UNIVERSIDADE FEDERAL DE SANTA CATARINA

CENTRO SÓCIO-ECONÔMICO

DEPARTAMENTO DE CIÊNCIAS CONTÁBEIS

Abordagem Simplificada da Análise de Regressão/Correlação: Uma aplicação Analítica como Alternativa de Selecionar Critérios de Rateio

JADER JULIO PIRES DA SILVA

FLORIANÓPOLIS – SANTA CATARINA
1996

UNIVERSIDADE FEDERAL DE SANTA CATARINA

CENTRO SÓCIO-ECONÔMICO

DEPARTAMENTO DE CIÊNCIAS CONTÁBEIS

Abordagem Simplificada da Análise de Regressão/Correlação: Uma aplicação Analítica como Alternativa de Selecionar Critérios de Rateio

Trabalho de Conclusão de Curso submetido ao Departamento de Ciências Contábeis do Centro Sócio Econômico, da Universidade Federal de Santa Catarina, como requisito parcial para obtenção do grau de Bacharel em Ciências Contábeis

Acadêmico: JADER JULIO PIRES DA SILVA
Orientador: Prof. Msc. José Alonso Borba

FLORIANÓPOLIS – SANTA CATARINA
1996

Abordagem Simplificada da Análise de Regressão/Correlação: Uma aplicação Analítica como Alternativa de Selecionar Critérios de Rateio

Autor: Acadêmico JADER JULIO PIRES DA SILVA

Esta monografia foi apresentada como trabalho de conclusão do curso de Ciências Contábeis da Universidade Federal de Santa Catarina, obtendo a nota média de ...9,5..., atribuída pela banca examinadora integrada pelos professores abaixo nominados.

Florianópolis, 03 de julho de 1996.

Prof. Adalberto Nienkötter
Coordenador de Monografia do CCN

Professores que compuseram a banca examinadora:

Prof. Msc. José Alonso Borba
Presidente

Prof. Erves Ducati
Membro

Prof. Msc. Joisse Lorandi
Membro

RESUMO

Esse trabalho visa apresentar, de forma simples e didática, os caminhos a serem percorridos para se atingir um grau de conhecimento preliminar para aplicação da análise de regressão/correlação em casos práticos de contabilidade de custos. Um tipo de problema específico foi selecionado para exemplificar a sua utilização, onde analisar-se-á a seleção de bases de alocação de CIF[1] em função do melhor relacionamento dos mesmos com os produtos. Esse exemplo foi escolhido porque, além de ser um assunto de extrema relevância para os contadores, reflete uma das grandes dificuldades da contabilidade de custos: a arbitrariedade das bases de alocação.

Para atingir o objetivo do trabalho abordar-se-á de maneira simples aspectos básicos de estatística mais especificamente, de correlação/regressão, assim como os aspectos referentes a utilização, aplicação e análise necessários a resolução do problema específico.

---

[1] Adotar-se-á a abreviatura CIF para Custo Indireto de Fabricação

SUMÁRIO

1 INTRODUÇÃO

Como conseqüência do progresso tecnológico, a mão-de-obra vêm sendo, gradativamente, substituída por ambientes automatizados. Essa substituição, implica inversão da composição dos custos totais das empresas onde os custos diretos diminuem e os custos indiretos aumentam.

Segundo o Bol IOB/TCB (47/95, p. 428):

"Os custos indiretos de fabricação tornam-se cada vez mais significativos em relação ao custo total do processo produtivo ou as taxas de custos gerais vêm crescendo durante os últimos anos. A tendência no mundo atual é a substituição da mão-de-obra pela tecnologia."

Johnson & Kaplan (1993, p. 192) dizem que *"os custos de despesas gerais representam uma fração bem maior dos custos totais precisando, por isso, ser compreendido e controlados com bem mais cuidado do que no passado."*

É provável que com o aumento global dos CIF aumentem as distorções na alocação dos custos, ou pelo menos que aumentem as preocupações em gerar informações mais acuradas que respondam aos desejos dos tomadores de decisões.

Entretanto, qualquer tentativa de alocação de CIF tem que ser feita de maneira estimada, subjetiva, proporcional a uma base de rateio que tenha fiel relação com o custo indireto.

Carsberg (1972, p. 120) cita que a maioria dos textos sobre custos sugere que a alocação deve-se basear nos recursos que ocupam os papéis dominantes na produção, mas eles não oferecem uma razão convincente de um critério ser bom para a seleção de um método de alocação de CIF.

Portanto, alguns autores de custos tratam a alocação de custos num contexto fora da realidade atual, onde, por exemplo, Martins (1979, p. 64) aloca todo o CIF em função de únicas e arbitrárias bases de rateio.

Sobre a razão convincente de um critério ser bom para a seleção de um método de alocação de CIF, Dutra (1991, p. 105) diz: "*que a melhor base de rateio para determinado custo é aquela que se supõe que o custo ocorra na mesma proporção dela, ou seja, o custo indireto deve guardar estreita correlação com os dados escolhidos como base de rateio.*"

A escolha do tema a ser abordado foi escolhido face a importância da utilização de recursos matemáticos como instrumento de fornecimento de informações contábeis.

De acordo com Iudícibus (1986, p. 283):

"Um dos desenvolvimentos mais acentuados notados na gestão moderna das empresas consiste no emprego cada vez mais crescente de matemática e estatística, bem como de outras metodologias mais avançadas, na solução de problemas empresariais."

Já o boletim IOB/TCB 07/95 que aborda a análise de regressão como um destes recursos matemáticos, afirma que:

"A análise de regressão é utilizada com certa freqüência por profissionais de várias áreas, além de contábil, como economistas, administradores e engenheiros, sendo cada vez maior sua adoção em perícias contábeis, inclusive em avaliações, assim como, especialmente, em análises de balanços, na contabilidade de custos, na contabilidade gerencial e em controladoria, apesar de ainda ser desconhecida ou pouco aplicada por grande número de profissionias da área contábil..."

A não muito tempo atrás, o contador mergulhava em papéis, fichas, livros e relatórios, a serem preenchidos para que o tratamento das informações contábeis chegassem dentro dos prazos previstos para a tomada de decisões.

Com o advento da informatização do fluxo de informações contábeis, que faz parte do crescente desenvolvimento tecnológico e da modernização que domina não só setores empresariais, houve, de um modo geral, uma redução das tarefas que antes eram executadas pelo contador.

Na busca da manutenção de sua própria existência e para minimizar essa drástica redução no campo do profissional de contabilidade no mercado de trabalho, a atuação em áreas de apoio, controle e decisão tornou-se mais presente, já que o contador passou a dispor de mais tempo para raciocinar, analisar e interpretar os informes contábeis.

Com a elaboração de dados contábeis cada vez mais depurados e com a utilização de técnicas cada vez mais sofisticadas como instrumento de auxílio a tomada de decisões, o contador desponta como uma peça fundamental na estrutura de qualquer organização. Dentre essas técnicas capazes de fornecer subsídios importantes para a solução de problemas que auxiliem no tratamento das informações contábeis existe a análise de regressão, que tem por objetivo estabelecer uma equação que descreva o relacionamento entre duas variáveis.

2 OBJETIVOS

Os objetivos que se pretendem atingir, são:

2.1 Objetivo geral

O objetivo deste estudo, como contribuição a área contábil, é demonstrar os princípios relacionados com a aplicação das técnicas estatísticas da análise de regressão/correlação como uma alternativa de minimizar arbitrariedades na seleção de critérios para rateio de CIF.

2. 2 Objetivos específicos

Os resultados específicos que se pretende atingir, são:

- Revisar a teoria estatística básica necessária a compreensão da análise de regressão /correlação.
- Descrever a análise de regressão/correlação através de uma abordagem simples e didática, elucidando a sua importância na área contábil, sua utilização, seus cálculos e suas particularidades;
- Estabelecer, baseado em um estudo de caso, uma ou mais variáveis que expliquem a variação do comportamento dos CIF;
- Verificar se a análise de regressão/correlação serve como uma tentativa de seleção de critérios de rateio.

3 METOLOGIA

Segundo Trujillo (1974): *"A ciência é todo um conjunto de atitudes e atividades racionais, dirigidas ao sistemático conhecimento com objeto limitado, capaz de ser submetido a verificação".*

A respeito da ligação da ciência com os métodos científicos Lakatos & Marconi (1985) afirmam que todas as ciências caracterizam-se pela utilização de métodos científicos, que significam o conjunto das atividades sistemáticas e racionais que com maior segurança e economia, permite alcançar o objetivo – conhecimento válidos e verdadeiros – traçando o caminho a ser seguido, detectando erros e auxiliando as decisões do cientista.

A propósito, este capítulo traçará a metodologia utilizada na elaboração desta pesquisa, elucidando a delimitação do universo bem como o tipo, o delineamento, as limitações e as delimitações da pesquisa.

3.1 Problemática

A justificativa da elaboração do presente trabalho decorre dos seguintes questionamentos.

Com a utilização de variáveis independentes, que tenham influência nos CIF, é possível aplicar a análise de regressão/correlação para estabelecer uma ou mais variáveis explicativas do comportamento dos CIF?

Esse modelo representa uma alternativa para selecionar quais critérios de rateio melhor relacionam os CIF com os produtos?

3.2 Delimitação e tipo da pesquisa

Cabe aqui a mesma dúvida de Iudícibus (1988, p. 11) quanto ao método de abordagem, profundidade e a delimitação do tratamento. Quanto a abordagem a utilizar, se a tradicional, descritiva e dedutiva, ou uma outra mais pragmática e indutiva. Quanto a profundidade, se utilizar um modelo simplificado, sem muitas variáveis para melhorar a didática do trabalho, ou um modelo mais complexo, de modo a poder abranger uma gama maior de aspectos e restrições.

A combinação destes fatores geram trabalhos com composições das mais variadas, onde determinada escolha possa, talvez, agradar a uns e a outros não. Na escolha de fatores, visando apresentar um trabalho eficaz de propósitos, foi selecionada uma abordagem simplificada através de um estudo de caso (monográfico). Sempre que for necessário uma forma de tratamento quantitativo dos problemas será realizado, enfatizando, todavia, a análise em si, sendo a metodologia empregada apenas um meio mais eficiente de se atingir os fins.

3.3 População

A estruturação da análise terá embasamento num caso prático onde foi selecionada uma empresa situada na região da grande Florianópolis, classificada como indústria de transformação do gênero material elétrico e de comunicações.

Por motivos de proteção e segurança das informações relatadas o nome da empresa será omitido. Apesar de aparentemente não fazer muito sentido essa condição da empresa para fornecimento dos dados, ela afirma que existe uma verdadeira batalha campal das concorrentes para obtenção de informações sigilosas. Mas, para superar essa condição, adotar-se-á para fins didáticos a nomeclatura: Cia GranFpolis.

3.4 Delineamento da pesquisa

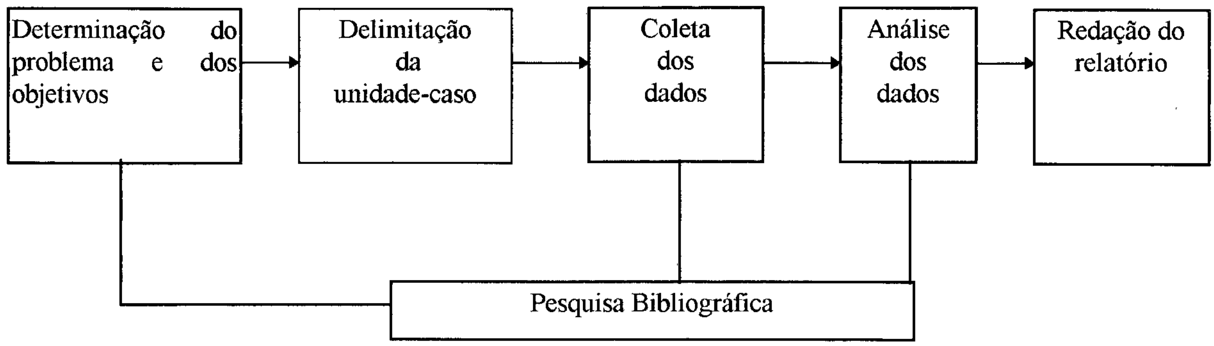
A previsão de análise e interpretação de coleta de dados seguirá um caminho flexível, onde se utilizará mais de um procedimento.

A coleta de dados desta pesquisa será com base na seleção de bibliografias e análise de dados da empresa obtido através de entrevistas e documentos.

Quanto ao processo de análise e interpretação dos dados, não se pode falar em etapas para o estudo de caso (Gil, 1993).

A análise compreenderá, somente, aspectos essenciais do tema proposto, afim de não tornar excessivamente complexo a resolução do estudo. Convém ainda mencionar que a análise se utilizará de categorias analíticas derivadas de teorias que gozem de razoável grau de aceitação (Gil, 1993).

O delineamento da pesquisa não só se refere a previsão de análise e interpretação de coleta de dados (Gil, 1993), mas também a sua diagramação. Tendo como premissa que um estudo de caso caracteriza-se por grande flexibilidade, sendo impossível estabelecer um roteiro rígido que determine com precisão como deverá ser desenvolvida a pesquisa (Gil, 1993), para que haja um acompanhamento do desenvolvimento da pesquisa, elaborou-se o seguinte fluxograma da pesquisa:



3.5 Limitações da pesquisa

Existem vários fatores limitativos nesta pesquisa, que não invalidam no total ou em partes a generalização de suas finalidades.

Algumas limitações são referentes a dificuldade de materiais bibliográficos referentes ao assunto e ao tempo disponível para execução da pesquisa.

Outras limitações são conseqüentes:

- Do relacionamento dos CIF com as bases de rateio, já que devem estar relacionados em cada observação;
- Da ineficiência dos sistemas de custos: se os sistemas de custos da empresa operam com ineficiência, as distorções estabelecidas pelos dados coletados estarão incluídas na análise;
- Do estudo de caso envolvendo apenas uma empresa e com poucas variáveis, no sentido de que deve ser interpretado à luz das condições concretas e que foram obtidos os dados;
- Do coeficiente de correlação que não é uma medida absoluta, no sentido de que exprime uma relação independente da amostra que foi obtido (Nick & Kellner, 1971, p. 161);

– Do intervalo: a amostra só pode extrair informações dentro do intervalo relevante pois fora dele as informações são pouco confiáveis.

4 REVISÃO DA LITERATURA

4.1 Sistema de contabilidade de custos e principais filosofias adotadas

Resumidamente, os objetivos da contabilidade de custos, segundo Martins (1975), são:

– Determinação do lucro, auxiliando na avaliação de estoques e lucros globais;

– Controle das operações;

– Tomada de decisões.

Custeio, segundo Martins (1979, p. 208), significa, forma de apropriação de custos. Além disso, os métodos de custeio, são:

– *Custeio por Absorção: Todos os custos de produção são apropriados aos bens elaborados;*

– *Custeio Variável: Somente os custos variáveis são apropriados aos bens elaborados.*

Paranhos (1992, p. 12) diz que para se calcular um custo é preciso que, antes, sejam definidos a finalidade para a qual se destina e, como conseqüência, os critérios a serem seguidos no seu cálculo.

Quanto a apuração os custos podem ser classificados em diretos e indiretos. Segundo Martins (1979, p. 55) o custo direto é aquele que pode ser diretamente apropriado a cada tipo de bem ou órgão, no momento de sua ocorrência, isto é, está ligado diretamente a cada tipo de bem ou função de custo. Sobre o custo indireto, Dutra (1991, p. 35) diz que é aquele que não se pode apropriar diretamente a cada tipo de bem ou função de custo no momento da sua ocorrência. É um custo comum a muitos tipos diferentes de bens, sem que se possa separar a parcela referente a cada um, no momento da sua ocorrência. Pela falta de condição de uma medida objetiva, tal separação é efetuada através de rateio que, segundo Dutra (1991, p. 35), é uma divisão proporcional por uma base que tenha valores conhecidos em cada função e que se julga que o custo ocorre nas mesmas proporções da base.

4.1.1 Critérios de rateio nos CIF

Martins, (1979, p. 90) considera que:

"Todos os custos indiretos só podem ser apropriados, pela sua própria definição, de forma indireta aos produtos, isto é, mediante estimativas, critérios de rateio, previsão de comportamento dos custos, etc. Todas essas formas de distribuição contêm, em menor ou maior grau, um certo subjetivismo; Portanto, a arbitrariedade sempre vai existir nessas alocações, sendo que às vezes ela existirá em nível bastante aceitável, e em outras oportunidades só a aceitamos por não haver alternativas melhores."

Portanto, qualquer tentativa de alocação tem que ser feita de maneira estimada, subjetiva, proporcional a uma base de rateio que tenha relação fiel com o custo indireto.

Na escolha de uma ou outra base de rateio altera-se por completo os resultados de CIF, conseqüentemente, altera-se, também, os resultados de custos totais e do lucro, provocando distorções nas análises da empresa. Desse modo, Martins (1979, p. 65) diz que:

"Valores de custos indiretos diferentes e conseqüentes custos totais também diferentes para cada produto podem não só provocar análises distorcidas, como também diminuir o grau de credibilidade com relação as informações de custos". Do mesmo modo, Dearden (1976, p. 57) diz que *"As técnicas para alocar os custos podem ter uma influência significativa nas informações geradas pelo sistema de contabilidade de custos; Conseqüentemente, considerações cuidadosas devem ser feitas a respeito do tipo de técnicas de alocação que são usadas."*

4.1.2 Formas de rateio

Dutra (1991, p. 35) observa que uma das maneiras de efetuar o rateio é: eleger a melhor base de distribuição, entre as disponíveis para o custo; dividir o total a ser rateado pelo total da base escolhida, para obter o coeficiente de rateio; multiplicar o coeficiente de rateio por cada um dos valores componentes do total da base, obtendo-se a parcela do custo indireto atribuível a cada função.

Outra forma de rateio é o rateio em cascata,

"Onde a soma dos custos específicos de cada função, quando existem mais de um custo, é rateada tomando-se como base os valores das colunas seguintes, existentes na mesma linha, e as parcelas obtidas são somadas aos valores existentes em cada coluna (funções) desta linha. Em seguida, o novo valor da próxima coluna será rateado pelos valores da linha nas demais colunas. Estes valores são somados aos já existentes na linha e que serviram de base de rateio. O novo valor da coluna seguinte será rateado pelos valores das colunas restantes na linha, e assim

sucessivamente, até que todos os custos se concentrarão nas últimas colunas, correspondentes as últimas funções. (Dutra, 1991, p. 109).”

Uma outra maneira de efetuar o rateio é através de taxas pré-determinadas de custos indiretos muito usado quando a produção é nova ou em encomendas que a empresa tem necessidade de fixar por antecipação os CIF e o volume de produção. Os coeficientes são estimados, Florentino (1988, p. 174) tendo em vista produções passadas ou cálculos e orçamentos efetuados.

Vale lembrar que o "feeling" do contador é de suma importância na escolha da melhor base de rateio, assim como a experiência e o conhecimento de todo processo de estruturação dos custos.

Complementando, Dutra (1991, p. 105) faz um comentário importante para o prosseguimento da pesquisa:

"*Entre as bases disponíveis elege-se aquela considerada como a melhor para o custo que será rateado, em função da afinidade entre os dois, por se supor que o custo tem 'correlação' direta com a base escolhida.*"

4.2 Conceitos básicos de estatística

Neste tópico, tentar-se-á desenvolver os aspectos mais relevantes de estatística visando dar sustentação ao objetivo da pesquisa.

4.2.1 Definição de estatística e sua divisão

Segundo Cunha (1978, p. 12) a estatística significa o conjunto de processos usados na coleta, organização, resumo, comunicação, análise e interpretação de dados.

Spiegel (1985, p. 13) diz que: *"A estatística está interessada nos métodos científicos para coleta, organização, resumo, apresentação e análise de dados bem como na obtenção de conclusões válidas e na tomada de decisões razoáveis baseadas em tais análises".*

A estatística divide-se, basicamente, em dois ramos:

- Estatística descritiva: destina-se a organização, resumo, simplificação de informações, e segundo Spiegel (1985, p. 14) descrevendo e analisando um certo grupo, sem tirar quaisquer conclusões ou inferências sobre um grupo maior;
- Inferência estatística: diz respeito a análise e interpretação de dados amostrais. Segundo Spiegel (1985, p. 14) se uma amostra é representativa de uma população, conclusões importantes sobre a população podem ser inferidas de sua análise. É a parte da estatística que trata das condições sob as quais essas inferências são válidas.

Existe ainda um outro grupo entre esses dois ramos, a probabilidade, que proporciona uma base racional para lidar com situações influenciadas por fatores relacionados com o acaso. Spiegel comenta que como a inferência não pode ser absolutamente certa, a linguagem da probabilidade é muitas vezes usada, no estabelecimento, das conclusões.

"A estatística é aplicável a qualquer ramo do conhecimento onde se manipulem dados experimentais." (Costa Neto, 1977, p. 1).

Portanto, a estatística como um conjunto de métodos que podem ser aplicados a qualquer ciência, é uma ferramenta poderosa diante da incerteza.

4.2.2 Medidas de posição

As medidas de posição mais importantes são as medidas de tendência central; Elas são usadas para descrever um único número que represente melhor um conjunto particular de dados.

As três medidas mais usadas, são:

A média aritmética: É a soma dos itens de uma série de números dividida pela quantidade de itens dessa série.

$\bar{x} = \frac{\sum xi}{n}$ Para dados não agrupados $\bar{x} = \frac{\sum xifi}{n}$ Para dados agrupados

Propriedades da média:

A) A média de um conjunto de números pode ser sempre calculada;
B) Para um dado conjunto de números, a média é única;
C) A média é sensível a todos os valores do conjunto, isto é, se alterarmos qualquer item na série a média mudará;
D) Somando-se uma constante a cada valor do conjunto, a média ficará aumentada do valor dessa constante;
E) A soma dos desvios dos números de um conjunto a contar da média é zero.

Exemplo:

xi	$\bar{x}$	xi - $\bar{x}$
0	5	– 5
0	5	– 5
5	5	0
10	5	5
10	5	5
	Σ	0

Mediana: Ela representa o ponto central de uma série de números, dividindo-o em dois grupos onde um terá valores inferiores a mediana, e o outro grupo superiores a mediana.

A posição da mediana pode ser determinada pela fórmula: $\frac{n+1}{2}$

Exemplo: Dado a série de números, qual a sua mediana?

1 2 3 4 6 7 8 9 9 $\quad R = \frac{n+1}{2} = \frac{9+1}{2} = 5$

portanto, o 5º elemento da série é a mediana.

Moda: Ela representa o valor que ocorre com maior freqüência numa série de números.

Exemplo: Dada a série abaixo, qual a sua moda?

2, <u>5</u>, 8, <u>5</u>, 6, <u>5</u>, 6, 4, <u>5</u>

R: Como o número 5 repete com maior freqüência (4 vezes), ele é a moda.

4.2.3 Medidas de dispersão

As medidas de dispersão indicam em uma série se os itens estão relativamente próximos uns dos outros ou separados.

Amplitude total ou intervalo: É a diferença entre o maior e o menor número de uma série.

Exemplificando, dada a série 5 7 9 10 11, qual a sua amplitude total?

R: 11 – 5 = 6

Desvio médio absoluto (D. M. A): é o desvio médio absoluto dos itens em relação a média de uma série.

$$D.M.A = \frac{\sum |xi - \bar{x}|}{n}$$

Exemplo: Seja a série 2, 4, 6, 8, 10, calcule o D. M. A.

Primeiro: acha-se a média da série.

$$\bar{x} = \frac{\sum x}{n} = \frac{30}{5} = 6$$

Segundo: constrói-se a tabela subtraindo de cada item a média aritmética e depois soma-se os desvios absolutos.

xi	$\|xi - \bar{x}\|$
2	4
4	2
6	0
8	2
10	4
	12

Terceiro: divide-se $\sum |xi - \bar{x}| = 12$ por n:

$$DMA = \frac{\sum |xi - \bar{x}|}{n} = \frac{12}{5} = 2,41$$

Variância (Sx^2): É o quadrado dos desvios dos itens em relação a média.

$$Sx^2 = \frac{\sum (xi - \bar{x})^2}{n}$$

Desvio-padrão: é a raiz quadrada da variância.

$$Sx = \frac{\sqrt{\sum (xi - \bar{x})^2}}{n}$$

4.2.4 Probabilidade

Segundo Stevenson as probabilidades são utilizadas para exprimir a chance de ocorrência de determinado evento.

P (A): probabilidade de um evento A, representada por um número de 0 a 1, que indica a chance de ocorrência do evento A.

Neto (p. 53) diz que *"a chance pode ser encarada como a interação de grande número de fatores, que influem coletivamente no resultado de um experimento".* Num lançamento de moeda existem fatores que podem influenciar no resultado, como por exemplo a força com que se lança.

Neto complementa, ainda que, admitindo-se que os mesmos fatores atuam, na maioria das vezes, de maneira análoga em observações repetidas, certos resultados podem ter mais probabilidade de ocorrência que outros.

4.2.5 Variável aleatória

"Uma variável aleatória é uma função com valores numéricos cujo valor é determinado por fatores de chance." (Neto, p. 54).

Portanto, é uma variável quantitativa que representa os resultados obtidos de uma amostra.

As variáveis aleatórias podem ser discretas ou contínuas. São discretas quando os valores possíveis formam um conjunto finito e podem ser contados; são contínuas quando os valores possíveis não podem ser contados, somente medidos. Portanto, os valores possíveis formam um conjunto infinito.

4.2.6 Distribuição de probabilidade

"É uma distribuição de freqüências relativas para os resultados possíveis de uma variável aleatória. Representa a probabilidade da variável aleatória assumir cada um dos diversos valores." (Stevenson, 1981, p. 101).

São classificadas como distribuições discretas de probabilidade, entre outras, a distribuição de Bernouilli, distribuição de Poisson, e a distribuição binomial, e são distribuições de probabilidade contínuas, a distribuição uniforme, a distribuição exponencial e a distribuição normal.

De acordo com Neto (p. 56) a aplicação de determinada distribuição a um problema depende do grau de aproximação entre a situação real e o conjunto de condições admitidas na distribuição de probabilidades.

A distribuição normal destacar-se-á pela importância que tem sobre a pesquisa.

4.2.6.1 Distribuição normal

A distribuição normal também é conhecida como distribuição de Gauss, Laplace ou distribuição Laplace-Gauss.

É utilizada tanto para desenvolvimento teórico quanto aplicado da estatística. (Neto, p. 63).

Na distribuição normal a partir de uma amostra são inferidos os parâmetros, isto é, os valores da população. (Cunha, 1978).

Características da distribuição normal:

A) Forma de sino;

B) Simétrica em relação a média;

C) Prolonga-se de menos infinito para mais infinito;

D) É completamente especificada pela média e desvio-padrão;

E) A área sob a curva representa 100% das probabilidades associadas a variável;

F) A probabilidade da variável aleatória assumir exatamente um determinado valor é zero. Assim, as probabilidades se referem a intervalos de valores.

G) A área sob a curva entre a média e um ponto qualquer é igual a função da distância entre a média e o ponto, em número de vezes o desvio-padrão.

H) A probabilidade da variável aleatória assumir um valor de um intervalo é igual a área sob a curva naquele intervalo.

4.2.6.2 Distribuição normal reduzida

Admita-se que, sob uma certa hipótese, a distribuição normal de uma estatística X é normal com média μ e desvio-padrão σ. Então, a distribuição da variável reduzida dado por $Z = \frac{x-u}{\sigma}$ é a distribuição normal reduzida. (Neto, p. 85).

A utilização da forma normal reduzida conta com a vantagem do valor de Z ser tabelado.

Nesse caso, diz-se que Z é normalmente distribuído com média 0 e variância 1.

Exemplificando: Seja x: N (100, 25) calcular P (100<x<106).

Solução:

Se $\sigma^2 = 25$ então $\sigma = 5$ e $\mu = 100$

para $x_1 = 100$ então $Z = \frac{100 - 100}{5} = 0$

para $x_2 = 106$ então $Z = \frac{106 - 100}{5} = 1{,}2$

Da tabela para Z=1,2 então: 0,3849

Logo, P(100<x<106) = P(0<Z<1,2) = 0 + 0,3849 = 0,3849

4.2.7 Testes de hipóteses

O objetivo do teste de hipóteses é avaliar afirmações sobre parâmetros populacionais baseando-se nos resultados de uma amostra. Neto (p. 82) afirma que o problema fundamental da estatística é o teste de hipóteses sobre as populações, ou seja, avaliar afirmações sobre os valores de parâmetros da população.

Ao tentar a fixação de decisões e conveniente a formulação de suposições acerca das populações interessadas. Essas suposições, que podem ser verdadeiras ou falsas, são denominadas hipóteses estatísticas e, em geral, constituem considerações acerca das distribuições de probabilidade das populações (Neto, p. 82).

Neto diz que:

"Em alguns casos, formula-se uma hipótese estatística com único propósito de rejeitá-la. Por exemplo, se se deseja decidir se uma moeda é viciada, formula-se a hipótese de que ela não o seja, isto é, p=0,5, em que é a probabilidade de caras. De modo semelhante, se se deseja decidir se um

processo é melhor que outro, formula-se a hipótese que não há diferença entre eles."

Essas hipóteses que dizem que a afirmação é verdadeira são denominadas de hipótese nula. E qualquer hipótese diferente da hipótese nula é chamada de hipótese alternativa.

Exemplo: Uma firma de auditoria deveria verificar pelo menos 10% das contas de uma grande companhia. Para se acompanhar o trabalho dos auditores foi retirada uma amostra de 200 contas, das quais 15 tinham sido verificadas por eles. Pode-se dizer que os auditores estejam checando menos que o necessário com 8% de significância?

Solução:

1º passo: Formulação das hipóteses:

$$\begin{cases} \text{hipótese nula: H0: } p = 0{,}05 \\ \text{hipótese alternativa: H1 : } p > 0{,}05 \end{cases}$$

Obs: H_0: p = 0,05 pega-se sempre a pior situação possível para a hipótese nula. Não se pega $p \leq 0{,}05$.

2º passo: Definição do nível de significância do teste.

α = nível de significância – é a probabilidade da hipótese nula ser rejeitada quando verdadeira (Aceito H_0).

Na tabela t: $\begin{cases} \alpha = 16\%(0{,}08 \times 2) \\ G.L = \infty \end{cases}$ Então, $Z\alpha = -1{,}41$

onde: G.L é o número de graus de liberdade.

3º passo: Comparação Z_{teste} (Amostra) com $Z\alpha$ (Significância).

$$Z_{teste} = \frac{\overline{P} - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}}$$ onde, P_0 = P da hipótese nula H_0.

$$\begin{cases} Se\ Z_{teste} > Z\alpha & aceita - se\ H_0 \\ Se\ Z_{teste} < Z\alpha & rejeita - se\ H_0 \end{cases}$$

No exemplo, $Z_{teste} = \dfrac{0{,}075 - 0{,}10}{\sqrt{\dfrac{0{,}10 \times 0.9}{n}}} = -1{,}1785$

Portanto, $Z_{teste} > Z\alpha$ então aceito H_0 com 8% de significância, ou seja, não se pode dizer que eles estão checando menos contas que o necessário.

4.2.7.1 Testes unilaterais e bilaterais

Para qualquer tipo de hipóteses tem-se três testes $\begin{cases} \text{ou p} > \\ \text{ou p} < \\ \text{ou p} \neq \end{cases}$

– Teste unilateral à direita $\begin{cases} \text{H0}: p = 5\% \\ \text{H1}: p > 5\% \end{cases}$

Verifica-se, neste caso, se o padrão máximo não foi extrapolado.

Então, $\begin{cases} \text{Se } Z_{teste} < Z\alpha & \text{aceita - se } H_0 \\ \text{Se } Z_{teste} > Z\alpha & \text{rejeita - se } H_0 \end{cases}$

– Teste unilateral à esquerda $\begin{cases} \text{H0}: p = 5\% \\ \text{H1}: p < 5\% \end{cases}$

Verifica-se, neste caso, se o padrão mínimo não foi extrapolado.

Então, $\begin{cases} \text{Se } Z_{teste} > Z\alpha & \text{aceita - se } H_0 \\ \text{Se } Z_{teste} < Z\alpha & \text{rejeita - se } H_0 \end{cases}$

– Teste bilateral

$$\begin{cases} H_0 : p = 5\% \\ H_1 : p \neq 5\% \end{cases}$$

Verifica-se, neste caso, variações em ambas as direções.

Então, $\begin{cases} Se\ |Zteste| < Z\alpha/2 & aceita-se\ H_0 \\ Se\ |Zteste| > Z\alpha/2 & rejeita-se\ H_0 \end{cases}$

4.2.7.2 Distribuição t de Student

Quando o desvio padrão da população não é conhecido usa-se o desvio-padrão da amostra como estimativa, substituindo-se α por λ.

Como isso, a distribuição amostral passa a ser uma distribuição t de Student.

Isso não acarreta maiores dificuldades, pois o desvio-padrão amostral dá uma aproximação bastante razoável do verdadeiro valor, na maioria dos casos.

Sabe-se que quando o tamanho da amostra é superior a 30 a distribuição das médias é quase normal. Entretanto, para amostras menores a aproximação normal não é indicada. Para resolver esse problema é que usamos a distribuição t de Student.

Apesar da semelhança com a distribuição normal, a distribuição t tem maior área nas caudas e mais dispersa pois depende do tamanho da amostra.

4.2.8 Correlação e regressão

Existem métodos estatísticos que lidam com uma única variável (x) como os vistos até o momento. Entretanto, muitos problemas de estatística envolvem duas ou mais variáveis. Estudar-se-á num primeiro momento a relação que se estabelece entre as duas variáveis quantitativas.

Um problema de correlação surge quando se quer saber se há alguma relação entre um par de variáveis (Neto, p. 100). Ilustrando, peso e estatura de um indivíduo são duas variáveis. Sem dúvida, existe relação entre peso e estatura, pois quanto mais alto for o indivíduo maior será seu peso. Nota-se na relação das duas variáveis que a medida que aumenta a estatura, aumenta o peso. Diz-se, então, que a variável peso é a variável dependente, pois o peso depende da estatura. E a variável estatura é a variável independente.

Entretanto, existem muitas exceções e restrições a regra. Existem pessoas de baixa estatura com pesos maiores. Existem pessoas altas com pouco peso. Dessa forma, as correlações medem a força ou o grau de relacionamento entre as duas variáveis.

A teoria da correlação e da regressão estão intimamente ligadas. Segundo Neto (p. 101): *"A correlação mede a força ou grau de relacionamento entre duas variáveis e a regressão dá uma equação que descreve o relacionamento em termos matemáticos."*

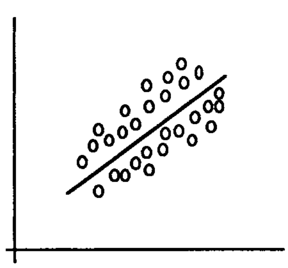
4.2.8.1 Tipos de relacionamento entre duas variáveis (x, y)

Levin (1987, p. 276) afirma que pode-se visualizar as diferenças quanto ao grau de correlação, por meio de um diagrama de dispersão, que é um gráfico capaz de mostrar a maneira pela qual os valores de duas variáveis, x e y, distribuem-se ao longo da faixa dos possíveis resultados.

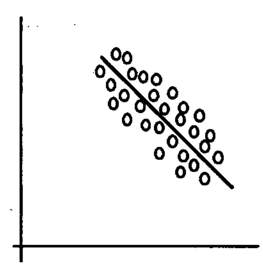
A vantagem de se construir o diagrama de dispersão está em que, muitas vezes, sua simples observação já mostra uma idéia razoável de como as duas variáveis se correlacionam. (Costa Neto, 1977, p. 178).

Diz-se que o grau de relacionamento entre x e y aumenta a medida que os pontos, no diagrama de dispersão, se agrupam mais compactados em torno de uma linha reta imaginária (Levin, 1987, p. 276). Portanto, existe relacionamento entre as variáveis x e y quando y varia de acordo com x.

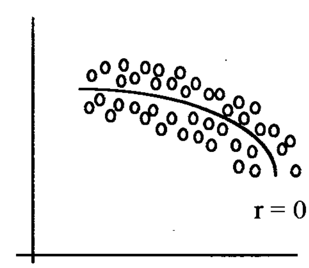
Quanto ao sentido da correlação, ela pode ser positiva quando y tende a variar proporcionalmente a x, isto é, para maiores valores de x a tendência é observarem-se maiores valores de y. Ou ela pode ser negativa quando y tender a variar inversamente proporcional a x, isto é, para maiores valores de x a tendência é observarem-se menores valores de y e vice-versa.



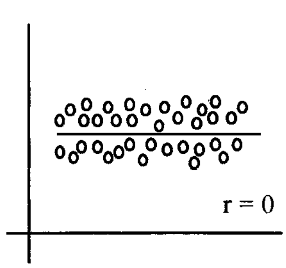
(a) correlação linear positiva



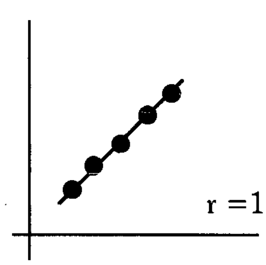
(b) correlação linear negativa



(c) correlação não-linear



(d) não existe correlação



(e) correlação perfeita

Diagrama de Dispersão

"Pode-se determinar de modo qualitativo, quão bem uma certa curva ou reta representa a relação entre as variáveis, mediante a observação direta, do próprio diagrama de dispersão. Essa curva ou reta é denominada de ajustamento então diz-se que há uma relação não-linear conforme o caso, se o ajustamento se aproxima a uma curva ou a uma reta." (Neto, p. 100).

Para que se possa expressar o grau e o sentido da correlação em termos numéricos utilizar-se-á o coeficiente de regressão linear de Pearson, que nada mais é do que um índice de correlação obtido a partir dos desvios reduzidos. Nick & Kellner (1971, p. 147) dizem que do conceito de correlação deduz que quanto maior a correlação mais próximos devem ser os desvios reduzidos do mesmo indivíduo em ambas as variáveis. Com efeito, uma correlação positiva perfeita implica que o mesmo indivíduo esteja igualmente afastado da média em ambas as variáveis, isto é, que os seus desvios reduzidos sejam iguais. *"A medida que diminui a correlação, maior é a discrepância entre os desvios reduzidos em ambas as variáveis."* Nick & Kellner (1971, p. 147).

O coeficiente de correlação r ou r(x, y) introduzido por Karl Pearson, calculado a partir de uma amostra de n pares de observações de x e y, mede a quantidade de dispersão em torno da equação linear ajustada através do método dos mínimos quadrados, ou o grau de relação das variáveis, na amostra. Fonseca et alli (1985, p. 11).

A fórmula do coeficiente de correlação r depois de desenvolvida, é: $r = \frac{Sxy}{\sqrt{Sxx \cdot Syy}}$

onde: $Sxy = \sum xy - \frac{\sum x \sum y}{n}$

$$Sxx = \sum x^2 - \frac{\left(\sum x\right)^2}{n}$$

$$Syy = \sum y^2 - \frac{\left(\sum y\right)^2}{n}$$ onde $-1 \leq r \leq 1$.

O coeficiente de correlação r está compreendido no intervalo que varia de -1 a $+1$. Quanto maior o ajustamento da reta proposta aos pontos do diagrama de dispersão mais próximo de $+1$ ou -1 estará o valor de r. E quanto mais próximo de zero estiver, menor será a relação linear.

4.2.8.2 Propriedades de r

- O coeficiente de correlação é adimensional, isto é, independe das unidades de medida das variáveis x e y;
- O coeficiente de correlação independe da origem em relação a qual os valores que o compõem são calculados, ou seja, alterando-se com um valor constante uma variável o coeficiente de correlação não se altera.

4.2.8.3 Teste de significância de r

Quando se extrai uma amostra aleatória de uma particular população, é necessário verificar se a associação obtida entre x e y existe de fato na população, e não resulta de erro amostral. Portanto, para testar a significância de uma medida de correlação geralmente estabelece-se a hipótese de que não existe correlação na população. Com isso, basta consultar a tabela F, onde figura uma lista de valores significantes do r aos níveis de significância de 0,05 e 0,01, associados a graus de liberdade que vão de 1 a 90.

Comparando o resultado entre o valor do r calculado e o valor do r tabelado (Tabela F):

Se $r_{calculado} < r_{tabelado}$ então aceita-se a hipótese nula de que $r=0$

Se $r_{calculado} \geq r_{tabelado}$ então rejeita-se a hipótese nula e aceita-se a hipótese de que existe correlação na população. (Levin, 1987, p. 286).

4.2.8.4 Requisitos para o uso do coeficiente r

Levin (1987, p. 288), diz que o coeficiente r precisa de alguns requisitos a fim de permitir inferências:

– Correlação linear: o r só se aplica a correlações lineares entre x e y;

– Dados intervalares: as variáveis x e y devem ser mensuradas, no mínimo, a nível intervalar, de sorte que seja possível trabalhar com escores.

– Amostragem casual: os sujeitos amostrais devem ter sido extraídos aleatoriamente de uma dada população. Se assim não for, não terá nenhum sentido a prova de significância do coeficiente obtido.

– Variáveis distribuídas normalmente: quando as amostras são pequenas ($n < 30$) qualquer descuido na observância dessa normalidade de distribuição pode comprometer seriamente a validade do r de pessoas.

4.2.8.5 Análise de regressão

"Muitas vezes a posição dos pontos no diagrama de dispersão sugere a existência de uma relação funcional entre as duas variáveis. Surge então o problema de se determinar uma função que exprima esse relacionamento." (Costa Neto, 1977, p. 188).

Suponha-se a existência de duas variáveis, por exemplo, x = estatura e y = peso. É notório que existe correlação entre elas, quanto maior a estatura, maior será o peso. É notório também que os indivíduos de uma estatura determinada, variarão menos de peso entre si, do que os indivíduos de todas as estaturas. Por exemplo, 12 indivíduos que tem 1,65m de estatura, seus pesos variarão menos do que 12 indivíduos de todas as estaturas.

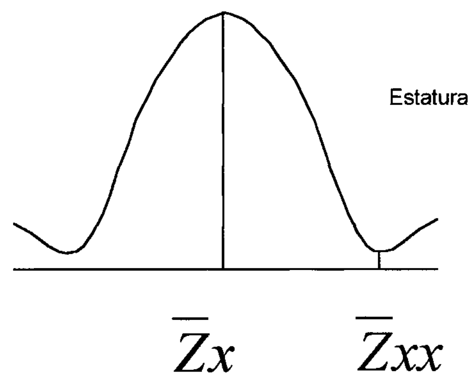
Considere como amostra um grupo de pessoas de estaturas bem elevadas. Investigando quais os pesos deste grupo de pessoas, encontrar-se-á uma distribuição tal que os pesos se concentrem em torno de um valor médio. Mas esse valor médio para pesos será maior do que a média geral de pesos considerados para todas as estaturas.

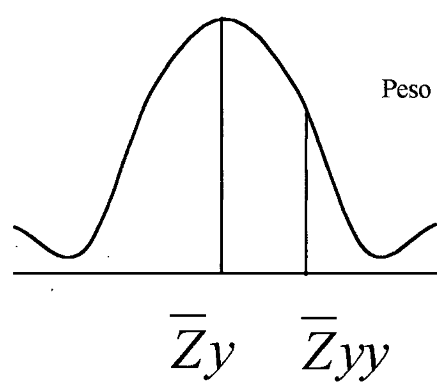
Cabe agora perguntar: esse valor médio para pesos estará tão afastado em relação a média geral de pesos, quanto o valor médio de estaturas bem elevadas em relação a média de todas as estaturas?

Não! Pois o valor médio para pesos desta amostra estará mais próxima da média geral de pesos do que a estatura desta amostra estará para todas as estaturas.

Este fenômeno por Galton recebe o nome de Regressão.

Graficamente:





onde: $\overline{Z}x =$ Média dos desvios reduzidos em relação à estatura total.

$\overline{Z}xx =$ Desvio reduzido correspondente a estatura elevada.

$\overline{Z}y =$ Média dos desvios reduzidos em relação ao peso total.

$\overline{Z}yy =$ Média dos desvios reduzidos correspondentes aos pesos de todos os indivíduos com estatura elevada.

Analisando o gráfico, a distância $\overline{Z}xx - \overline{Z}x$ é maior que a distância $\overline{Z}yy - \overline{Z}y$, portanto a média dos pesos manifesta regressão para a média total.

Se houvesse correlação perfeita as duas distâncias seriam iguais, e não haveria regressão.

Se houvesse regressão absoluta, a média $\overline{Z}xx$ seria igual a $\overline{Z}x$, a média $\overline{Z}yy$ seria igual a $\overline{Z}y$.

4.2.8.6 Regressão linear simples

Para conhecer o valor da regressão utiliza-se a função que melhor descreva a relação entre as variáveis. Várias funções satisfazem tais condições e a análise estatística indica qual a melhor. (Cunha, 1978, p. 180).

A relação mais simples entre duas variáveis é a linear: $y = a + bx$, onde a e b são parâmetros desconhecidos que indicam o intercepto e a declividade.

Na Regressão, dado um conjunto de pontos dispostos no diagrama de dispersão, imagina-se uma reta representativa desses pontos e utiliza-se a equação da reta $y_c = a + bx$ para descrever a reta que melhor represente esses pontos.

Portanto, na reta de regressão $y_c = a + bx$ tem-se que:

a = intercepto, que indica quando a reta corta o eixo y (valor de y quando $x = 0$)

b = declividade, indica a inclinação da reta

y_c = variável dependente: variável que se pretende estimar.

X = variável independente: causa da variação de y no modelo.

Para a determinação da reta de regressão utiliza-se o método dos mínimos quadrados que estima os parâmetros a e b de forma a minimizar a soma dos desvios ao quadrado, cujas fórmulas são:

$$b = \frac{Sxy}{Sxx}; \qquad a = \bar{y} - b\bar{x}$$

Onde:

$$Sxy = \sum(x-\bar{x})(y-\bar{y}), \qquad \bar{x} = \frac{\sum x}{n}$$

$$Sxx = \sum(x-\bar{x})^2, \qquad \bar{y} = \frac{\sum y}{n}$$

4.2.8.7 Teste de hipótese t para existência de regressão

O procedimento para se efetuar os testes de hipótese:

– Enunciar as hipóteses: $\begin{cases} H_0 : \beta = 0 \\ H_1 : \beta \neq 0 \end{cases}$

– Fixação do nível de significância (α) e escolha da variável t de Student com n - 2 graus de liberdade.

– Cálculo da variável $t = \dfrac{b}{\dfrac{S}{\sqrt{Sxx}}}$

Lembre-se que $b = \dfrac{Sxy}{Sxx}$ e $S^2 = \dfrac{Syy - bSxy}{n-2}$

– Conclusão:

Se $-t\alpha/2 < t < t\alpha/2$ aceita-se H_0, e conclui-se com risco α que não há regressão.

Se $|t| > t\alpha/2$ rejeita-se H_0, e conclui-se que existe regressão.

4.2.8.8 Intervalos de confiança

Intervalo de confiança para β:

$$P(b - \sigma t\alpha/2 \leq \beta \leq b + \sigma_2 t\alpha/2) = 1-\alpha \quad onde\ \sigma = \frac{S}{\sqrt{Sxx}}$$

Intervalo de confiança para α:

$$P(a - \sigma \alpha t\alpha / 2 \leq \alpha \leq a + \sigma_2 t\alpha / 2) = 1 - \alpha \quad onde\ \sigma = S\sqrt{\frac{1}{n} + \frac{\bar{x}}{Sxx}}$$

Intervalo de confiança para um valor individual y_i, dado um x_i:

$$P(y_i - \sigma \alpha t\alpha / 2 \leq \alpha \leq y_i + \sigma_2 t\alpha / 2) = 1 - \alpha$$

$$onde\ \sigma_y = S\sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{Sxx}}$$

4.2.8.9 Coeficiente de determinação

Além dos testes de hipóteses e dos intervalos de confiança, outro indicador que fornece elementos para a análise do modelo adotado é o coeficiente de determinação (Fonseca, 1985, p. 100), definido por:

$$R^2 = \frac{b^2 Sxx}{Syy}$$

Este coeficiente indica quantos por cento a variação explicada pela regressão representa da variação total, isso porque uma parcela da variação total é devida a fatores aleatórios. Devemos ter: $0 \leq R^2 \leq 1$

Portanto, no caso em que $R^2 = 1$ a reta estimada se ajusta perfeitamente aos pontos. Dir-se-á que o ajuste é perfeito. As variações de y são 100% explicadas pelas variações de x através da função especificada, não havendo desvios em torno da função estimada.

Entretanto, se $R^2 = 0$, conclui-se que as variações de y são exclusivamente aleatórias, e a introdução da variável x no modelo não incorporará informação alguma sobre as variações.

Então, quanto mais próximo de 1 maior será a variação explicada pela regressão.

Além desses indicadores, segundo Iudícibus (1986, p. 287), a análise de regressão precisa de algumas premissas afim de permitir inferências:

– Deve existir linearidade entre x e y;

– O desvio entre o valor real de y e o calculado pela função deve ter um valor médio esperado igual a 0 (zero);

– O desvio-padrão e a variância dos (x e y) devem ser constantes, onde y = reta calculada.

– O desvio de um ponto sobre o diagrama de dispersão não está relacionado com o desvio de outro ponto (não existe correlação serial);

– Os pontos ao redor da linha de regressão são normalmente distribuídos;

Exemplo de aplicação: (Fonseca, 1985, p. 100-104)

Suponha que exista uma relação linear entre as variáveis: x= despesas com propaganda e y = vendas de um certo produto.

Considerando os dados da tabela abaixo, determine:

a) A reta de mínimos quadrados;

b) Os testes de hipótese, intervalos e coeficiente de explicação.

x = milhões de reais

y = milhares de unidade

	x	y	xy	x^2	y^2
	1,5	120	180	2,25	14400
	5,5	290	1045	30,25	36100
	10,0	240	2400	100,00	57600
	3,0	140	420	9,00	19600
	7,5	180	1350	56,25	32400
	5,0	150	750	25,00	22500
	13,0	280	3640	169,00	78400
	4,0	110	440	16,00	12100
	9,0	210	1890	81,00	44100
	12,5	220	2750	156,25	48400
	15	310	4650	225,00	96100
Σ	86,0	2150	19515	870,00	461700

$$\bar{y} = \frac{\sum y}{n} = \frac{2150}{11} = 195,45 \qquad \bar{x} = \frac{\sum x}{n} = \frac{86}{11} = 7,82$$

$$Sxy = \sum xy - \frac{\sum x \sum y}{n} = 19515 - \frac{(86)(2150)}{11} = 2705,91$$

$$Sxx = \sum x^2 - \frac{\left(\sum x\right)^2}{n} = 870 - \frac{(86)^2}{11} = 197,64$$

$$Syy = \sum y^2 - \frac{\left(\sum y\right)^2}{n} = 461700 - \frac{(2150)^2}{11} = 41472,73$$

Logo: $b = \frac{Sxy}{Sxx} = \frac{2705,91}{197,64} = 13,69$

a = $\bar{y} - b\bar{x}$ = 195,45 – (13,69)(7,82)=88,39

$$\boxed{y = 88,39 + 13,69x}$$

é a reta estimada

Quanto ao teste t para existência de regressão:

I) $H_0 : \beta = 0$

$H_1 : \beta \neq 0$

II) Fixando-se $\alpha = 5\%$ e escolhendo a variável t com 9 graus de liberdade, pois n = 11 e G.L = n – 2.

Então, na tabela $t_{\alpha\text{-}2} = 2{,}2622$

III) Cálculo da variável $t = \dfrac{b}{\dfrac{S}{\sqrt{Sxx}}}$

Então: $S^2 = \dfrac{Syy - bSxy}{n-2} = \dfrac{41472{,}43 - (13{,}69)(2705{,}91)}{9} = 492{,}06$

Logo: $S = \sqrt{492{,}06} = 22{,}18$

Então: $t = \dfrac{13{,}69}{\dfrac{22{,}18}{\sqrt{197{,}64}}} = 8{,}71$

IV) Conclusão: Como $|8{,}71| > 2{,}2622$, rejeita-se H_0, e conclui-se com risco de 5% que existe regressão.

Intervalo de confiança: $1 - \alpha = 95\%$ pois $\alpha = 5\%$.

I.C para β:

$\sigma = \dfrac{S}{\sqrt{Sxx}} = \dfrac{22{,}18}{\sqrt{197{,}84}} = 1{,}57$, como:

$P(b - \sigma t\alpha/2 \leq \beta \leq b + \sigma t\alpha/2) = 1 - \alpha$, então:

$P(13{,}69 - 1{,}57 \times 2{,}2622 \leq \beta \leq 13{,}69 + 1{,}57 \times 2{,}2622) = 95\%$, então:

$P(10{,}14 \leq \beta \leq 17{,}24) = 95\%$

O intervalo acima contém o verdadeiro coeficiente de regressão com 95% de confiança.

I.C. para α:

$\sigma\alpha = S\sqrt{\dfrac{1}{n} + \dfrac{\bar{x}^2}{Sxx}} = 22{,}18\sqrt{\dfrac{1}{11} + \dfrac{(7{,}82)^2}{197{,}64}} = 13{,}97$

$P(b - \sigma t\alpha / 2 \leq \alpha \leq b + \sigma t\alpha / 2) = 1 - \alpha$, então:

$P(88{,}39 - 13{,}97 \times 2{,}26 \leq \alpha \leq 88{,}39 + 13{,}97 \times 2{,}26) = 95\%$, então:

$P(56{,}81 \leq \alpha \leq 119{,}97) = 95\%$

O intervalo acima contém o verdadeiro intercepto da reta com 95% de confiança.

I.C para y:

Neste caso, calcularemos I.C. para y quando x = 7, ou seja, o I.C. para a venda quando investe-se 7 milhões de reais:

$$\sigma y = S\sqrt{1 + \frac{1}{n} + \frac{(xi - x)^2}{Sxx}} = 22{,}18\sqrt{1 + \frac{1}{11} + \frac{(7 - 7{,}82)^2}{197{,}64}} = 23{,}07$$

$P(y - \sigma t\alpha / 2 \leq y_1 \leq y_1 + \sigma t\alpha / 2) = 1 - \alpha$, então:

$P(88{,}39 - 13{,}97 \times 2{,}26 \leq y \leq 88{,}39 + 13{,}97 \times 2{,}26) = 95\%$, então:

$P(56{,}81 \leq y \leq 119{,}97) = 95\%$

Coeficiente de determinação:

$$R^2 = \frac{b^2 Sxx}{Syy} = (13{,}69)^2 * \frac{197{,}64}{41.472{,}73} = 0{,}89 \; ou \; 89\%$$

Portanto, esse modelo explica 89% da variação total de y.

4.2.8.10 Regressão linear múltipla

A regressão linear múltipla envolve mais de duas variáveis, onde uma única variável é dependente e as outras restantes independentes. Toda a teoria é uma extensão da análise de regressão linear simples.

Adotar-se-á a abreviatura I.C. para intervalo de confiança.

"Sempre a análise tem por objetivo estabelecer uma equação que possa ser usada para predizer valores de y para valores dados das diversas variáveis independentes. A finalidade das variáveis independentes adicionais é melhorar a capacidade de predição em confronto com a regressão linear simples." (Neto, p. 105).

A equação de regressão tem a forma:

$y = a + b_1x_1 + b_2x_2 + ... + b_nx_n$

Onde n = número de variáveis independentes

4.2.8.11 Regressão não-linear

Além da regressão linear $y = a + bx$ existem outras regressões não-lineares. Todavia, as relações não-lineares podem, às vezes, para facilitar, serem transformadas em lineares mediante simples transformação de suas variáveis. (Neto, p. 106). As mais usadas para linearizar a relação entre as variáveis são:

$y = ax^b$ curva geométrica

$y = ab^x$ curva exponencial

$y = a + b_1x + b_2x^2$ parábola

$y = \frac{1}{\alpha + \beta x}$ hipérbole

5 ANÁLISE E INTERPRETAÇÃO DOS DADOS

5.1 Informações da empresa

Para a estruturação da análise torna-se necessário a descrição de dados obtidos através de entrevistas dentro da empresa GranFpolis.

A empresa GranFpolis caracteriza-se como uma indústria nacional de transformação que apresenta fluxo de produção contínua e tecnologia simples. Sua produção e as vendas tem aumentado relativamente nos últimos anos e deve permanecer para os próximos anos.

Quanto a sua atuação, ela atua somente no mercado nacional, com preços estáveis, sem problemas de sazonalidades. Além disso, a empresa em questão, apresenta processos tecnológicos de fabricação padronizados e um sistema contábil capaz de gerar e processar os dados necessários a análise.

Quanto a mecânica da produção, ela se processa com linha de montagem convergente.

Quanto aos seus produtos, ela fabrica três produtos diferenciados, com ciclo de vidas curtos.

Quanto a filosofia de custeio ela utiliza o custeio por absorção.

Quanto ao método de custeio a empresa não emprega nenhum método mais apurado e nem possui departamentos definidos em função dos produtos. A apropriação dos custos diretos baseiam-se na utilização de apontamentos para a mão-de-obra e na utilização de requisições para a matéria-prima. Já os custos indiretos são feitos rateios diretos item por item de custo aos produtos, e sua composição média para o ano de 1994 ficou assim compreendida:

Custo indireto	%aproximado
Mão-de-obra indireta	30
Depreciação	20
Energia Elétrica	26
Aluguéis e impostos	14
Materiais diversos	6
Outros custos	4
	100

A M.O.I. da empresa é referente a pagamento de honorários (35%) e salários os chefes de equipes de produção (65%), e é rateada com base no valor da mão-de-obra direta.

As informações de custos relativos ao ano de 1994, são:

Mês	MP	MOD	CIF	Custos Totais	Horas/ MOD	Horas/ MOTot	MOI	Total M.O	A	B	C	Total
1	12.310	13.060	5.730	21.100	612	115.5	1.770	14.830	320	388	302	1010
2	12.500	13.270	5.800	21.570	654	125.0	1.855	15.125	333	399	310	1042
3	12.080	13.220	5.690	20.990	644	104.0	1.710	14.930	316	372	302	990
4	12.670	13.310	5.860	21.840	662	133.5	1.760	15.070	341	408	317	1066
5	13.040	13.480	7.130	23.650	696	152.0	2.130	15.610	391	462	376	1229
6	13.120	13.570	7.180	23.870	714	156.0	2.150	15.720	396	472	381	1249
7	12.920	13.360	6.920	23.200	672	146.0	2.420	15.780	384	452	353	1189
8	12.980	13.410	7.040	23.430	682	149.0	1.970	15.380	388	492	329	1209
9	13.200	13.690	7.230	24.120	738	160.0	2.460	16.150	400	486	386	1272
10	13.380	13.840	7.400	24.620	768	169.0	2.370	16.210	424	506	387	1317
11	13.340	13.620	7.330	24.290	724	167.0	2.270	15.890	418	480	389	1287
12	13.280	13.750	7.290	24.320	750	164.0	2.040	15.790	418	492	380	1290
Total	154.820	161.580	80.600	397.000	8.316	17.410	24.905	186.485	4.529	5.409	4.212	14150

5.2 A escolha da relação de critérios afins aos CIF

Inicialmente, derivado de literaturas de custos, "feeling" do contador e informações da empresa elege-se os critérios de rateio que mais tenham afinidade com o tipo de CIF.

Adotar-se-á as abreviaturas MOD e MOI para mão-de-obra direta e mão-de-obra indireta, respectivamente.

No caso da MOI da GranFpolis constata-se que os melhores critérios são:

– percentagem de mão-de-obra total em reais

– percentagem MOD em reais

– taxa-hora – MOD

A subjetividade de escolha dos critérios de rateio que melhor representem os CIF é visivelmente minimizada em função da capacidade e experiência do contador.

Portanto, a concentração e potencialização do trabalho do contador deve basear nas providências possíveis para assegurar que os fatores excluídos do modelo não afetam materialmente o rateio dos CIF.

Todavia, é impossível a determinação de um critério específico que leve em consideração todos os fatores que possam ser relevantes a determinado tipo de CIF.

5.3 A utilização do modelo $y_n = a + bx$

Segundo Carsberg (1972, p.16) um modelo, essencialmente, é um artifício usado para simular um processo do mundo real. Se o modelo é bom, ele auxilia a prever um estado do mundo. Alguns modelos habilitam a fazer previsões porque representam uma analogia física aproximada com uma situação real.

Portanto, tentar-se-á utilizar a correlação/regressão para construir um modelo $y_n = a + bx$ capaz de selecionar critérios de rateio de CIF para a empresa GranFpolis.

Carsberg (1972, p. 17) diz ainda que um modelo pode ser mais abstrato em sua natureza quando toma a forma de um gráfico ou equação matemática utilizada para afirmar a existência de relação entre certas variáveis. Mais abstrato porque um modelo matemático não é uma definição que expresse uma relação que deveria vigorar exatamente e sim uma relação para dar uma idéia mais ou menos aproximada da situação real. É o que ocorre com a correlação/regressão no modelo $y_n = a + bx$.

A utilidade de um modelo depende de sua representatividade como simplificação das relações que abrangem alguma situação real (Carsberg, 1972, p. 18).

A representatividade do modelo $y_n = a + bx$ na relação CIF X Critério de rateio baseia-se na escolha de alguns critérios de rateio que tenham mais afinidades com os CIF. Para a escolha desses critérios é imprescindível o "feeling" do contador, a sua experiência e o seu conhecimento íntimo das informações que afetam os CIF. Portanto, a causa da relação é detectada anteriormente pelo contador, significando que os CIF podem ser explicados por qualquer dos critérios escolhidos.

O modelo entra em cena somente para selecionar, dentre os critérios escolhidos, o que melhor se correlaciona com os CIF.

Entretanto, Carsberg (1972, p. 17) diz ainda que as conclusões baseadas no desempenho de um modelo não devem ser aceitas , no entanto, a menos, que exista uma boa razão para acreditar que seja ele digno de confiança (num nível satisfatório de aproximação).

Mattessich (1966, p. 516) sugere que as hipóteses em contabilidade e assuntos correlatos deveriam depender da avaliação das diversas considerações aplicadas ao teste de hipóteses científicas.

Uma hipótese científica é refutada por uma instância em que ela se provar válida. Uma hipótese de contabilidade deve ser encarada como inválida somente quando uma hipótese concorrente possa ser encontrada.

Quanto a validade da utilização do modelo nas conclusões da análise, ainda, vale observar que:

- A seleção de critérios através do modelo pode ser interpretada apenas como a escolha da melhor base de rateio para a empresa tendo em vista a série de critérios efetivamente considerados.
- As incerteza associadas aos diversos critérios de rateio tornam praticamente impossível evidenciar conclusivamente qual a base de rateio ótima.

5.4 Seleção dos critérios de rateio escolhidos

A construção do primeiro modelo a ser analisado, percentagem de mão-de-obra direta em reais, será todo feito manualmente para fins didáticos.

x = mil reais y = mil reais

n = mês	x = M.O.I	y = M.O.D	x.y	x^2	y^2
1	1,770	13,06	23,116	3,133	170,56
2	1,855	13,27	24,616	3,441	176,09
3	1,710	13,22	22,606	2,924	174,77
4	1,760	13,31	23,426	3,098	177,16
5	2,130	13,48	28,712	4,537	181,71
6	2,150	13,57	29,175	4,623	184,14
7	2,420	13,36	32,331	5,856	178,49
8	1,970	13,41	26,418	3,881	179,83
9	2,460	13,69	33,677	6,052	187,42
10	2,370	13,84	32,801	5,612	191,52
11	2,270	13,62	30,917	5,153	185,50
12	2,040	13,75	28,050	4,162	189,06
Σ	24,905	161,58	335,835	52,472	2176,28

Cálculo das médias:

$$\bar{y} = \frac{\sum y}{n} = \frac{161,58}{12} = 13,465$$

$$\bar{x} \frac{\sum x}{n} = \frac{24,905}{12} = 2,075$$

Cálculo do coeficiente de correlação r:

$$Sxx = \sum x^2 - \frac{\left(\sum x\right)^2}{n} = 52,472 - \frac{(24,905)^2}{12} = 0,784$$

$$Syy = \sum y^2 - \frac{\left(\sum y\right)^2}{n} = 2176,28 - \frac{(161,58)^2}{12} = 0,605$$

$$Sxy = \sum xy - \frac{\left(\sum x \sum y\right)^2}{n} = 335,835 - \frac{(24,905)(161,58)}{12} = 0,489$$

$$r = \frac{Sxy}{\sqrt{SxxSyy}} = \frac{0,489}{\sqrt{0,784 \times 0,605}} = 0,710$$

5.5 Teste de significância de r

Para testar a significância de r basta consultar a tabela F, para G.L = (n – 2) = 12 – 2 = 10 e nível de 0,05, verifica-se que um r para ser significante e, assim levar à rejeição da hipótese nula, deve ser no mínimo, igual a: 0,576.

r calculado = 0,71

r tabelado = 0,576

Como r calculado > r tabelado então rejeita-se a hipótese nula e aceita-se a hipótese alternativa de que existe correlação na população, e não resulta de erro amostral.

5.6 Análise de regressão

Agora para se achar a reta de regressão, utiliza-se os parâmetros a e b:

$$b = \frac{Sxy}{Sxx} = \frac{0,489}{0,784} = 0,6237$$

portanto, $\boxed{\bar{y} \cong 12,1708 + 0,6237\bar{x}}$ é a reta de regressão.

Quanto ao teste t para existência de regressão:

Seja $H_0 : \beta = 0$

Fixando α = 5% e escolhendo a variável t com 10 graus de liberdade, pois n = 12 e G.L = n – 2.

Então, na tabela t de Student, $t\alpha/2$ = 2,228.

Calculando-se: $S^2 = \frac{Syy - bSxy}{n-2} = \frac{0,605 - (0,6237)(0,489)}{12-2} = 0,03$

Logo, $S = \sqrt{0,03} = 0,1732$

Então: $t = \frac{b}{\frac{S}{\sqrt{Sxx}}} = \frac{0,6237}{\frac{0,1732}{\sqrt{0,784}}} = 3,188$

Portanto, como 3,188 > 2,228, rejeita-se H_0, e conclui-se com risco de 5% que existe regressão.

5.7 Coeficiente de determinação

$R^2 = 0,710^2 = 0,5041$

Portanto, esse modelo explica 50,41% da variação total de y, e 40,59% é devida a fatores aleatórios.

Conclusão da análise do modelo: percentagem de MOD

Algumas bases para aceitar ou não as retas estimadas como válidas são as seguintes de acordo com Iudícibus (1986, p. 293):

– O coeficiente de determinação (r^2) deve ser alto;

No modelo o r^2 é razoável, cerca de 50% da variação da MOI explica-se pela variação da MOD.

– O valor de t deve ser alto, pelo menos igual a 2.

No modelo o valor de t é considerado bom, afiançando o coeficiente b.

Portanto, o modelo indica que a percentagem de MOD é um bom critério para rateio da mão-de-obra indireta.

A construção do 2º modelo, percentagem da mão-de-obra total em reais, será elaborada através da calculadora HP-12c da Hewlett-Packard.

x = mil reais y = mil horas

n = mês	x = MOI	y = M.O. Total
1	1,770	14,830
2	1,855	15,125
3	1,710	14,930
4	1,760	15,070
5	2,130	15,610
6	2,150	15,720
7	2,420	15,780
8	1,970	15,380
9	2,460	16,150
10	2,370	16,210
11	2,270	15,890
12	2,040	15,790

Para armazenar os dados de entrada, digita-se:

14830	ENTER	1770	Σ+
15125	ENTER	1855	Σ+
14930	ENTER	1710	Σ+
15070	ENTER	1760	Σ+
15610	ENTER	2130	Σ+
15720	ENTER	2150	Σ+
15780	ENTER	2420	Σ+
15380	ENTER	1970	Σ+
16150	ENTER	2460	Σ+
16210	ENTER	2370	Σ+
15890	ENTER	2270	Σ+
15790	ENTER	2040	Σ+

Para obter o coeficiente r:

Digita-se:	No visor aparecerá:
x><y	0,939

Para obter o parâmetro b:

Digita-se:	No visor aparecerá:
o g y, r	1,642

Para obter o parâmetro a:

Digita-se:	No visor aparecerá:
STO 00 g x, r	12,133
CHS RCL 0 x < >y ÷	

Conclusão da análise do modelo: mão-de-obra total.

A reta de regressão será:

$$\boxed{y = 12{,}133 + 1{,}642x}$$

O coeficiente de determinação (r^2) será: 0,882, um ótimo indicativo de variação total.

O valor de t será: 8,706, também reflete um ótimo coeficiente b.

Portanto, sem sombra de dúvidas, abandona-se o 1º modelo (percentagem de MOD) e adota-se esse novo modelo como critério de rateio da MOI.

Apesar de já se ter um bom modelo representativo da MOI, falta analisar a taxa-hora-mod. A resolução desse modelo efetuar-se-á através do programa Excel. Todavia, outras planilhas de cálculo, como por exemplo Lotus e Quatro-pró, chegam a resultados idênticos.

O painel de indicadores fornecidos pelo micro é bem maior que o de uma calculadora. Existem vários aplicativos para micro computadores para problemas desse tipo. Alguns são mais sofisticados do que o apresentado, outros, menos. O aqui reproduzido é um bom programa, faltando a parte gráfica.

Saída de Programa Excel

Coeficiente

x = 2 = taxa – hora – MOD
Y = 1 = CIF

R – coeficiente de correlação

1	0,799
0,799	1

Regressão

Coef	B
0	0,1384
2	0,6326

$R^2 = 0,6384$
Valor – t = 4,208

Conclusão da análise do modelo: taxa-hora-MOD

O modelo apresenta bons resultados tanto para o coeficiente de determinação (0,6384) quanto para o valor-t (4,208).

Conclusão: Em função de toda MOD empregada não ser paga proporcional à taxa-hora-total, o resultado não foi parecido ao segundo modelo. O valor de R e o valor – t foram piores que do segundo modelo, concluindo-se que o segundo modelo, realmente, é o mais indicado como representante do critério de rateio da MOI.

Os sistemas de alocação aqui mostrados são bem mais complexos do que aqueles normalmente empregados na prática que envolvem poucas ou somente uma base de rateio para todo CIF. De acordo com o BOL

IOB/TCB 47/95, todas as operações não são realmente iguais. Cada uma possui uma estrutura diferente, e diferentes fatores influenciam a sua estrutura.

Mas, a montagem de um sistema de alocação mais sofisticado envolvendo uma malha de equações representativas de cada operação, centro de custos ou departamentos hoje em dia não é só possível, mas também viável. Johnson & Kaplan (1993, p. 214) afirmam que

"A enorme expansão da capacidade de computação dotou os atuais projetistas de sistemas de contabilidade gerencial de oportunidades com que seus predecessores não poderiam sequer ter sonhado. Com muitos processos de produção sob controle direto de computadores digitais, a informação pode ser registrada em tempo real, para análise do desempenho operacional. Em ambiente altamente automatizado, praticamente toda transação pode ser captada na análise subseqüente... Ou seja, são agora viáveis os sistemas abrangentes para medir e determinar com precisão as demandas de recursos de cada produto em diferentes linhas."

Além disso eles, Johnson & Kaplan (1993, p. 192) dizem que *"Os custos de despesas gerais representam uma fração bem maior dos custos totais precisando, por isso, ser compreendidos e controlados com bem mais cuidado do que no passado".*

Convém, ainda, enfatizar as seguintes colocações finais sobre essas alocações:

- O padrão básico do comportamento dos CIF é linear;
- Para longos períodos há sempre a possibilidade de que modifiquem as relações causais identificadas no presente;

- A tendência no mundo atual é a substituição da MOD pela tecnologia;
- Um sistema de custos quanto mais representativo da realidade melhor estimará os custos de produção de longo prazo.

6 CONCLUSÃO

Esta pesquisa tornou-se um grande desafio em função das potencialidades que o assunto tem a oferecer. Partindo da premissa que este é um trabalho acadêmico a nível de graduação, a racionalização imperou sobre a sua abrangência.

O aspecto simplificado que foi adotado na pesquisa visa um melhor acompanhamento da utilização da análise de regressão/correlação. Contudo, deficiências inerentes a essa simplificação eram de se esperar, onde algumas vezes, tópicos, demonstrações ou particularidades foram omitidos ou pouco comentados.

Sobre o relacionamento dos fatos verificados com a teoria, Roussel (1978, p. 181) alerta que *"numa ciência há dois tipos de proposições empíricas. As que se referem a especiais questões de fato e as que se referem as leis induzidas das questões de fato."* Neste caso, a aplicação da técnica de regressão/correlação mesclou as proposições para obtenção da razão sobre os fatos, na medida em que foram sendo observados.

A estruturação da pesquisa foi delineada com os conhecimentos teóricos necessários ao acompanhamento da resolução do problema da Cia GranFpolis de selecionar a melhor base de rateio para M.O.I. A minimização das arbitrariedades é resultado da aplicação da análise de regressão/correlação que seleciona, dentre os critérios pré-estabelecidos como influentes, aquele que melhor relaciona-se com os CIF.

Os critérios para serem considerados influentes dependem, em última instância, da avaliação do contador, que é baseada em análises e informações da empresa. Dependendo de cada caso esses critérios serão selecionados para cada item de CIF, por centros ou por departamentos onde montando-se uma malha dessas informações através do uso do computador resultará num programa com os CIF e suas respectivas parcelas de rateio.

Porém, a pesquisa limitou-se a fazer a seleção de apenas um item dos custos indiretos, onde, no entanto, os argumentos não são afetados pela complexidade dos itens de custos selecionados.

Concluindo, almejou-se nesta pesquisa contribuir de alguma forma para uma melhor utilização dos instrumentos de tratamento de dados na área contábil.

REFERÊNCIAS BIBLIOGRÁFICAS

Boletim IOB/TCB. **Contabilidade e regressão linear.** nº 7 e 8/95.

Boletim IOB/TCB. **Identificação dos problemas do sistema de custos.** nº 47/95.

CARSBERG, Bryan. **Introdução a programação matemática para administração financeira.** Ed. Vozes, 1972.

COSTA NETO, Pedro Luiz de Oliveira. **Estatística**. São Paulo: Edgar Blüsher, 1977.

CUNHA, Suzana Ezequiel da. **Estatística descritiva na psicologia e na educação.** Rio de Janeiro: Florense – Universitária, 1988.

DEARDEN, John. **Análise de custos e orçamentos nas empresas.** Tradução por Emmanuel Rothenberg. 3 ed. Rio de Janeiro: Zahar, 1976.

DUTRA, René Gomes. **Custos: uma abordagem prática**. 3 ed. São Paulo: Atlas, 1971, p. 1971, p. 194-213.

FLORENTINO, Américo Matheus. **Custos: princípios, cálculo e contabilização.** 12 ed. Rio de Janeiro: Editora da Fundação Getúlio Vargas, 1988, p. 165-175.

FONSECA, Jairo Simon da ; MARTINS, Gilberto de Andrade; TOLEDO, Geraldo Luciano. **Estatística aplicada.** 2 ed. São Paulo: Atlas, 1985, p. 267.

GIL, Antônio Carlos. **Como elaborar projetos de pesquisa.** 3 ed. São Paulo: Atlas, 1991.

HORNGREN, Charles T. **Introdução a contabilidade gerencial**. Rio de Janeiro: Prentice-Hall do Brasil, 1985, p. 168-187.

_____. **Cost accounting: a managerial emphasis**. 4 ed. Englewood Cliffs: Prentice-Hall, 1977, cap. 25 e 27.

IUDÍCIBUS, Sérgio de. **Contabilidade Gerencial**. São Paulo: Atlhas, 1986, p. 309.

_____.**Análise de custos**. São Paulo: Atlas, 1988, p. 169.

JOHNSON, H. Thomas; KAPLAN, Robert S. **Contabilidade Gerencial: a restauração da relevância da contabilidade nas empresas.** Rio de Janeiro: Campus, 1993.

KAPLAN, Robert S. **Advanced management accouting.** Englewood Cliffs: Prentice-Hall, 1982. Cap. 3 e 4.

KLIEMMAN NETO, Francisco José. **A falência da contabilidade tradicional como instrumento gerencial face aos novos paradigmas da organização da produção** . 1994.

LAKATOS, Eva Maria; MARCONI, Marina de Andrade. **Fundamentos de metodologia científica**. São Paulo: Atlas, 1985.

LEVIN, Jack. **Estatística aplicada a ciências humanas.** 2 ed. São Paulo: Harbra, 1987.

MARION, José Carlos; SILVA, Laércio Batista da. **Contabilometria – revisão da teoria estatística básica**. Trabalho Set/84.

MARTINS, Eliseu. **Contabilidade de custos**. 1 ed. São Paulo: Atlas, 1979, p. 367.

MATTESSICH, R. **Modern accounting theory**. Engliwood Cliffs: M. Blacker, 1966, p. 516-517.

NETO, Hermenegildo Milani. **Introdução a contabilometria e revisão da teoria estatística básica.** Trabalho FEA/USP.

NICK, Eva; KELLNER, Sheila R. de. **Fundamentos de estatística para as ciências do comportamento.** 2 ed. Rio de Janeiro: Renes, 1971.

PARANHOS, José Luiz B. **Contabilidade decisorial: análise gerencial de custos e resultados.** São Paulo: STS, 1992.

PETERS, William S; SUMMERS, George W. **Análise estatística e processo decisório**. Rio de Janeiro: Fundação Getúlio Vargas, 1973. P. 686. Il. 21cm.

PYNDICK, Robert S.; RUBINFELD, Daniel L. **Econometric models and economic forecasts.**

POPPER, Karl. **A lógica da pesquisa científica**. São Paulo: Cultrix, 1972, p. 34.

PORTER, Michael E. **Estratégia competitiva**: **técnicas para análise de indústrias e da concorrência.** 5 ed. Rio de Janeiro: Campus, 1991.

ROUSSEL, Bertrand. **A análise da matéria**. Rio de Janeiro: Zahar, 1978, p. 181.

SEVERINO, Antônio J. **Metodologia do trabalho científico.** 19 ed. São Paulo: Cortez, 1993.

SPIEGEL, Murray R. **Estatística**. 2 ed. São Paulo: McGraw-Hill do Brasil, 1985.

STEVENSON, William J. **Estatística aplicada à administração.** São Paulo: Harper & Row.

TRUJILLO FERRARI, Alfonso. **Metodologia da ciência.** 3 ed. Rio de Janeiro: Kennedy, 1974.

ANEXO

Condução da análise na Empresa GRANFPOLIS

Segundo Porter (1991, p. 335) existem basicamente dois tipos de dados sobre indústrias: dados publicados e aqueles coletados em entrevistas com participantes e observadores da indústria (dados de campo).

Como fontes publicadas foram utilizados documentos e prospectos da Empresa. Como coleta de dados de campo utilizou-se a conferência cruzada e a verificação de dados de diferentes fontes interna.

Para resultado final das entrevistas, elaborou-se um questionário sobre informações gerais da Empresa.

Questionário

1. **Dados Históricos**

1.1 Identificação da Empresa

1.2 Constituição da Empresa

1.3 Classificação da Empresa quanto:

1.3.1 Ao meio social, político e legal;

1.3.2 Ao meio macroeconômico;

2 **Dados contábeis**

2.1 Sistema contábil

2.2 Sistema de custos

2.3 Almoxarifado

3 **Dados de produção**

3.1 Sistema produtivo

3.2 Fornecedores/compradores

3.3 Tecnologia de produção e distribuição

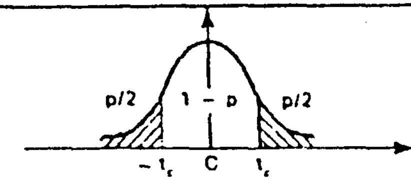
TABELA F Valores de *r* aos Níveis de Significância de 0,05 e 0,01.

gl	.05	.01
1	.99692	.999877
2	.95000	.990000
3	.8783	.95873
4	.8114	.91720
5	.7545	.8745
6	.7067	.8343
7	.6664	.7977
8	.6319	.7646
9	.6021	.7348
10	.5760	.7079
11	.5529	.6835
12	.5324	.6614
13	.5139	.6411
14	.4973	.6226
15	.4821	.6055
16	.4683	.5897
17	.4555	.5751
18	.4438	.5614
19	.4329	.5487
20	.4227	.5368
25	.3809	.4869
30	.3494	.4487
35	.3246	.4182
40	.3044	.3932
45	.2875	.3721
50	.2732	.3541
60	.2500	.3248
70	.2319	.3017
80	.2172	.2830
90	.2050	.2673

TÁBUA V

Distribuição de Student: St(n)

Valores críticos de t tais que $P(-t_C < t < t_C) = 1 - p$



Graus de liberdade	p = 90%	80%	70%	60%	50%	40%	30%	20%	10%	5%	4%	2%	1%	0.2%	0.1%	Graus de liberdade
1	0.156	0.325	0.510	0.727	1,000	1,376	1,963	3,078	6,314	12,706	15,894	31,821	63,657	318,309	636,619	1
2	0.142	0.289	0.445	0.617	0.816	1,061	1,386	1,886	2,920	4,303	4,849	6,965	9,925	22,327	31,598	2
3	0.137	0.277	0.424	0,564	0,765	0,978	1,250	1,638	2,353	3,182	3,482	4,541	5,841	10,214	12,924	3
4	0.134	0.271	0.414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	2,998	3,747	4,604	7,173	8,610	4
5	0,132	0.267	0.408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	2,756	3,365	4,032	5,893	6,869	5
6	0.131	0.265	0.404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	2,612	3,143	3,707	5,208	5,959	6
7	0.130	0.263	0.402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,517	2,998	3,499	4,785	5,408	7
8	0.130	0.262	0.399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,449	2,896	3,355	4,501	5,041	8
9	0.129	0.261	0.398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,398	2,821	3,250	4,297	4,781	9
10	0.129	0.260	0.397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,359	2,764	3,169	4,144	4,587	10
11	0.129	0.260	0.396	0.540	0.697	0.876	1,088	1,363	1,796	2,201	2,328	2,718	3,106	3,025	4,437	11
12	0.128	0.259	0.395	0.539	0.695	0.873	1,083	1,356	1,782	2,179	2,303	2,681	3,055	3,930	4,318	12
13	0.128	0.259	0,394	0.538	0.694	0.870	1,079	1,350	1,771	2,160	2,282	2,650	3,012	3,852	4,221	13
14	0.128	0.258	0.393	0.537	0.692	0.868	1,076	1,345	1,761	2,145	2,264	2,624	2,977	3,787	4,140	14
15	0.128	0.258	0.393	0.536	0.691	0.866	1,074	1,341	1,753	2,131	2,248	2,602	2,947	3,733	4,073	15
16	0.128	0.258	0.392	0.535	0,690	0,865	1,071	1,337	1,748	2,120	2,235	2,583	2,921	3,686	4,015	16
17	0.128	0.257	0.392	0.534	0,689	0,863	1,069	1,333	1,740	2,110	2,224	2,567	2,898	3,646	3,965	17
18	0.127	0.257	0.392	0.534	0,688	0,862	1,067	1,330	1,734	2,101	2,214	2,552	2,878	3,610	3,922	18
19	0.127	0.257	0.391	0.533	0,688	0,861	1,066	1,328	1,729	2,093	2,205	2,539	2,861	3,579	3,883	19
20	0.127	0.257	0.391	0.533	0,687	0,860	1,064	1,325	1,725	2,086	2,197	2,528	2,845	3,552	3,850	20
21	0.127	0.257	0.391	0.532	0,686	0,859	1,063	1,323	1,721	2,080	2,189	2,518	2,831	3,527	3,819	21
22	0.127	0.256	0.390	0.532	0,686	0,858	1,061	1,321	1,717	2,074	2,183	2,508	2,819	3,505	3,792	22
23	0.127	0.256	0.390	0.532	0,685	0,858	1,060	1,319	1,714	2,069	2,177	2,500	2,807	3,485	3,768	23
24	0.127	0.256	0.390	0.531	0,685	0,857	1,059	1,318	1,711	2,064	2,172	2,492	2,797	3,467	3,745	24
25	0.127	0.256	0.390	0.531	0,684	0,856	1,058	1,316	1,708	2,060	2,166	2,485	2,787	3,450	3,725	25
26	0.127	0.256	0.390	0.531	0,684	0,856	1,058	1,315	1,706	2,056	2,162	2,479	2,779	3,435	3,707	26
27	0.127	0.256	0.389	0.531	0,684	0,855	1,057	1,314	1,703	2,052	2,158	2,473	2,771	3,421	3,690	27
28	0.127	0.256	0.389	0.530	0,684	0,855	1,056	1,313	1,701	2,048	2,154	2,467	2,763	3,408	3,674	28
29	0.127	0.256	0.389	0.530	0,683	0,854	1,055	1,311	1,699	2,045	2,150	2,462	2,756	3,396	3,659	29
30	0.127	0.256	0.389	0.530	0,683	0,854	1,055	1,310	1,697	2,042	2,147	2,457	2,750	3,385	3,646	30
35	0.126	0.255	0.388	0.529	0,682	0,852	1,052	1,306	1,690	2,030	2,133	2,438	2,724	3,340	3,591	35
40	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1,050	1,303	1,684	[illegible]	[illegible]	[illegible]	2,704	[illegible]	[illegible]	40
50	0.126	0,254	0.387	0.528	0.679	0.849	1,047	1,299	1,676	2,009	2,109	2,403	2,678	3,261	3,496	50
60	0.126	0.254	0.387	0.527	0.679	0.848	1,045	1,296	1,671	2,000	2,099	2,390	2,660	3,232	3,460	60
120	0,126	0.254	0.386	0.526	0,677	0,845	1,041	1,289	1,658	1.980	2,076	2,358	2,617	3,160	3,373	120
∞	0.126	0,253	0,385	0.524	0.674	0,842	1,036	1,282	1,645	1,960	2,054	2,326	2,576	3,090	3,291	∞
	p = 90%	80%	70%	60%	50%	40%	30%	20%	10%	5%	4%	2%	1%	0.2%	0.1%	