

Aula 8 –Planejamento e Análise de Experimentos

Professores

Miguel Antonio Sovierzoski, Dr.

miguelaso@utfpr.edu.br;

Vicente Machado Neto, Dr.

vmachado@utfpr.edu.br;

Testes baseados em atributos

A maioria das medições industriais usam escalas, micrometros e outros dispositivos que fornecem diretamente a medida física de um produto. Contudo, as vezes as medidas são subjetivas, classificações e avaliações, por exemplo: a) Classificação da fábrica pelo tipo de defeito; b) Classificação da solda como boa ou ruim; c) Ranking da qualidade do vinho na escala de 1 a 10; d) Ranking das bolhas de ar no vidro na escala de 1 a 5;

Usa-se Análise de atributos para avaliar a consistência e a correção em rankings subjetivos.

- Avaliações da taxa de acerto do avaliador em relação a um padrão;
- Avaliação da concordância de respostas entre os avaliadores;
- Avaliação da concordância das avaliações de um mesmo avaliador.

Testes baseados em atributos

Nome	Descrição	Exemplo				
Dicotômicas	Questões com duas respostas possíveis.	"Eu gostei do curso de Estatística Aplicada". Sim [] Não []				
Múltipla Escolha	Questões com três ou mais respostas possíveis.	"Com quem você pretende fazer este curso?" Ninguém [] Colegas de trabalho [] Familiares / amigos []				
Escala Likert	Afirmção na qual o entrevistado mostra o quanto concorda ou discorda.	"Este curso ajudou a aprimorar meus conhecimentos de estatística e me tornou mais apto a utilizar ferramentas para a tomada de decisão."				
		Discordo totalmente []	Discordo parcialmente []	Nem concordo, nem discordo []	Concordo parcialmente []	Concordo totalmente []
Diferencial Semântico	Escala inserida entre duas palavras antônimas e o entrevistado seleciona o ponto que representa a direção e a intensidade de seus sentimentos.	O professor, em sua opinião, é: Ágil <u> x </u> : _____ : _____ : _____ : _____ : _____ : Moroso Experiente _____ : _____ : _____ : _____ : <u> x </u> : _____ : Inexperiente Cativante _____ : <u> x </u> : _____ : _____ : _____ : _____ : Não Cativante				
Escala de Importância	Escala que avalia a importância de alguns atributos, variando de "nada importante" a "extremamente importante".	"Eu considero o conteúdo ministrado no curso...":				
		Extremamente importante []	Muito importante []	Razoavelmente importante []	Não muito importante []	Nada importante []
Escala de Gradação	Escala que avalia certos atributos, de "fracos" a "excelentes".	"A variedade de alimentos no <i>coffe break</i> está":				
		Excelente []	Muito boa []	Boa []	Razoável []	Fraca []

Testes baseados em atributos

Uma das preocupações em testes baseados em atributos é a consistência dos avaliadores. Por exemplo um mesmo avaliador pode avaliar um vinho com alta pontuação e em um segundo momento, avaliar o mesmo vinho com baixa pontuação.

Caso não se confie na consistência dos avaliadores, não se pode confiar nos dados que eles reportam.

Uma análise de atributos pode ajudar a decidir se os avaliadores são consistentes nas suas avaliações.

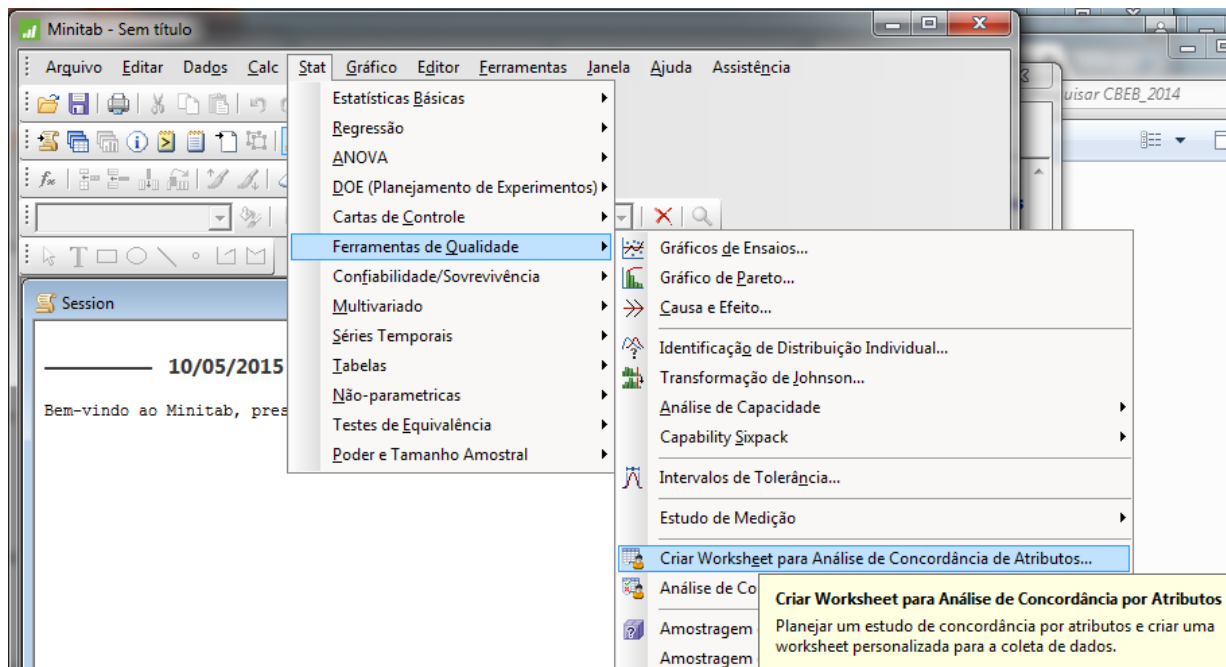
Análises também podem ser feitas levando-se em conta padrões estabelecidos e as avaliações dos avaliadores.

Testes baseados em atributos

Tipos de dados

As variáveis devem ser colocadas em colunas de dados numéricos ou texto. Os dados podem ser tipo binário (good/bad), nominal (blue/yellow/red) ou ordinal (ruim/razoável/bom). Deve-se distinguir as respostas de acordo com o avaliador e rodada do experimento.

Para se criar uma planilha de atributos no Minitab, use em ferramentas de qualidade, criar uma folha de dados para análise de concordância por atributos.



Testes baseados em atributos

Exercício

Os avaliadores de uma empresa de impressão em tecidos fazem avaliações em uma escala de 1 a 5. O engenheiro de qualidade quer saber a consistência e correção dos avaliadores. O engenheiro pede para 4 avaliadores classificarem 50 amostras de impressão em tecido por duas vezes, em ordem aleatória. As amostras já possuem uma classificação prévia aceita como padrão.

Abrir arquivo excel Dados_TextilePrinting para copiar os dados para o Minitab.

Avaliadores: Mike, Amanda, Eric e Britt

Amostras: 1 até 50

Rodadas: 1 e 2

Padrão

Testes baseados em atributos

Exercício

Os avaliadores avaliaram 50 amostras, cada uma duas vezes. Tem-se um padrão de referência, para comparar os resultados.

The image shows a screenshot of the Minitab software interface. The main window displays the 'Stat' menu with 'Ferramentas de Qualidade' selected, and 'Análise de Concordância de Atributos...' highlighted. A dialog box titled 'Análise de Concordância por Atributos' is open, showing the following settings:

- Dados organizados como:**
 - Coluna de atributo: Response
 - Amostras: Sample
 - Avaliadores: Appraiser
- Múltiplas colunas:
- Atributo/padrão conhecido:** Standard (Opcional)
- As categorias dos dados de atributo estão ordenadas

The dialog box also includes buttons for 'Selecionar', 'Ajuda', 'Informações...', 'Opções...', 'Gráficos...', and 'Resultados...'. The background shows the Minitab menu structure and a session window with the date 10/05/2015.

Testes baseados em atributos

Exercício

Os avaliadores avaliaram 50 amostras, cada uma duas vezes. Tem-se um padrão de referência, para comparar os resultados.

The image displays five overlapping dialog boxes from a software application, likely for statistical analysis of attribute-based concordance. The windows are:

- Análise de Concordância por Atributos:** The main configuration window. It features a list of categories (C1 Appraiser, C2 Response, C3 Sample, C4 Trial, C5 Standard) on the left. The main area is divided into sections for 'Dados organizados como' (with 'Coluna de atributo' set to 'Responset'), 'Amostras' (set to 'Sample'), and 'Avaliadores' (set to 'Appraiser'). There are also fields for 'Número de avaliadores', 'Número de ensaios', and 'Atributo/padrão conhecido' (set to 'Standard').
- Análise de Concordância por Atributos: Opções:** A sub-dialog for options. It has checkboxes for 'Calcular kappa de Cohen se apropriado' and 'Exibir tabela de discordância', both of which are checked. The 'Nível de confiança' is set to 95,0.
- Análise de Concordância por Atributos: Gráficos:** A sub-dialog for graphics. It has checkboxes for 'Distribuir a porcentagem de concordância da avaliação entre' (checked), 'Ensaio por avaliador (não exibe gráfico para ensaio único)' (checked), 'Ensaio por avaliador e padrão/atributo conhecido' (checked), and 'Exibir figuras em um gráfico' (checked).
- Análise de Concordância por Atributos: Informação:** A sub-dialog for information. It contains text input fields for 'Data do estudo' (13 de maio de 2015), 'Reportado por' (Vicente), 'Nome do produto' (Impressão em Tecido), and 'Diversos'.
- Análise de Concordância por Atributos: Resultados:** A sub-dialog for results. It has radio buttons for 'Controlar a Exibição dos Resultados', with the selected option being 'Além disso, coeficientes de Kendall (dados ordinais) e kappa'.

Testes baseados em atributos

Concordância entre as respostas de um mesmo avaliador

Caso os avaliadores tenham duas ou mais avaliações do mesmo item, pode-se avaliar a consistência do avaliador. Deve-se observar que aqui não se faz uma comparação com o padrão, assim mesmo que as avaliações sejam consistentes, elas não são necessariamente corretas.

Analisando a consistência dos resultados dos quatro avaliadores.

Análise de Concordância por Atributos para Response

Data do estudo: 13 de maio de 2015
Informado por: Vicente
Nome do produto: Impressão em Tecido
Div:

Dentro dos Avaliadores

Concordância de Avaliação

Avaliador	Nº de Inspeccionados	Nº de Correspondências	Percentual	IC de 95%
Amanda	50	50	100,00	(94,18; 100,00)
Britt	50	48	96,00	(86,29; 99,51)
Eric	50	43	86,00	(73,26; 94,18)
Mike	50	45	90,00	(78,19; 96,67)

Nº de Concordâncias: O avaliador concorda com os ensaios.

No exercício houve coincidência de :

Amanda 100% IC(94,18; 100,00)
Britt 96% IC(86,29; 99,51)
Eric 86% IC(73,26; 94,18)
Mike 90% IC(78,19; 96,67)

Testes baseados em atributos

Concordância entre as respostas de um mesmo avaliador

Utilizando o coeficiente kappa para comparar os resultados dos avaliadores, tem-se que $\text{kappa}=1$ caso haja perfeita concordância dos resultados, caso $\text{kappa} = 0$ a concordância é aleatória. Valores negativos é uma concordância menor do que a aleatória. Valores de kappa menores do que 0,7, indicam que o sistema de mensuração precisa melhorias. Valores de kappa maiores que 0,9 são considerados excelentes.

Estatísticas de Kappa de Fleiss

Avaliador	Resposta	Kappa	EP de Kappa	Z	P(vs > 0)
Amanda	1	1,00000	0,141421	7,0711	0,0000
	2	1,00000	0,141421	7,0711	0,0000
	3	1,00000	0,141421	7,0711	0,0000
	4	1,00000	0,141421	7,0711	0,0000
	5	1,00000	0,141421	7,0711	0,0000
	Global		1,00000	0,071052	14,0741
Britt	1	1,00000	0,141421	7,0711	0,0000
	2	0,89605	0,141421	6,3360	0,0000
	3	0,86450	0,141421	6,1129	0,0000
	4	1,00000	0,141421	7,0711	0,0000
	5	1,00000	0,141421	7,0711	0,0000
	Global		0,94965	0,071401	13,3002
Eric	1	0,83060	0,141421	5,8733	0,0000
	2	0,84000	0,141421	5,9397	0,0000
	3	0,70238	0,141421	4,9666	0,0000
	4	0,70238	0,141421	4,9666	0,0000
	5	1,00000	0,141421	7,0711	0,0000
	Global		0,82354	0,071591	11,5034
Mike	1	1,00000	0,141421	7,0711	0,0000
	2	0,83060	0,141421	5,8733	0,0000
	3	0,81917	0,141421	5,7924	0,0000
	4	0,86450	0,141421	6,1129	0,0000
	5	0,86450	0,141421	6,1129	0,0000
	Global		0,87472	0,070945	12,3295

Testes baseados em atributos

Concordância entre as respostas de um mesmo avaliador

Tem-se a comparação estatística conforme a estatística kappa para cada resposta e geral, assim pode-se responder à questão: “Os avaliadores tem dificuldades com uma resposta em particular?”

Um teste de hipóteses é feito com as hipóteses:

H0: A concordância do avaliador é igual a aleatória.

H1: A concordância do avaliador não é igual a aleatória.

Estadísticas de Kappa de Fleiss

Avaliador	Resposta	Kappa	EP de Kappa	Z	P(vs > 0)
Amanda	1	1,00000	0,141421	7,0711	0,0000
	2	1,00000	0,141421	7,0711	0,0000
	3	1,00000	0,141421	7,0711	0,0000
	4	1,00000	0,141421	7,0711	0,0000
	5	1,00000	0,141421	7,0711	0,0000
	Global		1,00000	0,071052	14,0741
Britt	1	1,00000	0,141421	7,0711	0,0000
	2	0,89605	0,141421	6,3360	0,0000
	3	0,86450	0,141421	6,1129	0,0000
	4	1,00000	0,141421	7,0711	0,0000
	5	1,00000	0,141421	7,0711	0,0000
	Global		0,94965	0,071401	13,3002
Eric	1	0,83060	0,141421	5,8733	0,0000
	2	0,84000	0,141421	5,9397	0,0000
	3	0,70238	0,141421	4,9666	0,0000
	4	0,70238	0,141421	4,9666	0,0000
	5	1,00000	0,141421	7,0711	0,0000
	Global		0,82354	0,071591	11,5034
Mike	1	1,00000	0,141421	7,0711	0,0000
	2	0,83060	0,141421	5,8733	0,0000
	3	0,81917	0,141421	5,7924	0,0000
	4	0,86450	0,141421	6,1129	0,0000
	5	0,86450	0,141421	6,1129	0,0000
	Global		0,87472	0,070945	12,3295

Testes baseados em atributos

Concordância entre as respostas de um mesmo avaliador

Analisando os resultados abaixo pode-se observar que os avaliadores apresentam concordância entre as suas avaliações de uma maneira global P Valor = 0,000 , descartando a hipótese de aleatoriedade.

Estadísticas de Kappa de Fleiss

Avaliador	Resposta	Kappa	EP de Kappa	Z	P(vs > 0)
Amanda	1	1,00000	0,141421	7,0711	0,0000
	2	1,00000	0,141421	7,0711	0,0000
	3	1,00000	0,141421	7,0711	0,0000
	4	1,00000	0,141421	7,0711	0,0000
	5	1,00000	0,141421	7,0711	0,0000
	Global	1,00000	0,071052	14,0741	0,0000
Britt	1	1,00000	0,141421	7,0711	0,0000
	2	0,89605	0,141421	6,3360	0,0000
	3	0,86450	0,141421	6,1129	0,0000
	4	1,00000	0,141421	7,0711	0,0000
	5	1,00000	0,141421	7,0711	0,0000
	Global	0,94965	0,071401	13,3002	0,0000
Eric	1	0,83060	0,141421	5,8733	0,0000
	2	0,84000	0,141421	5,9397	0,0000
	3	0,70238	0,141421	4,9666	0,0000
	4	0,70238	0,141421	4,9666	0,0000
	5	1,00000	0,141421	7,0711	0,0000
	Global	0,82354	0,071591	11,5034	0,0000
Mike	1	1,00000	0,141421	7,0711	0,0000
	2	0,83060	0,141421	5,8733	0,0000
	3	0,81917	0,141421	5,7924	0,0000
	4	0,86450	0,141421	6,1129	0,0000
	5	0,86450	0,141421	6,1129	0,0000
	Global	0,87472	0,070945	12,3295	0,0000

Testes baseados em atributos

Concordância entre as respostas de um mesmo avaliador

Analisando os resultados pelas estatísticas de Kappa de Cohen abaixo pode-se observar que os avaliadores Jim e Vanessa apresentam resultados concordantes. A média geral dos quatro avaliadores tem o p valor é de 0,0000 valor este que permite descartar a hipótese de aleatoriedade dos resultados obtidos.

Estatísticas de Kappa de Cohen

Avaliador	Resposta	Kappa	EP de Kappa	Z	P(vs > 0)
Amanda	1	1,00000	0,141421	7,0711	0,0000
	2	1,00000	0,141421	7,0711	0,0000
	3	1,00000	0,141421	7,0711	0,0000
	4	1,00000	0,141421	7,0711	0,0000
	5	1,00000	0,141421	7,0711	0,0000
	Global	1,00000	0,071052	14,0741	0,0000
Britt	1	1,00000	0,141421	7,0711	0,0000
	2	0,89627	0,140658	6,3719	0,0000
	3	0,86486	0,140124	6,1721	0,0000
	4	1,00000	0,141421	7,0711	0,0000
	5	1,00000	0,141421	7,0711	0,0000
	Global	0,94970	0,071157	13,3465	0,0000
Eric	1	0,83070	0,141196	5,8833	0,0000
	2	0,84009	0,141220	5,9488	0,0000
	3	0,70588	0,135166	5,2223	0,0000
	4	0,70588	0,135166	5,2223	0,0000
	5	1,00000	0,141421	7,0711	0,0000
	Global	0,82430	0,070470	11,6972	0,0000
Mike	1	1,00000	0,141421	7,0711	0,0000
	2	0,83070	0,141196	5,8833	0,0000
	3	0,81928	0,141165	5,8037	0,0000
	4	0,86450	0,141421	6,1129	0,0000
	5	0,86450	0,141421	6,1129	0,0000
	Global	0,87475	0,070885	12,3404	0,0000

Testes baseados em atributos

Concordância entre as respostas de um mesmo avaliador

Há concordância dos resultados obtidos pela estatística de Kappa Fleiss e Kappa Cohen.

Estadísticas de Kappa de Fleiss

Avaliador	Resposta	Kappa	EP de Kappa	Z	P(vs > 0)
Amanda	1	1,00000	0,141421	7,0711	0,0000
	2	1,00000	0,141421	7,0711	0,0000
	3	1,00000	0,141421	7,0711	0,0000
	4	1,00000	0,141421	7,0711	0,0000
	5	1,00000	0,141421	7,0711	0,0000
	Global	1,00000	0,071052	14,0741	0,0000
Britt	1	1,00000	0,141421	7,0711	0,0000
	2	0,89605	0,141421	6,3360	0,0000
	3	0,86450	0,141421	6,1129	0,0000
	4	1,00000	0,141421	7,0711	0,0000
	5	1,00000	0,141421	7,0711	0,0000
	Global	0,94965	0,071401	13,3002	0,0000
Eric	1	0,83060	0,141421	5,8733	0,0000
	2	0,84000	0,141421	5,9397	0,0000
	3	0,70238	0,141421	4,9666	0,0000
	4	0,70238	0,141421	4,9666	0,0000
	5	1,00000	0,141421	7,0711	0,0000
	Global	0,82354	0,071591	11,5034	0,0000
Mike	1	1,00000	0,141421	7,0711	0,0000
	2	0,83060	0,141421	5,8733	0,0000
	3	0,81917	0,141421	5,7924	0,0000
	4	0,86450	0,141421	6,1129	0,0000
	5	0,86450	0,141421	6,1129	0,0000
	Global	0,87472	0,070945	12,3295	0,0000

Estadísticas de Kappa de Cohen

Avaliador	Resposta	Kappa	EP de Kappa	Z	P(vs > 0)
Amanda	1	1,00000	0,141421	7,0711	0,0000
	2	1,00000	0,141421	7,0711	0,0000
	3	1,00000	0,141421	7,0711	0,0000
	4	1,00000	0,141421	7,0711	0,0000
	5	1,00000	0,141421	7,0711	0,0000
	Global	1,00000	0,071052	14,0741	0,0000
Britt	1	1,00000	0,141421	7,0711	0,0000
	2	0,89627	0,140658	6,3719	0,0000
	3	0,86486	0,140124	6,1721	0,0000
	4	1,00000	0,141421	7,0711	0,0000
	5	1,00000	0,141421	7,0711	0,0000
	Global	0,94970	0,071157	13,3465	0,0000
Eric	1	0,83070	0,141196	5,8833	0,0000
	2	0,84009	0,141220	5,9488	0,0000
	3	0,70588	0,135166	5,2223	0,0000
	4	0,70588	0,135166	5,2223	0,0000
	5	1,00000	0,141421	7,0711	0,0000
	Global	0,82430	0,070470	11,6972	0,0000
Mike	1	1,00000	0,141421	7,0711	0,0000
	2	0,83070	0,141196	5,8833	0,0000
	3	0,81928	0,141165	5,8037	0,0000
	4	0,86450	0,141421	6,1129	0,0000
	5	0,86450	0,141421	6,1129	0,0000
	Global	0,87475	0,070885	12,3404	0,0000

Testes baseados em atributos

Concordância entre as respostas de um mesmo avaliador

Se você tem duas ou mais avaliações por avaliador, é possível avaliar a consistência de cada avaliador nas suas avaliações.

Coeficiente de concordância de Kendall expressa o grau de concordância entre as várias avaliações feitas por um avaliador. Ele usa informações sobre as classificações relativas e é sensível à gravidade do erro de classificação.

O coeficiente de concordância de Kendall é possível de ser obtido quando o avaliador faz duas ou mais avaliações da mesma unidade. Com 3 ou mais níveis o coeficiente Kendall pode ser calculado.

O coeficiente de concordância de Kendall pode variar de 0 a 1, quanto maior o coeficiente de Kendall mais forte é a concordância.

O p valor testa as hipóteses:

H_0 : Não há concordância entre as avaliações de um mesmo avaliador;

H_1 : As avaliações de um mesmo avaliador são concordantes entre si.

Testes baseados em atributos

Concordância entre as respostas de um mesmo avaliador

Se você tem duas ou mais avaliações por avaliador, é possível avaliar a consistência de cada avaliador nas suas avaliações.

Coeficiente de concordância de Kendall

O p valor testa as hipóteses:

H_0 : Não há concordância entre as avaliações de um mesmo avaliador;

H_1 : As avaliações de um mesmo avaliador são concordantes entre si.

Coeficiente de Concordância de Kendall

Avaliador	Coef	Qui-Quadrado	GL	P
Amanda	1,00000	98,0000	49	0,0000
Britt	0,99448	97,4587	49	0,0000
Eric	0,98446	96,4769	49	0,0001
Mike	0,98700	96,7256	49	0,0001

Testes baseados em atributos

Cada avaliador versus Padrão

Observa-se que embora as avaliações dos avaliadores possam ser concordantes, elas podem não ser corretas, já que não se faz um comparação com o padrão estabelecido.

Caso tenha-se um padrão pode-se avaliar a assertividade de cada avaliador.

Cada Avaliador versus Padrão

Concordância de Avaliação

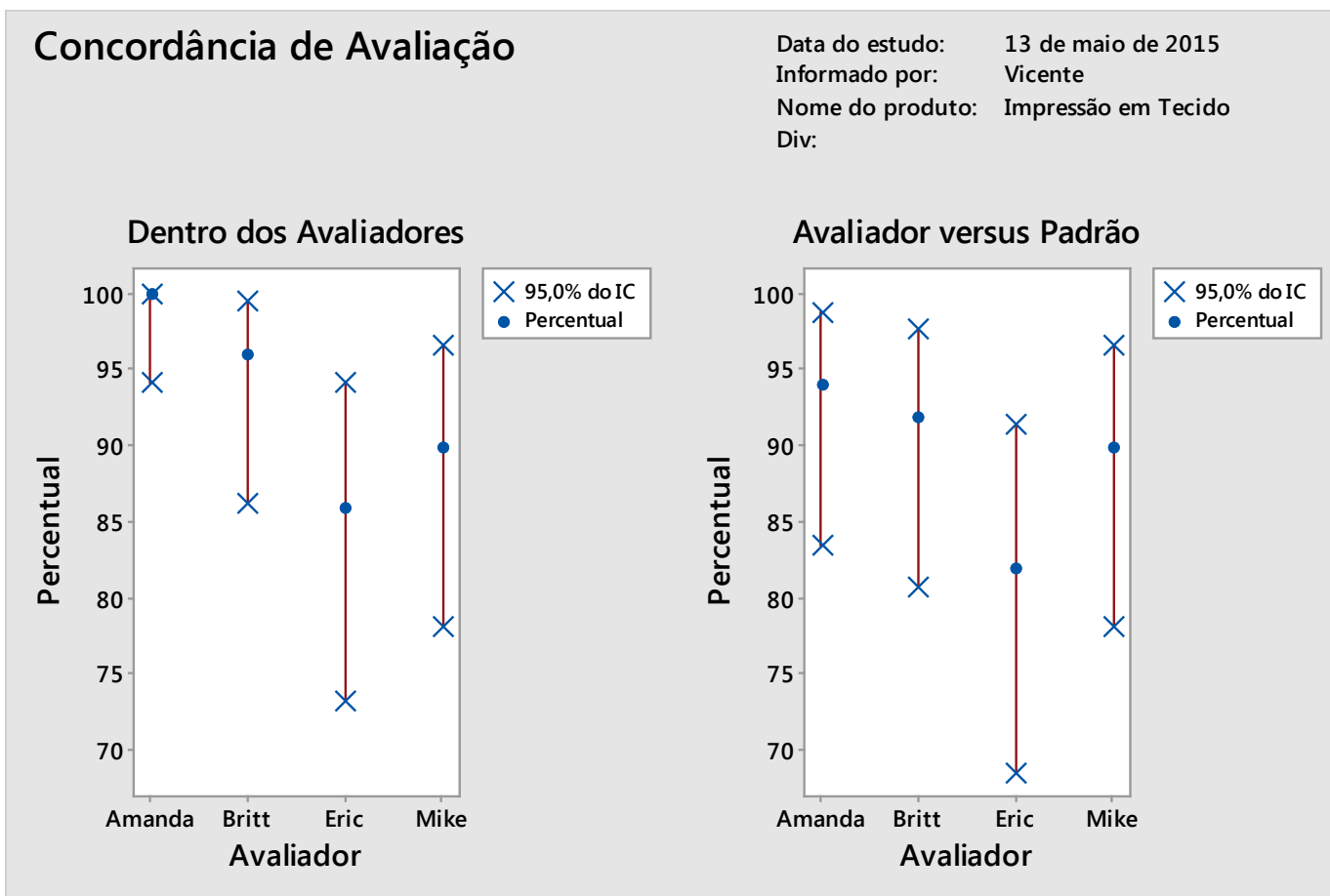
Avaliador	Nº de Inspeccionados	Nº de Correspondências	Percentual	IC de 95%
Amanda	50	47	94,00	(83,45; 98,75)
Britt	50	46	92,00	(80,77; 97,78)
Eric	50	41	82,00	(68,56; 91,42)
Mike	50	45	90,00	(78,19; 96,67)

Nº de Correspondências: A avaliação do avaliador ao longo dos ensaios concorda com o padrão conhecido.

Testes baseados em atributos

Gráficos

Concordância entre avaliações de um mesmo avaliador
Cada avaliador versus Padrão

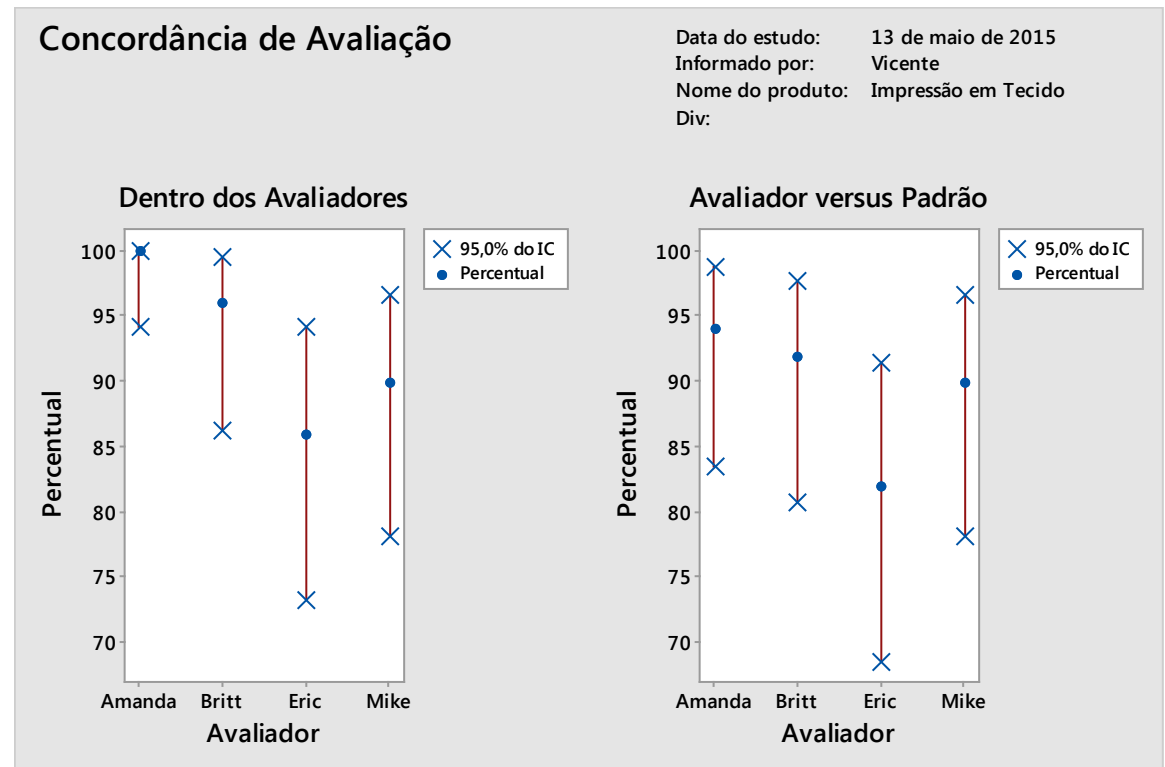


Testes baseados em atributos

Gráficos

Concordância entre avaliações de um mesmo avaliador Cada avaliador versus Padrão

O círculo azul indica o percentual de acerto atual;
A linha vermelha mostra o intervalo de confiança do acerto;
E os Xs indicam os limites inferior e superior do intervalo de confiança.



Testes baseados em atributos

Cada avaliador versus Padrão

Utilizando a estatística de Fleiss'Kappa para avaliar a assertividade dos avaliadores, teremos $\kappa = 1$ caso haja uma perfeita concordância, se $\kappa = 0$ a concordância é a mesma esperada para a aleatoriedade. Kappa menores que 0,7 indicam que o sistema de medição precisa de melhorias e Kappa acima de 0,9 indicam que o sistema é excelente.

O p valor testa as hipóteses:

H_0 : A concordância entre os avaliadores e o padrão é aleatória;

H_1 : A concordância entre os avaliadores e o padrão não é aleatória.

Testes baseados em atributos

Cada avaliador versus Padrão

Pelos valores de Kappa pode-se observar para quais respostas cada um dos avaliadores apresentam melhores e piores concordância com o padrão. Kappa igual a 1 significa perfeita concordância com o padrão.

Estadísticas de Kappa de Fleiss

Avaliador	Resposta	Kappa	EP de Kappa	Z	P(vs > 0)
Amanda	1	1,00000	0,100000	10,0000	0,0000
	2	0,83060	0,100000	8,3060	0,0000
	3	0,81917	0,100000	8,1917	0,0000
	4	1,00000	0,100000	10,0000	0,0000
	5	1,00000	0,100000	10,0000	0,0000
	Global	0,92476	0,050257	18,4006	0,0000
Britt	1	1,00000	0,100000	10,0000	0,0000
	2	0,83838	0,100000	8,3838	0,0000
	3	0,80725	0,100000	8,0725	0,0000
	4	1,00000	0,100000	10,0000	0,0000
	5	1,00000	0,100000	10,0000	0,0000
	Global	0,92462	0,050396	18,3473	0,0000
Eric	1	0,91159	0,100000	9,1159	0,0000
	2	0,81035	0,100000	8,1035	0,0000
	3	0,72619	0,100000	7,2619	0,0000
	4	0,84919	0,100000	8,4919	0,0000
	5	1,00000	0,100000	10,0000	0,0000
	Global	0,86163	0,050500	17,0622	0,0000
Mike	1	1,00000	0,100000	10,0000	0,0000
	2	0,91694	0,100000	9,1694	0,0000
	3	0,90736	0,100000	9,0736	0,0000
	4	0,92913	0,100000	9,2913	0,0000
	5	0,93502	0,100000	9,3502	0,0000
	Global	0,93732	0,050211	18,6674	0,0000

Testes baseados em atributos

Cada avaliador versus Padrão

Para as mesmas análises existe as estatísticas de Kappa de Cohen que tem a mesma interpretação das de Kappa de Fleiss.

Estatísticas de Kappa de Cohen

Avaliador	Resposta	Kappa	EP de Kappa	Z	P(vs > 0)
Amanda	1	1,00000	0,100000	10,0000	0,0000
	2	0,83070	0,099841	8,3203	0,0000
	3	0,81928	0,099818	8,2077	0,0000
	4	1,00000	0,100000	10,0000	0,0000
	5	1,00000	0,100000	10,0000	0,0000
	Global	0,92477	0,050214	18,4167	0,0000
Britt	1	1,00000	0,100000	10,0000	0,0000
	2	0,83848	0,099731	8,4075	0,0000
	3	0,80743	0,099542	8,1114	0,0000
	4	1,00000	0,100000	10,0000	0,0000
	5	1,00000	0,100000	10,0000	0,0000
	Global	0,92465	0,050309	18,3793	0,0000
Eric	1	0,91174	0,099570	9,1567	0,0000
	2	0,81039	0,099929	8,1097	0,0000
	3	0,72794	0,097813	7,4421	0,0000
	4	0,84965	0,098852	8,5952	0,0000
	5	1,00000	0,100000	10,0000	0,0000
	Global	0,86191	0,050156	17,1846	0,0000
Mike	1	1,00000	0,100000	10,0000	0,0000
	2	0,91696	0,099920	9,1769	0,0000
	3	0,90738	0,099909	9,0820	0,0000
	4	0,92918	0,099749	9,3152	0,0000
	5	0,93506	0,099789	9,3704	0,0000
	Global	0,93734	0,050145	18,6926	0,0000

Testes baseados em atributos

Cada avaliador versus Padrão

O coeficiente de correlação médio de Kendall é um coeficiente de correlação entre todas as avaliações de um mesmo avaliador.

Coeficiente de concordância de Kendall expressa o grau de associação entre as várias avaliações feitas por um avaliador. Ele usa informações sobre as classificações relativas e é sensível à gravidade do erro de classificação.

O coeficiente de Kendall pode variar de -1 até 1. Um valor positivo indica uma associação positiva e um valor negativo indica uma associação negativa. Quanto maior o coeficiente maior a associação.

O p valor testa as hipóteses:

H_0 : A concordância das avaliações (padrão x avaliador) são devido a aleatoriedade;

H_1 : A concordância das avaliações (padrão x avaliador) não são devido a aleatoriedade.

Testes baseados em atributos

Cada avaliador versus Padrão

Observa-se para o exemplo que com os p valor são zero (menor que α) assim podemos descartar a hipótese da aleatoriedade de H_0 .

Coeficiente de Correlação de Kendall

Avaliador	Coef	EP de Coef	Z	P
Amanda	0,967386	0,0690066	14,0128	0,0000
Britt	0,967835	0,0690066	14,0193	0,0000
Eric	0,951863	0,0690066	13,7879	0,0000
Mike	0,975168	0,0690066	14,1256	0,0000

O p valor testa as hipóteses:

H_0 : A concordância das avaliações (padrão x avaliador) são devido a aleatoriedade;

H_1 : A concordância das avaliações (padrão x avaliador) não são devido a aleatoriedade.

Testes baseados em atributos

Concordância entre Avaliadores

No exercício pode-se determinar o quanto cada avaliador concorda com o outro.

Entre Avaliadores

Concordância de Avaliação

N° de Inspeccionados	N° de Correspondências	Percentual	IC de 95%
50	37	74,00	(59,66; 85,37)

N° de Correspondências: As avaliações de todos os avaliadores concordam entre si.

Observa-se um índice de 74% de concordância entre os avaliadores $IC_{95\%}$ (59,66%; 85,37%).

Testes baseados em atributos

Concordância entre Avaliadores

No exercício pode-se determinar a concordância entre as avaliações dos avaliadores para as 5 respostas.

Estatísticas de Kappa de Fleiss

Resposta	Kappa	EP de Kappa	Z	P(vs > 0)
1	0,954392	0,0267261	35,7101	0,0000
2	0,827694	0,0267261	30,9695	0,0000
3	0,772541	0,0267261	28,9058	0,0000
4	0,891127	0,0267261	33,3429	0,0000
5	0,968148	0,0267261	36,2248	0,0000
Global	0,881705	0,0134362	65,6218	0,0000

Estatísticas de Kappa de Cohen

Você deve ter dois avaliadores e um único ensaio por avaliador para calcular kappa.

No exemplo observa-se que a resposta 3 é a que apresenta pior concordância.

O p valor testa as hipóteses:

H_0 : A concordância entre os avaliadores é devido a aleatoriedade;

H_1 : A concordância entre os avaliadores não é devido a aleatoriedade.

Testes baseados em atributos

Concordância entre Avaliadores

No exercício pode-se determinar a concordância entre as avaliações dos avaliadores para as 5 respostas. O coeficiente de Kendall avalia a concordância geral.

Coeficiente de Concordância de Kendall

Coef	Qui-Quadrado	GL	P
0,976681	382,859	49	0,0000

O p valor testa as hipóteses:

H_0 : A concordância entre os avaliadores é devido a aleatoriedade;

H_1 : A concordância entre os avaliadores não é devido a aleatoriedade.

Testes baseados em atributos

Concordância geral de todos avaliadores e o padrão

No exercício pode-se avaliar a concordância geral entre os avaliadores e o padrão estabelecido.

Todos os Avaliadores versus Padrão

Concordância de Avaliação

Nº	Nº	Percentual	IC de 95%
de Inspeccionados	de Correspondências		
50	37	74,00	(59,66; 85,37)

Nº de Correspondências: As avaliações de todos os avaliadores concordam com o padrão conhecido.

Testes baseados em atributos

Concordância todos avaliadores e padrão

No exercício pode-se avaliar a concordância geral entre os avaliadores e o padrão estabelecido pela estatística de Kappa de Fleiss.

Estadísticas de Kappa de Fleiss

Resposta	Kappa	EP de Kappa	Z	P (vs > 0)
1	0,977897	0,0500000	19,5579	0,0000
2	0,849068	0,0500000	16,9814	0,0000
3	0,814992	0,0500000	16,2998	0,0000
4	0,944580	0,0500000	18,8916	0,0000
5	0,983756	0,0500000	19,6751	0,0000
Global	0,912082	0,0251705	36,2362	0,0000

O p valor testa as hipóteses:

H_0 : A concordância entre os avaliadores e o padrão é devido a aleatoriedade;

H_1 : A concordância entre os avaliadores e o padrão não é devido a aleatoriedade;

Testes baseados em atributos

Concordância todos avaliadores e padrão

No exercício pode-se avaliar a concordância geral entre os avaliadores e o padrão estabelecido pela estatística de Kappa de Cohen.

Estatísticas de Kappa de Cohen

Resposta	Kappa	EP de Kappa	Z	P (vs > 0)
1	0,977935	0,0499464	19,5797	0,0000
2	0,849134	0,0499276	17,0073	0,0000
3	0,815508	0,0496373	16,4293	0,0000
4	0,944707	0,0498256	18,9603	0,0000
5	0,983766	0,0499736	19,6857	0,0000
Global	0,912167	0,0251030	36,3370	0,0000

O p valor testa as hipóteses:

H_0 : A concordância entre os avaliadores e o padrão é devido a aleatoriedade;

H_1 : A concordância entre os avaliadores e o padrão não é devido a aleatoriedade;

Testes baseados em atributos

Concordância todos avaliadores e padrão

No exercício pode-se avaliar a concordância geral entre os avaliadores e o padrão estabelecido pelo coeficiente de correlação de Kendall.

Coeficiente de Correlação de Kendall

Coef	EP de Coef	Z	P
0,965563	0,0345033	27,9817	0,0000

O p valor testa as hipóteses:

H_0 : A concordância entre os avaliadores e o padrão é devido a aleatoriedade;

H_1 : A concordância entre os avaliadores e o padrão não é devido a aleatoriedade;

Testes baseados em atributos

Concordância todos avaliadores e padrão

No exercício pode-se avaliar o percentual de discordância de cada um dos avaliadores e as amostras com o padrão.

Sumário Desacordo da Avaliação com Padrão

Appraiser		Amanda		Britt		Eric	
Amostra	Padrão	Frequência	Percentual	Frequência	Percentual	Frequência	Percentual
1	5	0	0,00	0	0,00	0	0,00
2	3	0	0,00	0	0,00	1	50,00
3	1	0	0,00	0	0,00	0	0,00
4	2	0	0,00	0	0,00	0	0,00
5	1	0	0,00	0	0,00	0	0,00
6	3	2	100,00	2	100,00	2	100,00
7	4	0	0,00	0	0,00	0	0,00
8	5	0	0,00	0	0,00	0	0,00
9	2	0	0,00	0	0,00	0	0,00
10	1	0	0,00	0	0,00	0	0,00
11	5	0	0,00	0	0,00	0	0,00
12	4	0	0,00	0	0,00	1	50,00
13	1	0	0,00	0	0,00	0	0,00
14	3	0	0,00	0	0,00	0	0,00
15	2	0	0,00	0	0,00	0	0,00
16	4	0	0,00	0	0,00	0	0,00
17	1	0	0,00	0	0,00	0	0,00
18	2	0	0,00	1	50,00	0	0,00
19	3	0	0,00	0	0,00	0	0,00
20	5	0	0,00	0	0,00	0	0,00

Testes baseados em atributos

Concordância todos avaliadores e padrão

No exercício pode-se avaliar o percentual de discordância de cada um dos avaliadores e as amostras com o padrão.

Appraiser	Amostra	Padrão	Amanda		Britt		Eric	
			Frequência	Percentual	Frequência	Percentual	Frequência	Percentual
21	5		0	0,00	0	0,00	0	0,00
22	4		0	0,00	0	0,00	0	0,00
23	3		0	0,00	0	0,00	0	0,00
24	2		0	0,00	0	0,00	0	0,00
25	2		0	0,00	0	0,00	1	50,00
26	4		0	0,00	0	0,00	0	0,00
27	1		0	0,00	0	0,00	0	0,00
28	2		0	0,00	0	0,00	0	0,00
29	5		0	0,00	0	0,00	0	0,00
30	3		0	0,00	0	0,00	0	0,00
31	5		0	0,00	0	0,00	0	0,00
32	2		2	100,00	1	50,00	1	50,00
33	1		0	0,00	0	0,00	0	0,00
34	3		0	0,00	2	100,00	1	50,00
35	1		0	0,00	0	0,00	0	0,00
36	2		2	100,00	0	0,00	0	0,00
37	5		0	0,00	0	0,00	0	0,00
38	4		0	0,00	0	0,00	0	0,00
39	2		0	0,00	0	0,00	0	0,00
40	3		0	0,00	0	0,00	2	100,00
41	1		0	0,00	0	0,00	0	0,00
42	2		0	0,00	0	0,00	0	0,00
43	2		0	0,00	0	0,00	1	50,00
44	3		0	0,00	0	0,00	0	0,00
45	3		0	0,00	0	0,00	0	0,00
46	5		0	0,00	0	0,00	0	0,00
47	4		0	0,00	0	0,00	0	0,00
48	5		0	0,00	0	0,00	0	0,00
49	4		0	0,00	0	0,00	1	50,00
50	1		0	0,00	0	0,00	0	0,00

Testes baseados em atributos

Concordância todos avaliadores e padrão

No exercício pode-se avaliar o percentual de discordância de cada um dos avaliadores e as amostras com o padrão.

Appraiser		Mike		Appraiser		Mike	
Amostra	Padrão	Frequência	Percentual	Amostra	Padrão	Frequência	Percentual
1	5	0	0,00	26	4	0	0,00
2	3	0	0,00	27	1	0	0,00
3	1	0	0,00	28	2	0	0,00
4	2	0	0,00	29	5	0	0,00
5	1	0	0,00	30	3	0	0,00
6	3	1	50,00	31	5	0	0,00
7	4	0	0,00	32	2	0	0,00
8	5	0	0,00	33	1	0	0,00
9	2	0	0,00	34	3	0	0,00
10	1	0	0,00	35	1	0	0,00
11	5	0	0,00	36	2	1	50,00
12	4	0	0,00	37	5	1	50,00
13	1	0	0,00	38	4	0	0,00
14	3	0	0,00	39	2	0	0,00
15	2	0	0,00	40	3	0	0,00
16	4	0	0,00	41	1	0	0,00
17	1	0	0,00	42	2	0	0,00
18	2	0	0,00	43	2	1	50,00
19	3	0	0,00	44	3	0	0,00
20	5	0	0,00	45	3	0	0,00
21	5	1	50,00	46	5	0	0,00
22	4	0	0,00	47	4	0	0,00
23	3	0	0,00	48	5	0	0,00
24	2	0	0,00	49	4	0	0,00
25	2	0	0,00	50	1	0	0,00