente como $\sum$.

1ª Propriedade da Média: $\sum (Xi - \bar{X}) = 0$, ou seja, a soma dos desvios de todas as observações com relação à média $(\bar{X})$ é igual a zero.

2ª Propriedade da Média: $\sum (Xi - X)^2$ = mínimo, ou seja, a soma de todos os desvios quadrados, de cada observação com relação à média ($\overline{X}$) é menor que a soma desses mesmos desvios com relação a qualquer outro número. Essa propriedade é chamada de "mínimos quadrados" e é muito utilizada como medida de variação total ou heterogeneidade de uma distribuição, como veremos mais adiante.

Como exemplo vamos calcular a média dos dados do Quadro – 1:

N = 100

∑ Xi = 773220

$X = \frac{\sum Xi}{N} = \frac{773220}{100}$ portanto $\overline{X} = 77.732,2$

II.3.2 - A Média aritmética para dados Agrupados e a Média Aritmética Ponderada.

Num conjunto de dados agrupados em categorias o cálculo da média aritmética deve ser efetuado da seguinte forma:

$X = \frac{\sum fi\ \overline{X}i}{\sum fi}$, onde f : freqüência da classe i

$\overline{X}$ i : ponto médio da classe i, ou seja, média aritmética do limite inferior e superior da classe i.

Como exemplo tomemos os dados do Quadro – 2 e calculemos o ponto médio de cada intervalo de classe.

Quadro – 4

Classes	Ponto Médio ($\overline{X}$)	Freqüência (f)	f. $\overline{X}$
3600 - 4800	4200	9	37800
4800 - 6000	5400	9	48600
6000 - 7200	6600	21	138600
7200 - 8400	7800	32	249600
8400 - 9600	9000	12	108000
9600 -10800	10200	6	61200
10800 -12000	11400	3	34200
12000 -13200	12600	3	37800
13200 -14400	13800	3	41400
14400 -15600	15000	2	30000
	TOTAL	**100**	**787200**

Portanto: $\sum$ fi xi = 787200

$\sum$ fi = 100

$\overline{X}$ = 7872

Como podemos observar o segundo resultado ($\overline{X}$ = 7872) é diferente do primeiro que calculamos ($\overline{X}$ = 7732,2). Tal diferença se explica pelo fato, já comentado, de que quando trabalhamos com dados agrupados, perdemos inevitavelmente algumas informações. Sabemos, por exemplo, que existem 21 indivíduos que possuem renda anual entre $ 6000 e $ 7200, mas não sabemos exatamente como as rendas estão distribuídas dentro do intervalo citado. Nosso raciocínio supôs que os 21 indivíduos tivessem rendas iguais a 6600, que é o ponto médio do intervalo. Esse raciocínio é equivalente a supor que as rendas estivessem distribuídas no intervalo de maneira uniforme.

A média aritmética calculada dessa maneira, para dados agrupados, é um caso particular da Média Aritmética Ponderada onde as freqüências fizeram o papel dos Pesos Relativos ao número de ocorrência de um mesmo valor num conjunto de observações.

A Média Aritmética Ponderada é calculada da seguinte forma:

$$X = \frac{\sum W_i X_i}{\sum W_i}$$

onde X_i : valor da observação i

W_i : Número de ocorrências de observação i no conjunto das observações

II.3.3. - A MEDIANA

Num conjunto de observações a mediana é definida como sendo um valor que tem a propriedade de possuir o mesmo número de observações com valores maiores do que ele.

Como se nota, pressupõe-se que os dados estejam ordenados segundo sua magnitude. Se temos um número ímpar de observações ordenadas, a mediana será a observação de ordem (N+1) / 2, onde N é o número total das observações. Se N for par, a mediana será a média aritmética ética das observações, de ordem N/2 e N + 1.
2

Como exemplo sejam os dois seguintes conjuntos de observações:

57, 69, 72, 81, 86

12, 12, 5, 14, 15, 16, 18

No primeiro conjunto a mediana é a observação central de ordem 3, pois, N=5, portanto Md = 72.

No segundo caso a mediana vale:

Md = $\frac{14 + 15}{2}$ = 14,5, pois, N = 6 e a mediana é a média aritmética da 3ª e 4ª observações.

1ª - Propriedade da mediana: $\sum$ Xi-Md = mínimo, ou seja a soma de todos os desvios, tomados em seu valor absoluto (sem considerar os sinais), com a relação à mediana é menor que a soma dos desvios tomados da mesma forma com relação a qualquer outro valor.

II.3.4 - A MEDIANA PARA DADOS AGRUPADOS

O primeiro passo do cálculo da mediana para dados agrupados é encontrar a classe mediana (ou seja, a classe que contém a mediana). Isso pode ser feito com o auxílio da distribuição de freqüências acumuladas.

Em seguida podemos aplicar uma das duas fórmulas.

1) Md = ℓ + $\frac{N/2 - F1}{f}$. i L+

2) Md = M – $\frac{F2 - N/2}{f}$ .i U -

Md = Mediana

ℓ = Limite inferior da classe que contém a mediana

N = Número total de observações

F1 = Freqüência acumulada correspondente ao limite inferior.

I = Amplitude do intervalo que contém a mediana

f = Freqüência da classe que contém a mediana

U = Limite superior da classe mediana

F2 = Freqüência acumulada correspondente ao limite superior

Como exemplo vamos calcular a mediana dos dados do quadro – 2.

Com ajuda do quadro 3, podemos constatar que a mediana está contida na classe com limite inferior igual a $ 7200 e limite superior igual a $ 8400, pois, temos um número par de observações (100) e as observações de ordem 50ª e 51ª estão nesse intervalo.

Como: N = 100

1 = 39. F2 = 7.1

f = 32; ℓ = 7200

i = 1200 : M = 8400

Temos $Md = 7200 + \frac{50 - 39}{32} \cdot 1200 = 7200 + 412,5$

$\boxed{Md = 7612,5}$ ou

$Md = 8400 - \frac{71 - 50}{32} \cdot 1200 = 8400 - 787,5$

Da mesma forma temos $\boxed{Md = 7612,5}$

II.3.5 - <u>A Moda</u>

Num conjunto de observações, uma outra medida de tendência central utilizada é a <u>moda</u> definida como a observação que tem maior freqüência.

Para dados agrupados a definição é análoga: é a classe mais “populosa “, ou seja, a que tem freqüência máxima.

Para explicar tomemos as seguintes seqüências numéricas:

1) 71, 75, 83, 75, 61, 63
2) 71, 75, 83, 74, 61, 68
3) 71, 75, 83, 75, 83, 68

A primeira tem moda igual a 75 que é o valor mais freqüênte. A Segunda não possui moda e a terceira tem duas modas iguais a 75 e 83 (ambos c/ freqüência igual a 2).

Nos casos de dados agrupados tomemos o quadro – 2 como exemplo. A classe de 7200 até 8400 é a classe modal pois sua freqüência é a máxima da distribuição dada.

A aplicação dessa medida, par descrever uma distribuição de freqüência, adquire maior importância e consistência quando estamos trabalhando com um número bastante grande de observações.

II.3.6 - DECÍS, QUARTÍS E PERCENTÍS

Existem outras medidas, chamadas de posição (analogamente à mediana) que podemos utilizar para fixar a posição de dados maiores ou menores do que uma proporção determinada dos caos examinados.

Podemos dividir uma distribuição em quatro quartis, por exemplo, sendo o primeiro quartil o valor que possui a propriedade de que um quarto dos dados seja de menor magnitude que a sua. (De maneira semelhante o 2^o e 3^o quartis).

Analogamente podemos dividir a distribuição em dez decís ou em cem percentís.

O cálculo dos decís, quartís, percentís, segue exatamente o mesmo raciocínio usando para o cálculo da mediana..

II.3.7 - A MÉDIA GEOMÉTRICA E A MÉDIA HARMÔNICA

Menos utilizada, existem ainda como medidas de tendência central a média geométrica e a média harmônica.

$$\text{Média Geométrica} = M.G. = \sqrt[n]{X_1 . X_2 \ldots\ldots X_n}$$

$$\text{Média Harmônica} = M.H = \frac{1}{\sum \frac{1}{X_i}}$$

II.3.8 - COMPARAÇÃO ENTRE A MÉDIA ARITMÉTICA E A MEDIANA

Para estabelecer a comparação entre essas duas medidas, vamos trabalhar com algumas seqüências de observações:

1) 25.30. 35. 40. 50
2) 5, 30. 35. 40. 50
3) 25. 30. 35. 40. 105

$\overline{X}$ (1) = 36 ; $\overline{X}$ (2) = 32 ; $\overline{X}$ (3) = 47

Md (1) = 35 , Md (2) = 35 , Md (3) = 35

Através da observação dos valores acima podemos tirar algumas conclusões:

A) A média é bastante afetada pelo mudança dos valores extremos da seqüência dos dados, do passo que a mediana não altera (a menos que se mude o valor do caso médio).

B) No cálculo da média são utilizados todos os casos da seqüência, enquanto que para o cálculo da mediana é utilizado apenas o valor do caso médio

II.4 - Medidas de Dispersão

As medidas de tendência central, estudadas até agora, fornecem informações apenas sobre o tamanho do valor central de um conjunto de observações ou de uma distribuição de freqüência dessas observações. Nada nos informa, porém, a respeito de como estão dispersas as observações em torno desse valor central (representativo desse conjunto de observações). Para isso, temos que calcular novas medidas chamadas de Dispersão ou Homogeneidade, que vão nos indicar, com algum detalhe, de que forma o conjunto das observações se comporta em função do (s) valor (es) central (ais) calculado (s) o qual (ais) tipifica (m) esse conjunto de dados ou observações.

II.4.1 - A amplitude

A mais simples das medidas de dispersão é a amplitude definida como a diferença entre as observações de maior valor e a de menor.

Embora a amplitude nos forneça uma indicação bruta da dispersão, a sua utilização nem sempre é conveniente. Na prática quase sempre trabalharemos com amostras ao invés da população total e geralmente com uma amostra dessa população.

Assim sendo, como a amplitude se baseia, exclusivamente, nos dois casos extremos da amostra (o maior e o menor), teríamos que colocar amostras de um número muito grande para aumentar a possibilidade de incluir nela um dos casos extremos.

II.4.2 - O Desvio Quartil

O desvio quartil é um tipo de amplitude que se define como a metade da distância entre o primeiro e terceiro quartis, ou sendo:

$$D.Q. = \frac{Q_3 - Q_1}{\quad}$$ onde D.Q. = Desvio de Quartil

2 Q_3 = 3 quartil
Q_1 = 1 quartil

Podemos explicar calculando o desvio quartil da distribuição de freqüência do quadro – 2:

$$Q_1 = 6000 + \frac{25-18}{21} \cdot 1200 = 600 + 400 = 6400$$

$$Q_3 = 8400 + \frac{75-71}{12} \cdot 1200 = 8400 + 400 = 8800$$

$$D.Q = \frac{Q_3 - Q_1}{2} = \frac{8800 - 6400}{2} = \frac{2200}{2} = 1100$$

O desvio quartil mede a amplitude ocupada pela metade central dos dados, ou seja, mede a concentração dos valores em torno da mediana. Tal concentração é inversamente proporcional ao tamanho do desvio quartil.

Essa medida de dispersão embora seja mais estável de que a amplitude também não se utiliza para seu cálculo do conjunto das informações. Além disso, com ela não medimos a variabilidade entre os casos centrais e não levamos em consideração os casos extremos da distribuição.

II.4.3 - O Desvio Médio

O desvio médio é definido como a média aritmética das diferenças absolutas de cada observação com respeito à média..

$$D.M. = \frac{\sum X_i - \overline{X}}{N}$$

A principal vantagem que esse desvio apresenta é uma interpretação intuitiva mais direta, pois, significa a dimensão que, em média, as observações diferem da média $\overline{X}$.

II.4.4 - A Variância e o Desvio Padrão

A mais útil das medidas de dispersão e de utilização mais freqüente é o desvio padrão, definido como a raiz quadrada da média aritmética dos desvios quadrados de cada observação com relação à média X, ou seja:

$$S = \sqrt{\frac{\sum (Xi - \overline{X})^2}{N}}$$

Exemplo : Vamos calcular o desvio padrão do seguinte conjunto de observações:

72 = 81, 86, 69, 57

$\overline{X}$ = 73

Xi	(Xi – $\overline{X}$)	(Xi - $\overline{X}$)^2
72	-1	1
81	8	64
86	13	169
69	-4	16
57	-16	256

$$S = \sqrt{506/5} = \sqrt{101,2} = 10,06$$

Observamos do cálculo acima que quanto maior for a dispersão torno da média, maior será o desvio padrão. Por outro lado, os desvios extremos com relação à média, pesam muito mais no momento de determinar o desvio-padrão. Nesse sentido, numa distribuição que apresenta poucos casos muito extremos, o desvio padrão pode nos conduzir a resultados enganosos. Em tais casos é mais adequado utilizar como medida de tendência central a mediana, e como medida de dispersão o desvio quartil.

Podemos nos utilizar ainda, como medida de dispersão, do quadrado do desvio padrão ou variância, assim definida:

$$S^2 = \frac{\sum (Xi - \overline{X})^2}{N}$$

OBS: - No cálculo do desvio padrão podem ser utilizadas fórmulas que simplificam o cálculo, tais como:

$$S = \frac{Xi^2}{N} - \overline{X}^2 \quad \text{ou}$$

$$S = 1 \quad N \quad Xi^2 - (Xi)^2$$

II . 4. 5 - Coeficiente de Variação - Dispersão Relativa

Vamos definir o coeficiente de variação através de um exemplo. Seja um grupo de pessoas formado por homens e mulheres e suponhamos que seus salários mensais e suas idades estejam assim expressos:

Salários – mulheres $\overline{X}$ = \$ 450 e S = \$ 67,5

Salários – homens $\overline{X}$ = \$ 600 e S = \$ 72

Idades: $\overline{X}$ = 45 e S = 9

No caso salários podemos comparar o desvio padrão pois a unidade é a mesma. A dispersão absoluta é maior para os salários dos homens. Se relacionarmos, porém, essa dispersão com a média, usando a expressão $C.V = \frac{S}{\overline{X}}$ obteremos uma medida da dispersão relativa à média chamada coeficiente de variação.

Calculando : Salário mulher - $C.V = \frac{67,5}{450} = 15\%$

$$\text{Salário Homem} - \quad C.V = \frac{72}{600} = 12\%$$

$$\text{Idades} - \quad C.V = \frac{9}{45} = 20\%$$

Essa medida nos permite comparar se mês de valores com unidades diferentes tais como os salários e as idades.

II. 5 - Medidas de Assimetria

Além das medidas de tendência central e das medidas de dispersão estudadas até agora, existem as medidas de assimetria que são também utilizadas para descrever e resumir uma distribuição de freqüência.

Uma distribuição é simétrica quando a média for igual à mediana. Quanto maior a diferença entre a média e a mediana uma distribuição maior será sua assimetria.

Se chamarmos a medida de assimetria de A podemos escrever:

$$A = \frac{3(\overline{X} - C)}{S}$$

Se $\overline{X}$ > Md temos A>O, a assimetria será positiva significa que a distribuição está desviada para a direita, ou seja, há mais observações com valores maiores do que com valores menores.

Se $\overline{X}$ < Md temos A < O, a assimetria será negativa.

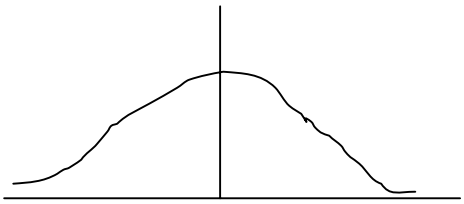
III. 6 - A Distribuição Normal

A distribuição normal (ou curva normal) é um tipo muito importante de distribuição de freqüências. A grande utilidade da distribuição normal é devida ao fato de que um grande número de distribuições encontradas na prática é aproximadamente normal, e, além disso, devido a seu significado teórico na estatística indutiva (como veremos mais adiante).

A curva normal é um tipo especial de curva contínua e simétrica que se baseia num número infinitamente grande de observações.

A forma geral da distribuição normal é a seguinte:

Figura 1 – Distribuição Normal



Na distribuição normal $\overline{X}$ = Md = Mo+

A equação matemática da curva normal relaciona a sua altura (Y) com cada valor das observações (X).

$$Y = \frac{1}{S\sqrt{2r}} \cdot e^{-\frac{1}{2}\left(\frac{X - \overline{X}}{S}\right)^2}$$

Y e X estão relacionados através de apenas duas medidas: uma de tendência central (a média $\overline{X}$) e outra de dispersão (s).

No anexo I podemos ver várias comparações entre diferentes curvas normais obtidas variando a média ($\overline{X}$) e/ou o desvio padrão (s).

Muitas vezes teremos necessidade de determinar a proporção dos casos que se encontra no interior de um determinado intervalo.

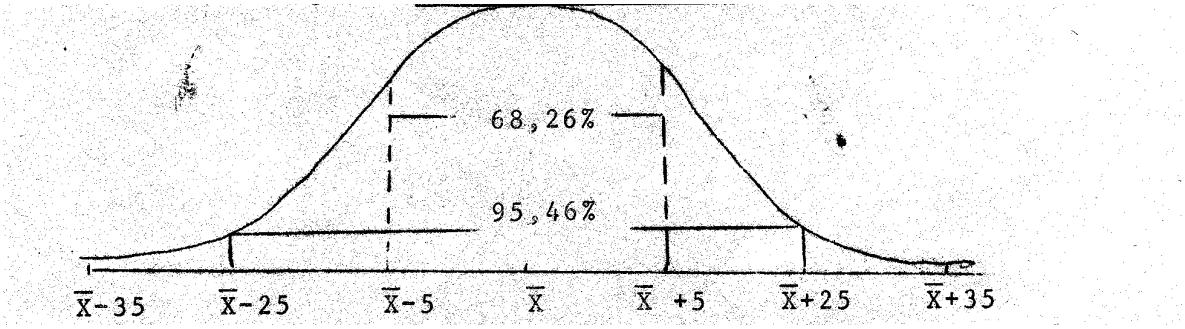
A curva normal possui uma importante propriedade que simplifica essa tarefa: “Há sempre, abaixo da curva normal, uma área constante (ou proporção de casos) compreendida entre a média e uma abscissa) que difira da média de um certo valor medido em unidades de desvio-padrão.

Dessa forma, se tomarmos uma ordenada à direita que seja X = $\overline{X}$ + s, teremos sempre 34,13% da área total, sob a curva, aí incluída.

A partir daí podemos dizer, em função da simetria da curva, que se X1 = $\overline{X}$ -S, a proporção de casos incluída entre XI e $\overline{X}$ será também de 34,13%.

A figura 2 ilustra essa propriedade:

Fig. 2 - Áreas Sob a Curva Normal



OBS: - É possível determinar áreas debaixo de curva normal mesmo que a distância entre a média e as de abscissas, que estivermos usando, não foram múltiplos inteiros do desvio padrão. (S).

Essa propriedade oferece uma interpretação do desvio padrão e um método para representar em forma visual o significado desta medida de dispersão.

É possível tomar qualquer curva normal e transformar seus valores numéricos de tal forma que se possa utilizar uma tabela padronizada para avaliar a proporção de casos no interior de qualquer intervalo desejado.

Esse processo consiste em transformar uma distribuição normal com média $\overline{X}$ e desvio padrão S, numa distribuição na “Forma Padrão” ou “Reduzida” com média igual a zero e desvio padrão igual a um. Essa distribuição padrão é que é tabelada.

Vamos explicar através de um exemplo:

Seja X uma curva normal com média $\overline{X}$ = 50 e desvio padrão S=10. Queremos saber a proporção dos casos no intervalo de 50 a 65.

Primeiro passo, determinamos a quantas unidades de desvio padrão se encontra 65 da média 50. Vamos chamar Z essa diferença medida em unidades de desvio padrão

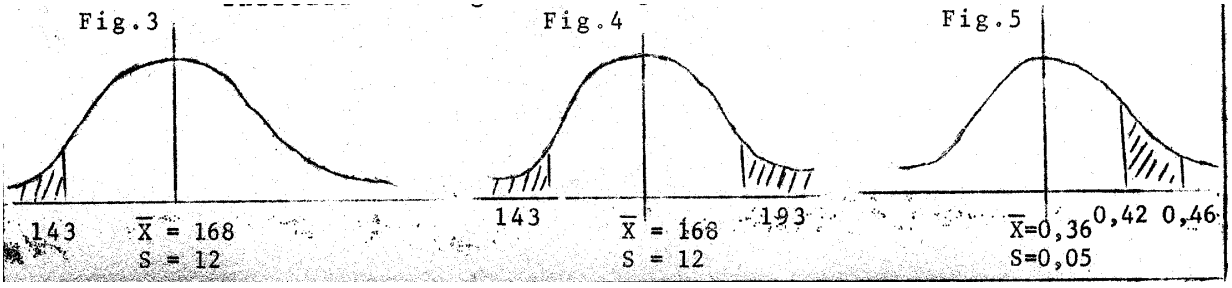
$$Z = \frac{X - \overline{X}}{S} \quad \text{ou} \quad Z = \frac{65 - 50}{10} \quad \therefore \; Z = 5$$

Esse primeiro resultado significa que o valor 65 está a 1,5 desvios padrão da média.

Segundo passo: procuramos na tabela (ver anexo – II) o valor da área correspondente a Z = 1,5. O resultado é 0,4332, o que significa que 43,32% dos casos descritos pela distribuição X estão entre o valor da média dessa distribuição, $\overline{X}$ = 50, e o valor pesquisado X = 65.

Outros Exemplos de Aplicação da Tabela Normal

Suponhamos que queremos achar a área indicada nas seguintes figuras:



Na figura –3 queremos determinar a proporção dos casos com valores menores ou iguais a 143.

1 passo: Padronização da Variável X

$$Z = \frac{X - \overline{X}}{S} = \frac{143 - 168}{12} = 2{,}08$$

O fato de Z ser negativo representa que a área que queremos está à esquerda da média. Como a curva normal é perfeitamente simétrica, ao utilizarmos a tabela podemos desprezar o sinal de Z.

2 passo : Pesquisa na tabela (ver anexo - II)

O resultado encontrado nos diz que uma proporção de 0,4812 dos casos, está compreendida entre a média e o valor 143. Mas, o que estávamos procurando não era isso, e sim a proporção dos casos Abaixo de 143.

Para encontrar a proporção que queremos, basta lembrarmos que a área total abaixo da curva normal é igual a 1 (ou 100% dos casos). Portanto, a área a esquerda da média deve ser igual a 0,5 (por simetria).

(Proporção de casos $\leq$ 143 = 0,5 - (proporção de casos entre a média e 143) ou

(Proporção de casos > 143) = 0,5 – 0,4812 = 0,0118

Na figura – 4, queremos encontrar a proporção dos casos que se encontram fora do intervalo 168 + 25 ou seja a proporção dos casos Abaixo de 143 e acima de 193.

Como a curva é perfeitamente simétrica e já conhecemos a proporção abaixo de 143, basta multiplicarmos esse valor por dois, ou seja:

(proporção dos casos abaixo de 143 e acima de 193) = 2 x (0,5 – 0,4812) = 0,0236.

O exemplo da figura – 5 deve ser feito como exercício.

III - A Estatística Indutiva

III – 1 - Estatística e Parâmetros

O Objetivo das Generalizações estatísticas consiste em afirmar algo sobre as diversas características de uma população estudada, com base em fatos conhecidos através de uma amostra tirada dessa população ou universo. Designaremos as características da população (entendida como o Conjunto de todos os elementos que possuem uma determinada propriedade) como parâmetros, e por outro lado, chamaremos de estatística, as características da amostra (entendido, como um Subconjunto formado exclusivamente por elementos de uma determinada população).

Desde logo vamos estabelecer uma distinção entre parâmetros e estatísticas. Os parâmetros são valores fixos referidos à população, que, geralmente, não se conhecem. Por exemplo: Em qualquer momento dado, a idade média dos estudantes de uma universidade ou a quantidade média de produtos vendidos pelas empresas de um determinado ramo. Em contraste, as estatísticas de uma amostra são conhecidas ou podem ser calculadas.

Mas não sabemos o quanto representativo da população é a amostra, ou até que ponto a estatística obtida se aproxima ao parâmetro correspondente, que é desconhecido.

Em todos os casos o que nos interessa efetivamente é a população e não uma amostra particular dela. Geralmente escolhemos uma amostra por questão de conveniência, mas as estatísticas que calculamos numa amostra não têm importância em si mesma.

Nas verificações de hipóteses, formulamos suposições a respeito dos parâmetros desconhecidos, e, em seguida, perguntamos como seriam nossas estatísticas específicas se tais suposições fossem corretas. Isso significa que temos que decidir, racionalmente, se os valores supostos de tais parâmetros são ou não razoáveis em função da evidência de que dispomos.

Para melhor distinguir entre as características que se referem à população e aquelas que dizem respeito à amostra, são usada, geralmente, a seguinte notação: Para a população – Letras Gregas – Ex. média ( ), desvio padrão (j), etc.

Para as amostras – letras latinas – Ex: média ($\overline{X}$), desvio padrão (s), etc....

III.2 - Etapas da Verificação de uma Hipótese

A hipótese deve ser entendida como um enunciado a respeito de um acontecimento futuro, ou de um acontecimento cujo resultado se desconhece no momento da previsão, formulado de tal maneira que se possa inclusive rejeitá-lo.

Podemos dizer que se comprovou uma hipótese cada vez que tenham efetuado os seguintes passos:

1. Sejam antecipados, anteriormente à verificação, todos os resultados possíveis do experimento ou observação.

2. Seja tomada uma decisão, antes de proceder à verificação, respeito das operações ou procedimentos que serão empregados na determinação de quais resultados possam ser produzidos efetivamente.

3. Tenha se decidido, previamente, quais dos resultados implicarão, no caso de se efetivarem, na rejeição da hipótese ou na sua confirmação. A rejeição da hipótese deve ser sempre um dos resultados possíveis.

4. Tenha sido realizado o experimento ou observado o acontecimento, se registrou os resultados e se decidiu se a hipótese foi aceita ou rejeitada.

III.3 - A Forma das Hipóteses Estatísticas

As hipóteses estatísticas são constituídas por uma teoria (que vamos chamar de T) e um determinado conjunto de resultados possíveis (que chamaremos de R).

A teoria T consta de um número de suposições a respeito do caráter da população e dos procedimentos relativos à seleção de amostras. Além disso, consta de

T alguns enunciados de probabilidade a respeito da ocorrência de particulares resultados da amostra.

Os resultados B estão representados por uma certa amplitude de resultados da amostra.

Ao decidir pela amplitude ou extensão R, temos que levar em conta o risco de ocorrer dois tipos de erros:

Erro do tipo I ou = os resultados caem fora de R, nós rejeitamos T, e ela é verdadeira.

Erro do tipo II ou os resultados caem no interior de R, nós aceitamos T e ela é falsa.

Num teste de hipótese, a probabilidade de incorrer num erro do tipo I é chamada Nível de Significância do teste.

III .4 - Probabilidade

De uma maneira simplificada, podemos definir a probabilidade matemática de um determinado acontecimento (ou evento). E, como o valor limite alcançado pela seguinte razão:

$$\frac{\text{Nº. de ocorrências do evento E}}{\text{Nº total de experimentos realizados}} = P(E) = \frac{n}{N}$$

Devemos considerar que em cada um dos N experimentos realizados o evento E era um dos resultados possíveis. Além disso o valor limite é alcançado quando o número de experimentos é extremamente grande.

Outra maneira também simples de definir probabilidade é a seguinte:

$$P(E) = \frac{\text{número de maneiras pelas quais o evento E pode ocorrer}}{\text{número de todos os resultados possíveis do experimento}} = \frac{r}{N}$$

Ex: Num lançamento de um dado há 6 resultados possíveis.

Se E: Face voltada para cima = 2

Então $P(E) \frac{n}{N} = \frac{1}{6}$

Na prática, o pesquisador só pode obter proporções, pois o número de experimentos ou casos será sempre finito.

Principais Propriedades da Probabilidade Matemática

1. O ≤ P (A) ≤, para qualquer acontecimento A
2. P (A) ou B) = P (A) + P (B), se A e B são dois eventos mutuamente exclusivos.
3. P (A) E B) = P (A) . P (B), se A e B são independentes.

As provas estatísticas (teste de hipóteses) que descreveremos a seguir partem do pressuposto que existe independência de seleção b no interior de uma amostra, não tendo a seleção de um indivíduo (ou uma observação) nenhuma influência sobre a seleção de outro incluído na mesma amostra.

III.5 - Testes de Hipóteses

As provas estatísticas comportam um certo número de procedimentos específicos:

1. Formular suposições a respeito da população. Por exemplo, identificar o modelo a ser aplicado e formular as hipóteses Ho e H1. Ho é a hipótese nula formulada com o propósito de ser rejeitada e H1 é a hipótese alternativa.
2. Obter a distribuição da amostra.
3. Selecionar o nível de significância e a região crítica.
4. Calcular a estatística do teste.
5. Tomar uma decisão sobre as hipóteses.

Alguns dos principais testes estatísticos são:

A) Teste da média amostral ($\overline{X}$) em relação à populacional (m) desconhecendo-se o desvio padrão (j) da população.

A estatística utilizada é a “distribuição” T de “student”:

$$T = \frac{\overline{X} - u}{S/\sqrt{N-1}}$$

Onde S = desvio padrão da amostra está sendo usado como uma estimativa justa ($\hat{\sigma}$) odo desvio padrão da população

Portando: $\hat{\sigma} = \sqrt{\frac{(xi - \bar{x})^2}{N - 1}}$

B) Testes de duas Amostras

B.1 Teste de diferença das médias

Estatística utilizada: $Z = \frac{\bar{X}1 - \bar{X}2}{\hat{\sigma}\ \bar{X}1 - \bar{X}2}$

Supondo que $\sigma 1 \neq \sigma 2$, a estimativa $\hat{\sigma}\ \bar{X}1 - \bar{X}2$ fica

$$\hat{\sigma}_{\bar{X}1 - \bar{X}2} = \sqrt{\frac{S1^2}{N1-1} + \frac{S2^2}{N2-1}}$$

B.2 Teste da diferença das proporções

Esse teste de refere a eventos Dicotomicos

Suposição - $\sigma 1 = \sigma 2$

A estatística utilizada é $Z = \frac{P1 - P2}{\hat{\sigma}\ P1 - P2}$, onde

$$\hat{\sigma}_{P1-P2} = \hat{\sigma} \frac{N1 + N2}{N1.N2}$$

e $\hat{\sigma} = \sqrt{\hat{p}_\mu . \hat{q}_\mu}$, sendo que $\sigma = \sqrt{p_\mu . q_\mu}$

p_μ e q_μ são as proporções em que se dividem as duas amostras, ou seja: Amostra-1- $\begin{cases} p_{\mu 1} \\ q_{\mu 1} \end{cases}$

Amostra-2 $\begin{cases} p_{\mu 2} \\ q_{\mu 2} \end{cases}$

$$e\ \hat{p}_\mu = \frac{N_1 p_{\mu 1} + N_2 p_{\mu 2}}{N_1 + N_2}$$

Além desses testes existem outros que não descreveremos nesse textos.

III.6 Regressão e Correlação

Em muitos situações práticas, estamos interessados não só em descrever isoladamente duas ou mais variáveis: Podemos nos interessar, particularmente. Em encontrar alguma forma

de medir a relação entre duas variáveis x e y.

Vamos explicar nosso objetivo através de um exemplo:

Suponhamos que um levantamento feito em 11 residências tenha nos fornecido os seguintes dados para a idade da casa (x) e o valor do aluguel mensal (y).

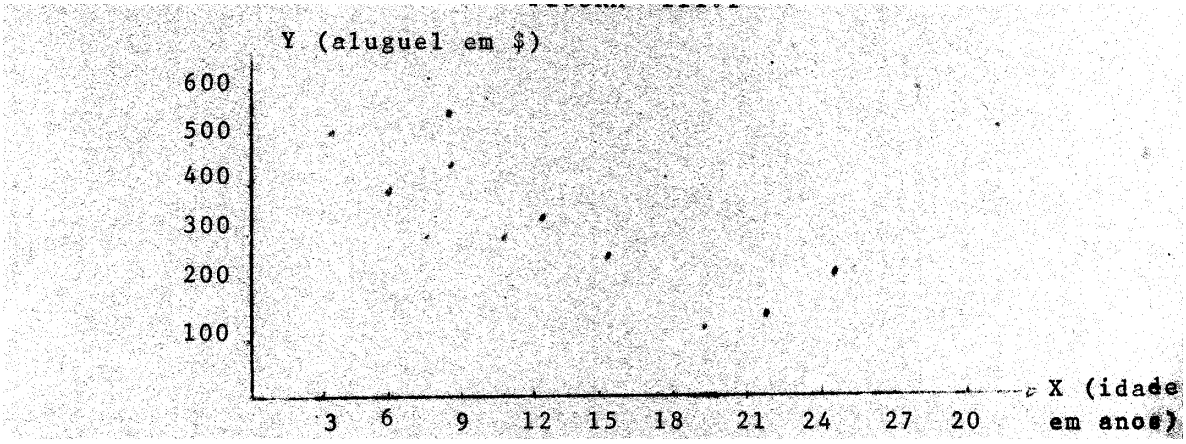
Idade (anos)	Aluguel Mensal ($)
X	Y
3	500
12	320
5	400
7	330
8	450
19	130
10	300
22	140
15	280
8	510
25	260

Queremos encontrar alguma forma de medir a relação entre X e Y de tal maneira que a medida apropriada amostre:

1. Se há relação entre X e Y e, caso haja, se é forte ou fraca.
2. Que Y, em médio, está associado a cada X.
3. Se a relação encontrada pode ou não ser usada para fins de predição, ou seja, que valor de Y pode ser esperado para cada x dado.

A primeira idéia que podemos ter sobre a existência de alguma relação é colocando os dados num diagrama de dispersão. Mediremos X no eixo horizontal e Y no eixo vertical.

FIGURA III.1



Podemos tirar uma primeira conclusão do diagrama: existe uma relação entre X e Y.

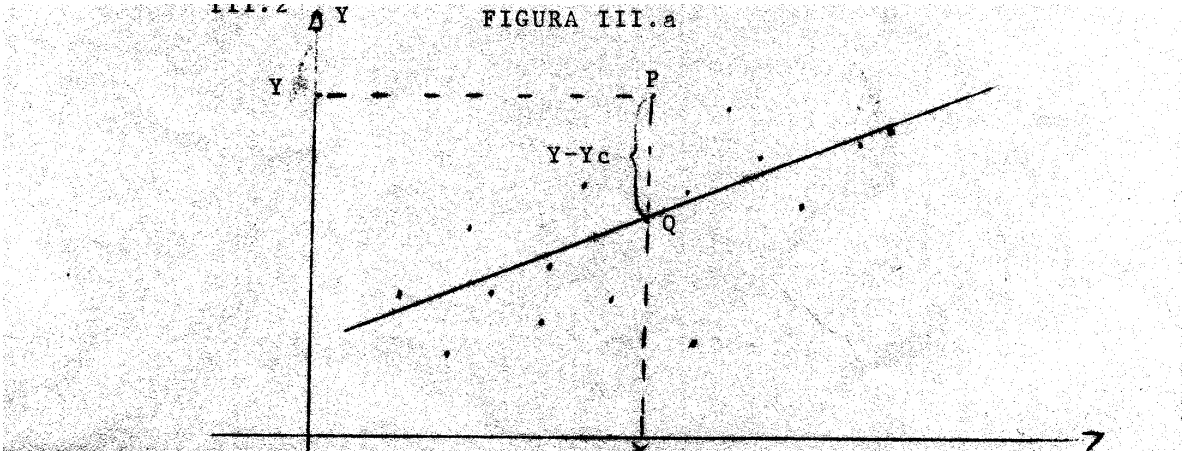
Vamos considerar aqui somente o caso das relações lineares isto é, segundo uma reta. Suponhamos que dispomos de N pares de observações sobre duas variáveis X e Y: X^1 , Y^1 , X^2 , Y^2 ..... XN YN. Se existir uma relação linear entre x e y poderemos escrever.

$$Yc = a + bX \qquad (I)$$

Onde: a e b são constantes que determinam a posição da reta Yc: valor de Y para um dado X, Resultante da Relação.

A relação expressa pela reta (I) não é perfeita e expressa tão somente uma tendência geral de associação entre x e y. Isso pode ser visto na figura III.2

FIGURA III. a



UNICAMP DEPE CTAE	ELEMENTOS DE ESTATÍSTICA GERAL	ANEXO I

COMPARAÇÃO ENTRE DIFERENTES CURVAS NORMAIS

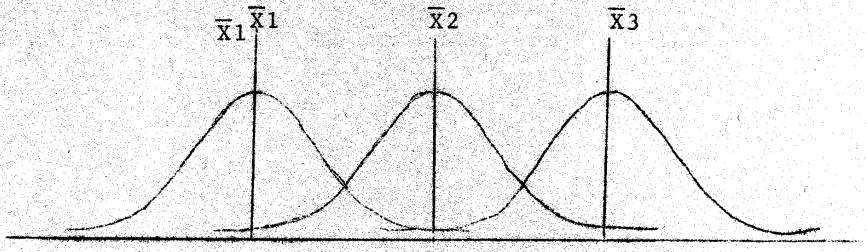
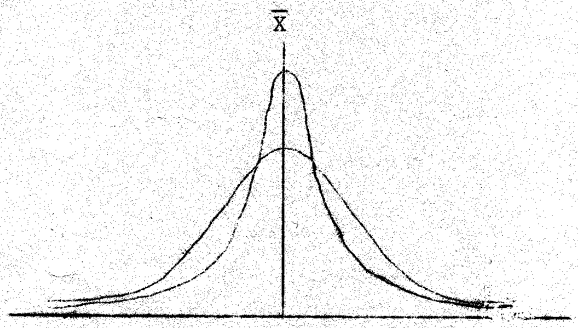


Figura 1: Comparação de Curvas normais de mesmo desvio padrão e diferentes médias ($\overline{X}1$, $\overline{X}2$, $\overline{X}3$)



igura 2: Comparação de Curvas Normais de médias iguais ($\overline{X}$), e diferentes desvio-padrão.

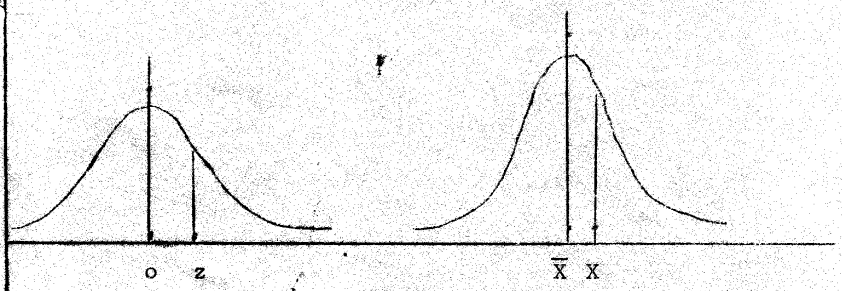


Figura 3: Comparação entre a forma padrão e a forma geral da curva normal

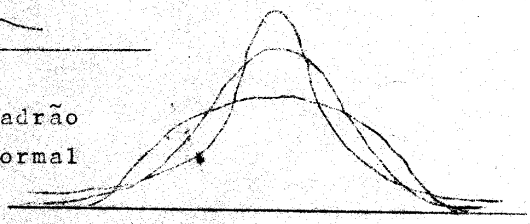


Fig.4 - Comparação entre curvas normais de mesma média mesmo desvio padrão mas com diferentes pontos de frequência. máxima.

A equação (I) expressa a relação média entre Y e um conjunto de dados X, e é chamada de Equação de Registro de Regressão Linear de Y sobre X.

A constante b é conhecida como coeficiente de regressão de Y sobre X.

Nosso problema agora é como traçar a reta, ou seja, como determinar as constantes a e b.

Estamos interessados em determinar a reta de tal forma que ela esteja o mais próximo possível dos pontos observados, isto é, de tal forma que possamos minimizar a discrepância total entre os pontos marcados e a reta.

Para cada ponto teremos um desvio Y – Yc como mostra a figura III – 2

Queremos minimizar o desvio total, ou seja, $\sum (Y_i - Yc)$. Como já vimos, $\sum (Y_i - Yc) = 0$ (Análogo a $\sum (X_i - \overline{X}) = 0$). Usaremos portanto como medida de discrepância total $\sum (Y_i - Yc)^2$. Essa forma de ajuste de uma reta de regressão é chamada de "Método dos mínimos Quadrados".

Nosso objetivo portanto será ajustar uma reta Yc = a + bx aos pontos marcados de tal forma que $\sum (Y_i - Yc)^2$ seja mínima.

Minimizando a função $W = \sum (Y_i - Yc)^2 = \sum (Y_i - a - bx_i)^2$ obtemos para a e b os seguintes valores:

$$A = \overline{Y} - B\overline{x}$$

$\overline{Y}$ = média das observações Y

$\overline{X}$ = média das observações X

$$b = \frac{\sum X_i - Y_i - \overline{N}\,\overline{X}\,\overline{Y}}{\sum X_i^{\,2} - N\,X^{\,2}}$$

Existem formas mais simples para o cálculo de b

$$B = \frac{\sum x_i - y_i}{\sum X_i^{\,2}} \qquad \text{onde} \quad x_i = X_i - \overline{X}$$

$$y_i = Y_i - \overline{Y}$$

ou

$$b = \frac{\sum x' - y' - N\overline{x'}\,\overline{y'}}{\sum x'^2 - N\,x'^{\,2}} \qquad \text{onde} \quad x' = X - A$$

$$y' = Yi - Y$$

A e B são origens arbitrárias.

Para termos idéia da força da relação entre X e Y usamos uma medida chamada Coeficiente de Correlação Linear (r) assim definido:

$$r - \sqrt{\frac{\sum (Y_c - \bar{Y})^2}{\sum (Y - Y)^2}}$$

ou

$$r = \sqrt{\frac{\sum' xy}{\sum' x^2 \cdot \sum' y^2}}$$

Para sabermos o grau de dispersão absoluta dos valores de Y em relação à reta, usamos uma medida chamada Erro Padrão da Estimativa (Sy (e)

$$Sy\ (e) = \sqrt{\frac{\sum' (Y - Ye)^2}{N - 2}} \quad ou$$

$$SY\ (e) = \sqrt{\frac{(1-n^2)\ \sum' Y^2}{N - 2}}$$

Podemos expressar a variabilidade dos Y em relação à sua própria média:

$$\sum (Y-Y)^2 = \sum (Y-Y_c)^2 + \sum (Yc - Y)^2$$

A primeira componente mede a parte variabilidade que não é explicada pela reta de regressão. Se a relação é perfeita termos $\sum (Y-Yc)^2 = 0$.

A segunda mede a parte explicada pela reta, ou seja, traduz a variabilidade dos X, na variabilidade dos Y.

BIBLIOGRAFIA

BLALOCK, HUBERT M. - "*Estatística Social" – Fundo de Cultura Econômica* – México – Buenos Aires – 1966

KARMEL, P. H. e POLASER M. "*Estatística Geral e Aplicada para Economistas*" – Editora Atlas S/A – Usp

SPIEGEL, MURROY R. – *"Estatística*" – Editora Mc Graw-Hill Do Brasil Ltda. Coleção SCHAUM – 1971

MONTELLO, JESSÉ - *"Estatística para Economistas*" – Apec Editora S/A – 1970

MEYER, PAUL L. – *"Probabilidade – Aplicação à Estatística*" ao livro Técnico S/A e Editora Da Usp – 1969

MACHLINE, CLAUDE - *"Manual de Administração da Produção*" - (cap. XV – Estatística Industrial) Fundação Getulio Vargas – R. de Janeiro – 1971.

EXERCÍCIO

1) Calcular a média, o desvio padrão e a variância da distribuição que consta do anexo III.
2) Agrupar os dados do anexo III em uma distribuição de freqüências e calcular todas às medidas de tendência central da distribuição, bem como todos os tipos de desvio que foram estudados.
3) Supondo que os dados do anexo III correspondam aos gastos anuais de 100 indivíduos na compra de camisas e que o quadro – 1 represente as respectivas rendas anuais desses mesmos 100 indivíduos, calcular a regressão linear dos gastos sobre as rendas. Calcular ainda o coeficiente de correlação da regressão.
4) As vendas mensais de uma empresa numa certa região apresentavam uma média de 1800 unidades de um certo produto, com desvio padrão de 100 unidades.
   Mediante uma modificação em seu formato, acredita-se que o produto tenha maior saída. Examinando-se uma amostra de 40 compradores usuais do produto, verificou-se que as vendas apresentavam uma média de 1850 unidades. A um nível de significância introduzida tenha realmente aumentado as vendas do produto?
5) Para o lançamento de um produto no mercado foram selecionados ao acaso, dois grupos de donas de casa, constituído cada um de 100 pessoas. Foi entregue a cada pessoa uma unidade do produto
   A média de pessoas que se mostra favorável foi de 74 no grupo A e 78 no grupo B, com desvios padrão de 8 para o grupo A e 7 para o grupo B.
   Testar a hipótese de que há diferença significativa entre os dois grupos no que se Refere a aceitação do produto, a um nível de significância de 0,05.
5) Uma pesquisa entre 300 consumidores do bairro A e 200 do bairro B indicou que 56% e 48% respectivamente, manifestaram-se favoráveis a um certo produto s ser lançado no mercado. A um nível de significância de 0,05 testar nas hipóteses:

   a) Existe diferença entre a proporção de aceitação do produto nos dois bairros?
   b) A proporção de aceitação é maior no bairro A?