

UFRGS

INSTITUTO
DE
MATEMÁTICA

Departamento
de
Estatística

Apostila

Testes
de
Hipóteses
Não
Paramétricos



Prof. Lorí Viali, Dr.

Porto Alegre, 2008



Sumário

Sumário	1
1. Introdução	3
1.1. Generalidades	3
1.2. Alguns motivos para o seu uso	3
1.3. Algumas restrições ao seu uso	3
1.4. A escolha do teste estatístico	3
1.5. Mensuração	4
1.5. A distribuição amostral	5
1.6. Etapas do teste de hipóteses	5
1.7. Tipos de testes não-paramétricos	6
2. Testes para uma amostra	7
2.1. O teste Qui-Quadrado	7
2.1.1. Função	7
2.1.2. Método	7
2.1.3. Pequenas Frequências Esperadas	8
2.1.4. O teste qui-quadrado relacionado com outros testes	8
2.2. O teste K-S (Kolmogorov-Smirnov)	9
2.2.1. Função e fundamentos lógicos	9
2.2.2. Método	9
3. Testes para duas amostras relacionadas	11
3.1. O teste de McNemar para a Significância de Mudanças	11
3.1.1. Função	11
3.1.2. Método e fundamentos lógicos	11
3.1.3. Correção de continuidade	11
3.1.4. Pequenas frequências esperadas	12
3.2. O teste de Wilcoxon	12
3.2.1. Função	12
3.2.2. Fundamentos lógicos e método	13
3.2.3. Empates	13
3.2.4. Pequenas Amostras	13
3.2.5. Grandes Amostras	14
4. Testes para duas amostras independentes	15
4.1. O teste Qui-Quadrado	15
4.1.1. Função	15
4.1.2. Método	15
4.1.3. Tabelas de Contingência 2X2	16
4.1.4. Quando usar o teste	17
4.2. O teste U de Mann-Whitney	17
4.2.1. Função	17
4.2.2. Método	18
4.2.3. Amostras muito pequenas	18
4.2.4. Amostras médias (n entre 9 e 20)	19
4.2.5. Grande amostras (n > 20)	20
4.2.6. Empates	20



4.3. O teste de Kolmogorov-Smirnov	21
4.3.1. Função e fundamentos lógicos	21
4.3.2. Método	21
4.3.3. Pequenas amostras	21
4.3.4. Grandes amostras: prova bilateral	22
4.3.5. Grandes amostras: prova unilateral	23
5. Testes para k amostras relacionadas	25
O teste de Friedman (Análise de variância de dupla classificação por postos)	25
5.1. Função	25
5.2. Fundamentos lógicos do método	25
6. Testes para k amostras independentes	28
6.1. O teste qui-quadrado	28
6.1.1. Função	28
6.1.2. Método	28
6.1.3. Quando usar a prova do qui-quadrado	29
6.2. O teste de Kruskal-Wallis (Análise de variância de uma classificação por postos)	29
6.2.1. Função	29
6.2.2. Método	29
6.2.3. Empates	31
7. Medidas de correlação e seus testes de significância	32
7.1. O coeficiente de contingência: C	32
7.1.1. Função	32
7.1.2. Método	32
7.1.3. A prova de significância do coeficiente de contingência	33
7.1.4. Limitações do coeficiente de contingência	33
7.2. O coeficiente V de Cramer	34
7.3. O coeficiente de correlação de Postos de Spearman: r_s	35
7.3.1. Função	35
7.3.2. Fundamentos lógicos	35
7.3.3. Observações empatadas	37
7.3.4. Teste de significância para o coeficiente de correlação de Spearman	37
7.4. O coeficiente de concordância de Kendall: W	38
7.4.1. Função	38
7.4.2. Fundamentos lógicos	38
7.4.3. Método	39
7.4.4. Empates	39
7.4.4. Teste de significância para W	40
7.4.5. Interpretação de W	41
7.5. Conclusão	41
8. BIBLIOGRAFIA	42



1. Introdução

1.1. Generalidades

Um dos principais assuntos da Estatística moderna é a *inferência estatística*. A inferência estatística é dividida em dois grandes tópicos: a estimação de parâmetros de uma população e os testes de hipóteses.

No desenvolvimento dos métodos da estatística moderna, as primeiras técnicas de inferência que apareceram foram as que faziam diversas hipóteses sobre a natureza da população da qual se extraíam os dados. Como os valores relacionados com a população são denominados “parâmetros”, tais técnicas estatísticas foram denominadas de paramétricas.

A Estatística Não-Paramétrica é tão recente, que o aparecimento dos primeiros testes, neste área, datam do início do século. O seu maior crescimento ocorreu nos últimos 40 anos. Um teste não-paramétrico é aquele cujo modelo não especifica condições sobre os parâmetros da população da qual a amostra foi obtida. Mesmo quando existem certas pressuposições, estas são mais brandas do que aquelas associadas aos testes paramétricos.

1.2. Alguns motivos para o seu uso

O uso freqüente dos testes não-paramétricos dará ao pesquisador outras vantagens, além das seguintes:

- São menos exigentes do que os paramétricos. Dispensam, por exemplo, a normalidade dos dados.
- Em geral, as probabilidades das afirmativas obtidas na maioria dos testes não-paramétricos, são exatas, salvo quando se usam aproximações para grandes amostras.
- Independem da forma da população da qual a amostra foi obtida.
- São, em geral, de mais fácil aplicação e exigem, quase sempre, menor volume de cálculos.
- Existem testes não-paramétricos que nos permitem trabalhar com dados de diferentes populações, o que não é possível com os paramétricos.
- São úteis nos casos em que é difícil estabelecer uma escala de valores quantitativos para os dados. o pesquisador pode apenas dizer que um dado tem *mais* ou *menos* da característica que está sendo analisada, sem poder precisar ou quantificar as diferenças. Os dados se encontram numa certa ordem de classificação: *mais* ou *menos*; *melhor* ou *pior*; *maior* ou *menor*; etc.
- São mais eficientes do que os paramétricos, quando os dados da população não têm uma distribuição normal. E quando a população é normalmente distribuída, sua eficiência, em alguns casos, é levemente inferior à dos concorrentes.

1.3. Algumas restrições ao seu uso

- Em geral não levam em consideração a magnitude dos dados. É muito comum transformar os dados, de valores para simples ordem ou sinais. Em muitos casos isso se traduz num desperdício de informações.
- Quando todas as exigências do modelo estatístico estão satisfeitas, o teste paramétrico tem mais poder. Para se obter a mesma eficiência com um teste não-paramétrico é necessário uma amostra maior.
- Em, geral, não permitem testar interações, exceto a aditividade em condições especiais. Isto restringe a sua aplicação aos modelos mais simples.
- A obtenção, utilização e interpretação das tabelas (distribuições de probabilidade) são em geral, mais complexas.

1.4. A escolha do teste estatístico

Existem inúmeros testes estatísticos tanto paramétricos quanto não paramétricos. Alguns itens devem ser levados em conta na escolha da prova estatística para determinada situação. A maneira como a amostra foi obtida, a natureza da população da qual se extraiu a amostra e o tipo de mensuração ou escala empregado



nas definições operacionais das variáveis envolvidas, isto é, o conjunto de valores numéricos e ainda o tamanho da amostra disponível.

Uma vez determinados a natureza da população e o método de amostragem ficará estabelecido o modelo estatístico. Associado a cada teste estatístico tem-se um modelo estatístico e condições de mensuração, o teste é válido sob as condições especificadas no modelo e pelo nível da escala de mensuração. Nem sempre é possível verificar se todas as condições do modelo foram satisfeitas e neste caso tem-se que admitir que estas condições foram satisfeitas. Estas condições do modelo estatístico são denominadas *suposições* ou *hipóteses* do teste. Qualquer decisão tomada através de um teste estatístico somente terá validade se as condições do modelo forem válidas.

É óbvio que quanto mais fracas forem as suposições do modelo mais gerais serão as conclusões. No entanto, as provas mais poderosas, isto é, apresentam maior probabilidade de rejeitar H_0 quando for falsa, são as que exigem as suposições mais fortes ou mais amplas.

1.5. Mensuração

O processo de selecionar um modelo matemático ou estatístico a ser utilizado com uma dada técnica de pesquisa envolve algumas decisões importantes. A escolha do modelo a ser aplicado é precedida pela mensuração do fenômeno envolvido. E a primeira dificuldade surge já na necessidade de definirmos o que é mensuração. Se ela se referir somente aqueles tipos de medidas comumente utilizados em ciências tais como a Física (por exemplo: medidas de comprimento, massa ou tempo) não haverá muitos problemas na escolha do sistema matemático. Agora se o conceito de medida for ampla o suficiente para incluir certos procedimentos de categorização, normalmente utilizados em Ciências Sociais, então o problema torna-se mais complexo. Pode-se distinguir diversos níveis de mensuração e, para cada um, existem diferentes modelos estatísticos apropriados. As operações possíveis em um determinado conjunto numérico dependem do nível de mensuração atingido.

As quatro formas de mensuração ou tipos de medidas são: *nominal*, *ordinal*, *intervalar* e de *razão*.

Nível nominal. Os termos nível nominal de medida ou escala nominal são utilizadas para se referir a àqueles dados que só podem ser categorizados. No sentido estrito, não existe uma medida ou escala envolvida, o que existe é apenas uma contagem. Variáveis que podem ser ditas nominais são: a classificação das pessoas quanto à religião, sexo, estado civil, etc. Não existe uma ordem particular entre as categorias ou grupos e além disso duas categorias quaisquer são mutuamente excludentes, isto é, uma pessoa não pode ser ao mesmo tempo católico e protestante. Além disso as categorias são exaustivas, significando que um determinado elemento deve aparecer em uma e somente uma das categorias. Veja-se um exemplo na tabela 1.1.

Tabela 1.1 - Exemplo de uma variável nominal

Estado civil	Número de pessoas
Casados	340
Solteiros	250
Viúvos	40
Divorciados	50
Total	700

Na classificação os nomes das categorias são atribuídos arbitrariamente, como rótulos de conveniência. Por exemplo, colocam-se católicos e protestantes em categorias diferentes, mas isto não significa que uma é melhor ou maior que a outra. Como as categorias são exaustivas (incluem todos os casos) e mutuamente exclusivas (não há sobreposição, um elemento pertence a uma e somente uma categoria) têm-se as condições mínimas para a aplicação de procedimentos estatísticos. O termo escala nominal é utilizado para indicar o nível mais baixo de mensuração.

Tabela 1.2 - Exemplo de uma variável ordinal

Conceitos	Número de alunos
A	4
B	6
C	14
D	3
E	2
Total	30

As estatísticas possíveis de serem calculadas quando se tem uma escala nominal são: a moda e a contagem de freqüências. Sob certas condições, pode-se comprovar hipóteses utilizando-se o teste χ^2 (qui-quadrado). A medida de associação mais comum para dados nominais é o coeficiente de contingência C.

Nível ordinal. O nível (ou escala) ordinal é o tipo (ou escala) nominal em que se pode ordenar as categorias. A única diferença entre os dois níveis é a relação de ordem que se pode estabelecer entre as categorias. No entanto,



não é possível afirmar o quanto uma categoria é maior do que a anterior, isto é, não se pode afirmar o quanto uma categoria possui da característica. A avaliação do desempenho escolar, através de conceitos, é um exemplo de escala ordinal. No entanto, com este tipo de medida, não se pode afirmar que quem tirou A é ou teve um número de acertos duas vezes maior que quem tirou C. A única coisa que se sabe é que tem A acertou mais questões do que tem B e este de quem tem C e assim por diante.

A estatística mais adequada para a descrição da tendência central dos valores em uma escala ordinal é a mediana, pois ela não é afetada por modificações de quaisquer valores acima ou abaixo dela, desde que o número de observações acima ou abaixo permaneça o mesmo. Numa escala ordinal pode ser utilizado qualquer teste que envolva ordenações ou “postos”. São adequados os coeficientes de correlação baseados em postos, como por exemplo: o coeficiente r_s de Spearman. A tabela 1.2 apresenta um exemplo deste tipo de medida.

Nível intervalar. A escala de medida intervalar é uma escala nominal em que a distância entre as categorias ao contrário da ordinal é sempre a mesma. Ou seja ela possui todas as características da escala ordinal mais o fator de que a distância entre as diversas categorias (ou valores) é sempre constante. As escalas de medir temperaturas como a Fahrenheit e a Centígrada são exemplos de escalas de intervalo. No entanto, não se pode afirmar que uma temperatura de 40 graus é duas vezes mais quente que uma de 20 graus, embora se possa dizer que a diferença entre 20 graus e 40 graus é a mesma que entre 75 graus e 95 graus. Isto porque este tipo de escala não possui um zero absoluto. Ou seja o valor zero na escala é apenas um ponto de referência e não significa a ausência de calor.

A escala de intervalo é a primeira verdadeiramente qualitativa encontrada até agora. Todas as estatísticas paramétricas comuns como: médias, desvios-padrão, correlação de Pearson, etc. são aplicáveis a dados nesta escala, assim como os testes paramétricos comuns como o t e o F .

Nível de razão. Este é o mais alto nível de medida. É caracterizado por apresentar todas as propriedades da escala intervalar mais um zero absoluto. Isto é, aqui o zero pode e deve ser entendido como a ausência da característica e as comparações de valor (razão) tem sentido. Um exemplo de variável deste tipo é o peso. Um valor igual a zero significa ausência de peso e um valor de 20 kg é o duas vezes mais pesado que um de 10 kg. Os valores de uma escala de razão são números verdadeiros e com um zero verdadeiro (absoluto), então qualquer estatística é aplicável a este tipo de escala.

1.5. A distribuição amostral

A distribuição amostral é uma distribuição de probabilidade, isto é, é uma distribuição teórica que descreve o comportamento de uma determinada estatística ou estimador. As principais estatísticas utilizadas nos testes de hipóteses possuem modelos conhecidos. Têm-se a distribuição normal, a distribuição t (de Student) a distribuição χ^2 (qui-quadrado), a distribuição F (de Snedkor) como as principais.

1.6. Etapas do teste de hipóteses

Qualquer teste de hipóteses não-paramétrico segue os seguintes passos:

1. Formular as hipóteses. Estabelecer as hipóteses nula e alternativa. A construção de um teste de hipóteses pode ser colocado de forma geral do seguinte modo. Toma-se uma amostra da variável (ou das variáveis) X (no caso) de uma dada população, de onde se tem uma hipótese sobre um determinado parâmetro, por exemplo: θ . Esta hipótese é a hipótese nula ou hipótese de igualdade:

$$H_0: \theta = \theta_0$$

Tendo formulado a hipótese nula é conveniente determinar qual será a hipótese aceita caso a hipótese nula seja rejeitada, isto é, convém explicitar a hipótese alternativa. A hipótese alternativa vai depender de cada situação mas de forma geral tem-se:

$H_1: \theta = \theta_1$ (hipótese simples), ou então o que é mais comum, hipóteses compostas:

$H_1: \theta > \theta_0$ (teste unilateral ou unicaudal à direita)

$\theta < \theta_0$ (teste unilateral ou unicaudal à esquerda)

$\theta \neq \theta_0$ (teste bilateral ou bicaudal)



2. Estabelecer a estatística (estimador) a ser utilizado. Após fixar as hipóteses é necessário determinar se a diferença entre a estatística amostral e o suposto valor do parâmetro da população é suficiente para rejeitar a hipótese. A estatística utilizada deve ser definida e sua distribuição teórica determinada.

3. Fixar o nível de significância do teste. Fixar a probabilidade de ser cometer erro do tipo I, isto é, estabelecer o nível de significância do teste. Fixado o erro do tipo I, é possível determinar o valor crítico, que é um valor lido na distribuição amostral da estatística considerada (tabela). Este valor vai separar a região de crítica (de rejeição) da região de aceitação.

4. Calcular a estatística teste (a estimativa). Através da amostra obtida calcular a estimativa que servirá para aceitar ou rejeitar a hipótese nula. Dependendo do tipo de hipótese alternativa este valor servirá para aceitar ou rejeitar H_0 .

5. Tomar a decisão. Se o valor da estatística observada na amostra estiver na região crítica rejeitar H_0 , caso contrário aceitar H_0 .

6. Conclusão. Enunciar a tomada da decisão em termos do problema sendo testado.

1.7. Tipos de testes não-paramétricos

Os testes não-paramétricos podem ser divididos em testes para:

- Uma amostra
- Duas amostras emparelhadas (dependentes)
- Duas amostras independentes
- Várias amostras emparelhadas (dependentes)
- Várias amostras independentes

Abaixo segue um resumo dos principais testes estatísticos não-paramétricos classificados de acordo com o nível de medida utilizado e de acordo com o(s) tipo(s) de amostra utilizados.

Tabela 1.3 - Resumo dos testes não-paramétricos

Nível de mensuração	TESTES ESTATÍSTICOS NÃO-PARAMÉTRICOS					Medidas de correlação não-paramétricas
	Caso de uma amostra	Caso de duas Amostras		Caso de k amostras		
		Amostras relacionadas	Amostras independentes	Amostras relacionadas	Amostras independentes	
Nominal	Binomial e χ^2	McNemar	Fisher e χ^2	Q de Cochran	χ^2	De contingência
Ordinal	Kolmogorov-Smirnov Iterações	Sinais Wilcoxon	Mediana U de Mann-Whitney Kolmogorov-Smirnov Iterações de Wald-Wolfowitz Moses	Friedman	Extensão da mediana Kruskal-Wallis	Por postos de Spearman Por postos de Kendall Parcial de postos de Kendall Concordância de Kendall
Intervalar		Walsh Aleatoriedade	Aleatoriedade			

Alguns destes testes serão vistos na disciplina. Para os demais recomenda-se o livro do Siegel citado na bibliografia.



2. Testes para uma amostra

2.1. O teste Qui-Quadrado

2.1.1. Função

A prova χ^2 de uma amostra é aplicada quando o pesquisador está interessado no número de indivíduos, objetos ou respostas que se enquadram em várias categorias que podem ser duas ou mais. Usa-se a técnica do tipo de prova de aderência, ou seja, deve comprovar se existe diferença significativa entre o número **observado** de indivíduos, ou de respostas, em determinada categoria, e o respectivo número **esperado**, baseado na hipótese de nulidade.

2.1.2. Método

O método usado é o da **comparação**, ou seja, comparar um **grupo observado** com um **grupo esperado** de freqüências. Mas antes deve-se determinar as freqüências esperadas. Para isso, usa-se a **hipótese de nulidade**, que dará a proporção de indivíduos, ou objetos, que se enquadram em cada uma das diferentes categorias em que a população está presumidamente classificada. A hipótese de nulidade pode ser testada por:

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}, \text{ onde:}$$

O_i = número de casos observados classificados na categoria i .

E_i = número de casos esperados na categoria i sob H_0 , onde k = número de categorias.

Se há concordância entre os valores observados e os esperados, as diferenças ($O_i - E_i$) serão pequenas e, conseqüentemente, χ^2 será também pequeno. Se as divergências, entretanto, forem grandes, o valor de χ^2 , será também grande. Pode-se mostrar que a distribuição amostral de χ^2 , sob H_0 , calculada pela fórmula acima, segue a distribuição qui-quadrado com um número de graus de liberdade igual a "k-1" onde "k" é igual ao número de categorias em que a variável foi classificada.

Existem muitas distribuições qui-quadrado diferentes, uma para cada grau de liberdade. O grau de liberdade, anotado por gl reflete o *número de observações livres* (que podem variar) após feitas certas restrições sobre os dados. Por exemplo, se forem classificados em duas categorias dados relativos a 50 casos, tão logo se saiba que, digamos, 35 casos se enquadram em uma das categorias, automaticamente fica-se sabendo que 15 casos se enquadrarão na outra. Tem-se, então que $gl = 1$, porque com duas categorias e qualquer n fixo, tão logo se conheça o número de casos em uma categoria a outra estará automaticamente determinada. Em geral, no caso de uma amostra, quando H_0 especifica plenamente os valores esperados o número de graus de liberdade será: $gl = k - 1$, onde k representa o número de categorias usadas na classificação dos dados.

Para empregar a prova χ^2 na comprovação de uma hipótese, deve-se enquadrar cada observação em uma das k células. O número total dessas observações deve ser n (número de elementos da amostra considerada). Isto é, cada observação deve ser independente de qualquer outra. Não se pode, portanto, fazer várias observações sobre o mesmo indivíduo e considerá-las como sendo independentes. Deve-se também determinar a freqüência esperada para cada uma das k células. Se H_0 especificar que a proporção de elementos em cada categoria seja a mesma, então $E_i = n / k$.

Exemplo:

Em corridas de cavalos é ponto de vista comum entre os apostadores que, em uma pista circular, as chances são mais favoráveis a cavalos em determinadas posições (raias). A raia 1 é a mais próxima do lado interno da pista, e a 8 o mais afastada (numa corrida com 8 cavalos). Pode-se comprovar os efeitos das raias, analisando-se os resultados das corridas, dados em função das raias. No exemplo, coletou-se os resultados do primeiro mês da temporada de 1955 (*conforme o New York Post, Ago. 30, 1955, pág. 42*) em uma pista circular.



Tabela 2.2 - Número de vitórias de cavalos e seus respectivos postos.

	Posto								Total
	1	2	3	4	5	6	7	8	
Número de vitórias	29	19	18	25	17	10	15	11	144

1. Hipóteses: H_0 : Não há diferença entre o número esperado de ganhadores em relação a cada posto. H_1 : Existe diferença entre o número de ganhadores de cada posto.

2. Prova Estatística. Como se está comparando os dados de uma população presumida, usa-se uma prova unilateral. Emprega-se a prova χ^2 porque a hipótese em estudo se refere à comparação de frequências observadas e esperadas em categorias discretas. (As categorias são os oito postos).

3. Distribuição amostral. A distribuição amostral de χ^2 tal como calculada, pela expressão dada acima, segue a distribuição qui-quadrado com $gl = k - 1$.

4. Região de Rejeição. H_0 será rejeitada se o valor observado de χ^2 , calculado pela expressão acima, for maior que o valor tabelado, a um nível de significância dado α .

5. Decisão. A amostra de 144 ganhadores forneceu os dados exibidos na tabela 2.2 acima. O cálculo do valor observado do qui-quadrado é dado por:

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} = 16,30$$

A tabela fornece um valor χ^2 igual a 18,475 para $gl = 7$ e um nível de significância de 1%. Neste caso, não é possível rejeitar H_0 , isto é, não é possível afirmar a 1% de significância que o número de vitórias depende do posto.

2.1.3. Pequenas Frequências Esperadas

Quando $gl = 1$, isto é, quando $k = 2$, cada frequência esperada não deve ser inferior a 5. Quando o grau de liberdade for maior do que um, isto é, quando $k > 2$, a prova χ^2 não deve ser usada se mais de 20% das frequências esperadas forem inferiores a 5 ou se qualquer frequência esperada é inferior a 1. As frequências esperadas podem eventualmente ser aumentadas combinando-se categorias adjacentes. Isto naturalmente só deve ser feito se as combinações forem significativas.

Por exemplo, pode-se classificar um grupo de pessoas quanto à sua atitude em relação a determinada opinião em: "apoia fortemente", "apoia", "indiferente", "é contra" e "é fortemente contra". Como forma de aumentar as frequências esperadas as categorias poderiam ser reclassificadas em: "apoia", "indiferente" e "é contra".

2.1.4. O teste qui-quadrado relacionado com outros testes

A tabela 2.3, relaciona o teste χ^2 com outros testes não paramétricos e paramétricos.

Tabela 2.3 - Relacionamento entre testes paramétricos e não-paramétricos

	Teste não-paramétrico		Teste paramétrico
	Dados nominais	Dados ordinais	
Uma amostra	Qui-quadrado de aderência		Teste z de $H_0: P = a$ Teste t de $H_0: \mu = a$
Duas amostras independentes	Qui-quadrado de homogeneidade	Teste da mediana e Teste Mann-Whitney	Teste z de $H_0: P_1 = P_2$ Teste t de $H_0: \mu_1 = \mu_2$
Duas amostras relacionadas	Teste de Mc-Nemar	Teste de Wilcoxon	Teste z de $H_0: P_1 = P_2$ Teste t de $H_0: d = 0$
k amostras	Qui-quadrado de homogeneidade	Teste de Kruskal-Wallis	ANOVA de uma classificação



2.2. O teste K-S (Kolmogorov-Smirnov)

2.2.1. Função e fundamentos lógicos

A prova de Kolmogorov-Smirnov (K-S) é uma prova de aderência. Isto é, avalia o grau de concordância entre a distribuição de um conjunto de valores amostrais (valores observados) e determinada distribuição teórica específica. A prova testa se os valores amostrais podem provavelmente serem considerados como oriundos de uma população com uma suposta distribuição teórica.

A prova utiliza as distribuições acumuladas, isto é, ela compara a distribuição de freqüências acumulada que deveria ocorrer sob a suposta distribuição (sob H_0) com a distribuição de freqüências acumuladas dos valores observados (amostrais). A estatística teste é o ponto de maior diferença (em valor absoluto) entre as duas distribuições.

2.2.2. Método

A distribuição teórica acumulada (sob H_0) é representada por $F_0(x)$ e a distribuição de freqüências dos valores amostrais por $S_n(x)$. Como H_0 supõe que a amostra tenha sido obtida da distribuição $F_0(x)$ é razoável esperar que, para cada valor de x , $S_n(x)$ esteja próximo de $F_0(x)$, isto é, sob H_0 , espera-se que as diferenças entre $S_n(x)$ e $F_0(x)$ sejam pequenas. O teste K-S toma a maior destas diferenças em módulo que é denominada de *desvio máximo* e é anotada por **D**.

Assim:

$$D = |F_0(x) - S_n(x)|$$

A distribuição amostral de **D**, sob H_0 , é conhecida e se encontra tabelada (tabela E, Siegel, pg. 282) em função de "n".

Exemplo:

Suponha que um dado é jogado 150 vezes e que o número obtido de cada face seja anotado e forneçam os resultados apresentados na tabela 2.3. Testar ao nível de 1% de significância a hipótese de que o dado é equilibrado.

Tabela 2.3 - Número de faces na jogada de um dado 150 vezes.

Número de vezes que a face apareceu	Faces					
	1	2	3	4	5	6
	29	19	19	27	26	30

Hipóteses: H_0 : O dado é equilibrado.

H_1 : O dado não é equilibrado.

Prova Estatística. Emprega-se a prova K-S porque o pesquisador deseja comparar uma distribuição observada de escores em escala ordinal com uma distribuição teórica.

Nível de significância. Seja $\alpha = 0,01$.

Distribuição amostral. A tabela E (Siegel, pg. 282) apresenta vários valores críticos de **D** (valores da distribuição amostral) com as respectivas probabilidades de ocorrência sob H_0 .

Região de Rejeição. A região de rejeição consiste de todos os valores de **D** tão grandes que a probabilidade associada à sua ocorrência, sob H_0 , seja menor ou igual a 0,01.

Neste caso, $F_0(x)$ é a distribuição acumulada teórica, sob H_0 , onde H_0 é a hipótese de que cada uma das cinco cópias tenha precisamente $1 / 6 = 16,67\%$ das preferências. $S_{10}(x)$ é a distribuição acumulada das freqüências observadas dos 150 lançamentos realizados. A última linha da tabela 2.4 fornece o valor absoluto do desvio de cada valor amostral em relação ao correspondente valor teórico.



Tabela 2.4 - Cálculos ilustrando a obtenção da estatística de K-S.

Número de vezes	Fases					
	1	2	3	4	5	6
Número de vezes	29	19	19	27	26	30
$F_0(x)$	1/6	2/6	3/6	4/6	5/6	1
$S_{10}(x)$	29/150	48/150	67/150	94/150	120/150	1
$ F_0(x) - S_{10}(x) $	0,0267	0,0133	0,0533	0,0400	0,0333	0

Observando-se a última linha da tabela 2.4, tem-se que $D = 0,053$. Observando-se a tabela E (Siegel, pg. 282), a $\alpha = 1\%$, verifica-se que o valor de D é $1,630/\sqrt{150} = 0,133$. Como o D calculado não é maior que o valor tabelado a conclusão é: aceitar H_0 ao nível de significância de 1% , isto é, não se pode afirmar que o dado é desequilibrado.



3. Testes para duas amostras relacionadas

3.1. O teste de McNemar para a Significância de Mudanças

3.1.1. Função

O teste de McNemar para a significância de mudanças é particularmente aplicável aos experimentos do tipo "antes e depois" em que cada sujeito é utilizado como seu próprio controle e a medida é efetuada em escala nominal ou ordinal.

3.1.2. Método e fundamentos lógicos

Para testar a significância de qualquer mudança observável, através deste método, é necessário construir uma tabela de freqüências "dois por dois" para representar o primeiro e o segundo conjunto de respostas dos mesmos indivíduos. As características gerais de tal tabela encontram-se ilustradas abaixo, onde os valores + e - são utilizados para representar respostas diferentes.

Tabela 3.1 - Tabela 2x2 utilizada para testar a significância de mudanças no teste de McNemar

		Depois	
		-	+
Antes	+	A	B
	-	C	D

Note-se que aqueles casos que mostram mudanças entre a primeira e a segunda resposta aparecem nas células A e D. Um sujeito é contado na célula A se ele muda de + para - e é contado na D se ele muda de - para +. Se nenhuma mudança é ocorre ele é contado nas células A (resposta + antes e depois) e C (resposta - antes e depois).

Como A + D representa o número total de elementos que acusaram alguma modificação, a expectativa, sob a hipótese de nulidade, é de que $1/2 (A + D)$ acuse modificações em um sentido e $1/2 (A + D)$ no outro sentido.

Do teste qui-quadrado tem-se que:

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}, \text{ onde } O_i \text{ é o número observado de casos na categoria "i" e } E_i \text{ é o número esperado}$$

de casos nesta mesma categoria.

Neste teste, as células de interesse são somente a A e a D. Desta forma, se A é o número de casos observados na célula A e D é o número observado de casos na célula D e $(A + D) / 2$ é o número esperado de casos em cada uma das células, então vem:

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} = \frac{(A - \frac{A+D}{2})^2}{\frac{A+D}{2}} + \frac{(D - \frac{A+D}{2})^2}{\frac{A+D}{2}}. \text{ Simplificando, vem:}$$

$$\chi^2 = \frac{(A-D)^2}{A+D} \text{ com grau de liberdade (isto é, linha da tabela) igual a 1.}$$

3.1.3. Correção de continuidade

A aproximação da distribuição acima pela distribuição qui-quadrado torna-se excelente, se for executada uma correção de continuidade. A correção torna-se necessária porque uma distribuição contínua, no caso, o qui-quadrado está sendo usada para aproximar uma distribuição discreta. Quando todas as freqüências esperadas são pequenas, esta aproximação pode não ser boa. A correção de continuidade (de Yates) é uma tentativa de remover esta fonte de erro. A expressão acima incluindo a correção de Yates fica:



$$\chi^2 = \frac{(|A-D|-1)^2}{A+D}$$

Exemplo:

Um psicólogo infantil está interessado em observar a iniciação de contatos sociais em crianças. Ele observou que crianças que são novas em uma escola maternal estabelecem contatos interpessoais com adultos ao invés de com outras crianças. Ele prevê que à medida que se familiarizam com o ambiente as crianças estabelecem contatos interpessoais com outras crianças ao invés de com adultos. Para testar esta hipótese ele observa 25 crianças nos seus primeiros dias em uma escola maternal e então categoriza suas primeiras iniciações de contatos sociais em: se foi dirigido a um adulto ou se foi dirigido a outra criança. Ele, então, observa cada uma das 25 crianças depois de elas estarem na escola por um mês, fazendo a mesma classificação. Os dados estão colocados na tabela 3.2 abaixo.

Tabela 3.2 - Tipo de iniciação social de crianças de uma escola maternal

Objeto de iniciação no primeiro dia	Objeto de iniciação no trigésimo dia	
	Criança	Adulto
Adulto	14	4
Criança	3	4

Hipóteses: **H₀:** Para aquelas crianças que mudam, a probabilidade de que uma criança mude o seu objeto de iniciação de um adulto para criança (isto é, P_A) é igual a probabilidade de que ela mude seu objeto de iniciação de criança para adulto (isto é, P_B) e é igual a 50%, ou seja: P_A = P_B = 1/2.

H₁: P_A > P_B

Prova Estatística. Prova de McNemar para a significância de mudanças porque o estudo utiliza duas amostras relacionadas e utiliza mensuração nominal.

Nível de significância. Sejam $\alpha = 0,05$ e $n = 25$, o número de crianças observadas no primeiro e no trigésimo dia na escola maternal.

Distribuição amostral. Qui-quadrado com 1 grau de liberdade.

Região de Rejeição. Consiste de todos os valores da distribuição χ^2 obtidos dos dados tal que a probabilidade de ocorrência de um valor mais extremo é menor que 0,05.

Decisão. Os dados hipotéticos do exemplo estão mostrados na tabela 3.2 acima. De acordo com eles o valor de qui-quadrado calculado é:

$$\chi^2 = \frac{(|A-D|-1)^2}{A+D} = \frac{(|14-4|-1)^2}{14+4} = 4,50$$

Uma consulta à tabela mostra que o valor da distribuição qui-quadrado com “um” grau de liberdade e com probabilidade de 5% é 3,84. Como o valor calculado é maior do que o valor tabelado rejeita-se H₀, isto é, pode-se afirmar que as crianças apresentam tendência significativa para mudar o objeto de seu interesse, de adulto para outra criança, após 30 dias de freqüência à escola maternal.

3.1.4. Pequenas freqüências esperadas

Se a freqüência esperada, isto é, 1/2 (A + D) é muito pequena (menor do que 5), deve ser usada a prova Binomial no lugar da prova de McNemar. Para o teste Binomial $n = A + D$ e $x =$ menor das duas freqüências observadas A ou D.

3.2. O teste de Wilcoxon

3.2.1. Função

O teste de Wilcoxon é o mais poderoso para o pesquisador do comportamento. Com dados comportamentais não é de todo incomum que o pesquisador possa: (a) dizer qual membro do par é “maior”, isto é, determinar o sentido da diferença dentro do par e (b) ordenar estas diferenças no sentido de seu valor absoluto.



3.2.2. Fundamentos lógicos e método

Seja d_i = valor da diferença dentro do par "i". Para realizar o teste de Wilcoxon deve-se:

- Atribuir postos a cada d_i , independentemente de sinal. Ao menor d_i , atribuir o posto 1; ao próximo o posto 2 e assim por diante.
- A cada posto deve-se atribuir o sinal da diferença, isto é, indique quais postos decorrem de diferenças negativas e quais de diferenças positivas.

Se as duas classificações são equivalentes, isto é se H_0 é verdadeira, é de se esperar que algumas das maiores diferenças sejam positivas e outras negativas. Desta forma, se forem somados os postos com sinal mais e os postos com sinal menos, deve-se esperar somas aproximadamente iguais. Se houver diferença entre estas duas somas é sinal de que as duas classificações (ou tratamentos) não se equivalem e deve-se então rejeitar a hipótese nula.

3.2.3. Empates

Eventualmente os escores de dois pares serão iguais. Neste caso eles são excluídos da análise. É o mesmo procedimento adotado no teste dos sinais. Da mesma forma o valor de n será reduzido na mesma quantidade de valores em que a diferença for nula.

Pode ocorrer, ainda, um outro tipo de empate. Duas ou mais diferenças podem ter o mesmo valor absoluto. Neste caso, atribui-se o mesmo posto aos empates. Este posto é a média dos postos que teriam sido atribuídos se as diferenças fossem diferentes. Por exemplo, se três pares acusam as diferenças: -1, -1 e +1, a cada par será atribuído o posto 2, que é a média entre 1, 2 e 3. O próximo valor, pela ordem, receberia o valor 4, porque já teriam sido utilizados os postos 1, 2 e 3.

3.2.4. Pequenas Amostras

Seja T a menor soma dos postos de mesmo sinal (negativos ou positivos), isto é, ou a soma dos postos positivos ou a soma dos postos negativos (a que for menor). A tabela G (Siegel, pg. 285) fornece vários valores de T com os respectivos níveis de significância. Se um T observado não supera o valor dado na tabela G sob determinado nível de significância para uma amostra de tamanho n , rejeita-se a hipótese de nulidade àquele nível.

A tabela G pode ser usada tanto para testes unilaterais quanto bilaterais. Utiliza-se uma prova unilateral se o pesquisador pode prever, antes de examinar os dados, o sinal da menor soma de postos. Isto é, tal como no caso de todas as provas unilaterais, ele deve poder prever o sentido da diferença.

Exemplo:

Suponha-se que um psicólogo está interessado em testar se a frequência a uma escola maternal tem algum efeito sobre os escores de perceptividade social das crianças. Oito pares de gêmeos são os objetos da sua observação. Um dos gêmeos freqüentará a escola por um período e o outro permanecerá em casa. Ele classifica a percepção através da atitude da criança em relação a um conjunto de figuras que ilustram uma diversidade de situações sociais, formulando um grupo padrão de perguntas sobre cada figura. Assim ele pode obter um escore entre "0" e "100" para cada criança. Ao fim do período escolar, as 16 crianças são submetidas ao teste de perceptividade social.

Hipótese de nulidade: H_0 : Não há diferença entre os graus de perceptividade das crianças que ficaram em casa e das que freqüentaram a escola, ou seja, a soma dos postos negativos é igual a soma dos postos positivos.

H_1 : Os graus de perceptividade social dos dois grupos de crianças são diferentes, isto é, a soma dos postos negativos é diferente da soma dos postos positivos.

Prova Estatística: Escolhe-se a prova de Wilcoxon, pois é um caso de duas amostras relacionadas e proporciona escores de diferenças que podem ser ordenados segundo seus valores absolutos.

Nível de significância: Sejam $\alpha = 0,05$ e n = número de pares (8) menos o número em que eventualmente se tenha $d = 0$.



Região de rejeição. É bilateral, pois não se prevê o sentido da diferença. A região de rejeição consiste de todos os valores de T tão pequenos que a probabilidade de ocorrência, sob H_0 , não seja superior a $\alpha = 0,05$ para uma prova bilateral. .

Decisão: Neste caso, os 8 pares de crianças de “casa” e da “escola” são submetidos ao teste após o segundo grupo ter permanecido na escola durante um período escolar. A tabela 3.5 apresenta os escores obtidos.

Tabela 3.5 - Escores de perceptividade social de 8 pares de crianças.

Pares	Escola	Casa	d	Posto de d
a	82	63	19	7
b	69	42	27	8
c	73	74	-1	-1
d	43	37	6	4
e	58	51	7	5
f	56	43	13	6
g	76	80	-4	-3
h	85	82	3	2

Apenas 2 pares apresentam diferença no sentido de maior perceptividade das crianças que ficaram em casa. E estas diferenças de escore estão entre os menores. e sua soma $T = 1 + 3 = 4$. A tabela G (Siegel, pg. 285) mostra que para $n = 8$ um valor de T igual a 4 permite rejeitar a hipótese ao nível de significância de 5% para um teste bilateral. Desta forma, pode-se concluir que a escola maternal afeta a perceptividade social das crianças.

3.2.5. Grandes Amostras

Quando n é maior do que 25 a tabela G não pode ser utilizada. No entanto, pode ser mostrado que a soma dos postos, T , é aproximadamente normal, com

$$\text{Média} = \mu_T = n(n + 1) / 4 \text{ e desvio padrão } \sigma_T = \sqrt{\frac{n(n + 1)(2n + 1)}{24}}$$

$$\text{Desta forma, } Z = \frac{T - \mu_T}{\sigma_T} = \frac{T - \frac{n(n + 1)}{4}}{\sqrt{\frac{n(n + 1)(2n + 1)}{24}}} \text{ é aproximadamente } N(0, 1).$$

Para mostrar que a aproximação é excelente, mesmo para pequenas amostras, considere o caso anterior em que $n = 8$ e $T = 4$. Colocando estes valores na expressão acima tem-se:

$$z = \frac{4 - \frac{8 \cdot 9}{4}}{\sqrt{\frac{8 \cdot 9 \cdot 17}{24}}} = -1,96$$

Pela tabela da normal, pode-se verificar que a significância deste valor é $p = 5\%$ para um teste bilateral. Este mesmo valor é encontrada na tabela G (Siegel, pg. 285).



4. Testes para duas amostras independentes

4.1. O teste Qui-Quadrado

4.1.1. Função

Utiliza-se esta prova quando os dados da pesquisa se apresentam sob forma de freqüências em categorias discretas. Pode aplicar a prova χ^2 para determinar a significância de diferenças entre dois grupos independentes e conseqüentemente, com respeito a freqüências relativas com que os componentes do grupo se enquadram nas diversas categorias.

4.1.2. Método

A hipótese da nulidade pode ser testada mediante:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \text{ onde:}$$

O_{ij} = n de casos observados na linha "i" da coluna "j".

E_{ij} = n de casos esperados, sob H_0 , na linha "i" da coluna "j".

$\sum_{i=1}^r \sum_{j=1}^k$ indica somatório sobre todas as "r" linhas e todas as "k" colunas.

Os valores de χ^2 obtidos pela fórmula acima, tem distribuição aproximadamente qui-quadrado com $gl = (r - 1)(k - 1)$, onde r = número de linhas e k é o número de colunas.

Para obter a freqüência esperada E_{ij} em cada célula, multiplicam-se os totais marginais comuns a uma determinada célula e divide-se produto por n = total de casos.

Exemplo:

Pode-se ilustrar o método mediante um exemplo simples, com dados fictícios. Suponha-se que se deseje comprovar se existe diferença de qualidade de liderança entre pessoas altas e pessoas baixas. A tabela 4.6 mostra os resultados da classificação de 95 pessoas, que foram divididos entre "altos" e "baixos" de um lado e por outro lado como "líderes", "liderados" e "não-classificáveis". A hipótese de nulidade é de que a altura é independente da classificação como líder ou liderado, isto é, a proporção de líderes "altos" e a mesma que a de líderes "baixos" e o mesmo se verificando entre os liderados.

Tabela 4.6 - Altura e liderança

	Baixo	Alto	Total
Líder	12	32	44
Liderado	22	14	36
Não-classificável	9	6	15
Total	43	52	95

Se as freqüências observadas estão próximas das freqüências esperadas o valor do χ^2 também será pequeno. Com um pequeno valor de χ^2 não podemos rejeitar a hipótese de nulidade, de que os dois conjuntos de características sejam independentes um do outro. Todavia, se algumas ou muitas das diferenças são grandes, o valor de χ^2 será também grande. Quanto maior o valor de χ^2 , tanto maior a probabilidade de que os dois grupos difiram em relação as classificações adotadas. Pode-se mostrar que a distribuição de χ^2 , tal como definida pela fórmula acima, tem distribuição aproximadamente qui-quadrado com:

$$gl = (r - 1)(k - 1)$$

A tabela 4.7 ilustra o cálculo das freqüências esperadas para os dados da tabela 4.6. Assim, por exemplo, a freqüência esperada para a célula 3x2, isto é, E_{32} é: $(52.15) / 95 = 8,2$.



Tabela 4.7 - Altura e liderança

	Baixo	Alto	Total
Líder	12 (19,9)	32 (24,1)	44
Liderado	22 (16,3)	14 (19,7)	36
Não-classificável	9 (6,8)	6 (8,2)	15
Total	43	52	95

O valor do χ^2 é dado por:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = (12 - 19,9)^2 / 19,9 + (32 - 24,1)^2 / 24,1 + (22 - 16,3)^2 / 16,6 + (14 - 19,7)^2 / 19,7 + (9 - 6,8)^2 / 6,8 + (6 - 8,2)^2 / 8,2 = 3,14 + 2,59 + 1,99 + 1,65 + 0,71 + 0,59 = 10,67$$

Para determinar a significância de $\chi^2 = 10,67$, quando $gl = (3 - 1) \cdot (2 - 1) = 2$, utiliza-se a tabela 3 que mostra que este valor é significativo além do nível $\alpha = 0,01$. Portanto, pode-se rejeitar a hipótese de nulidade ao nível de 1%. De fato, neste caso, seria possível rejeitar também ao nível de 0,5%.

4.1.3. Tabelas de Contingência 2X2

Talvez a aplicação mais comum do teste χ^2 consista em comprovar se uma distribuição de valores em uma tabela de contingência 2x2 pode ter ocorrido sob H_0 . Neste caso, a fórmula acima assume a seguinte expressão particular:

$$\chi^2 = \frac{n \left(|AD - BC| - \frac{n}{2} \right)^2}{(A + B)(C + D)(A + C)(B + D)} \text{ com } gl = 1$$

Esta expressão é um pouco mais simples de aplicar do que a fórmula anterior, pois requer apenas uma divisão. Tem ainda a vantagem de incorporar uma correção de continuidade que melhora sensivelmente a aproximação do χ^2 calculado pela distribuição qui-quadrado.

Exemplo:

Adams estudou a relação entre os interesses vocacionais e a escolha do currículo com a taxa de desistência do curso universitário por parte de estudantes superdotados. Os indivíduos observados eram estudantes classificados no percentil 90 nos teste de admissão e que haviam resolvido mudar de carreira após a matrícula. O pesquisador comparou os estudantes destacados cuja escolha curricular se manteve na linha considerada desejável à vista do resultado obtido no teste vocacional de Strong (tais casos sendo considerados como "positivos") com os estudantes destacados cuja escolha curricular se processou em sentido diverso do indicado pelo T teste de interesses. A hipótese do pesquisador é que os estudantes cuja escolha foi considerada "positiva" acusam maior frequência de permanência na faculdade ou curso universitário inicialmente escolhido.

Hipóteses: H_0 : Não há diferença entre os dois grupos (escolha "positiva" e escolha "negativa" de currículo) no que diz respeito a proporção dos estudantes que permanecem na faculdade.

H_1 : A porcentagem de permanência na faculdade é maior entre os estudantes cuja escolha de currículo foi considerada "positiva".

Prova Estatística. Escolhe-se a prova χ^2 para duas amostras independentes porque os dois grupos considerados "positivo" e "negativo" são independentes e porque os escores que estão sendo estudados consistem de frequências em categorias discretas (permanência na faculdade ou afastamento dela).

Nível de Significância. Sejam $\alpha = 0,05$ e $n =$ número de estudantes na amostra = 80.

Distribuição Amostral. χ^2 tal como calculado pela fórmula do exemplo tem distribuição amostral aproximadamente qui-quadrado com $gl = 1$. A tabela 3 dá os valores críticos do qui-quadrado.

Região de Rejeição. A região de rejeição consiste de todos os valores de χ^2 que são tão grandes que a probabilidade associada à sua ocorrência, sob H_0 , não supere $\alpha = 0,05$. Como H_1 prevê o sentido da di-



ferença entre os dois grupos, a região de rejeição é unilateral. A tabela 3 indica que, uma prova unilateral, quando $gl = 1$, $\chi^2 = 3,84$. Portanto, a região de rejeição consiste de todos os $\chi^2 > 3,84$ se o sentido dos resultados é o previsto em H_1 .

Decisão. A tabela dá os resultados obtidos por Adams. Por ali se vê que de 56 estudantes superdotados que fizeram escolha "positiva", 10 se afastaram da universidade, e 46 permaneceram nela. Dos 24, que fizeram escolha "negativa", 11 se afastaram da universidade e 13 permaneceram nela.

Tabela 4.8 - Escolha de currículo e afastamento da Universidade entre os estudantes super dotados

Sentido da escolha curricular			
	Positivo	Negativo	Total
Afastamento	10	11	21
Permanência	46	13	59
Total	56	24	80

O valor do χ^2 para este dados é:

$$\chi^2 = \frac{n \left(|AD - BC| - \frac{n}{2} \right)^2}{(A+B)(C+D)(A+C)(B+D)} = \frac{80 \left(|10 \cdot 13 - 11 \cdot 46| - \frac{80}{2} \right)^2}{(21)(59)(56)(24)} = 5,42$$

A probabilidade de ocorrência, sob H_0 , de $\chi^2 = 5,42$ com $gl = 1$ é menor que 5%. Como este valor é superior ao valor tabelado, a decisão é rejeitar H_0 em favor de H_1 . Conclui-se, pois que os estudantes superdotados cuja a escolha de currículo foi considerado "positiva" acusam maior freqüência de permanência na universidade do que estudantes superdotados cuja escolha foi considerada "negativa".

4.1.4. Quando usar o teste

A prova χ^2 exige que as freqüências esperadas em cada célula não sejam muito pequenas. Quando elas são inferiores ao mínimo exigido, a aplicação da prova pode se tornar inadequada ou mesmo inútil.

O caso 2x2. Se as freqüências se dispõem em uma tabela de contingência 2x2 a decisão quanto ao uso da prova χ^2 deve basear-se nas seguintes considerações:

- Quando $n > 40$ utilizar a prova com correção de continuidade.
- Quando $20 \leq n \leq 40$ a prova pode ser aplicada desde que nenhuma das freqüências esperadas seja inferior a 5. Se a menor freqüência esperada for inferior a 5, utilizar a prova de Fisher (SIE56).
- Quando $n < 20$ utilizar a prova de Fisher em qualquer caso.

Tabelas de contingência com gl superior a um. Quando $k > 2$ (e conseqüentemente $gl > 1$), a prova χ^2 pode ser aplicada somente se o número de células com freqüência esperada inferior a 5 é inferior a 20% do total de células e se nenhuma célula tem freqüência esperada inferior a 1. Se essas condições não são satisfeitas pelos dados da forma em que foram coletados originalmente, o pesquisador deve combinar categorias adjacentes de modo a aumentar as freqüências esperadas nas diversas células. Somente após feita a combinação de categorias de forma a satisfazer as exigências acima é que a prova χ^2 pode ser validamente aplicada.

Quando $gl > 1$ a prova χ^2 é insensível ao efeito de ordem. Por isso, quando determinada hipótese leva em conta a ordem, a prova qui-quadrado pode não ser a melhor opção.

4.2. O teste U de Mann-Whitney

4.2.1. Função

Desde que o grau de mensuração seja pelo menos ordinal, pode-se aplicar a prova U de Mann-Whitney para comprovar se dois grupos independentes foram ou não extraídos da mesma população. Trata-se de uma das mais poderosas provas não-paramétricas e constitui uma alternativa extremamente útil da prova



paramétrica t, quando se deseja evitar as hipóteses exigidas por ela ou quando a mensuração exigida é inferior à de escala de intervalos.

Suponha-se que existam duas amostras extraídas das populações A e B. A hipótese de nulidade é que A e B tenham a mesma distribuição. A hipótese alternativa, H_1 , é que A é maior do que B (teste unilateral). Pode-se rejeitar H_0 se a probabilidade de um escore de A ser maior do que um escore de B é maior do que 1/2. Isto é, se **a** é uma observação da população A e **b** é uma observação da população B, então a hipótese alternativa é que $P(a > b) > 1/2$. Se a evidência apóia H_1 , isto implica que o “grosso” da população A é superior ao “grosso” da população B. Para uma prova bilateral, H_1 seria que $P(a > b) \neq 1/2$.

4.2.2. Método

Seja n_1 = número de casos no menor dos dois grupos independentes e n_2 = número de casos no maior grupo. Para aplicar o teste U, primeiramente combinam-se as observações ou escores de ambos os grupos, relacionando-os por ordem ascendente. Nessa ordenação ascendente, consideram-se os valores algébricos, isto é, os postos mais baixos são atribuídos aos maiores números negativos (se houver).

Focaliza-se agora um dos grupos, seja o grupo que apresentar n_1 casos. O valor de U (a estatística utilizada na prova) é obtido pelo número de vezes que um escore no grupo com n_2 casos precede um escore no grupo com n_1 casos no grupo ordenado crescentemente.

Por exemplo, suponha-se um grupo experimental com 3 casos e um grupo de controle com 4 casos. Aqui $n_1 = 3$ e $n_2 = 4$. Admita-se que os escores sejam os seguintes:

Escores E	9	11	15	
Escores C	6	8	10	13

Para determinar U, ordenam-se primeiro os escores em ordem crescente, tendo o cuidado de identificar a qual grupo cada um pertence (E ou C):

6	8	9	10	11	13	15
C	C	E	C	E	C	E

Considera-se agora o grupo de controle, C, e conta-se o número de escores E que precedem cada escore deste grupo. Nenhum escore E precede o escore C igual a 6. Isto também é verdade para o escore C = 8. O próximo escore C é 10 e é precedido por um escore E. O último escore C, o 13, é antecedido por dois escores E. Assim, $U = 0 + 0 + 1 + 2 = 3$. O número de vezes que um escore E vem antes de um escore C é igual a 3, isto é, $U = 3$.

A distribuição amostral de U, sob H_0 , é conhecida e pode-se então determinar-se a probabilidade associada à ocorrência, sob H_0 , de qualquer valor de U tão extremo quanto o valor observado.

4.2.3. Amostras muito pequenas

Quando nem n_1 e nem n_2 são superiores a 8, pode-se utilizar a tabela J (Siegel, pg. 302-04) para determinar a probabilidade exata associada à ocorrência, sob H_0 , de qualquer U tão extremo quanto o valor observado.

A tabela J é constituída de 6 subtabelas separadas, uma para cada valor de n_2 , de $n_2 = 3$, a $n_2 = 8$. Para determinar a probabilidade, sob H_0 , associada aos dados é necessário saber o valor de n_1 , de n_2 e de U.

No exemplo acima, tem-se: $n_1 = 3$, $n_2 = 4$ e $U = 3$. A subtabela para $n_2 = 4$ da tabela J mostra que $U \leq 3$ tem probabilidade de ocorrência, sob H_0 , de $p = 0,20 = 20\%$.

As probabilidades fornecidas na tabela J são unilaterais. Para um teste bilateral, deve-se duplicar o valor de **p** constante na tabela.

Caso o valor observado de U seja grande e não conste da tabela, existe a possibilidade de ter-se tomado o grupo “errado” para a determinação de U. Neste caso, usa-se a transformação:

$$U = n_1 \cdot n_2 - U', \text{ onde } U' \text{ é o valor não encontrado na tabela.}$$



Exemplo:

Solomon e Coles¹ estudaram se os ratos seriam capazes de generalizar uma imitação aprendida, quando colocados sob nova impulsão (*drive*) e em nova situação. Cinco ratos foram treinados para imitar ratos líderes em um labirinto T. Foram treinados para seguir seus líderes quando estivessem com fome, a fim de atingir o alimento. Em seguida, os cinco ratos foram transferidos para uma nova situação de esquiva ao choque elétrico. Seu comportamento na situação de esquiva ao choque foi então comparado ao de quatro controles que não tinham nenhum treinamento prévio para seguir seus líderes. A hipótese era de que os 5 ratos que já tinham sido treinados para imitar seus líderes transfeririam este treinamento para a nova situação e, assim, aprenderiam a evitar o choque mais depressa do que os 4 ratos de controle. A comparação se fez em termos de quantas tentativas foram precisas para cada rato atingir um critério de 10 respostas corretas em 10 tentativas.

Hipóteses: **H₀:** O número de tentativas para atingir o critério desejado na situação de esquiva ao choque é o mesmo tanto para os ratos previamente treinados a seguir um líder na busca do alimento, quanto para os ratos sem nenhum treinamento prévio. **H₁:** Os ratos com treinamento prévio para seguir um líder na busca de alimento atingirão o critério desejado na nova situação de esquiva ao choque mais rapidamente que os ratos não treinados.

Prova Estatística. Escolhe-se a prova U de Mann-Whitney porque o estudo utiliza duas amostras independentes, pequenas e mensuração (número de tentativas para atingir o critério desejado como índice de velocidade de aprendizagem) provavelmente em escala ordinal, na melhor das hipóteses.

Nível de Significância. Sejam $\alpha = 0,05$ e $n_1 = 4 =$ ratos de controle e $n_2 = 5 =$ ratos experimentais.

Distribuição Amostral. A tabela J (Siegel, pg. 302-04) fornece as probabilidades associadas à ocorrência, sob H_0 , de valores tão pequenos quanto determinado U observado para $n_1, n_2 \leq 8$.

Região de Rejeição. Como H_1 prediz o sentido da diferença, a região de rejeição será unilateral. Consiste de todos os valores de U tão pequenos que a probabilidade associada a sua ocorrência, sob H_0 , não supera $\alpha = 0,05$.

Decisão. Foram os seguintes os números de tentativas necessárias para os ratos E = experimental e C = controle atingirem o critério desejado:

Ratos E	78	64	75	45	82
Ratos C	110	70	53	51	

Dispondo os escores em ordem crescente e mantendo a identidade de cada um, vem:

45	51	53	64	70	75	78	82	110
E	C	C	E	C	E	E	E	C

Obtém-se o valor de U, contando o número de escores E que precedem cada escore C. Assim:

$$U = 1 + 1 + 2 + 5 = 9$$

Na tabela J verifica-se que na subtabela para $n_2 = 5$, $U \leq 9$, quando $n_2 = 4$ tem probabilidade de ocorrência, sob H_0 , de $p = 0,452$. A decisão é que os dados não mostram evidência que justifique a rejeição de H_0 , ao nível dado. Assim, não é possível afirmar que o treinamento prévio para imitar se generalize através de novas situações e novas impulsões.

4.2.4. Amostras médias (n entre 9 e 20)

Se n_2 representar o tamanho da maior das duas amostras e for maior do que 8, a tabela J não poderá mais ser utilizada. Quando $9 \leq n_2 \leq 20$, pode-se aplicar a prova de Mann-Whitney utilizando a tabela K (Siegel, pg. 305-08) que fornece valores críticos de U para os níveis de significância de 0,001, 0,01, 0,025 e 0,05 para um teste unilateral. Para um teste bilateral, os níveis de significância são dados por: 0,002, 0,02, 0,05 e 0,10. Note-se que este conjunto de tabelas fornece valores críticos de U e não probabilidades exatas (como as tabelas J). Isto é, se um valor observado de U, para um dado $n_1 \leq 20$ e n_2 entre 9 e 20, não supera o valor dado na tabela, pode-se rejeitar H_0 , a um dos níveis de significância indicados no cabeçalho da tabela.

¹ SOLOMON, R. L., COLES, M. R. A case of failure of generalization of imitation across drives and across situations. J. Abnorm. Soc. Psychol., 49, 7-13, 1954.



Por exemplo, se $n_1 = 6$ e $n_2 = 13$, um valor de $U = 12$ permite rejeitar H_0 ao nível de $\alpha = 0,01$ em uma prova unilateral e rejeitar H_0 ao nível $\alpha = 0,02$ em uma prova bilateral.

Determinação do valor U. Para valores razoavelmente grandes de n_1 e n_2 , o método para determinar o valor de U pode ser bastante trabalhoso. Um processo alternativo, que dá resultados idênticos, consiste em atribuir posto 1 ao mais baixo escore do grupo combinado de $(n_1 + n_2)$ escores, o posto 2 ao escore seguinte, etc. Então:

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 \text{ ou, de forma equivalente } U = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2$$

onde R_1 = soma dos postos atribuídos ao grupo cujo tamanho de amostra é n_1 ,

R_2 = soma dos postos atribuídos ao grupo cujo tamanho de amostra é n_2 .

Por exemplo, poder-se-ia ter utilizado este processo para determinar o valor de U no caso de pequenas amostras tratado acima. Os escores E e C, bem como seus postos, são apresentados novamente na tabela 4.12.

Aplicando a fórmula acima vem:

$$U = 4.5 + 5.(5 + 1) / 2 - 26 = 9$$

9

Tabela 4.12 - Tentativas dos ratos para atingir o critério desejado

Escore E	Posto	Escore C	Posto
78	7	110	9
64	4	70	5
75	6	53	3
45	1	51	2
82	8		
Soma	$R_2 = 26$	Soma	$R_1 = 19$

O menor dos dois valores de U é aquele cuja distribuição amostral constitui a base da tabela K (Siegel, pg. 305-08).

4.2.5. Grande amostras ($n > 20$)

Nem a tabela J e nem a K podem ser utilizadas quando $n_2 > 20$. Todavia, Mann e Whitney mostraram (1947), que à medida que n_1 e n_2 aumentam, a distribuição amostral de U tende rapidamente para a distribuição normal, com:

$$\text{Média} = \mu_U = (n_1 n_2) / 2$$

e desvio padrão

$$\sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}, \text{ isto é, quando } n_2 > 20, \text{ o valor de: } Z = \frac{U - \mu_U}{\sigma_U} = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} \text{ tem dis-}$$

tribuição aproximadamente $N(0, 1)$.

4.2.6. Empates

A prova de Mann-Whitney supõe que os escores representem uma distribuição basicamente contínua. Numa distribuição contínua a probabilidade de um empate é zero. Todavia, como a mensuração tem uma precisão limitada, os empates podem ocorrer. Admite-se que as observações que estejam empatadas, tenham, na realidade, escores diferentes, e que esta diferença é muita pequena para ser detectada pelo instrumento de medida.

Quando ocorrem empates atribuí-se a cada um dos valores empatados a média dos postos que lhes seriam atribuídas se não houvesse empate.

Se os empates ocorrem entre dois ou mais valores do mesmo grupo, o valor de U não é afetado. Mas se os empates ocorrem entre duas ou mais observações envolvendo os dois grupos, então o valor de U é afetado. Embora, os efeitos práticos dos empates sejam desprezíveis existe uma correção para empates que deve ser utilizada com a aproximação normal para grandes amostras. O efeito dos postos empatados modifica a variabilidade do conjunto de postos. Assim, a correção deve ser aplicada ao desvio padrão da distribuição amostral de U . Com esta correção o desvio padrão é dado por:



$$\sigma_U = \sqrt{\left(\frac{n_1 n_2}{n(n-1)}\right) \left(\frac{n^3 - n}{12} - \sum T\right)}, \text{ onde } n = n_1 + n_2 \text{ e } T = (t^3 - t) / 12 \text{ (} t = \text{número de escores empatados}$$

para um determinado posto).

Pode-se verificar que não houver empates a expressão acima se reduz a anterior.

4.3. O teste de Kolmogorov-Smirnov

4.3.1. Função e fundamentos lógicos

A prova de Kolmogorov-Smirnov (K-S) de duas amostras comprova se elas foram extraídas da mesma população (ou de populações com a mesma distribuição). A prova bilateral é sensível a qualquer diferença nas distribuições das quais se extraíram as amostras - diferenças na posição central, na dispersão, na assimetria, etc. A prova unilateral é utilizada para determinar se os valores da população da qual se extraiu uma das amostras são, ou não, estocasticamente maiores do que os valores da população que originou a outra amostra, por exemplo, para testar a hipótese de que os escores de um grupo experimental serão “melhores” do que os escores do grupo de controle.

Tal como a prova de K-S para uma amostra (item 2.3) o teste utiliza as distribuições acumuladas. A prova de uma amostra se refere a concordância entre a distribuição de um conjunto de valores amostrais e determinada distribuição teórica. A prova de duas amostras visa a concordância entre dois conjuntos de valores amostrais.

Se as duas amostras foram extraídas da mesma população, então é de se esperar que as distribuições acumuladas das duas amostras sejam bastante próximas uma da outra, acusando apenas desvios casuais em relação à distribuição da população. Se as distribuições acumuladas são “diferentes” ou “distantes” uma da outra em qualquer ponto, isto sugere que as amostras provenham de populações também distintas. Assim um desvio grande pode levar a rejeição da hipótese de nulidade.

4.3.2. Método

Para aplicar a prova de Kolmogorov-Smirnov de duas amostras, constrói-se a distribuição das frequências acumuladas relativas de cada uma das amostras, utilizando os mesmos intervalos (amplitude de classes) para cada uma delas. Em cada intervalo subtraí-se uma função da outra. A prova utiliza como estatística o maior destas diferenças.

Sejam $S_{n_1}(x)$ = função acumulada observada para a primeira amostra, isto é, $S_{n_1}(x) = k / n_1$, onde k = número de escores não superiores a x . Seja $S_{n_2}(x)$ = função acumulada observada da segunda amostra, isto é, $S_{n_2}(x) = k / n_2$. O teste K-S toma a maior destas diferenças em módulo que é denominada de *desvio máximo* e é anotada por **D**. Assim:

$$D = [S_{n_1}(x) - S_{n_2}(x)] \text{ para uma prova bilateral e } D = |S_{n_1}(x) - S_{n_2}(x)| \text{ para uma prova unilateral.}$$

A distribuição amostral de D , sob H_0 , para uma prova bilateral é conhecida (*Smirnov, 1948, Massey, 1951*) e se encontra tabelada.

4.3.3. Pequenas amostras

Quando $n_1 = n_2$ e não são superiores a 40, então pode-se utilizar a tabela L (Siegel, pg. 309) para comprovar a hipótese de nulidade. O corpo da tabela fornece vários valores da quantidade K_D , que é definida como o numerador da maior diferença entre as duas distribuições acumuladas, isto é, o numerador D . Para ler um valor nesta tabela, deve-se conhecer o valor de $n = n_1 = n_2$ e o valor de K_D . Por exemplo, em uma prova unilateral com $n = 14$, se $K_D \geq 8$, pode-se rejeitar H_0 ao nível de $\alpha = 0,01$.



Exemplo:

Lepley² comparou o aprendizado serial de 10 alunos do sétimo grau com o aprendizado serial de 10 alunos do décimo primeiro grau, para comprovar a hipótese de que o efeito de primazia é menos predominante no aprendizado de estudantes mais jovens. O efeito de primazia é a tendência para reter a matéria aprendida no começo de determinada série mais facilmente do que a matéria aprendida no fim daquela série. Lepley comprovou sua hipótese comparando a percentagem de erros cometidos pelos dois grupos na primeira metade da série, prevendo que o grupo mais velho (alunos do décimo primeiro grau) cometeria relativamente menos erros do que o grupo mais jovem, ao evocar a primeira metade da série.

Tabela 4.14 - Percentagem de erros totais

Alunos do 7º grau	Alunos do 11º grau
39,1	35,2
41,2	39,2
45,2	40,9
46,2	38,1
48,4	34,4
48,7	29,1
55,0	41,8
40,6	24,3
52,1	32,4
47,2	32,6

Hipóteses: **H₀:** Não há diferença na proporção de erros cometidos ao evocar a primeira metade da série, entre os alunos dos dois graus.

H₁: Os alunos do décimo primeiro grau cometem relativamente menos erros do que os de sétimo grau ao evocarem a primeira metade da série.

Prova Estatística. Emprega-se a prova K-S porque o pesquisador deseja comparar duas amostras pequenas e independentes, de mesmo tamanho.

Nível de significância. Sejam $\alpha = 0,01$ e $n_1 = n_2 = n =$ número de indivíduos em cada grupo = 10.

Distribuição amostral. A tabela L (Siegel, pg. 309) apresenta vários valores críticos de k_D para $n_1 = n_2$ quando n_1 e n_2 são inferiores a 40.

Região de Rejeição. Como H_1 prediz o sentido da diferença, a região de rejeição é unilateral. H_0 será rejeitada se o valor de k_D do maior desvio na direção prevista for tão grande que a probabilidade associada à sua ocorrência, sob H_0 , não seja superior a 0,01.

Decisão. A tabela 4.14 fornece a percentagem dos erros cometidos por cada aluno ao evocar a primeira metade da série. Para análise pela prova K-S os dados foram dispostos em duas distribuições de frequências acumuladas, apresentadas na tabela 4.15.

Note-se que a maior diferença entre as duas séries é 7/10. Assim $k_D = 7$, que é o numerador desta diferença máxima. Utilizando a tabela L, para $n = 10$, vê-se que este valor é significativo ao nível $\alpha = 1\%$ para uma prova unilateral. A decisão é, portanto, rejeitar H_0 em favor de H_1 . Conclui-se que os alunos do décimo primeiro grau cometem proporcionalmente menos erros do que os do sétimo grau, ao evocar a primeira metade da série.

Tabela 4.15 - Distribuições acumuladas dos dados da tabela 4.14

Classes	% erros na primeira metade da série							
	24-27	28-31	32-35	36-39	40-43	44-47	48-51	52-55
$Sn_1(x)$	1/10	2/10	5/10	7/10	10/10	10/10	10/10	10/10
$Sn_2(x)$	0/10	0/10	0/10	0/10	3/10	5/10	8/10	10/10
$Sn_1(x) - Sn_2(x)$	1/10	2/10	5/10	7/10	7/10	5/10	2/10	0

4.3.4. Grandes amostras: prova bilateral

Quando tanto n_1 quanto n_2 são maiores do que 40, pode-se utilizar a tabela M (Siegel, pg. 310) para realizar o teste K-S de duas amostras independentes. Neste caso, não é necessário que n_1 seja igual a n_2 .

Para utilizar a tabela M, determina-se o valor de D para os dados observados, por meio da expressão $D = [Sn_1(x) - Sn_2(x)]$. Compara-se então esse valor observado com o valor crítico que se obtém substituindo os valores observados de n_1 e n_2 na expressão dada pela tabela M. Se o D observado é no mínimo igual ao calcu-

^LEPLEY, W. M., *Serial reactions considered as considered as reactions*. Psychol. Monogr., 46, n. 205, 1934



lado por esta expressão, então, ele pode ser rejeitado ao nível de significância (bilateral) associado àquela expressão.

Por exemplo, suponha-se que $n_1 = 55$ e $n_2 = 60$ e que se queira uma prova bilateral ao nível de 5%. Na coluna da tabela M correspondente a $\alpha = 5\%$, será encontrado o valor de D que os dados devem pelo menos igualar, para que se possa rejeitar a hipótese de nulidade. Efetuando os cálculos, verifica-se que D deve ser maior ou igual a 0,254 para a rejeição de H_0 , pois:

$$1,36 \sqrt{\frac{n_1 + n_2}{n_1 n_2}} = 1,36 \sqrt{\frac{55 + 60}{(55)(60)}} = 0,254$$

4.3.5. Grandes amostras: prova unilateral

Quando n_1 e n_2 são grandes e independentemente de ser $n_1 = n_2$, não se pode aplicar uma prova unilateral considerando $D = [Sn_1(x) - Sn_2(x)]$.

Neste caso deve ser utilizada a expressão:

$$\chi^2 = 4D^2 \frac{n_1 n_2}{n_1 + n_2}$$

Goodman (1954) mostrou que esta expressão tem uma distribuição amostral aproximadamente qui-quadrado com $gl = 2$. Ou seja, pode-se determinar a significância de um valor observado D, tal como calculado pela expressão de D acima, aplicando esta expressão em relação aos valores observados de D, n_1 e n_2 e recorrendo a tabela do qui-quadrado, ao invés da tabela L ou M.

Exemplo:

Em um estudo dos correlatos da estrutura da personalidade autoritária³ formulou-se a hipótese de que as pessoas com alto grau de autoritarismo apresentariam maior tendência para possuir estereótipos sobre membros de diversos grupos nacionais e étnicos, do que pessoas com baixo grau de autoritarismo. A hipótese foi comprovada em um grupo de 98 estudantes universitárias selecionadas ao acaso. Cada uma recebeu 20 fotografias e foi solicitada a "i-

Tabela 4.16 - Nº de fotos identificadas por 98 universitárias

Nº de fotos identificadas	Escore baixo	Escore alto
0 - 2	11	1
3 - 5	7	3
6 - 8	8	6
9 - 11	3	12
12 - 14	5	12
15 - 17	5	14
18 - 20	5	6

dentificar" aquelas cuja nacionalidade reconhecia, "casando" a fotografia apropriada com o nome do grupo nacional. Não havia restrição quanto ao número de fotos que pudessem identificar pelo processo descrito. Acontece que (sem que as estudantes soubessem) todas as fotos eram de pessoas de nacionalidade mexicana - ou candidatas à legislatura mexicana ou vencedoras de um concurso de beleza mexicana; e como a lista de 20 nacionalidades não incluía a nacionalidade mexicana, o número de fotos que cada um identificasse constituiria um índice de sua tendência à estereotipia. O grau de autoritarismo, medido pela escala F (de Adorno et al., 1950), foi classificado como "alto" ou "baixo". Escores considerados altos foram os situados acima da mediana e baixos os situados abaixo da mediana.

Hipóteses: **H₀:** As universitárias com baixo grau de autoritarismo identificariam tantos fotos quanto as universitárias com alto grau de autoritarismo.

H₁: As universitárias com alto grau de autoritarismo identificariam maior número de fotos do que as universitárias com baixo grau.

Prova Estatística. Como a prova envolve amostras independentes foi escolhida o teste K-S.

Nível de significância. Seja $\alpha = 0,01$. Os tamanhos de n_1 e n_2 só podem ser determinados após a coleta dos dados, pois as pessoas serão agrupadas conforme o escore esteja acima ou abaixo do escore mediano do grupo todo.

³ SIEGEL, S. *Certain determinants and correlates of authoritarianism*. Genet. Psychol. Monogr., 49, 187-229, 1954.



Distribuição amostral. Para grandes amostras a distribuição adequada é a qui-quadrado.

Região de Rejeição. Como H_1 não prevê o sentido da diferença entre os dois grupos, utiliza-se uma prova unilateral.

Decisão. Para aplicar a prova K-S os dados são reagrupados nas distribuições de freqüências acumuladas conforme tabela 4.17. A maior das diferenças que se pode verificar é 0,41 (coluna 4), isto é, $D = 0,41$.

Tabela 4.17 - Dados da tabela 4.16 dispostos para aplicação da prova K-S

Classes	Número de fotos "identificadas"						
	0 - 2	3 - 5	6 - 8	9 - 11	12 - 14	15 - 17	18 - 20
$Sn_1(x)$	11/44	18/44	26/44	29/44	34/44	39/44	44/44
$Sn_2(x)$	1/54	4/54	10/54	22/54	34/54	48/54	54/54
$Sn_1(x) - Sn_2(x)$	0,23	0,34	0,41	0,25	0,14	-0,03	0

O valor do qui-quadrado é então obtido por:

$$\chi^2 = 4D^2 \frac{n_1 n_2}{n_1 + n_2} = 4 \cdot (0,41)^2 [44 \cdot 54 / (44 + 54)] = 15,97$$

A tabela do qui-quadrado indica que a probabilidade associada a $\chi^2 = 15,97$ para $gl = 2$ é $p = 0,005$ (prova unilateral). Como este valor é inferior a $\alpha = 0,01$, pode-se rejeitar H_0 . Concluí-se que as mulheres universitárias com alto grau de autoritarismo acusam maior tendência à estereotipia do que as com baixo grau de autoritarismo.



5. Testes para k amostras relacionadas

O teste de Friedman (Análise de variância de dupla classificação por postos)

5.1. Função

Quando os dados de k amostras correspondentes se apresentam pelo menos em escala ordinal, a prova de Friedman é útil para comprovar a hipótese de nulidade, de que as k amostras tenham sido extraídas da mesma população. Como as k amostras estão em correspondência, o número de casos é o mesmo para cada uma delas. A correspondência pode ser estabelecida, quando se estuda o mesmo grupo de indivíduos sob cada uma das k condições. Ou pode-se obter vários conjuntos, cada um deles com k indivíduos em correspondência, associado-se, em seguida, aleatoriamente, um indivíduo em cada conjunto à primeira condição, um indivíduo em cada conjunto à segunda condição, etc. Por exemplo, desejando estudar as diferenças no aprendizado sob quatro métodos de ensino, pode-se obter n conjuntos de $k = 4$ alunos, cada conjunto constituído de alunos que se correspondem segundo variáveis relevantes (idade, aprendizado prévio, inteligência, situação sócio-econômica, etc.) associando-se em seguida, aleatoriamente, um aluno de cada um dos n conjuntos ao método de ensino A, outro de cada conjunto ao método B, outro ao método C e o quarto ao método D.

5.2. Fundamentos lógicos do método

Para a prova de Friedman, os dados se dispõem em uma tabela de dupla entrada com n linhas e k colunas. As linhas representam os vários indivíduos ou conjuntos correspondentes de indivíduos, e as colunas representam as diversas condições. Se estão sendo estudados os escores de indivíduos observados sob todas as condições, então cada linha dá os escores de um indivíduo sob as k condições.

Os dados da prova são postos. Aos escores de cada linha atribuem-se postos separadamente. Isto é, com k condições em estudo, os postos em qualquer linha vão de 1 a k . A prova de Friedman determina se é provável que as diferentes colunas de postos (amostras) provenham da mesma população. Por exemplo, suponha-se que se queira estudar os escores de 3 grupos sob 4 condições. Aqui $k = 4$ e $n = 3$. Cada grupo contém 4 indivíduos correspondentes, um associado a cada uma das 4 condições. Suponha-se que os escores obtidos sejam os da tabela 5.1.

Tabela 5.1 - Escores de três grupos correspondentes sob quatro condições

	Condições			
	I	II	III	IV
Grupo A	9	4	1	7
Grupo B	6	5	2	8
Grupo C	9	1	2	6

Para aplicar a prova de Friedman a estes dados, primeiro atribuí-se postos aos escores em cada linha. Ao mais baixo escore em cada linha pode-se atribuir o posto 1, ao seguinte em cada linha o posto 2, etc. Obtém-se assim os dados mostrados na tabela 5.2. Note-se que os postos em cada linha da tabela vão de 1 a $k = 4$.

Tabela 5.2 - Postos de três grupos correspondentes sob quatro condições

	Condições			
	I	II	III	IV
Grupo A	4	2	1	3
Grupo B	3	2	1	4
Grupo C	4	1	2	3
R_i	11	5	4	10

Se a hipótese de nulidade (de que todas as amostras - colunas - provenham da mesma população) é, de fato, verdadeira, então a distribuição de postos em cada coluna será aleatória, sendo então de se esperar que os postos 1, 2, 3 e 4 apareçam em todas as colunas com freqüências aproximadamente igual. Isto indica



que, para qualquer grupo, é uma questão de acaso sob que condição ocorre o menor escore, o que seria o caso se as condições realmente não diferissem entre si. Se os escores fossem dependentes das condições (isto é, se H_0 fosse falsa), então os totais de postos variariam de uma coluna para outra. Como as colunas contém, todas elas, o mesmo número de casos, uma afirmativa equivalente seria que, sob H_0 , os postos médios das várias colunas seriam aproximadamente iguais.

A prova de *Friedman* determina se os totais dos postos (R_j) diferem significativamente. Para aplicar o teste, calcula-se o valor de uma estatística que *Friedman* representou por χ_r^2 .

Quando o número de linhas e/ou colunas não é muito pequeno, pode-se mostrar (*Friedman*, 1937) que χ_r^2 tem uma distribuição aproximadamente qui-quadrado, com $gl = k - 1$, sendo:

$$\chi_r^2 = \frac{12}{nk(k+1)} \sum_{j=1}^k R_j^2 - 3n(k+1), \text{ onde}$$

n = número de linhas,

k = número de colunas,

R_j = soma dos postos da coluna j

Note-se que χ_r^2 tem distribuição aproximadamente qui-quadrado com $gl = k - 1$ somente quando o número de linhas e/ou colunas não é muito pequeno. Quando o número de linhas ou de colunas é inferior ao mínimo, existem tabelas com as probabilidades exatas que devem ser utilizadas. A tabela N (*Siegel*, pg. 311-12) dá as probabilidades exatas associadas a valores tão grandes quanto um χ_r^2 observado, para $k = 3$ e n variando de 2 a 9 e $k = 4$ e n variando de 2 a 4. Se os valores de n e k são superiores aos valores fornecidos na tabela N, pode-se então utilizar a expressão acima e utilizar a tabela do qui-quadrado.

Para ilustrar o uso da tabela N, considere-se os valores do exemplo acima. Aplicando a expressão tem-se:

$$\chi_r^2 = \frac{12}{nk(k+1)} \sum_{j=1}^k R_j^2 - 3n(k+1) = \frac{12}{3 \cdot 4(4+1)} [1^2 + 5^2 + 4^2 + 10^2] - 3 \cdot 3 \cdot (4+1) = 7,40$$

Pode-se determinar a probabilidade ocorrência, sob H_0 , de $\chi_r^2 \geq 7,40$, verificando a tabela N_{11} que fornece a probabilidade exata, associada a valores tão grandes quanto um χ_r^2 observado para $k = 4$, que, neste caso, é $p = 0,033$. Pode-se, portanto, com tais dados, rejeitar a hipótese de nulidade de que as 4 amostras tenham sido extraídas da mesma população com respeito à locação (postos médios) ao nível de significância de 3%.

Exemplo: (para n e k grandes)

Em um estudo do efeito de três padrões diferentes de reforço sobre a extensão do aprendizado discriminativo entre ratos, treinaram-se três amostras correspondentes ($k = 3$) de 18 ratos ($n = 18$) sob três tipos de reforço. Estabeleceu-se a correspondência utilizando-se 18 conjuntos de ratos de mesma cria, três em cada conjunto. Conquanto todos os 54 ratos tenham recebido a mesma quantidade de reforço (recompensa), o modo de administrar esse reforço foi diferente para cada um dos grupos. Um grupo foi treinado com 100% de reforço (RR), outro grupo foi treinado sob um reforço parcial em que cada seqüência de tentativas terminava com uma tentativa não re-

Tabela 5.34 - Postos de dezoito grupos correspondentes no estudo de transferência de aprendizado após treinamento sob três condições diferentes de reforço

Grupo	Tipo de reforço		
	RR	RU	UR
1	1	3	2
2	2	3	1
3	1	3	2
4	1	2	3
5	3	1	2
6	2	3	1
7	3	2	1
8	1	3	2
9	3	1	2
10	3	1	2
11	2	3	1
12	3	2	1
13	3	2	1
14	2	3	1
15	2,5*	2,5*	1
16	3	2	1
17	3	2	1
18	2	3	1
R_j	39,5	42,5	26,0



compensada (RU) e o terceiro grupo foi treinado sob recompensa parcial, cada seqüência de tentativas terminando com uma tentativa recompensada (UR).

Ao cabo desse treinamento, mediu-se a extensão do aprendizado pela rapidez com que os diversos ratos adquiriram um hábito "oposto": embora treinados para correrem em direção ao branco, eram agora estimulados a correr em direção do preto. Quanto melhor tivesse sido o aprendizado inicial, mais lenta deveria ser essa transferência de aprendizado. Predição: os diferentes tipos de reforço (recompensa) utilizados resultariam em diferentes graus de capacidade de transferência de aprendizado.

Hipóteses: Ho: Os diversos tipos de reforço não têm efeito diferencial.

H₁: Os diversos tipos de reforço têm efeito diferencial.

Prova estatística: Como o número de erros na transferência de aprendizado não é provavelmente uma medida intervalar da força do aprendizado original, escolheu-se a prova de *Friedman* (não-paramétrica) ao invés da prova paramétrica correspondente (**análise de variância**). Além disso, não se pode utilizar a análise de variância porque os escores acusaram possível falta de homogeneidade de variância, e, desta forma, os dados indicam que uma das suposições básicas para aplicação da prova F (de *Snedcor*) não foi satisfeita.

Nível de significância: Sejam $\alpha = 0,05$ e $n = 18$ = número de ratos em cada um dos 3 grupos correspondentes.

Distribuição Amostral: A distribuição qui-quadrado com $gl = k - 1$.

Região de rejeição Consiste de todos os valores χ^2 tais que a probabilidade de sua ocorrência, sob H_0 , não supere $\alpha = 0,05$.

Decisão: Determinou-se o número de erros cometidos por cada rato na situação de transferência de aprendizado, dispondo-se os escores em postos para cada um dos 18 conjuntos de 3 ratos correspondentes. A tabela 5.3 fornece estes postos.

Note-se que a soma dos postos para o grupo RR é 39,5, a soma dos postos para o grupo RU é 42,5 e a soma dos postos para o grupo UR é 26,0. Um posto baixo indica elevado número de erros na transferência, isto é, forte fixação do aprendizado original. Pode-se calcular χ^2 substituindo os valores observados na expressão:

$$\chi^2 = \frac{12}{nk(k+1)} \sum_{j=1}^k R_j^2 - 3n(k+1) = \frac{12}{18 \cdot 3(4+1)} [39,5^2 + 42,5^2 + 26^2] - 3 \cdot 18 \cdot (3+1) = 8,40$$

A tabela qui-quadrado indica que $\chi^2 = 8,40$ quando $gl = k - 1 = 3 - 1 = 2$ é significativo entre os níveis 0,025 e 0,01. Como $p < 0,02$ é inferior ao nível de significância $\alpha = 0,05$, rejeita-se H_0 , concluindo que os escores de transferência de aprendizado dos ratos dependem do tipo de reforço (recompensa) utilizado nas tentativas de aprendizado original.

Empates: No grupo 15, assinalado com asterisco na tabela 5.3, os animais RR e RU obtiveram escores iguais, empatando nos postos dois e três. Neste caso, foi atribuído a ambos o posto 2,5 (média daqueles postos). Friedman afirma que a substituição de valores empatados pelo seu valor médio não afeta a validade do teste.



6. Testes para k amostras independentes

6.1. O teste qui-quadrado

6.1.1. Função

Quando os dados de um levantamento consistem de freqüências em categorias discretas (nominais e ordinais), pode-se usar a prova χ^2 para determinar a significância das diferenças entre **k** grupos independentes. A prova χ^2 para **k** amostras independentes é uma extensão direta da prova qui-quadrado para duas amostras independentes. Em geral, o teste é o mesmo, tanto para duas, como para **k** amostras independentes.

6.1.2. Método

Dispõem-se as freqüências em uma tabela **kxr**. A hipótese de nulidade é que as **k** amostras de freqüência ou proporções provenham da mesma população ou de populações idênticas. Esta hipótese, de que as **k** amostras não difiram entre si, pode ser comprovada aplicando-se a seguinte expressão:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}}, \text{ onde:}$$

O_{ij} = número de casos observados classificados na linha "i" da coluna "j" e

E_{ij} = número de casos esperados, sob H_0 , na linha "i" da coluna "j",

Sob H_0 , pode-se mostrar que a distribuição amostral de χ^2 , tem distribuição aproximadamente qui-quadrado com $gl = (k - 1)(r - 1)$, onde **k** = número de colunas e **r** = número de linhas. Assim, a probabilidade associada à ocorrência de valores tão grandes quanto um valor observado de χ^2 é igual a, ou maior do que, o valor dado na tabela do qui-quadrado, para determinado nível de significância e $gl = (k - 1)(r - 1)$, então H_0 , pode ser rejeitada àquele nível.

Exemplo:

Em uma investigação da natureza e conseqüências da estratificação social em uma pequena comunidade do Oeste Médio americano,⁴ Hollingshead constatou que os membros da comunidade se dividiam entre si em cinco classes sociais, I, II, III, IV e V. Sua pesquisa centrou-se nos correlatos dessa estratificação entre os jovens da comunidade. Uma de suas predições era que os adolescentes de diferentes classes sociais se matriculariam em diferentes cursos (preparatório para a universidade, comercial ou geral) no ensino de segundo grau de Elmtown. Hollingshead testou sua hipótese identificando a classe social de 390 alunos e determinando o curso em cada um se havia matriculado.

Hipóteses: H_0 : A proporção de estudantes matriculados nos três diferentes é a mesma em todas as classes sociais. H_1 : A proporção de estudantes matriculados nos cursos distintos difere de classe para classe.

Prova estatística: Como os grupos em estudo são independentes e em número superior a 2, emprega-se uma prova estatística para **k** amostras independentes. E como os dados se apresentam em categorias discretas, a prova χ^2 é a prova arqueada.

Nível de significância: Sejam $\alpha = 0,01$ e $n = 390$ = número de estudantes cujas classes sociais e tipos de cursos foram estudados.

Distribuição Amostral: A distribuição qui-quadrado com $gl = (k - 1)(r - 1)$.

Região de rejeição: Consiste de todos os valores χ^2 tais que a probabilidade de sua ocorrência, sob H_0 , não supere $\alpha = 0,01$.

Decisão: A tabela 6.1 fornece as matrículas por curso dos 390 alunos de *Elmtown* estudados por *Hollingshead*. As classes sociais I e II foram agrupadas em razão do pequeno número de componentes de cada uma delas, particularmente na classe I. A tabela 6.1 fornece, também, em itálico, o número de jovens cuja matrícula em cada um dos três cursos era esperada sob H_0 , ou seja, as matrículas esperadas se não houvesse

⁴ HOLLINGSHEAD, A. B. *Elmtown's youth: The impact of social classes on adolescents*. New York: Willey, 1949.



realmente diferença nas preferências entre as diversas classes sociais. Por exemplo, enquanto que a matrícula efetiva no curso preparatório para a universidade, nas classes I e II, foi de 23, a matrícula esperada seria de apenas 7,3.

Tabela 6.1 - Frequência de matrícula de jovens de Elmtown de 5 classes sociais em 3 cursos

Curso	Classe				Total
	I e II	III	IV	V	
Preparatório para a universidade	23 7,2692	40 30,2330	16 38,0076	2 5,4000	81
Geral	11 18,5769	75 77,4923	107 97,1307	14 13,8000	207
Comercial	1 9,1538	31 38,1846	60 47,8615	10 6,8000	102
Total	35	146	183	26	390

O tamanho de χ^2 reflete a magnitude da discrepância entre os valores observados e os esperados em cada uma das células da tabela. Pode-se calcular χ^2 para os valores da tabela 6.1, através da expressão definida anteriormente:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = 33,8 + 3,1 + 12,7 + 2,1 + 3,1 + 0,08 + 1,0 + 0,003 + 7,3 + 1,4 + 3,1 + 1,5 = 69,2$$

Assim, para estes dados o valor do $\chi^2 = 69,2$ com $gl = (k - 1)(r - 1) = (4 - 1)(3 - 1) = 6$

Verificando uma tabela do qui-quadrado, pode-se constatar que este valor é significativo muito além do nível de 0,005. Como $p < 0,005$ é inferior ao nível de significância estabelecido (de 1%) a decisão é rejeitar H_0 . Conclui-se, pois que a escolha da matrícula pelos alunos de *Elmtown* não é independente das respectivas classes sociais.

6.1.3. Quando usar a prova do qui-quadrado

A prova do qui-quadrado exige que as frequências esperadas (E_{ij}) em cada célula não sejam muito pequenas. Quando tal exigência não é cumprida, os resultados do teste não são válidos. *Cochran* (1954) recomenda que, quando k ou r é maior do que 2, ou seja, no caso de testes onde o grau de liberdade é superior a um, o teste qui-quadrado somente seja aplicado se pelo menos 80% das células tenham frequência esperada superior a 5 e nenhuma das células tenha frequência esperada inferior a 1 (um).

6.2. O teste de Kruskal-Wallis (Análise de variância de uma classificação por postos)

6.2.1. Função

O teste de *Kruskal-Wallis*, é uma prova útil para decidir se k amostras independentes provêm de populações diferentes. Os valores amostrais quase que invariavelmente diferem entre si e o problema é decidir se essas diferenças entre as amostras significam diferenças efetivas entre as populações, ou se representam apenas variações casuais, que podem ser esperadas entre amostras aleatórias de uma mesma população. O teste supõe que a variável em estudo tenha distribuição contínua e exige mensuração no mínimo ao nível ordinal.

6.2.2. Método

No cálculo da prova de *Kruskal-Wallis* cada uma das n observações é substituída por um posto. Isto é, todos os escores de todas as k amostras combinadas são dispostos em uma única série de postos. Ao menor escore atribui-se o posto 1, ao seguinte o posto 2 e assim sucessivamente até o maior posto que é n , onde n = número total de observações independentes nas k amostras.

Feito isso, determina-se a soma dos postos em cada amostra (coluna). A prova então testa se estas somas são tão diferentes entre si que não seja provável que tenham sido todas retiradas de uma mesma população.



Pode-se mostrar que se as k amostras forem efetivamente retiradas de uma mesma população, isto é, se H_0 é verdadeira, então H (estatística de *Kruskal-Wallis* calculada abaixo) tem uma distribuição qui-quadrado com $gl = k - 1$, desde que os tamanhos das k amostras não sejam muito pequenos. Isto é:

$$H = \frac{12}{n(n+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(n+1), \text{ onde}$$

k = número de amostras,

n_j = número de elementos na amostra “ j ”,

R_j = soma dos postos na amostra (coluna) “ j ”,

$n = \sum n_j$ = número total de elementos em todas as amostras combinadas,

tem distribuição aproximadamente qui-quadrado com $gl = k - 1$, para tamanhos de amostras (n_j) suficientemente grandes.

Quando existem mais de cinco elementos em cada amostra, isto é, $n_j > 5$, a probabilidade associada à ocorrência, sob H_0 , de valores tão grandes quanto um H observado pode ser determinada com o auxílio da tabela qui-quadrado, para um nível de significância fixado e para $gl = k - 1$, então H_0 pode ser rejeitada a este nível.

Quando $k = 3$ e o número de casos em cada uma das 3 amostras é 5 ou menos, a aproximação pelo qui-quadrado da distribuição de H não é boa. Para tais casos, deve ser utilizado a tabela O (Siegel, pg. 313-14). A primeira coluna desta tabela fornece o número de elementos em cada uma das 3 amostras, isto é, os diversos valores possíveis para n_1 , n_2 e n_3 . A segunda coluna fornece diversos valores de H , calculados pela expressão acima. A terceira fornece a probabilidade associada à ocorrência, sob H_0 , de valores tão grandes quanto um H observado.

Por exemplo, se $H \geq 5,83$ quando as 3 amostras contêm 4, 3 e 1 elementos, a tabela O mostra que a hipótese de nulidade pode ser rejeitada ao nível de significância de 0,021.

Exemplo: (para pequenas amostras)

Suponha que se deseje comprovar a hipótese de que administradores escolares são tipicamente mais autoritários do que os professores. Sabe-se, no entanto, que os dados para testar esta hipótese podem ser tendenciosos, pois vários professores tem aspirações administrativas. Para evitar esta tendenciosidade, planeja-se dividir os 14 valores em 3 grupos: professores (professores que pretendem continuar nesta posição) professores/administradores (professores que tem aspirações administrativas) e administradores. O autoritarismo é medido através da escala F^5 e a hipótese é de que os três grupos vão diferir quanto as médias na escala F .

Hipóteses: H_0 : Não existe diferença nos escores F entre os três grupos.

H_1 : Os três grupos diferem quanto as escores F (de autoritarismo).

Prova estatística: Como são três grupos sendo estudados, um teste para k amostras é adequado. A escala F (de autoritarismo) pode ser considerado uma medida pelo menos ordinal, tornado o teste de *Kruskal-Wallis* adequado.

Nível de significância: Sejam $\alpha = 0,05$ e $n = 14$ = número total de educadores testados, $n_1 = 5$ (professores), $n_2 = 5$ (professores/administradores) e $n_3 = 4$ (administradores).

Distribuição Amostral: Para $k = 3$ e n_j pequenos a tabela O dá a probabilidade associada com a ocorrência, sob H_0 , para valores tão grandes quanto um H observado.

Região de rejeição: A região de rejeição consiste de todos os valores de H tão grandes que a probabilidade associada com sua ocorrência sob H_0 , é igual ou menor que $\alpha = 0,05$.

Decisão: Os escores F são apresentados na tabela 6.2.

⁵ Apresentada em: ADORNO, T. W. et al. *The authoritarian personality*. New York, Harper, 1950.



Tabela 6.2 - Escores de autoritarismo de 3 grupos de educadores

Professores	Professores/Administradores	Administradores
96	82	115
128	124	149
83	132	166
61	135	147
101	109	

Se estes dados forem colocados em postos e estes postos ordenados de forma crescente então se terá a tabela 6.3. Estes postos são somados e os resultados (R_i) estão no final da tabela.

Tabela 6.3 - Postos de autoritarismo de 3 grupos de educadores

Professores	Professores/Administradores	Administradores
4	2	7
9	8	13
3	10	14
1	11	12
5	6	
$R_1 = 22$	$R_2 = 37$	$R_3 = 46$

Agora é possível, então, determinar o valor da estatística H:

$$H = \frac{12}{n(n+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(n+1) = \frac{12}{14(14+1)} \left[\frac{22^2}{5} + \frac{37^2}{5} + \frac{46^2}{4} \right] - 3(14+1) = 6,40$$

Observando a tabela O para os valores 5, 5 e 4, tem-se que $H \geq 6,4$, tem probabilidade de ocorrência, sob H_0 , de $p < 0,049$. Como este valor é menor que $\alpha = 0,05$, a decisão é rejeitar H_0 . Pode-se concluir, então, que os 3 grupos de educadores diferem quanto ao grau de autoritarismo.

6.2.3. Empates

Quando ocorrem empates entre dois ou mais escores, cada escore recebe a média dos postos que deveriam receber se não houvesse empate. Como o valor de H é afetado pelos empates, uma correção deve ser feita na expressão do cálculo de H, que consiste em dividi-la pelo fator:

$$1 - \frac{\sum T}{n^3 - n}, \text{ onde } T = t^3 - t \text{ (onde } t \text{ é o número de valores empatados) e } n = \sum n_j.$$

Deste modo, a expressão geral para o cálculo de H, com a correção para empates é dada por:

$$H = \frac{\frac{12}{n(n+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(n+1)}{1 - \frac{\sum T}{n^3 - n}}$$

O efeito da correção para empates é aumentar o valor de H e assim tornar o resultado mais significativo do que seria se a correção não fosse realizada. Em muitos casos esta correção é tão pequena que pode ser desprezada. Se não mais do que 25% das observações estiverem empatadas, a probabilidade associada com um H calculado sem correção para empates é raramente alterada em mais de 10% por cento do que quando calculada com a corrigida.



7. Medidas de correlação e seus testes de significância

Em muitas situações é necessário saber se dois conjuntos de dados estão de alguma forma relacionados e com que intensidade ocorre esta relação. Medidas destinadas a determinar o grau de relacionamento entre duas ou mais variáveis são denominadas medidas de correlação. Estas medidas são expressas através de um número, que geralmente varia no intervalo de -1 a 1 e são denominados de **coeficientes de correlação**.

7.1. O coeficiente de contingência: C

7.1.1. Função

O coeficiente de contingência C é uma medida de correlação entre dois conjuntos de atributos. É útil quando se dispõem apenas de dados apresentados em escala nominal em um ou nos dois conjuntos de atributos. Para determinar esta medida não é necessário dispor as variáveis em uma determinada maneira. Não importa quem seja linha e quem seja coluna, o valor obtido será o mesmo.

7.1.2. Método

Para calcular o coeficiente de contingência C os dados devem ser apresentados em uma de contingência como a ilustrada em 7.1. Os dados podem ser divididos em qualquer número de categorias, isto é, a tabela pode ser do tipo Kr, onde k = número de colunas e r = número de linhas.

Tabela 7.1 - Tabela de contingência para o cálculo do coeficiente C

	A ₁	A ₂	...	k _A	Total
B ₁	A ₁ B ₁	A ₂ B ₁	...	A _k B ₁	
B ₂	A ₁ B ₂	A ₂ B ₂	...	A _k B ₂	
...
B _r	A ₁ B _r	A ₂ B _r	...	A _k B _r	
Total			...		n

O coeficiente de contingência pode, então, ser obtido através da seguinte expressão:

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}}, \text{ onde } \chi^2 = \sum_{i=1}^r \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \text{ e o qui-quadrado é calculado, conforme já visto.}$$

Exemplo:

Considere-se os valores apresentados na prova do qui-quadrado para k amostras independentes, onde foi testado se os cursos universitários escolhidos pelos jovens de Elmtown dependiam das classes sociais a que estes jovens pertenciam. Aqui, se tem uma associação entre uma variável nominal (curso) e uma variável ordinal (classe social). Os dados são repetidos na tabela 7.2.

Tabela 7.2 - Frequência de matrícula de jovens de Elmtown de 5 classes sociais em 3 cursos

Curso	Classe				Total
	I e II	III	IV	V	
Preparatório para a universidade	23	40	16	2	81
Geral	11	75	107	14	207
Comercial	1	31	60	10	102
Total	35	146	183	26	390

O valor do χ^2 calculado para estes dados foi de 69,20.

O valor do coeficiente de contingência será então:



$$C = \sqrt{\frac{\chi^2}{n + \chi^2}} = \sqrt{\frac{69,2}{390 + 69,2}} = 0,39$$

Logo a correlação entre a escolha do curso de nível e a classe social entre os jovens de *Elmtown* é de 0,39.

7.1.3. A prova de significância do coeficiente de contingência

Uma vez observado uma correlação entre dois conjuntos de atributos em amostras, quer-se determinar se é plausível concluir pela associação desses mesmos atributos na população de onde foram retiradas as amostras.

Ao se testar a significância de uma medida de associação, está-se na realidade testando a hipótese de nulidade de que não existe correlação na população, isto é, que o valor da medida de associação observada poderia ter ocorrido aleatoriamente entre as amostras se as populações não apresentam correlação.

Para testar a hipótese de nulidade, determina-se a distribuição amostral da estatística, neste caso, a medida de associação, sob H_0 . Utiliza-se, então, uma prova estatística adequada para determinar, a um nível de significância pré-fixado, se o valor observado pela estatística considerada pode ter provavelmente ocorrido sob H_0 .

Embora, muitas estatísticas de associação possam ser determinadas por este método o coeficiente de contingência C , constitui um caso especial. Uma das razões por que não se pode utilizar a distribuição amostral de C para testar um determinado valor observado, reside na considerável complexidade matemática de tal procedimento. Outra razão é que no desenvolvimento do cálculo de C , já se calcula de forma intermediária uma estatística que constitui uma indicação simples e adequada da significância de C . Tal estatística é o χ^2 . Pode-se determinar se um valor de C difere significativamente de um valor causal simplesmente determinando se um valor de χ^2 é significativo.

Para qualquer tabela de contingência $k \times r$ pode-se determinar a significância do grau de associação pela estatística C , determinando a probabilidade de ocorrência, sob H_0 , de valores tão grandes quanto o valor observado de χ^2 , com $gl = (k - 1)(r - 1)$. Se essa probabilidade não supera α , pode-se rejeitar a hipótese de nulidade, àquele nível. A tabela do qui-quadrado. Se o qui-quadrado baseado nos valores amostrais é significativo, pode-se concluir que, na população, a associação entre os dois conjuntos é diferente de zero.

Exemplo:

No exemplo acima foi mostrado que o coeficiente de correlação C entre as variáveis: classe social e opção curricular é $C = 0,39$. Para chegar a este valor foi utilizado o valor $\chi^2 = 69,20$. É este valor que vai ser usado para testar a significância de C . Verificando uma tabela qui-quadrado vê-se que $\chi^2 \geq 69,20$ com $gl = (4 - 1)(3 - 1) = 6$ tem probabilidade de ocorrência, sob H_0 , inferior a 0,001. Pode-se, assim, rejeitar a hipótese de nulidade, ao nível de 0,001 e concluir que o estatus social e a opção curricular acusam relacionamento na população da qual o grupo de *Elmtown* constitui uma amostra. Isto é, conclui-se que $C = 0,39$ é significativamente diferente de zero.

7.1.4. Limitações do coeficiente de contingência

A grande aplicabilidade e a determinação relativamente fácil de C podem dar a entender que se trata de uma medida ideal de associação. Este não é o caso, no entanto, em razões das limitações desta estatística.

Em geral, pode-se dizer que os coeficientes de correlação devem apresentar pelo menos as seguintes características:

- Onde houver completa falta de associação o coeficiente deve dar zero.
- Quando as variáveis são completamente dependentes entre si, isto é, estão perfeitamente correlacionadas o coeficiente deve ser igual a 1.

O coeficiente C tem a primeira destas características, mas não a segunda. Ele é zero quando não existe associação, mas não atinge o valor um, quando a correlação é perfeita, sendo esta a primeira limitação do coeficiente de contingência C .



O limite superior de C é função do número de categorias. Quando $k = r$, o limite superior de C, isto é, o valor que deveria ocorrer se as variáveis tivessem correlação perfeita é:

Por exemplo, o limite superior de C para uma tabela 2×2 é $\sqrt{1/2} = 0,71$. Para uma tabela 3×3 , o máximo que C pode atingir é $\sqrt{2/3} = 0,82$. O fato de o valor máximo de C, depender de k e r constitui uma segunda limitação de C, pois dois coeficientes de contingência só serão comparáveis se provierem de tabelas com o mesmo número de linhas e colunas.

Uma terceira limitação de C é que os dados devem se prestar para o cálculo do χ^2 antes que C possa ser convenientemente utilizado, isto é, o cálculo de C sofre das mesmas limitações do cálculo do qui-quadrado.

Uma última limitação de C e que ele não é diretamente comparável com nenhuma outra medida de correlação, como por exemplo, o coeficiente de Pearson ou o de Spearman.

A despeito destas limitações o coeficiente de contingência é uma medida útil pela sua larga aplicabilidade, pois não exige suposições sobre a forma da população de escores, não exige continuidade da variável em estudo e requer apenas mensuração nominal. Isto faz do C uma medida que pode ser aplicada em situações em que nenhuma outra possa ser aplicada.

7.2. O coeficiente V de Cramer

Apesar de sua popularidade o coeficiente de contingência tem a desvantagem de que o número de linhas e colunas influencia o resultado. A alternativa é utilizar o coeficiente V (de Cramer), definido por:

$$V = \sqrt{\frac{\chi^2}{n \cdot (k - 1)}}, \text{ onde:}$$

n = tamanho da amostra e $k = \min \{\text{número de linhas, número de colunas}\}$.

Exemplo:

Consumo de drogas	Consumo de álcool			Total
	Alto	Moderado	Baixo	
Alto	5	7	20	32
Moderado	10	8	15	33
Baixo	15	6	14	35
Total	30	21	49	$n = 100$

Considerando a tabela acima que cruza o consumo de álcool com o consumo de drogas, determine o coeficiente V (de Cramer).

A tabela abaixo mostra os cálculos dos valores esperados:

Consumo de drogas	Consumo de álcool			Total
	Alto	Moderado	Baixo	
Alto	9,60	6,72	15,68	30,00
Moderado	9,90	6,93	16,17	33,00
Baixo	10,50	7,35	17,15	35,00
Total	30,00	21,00	49,00	100,00

O valor do χ^2 calculado para os valores da tabela acima será:

$$\chi^2 = (5 - 9,60)^2/9,60 + (7 - 6,72)^2/6,72 + (20 - 15,80)^2/15,80 + (10 - 9,90)^2/9,90 + (8 - 6,93)^2/6,93 + (15 - 16,67)^2/16,67 + (15 - 10,50)^2/10,50 + (6 - 7,35)^2/7,35 + (14 - 17,15)^2/17,15 = 6,41$$



7.3. O coeficiente de correlação de Postos de Spearman: r_s

7.3.1. Função

Dentre todas as estatísticas com base em postos, o coeficiente de correlação por postos de Spearman foi a que surgiu primeiro e é talvez a mais conhecida hoje. Esta estatística, por vezes designada “rho” (ρ), é representada, aqui por r_s . É uma medida de associação que exige que as duas variáveis tenham mensuração a nível pelo menos ordinal, para que se possa ordenar, isto é, determinar seus postos.

7.3.2. Fundamentos lógicos

Suponha-se que n indivíduos ordenados em postos segundo duas variáveis. Por exemplo, um grupo de estudantes ordenado de acordo com suas notas no vestibular de uma universidade e também de acordo com sua classificação escolar ao fim do primeiro ano. Denotando os escores do vestibular por:

X_1, X_2, \dots, X_n , e os escores da classificação escolar ao fim do primeiro ano por:

Y_1, Y_2, \dots, Y_n , pode-se utilizar uma medida de correlação por postos para determinar o relacionamento entre as duas variáveis.

A correlação entre a classificação no vestibular e a classificação ao fim do primeiro ano seria perfeita se e somente se $X_i = Y_i$ para todo “ i ”. Portanto, parece lógico usar as diversas diferenças: $d_i = X_i - Y_i$, como indicativo da diferença entre os dois conjuntos de postos. Suponha-se que o aluno A tenha obtido o primeiro lugar no vestibular, mas ao fim do primeiro ano esteja em 6º lugar. Neste caso, $d = 1 - 6 = -5$. Um aluno B, por outro lado, ficou em nono lugar no vestibular e agora, ao final do primeiro ano, é o segundo colocado. O valor de d para ele é então: $d = 9 - 2 = 7$. O valor das diversas diferenças “ d ” fornece uma idéia do relacionamento entre a classificação no vestibular e no fim do primeiro ano escolar. Se a relação entre os dois conjuntos de postos fosse perfeita, todos os valores de “ d ” seriam zero. Quanto maiores os diversos valores de “ d ”, menor será a associação entre as duas variáveis.

A utilização direta dos valores das diferenças (d) para o cálculo do coeficiente de correlação acarreta dificuldades. Por exemplo, os valores negativos se cancelam com os positivos se fosse somados para fornecer a diferença total. Por isso é utilizado o valor de d ao quadrado, d^2 , para eliminar esta dificuldade.

A obtenção da expressão para o cálculo do coeficiente de correlação de Spearman é baseada no cálculo do coeficiente de Pearson (estatística paramétrica) r , onde:

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}, \text{ onde } x = X - \bar{X} \text{ e } y = Y - \bar{Y}$$

Mas quando X e Y são postos, $r = r_s$, e a soma de n inteiros: $1, 2, \dots, n$ é:

$$\sum X = \frac{n(n+1)}{2} \text{ e a soma de seus quadrados, } 1^2, 2^2, \dots, n^2 \text{ é: } \sum X^2 = \frac{n(n+1)(2n+1)}{6}$$

$$\text{Como } \sum x = \sum (X - \bar{X}) = \sum X^2 - \frac{(\sum X)^2}{n} \text{ vem:}$$

$$\sum x^2 = \frac{n(n+1)(2n+1)}{6} - \frac{n^2(n+1)^2}{4} = \frac{n^3 - n}{12} \text{ e de forma análoga segue que } \sum y^2 = \frac{n^3 - n}{12}$$

Mas $d = x - y =$, então $d^2 = (x - y)^2 = x^2 + y^2 - 2xy$ e $\sum d^2 = \sum x^2 + \sum y^2 - 2\sum xy$.

Pela expressão do cálculo do coeficiente de correlação de Pearson, tem-se:

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}} = r_s, \text{ quando as observações são medidas em postos. Portanto}$$

$$\sum d^2 = \sum x^2 + \sum y^2 - 2\sum xy = \sum x^2 + \sum y^2 - 2r_s \sqrt{\sum x^2 \sum y^2} \text{ e assim:}$$



$$r_s = \frac{\sum x^2 + \sum y^2 - \sum d^2}{2\sqrt{\sum x^2 \sum y^2}}.$$

Substituindo $\sum x^2$ e $\sum y^2$ pelos seus valores em termos de postos e fazendo as simplificações necessárias vem:

$$r_s = 1 - \frac{6\sum d^2}{n^3 - n},$$

que é a expressão mais conveniente para o cálculo do coeficiente r_s de Spearman.

Exemplo:

Em um estudo sobre o efeito das pressões grupais sobre um indivíduo para uma atitude conformista em uma situação que envolva risco monetário, os pesquisadores⁶ aplicaram a 12 estudantes universitários a escala F (medida de autoritarismo) e uma escala destinada a medir as aspirações de status social. Desejava-se uma informação sobre a correlação entre os escores relativos ao autoritarismo e os escores referentes às aspirações de status social. (Tais aspirações foram definidas de acordo com os pontos de vista “O indivíduo não deve casar-se com pessoa de nível social inferior ao seu”, ou “Para um encontro, é melhor uma demonstração eqüestre do que um jogo de baseball”, ou ainda, “É interessante verificar sua genealogia”. A tabela 7.3 fornece os escores de cada um dos 12 estudantes nas duas escalas.

Tabela 7.3 - Escores de autoritarismo e aspiração de status social

Estudante	Escore	
	Autoritarismo	Aspiração
A	82	42
B	98	46
C	87	39
D	40	37
E	116	65
F	113	88
G	111	86
H	83	56
I	85	62
J	126	92
K	106	54
L	117	81

Para calcular o coeficiente de correlação por postos, de Spearman, para estes dois conjuntos de valores é necessário colocá-los, inicialmente em duas séries de postos. Estes postos são apresentados na tabela 7.4, juntamente com as diferenças entre eles e as diferenças ao quadrado.

Através destes dados então, pode-se calcular o coeficiente de correlação r_s , através da expressão mostrada acima. Assim:

$$r_s = 1 - \frac{6\sum d^2}{n^3 - n} = 1 - \frac{6.52}{12^3 - 12} = 0,82.$$

Tabela 7.4 - Postos referentes a autoritarismo e aspiração de status social

Estudante	Escore		d_i	d_i^2
	Autoritarismo (Posto)	Aspiração (Posto)		
A	2	3	-1	1
B	6	4	2	4
C	5	2	3	9
D	1	1	0	0
E	10	8	2	4
F	9	11	-2	4
G	8	10	-2	4
H	3	6	-3	9
I	4	7	-3	9
J	12	12	0	0
K	7	5	2	4
L	11	9	2	4
				$\sum d_i^2 = 52$

⁶ SIEGEL, S., FAGAN, Joen. *The Asch effect under conditions of risk*. Dados extraídos de um estudo piloto, não publicado.



7.3.3. Observações empatadas

Ocasionalmente podem ocorrer empates entre os escores de dois indivíduos na mesma variável. Quando isto ocorre, a cada um deles é atribuído a média dos postos que seriam atribuídos caso o empate não ocorresse, isto é, adota-se o procedimento usual.

Se a proporção de empates não é grande seu efeito sobre o coeficiente de correlação é desprezível. Quando a proporção de empates é grande torna-se necessário a utilização de um fator de correção.

O efeito de postos empatados na variável X, consiste em reduzir a soma dos quadrados. Portanto, quando houver empates em X é necessário corrigir a soma dos quadrados pelo fator:

$$T = \frac{t^3 - t}{12}, \text{ onde } t = \text{número de observações empatadas em determinado posto.}$$

A soma dos quadrados corrigida será então:

$\sum x^2 = \frac{n^3 - n}{12} - \sum T$, onde a soma de T, indica o somatório sobre os vários valores de T para todos os grupos de observações empatadas.

Assim se o número de empates for considerável o cálculo do coeficiente de correlação de Spearman deve ser realizado através de:

$$r_s = \frac{\sum x^2 + \sum y^2 - \sum d^2}{2\sqrt{\sum x^2 \sum y^2}}, \text{ onde } \sum x^2 = \frac{n^3 - n}{12} - \sum T_x \text{ e } \sum y^2 = \frac{n^3 - n}{12} - \sum T_y.$$

7.3.4. Teste de significância para o coeficiente de correlação de Spearman

Se as amostras utilizadas no cálculo do coeficiente de correlação de Spearman são selecionadas aleatoriamente, então pode-se utilizar os seus valores para testar se as variáveis correspondentes estão associadas na população, isto se r_s pode ser considerado diferente de zero.

Pequenas amostras. Suponha-se verdadeira a hipótese de nulidade, isto é, suponha-se que não exista relacionamento na população de onde foram extraídas as amostras. Se são extraídas uma amostras de escores X e uma de escores Y ao acaso desta população, então para uma dada ordem dos escores de X, todas as ordens possíveis dos escores Y tem a mesma probabilidade. Para n indivíduos existe n! ordenações possíveis dos escores X que podem ocorrer com qualquer ordenação dos escores Y. Como esses são igualmente prováveis, a probabilidade de ocorrência de determinada ordenação dos escores X conjuntamente com dada ordenação dos escores Y é $1/n!$.

A cada uma das possíveis ordenações de Y está associado um valor de r_s . A probabilidade de ocorrência, sob H_0 , de qualquer valor particular de r_s é assim, proporcional ao número de permutações que originam aquele valor.

Aplicando a fórmula do cálculo de r_s pode ver que:

Se $n = 2$, então r_s só pode assumir os valores -1 e +1. Cada um destes valores tem probabilidade 1/2.

Se $n = 3$, então os possíveis valores de r_s são -1, -1/2, +1/2 e +1. Cada um destes valores tem probabilidade de ocorrência, sob H_0 , respectivamente de: 1/6, 1/3, 1/3 e 1/6.

A tabela P (Siegel, pg. 315) fornece os valores críticos de r_s , obtidos por este método. Para n variando de 4 a 30, a tabela fornece o valor de r_s com probabilidade associada, sob H_0 , $p = 0,05$, e o valor de r_s com probabilidade associada, sob H_0 , $p = 0,01$. A tabela é unilateral.

Exemplo:

No exemplo anterior o coeficiente de correlação foi determinado como sendo igual a $r_s = 0,82$, para um valor de $n = 12$. Pela tabela P vê-se que um valor tão grande quanto este é significativo ao nível $p < 0,01$ (teste unilateral). Se poderia então rejeitar a hipótese ao nível de 1% de significância, concluindo que, na população estudada, o autoritarismo e as aspirações de estatus social estão associados.



Grandes amostras. Quando n é igual ou superior a 10, a significância de um valor obtido de r_s sob a hipótese de nulidade pode ser comprovado através de (Kendall, 1948):

$$t = r_s \sqrt{\frac{n-2}{1-r_s^2}}$$

Quer dizer, que para n grande, o valor de r_s , tem distribuição t com $gl = n - 2$.

7.4. O coeficiente de concordância de Kendall: W

7.4.1. Função

As medidas anteriores consideravam a correlação entre **dois** conjuntos de postos de n elementos. Agora será considerada uma medida de relação entre **vários** conjuntos de postos de n elementos. Quando se tem k conjuntos de postos pode-se determinar a associação entre eles utilizando o coeficiente de concordância de Kendall, W . Enquanto que r_s exprime o grau de associação entre duas variáveis transformadas em postos, W exprime o grau de associação dentre k destas variáveis. Tal medida pode ser especialmente útil em estudos de fidedignidade relativos a julgamentos ou testes e tem também aplicações no estudo de conglomerados de variáveis.

7.4.2. Fundamentos lógicos

Como solução do problema da determinação da concordância global entre “ k ” conjuntos de postos, poderia ser razoável determinar os r_s 's (ou r 's) entre todos os pares possíveis de postos e então calcular a média desses coeficientes para determinar a associação global. Se tal procedimento fosse adotado, seria necessário calcular $\binom{k}{2}$ coeficientes de correlação de postos e a menos que k seja pequeno, o processo se torna impraticável.

O cálculo de W é muito simples e W tem uma relação linear com o valor médio de r_s relativo a todos os grupos. Denotando por r_{sav} o valor médio dos coeficientes de correlação por postos de Spearman entre os $\binom{k}{2}$ pares possíveis de postos, Kendall mostrou que:

$$r_{sav} = (kW - 1) / (k - 1)$$

Outro processo consiste em imaginar como se apresentariam os dados caso não houvesse concordância alguma entre os conjuntos de postos, e em seguida, como se apresentariam se houvesse concordância perfeita. O coeficiente de concordância seria então um índice de divergência entre a concordância efetiva acusada pelos dados e a concordância máxima possível (perfeita). De modo aproximado, W é um coeficiente desta natureza.

Suponha-se que três chefes de pessoal sejam encarregados de entrevistar seis candidatos a emprego e de classificá-los em postos, separadamente, segundo a capacidade de cada um para preencher a vaga. A tabela 7.9 fornece os 3 conjuntos independentes de postos atribuídos pelos chefes X, Y, Z aos candidatos de “a” a “f”. A última linha da tabela dá as somas (R_j) dos postos atribuídos a cada candidato.

Tabela 7.6 - Postos atribuídos a seis candidatos a emprego por três chefes de pessoal.

	Candidato					
	a	b	c	d	e	f
Chefe X	1	6	3	2	5	4
Chefe Y	1	5	6	4	2	3
Chefe Z	6	3	2	5	4	1
R_j	8	14	11	11	11	8

Se todos os chefes de pessoal apresentassem *perfeita* concordância em seus julgamentos, isto é, se tivessem atribuído postos aos candidatos na mesma ordem, então um candidato teria recebido três postos 1 e assim sua soma de postos, R_j , seria: $1 + 1 + 1 = 3 = k$. O candidato que os chefes tivessem considerado em



segundo lugar receberia $R_j = 2 + 2 + 2 = 6 = 2k$. E o menos promissor dos candidatos teria: $R_j = 6 + 6 + 6 = 18 = nk$.

Na tabela acima, percebe-se que não houve concordância perfeita entre os três chefes, vê-se que o grau de concordância entre os k julgadores é refletido pelo grau de variância entre as n somas de postos.

7.4.3. Método

Para determinar W , determina-se a soma dos postos, R_j , em cada coluna de uma tabela $k \times n$. Em seguida, soma-se os R_j e divide-se a soma por "n", obtendo a média dos R_j . Cada R_j pode então ser expresso como um desvio a contar da média (quanto maior for este desvio, maior é a associação entre os k conjuntos de postos). Por fim, determina-se a soma dos quadrados desses desvios. Conhecidos estes valores, pode-se calcular W , como sendo:

$$W = \frac{s}{\frac{1}{12}k^2(n^3-n)}, \text{ onde}$$

$$s = \text{soma dos quadrados dos desvios observados a contar da média dos } R_j, \text{ isto é, } s = \sum \left(R_j - \frac{\sum R_j}{n} \right)^2$$

k = número de conjunto de postos, neste exemplo número de julgadores.

n = tamanho da amostra

$\frac{1}{12}k^2(n^3-n)$ = valor máximo da soma dos quadrados dos desvios, isto é, o valor de s que ocorreria no caso de concordância perfeita entre os k conjuntos de postos.

Para os valores da tabela 7.9 os totais de postos foram: 8, 14, 11, 11, 11 e 8. A média destes valores é 10,5. Para determinar s eleva-se ao quadrado o valor de cada desvio de destes valores em relação a média e soma-se, isto é:

$$s = (8 - 10,5)^2 + (14 - 10,5)^2 + (11 - 10,5)^2 + (11 - 10,5)^2 + (11 - 10,5)^2 + (8 - 10,5)^2 = 25,5$$

Conhecendo s , pode-se determinar W para os dados da tabela 7.9:

$$W = \frac{s}{\frac{1}{12}k^2(n^3-n)} = \frac{25,5}{\frac{1}{12}3^2(6^3-6)} = 0,16$$

$W = 0,16$ exprime o grau de concordância entre os três chefes ao atribuírem postos aos seis candidatos a emprego.

7.4.4. Empates

Quando ocorrem empates atribui-se a cada valor empatado a média dos postos que lhes caberia se não houvesse empates. É o tratamento usual que se dá aos escores empatados em postos. O efeito dos empates é reduzir o valor de W . Se a proporção de empates é pequena, o efeito pode ser desprezado. Se, no entanto, esta proporção for grande, deve-se utilizar uma correção que aumenta o valor de W . A correção utilizada é a mesma utilizada no coeficiente de correlação de Spearman:

$T = \frac{\sum(t^3-t)}{12}$, onde: t = número de valores empatados em um grupo em relação a um determinado posto.

Com a correção de empates a expressão para o cálculo de W fica:

$$W = \frac{s}{\frac{1}{12}k^2(n^3-n) - k \frac{\sum T}{T}}, \text{ onde: } \sum T \text{ indica somatório sobre todos os valores de } T \text{ para todos os } k$$

conjuntos de postos.



Exemplo (com empates):

A tabela 7.7 mostra a classificação de 10 objetos em relação as variáveis X, Y e Z.

Tabela 7.7 - Postos de 10 objetos em relação a três variáveis

Variável	Objeto									
	a	b	c	d	e	f	g	h	i	j
X	1	4,5	2	4,5	3	7,5	6	9	7,5	10
Y	2,5	1	2,5	4,5	4,5	8	9	6,5	10	6,5
Z	2	1	4,5	4,5	4,5	4,5	8	8	8	10
R _j	5,5	6,5	9	13,5	12	20	23	23,5	22,5	26,5

A média dos R_j é 16,5. Para obter s, somam-se os quadrados dos desvios de cada R_j em relação a média:

$$s = (5,5-16,5)^2 + (6,5-16,5)^2 + (9-16,5)^2 + \dots + (26,5-16,5)^2 = 591.$$

Como a proporção de empates nos postos é grande, deve-se introduzir a correção para empates no cálculo de W. Nos postos de X existem dois conjuntos de empates: 2 objetos acham-se empatados em 4,5 e dois em 7,5. Para os dois grupos, t = número de valores empatados em um dado posto = 2. Desta forma:

$$T_X = \frac{\sum(t^3 - t)}{12} = \frac{(2^3 - 2)}{12} + \frac{(2^3 - 3)}{12} = 1$$

Nos postos de Y, existem três conjuntos de empates e cada conjunto contém duas observações. Aqui t = 2 em cada caso e:

$$T_Y = \frac{\sum(t^3 - t)}{12} = \frac{(2^3 - 2)}{12} + \frac{(2^3 - 3)}{12} + \frac{(2^3 - 3)}{12} = 1,5$$

Nos postos de Z, existem dois conjuntos de empates. Um deles empatado em 4,5, consiste de 4 valores e t = 4. O outro, empatado no posto 8, consiste de 3 valores e t = 3. Assim:

$$T_Z = \frac{\sum(t^3 - t)}{12} = \frac{(4^3 - 2)}{12} + \frac{(3^3 - 3)}{12} = 7$$

Conhecidos os valores de T para os conjuntos de postos de X, Y, e Z, pode-se calcular W com a correção para empates:

$$W = \frac{s}{\frac{1}{12}k^2(n^3 - n) - k\sum T} = \frac{591}{\frac{1}{12}3^2(10^3 - 10) - 3 \cdot 9,5} = 0,83$$

Se os empates não tivessem sido considerados o valor de W seria 0,80 ao invés do 0,83 obtido.

7.4.4. Teste de significância para W

Pequenas amostras. Pode-se comprovar a significância de qualquer valor observado de W determinando a probabilidade associada à ocorrência, sob H₀, de um valor tão grande quanto o **s** que está associado. Se for determinado a distribuição amostral de **s** para todas as permutações nos n postos em todas as maneiras possíveis nos k conjuntos, ter-se-á (n!)^k conjuntos de postos possíveis. Fazendo uso destes postos pode-se comprovar a hipótese de nulidade, de que os k conjuntos de postos são independentes, obtendo desta distribuição a probabilidade associada à ocorrência, sob H₀, de um valor tão grande quanto um **s** observado.

Por este método é que foi determinada a distribuição de **s** sob H₀ e foram tabelados certos valores críticos. A tabela R (Siegel, pg. 317) fornece estes valores. Esta tabela é aplicável para k de 3 a 20 e n de 3 a 7. Se um valor observado de **s** é igual ou superior ao valor exibido na tabela R, para um dado nível de significância, então H₀ pode ser rejeitada àquele nível.

Por exemplo, viu-se que quando k = 3 chefes de pessoal classificaram n = 6 candidatos a emprego, a concordância dos julgamentos foi W = 0,16. A tabela R, indica que o **s** associado àquele valor W (s = 25,5) não



é significativo. Para que a associação fosse significativa ao nível de 0,05, o valor de s deveria ser no mínimo igual a 103,9.

Grandes amostras. Quando n é maior que 7, então a probabilidade associada à ocorrência sob H_0 , de qualquer valor tão grande quanto um W observado tem distribuição aproximadamente qui-quadrado com $gl = n - 1$. Neste caso a significância pode ser determinada através da tabela do qui-quadrado.

Note-se que:
$$\frac{s}{\frac{1}{12}k^2(n^3 - n)} = k(n - 1)W$$
 e, portanto, $\chi^2 = k(n - 1)W$

Pode-se, então, utilizar esta expressão para determinar a probabilidade associada à ocorrência sob H_0 , de qualquer valor tão grande quanto um W observado, que é muito mais simples de calcular.

7.4.5. Interpretação de W

Um valor elevado ou significativo de W pode ser interpretado como indicando que os observadores ou juizes estão aplicando essencialmente os mesmos padrões ao atribuírem postos aos n elementos em estudo. No entanto, isto não significa dizer que as ordenações feitas sejam *corretas*. Na realidade, elas podem ser todas incorretas em relação a algum critério externo. É possível que diversos julgadores concordem quanto a ordenação ou classificação de indivíduos porque todos empregam o mesmo critério “errado”. Em tal caso, um valor alto de W significaria concordância na escolha do critério errado.

7.5. Conclusão

Foram apresentadas três técnicas não-paramétricas para medir o grau de correlação entre variáveis amostrais. E para cada uma delas foi apresentado o respectivo teste de significância da associação observada.

Uma destas técnicas, o coeficiente de contingência, é especialmente aplicável quando os dados se apresentam em escala nominal. Isto é, se a mensuração é tão elementar que as classificações em jogo não se apresentam relacionadas dentro de qualquer conjunto e assim não podem ser ordenadas.

Se as variáveis em estudo forem mensuradas no mínimo em escala ordinal, pode-se ainda empregar o coeficiente de contingência, mas um método adequado de correlação *por postos* utilizará melhor as informações contidas nos dados, sendo, por isso, preferível.

Para o caso bivariado foi apresentado o coeficiente r_s de *Spearman*. Este coeficiente é simples de calcular e tem a vantagem de estar linearmente relacionado com o coeficiente de concordância W .

O coeficiente de concordância de *Kendall*, W , mede a extensão da associação entre vários (k) conjuntos de postos de N entidades. É útil para determinar a concordância entre diversos julgamentos a respeito de associação entre três ou mais variáveis. Tem aplicação especial como método-padrão de ordenação de elementos de acordo com o consenso, quando não se dispõe de uma ordem objetiva dos mesmos. A tabela 7.8 mostra uma matriz relacionando os coeficientes de correlação e a escala de medida apropriada para as variáveis X e Y .

Tabela 7.8 - Coeficientes de correlação e escalas de medidas para as variáveis X e Y .

		Variável X		
		Nominal	Ordinal	Intervalar/Razão
Variável Y	Nominal	(1) a. Phi (ϕ) b. Coeficiente C c. V de Kramer d. λ e λ_v	(4)	(6)
	Ordinal	(4) Biserial por postos	(2) a. Tetrachoric b. ρ de Spearman	(5)
	Intervalar/Razão	(6) Biserial por ponto	(5) Biserial	(3) r de Pearson



8. BIBLIOGRAFIA

- BLALOCK, Hubert M. Jr. *Social Statistics*. Tokio: McGraw-Hill Kogakusha, 1972, 583 p.
- GIBBONS, Jean Dickinson. *Nonparametric Statistical Inference*. New York: McGraw-Hill, 1971, 306 p.
- GIBRA, Isaac N. *Probability and Statistical Inference for Scientists and Engineers*. Englewood Cliffs (NJ): Prentice-Hall, 1973, p. 596.
- HINKLE, Dennis E., WIERSMA, William, JURIS, Stephen G. *Applied Statistics for the Behavioral Sciences*. Boston: Houghton Mifflin Co. 1988, p. 682.
- LEVIN, Jack *Estatística Aplicada a Ciências Humanas*. São Paulo: Harbra, 1985, 392 p.
- NOETHER, Gottfried E. *Introduction to Statistics: A Nonparametric Approach*. Boston: Houghton Mifflin, 1976, 292 p.
- ROSCOE, John T. *Fundamental Research Statistics for the Behavioral Sciences*. New York: Holt, Rinehart and Wiston, 1975, 483 p.
- SIEGEL, Sidney. *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill Book Company, Inc., 1956, 312 p.
- WELKOWITZ, Joan, EWEN, Robert B., COHEN, Jacob. *Introductory Statistics for the Behavioral Sciences*. Orlando (FL): Harcourt Brace Jovanovich, 1982, 269 p.