

ANÁLISE QUANTITATIVA DE SUBJETIVIDADE: UM EXEMPLO DE CONCORDÂNCIA DE ATRIBUTOS

Elisa Maria Melo Silva

Instituto de Engenharia de Produção e Gestão, UNIFEI
Av. BPS, 1303, Bairro Pinheirinho, Itajubá - MG
lizzbr@gmail.com

David Custódio de Sena

Instituto de Engenharia de Produção e Gestão, UNIFEI
Av. BPS, 1303, Bairro Pinheirinho, Itajubá - MG
davidc.sena@gmail.com

Julio Cesar Mosquera

Instituto de Engenharia de Produção e Gestão, UNIFEI
Av. BPS, 1303, Bairro Pinheirinho, Itajubá - MG
jcmosqueragutierrez@gmail.com

José Henrique de Freitas Gomes

Instituto de Engenharia de Produção e Gestão, UNIFEI
Av. BPS, 1303, Bairro Pinheirinho, Itajubá - MG
ze_henriquefg@yahoo.com.br

RESUMO

A Análise de Concordância de Atributos avalia a concordância entre as medições feitas por avaliadores e seus conhecidos padrões, podendo determinar a precisão das medições e identificar quais itens tem as maiores taxas de erros de classificação. O objetivo deste trabalho é fazer uma revisão bibliográfica desta através de artigos nacionais e internacionais, desenvolver seu raciocínio e apresentar um exemplo de aplicação do método estatístico. O método de pesquisa utilizado foi a Pesquisa Axiomática em Gestão de Operações. Os resultados foram apresentados em três tabelas de avaliação de concordância: Cada Avaliador x Padrão, Entre Avaliadores e Todos os Avaliadores x Padrão. Em geral, os coeficientes Kappa e Kendall sugeriram uma boa concordância. Estes resultados foram importantes na comprovação de que dados subjetivos podem ser avaliados objetivamente através de ferramenta estatística com o uso de modelos matemáticos.

PALAVRAS-CHAVE Concordância de Atributos. Subjetividade. Coeficientes Kappa e Kendall.

ABSTRACT

The Attribute Agreement Analysis evaluates the correlation between the measurements taken by the appraisers and their known standard values which determine the accuracy of their measures and identify which items have the highest misclassification rates. This paper's objective is to review the existing literature through national and international articles, to develop their logic and to present an example of the application of the statistical method. The research method used was Axiomatic Research in Operations Management. The results were presented in three tables of agreement assessment: Each Appraiser x Standard, Between Appraisers and All Appraisers x Standard. In general, Kappa and Kendall coefficients suggested a good agreement. These results were important to verify that subjective data can be objectively assessed by statistical tools using mathematical models.

KEY-WORDS Attribute Agreement. Subjectivity. Kappa and Kendall Coefficients.

1. Introdução

A Análise de Concordância de Atributos é utilizada para avaliar a concordância entre as medições feitas por avaliadores e seus conhecidos padrões, podendo determinar a precisão das medições e identificar quais itens tem as maiores taxas de erros de classificação. A ferramenta tem a característica de avaliar dados comparando-os com seus padrões específicos, retirando o fator subjetividade da análise através de mecanismos estatísticos com o uso de um *software*.

A maioria das medições de processos industriais baseiam-se em escalas, micrômetros, entre outros objetos responsáveis pela mensuração das características físicas do produto. Durante a pesquisa para a nova versão de um automóvel, os clientes avaliam quais características podem permanecer ou mudar nesta nova versão do veículo. Em estudos de doenças psicossomáticas, onde exames médicos mais psicológicos são levados em consideração, o uso de ferramentas matemáticas se faz necessário para retirar o fator de subjetividade destes estudos para um melhor diagnóstico (CHATTOPADHYAY; ACHARYA, 2012). Todos os exemplos identificam fatores subjetivos destas análises, pesquisas, avaliações de características de qualidade tem uma importância no resultado final.

O objetivo deste trabalho é apresentar uma ferramenta capaz de analisar dados subjetivos na indústria ou na área de serviços e sua aplicação com um exemplo prático. Segundo Uchimura e Bosi (2002), a justificativa deste trabalho é que a qualidade em sua dimensão objetiva é mensurável (generalizável). O que não ocorre com a dimensão subjetiva já que esta habita o espaço das vivências, das emoções, do sentimento, não cabendo a quantificação pois expressam singularidades.

A organização deste trabalho se apresenta da seguinte maneira: primeiramente, uma introdução ao assunto na seção 1. A seguir (seção 2), uma revisão bibliográfica do tema Análise de Concordância de Atributos, sua relevância e suas aplicações. Na seção 3 uma breve descrição do método de pesquisa utilizado e na seção 4, o desenvolvimento. A seção 5 apresenta os resultados e a 6, a conclusão, respectivamente.

2. Revisão bibliográfica

Segundo Watson e Petrie (2010), uma questão comum em biologia reprodutiva é se, a medição de uma variável por dois métodos diferentes ou por duas pessoas diferentes que usam o mesmo método em tempos distintos ou uma pessoa que repete as medições em dois tempos diferentes, produzem essencialmente o mesmo resultado. A preocupação está no rigor (no modo como um valor observado corresponde ao valor real) e na precisão (conformidade observada na repetição das medições). Vários métodos foram utilizados para responder a questão com consequências insatisfatórias ou que levavam a conclusões enganosas.

Quando características de qualidade são consideradas robustas em estudos de *design*, elas apresentam uma natureza estática com valores fixados e associados a atributos físicos como peso, altura ou voltagem. Em alguns casos, estas propriedades podem ser sensoriais como cor ou aparência. Características de qualidade com atributos relacionados ao tempo são menos observadas frequentemente. E quando são estudadas, são relacionadas com propriedades intangíveis como durabilidade, confiabilidade, entre outras. Existem várias aplicações na indústria, entretanto, que requerem a investigação de características que fisicamente mudam ao longo do tempo. Por exemplo, engenheiros ambientais interessados na influência que diferentes filtros têm na concentração de impurezas encontradas na água durante um período de tempo. Ou a eficiência durante a implantação da ferramenta de qualidade 5S por um engenheiro de produção, qual o impacto desta nos operadores de uma indústria de autopeças durante um período de alguns meses após implantação. Em cada uma destas situações, a distribuição da característica de interesse muda ao longo de um determinado intervalo de tempo, portanto, as especificações, o valor alvo, bem como a média e variância do processo também estão mudando. O objetivo geral é, então, encontrar as configurações ideais dos fatores que minimizam os desvios em relação ao valor desejado ao longo do tempo ou perfil buscado (GOETHALS; CHO, 2011).

Para analisar de forma objetiva e controlar estas variáveis subjetivas, são utilizadas ferramentas estatísticas que com o uso de modelos matemáticos, conseguem chegar a uma

solução otimizada para questões como as descritas anteriormente. Segundo Rungtusanatham (2001), pode-se deduzir que o controle de processo estatístico e sua implantação representam mais do que uma intervenção tecnológica projetada para trazer melhorias no processo e na qualidade do produto. Para apreciar melhor essa inferência, percebe-se que qualquer processo de transformação é, fundamentalmente, uma sequência de atividades realizadas para converter insumos em produtos desejados. Na execução desta sequência de atividades de forma tão eficaz e eficientemente quanto possível, tanto a tecnologia quanto uma estrutura social ligando seus trabalhadores são essenciais. Portanto, quando um controle de processo estatístico é implantado em um processo de transformação, os seus potenciais efeitos, positivos ou negativos, não se limitam às alterações no subsistema tecnológico, e incluem alterações no subsistema social.

A ferramenta estatística, objeto de estudo deste trabalho, é a Análise de Concordância de Atributos. Esta avalia a uniformidade das respostas dentro de um grupo de avaliadores e entre eles e, em caso apropriado, comparar suas respostas com os valores padronizados ou referenciais (*standard values*). A análise utiliza as notas ou as classificações dos atributos. As medições são avaliações subjetivas feitas por pessoas e não por medições físicas diretas (ALVARADO, 2008).

Saleh, Hacker e Randhawa (2001) identificam um processo para descobrir os atributos de decisão críticos, que pode ser dividido em três etapas:

- Identificação dos atributos iniciais: o processo envolve o desenvolvimento de uma lista de atributos da literatura, reduzindo-os em subatributos (se apropriado), definição das operações e validação dessa lista com especialistas em decisão de capital da indústria ou do mundo acadêmico;
- Validação dos atributos na indústria: confirma ou modifica os subcomponentes iniciais identificados na decisão dos atributos na literatura baseados em dados reais da indústria;
- Determinação da importância relativa dos atributos: estes são classificados comparados aos outros atributos em termos de importância para decisões de investimento.

Ainda de acordo com Saleh, Hacker e Randhawa (2001), as duas últimas etapas são feitas através de pesquisas com grupos de gerentes industriais ligados ao processo, em que testaram a importância da decisão destes atributos descritos na literatura.

A ferramenta Análise de Concordância de Atributos busca a resposta as seguintes perguntas (ALVARADO, 2008):

- As respostas de um avaliador são uniformes?
Os diferentes avaliadores classificam os mesmos elementos de maneira uniforme? Existe concordância entre eles?
- As respostas dos avaliadores são exatas quando comparadas ao valor *standard* ou uma resposta correta?
- O sistema de medição de atributos é efetivo?
- Os operadores concordam com as notas de aprovação/reprovação sobre uma amostra de produto acabado?
- Estão corretas as classificações dos operários quando comparadas com o padrão fornecido pelo Coordenador da Qualidade?
- O sistema de inspeção visual é eficaz para detectar a falta de algum componente chave na montagem?

No entanto, para melhor interpretar estes resultados e melhor apoiar o poder de decisão, é utilizada esta ferramenta de um *software* estatístico.

3. Método

O método de pesquisa deste trabalho, segundo Bertrand e Fransoo (2002), é Pesquisa Axiomática em Gestão de Operações.

Entre as justificativas que comprovam este método, estão:

- Descrição condensada das características do processo operacional;
- Conceitos e termos aceitos como padrões na literatura científica de Gestão de Operações;

- A descrição do modelo conceitual se refere a recentes artigos que estudam os processos ou problemas estreitamente relacionados ao objeto de estudo;
- Relevância no estudo de um processo ou problema que já foi estudado, mas providencia uma nova, ou melhor, solução ao problema, ou aplicando novas técnicas ou alcançando melhores resultados com técnicas de soluções aceitáveis;
- Modelo científico: Simulação computadorizada;
- Provas matemáticas das análises.

4. Desenvolvimento

4.1 Análises estatísticas recomendadas

Medição de concordância, segundo Yang e Chinchilli (2011), é frequentemente utilizada na avaliação da reprodutibilidade de um novo ensaio, ou de um instrumento, a aceitabilidade de um processo novo ou genérico assim como no método de comparação. Exemplos incluem o acordo de dois métodos, dois avaliadores que fazem uma medição ao mesmo tempo, a concordância de valores observados com valores previstos, entre outros. Algumas boas práticas para estudos de *Gage R&R* (*Repeatability and Reproducibility*, que significa medidor de repetição e reprodutibilidade, é uma ferramenta estatística que mede a quantidade de variação no sistema de medição decorrente) por atributos são:

- Avaliadores devem avaliar aproximadamente o mesmo número de itens;
 - Estabelecer o padrão conhecido para cada item;
 - Selecionar pelo menos três avaliadores para o estudo;
 - Avaliadores devem avaliar os itens em condições normais;
 - Além de avaliar os itens em uma ordem aleatória para minimizar o viés.
- O *software* apresenta quatro tabelas com a análise de concordância por atributos:
- Entre os Avaliadores (*Within*) – aparece quando cada avaliador faz mais de um ensaio;
 - Cada Avaliador x Padrão – é apresentada quando o padrão é conhecido;
 - Entre Avaliadores (*Between*) – para quando mais de um avaliador apresenta uma resposta;
 - Todos Avaliadores x Padrão – aparece quando se especifica o padrão conhecido.

Todas as tabelas de concordância incluem número de inspeção, número de acerto, porcentagem de concordância e intervalo de confiança de 95%. A porcentagem de concordância é igual a $100 \frac{m}{N}$, onde m é o número de acertos e N é o número de inspeção.

O programa calcula o limite inferior da menor porcentagem do intervalo de confiança quando $\alpha = 0,05$ para a equação (1):

$$\frac{v_1 \cdot F_{v_1, v_2, 0,025}}{v_2 + v_1 \cdot F_{v_1, v_2, 0,025}} \quad (1)$$

Onde:

$$v_1 = 2m$$

$$v_2 = 2(N - m + 1)$$

$$F_{v_1, v_2, 0,025} = 2,5\% \text{ da distribuição F com graus de liberdade } v_1 \text{ e } v_2.$$

Se a porcentagem é igual à zero (sem concordância), o limite inferior é zero. Se a porcentagem é igual a um (concordância perfeita), utilize α ao invés de $\alpha/2$ na fórmula.

Para o limite superior do intervalo de confiança quando $\alpha = 0,05$ na equação (2), seguem:

$$\frac{v_1 \cdot F_{v_1, v_2, 0,975}}{v_2 + v_1 \cdot F_{v_1, v_2, 0,975}} \quad (2)$$

Onde:

$$v_1 = 2(m + 1)$$

$$v_2 = 2(N - 1)$$

$$F_{v_1, v_2, 0,975} = 97,5\% \text{ da distribuição F com graus de liberdade } v_1 \text{ e } v_2.$$

Se a porcentagem é igual à zero (sem concordância), utilize α ao invés de $\alpha/2$ na fórmula. Se a porcentagem é igual a um (concordância perfeita), o limite superior é igual a um.

No *software* utilizado, as análises de concordância entre atributos são feitas através de dois coeficientes, Kappa e Kendall. Cada qual apresenta características diferentes de avaliação apresentadas a seguir.

4.1.1 Estatística Kappa e sua interpretação

Conforme Burn e Weir (2011), a estatística Kappa é utilizada para avaliar à medida que a proporção de acordo dos observadores (*between*) ou entre os mesmos (*within*) é melhor que o acaso. Desta maneira, este índice é mais rigoroso que as correlações ou porcentagens brutas. Além disso, pode ser utilizado para avaliar a consistência e a concordância entre métodos alternativos de ensaio para medir a mesma variável ou entre avaliadores ao comparar suas aferições a padrões estabelecidos.

O valor Kappa representa o grau de concordância absoluta entre as qualificações, trata todas as qualificações incorretas da mesma maneira, sem levar em conta sua magnitude. E este é utilizado para classificações binárias, nominais ou ordinais (ALVARADO, 2008).

Encontrar uma boa aceitação, como numa porcentagem próxima a 100 e valor de Kappa próximo a 1, permite que as medições sejam feitas por diferentes avaliadores (ou pelo mesmo avaliador em momentos diferentes) com alguma confiança em sua consistência. No contrário, com baixas porcentagens de acordo ou valor de Kappa próximo ao zero, é útil para alertar os usuários sobre os sistemas de pontuação ou métodos de diagnóstico que exigem modificação, uma definição mais clara (BURN; WEIR, 2011). Segundo Alvarado (2008), valores de Kappa iguais ou superiores a 0,75 são considerados de bom a excelentes; valores menores que 0,40 indicam pouca concordância.

A fórmula (3) utilizada para Kappa é apresentada a seguir.

$$\frac{P_o - P_e}{1 - P_e} \quad (3)$$

Onde:

P_o : proporção de k vezes em que os avaliadores concordaram (equação (4)).

$$P_o = \frac{1}{Nn(n-1)} (\sum_{i=1}^N \sum_{j=1}^k x_{ij}^2 - Nn) \quad (4)$$

P_e : a proporção esperada de k vezes em que os avaliadores concordaram.

Calcula-se P_e para $\sum P_j^2$.

Onde:

P_j^2 : proporção esperada de concordância para cada categoria (equação (5));

$$P_j = \frac{1}{Nn} \sum_{i=1}^N x_{ij} \quad (5)$$

N: número de temas

n: número de avaliadores

k: número de categorias da escala

x_{ij} : número de avaliadores que atribuiu o tema *i* situado na categoria *j*

O coeficiente de Kappa para a concordância dos testes com o padrão conhecido é a média destes coeficientes Kappa.

4.1.2 Coeficiente de concordância de Kendall

O coeficiente de concordância de Kendall é uma medida de aquiescência entre alguns avaliadores que medem um determinado número de objetos. Dependendo do campo de aplicação, os avaliadores podem ser variáveis, pessoas, etc. (LEGENDRE, 2005).

Os valores de Kendall variam entre 0 e +1. Quanto maior seu valor, mais forte a associação entre os avaliadores ao medir as mesmas amostras. Geralmente, valores do coeficiente iguais ou maiores a 0,90 são considerados muito bons. Um alto valor deste índice significa que os avaliadores aplicam essencialmente o mesmo padrão quando avaliam as amostras.

Neste método, assume-se que uma única classificação em cada tema é feita por cada avaliador, e existem k avaliadores por tema. Então, para o cálculo de Kendall, os k avaliadores representam as classificações K para cada avaliador.

Suponha que os dados são dispostos numa tabela $k \times N$ com cada linha representando as classificações feitas por um avaliador específico para N temas. Quando o verdadeiro padrão não é conhecido, o programa estima o coeficiente de Kendall através da equação abaixo:

$$\frac{12 \sum_{i=1}^N R_i^2 - 3k^2N(N+1)^2}{k^2N(N^2-1) - k \sum_{j=1}^k T_j} \quad (6)$$

Onde:

N : número de temas

$\sum R_i^2$: a soma do quadrado das somas das classificações para cada N temas avaliados

K : número de avaliadores

T_j atribui a média das classificações às observações vinculadas; $T_j = \sum_{i=1}^{g_j} (t_i^3 - t_i)$ onde t_i é o número de classificações vinculadas ao grupo i de vínculos, e g_j é o número de grupos de vínculos no conjunto j de classificações.

4.1.3 Coeficiente de correlação de Kendall

Se for fornecida uma classificação conhecida para cada amostra, o *software* também calcula os coeficientes de correlação de Kendall, que são dados a cada avaliador para indicar a conformidade deste com o padrão conhecido e um coeficiente global representando todos avaliadores com os padrões.

O coeficiente da correlação auxilia na determinação da consistência de um avaliador, não analisando sua precisão.

Na descrição do método, sem perda de generalidade, assume-se que uma única classificação em cada tema é feita por cada avaliador, e há k avaliadores envolvidos por assunto. Então, para calcular o coeficiente de correlação de Kendall, k avaliadores representam k ensaios realizados por todos os avaliadores.

Quando o padrão verdadeiro é conhecido, o programa avalia o coeficiente de correlação de Kendall, calculando a média dos coeficientes do Kendall entre cada avaliador e o padrão. E o coeficiente de Kendall (7) é calculado entre cada ensaio e o padrão usando:

$$\frac{C-D}{SQRT \{ [n_{++}(n_{++}-1) \cdot 0,5 - T_X] [n_{++}(n_{++}-1) \cdot 0,5 - T_Y] \}} \quad (7)$$

Onde:

T_X : número de pares vinculados em $X = 0,5 \sum_i n_{i+} (n_{i+} - 1)$

T_Y : número de pares vinculados em $Y = 0,5 \sum_j n_{+j} (n_{+j} - 1)$

C : número de pares concordantes = $\sum_{i < k} \sum_{j < l} n_{ij} n_{kl}$

D : número de pares discordantes = $\sum_{i < k} \sum_{j > l} n_{ij} n_{kl}$

n_{i+} : número de observações na linha i

n_{+j} : número de observações na coluna j

n_{ij} : observação na célula correspondente à linha i e coluna j

n_{kl} : observação na célula correspondente à linha k e coluna l

n_{++} : número total de observações

O coeficiente de correlação de Kendall para a concordância entre os testes com o padrão conhecido é a média dos coeficientes de correlação de Kendall entre os ensaios.

4.2 Análise de concordância de atributos de dados binários

Para a análise do problema presente neste trabalho, foram utilizadas as funcionalidades do *software* Minitab 16.

4.2.1 Problema

Uma empresa de testes educacionais está treinando cinco novos avaliadores para a parte escrita de um teste padrão de determinada série. Cada avaliador irá corrigir quinze testes em uma escala de cinco pontos (-2, -1, 0, 1, 2). A empresa quer aferir a capacidade dos avaliadores aos padrões já estabelecidos nos testes.

Este estudo analisou 15 ensaios, cinco avaliadores, uma repetição e no total, 75 avaliações que foram exemplificadas na Tabela 1.

Tabela 1: Distribuição das amostras entre avaliadores.

Ordem da Avaliação	Amostra	Avaliador	Avaliação	Padrão
1	6	Duncan		1
2	5	Duncan		0
3	1	Duncan		2
4	10	Duncan		1
5	15	Duncan		1
...
73	15	Simpson		1
74	13	Simpson		2
75	10	Simpson		1

Fonte: Software estatístico

A empresa deseja avaliar, de acordo com Alvarado (2008):

- A uniformidade das respostas de cada avaliador;
- A uniformidade entre os diferentes avaliadores ao classificar as mesmas amostras;
- A exatidão das classificações dos avaliadores, de acordo com o padrão.

Para saber se os avaliadores estão qualificados, foi fornecida uma coluna com o padrão conhecido de cada amostra.

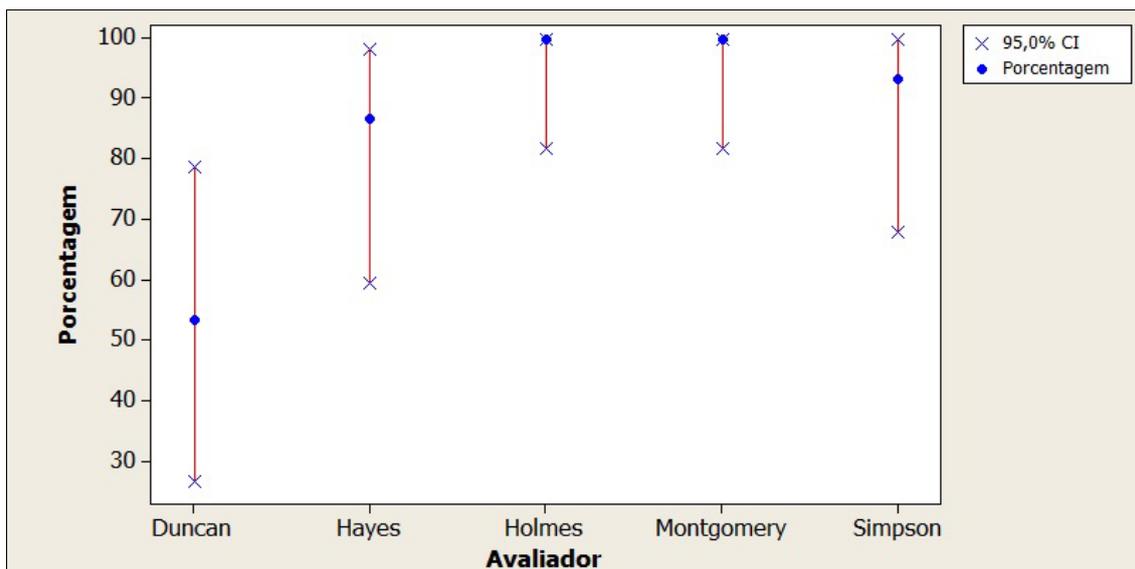
Os cinco avaliadores realizaram 15 ensaios uma vez cada, para um total de 75 avaliações. A ordem de ensaio lista a ordem em que as avaliações foram feitas (Tabela 1). A primeira avaliação é da amostra 6 feita por Duncan seguida da amostra 5. De forma padronizada, as avaliações foram aleatórias para um mesmo avaliador. Depois que o primeiro terminou suas avaliações, começou o segundo e assim, sucessivamente. Os valores de avaliação foram registrados na coluna Avaliação e comparados às normas da coluna Padrão, presentes na Tabela 1.

5. Interpretação dos resultados

O *software* apresentou três tabelas de avaliação de concordância: Cada Avaliador x Padrão, Entre Avaliadores e todos os Avaliadores x Padrão. Os coeficientes de Kappa e Kendall estão incluídos em cada uma destas tabelas. Em geral, estas estatísticas sugerem uma boa concordância. O coeficiente global de Kendall entre avaliadores é 0,966317 ($p = 0,0000$). O coeficiente de Kendall para todos os Avaliadores x Padrão é 0,958012 ($p = 0,0000$).

No entanto, um olhar mais atento na tabela Cada Avaliador x Padrão indica que as avaliações de Duncan e Hayes combinam muito pouco com o padrão. Já Holmes e Montgomery acertaram 15 de 15 avaliações, com 100% de acerto. Note que cada porcentagem de acertos está associada a um intervalo de confiança.

O gráfico Avaliador x Padrão (Figura 1) apresenta um *display* da avaliação da concordância para Cada Avaliador x Padrão. Pode-se facilmente comparar estas avaliações entre os 5 avaliadores.



Fonte: Software estatístico

Figura 1 – Gráfico Cada Avaliador x Padrão

Baseado nestas informações, Duncan, Hayes e Simpson precisam de treinamento adicional.

5.1 Cada Avaliador x Padrão

Comparando os valores reais aos padrões, Holmes e Montgomery tiveram 15 acertos em 15, representando uma porcentagem de 100% com um intervalo de confiança de 81,90 a 100,00 (Tabela 2).

Em ordem decrescente de porcentagem de acertos estão Simpson (93,33%), Hayes (86,67%) e Duncan (53,33%), presentes na Tabela 2. Os valores de Kappa Geral (Tabela 3) indicam que Duncan (0,41176) e Hayes (0,82955) cometeram mais erros em relação ao padrão.

Tabela 2 - Avaliação da Concordância

Avaliador	# Inspeção	# Acertos	Porcentagem	95% I.C
Duncan	15	8	53,33	(26,59; 78,73)
Hayes	15	13	86,67	(59,54; 98,34)
Holmes	15	15	100	(81,90; 100,00)
Montgomery	15	15	100	(81,90; 100,00)
Simpson	15	14	93,33	(68,05; 99,83)

Fonte: Software estatístico

Tabela 3: Estatística Kappa Fleiss

Avaliador	Resposta	Kappa	SE Kappa	Z	P (v > 0)
Duncan	-2	0,583330	0,258199	2,25924	0,0119
	-1	0,166670	0,258199	0,64550	0,2593
	0	0,440990	0,258199	1,70796	0,0438
	1	0,440990	0,258199	1,70796	0,0438
	2	0,423080	0,258199	1,63857	0,0507
	Geral	0,411760	0,130924	3,14508	0,0008
Hayes	-2	0,629630	0,258199	2,43855	0,0074
	-1	0,813660	0,258199	3,15131	0,0008
	0	1,000000	0,258199	3,87298	0,0001
	1	0,760000	0,258199	2,94347	0,0016
	2	0,813660	0,258199	3,15131	0,0008
	Geral	0,829550	0,130924	6,18307	0,0000
Holmes	-2	1,000000	0,258199	3,87298	0,0001
	-1	1,000000	0,258199	3,87298	0,0001
	0	1,000000	0,258199	3,87298	0,0001
	1	1,000000	0,258199	3,87298	0,0001
	2	1,000000	0,258199	3,87298	0,0001
	Geral	1,000000	0,130924	7,61584	0,0000
Montgomery	-2	1,000000	0,258199	3,87298	0,0001
	-1	1,000000	0,258199	3,87298	0,0001
	0	1,000000	0,258199	3,87298	0,0001
	1	1,000000	0,258199	3,87298	0,0001
	2	1,000000	0,258199	3,87298	0,0001
	Geral	1,000000	0,130924	7,61584	0,0000
Simpson	-2	1,000000	0,258199	3,87298	0,0001
	-1	1,000000	0,258199	3,87298	0,0001
	0	0,813660	0,258199	3,15131	0,0008
	1	0,813660	0,258199	3,15131	0,0008
	2	1,000000	0,258199	3,87298	0,0001
	Geral	0,915970	0,130924	6,99619	0,0000

Fonte: Software estatístico

Os valores do Coeficiente de Correlação de Kendall (Tabela 4) para cada avaliador determinam a consistência destes na análise mesmo que estejam errados.

Tabela 4 - Coeficiente de Correlação de Kendall

Avaliador	Coef	SE Coef	Z	P
Duncan	0,87506	0,19245	4,49744	0,0000
Hayes	0,94871	0,19245	4,88016	0,0000
Holmes	1,00000	0,19245	5,14667	0,0000
Montgomery	1,00000	0,19245	5,14667	0,0000
Simpson	0,96629	0,19245	4,97151	0,0000

Fonte: Software estatístico

5.2 Entre Avaliadores

Os cinco avaliadores tiveram suas qualificações baixas, com 6 acertos em 15 (40%). A Tabela 5 indica a uniformidade entre avaliadores e não sua exatidão nem como suas respostas se comparam aos padrões.

O valor Kappa Geral (0,682965) apresenta uma discordância entre os avaliadores (Tabela 6). Já o coeficiente de concordância de Kendall (0,966317) demonstra que os mesmos aplicam essencialmente o mesmo padrão ao avaliar as amostras (Tabela 7).

Tabela 5 - Avaliação da Concordância

# Inspeccionados	# Acertos	%	95% I.C
15	6	40,00	(16,34; 67,71)

Fonte: Software estatístico

Tabela 6 - Estatística Kappa Fleiss

Resposta	Kappa	SE Kappa	Z	P (v s> 0)
-2	0,680398	0,0816497	8,3331	0,0000
-1	0,602754	0,0816497	7,3822	0,0000
0	0,707602	0,0816497	8,6663	0,0000
1	0,642479	0,0816497	7,8687	0,0000
2	0,736534	0,0816497	9,0207	0,0000
Geral	0,672965	0,0412331	16,3210	0,0000

Fonte: Software estatístico

Tabela 7 - Coeficiente de Concordância de Kendall

Coef	Chi - Sq	DF	P
0,966317	67,6422	14	0,0000

Fonte: Software estatístico

5.3 Todos avaliadores x padrão

Os cinco avaliadores tiveram suas qualificações baixas comparadas aos padrões, também com 6 acertos em 15 (40%)(Tabela 8) indicando a uniformidade entre eles mas não sua exatidão.

O valor Kappa Geral (0,831455) apresenta uma concordância entre as respostas com os valores padrões (Tabela 9). O valor do coeficiente de correlação de Kendall (Tabela 10) apresenta uma consistência na análise dos avaliadores mesmo quando errados.

Tabela 8 - Avaliação da Concordância

# Inspeccionados	# Acertos	%	95% I.C
15	6	40,00	(16,34; 67,71)

Fonte: Software estatístico

Tabela 9 - Estatística Kappa Fleiss

Resposta	Kappa	SE Kappa	Z	P (v s> 0)
-2	0,842593	0,115470	7,2971	0,0000
-1	0,796066	0,115470	6,8941	0,0000
0	0,850932	0,115470	7,3693	0,0000
1	0,850932	0,115470	6,9536	0,0000
2	0,847348	0,115470	7,3383	0,0000
Geral	0,831455	0,058911	14,1136	0,0000

Fonte: Software estatístico

Tabela 10 - Coeficiente de Concordância de Kendall

Coef	SE Coef	Z	P
0,958012	0,0860663	11,1090	0,0000

Fonte: Software estatístico

Analisando os acertos dos avaliadores, dois se destacaram pela porcentagem de acertos de 100%, Holmes e Montgomery. Já Duncan e Hayes apresentaram resultados inconsistentes com o esperado, pois cometeram mais erros durante as medições.

Os cinco avaliadores demonstram alta associação entre as taxas, medições devido ao valor do coeficiente de Kendall de 0,958012. E também existe concordância entre os avaliadores e o padrão, de acordo com o Kappa igual a 0,831455.

O sistema de medição de atributos é efetivo, pois aponta quais avaliadores necessitam de mais treinamento, por exemplo. E mostra quais deles tiveram um ótimo desempenho. Com a

análise estatística destes dados e seus resultados, puderam-se avaliar matematicamente dados de valores subjetivos.

6. Conclusão

Este trabalho apresentou uma ferramenta estatística capaz de transformar dados subjetivos em objetivos, pois esses fazem parte do dia a dia de empresas de serviços, indústrias e até na criação de novos métodos de pesquisa e correção de testes como no exemplo estudado.

O exemplo da empresa que quer contratar 5 novos avaliadores para correção de provas demonstrou quais seriam os profissionais mais ou menos aptos ao trabalho; quais foram os mais coesos nas suas correções com os padrões dados das amostras e coerentes entre si; e a concordância de todos os avaliadores com os padrões. O resultado sugerido foi a demissão ou mais tempo de treinamento para dois deles, Duncan e Hayes.

Para demonstrar que dados subjetivos podem ser analisados matematicamente, o trabalho procurou explorar um exemplo de tomada de decisão estudando estes aspectos. Foi desenvolvido um estudo estatístico envolvendo análise de concordância por atributos, transformando dados subjetivos em objetivos. A análise de concordância por atributos desempenha um papel crucial para quantificar a subjetividade e existe uma gama de possibilidades para o uso da prática explorada nos mais diversos campos do conhecimento, onde decisões subjetivas são tomadas.

Agradecimentos

Os autores agradecem Fapemig, CAPES e CNPq.

Referências

Alvarado, F. P. Análisis de concordancia de atributos. *Tecnologia em Marcha*, v. 21, n. 4, p. 29–35, out-dez, 2008.

Bertrand, J. W. M.; Fransoo, J. C. Operations management research methodologies using quantitative modeling. *International Journal of Operations & Production Management*, v. 22, n. 2, p. 241–264, 2002.

Burn, C.; Weir, A. S. Using prevalence indices to aid interpretation and comparison of agreement ratings between two or more observers. *The Veterinary Journal*, v. 188, n. 2, p. 166–170, 2011.

Chattopadhyay, S.; Acharya, U. R. A novel mathematical approach to diagnose premenstrual syndrome. *Journal of medical systems*, v. 36, n. 4, p. 2177–86, 2012.

Goethals, P. L.; Cho, B. R. The development of a robust design methodology for time-oriented dynamic quality characteristics with a target profile. *Quality and Reliability Engineering International*, v. 27, n. 4, p. 403–414, 2011.

Legendre, P. Species associations: the Kendall coefficient of concordance revisited. *Journal of Agricultural, Biological and Environmental Statistics*, v. 10, n. 2, p. 226–245, 2005.

Rungtusanatham, M. Beyond improved quality: the motivational effects of statistical process control. *Journal of Operations Management*, v. 19, n. 6, p. 653–673, 2001.

Saleh, B.; Hacker, M.; Randhawa, S. Factors in capital decisions involving advanced manufacturing technologies. *International Journal of Operations & Production Management*, v. 21, n. 10, p. 1265–1288, 2001.

Sharp, J. P. Subjectivity. *International Encyclopedia of Human Geography*. Elsevier. 2009, p. 72–76.

Technical Support Document: Attribute Agreement Analysis, 1–2. Disponível em: <http://www.minitab.com/support/answers/answer.aspx?ID=2158>. Acesso em: 05 de julho de 2013.

Technical Support Document: Calculating Kendall's Correlation Coefficient for Each Appraiser vs Standard. Disponível em: <<http://www.minitab.com/en-BR/support/documentation/Answers/KendallsCorrCoeffEachAppraiser.pdf>>. Acesso em: 28 de setembro de 2013.

Technical Support Document: Methods of Calculating Kappa Coefficients. Disponível em: <<http://www.minitab.com/en-BR/support/documentation/Answers/KappaDesignDoc.pdf>>. Acesso em: 28 de setembro de 2013.

Technical Support Document: Methods of Calculating Kendall's Coefficients of Concordance for Ordinal Data. Disponível em: <<http://www.minitab.com/en-BR/support/documentation/answers/KendallDesignDoc.pdf>>. Acesso em: 28 de setembro de 2013.

Uchimura, K. Y.; Bosi, M. L. M. Qualidade e subjetividade na avaliação de programas e serviços em saúde. Cad. Saúde Pública, Rio de Janeiro, v. 18, p. 6, p. 1561–1569, nov-dez, 2002.

Watson, P. F.; Petrie, A. Method agreement analysis: a review of correct methodology. Theriogenology, v. 73, n. 9, p. 1167–79, 2010.

Yang, J.; Chinchilli, V. Fixed-effects modeling of Cohen's weighted kappa for bivariate multinomial data. Computational Statistics and Data Analysis, v. 55, n. 2, p. 1061-70, 2011.