

**PROPOSTA DE UM TESTE
NÃO-PARAMÉTRICO DE SINAL COM
POSTOS PARA DADOS INDEPENDENTES DE
DUAS POPULAÇÕES E ESTUDO DA TEORIA
DO TESTE DE GALTON**

PAULO CÉSAR DE RESENDE ANDRADE

2002

PAULO CÉSAR DE RESENDE ANDRADE

**PROPOSTA DE UM TESTE
NÃO-PARAMÉTRICO DE SINAL COM POSTOS PARA DADOS
INDEPENDENTES DE DUAS POPULAÇÕES E ESTUDO DA TEORIA
DO TESTE DE GALTON**

Dissertação apresentada à Universidade Federal de Lavras,
como parte das exigências do Curso de Mestrado em
Agronomia, área de concentração: Estatística e
Experimentação Agropecuária, para obtenção do título de
“Mestre”.

Orientador
Prof. Lucas Monteiro Chaves

**LAVRAS
MINAS GERAIS – BRASIL
2002**

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca Central da UFLA**

Andrade, Paulo César de Resende

Proposta de um teste não-paramétrico de sinal com postos para dados independentes de duas populações e estudo da teoria do teste de Galton / Paulo César de Resende Andrade. -- Lavras : UFLA, 2002.

87 p. : il.

Orientador: Lucas Monteiro Chaves.

Dissertação (Mestrado) – UFLA.

Bibliografia.

1. Teste de sinal com postos. 2. Distribuição nula. 3. Distribuição exata. 4. Binomial. 5. Galton. 6. Simulação Monte Carlo. I. Universidade Federal de Lavras. II. Título.

CDD-519.282

-519.532

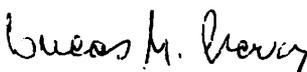
PAULO CÉSAR DE RESENDE ANDRADE

**PROPOSTA DE UM TESTE
NÃO-PARAMÉTRICO DE SINAL COM POSTOS PARA DADOS
INDEPENDENTES DE DUAS POPULAÇÕES E ESTUDO DA TEORIA
DO TESTE DE GALTON**

Dissertação apresentada à Universidade Federal de Lavras,
como parte das exigências do Curso de Mestrado em
Agronomia, área de concentração: Estatística e
Experimentação Agropecuária, para obtenção do título de
“Mestre”.

APROVADA em 28 de fevereiro de 2002

Prof. João Domingos Scalon	FUNREI
Prof. Daniel Furtado Ferreira	UFLA
Prof. Mário Javier Ferrua Vivanco	UFLA


Prof. Lucas Monteiro Chaves
(UFLA)
Orientador

**LAVRAS
MINAS GERAIS – BRASIL
2002**

OFEREÇO

A DEUS, pela vida e pela saúde

DEDICO,

À minha esposa Valéria

Aos meus filhos Guilherme e Júlia

Aos meus pais, Valter e Ana Maria

AGRADECIMENTOS

À Universidade Federal de Lavras e ao Departamento de Ciências Exatas (DEX), pelos conhecimentos.

Ao professor Lucas Monteiro Chaves, pela orientação, críticas, sugestões e, principalmente, pela amizade, compreensão e ensinamentos durante o curso.

Ao professor Daniel Furtado Ferreira, pelo profissionalismo, pela boa vontade, pela orientação, sugestões, críticas e ensinamentos.

Aos professores e funcionários do Departamento de Ciências Exatas da UFLA, pela amizade e ajuda, sem medir esforços.

Aos colegas de curso, Ceile, Douglas, Ednaldo, Flávio, José Marcelo, Lívia, Marcelo, Paulo José e Sérgio, pela amizade e solidariedade.

Aos amigos José Carlos, José Hortêncio, Sandro e Savinho, pela ajuda que prestaram.

Ao irmão Valter, pelo companheirismo, colaboração e amizade.

À irmã Adriana, aos sobrinhos Fabíola, Rafaela, Gabriela e Carlos, aos cunhados, aos sogros, pelo estímulo e carinho.

Aos meus pais Valter e Ana Maria, pelo amor, carinho, sacrifício e perseverança nos momentos difíceis e por tudo mais que puderam proporcionar.

Aos meus filhos Guilherme e Júlia, fontes de inspiração e incentivo para superar todas as dificuldades.

À minha esposa Valéria, pelo amor, pelo carinho, pela força, pelo estímulo e compreensão nos momentos de ausência.

A todos aqueles que direta ou indiretamente colaboraram e contribuíram para a realização deste trabalho.

SUMÁRIO

	Página
RESUMO.....	i
ABSTRACT.....	iii
1 INTRODUÇÃO.....	1
2 REFERENCIAL TEÓRICO.....	3
2.1 Erros tipo I, tipo II e poder do teste.....	3
2.2 Comparação de duas médias populacionais.....	7
2.3 Simulação e método de Monte Carlo.....	9
2.4 Estatística não-paramétrica.....	14
2.4.1 Testes para amostras dependentes (dados pareados).....	16
2.4.1.1 Teste dos sinais.....	18
2.4.1.2 Teste de sinais por postos.....	21
2.4.2 Testes para amostras independentes.....	25
2.4.2.1 Teste de Mann-Whitney.....	25
2.4.2.2 Teste da mediana.....	31
3 METODOLOGIA.....	34
3.1 Teste dos postos ordenados de Galton.....	34
3.2 Teste do sinal com postos para amostras independentes.....	35
3.3 Comparação dos testes.....	36
3.3.1 Taxas de erro tipo I.....	38
3.3.2 Poder dos testes.....	38
3.4 Exemplo de aplicação.....	41
4 RESULTADOS E DISCUSSÃO.....	42
4.1 Teste dos postos ordenados de Galton.....	42
4.2 Teste do sinal com postos.....	53
4.3 Simulação.....	65
4.4 Exemplo de aplicação.....	77
5 CONCLUSÕES.....	82
6 REFERÊNCIAS BIBLIOGRÁFICAS.....	83

RESUMO

ANDRADE, Paulo César de Resende. **Proposta de um teste não-paramétrico de sinal com postos para dados independentes de duas populações e estudo da teoria do teste de Galton.** Lavras, UFLA, 2002. 82p. (Dissertação-Mestrado em Agronomia/ Área de concentração em Estatística e Experimentação Agropecuária)*¹

O teste t de Student é o mais usado dentre os procedimentos empregados na comparação de duas amostras para inferir se existe diferença entre duas populações amostrais. Pressupõe-se que ambas as amostras vêm de populações normais com variâncias iguais. Se há desvios de normalidade e/ou desigualdade de variâncias, os testes não-paramétricos são recomendados. O presente trabalho tem como objetivos desenvolver a teoria de um teste do sinal na comparação de duas médias populacionais por meio de amostras independentes utilizando postos, avaliar sua performance sob diferentes condições quanto ao poder e taxas de erro tipo I, além de aplicá-lo a dados reais. Outro objetivo foi estudar a teoria do teste de Galton com objetivo específico de expor de forma didática seus aspectos combinatórios. Para desenvolver a teoria do teste proposto deve-se ordenar conjuntamente as duas amostras atribuindo postos e calcular a média dos postos de uma das amostras. A estatística S é definida como o número de variáveis da outra amostra que é menor ou igual a essa média. A distribuição exata de S, para valores pequenos de m e n, foi obtida computacionalmente exaurindo-se as permutações das amostras combinadas de ambas as populações. Foram computadas as médias, as variâncias e a proporção média esperada de sucessos. Apresentou-se uma tabela de valores críticos da estatística S para pequenos valores de m e n. A distribuição da estatística proposta apresentou boa aproximação à distribuição binomial com $p = 1/2$, conforme resultados obtidos da distribuição nula para vários valores de m e n e por simulação Monte Carlo. Para comparar o teste proposto com o de Mann-Whitney, Galton e t de Student, sob hipótese de igualdade de médias, foram realizadas 2000 simulações. A taxa de erro tipo I para cada teste foi computada estimando-se a proporção de análises que apresentaram significância para o nível nominal de 5%. Observa-se que os testes de Mann-Whitney e t não diferem muito entre si, oscilando em torno do valor de significância de 5%. O teste de Galton só apresenta significância a partir

* Comitê orientador: Lucas Monteiro Chaves – UFLA – (Orientador), Daniel Furtado Ferreira – UFLA – (Coorientador)

de amostras de tamanho 20. O teste de sinal com postos apresenta taxa de erro tipo I maior que o valor nominal de 5% até tamanho de amostras 35. A partir desse valor de m e n , as taxas são consideradas iguais à nominal, indicando que a aproximação binomial é adequada para essa situação. A comparação dos testes quanto ao poder foi realizada considerando diferenças específicas entre médias populacionais, fixadas parametricamente em função do erro padrão da diferença entre médias, em $k = 1/2, 1, 2$ e 4 erros padrão, tendo apresentado poder similares. Os testes apresentaram resultados bem próximos quando aplicados a dados reais. Chegou-se à conclusão que o teste de Galton controla as taxas de erro tipo I, sendo considerado rigoroso e mantém o poder elevado; deve ser recomendado para $n \geq 35$. O teste de sinal com postos apresenta controle da taxa de erro tipo I com amostras de mesmo tamanho e poder igual aos demais testes para $n \geq 40$, mas não deve ser recomendado para amostras de tamanhos diferentes.

SUMMARY

ANDRADE, Paulo César de Resende. **Proposal of a non-parametric test of signs with ranks for independent data and study of the theory of Galton's test.** Lavras, UFLA, 2002. 82p. (Dissertation-Master in Agronomy / major in Statistics and Agricultural Experimentation)²

The t test of Student is the most used one among the employed procedures when comparing two samples to conclude if there is any difference between two sample populations. It is presupposed that both samples come from normal populations with the same variances. If there are deviations of normality or inequality of variances, the non-parametric tests are recommended. The present work has as objective to develop the theory of a sign test when comparing two populations means through independent samples using ranks, to evaluate its power and type I error rates, besides applying it to real data. Another objective was to study the theory of Galton's test with the specific objective of exposing in a didactic way their combinatorial aspects. Developing the theory of the proposed test we should sort the two combined samples attributing their ranks and calculating the average of the ranks of one of the sample. The S statistics is defined as the number of variables of the other sample that is smaller or equal to that average. The exact distribution of S, for small values of m and n, was obtained computationally exhausting the permutations of the combined samples of both populations. The means, the variances and the expected proportion of successes were computed. A table of critical values of statistics S was shown for small values of m and n. The distribution of the statistics proposed presented a good approach to the binomial distribution with $p = 1/2$, according to the obtained results of the null distribution for several values of m and n and by Monte Carlo's simulation. Comparing the test proposed with the one of Mann-Whitney, Galton and t of Student, under hypothesis of equality of means, 2,000 simulations were done. The type I error rates for each test was computed being considered the proportion of analyses that presented significance for the nominal level of 5%. It is observed that the type I error rates do not differ a lot between the tests of Mann-Whitney and t, oscillating around the value of 5% nominal level of significance. Galton's test only presents significance starting from samples of size 20. The sign test with ranks presents type I error rates

² Guidance Committee: Lucas Monteiro Chaves – UFLA – (Major Professor), Daniel Furtado Ferreira – UFLA

pareados, ou seja, aqueles que são mensurados antes e após um tratamento ser empregado na mesma unidade amostral, ou ainda, aqueles em que dois tratamentos são aplicados em pares de unidades amostrais com grande uniformidade entre elas, os testes não-paramétricos empregados são o de Wilcoxon (1945, 1949) e o de sinais. Alternativas existem, considerando procedimentos de computação intensiva, como os testes de aleatorização e de Bootstrap, e ainda, os métodos Bayesianos.

O teste do sinal é uma alternativa também empregada nas inferências sobre médias, medianas e modas de uma população, e ainda, sobre tendências em uma seqüência temporal. O teste do sinal é um dos mais antigos testes estatísticos sendo citado há mais de 3 séculos (1710). Seu grande apelo é sua fácil implementação e tratabilidade, bem como seu amplo espectro de uso e sua flexibilidade para se adaptar a diferentes condições, uma vez que se baseia na distribuição binomial e poucas são as pressuposições requeridas ao seu uso. Apesar disso, nenhuma citação foi encontrada na literatura descrevendo o seu uso para o caso da comparação de médias a partir de duas amostras independentes, utilizando postos.

O objetivo desse trabalho foi de desenvolver a teoria para o emprego do teste do sinal para comparação de duas médias populacionais por meio de dados amostrais independentes, utilizando postos. Buscou-se avaliar a performance desse teste sob diferentes condições quanto ao poder e a taxas de erro tipo I em relação aos testes de Galton, Mann-Whitney e teste t de Student, além de buscar aplicar a teoria desenvolvida a dados reais. Também estudou-se a teoria do teste de Galton com objetivo específico de explorar e elucidar seus aspectos didáticos.

1 INTRODUÇÃO

A comparação de duas médias populacionais recebeu muita atenção pelos estatísticos nas décadas finais do milênio anterior (Behrens, 1929; Fisher, 1935; Jeffreys, 1940; Scheffé, 1943; Welch, 1947; McCullough et. al., 1960; Banerjee, 1961; Savage, 1961). Os testes para a hipótese de igualdade entre duas médias vêm sendo comparados em diversas situações. A razão fundamental dos diversos estudos realizados da viabilidade do emprego de diversas alternativas encontradas na literatura é a da violação das pressuposições requeridas pelos diferentes testes. O teste universalmente conhecido e utilizado é o teste paramétrico baseado na distribuição t de Student. Esse teste baseia-se nas pressuposições de independência e normalidade dos dados. Uma outra pressuposição refere-se à homogeneidade das variâncias das populações sob estudo. A violação dessa última exigência tem suscitado e motivado inúmeros trabalhos na literatura (Cochran, 1964; Cochran & Cox, 1957; Dixon & Massey, 1969; Fisher & Yates, 1963; Gill, 1971; Lee & Fineberg, 1991; Lee & Gurland, 1975; Satterthwaite, 1946), pois nesse caso a distribuição t de Student não é exata e o teste possui problemas quanto ao controle das taxas de erro tipo I e poder mais baixo, sendo conhecido como problema de Behrens-Fisher (Behrens, 1929; Fisher, 1939).

O teste t , quando os dados são qualitativos ordinais, não é aplicável, enquanto as alternativas não-paramétricas, nessas circunstâncias e em outras em que não há distribuição de probabilidade normal nos dados de cada amostra, são viáveis de serem empregadas. Para dados de amostras independentes podem-se citar os testes de Galton (citado por Feller, 1968) e de Mann-Whitney (1947). Já para dados

pareados, ou seja, aqueles que são mensurados antes e após um tratamento ser empregado na mesma unidade amostral, ou ainda, aqueles em que dois tratamentos são aplicados em pares de unidades amostrais com grande uniformidade entre elas, os testes não-paramétricos empregados são o de Wilcoxon (1945, 1949) e o de sinais. Alternativas existem, considerando procedimentos de computação intensiva, como os testes de aleatorização e de Bootstrap, e ainda, os métodos Bayesianos.

O teste do sinal é uma alternativa também empregada nas inferências sobre médias, medianas e modas de uma população, e ainda, sobre tendências em uma seqüência temporal. O teste do sinal é um dos mais antigos testes estatísticos sendo citado há mais de 3 séculos (1710). Seu grande apelo é sua fácil implementação e tratabilidade, bem como seu amplo espectro de uso e sua flexibilidade para se adaptar a diferentes condições, uma vez que se baseia na distribuição binomial e poucas são as pressuposições requeridas ao seu uso. Apesar disso, nenhuma citação foi encontrada na literatura descrevendo o seu uso para o caso da comparação de médias a partir de duas amostras independentes, utilizando postos.

O objetivo desse trabalho foi de desenvolver a teoria para o emprego do teste do sinal para comparação de duas médias populacionais por meio de dados amostrais independentes, utilizando postos. Buscou-se avaliar a performance desse teste sob diferentes condições quanto ao poder e a taxas de erro tipo I em relação aos testes de Galton, Mann-Whitney e teste t de Student, além de buscar aplicar a teoria desenvolvida a dados reais. Também estudou-se a teoria do teste de Galton com objetivo específico de explorar e elucidar seus aspectos didáticos.

2 REFERENCIAL TEÓRICO

2.1 Erros tipo I, tipo II e poder do teste

Quando se realiza um teste de hipótese, dois erros são possíveis, como apresentado por Mood et. al. (1974). Se a hipótese nula for rejeitada quando deveria ser aceita, diz-se que foi cometido erro tipo I. Se, por outro lado, for aceita uma hipótese que deveria ser rejeitada, diz-se que foi cometido erro tipo II. As probabilidades associadas a esses erros são denominadas α e β , respectivamente. A distinção entre os dois tipos fundamentais de erros estatísticos e o conceito de poder, vêm de um trabalho pioneiro de Neyman & Pearson (1928). O poder do teste é definido por esses mesmos autores como a probabilidade $(1 - \beta)$ de rejeitar a hipótese de nulidade quando ela é falsa.

Meyer (1984) comenta que não é possível evitar completamente esses erros, mas que se deve manter relativamente pequena a probabilidade de cometê-los, ou seja, os testes de hipóteses devem ser planejados de modo que os erros de decisão sejam reduzidos ao mínimo. Isso não é fácil, pois, para um dado tamanho de amostra, baixas probabilidades de se cometer um certo tipo de erro são associadas a altas probabilidades de se cometer outro tipo. Na prática, um tipo de erro pode ser mais importante que o outro, de modo que se deve procurar uma acomodação que favoreça a limitação do erro mais sério. Ambos podem ser reduzidos simultaneamente, mediante o aumento do tamanho da amostra, conforme Siegel (1975). De acordo com Zar (1999), para um dado α , amostras grandes resultarão em testes estatísticos com alto poder $(1 - \beta)$.

Tonhasca Jr. (1991), discutindo aspectos normalmente negligenciados no uso de metodologias estatísticas, diz que, estabelecida a hipótese alternativa, o maior problema do teste é calcular o valor de $1 - \beta$. Apesar disso, pesquisadores devem estar cientes das limitações impostas por experimentos com poucas repetições e analisados com os níveis usuais de α (5 e 1%), os quais freqüentemente fornecem baixos valores de $1 - \beta$, indicando que não haveria alta probabilidade de rejeitar a hipótese de nulidade quando esta hipótese é falsa. Ainda acrescenta que o poder de um teste depende do valor de α , da hipótese alternativa e do tamanho da amostra.

É possível encontrar, na literatura, estudos de alguns testes via simulação avaliando as taxas de erros empíricas, como o estudo sobre o erro tipo I e o poder de alguns testes de comparações múltiplas, feito por Perecin & Barbosa (1988). Utilizando o método de simulação de Monte Carlo, Silva et al.(1998) avaliaram as taxas de erro tipo I e o poder do teste de Scott-Knott e compararam os resultados com os testes de Tukey, t, Scheffé, Newman-Keuls modificado e t-bayesiano.

Borges & Ferreira (1999) buscaram avaliar o aumento de riscos de se tomarem decisões erradas (erro tipo I e tipo II) na comparação de duas médias populacionais. Para avaliar as probabilidades de se cometerem os erros, duas populações foram simuladas via método de Monte Carlo, no qual foi aplicado o teste t para verificar a diferença entre as médias amostrais. O teste t foi aplicado mediante as aproximações de Satterthwait (1946) e de Cochran & Cox (1957). Quando as variâncias populacionais eram iguais, as probabilidades de se cometer o erro tipo I foram praticamente as mesmas dos valores nominais estabelecidos. À medida que as diferenças entre as médias populacionais foram aumentadas, para o caso de variâncias

populacionais iguais, observou-se que a porcentagem de erro tipo II diminuiu consideravelmente, aumentando o poder do teste. Verificou-se também que, com o aumento do tamanho da amostra, houve uma redução nas probabilidades tanto do erro tipo I como do erro tipo II. As duas aproximações do teste t , usadas quando as populações apresentavam variâncias heterogêneas, não mostraram diferenças quanto à taxa de erro tipo I e tipo II empíricas.

Em seu trabalho, Leonel (2000) avalia um procedimento alternativo para comparação de tratamentos, nas análises em grupos de experimentos. Esse procedimento baseia-se na combinação das probabilidades dos contrastes de interesse de cada experimento. Os resultados obtidos por esses testes foram comparados com os resultados da análise de variância tradicional. Por meio do processo de simulação Monte Carlo, foi possível avaliar as taxas de erro tipo I e o poder do teste. Nas situações avaliadas, ambos os testes controlaram bem a taxa de erro tipo I, com os valores nominais de 5% e 1%. Quanto ao poder do teste verificou-se que o de significância coletiva apresenta baixo poder quando comparado com a análise de variância tradicional.

Santos (2000) propõe, em seu trabalho, uma modificação para o processo divisório do teste de Scott e Knott e novas alternativas de agrupamento. O comportamento do teste quanto às taxas de erro tipo I e o poder do teste, sob situações de hipótese de nulidade completa e parcial, foram avaliadas por meio de simulação Monte Carlo. Entre outras conclusões, o autor verifica que o teste de Scott Knott original controla adequadamente as taxas de erro tipo I, por comparação e por experimento, sob hipótese de nulidade completa, mantendo o poder elevado, mas não as controla adequadamente sob hipótese de nulidade parcial. O teste de Scott e Knott modificado apresenta os mesmos

resultados do original sob hipótese de nulidade completa, mas sob hipótese de nulidade parcial apresenta taxas de erro tipo I mais elevadas, embora tenha maior poder.

As características de um teste podem ser delineadas pelo conhecimento, para cada valor do parâmetro desconhecido θ , da probabilidade de rejeição da hipótese de nulidade. Uma tal função de θ é chamada função poder de um teste em questão, e simbolizada por $\pi(\theta)$. Ou seja, $1 - \pi(\theta)$ fornecerá a probabilidade de aceitação da hipótese de nulidade. Um teste ideal é aquele que rejeita a hipótese de nulidade com quase 100% de certeza quando ela é falsa, e a aceita igualmente com alta probabilidade somente quando θ está muito próximo de θ_0 .

Na Figura 1 apresenta-se a função poder do teste ($1 - \beta$) em função do parâmetro μ (medida de uma população normal) em um teste bilateral do tipo:

$$H_0: \mu = \mu_0$$

$$H_1: \mu \neq \mu_0$$

A análise da curva apresentada na Figura 1 permite verificar que: a) a altura da curva é a medida de seu poder e quanto mais elevada a curva em um dado ponto, maior é o poder do teste; b) a região de baixo poder está associada ao valor especificado em H_0 . Para que o poder aumente rapidamente, para valores de μ próximos de μ_0 deve-se aumentar o tamanho da amostra, que proporcionará um teste com maior poder e taxas de erro tipo I baixas.

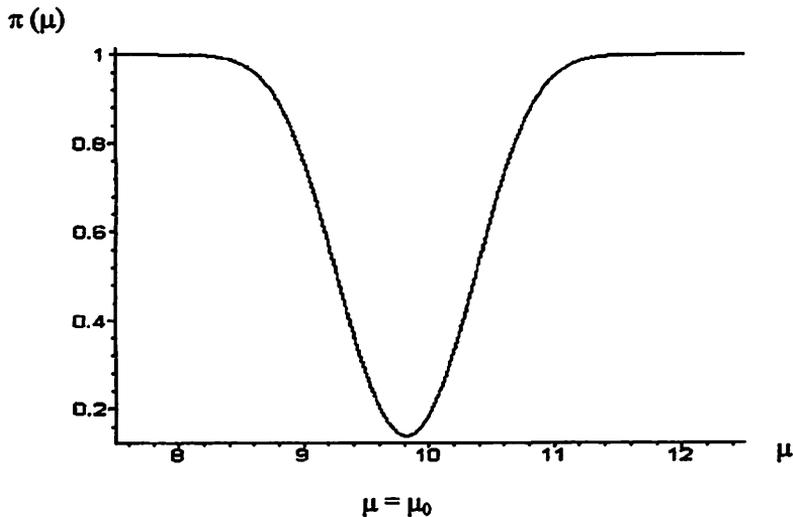


FIGURA 1. Curva do poder de um teste bilateral em que $\pi(\mu)$ é a função poder em relação a μ

2.2 Comparação de duas médias populacionais

Na prática é necessário, com muita frequência, tomar decisões a respeito das populações, com base nas informações das amostras. Para se tomarem decisões é conveniente a formulação de hipóteses. Essas hipóteses podem ser verdadeiras ou não. A tomada de decisão será então baseada no teste dessa hipótese.

Um dos principais testes, refere-se à necessidade de comparar duas médias populacionais distintas. Para comparar duas médias populacionais distintas, formulam-se as hipóteses de nulidade e alternativa:

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_1: \mu_1 - \mu_2 \neq 0 \text{ (bilateral);}$$

O teste de que a diferença entre as médias populacionais é igual a zero, equivale ao teste de que elas são iguais. Fixa-se o nível crítico, ou a probabilidade de se cometer o erro tipo I, que em geral é: $\alpha = 5\%$ ou $\alpha = 1\%$.

Para amostras da distribuição normal, a distribuição de amostragem das diferenças de médias é normal. Para amostras de população não-normais, as médias amostrais tendem a distribuir-se normalmente e as diferenças entre médias amostrais também tendem para uma distribuição normal à medida que n aumenta, conforme o teorema central do limite. Com essas propriedades, garante-se a aplicação do teste e determina-se o valor da estatística a ser utilizada em testes de hipóteses formuladas sobre a média de duas populações.

A diferença entre duas médias relaciona-se com a distribuição t . Sob H_0 de igualdade das médias, a estatística t_c , apresentada a seguir, distribui-se como t de Student com $v = m + n - 2$ graus de liberdade.

$$t_c = \frac{(\bar{X}_1 - \bar{X}_2)}{s_p \sqrt{\frac{1}{m} + \frac{1}{n}}}$$

em que:

\bar{X}_1 e \bar{X}_2 são as médias da amostra de tamanhos m e n das populações 1 e 2 e s_p é o desvio padrão combinado dado por:

$$s_p^2 = \frac{(m-1)s_1^2 + (n-1)s_2^2}{m+n-2}$$

em que:

s_1^2 e s_2^2 são os estimadores das variâncias das populações 1 e 2, respectivamente

Quando as variâncias populacionais são heterogêneas, a estatística t é apenas aproximada e é dada por

$$t_c = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}}$$

com v graus de liberdade de Satterthwaite (1946), dados por:

$$v = \frac{\left(\frac{s_1^2}{m} + \frac{s_2^2}{n}\right)^2}{\frac{\left(\frac{s_1^2}{m}\right)^2}{m-1} + \frac{\left(\frac{s_2^2}{n}\right)^2}{n-1}}$$

Convém salientar que nesse caso t_c tem distribuição apenas aproximada de t de Student. Estudos de simulação feitos por Borges & Ferreira (1996) demonstram que este teste mantém satisfatoriamente a probabilidade de se cometer o erro tipo I, mesmo para baixos tamanhos amostrais, mas o erro tipo II fica comprometido à medida que a magnitude da razão entre as variâncias populacionais aumenta.

2.3 Simulação e método de Monte Carlo

O uso da simulação de dados tem uma grande diversidade de áreas de aplicação. Na experimentação agrícola, é freqüente a instalação de ensaios em diversas áreas, sendo que, num estudo científico, esses

ensaios precisam ser repetidos sob as mais diversas condições para avaliar uma característica em estudo ou comparar algum método.

Segundo Dachs (1988), a idéia de construir modelos de simulação é muito antiga. As primeiras simulações feitas pelo homem talvez tenham sido na forma de maquetes para edificações e, mais tarde, em simulações de combates, usando pedras e galhos de árvores para representar batalhões e obstáculos. Possivelmente o avanço da ciência não teria alcançado o estágio em que a humanidade se encontra, ou pelo menos a teria por caminhos bastante diferentes, se o homem não tivesse a habilidade quase inata de criar modelos e simular o comportamento de fenômenos do mundo real por meio de sua utilização.

O mesmo autor define simulação como o processo para imitar o comportamento de um sistema real, para estudar seu funcionamento sob condições alternativas. Tornou-se cada vez mais freqüente o uso de métodos de simulação, para estudar novos procedimentos estatísticos ou para comparar o comportamento de diferentes técnicas estatísticas. O uso de métodos de simulação tem sido uma importante ferramenta para esse propósito. Com o crescente avanço do computador, é cada vez mais fácil e prática a simulação de variáveis ou de amostras baseadas em modelos estatísticos apropriados com parâmetros conhecidos, com o objetivo de se verificar a adequação de determinada metodologia ou na realização de comparações entre métodos.

Naylor et al. (1971) comentaram que o principal fundamento lógico para utilizar a simulação em qualquer assunto (seja em pesquisa operacional, científica, etc.) é o incessante desejo do homem de conhecer o futuro. Essa busca de conhecimento e o desejo de conhecer o futuro são tão velhos como a história da raça humana. Dizem ainda os autores que existem várias razões para usar a simulação em computadores, sendo

que uma delas considera que a simulação pode servir como uma primeira avaliação de um sistema para gerar novas políticas e regras de decisões antes de se correr o risco de experimentá-las no sistema real.

Para estudar ou avaliar um teste estatístico, muitas vezes torna-se bastante difícil obter, analiticamente, informações sobre o poder e taxas de erro tipo I de um teste. Uma maneira de se obterem as informações desejadas é mediante simulação. Na maioria dos trabalhos que envolvem simulação sempre está associado a ela o termo “Método de Monte Carlo”, que é muito utilizado para obter resultados de maneira mais simples.

O método de Monte Carlo pode ser definido como a representação da solução de um problema, utilizando para isso uma seqüência de números aleatórios com o objetivo de construir uma amostra da população. Assim, estimativas estatísticas de parâmetros representativos do problema podem ser obtidas.

Segundo Dachs (1988), o nome “Monte Carlo” é usado atualmente com freqüência crescente para referir-se ao uso de técnicas computacionais que geram amostras de acordo com determinadas distribuições teóricas conhecidas, visando a estudar novos comportamentos de diferentes técnicas estatísticas que poderiam ser empregadas num dado problema.

Para se ter uma idéia de simulação, é necessário definir o conceito de seqüência aleatória. Lehmer (1951) define que uma seqüência aleatória é um conceito vago, que engloba a idéia de uma seqüência em que cada termo é imprevisível. Na realidade, tais seqüências não são aleatórias no sentido restrito da palavra mas, para fins práticos, comportam-se como se fossem, comenta Dachs (1988). Essas seqüências são ditas pseudo-aleatórias.

O mesmo autor descreveu processos de geração de amostras aleatórias com base em uma distribuição uniforme (0, 1), utilizando a linguagem Pascal. Esse tipo de construção está baseado no teorema da probabilidade integral, que garante que é possível obter, a partir de uma distribuição uniforme (0, 1), uma amostra de qualquer outra distribuição. Esse teorema afirma que se U tem distribuição uniforme (0, 1); e se F é uma função de distribuição qualquer, a variável $X = F^{-1}(U)$ tem, distribuição F .

Analiticamente, tem-se que:

$$P(X \leq x) = P(F^{-1}(U) \leq x) = P(F[F^{-1}(U)] \leq F(x)) = P(U \leq F(x))$$

como U é uniforme e $F(x)$ está entre 0 e 1, então:

$$P(U \leq F(x)) = F(x)$$

de modo que:

$$P(X \leq x) = F(x),$$

ou seja, X tem função de distribuição F .

Por meio da Figura 2 é possível entender o teorema da probabilidade integral com maior facilidade. Repetindo-se o processo n vezes, pode-se gerar uma amostra aleatória com f.d.p. qualquer, a partir da distribuição uniforme (0, 1).

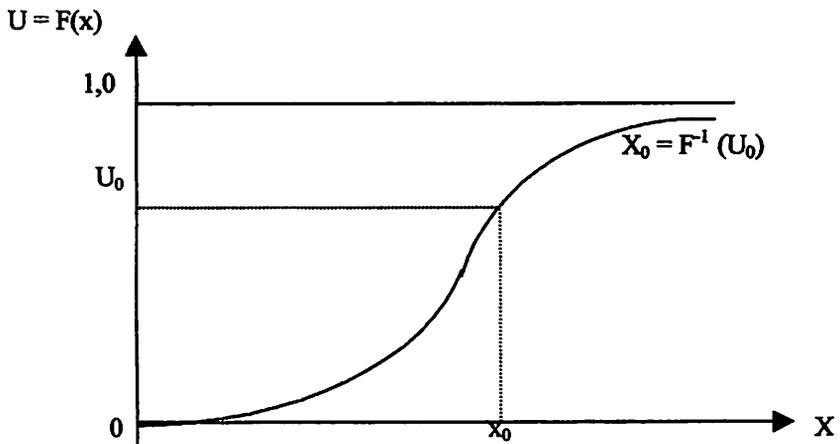


FIGURA 2. Processo ilustrativo da geração de um número aleatório (x_0) com uma função de distribuição $F(x)$ qualquer

O método de Monte Carlo, de uma maneira bastante simplificada, é um algoritmo que consiste em simular dados com base em uma seqüência pseudo-aleatória, obtida de uma distribuição uniforme (0, 1). Todo processo simulado que envolve um componente de qualquer distribuição é considerado como pertencente ao método de Monte Carlo. A única restrição para o uso desse método é a sua impraticabilidade para distribuições cuja função de distribuição seja desconhecida, ou cuja inversão não seja possível pela não-existência de algoritmos numéricos. Felizmente, na literatura, existem inúmeros algoritmos de inversão das funções de distribuições comumente usadas pelos estatísticos.

2.4 Estatística não-paramétrica

Os rudimentos da estatística não-paramétrica originaram-se na última parte do século XIX. Algumas contribuições iniciais são citadas em um trabalho de pesquisa de H. Scheffé (1943). Entretanto, somente com os trabalhos pioneiros de Wilcoxon (1945) e de Mann-Witney (1947) é que os fundamentos teóricos da estatística não-paramétrica começaram a ser montados.

Embora ainda apenas uma pequena parcela da extensa literatura estatística seja dedicada ao campo não-paramétrico, esse toma impulso dia-a-dia, constatando-se, atualmente, um crescimento muito grande de citações de aplicações de métodos não-paramétricos nos trabalhos de pesquisa. O desenvolvimento pós-guerra dos mesmos é descrito por Noether (1984).

Seu emprego, desde que não abusivo, constitui uma valiosíssima ferramenta de trabalho ao estatístico, por causa da sua grande diversidade e versatilidade. Pessoa (1977), diz que a abordagem dos métodos não-paramétricos tem sido utilizada como uma via didática de introdução dos conceitos básicos de Estatística, reduzindo bastante os pré-requisitos da Teoria da Probabilidade.

Zimmermann (1989) diz que eles são, na maior parte dos casos, muito facilmente trabalhados numericamente. Sua vantagem está no fato de que não se precisa ter, praticamente, nenhum conhecimento sobre a função de distribuição da população. São de distribuição livre, somente precisa-se estar certo de que as amostras aleatórias devem pertencer à mesma população (Walter, 1964; citado por Sachs, 1982). Esses procedimentos facilmente entendíveis podem ser aplicados a dados de postos e informações qualitativas.

Um teste livre de distribuição, quando aplicado a dados de uma família de distribuição conhecida, é sempre menos poderoso que o teste paramétrico correspondente (Pitman, 1949).

Um teste clássico de comparação de médias, o teste t de Student, pode ser aplicado sob as seguintes condições:

- 1- os dados devem ser independentes (amostras aleatórias);
- 2- a característica tem que ser mensurável em unidades de uma escala métrica;
- 3- as populações têm que ser (pelo menos aproximadamente) normalmente distribuídas;
- 4- as variâncias têm que ser iguais.

Campos (1983) define um teste não-paramétrico como aquele cujo modelo não especifica condições sobre os parâmetros da população da qual a amostra foi obtida. Mesmo quando existem certas pressuposições, estas são menos restritivas do que aquelas associadas aos testes paramétricos.

O mesmo autor apresenta as seguintes vantagens no uso dos métodos não-paramétricos:

- São menos exigentes do que os paramétricos, dispensando, por exemplo, a normalidade dos dados.
- As probabilidades das afirmativas obtidas da maioria dos testes não-paramétricos são, em geral, exatas, salvo quando se usam aproximações para grandes amostras.
- Independem da população em que foi obtida a amostra.
- São, em geral, de mais fácil aplicação, servem para pequenas amostras, são intuitivamente atraentes e exigem, quase sempre, menor quantidade de cálculos.

- Permitem trabalhar com diferentes populações, o que não é possível com os paramétricos.
- São aplicáveis a dados qualitativos.
- São mais eficientes do que os paramétricos, quando os dados não têm distribuição normal. E, quando a população é normalmente distribuída, sua eficiência, em alguns casos, é apenas levemente inferior.

O mesmo autor aponta algumas desvantagens:

- Em geral não levam em consideração a magnitude dos dados. É muito comum transformar os dados, de valores para simples ordens ou sinais. Em muitos casos isso se traduz numa perda de informações.
- Quando todas as exigências do modelo estatístico são satisfeitas, o teste paramétrico é mais poderoso. Para se obter a mesma eficiência, se faz necessário um maior tamanho de amostra para o não-paramétrico, Zar (1999).
- Há maior probabilidade de aceitar H_0 quando ela é falsa (erro tipo II).
- Em geral, não permite testar interações, salvo sob condições especiais sobre aditividade. Isso restringe o seu uso em modelos mais complicados.
- A obtenção, utilização e interpretação das tabelas são, em geral, mais complexas.

2.4.1 Testes para amostra dependentes (dados pareados)

A finalidade dos testes envolvendo amostras relacionadas é avaliar o efeito de algum “tratamento” numa variável de interesse. O

processo de avaliação acarreta a formação de dois grupos: um grupo de teste, que recebe o tratamento, e um grupo de controle, que não recebe o tratamento, ou que recebe um tratamento alternativo. O grupo de controle funciona como padrão de comparação para julgar a eficiência do tratamento. Para interpretar corretamente os resultados de tal comparação, é necessário emparelhar, tanto quanto possível, os dois grupos em termos de fatores importantes.

A necessidade de emparelhar os dois grupos pode causar problemas. Uma dificuldade freqüente é a capacidade de identificar corretamente fatores importantes como também a mensuração dessas variáveis, uma vez identificadas. Outra dificuldade é formar grupos de modo que tenham características emparelhadas, uma grande dispersão de dados poderá comprometer seriamente a validade das conclusões obtidas. A fim de contornar o problema, é recomendável, nesses casos, visando à redução da heterogeneidade, estruturar os dados aos pares, em que cada elemento receberá um dos tratamentos.

Independentemente da heterogeneidade dos dados, a própria natureza do problema poderá exigir o seu pareamento. O caso mais comum é usar cada indivíduo como seu próprio controle, constituindo o que denominamos de pré e pós tratamentos (Campos, 1983).

Dessa forma, todas as variáveis (ao menos em teoria) são emparelhadas, exceto a variável em estudo. Assim, fazem-se mensurações iniciais, os indivíduos se sujeitam a algum tratamento, e então obtêm-se um segundo conjunto de mensurações feitas sobre os mesmos indivíduos. Por exemplo, pode-se tomar o peso de um indivíduo antes de iniciar uma dieta, e após completá-la. A comparação dos dois conjuntos de mensurações dá informação quanto ao efeito do tratamento sobre os indivíduos.

Segundo Campos (1983) a aplicação do teste t às diferenças de resultados de cada par é possível de ser realizada. Entretanto, isso exige que sejam satisfeitas as exigências desse teste, o que nem sempre ocorre. Siegel (1975) cita alguns casos em que o teste t não é aplicável. Dentre eles:

- a) as exigências do teste não se coadunam com os dados da pesquisa;
- b) as diferenças entre os dados pareados são apenas ordinais, isto é, podemos apenas assinalar qual membro de cada par é maior do que, mas não podemos especificar o quanto é maior;
- c) os dados são qualitativos, e não quantitativos.

Nesses casos a aplicação de um adequado teste não-paramétrico é mais eficiente.

2.4.1.1 Teste dos Sinais

O teste do sinal é o mais antigo de todos os testes não-paramétricos (antecede 1710). Segundo Miller et al. (1999), o teste do sinal é usado como uma alternativa não-paramétrica ao teste t, para testar a hipótese de nulidade contra a hipótese alternativa.

O nome “Teste dos Sinais” se deve ao fato de o teste utilizar sinais “mais” ou “menos” em lugar de dados numéricos. Se existem dados numéricos, os valores são convertidos em sinais antes de iniciar o processo. Assim, desde que se possa estabelecer que o tratamento resultou em melhora (+), piora (-), ou não causou nenhuma modificação mensurável (0), o teste pode ser usado. Os dados do teste consistem em sinais “mais” ou “menos”, ignorando-se os zeros. A hipótese nula é a de que não há diferença entre os dois grupos, e a hipótese alternativa é a de que houve modificação. Ordinariamente usa-se um teste unilateral.

A questão é verificar se os testes sugerem mais do que uma variação simplesmente aleatória nas modificações observadas. Sugere-se a aleatoriedade quando os números de (+)'s e de (-)'s são aproximadamente iguais, sendo o tratamento eficiente se tiver muitos (+)'s e poucos (-)'s. O problema é conceitualmente idêntico a jogar uma moeda certo número de vezes e decidir se a moeda é honesta com base nos números de caras e coroas observados. E para grandes amostras, o modelo para essa situação pode ser a aproximação normal da binomial com $p = 1/2$.

Segundo Campos (1983), os dados consistem de n pares de observações (X_i, Y_i) , em que X_i representa uma situação pré e Y_i uma situação pós, ou então, (X_i, Y_i) são pareados de acordo com suas afinidades e objetivos da pesquisa.

Pressuposições:

- 1) os pares (X_i, Y_i) , $i = 1, 2, \dots, n$, são mutuamente independentes;
- 2) toma-se como modelo: $Z_i = \theta + e_i$ ($i = 1, 2, \dots, n$) em que θ é o efeito de tratamento e Z_i é a diferença $Y_i - X_i$;
- 3) os e_i são mutuamente independentes;
- 4) dentro de cada par, a escala de medida é ao menos ordinal (essa pressuposição torna o teste mais flexível e estende o seu campo de aplicação a dados qualitativos).

Método:

Preliminarmente define-se o indicador:

$$A_i = \begin{cases} 1 & \text{se } Y_i > X_i \text{ (} Z_i > 0 \text{)} \\ 0 & \text{se } Y_i < X_i \text{ (} Z_i < 0 \text{)} \end{cases}$$

Toma-se:

$$B = \sum_{i=1}^n A_i$$

onde B , sob $H_0: \theta = 0$, tem distribuição binomial com $p_0 = 1/2$.

Há um aspecto que deve merecer atenção: H_1 especifica tipicamente uma direção, que pode ser traduzida por um grande número de (+)'s ou (-)'s, ou por um pequeno número de (+)'s ou (-)'s. Quanto a testar os (+)'s, ou os (-)'s, é, na realidade, uma questão de pouca importância, pois a distribuição é simétrica, e o número de (+)'s e o de (-)'s somam n . Em outras palavras, obtém-se o mesmo resultado em qualquer dos dois casos.

Suponha que seja focalizado o número de (+)'s. Deve-se comparar o número observado de (+)'s com o número esperado e calcular um valor de B^* . Usando as fórmulas para o número de ocorrências, tem-se que o valor esperado e o desvio padrão do número de (+)'s, são, respectivamente:

$$\mu = np \qquad \sigma = \sqrt{np(1-p)}$$

Fazendo $p=1/2$:

$$\mu = \frac{n}{2} \quad e \quad \sigma = \frac{1}{2}\sqrt{n}$$

Determinação de B^* :

$$B^* = \frac{B - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}} = \frac{2B - n}{\sqrt{n}}$$

Assim, os testes de hipóteses ficam baseados na distribuição normal padrão e tem-se:

- a) $H_0: \theta = 0$ vs $H_1: \theta > 0$
 - rejeita-se H_0 se $B^* \geq z_{\alpha/2}$ onde $z_{\alpha/2}$ é o limite superior da distribuição normal, para o valor α de significância nominal.
- b) $H_0: \theta = 0$ vs $H_1: \theta < 0$
 - rejeita-se H_0 se $B^* \leq -z_{\alpha/2}$
- c) $H_0: \theta = 0$ vs $H_1: \theta \neq 0$
 - rejeita-se H_0 se $B^* \geq z_{\alpha/2}$ ou $B^* \leq -z_{\alpha/2}$

Quando n é 25 ou menos, pode-se utilizar tabelas binomiais para determinar a probabilidade exata.

2.4.1.2 O teste de sinais por postos

Esse teste foi introduzido por Wilcoxon (1945).

Quando os dados emparelhados não representam mensurações intervalares, há pouca alternativa, a não ser o uso do teste de sinais para avaliar o efeito de um tratamento. Todavia, quando dispomos de mensurações, a utilização do teste dos sinais pode levar a uma decisão incorreta. A razão é que o teste dos sinais perde informação, porque não leva em conta a magnitude da variação, mede apenas a direção da variação. Conseqüentemente, pode ocorrer uma situação em que os números de (+)'s e de (-)'s sejam aproximadamente iguais, mas as magnitudes de um tipo de variação sejam pequenas, ao passo que as do outro tipo são grandes. O teste do sinal poderia facilmente sugerir a aceitação de H_0 , quando a rejeição pareceria mais razoável.

Assim, segundo Miller et al. (1999), quando é possível determinar tanto a magnitude como a direção da variação para dados emparelhados, é mais vantajoso usar o teste de sinais por postos, porque esse teste perde menos informação. O teste focaliza diferenças por postos.

Conforme Campos (1983), consideram-se n pares de observações (X_i, Y_i) . Nesse caso, ao invés de se trabalhar com os “sinais” das diferenças Z_i , introduz-se o conceito de “rank” ou posto, e o teste se baseia exclusivamente nas ordens das diferenças positivas:

$$D_i = Y_i - X_i.$$

Pressuposições:

- 1) os pares (X_i, Y_i) são mutuamente independentes;
- 2) toma-se como modelo: $D_i = \theta + e_i$ ($i = 1, 2, \dots, N$);
- 3) cada D_i é uma variável contínua, com distribuição simétrica;
- 4) os D_i têm a mesma mediana.

Método:

Consideram-se:

- os valores absolutos das diferenças D_i ; ordena-se os valores D_i e atribui postos;
- se houver empate entre valores absolutos de D_i , toma-se para cada um deles, a média dos postos que seriam atribuídos caso não ocorresse o empate
- a variável aleatória A_i

$$A_i = \begin{cases} i & \text{se } Y_i > X_i \text{ (} D_i > 0 \text{)} \\ 0 & \text{se } Y_i < X_i \text{ (} D_i < 0 \text{)} \end{cases}$$

Define-se a estatística T:

$$T = \sum_{i=1}^N A_i$$

Observa-se que T representa a soma das ordens das diferenças positivas D_i . Segundo Zar (1999), o teste apresenta uma pressuposição de que as diferenças devam ser simétricas em relação à mediana.

Se a hipótese nula é verdadeira, é de esperar que os postos se repartam igualmente entre valores (+)'s e (-)'s e que as duas somas sejam aproximadamente iguais. O que deve ser determinado é se a soma de postos observada difere muito da soma esperada para ser apenas atribuível ao acaso.

A soma total de postos, quando se dispõem N objetos consecutivamente em postos, começando com 1 e terminando com N, é:

$$S = 1 + 2 + \dots + N = \frac{N(N+1)}{2}$$

Se H_0 é verdadeira, a soma esperada dos postos ((+)'s e (-)'s) (A_i), deve ser a metade do total :

$$E(A_i) = \frac{1}{2}i + \frac{1}{2}0 = \frac{i}{2}$$

$$E(T) = \frac{1}{2} \sum_{i=1}^N i = \frac{1}{2} \left[\frac{N(N+1)}{2} \right] = \frac{N(N+1)}{4}$$

$$V(A_i) = E(A_i^2) - [E(A_i)]^2 = \frac{1}{2}i^2 - \left(\frac{i}{2}\right)^2 = \frac{i^2}{4}$$

$$V(T) = \frac{1}{4} \sum_{i=1}^N i^2 = \frac{1}{4} \left(\frac{N(N+1)(2N+1)}{6} \right) = \frac{N(N+1)(2N+1)}{24}$$

O desvio-padrão é dado por:

$$\sigma = \sqrt{\frac{N(N+1)(2N+1)}{24}}$$

Determinação de Z:

$$Z = \frac{T - E(T)}{\sigma}$$

Assim, os testes de hipóteses ficam baseados na distribuição normal padrão e tem-se:

a) $H_0: \theta = 0$ vs $H_1: \theta > 0$

- rejeita-se H_0 se $Z \geq z_{\alpha/2}$ onde $z_{\alpha/2}$ é o limite superior da distribuição normal, para o valor α de significância nominal.

b) $H_0: \theta = 0$ vs $H_1: \theta < 0$

- rejeita-se H_0 se $Z \leq -z_{\alpha/2}$

c) $H_0: \theta = 0$ vs $H_1: \theta \neq 0$

- rejeita-se H_0 se $Z \geq z_{\alpha/2}$ ou $Z \leq -z_{\alpha/2}$

Algumas correções e aproximações foram efetuadas na estatística Z conforme Pratt (1959), Conover (1973), Iman (1974), Conover e Iman (1976).

Quando n é 30 ou menos, pode-se utilizar tabelas de valores críticos.

2.4.2 Testes para amostras independentes

A comparação de duas amostras independentes é feita por um teste conhecido como teste de Mann-Whitney (Mann & Whitney, 1947) e também como teste de Wilcoxon (1949). Formas equivalentes do mesmo teste apareceram na literatura sob diversos nomes, provavelmente por causa do apelo intuitivo do procedimento. Embora primariamente um teste para duas amostras, o teste de Mann-Whitney pode ser aplicado em muitas situações diferentes dessa.

Segundo Zimmermann (1989), a situação usual dos testes é aquela na qual o experimentador obteve duas amostras de possivelmente populações diferentes e deseja usar um teste estatístico para ver se a hipótese nula de que duas populações são idênticas pode ser rejeitada. Isto é, o experimentador deseja detectar diferenças entre as médias das duas populações com base nas amostras aleatórias das mesmas. Se os dados são do tipo ordinal, a diferença mais interessante é a média das duas populações.

2.4.2.1 Teste de Mann-Whitney

O teste de Wilcoxon-Mann-Whitney é usado para verificar se duas amostras independentes provêm de mesma população. Pode ser aplicado como alternativa ao teste t que exige que as populações tenham a mesma variância. Este teste não faz tal exigência. De fato, a única hipótese é que o nível de mensuração seja numa escala contínua, e mesmo essa hipótese não é absolutamente rígida. Apesar das pressuposições, o teste de Mann-Whitney é quase tão poderoso quanto o t .

O teste se baseia numa soma de postos. Dispõem-se os dados em postos, como se todas as observações fizessem parte de uma única amostra. Se H_0 é verdadeira, os postos baixos, médios e altos devem distribuir-se equilibradamente entre as duas amostras. Se H_0 é falsa, uma amostra tenderá a ter mais postos baixos (e, assim, uma soma de postos menor) enquanto a outra tenderá a ter maior soma de postos. Uma forma de analisar essa tendência é focalizar a soma de postos de uma das amostras e compará-la com a soma esperada de postos, supondo iguais médias. Pode-se testar qualquer dos dois conjuntos de postos. Todavia, é importante reconhecer que, se um conjunto de postos é mais alto do que se espera, o outro deve ser mais baixo que o esperado.

Pressuposições: Zimmermann (1989)

- 1) Ambas as amostras são amostras aleatórias das suas respectivas populações.
- 2) Em adição a independência dentro de cada amostra, existe independência mútua entre as duas amostras.
- 3) A escala de medida é pelo menos ordinal.

Método: Hollander & Wolfe (1973)

Sejam $X = \{x_1, x_2, \dots, x_m\}$ e $Y = \{y_1, y_2, \dots, y_n\}$ amostras independentes ($m \geq n$). Misturando as duas amostras, definem-se os postos, cuja soma é:

$$S = 1 + 2 + \dots + N = \frac{N(N+1)}{2}$$

em que $N = m + n$ (tamanho das amostras).

Considera-se a variável aleatória A_i baseada na amostra composta $X + Y$, da seguinte forma:

$$A_i = \begin{cases} i & \text{se o elemento} \in X \\ 0 & \text{se o elemento} \notin X \end{cases}$$

$$R = \sum_{i=1}^N A_i$$

Define-se a estatística R :

Observa-se que R representa a soma dos postos da amostra X .

Na posição i , a probabilidade de se ter um elemento da amostra X é m/N .

$$E(A_i) = \frac{m}{N} i$$

$$E(R) = \frac{m}{N} \sum_{i=1}^N i = \frac{m}{N} \left[\frac{N(N+1)}{2} \right] = \frac{m(N+1)}{2}$$

$$V(A_i) = E(A_i^2) - [E(A_i)]^2 = \frac{m}{N} i^2 - \left(\frac{m}{N} i \right)^2 = i^2 \left(\frac{m}{N} - \frac{m^2}{N^2} \right) =$$

$$V(A_i) = \left(\frac{mN - m^2}{N^2} \right) i^2 = \left(\frac{m(m+n) - m^2}{N^2} \right) i^2 = \left(\frac{m^2 + mn - m^2}{N^2} \right) i^2 = \frac{mn}{N^2} i^2$$

$$\begin{aligned}
\text{Cov}(A_i A_j) &= E(A_i A_j) - E(A_i)E(A_j) = \left(\frac{m(m-1)}{N(N-1)} \right) ij - \frac{m^2}{N^2} ij = \\
&= \left(\frac{m^2 - m}{N(N-1)} - \frac{m^2}{N^2} \right) ij = \left(\frac{N(m^2 - m) - m^2(N-1)}{N^2(N-1)} \right) ij = \\
&= \left(\frac{Nm^2 - Nm - Nm^2 + m^2}{N^2(N-1)} \right) ij = \left(\frac{-(m+n)m + m^2}{N^2(N-1)} \right) ij = \\
&= \left(\frac{-m^2 - mn + m^2}{N^2(N-1)} \right) ij = \left(\frac{-mn}{N^2(N-1)} \right) ij
\end{aligned}$$

$$\begin{aligned}
V(R) &= \sum V(A_i) + 2 \sum_{i < j} \text{Cov}(A_i A_j) = \frac{mn}{N^2} \sum_{i=1}^N i^2 + 2 \left(\frac{-mn}{N^2(N-1)} \right) \sum_{i < j} ij = \\
&= \frac{mn}{N^2} \left(\frac{N(N+1)(2N+1)}{6} \right) - \left(\frac{mn}{N^2(N-1)} \right) \left(\sum_{i,j=1}^N ij - \sum_{i=1}^N i^2 \right) = \\
&= \frac{mn}{N^2} \left(\frac{N(N+1)(2N+1)}{6} \right) - \left(\frac{mn}{N^2(N-1)} \right) \left(\sum_{i=1}^N \left[i \left(\sum_{j=1}^N j \right) \right] - \sum_{i=1}^N i^2 \right) = \\
&= \frac{mn}{N^2} \left(\frac{N(N+1)(2N+1)}{6} \right) - \left(\frac{mn}{N^2(N-1)} \right) \left(\left(\sum_{i=1}^N i \right) \left(\sum_{j=1}^N j \right) - \sum_{i=1}^N i^2 \right) =
\end{aligned}$$

O desvio padrão é dado por:

$$\sigma = \sqrt{\frac{mn(m+n+1)}{12}}$$

Determinação de Z:

$$Z = \frac{R - E(R)}{\sigma}$$

O teste de Mann-Whitney pode ser usado para testar $H_0: E(X_i) = E(Y_j) + d$, onde d é algum número especificado. Simplesmente adiciona-se o número d a cada Y_j e então usa-se o teste de Mann-Whitney sobre os originais X 's e os recém ajustados Y 's.

Assim, os testes de hipóteses ficam baseados na distribuição normal padrão e tem-se:

- a) $H_0: d = 0$ vs $H_1: d > 0$
 - rejeita-se H_0 se $Z \geq z_{\alpha/2}$ onde $z_{\alpha/2}$ é o limite superior da distribuição normal, para o valor α de significância nominal.
- b) $H_0: d = 0$ vs $H_1: d < 0$
 - rejeita-se H_0 se $Z \leq -z_{\alpha/2}$
- c) $H_0: d = 0$ vs $H_1: d \neq 0$
 - rejeita-se H_0 se $Z \geq z_{\alpha/2}$ ou $Z \leq -z_{\alpha/2}$

A aplicação dos testes unilaterais é recomendável quando, já a priori, espera-se um comportamento unidirecional de um dos tratamentos em relação ao outro. Na grande maioria dos casos não tem-se prévio

conhecimento de qual dos tratamentos é esperado ser melhor ou pior, e, conseqüentemente, deve-se aplicar o teste bilateral.

Segundo Mood (1954), o teste de Mann-Whitney é um dos testes não-paramétricos mais poderosos, entretanto, quando tanto o teste de Mann-Whitney, quanto o teste t for aplicável, o segundo tem maior poder. Quando as pressuposições do teste t são violadas, o teste de Mann-Whitney pode ser muito mais poderoso, segundo Hodges e Lecmann (1956).

Fix & Hodges (1955) descreveram uma aproximação para o teste de Mann-Whitney que é mais precisa do que a aproximação normal, mas requer esforço computacional. Hodges (1990) apresentou um método de aproximação que fornece bons resultados para probabilidades próximas de 0,1%.

Quando n é 20 ou menos, pode-se utilizar tabelas de valores críticos.

2.4.2.2 Teste da mediana

Sejam $X = \{x_1, x_2, \dots, x_m\}$ e $Y = \{y_1, y_2, \dots, y_n\}$ duas amostras independentes de tamanhos m e n , respectivamente, retiradas aleatoriamente de duas populações, objetos de estudo. Seja $W = \{w_{(1)}, w_{(2)}, \dots, w_{(N)}\}$ a amostra ordenada combinada formada pela junção dos conjuntos X e Y . Portanto $w_{(1)}$ é o menor valor, $w_{(2)}$ o segundo menor e assim por diante até $w_{(N)}$, que é o maior valor, com $N = m + n$. É fácil perceber que o índice i de um elemento de W corresponde ao posto de uma observação original que pode ser proveniente tanto de X quanto de Y . Dessa forma, é possível definir a

variável $I(X)_{(i)}$ para indicar se um elemento de W é proveniente do conjunto X por:

$$I(X)_{(i)} = \begin{cases} 1 & \text{se } w_{(i)} \in X \\ 0 & \text{se } w_{(i)} \notin X \end{cases}$$

Para apresentar a abordagem do teste do sinal, inicialmente, toma-se sem perda de generalidade $m \leq n$ e define-se o valor k por:

$$k = \begin{cases} \frac{m+n}{2} & \text{se } m+n \text{ for par} \\ \frac{m+n-1}{2} & \text{se } m+n \text{ for ímpar} \end{cases}$$

Definiu-se, ainda, a variável S como sendo igual ao número de valores de X menores ou iguais a $w_{(k)}$. Formalmente, definiu-se S , considerando a variável $I(X)_{(i)}$ apresentada anteriormente, por:

$$S = \sum_{i=1}^k I(X)_{(i)}$$

Assim, pode-se facilmente observar que a função de densidade de S , sob hipótese de nulidade, é dada pela hipergeométrica:

$$P(S=s) = \frac{\binom{m}{s} \binom{n}{k-s}}{\binom{m+n}{k}}, \quad s=0, 1, \dots, m$$

A média e a variância de S são dadas por:

$$\mu_s = \frac{m}{2} \quad \text{e} \quad \sigma_s^2 = \frac{mn}{4(m+n-1)}$$

Se $m = n$, logo $k = n$ e, portanto, S assumirá valores entre 0 e n (ou m), sendo que a função de densidade de S é dada por:

$$P(S=s) = \frac{\binom{n}{s}^2}{\binom{2n}{n}}$$

Nesse caso, a média e a variância de S são dadas por:

$$\mu_s = \frac{n}{2} \quad \text{e} \quad \sigma_s^2 = \frac{n^2}{4(2n-1)}$$

3 METODOLOGIA

Este trabalho foi realizado considerando duas partes, sendo que na primeira delas, em razão da complexidade dos pressupostos combinatórios e ausência de referências acessíveis, foi apresentado um estudo completo do teste de Galton. Na segunda parte, foi proposto o desenvolvimento da teoria de um novo teste do sinal para dados de amostras independentes, considerando os postos das observações. Foram realizadas simulações para comparar os testes quanto às taxas de erro tipo I e ao poder dos testes. Por fim, a teoria foi aplicada a um exemplo numérico.

3.1 Teste dos Postos Ordenados de Galton

O teste de Galton consiste em se tomar duas amostras aleatórias independentes $X = \{x_1, x_2, \dots, x_n\}$ e $Y = \{y_1, y_2, \dots, y_n\}$ de tamanho n , de duas populações. Seja $W = \{w_{(1)}, w_{(2)}, \dots, w_{(N)}\}$ a amostra ordenada combinada formada pela junção dos conjuntos X e Y , com $N = 2n$. Os valores observados são arranjados para cada amostra em ordem decrescente, sendo $a_1 > a_2 > \dots > a_n$ para a amostra X e $b_1 > b_2 > \dots > b_n$ para Y (a característica de interesse deve ser numérica).

Supondo que $X = \{x_1, x_2, \dots, x_n\}$ seja uma amostra de uma população controle e $Y = \{y_1, y_2, \dots, y_n\}$ uma amostra dessa mesma população após um tratamento. Deseja-se estudar a efetividade desse tratamento em relação à média. É de se esperar que se o tratamento for extremamente eficiente ter-se-á: $x_1 = b_1, \dots, x_n = b_n, x_{n+1} = a_1, \dots, x_{2n} = a_n$, ou seja, um tratamento extremamente bem sucedido pressupõe que todos os b 's precedem o primeiro a . Se o tratamento for inócua,

então a seqüência de a's e b's em (x_1, \dots, x_{2n}) deve ser aleatória. A eficiência do tratamento poderia ser avaliada através do número de b's que precedem os a's de mesmo posto, isto é, pelo número de k's tal que $b_k > a_k$.

Essa idéia foi usada por Galton, em 1876, em dados obtidos por Charles Darwin, segundo Hodges (1955). No caso n era igual a 15 e os b's lideravam 13 vezes. Galton concluiu ser o tratamento eficiente. No entanto, a resposta correta a essa questão depende do conhecimento da distribuição de probabilidade dos valores de k.

Apesar da simplicidade desse teste, a combinatória a ele subjacente é extremamente rica e com resultados surpreendentes, que em seu tempo enganou até a Galton.

Foi feito um estudo completo desse teste com objetivo didático e de referência.

3.2 Teste do sinal com postos para amostras independentes

Para apresentar o desenvolvimento do teste do sinal para dados de amostras independentes foi considerada uma abordagem que leva em consideração os postos das observações.

Sejam $X = \{x_1, x_2, \dots, x_m\}$ e $Y = \{y_1, y_2, \dots, y_n\}$ duas amostras independentes de tamanhos m e n, respectivamente, retiradas aleatoriamente de duas populações, objetos de estudo. Seja $W = \{w_{(1)}, w_{(2)}, \dots, w_{(N)}\}$ a amostra ordenada combinada formada pela junção dos conjuntos X e Y. Portanto $w_{(1)}$ é o menor valor, $w_{(2)}$ o segundo menor e assim por diante até $w_{(N)}$, que é o maior valor, com $N = m + n$. O índice i de um elemento de W corresponde ao posto de uma observação que pode ser proveniente tanto de X quanto de Y.

Inicialmente, serão definidas as variáveis $I(X)_{(i)}$ para indicar se um elemento de W é proveniente do conjunto X e $I(Y)_{(i)}$ para indicar se um elemento de W é proveniente do conjunto Y por:

$$I(X)_{(i)} = \begin{cases} 1 & \text{se } w_{(i)} \in X \\ 0 & \text{se } w_{(i)} \notin X \end{cases} \quad I(Y)_{(i)} = \begin{cases} 1 & \text{se } w_{(i)} \in Y \\ 0 & \text{se } w_{(i)} \notin Y \end{cases}$$

A distribuição exata de S , para valores pequenos de m e n , foi obtida computacionalmente exaurindo-se as permutações das amostras combinadas de ambas as populações. Foram computadas as médias e as variâncias e a proporção média esperada de sucessos, ou seja, da ocorrência de valores pertencentes a X na região determinada pela média de postos de Y (\bar{P}_Y).

Foi apresentada uma tabela de valores críticos da estatística S para pequenos valores de m e n e um estudo de uma aproximação para grandes amostras.

3.3 Comparação dos testes

Para comparar o teste proposto com o de Mann-Whitney, Galton e t de Student, sob hipótese de igualdade de médias, foram realizadas simulações computacionais. Foram feitas 2.000 simulações para cada combinação de tamanho de amostras e computadas as taxas de erro tipo I e poder dos testes.

O modelo estatístico adotado foi:

$$Y_{ij} = \mu + t_i + e_{ij},$$

com $i = 1, 2$ e $j = 1, 2, \dots, n_i$; em que $n_1 = m$ e $n_2 = n$;

- Y_{ij} é a resposta da variável simulada para a população i , na j -ésima repetição;
- μ é uma constante igual;
- t_i representa o efeito da população i ;
- e_{ij} é o erro aleatório, gerado para ter independência com distribuição normal ou exponencial.

Os valores do erro aleatório (e_{ij}) foram obtidos, por meio de simulação Monte Carlo, utilizando a função densidade de probabilidade da distribuição normal padronizada ou a função densidade de probabilidade da exponencial com $\lambda = 0,1$, representadas respectivamente por:

$$f(e) = \frac{1}{\sqrt{2\pi}} \exp\left(-e^2/2\right) \quad \text{e} \quad f(e) = 0,1 \exp(-0,1e)$$

A função de distribuição de probabilidade normal foi invertida para um valor uniforme, $U(0, 1)$, simulado por rotinas em Pascal apresentadas por Dachs (1988). Para a distribuição exponencial, a inversão da função de distribuição de probabilidade para um número U , uniforme entre 0 e 1, é:

$$e = \frac{-\ln(1-u)}{\lambda}$$

3.3.1 Taxas de erro tipo I

Para se avaliarem as taxas de erro tipo I, $\mu + t_i = \mu_i$ foi simulado como sendo igual a zero para $i = 1$ e para $i = 2$.

A taxa de erro tipo I para cada teste foi computada estimando-se a proporção de análises que apresentaram significância para o nível nominal de 5%. Foram feitas 2.000 simulações para diversos tamanhos de amostras (iguais e diferentes) da distribuição normal e exponencial.

Foram utilizadas as estatísticas representativas de cada teste para avaliar a significância dos testes para a hipótese H_0 de igualdade de médias.

Para comparar duas médias populacionais distintas foram testadas as hipóteses de nulidade e alternativa apresentadas a seguir:

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_1: \mu_1 - \mu_2 \neq 0 \text{ (bilateral);}$$

Foi fixado o valor crítico de significância em 5%, ou seja, $\alpha = 5\%$. O teste t foi realizado conforme descrição apresentada por Bussab & Morettin (1987).

3.3.2 Poder dos testes

Para comparar os testes quanto ao poder foram realizadas as mesmas situações descritas no item anterior, com exceção da forma de gerar os efeitos da média (μ_i). Para que os efeitos de médias fossem especificados, foram consideradas sempre diferenças específicas entre médias populacionais. Os valores dessas diferenças foram fixados

parametricamente em função do erro padrão da diferença entre médias, considerando-se diferenças dadas por

$$k \sigma_{(\bar{y}_1 - \bar{y}_2)}$$

sendo que k foi fixado em 1/2, 1, 2 e 4.

Para implementar esse estudo foi utilizado:

a) distribuição normal

O erro padrão da diferença entre médias é dado por:

$$\sigma_{(\bar{y}_1 - \bar{y}_2)} = \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}} = \sqrt{\sigma^2 \left(\frac{1}{m} + \frac{1}{n} \right)}$$

Foi considerado $\sigma^2 = 1$, sem perda de generalidade. Para a determinação da diferença de duas médias de k erros padrão, tem-se:

$$\mu_1 = 0 \quad \text{e} \quad \mu_2 = 0 + k \cdot \sqrt{\frac{1}{m} + \frac{1}{n}}$$

Portanto, a diferença de duas médias de k erros padrão da diferença entre médias é dado por:

$$\mu_1 - \mu_2 = k \cdot \sqrt{\frac{1}{m} + \frac{1}{n}}$$

b) distribuição exponencial

A média e a variância de uma distribuição exponencial são dadas, respectivamente por $1/\lambda$ e $1/\lambda^2$. Devido a média e a variância da distribuição exponencial serem relacionadas foi considerado que os valores de μ e de τ_i eram iguais a zero, com $i = 1$ ou 2 . Considerando uma diferença das médias de duas populações e do parâmetro da exponencial da primeira população dados por:

$$d = \mu_1 - \mu_2 = k \sigma_{(\bar{y}_1 - \bar{y}_2)} \text{ e } \lambda_1 = 0,1$$

$$\mu_1 = 10 \text{ e } \sigma_1^2 = 100$$

tem-se então:

$$\mu_2 = 10 + d \text{ e } \sigma_2^2 = (10 + d)^2$$

Assim,

$$\lambda_2 = \frac{1}{10 + d}$$

Portanto, k erros padrão da diferença entre médias é:

$$d = k \cdot \sqrt{\frac{100}{m} + \frac{(10 + d)^2}{n}}$$

Resolvendo a equação, tem-se d em função de m , n e k , assim representada:

$$d = 10 \cdot \left(\frac{1 + \sqrt{\frac{n}{m} [m + k \cdot (n - 1)]}}{(n - 1)} \right)$$

O número de análises que apresentaram significância, considerando o valor nominal $\alpha = 5\%$, para cada teste foi computado. A proporção de significância, em relação ao número total de simulações é uma medida empírica de Monte Carlo do poder.

3.4 Exemplo de aplicação

A teoria desenvolvida do teste do sinal com postos para comparação de duas médias populacionais por meio de dados amostrais independentes foi aplicada a um exemplo real. Foram utilizadas amostras retiradas de duas cultivares de batata de dois experimentos instalados na Universidade Federal de Lavras, em 1993 e 1994, respectivamente. São dados referenciais do artigo publicado por Bearzoti & Pinto (1999), na Revista Ciência e Agrotecnologia.

Foram selecionadas duas amostras de tamanho 25, de duas cultivares de batata, Achat e Monalisa, e aplicados os testes de Mann-Whitney, Galton, t de Student e sinal com postos, para verificação de significância dos mesmos.

4 RESULTADOS E DISCUSSÃO

4.1 Teste dos postos ordenados de Galton

Conforme exposto na metodologia, foi feito um estudo completo desse teste, ressaltando os aspectos combinatórios e utilizando uma abordagem geométrica.

A seqüência combinada (x_1, \dots, x_{2n}) foi transformada em uma seqüência de $(+1)$'s e (-1) 's da seguinte forma: se $x_i \in X$ será substituído por (-1) e caso $x_i \in Y$, por $(+1)$. Tem-se então que os resultados possíveis são todas as seqüências com n $(+1)$'s e n (-1) 's. Para calcular o número de tais seqüências basta supor a seqüência como $2n$ caixas vazias em que tem-se que escolher n destas caixas para colocar os $(+1)$'s, ficando assim, a posição dos (-1) 's determinada. Portanto tem-se um número de seqüências igual a:

$$\binom{2n}{n} = \frac{(2n)!}{n!n!}$$

Define-se então uma distribuição de probabilidade discreta $S = 0, 1, \dots, n$, onde:

$$P\{S = i\} = \frac{\text{número de seqüências tais que } b_k > a_k \text{ } i \text{ vezes}}{\binom{2n}{n}}$$

Dada uma $2n$ -upla de $(+1)$'s e (-1) 's, o objetivo é obter o número de índices k tais que $b_k > a_k$.

Exemplo: $n = 5$ e a seqüência $(-1, 1, 1, 1, -1, 1, -1, -1, 1, -1)$.

Neste caso, $S = 4$. Como detectar tal fato diretamente da seqüência? Seja S_j a soma dos j -ésimos primeiros elementos da seqüência: $S_j = x_1 + x_2 + \dots + x_j$. Para a seqüência dada, tem-se: $S_1 = -1$, $S_2 = 0$, $S_3 = 1$, $S_4 = 2$, $S_5 = 1$, $S_6 = 2$, $S_7 = 1$, $S_8 = 0$, $S_9 = 1$, $S_{10} = 0$.

Identifica-se a seqüência com um caminho, no qual cada $(+1)$ corresponde a um segmento para a direita e para cima e um (-1) um segmento para a direita e para baixo, conforme indicado na Figura 3.

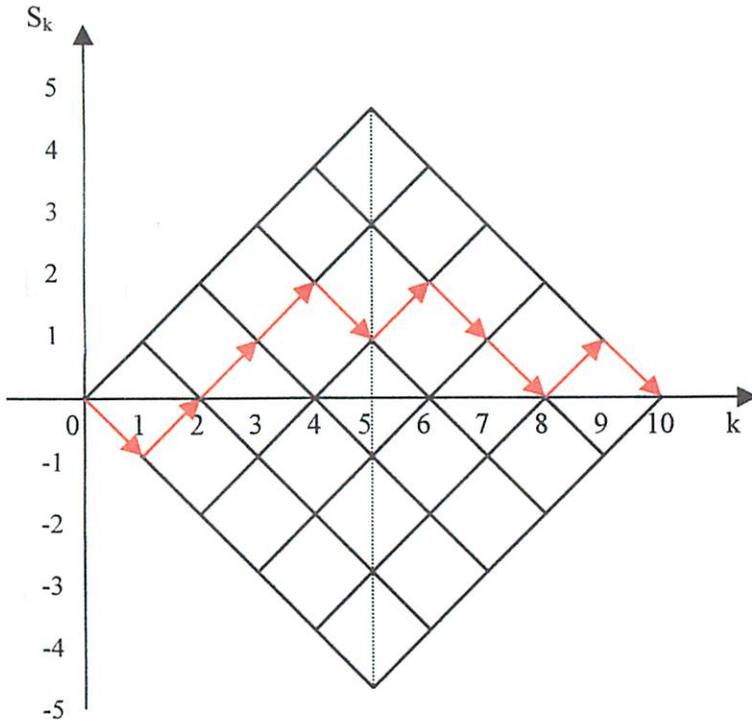


FIGURA 3. Representação da seqüência $(-1, 1, 1, 1, -1, 1, -1, -1, 1, -1)$

Se $S_{2k-1} > 0$ então tem-se que dos primeiros $2k - 1$ elementos da seqüência, pelo menos k deles é $(+1)$ e no máximo $k - 1$ é (-1) , o que implica que o $(+1)$ corresponde a b_k e o (-1) corresponde ao a_k , de onde segue que $b_k > a_k$. Vale a inversa, se $b_k > a_k$, então $S_{2k-1} > 0$,

isto é, $b_k > a_k$ se e somente se $S_{2k-1} > 0$. O fato de $S_{2k-1} > 0$ implica $S_{2k} \geq 0$. Então a desigualdade $b_k > a_k$ ocorre i vezes se e somente se tem i pares $(2k - 1, 2k)$ tais que $S_{2k-1} > 0$ e $S_{2k} \geq 0$. No exemplo anterior têm-se os pares $(3, 4)$, $(5, 6)$, $(7, 8)$ e $(9, 10)$. O problema assim colocado é de difícil resolução e bastante técnico. É proposto por Feller (1968) uma abordagem geométrica extremamente didática e que possibilita um ataque unificado a vários problemas combinatórios como o exposto.

Nota-se que as coordenadas de um ponto sobre o caminho é (k, S_k) .

As seqüências de $n (+1)$'s e $n (-1)$'s podem ser interpretadas como todos os caminhos de $(0, 0)$ até $(2n, 0)$. Pode-se agora interpretar que seqüências em que $b_k > a_k$ i vezes é equivalente aos caminhos com $2i$ lados acima do eixo Ox .

Na contagem do número de caminhos tem-se o seguinte fato fundamental, um teorema: (O Princípio da Reflexão)

“O número de caminhos de A até B que tocam ou cruzam o eixo Ox é igual ao número de caminhos que vão de A' até B , sendo A' a imagem refletida de A em relação ao eixo Ox ”, conforme Figura 4.

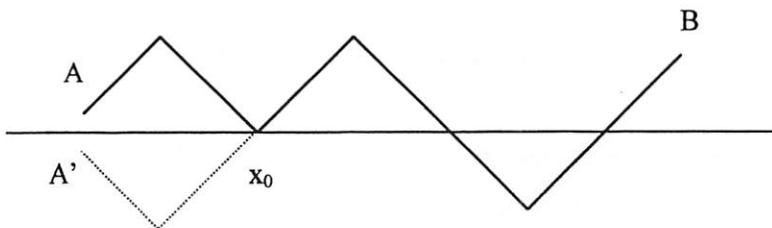


FIGURA 4. Princípio da Reflexão

Demonstração:

Seja α um caminho de A até B e seja x_0 o primeiro ponto do eixo Ox pertencente a α . Refletindo o caminho de A até x_0 em relação ao eixo Ox obtém-se um caminho α' de A' até x_0 . Após x_0 , α e α' são iguais. Da correspondência $\alpha \rightarrow \alpha'$ o resultado segue.

Um lema necessário é:

Lema I - O número de caminhos de um ponto (m, x) até (n, y) , $n > m$ é dado por:

$$\binom{n - m}{\frac{n - m + y - x}{2}}$$

Demonstração:

Como S_j = soma dos j primeiros elementos do caminho tem-se que $S_m = x$ e $S_n = y$. Sejam:

p_m = número de $(+1)$'s do caminho até o m -ésimo elemento

q_m = número de (-1) 's do caminho até o m -ésimo elemento

p_n = número de $(+1)$'s do caminho até o n -ésimo elemento

q_n = número de (-1) 's do caminho até o n -ésimo elemento

Portanto $p_n - p_m$ = número de $(+1)$'s do caminho do $(m+1)$ -ésima posição até a n -ésima posição. Idem para $q_n - q_m$.

Para determinar um caminho de (m, x) até (n, y) basta distribuir $p_n - p_m$ $(+1)$'s em $n - m$ posições e pode-se fazer isso com um número de maneiras dadas por:

$$\binom{n - m}{p_n - p_m}$$

Tem-se que:

$$\begin{cases} p_m + q_m = m \\ p_n + q_n = n \\ p_m - q_m = x \\ p_n - q_n = y \end{cases} \Rightarrow \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & -1 & 0 \end{bmatrix} \cdot \begin{bmatrix} p_n \\ p_m \\ q_n \\ q_m \end{bmatrix} = \begin{bmatrix} m \\ n \\ x \\ y \end{bmatrix} \Rightarrow \begin{cases} p_m = \frac{x + m}{2} \\ p_n = \frac{y + n}{2} \\ q_m = \frac{m - x}{2} \\ q_n = \frac{n - y}{2} \end{cases}$$

De onde segue que:

$$\binom{n - m}{p_n - p_m} = \binom{n - m}{\frac{n + y}{2} - \frac{m + x}{2}} = \binom{n - m}{\frac{n - m + y - x}{2}}$$

Obs: O número de caminhos de $(0, 0)$ até (n, y) é:

$$\binom{n}{\frac{n + y}{2}}$$

Teorema, Feller (1968): O número de caminhos de $(0, 0)$ até $(2n, 0)$ com exatamente $2i$ lados acima do eixo Ox é independente de i e é igual a:

$$\frac{\binom{2n}{n}}{n + 1}$$

Demonstração:

Consideram-se os casos $i = 0$ e $i = n$ separadamente.

Para $i = n$, precisa-se contar os caminhos de $(0, 0)$ até $(2n, 0)$ que estão acima do eixo Ox , isto é, caminhos que não tocam na reta $y = -1$. Na Figura 5, apresenta-se um desses caminhos.

Para contar estes caminhos, faz-se uso do princípio da reflexão, prolongando à esquerda o triângulo, de forma que o ponto $(0, 0)$ se torne o ponto $(1, 1)$ e a reta $y = 0$ a reta $y = 1$, conforme Figura 6.

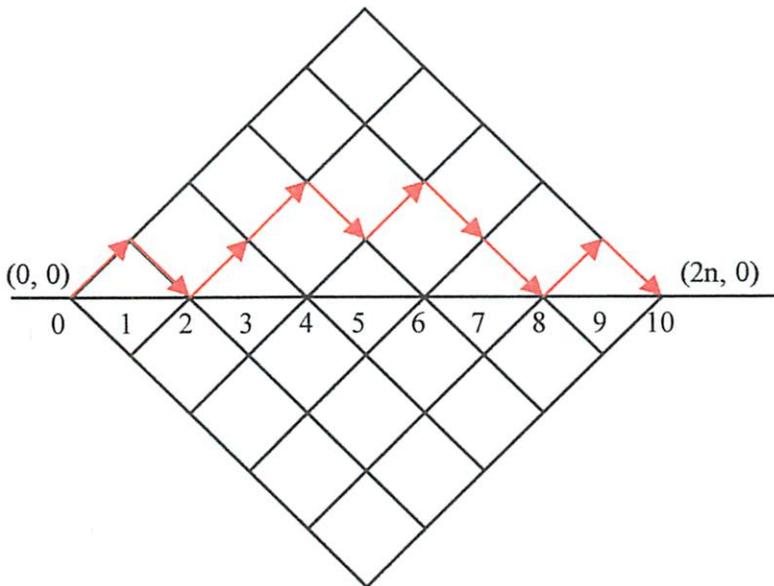


FIGURA 5. Caminho de $(0, 0)$ até $(2n, 0)$

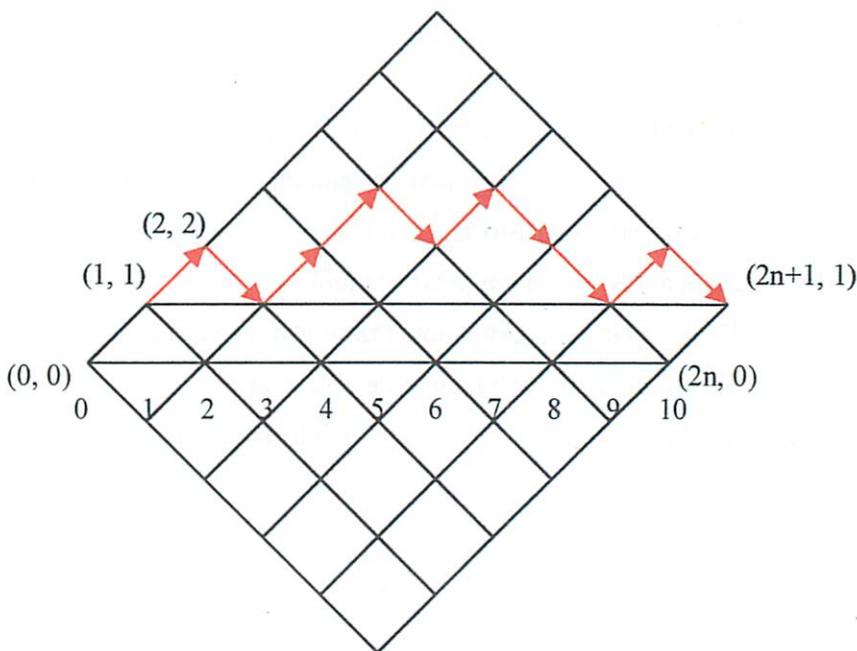


FIGURA 6. Uso do princípio da reflexão

Nesse caso, um caminho acima do eixo ligando $(0, 0)$ a $(2n, 0)$ se torna um caminho ligando $(1, 1)$ a $(2n + 1, 1)$ e que passa pelo ponto $(2, 2)$. Determinam-se então os caminhos de $(2, 2)$ até $(2n + 1, 0)$ acima do eixo $0x$. Pelo princípio da reflexão, o número destes é obtido calculando-se o número de caminhos de $(2, 2)$ até $(2n + 1, 1)$ e subtraindo do número de caminhos de $(2, -2)$ até $(2n + 1, 1)$. Pelo lema I, este número é:

$$\left(\frac{2n+1-2}{2n+1-2+1-2} \right) - \left(\frac{2n+1-2}{2n+1-2+1-(-2)} \right) = \left(\frac{2n-1}{2n-2} \right) - \left(\frac{2n-1}{2n+2} \right) =$$

$$\begin{aligned}
&= \binom{2n-1}{n-1} - \binom{2n-1}{n+1} = \frac{(2n-1)!}{(n-1)!n!} - \frac{(2n-1)!}{(n+1)!(n-2)!} = \\
&= \frac{2n!}{2n} \left(\frac{(n+1)}{(n-1)!(n+1)!} - \frac{n-1}{(n+1)!(n-1)!} \right) = \frac{2n!}{2n} \left(\frac{2}{(n+1)!(n-1)!} \right) = \\
&\quad \frac{1}{(n+1)} \frac{(2n)!}{n!n!} = \frac{1}{n+1} \binom{2n}{n}
\end{aligned}$$

O caso $i = 0$, segue do caso $i = n$ por simetria.

Para $1 \leq i \leq n-1$, o resultado será provado por indução em n .

Para $n = 1$ tem-se que:

$$\frac{1}{n+1} \binom{2n}{n} = \frac{1}{2} \binom{2}{1} = 1$$

Supondo que o resultado seja válido quando se consideram caminhos de comprimento menor que $2n$. Denota-se por $2r$ a coordenada em que pela primeira vez um caminho qualquer, com exatamente $2k$ lados acima do eixo, intercepta o eixo Ox . Têm-se duas possibilidades: a primeira, que no intervalo $[0, 2r]$ o caminho seja estritamente positivo. Nesse caso, tem-se $1 \leq r \leq k$ e no intervalo $[2r, 2n]$ o caminho tem $2k - 2r$ lados acima do eixo. Pela hipótese de indução, o número de caminhos de $(2r, 0)$ até $(2n, 0)$ com $2k - 2r$ lados acima do eixo Ox é de:

$$\frac{1}{n-r+1} \binom{2n-2r}{n-r} \quad 1 \leq r \leq k$$

Tem-se que calcular o número de caminhos de $(0, 0)$ a $(2r, 0)$ que estão estritamente acima do eixo Ox .

Lema II: O número de caminhos de $(0, 0)$ a $(2r, 0)$ estritamente acima do eixo Ox é dado por:

$$\frac{1}{r} \binom{2r-2}{r-1}$$

Demonstração:

Esses caminhos passam necessariamente pelos pontos $(1, 1)$ e $(2r-1, 1)$ e o número deles é igual ao número de caminhos estritamente positivos de $(1, 1)$ até $(2r-1, 1)$. Esse último pode ser calculado, pelo princípio da reflexão, como a diferença entre o número total de caminhos de $(1, 1)$ até $(2r-1, 1)$ e o número de caminhos de $(1, -1)$ até $(2r-1, 1)$. Pelo lema 1, tem-se:

$$\begin{aligned} & \left(\binom{2r-1-1}{\frac{2r-1-1+1-1}{2}} \right) - \left(\binom{2r-1-1}{\frac{2r-1+1-(-1)}{2}} \right) = \binom{2r-2}{r-1} - \binom{2r-2}{r} = \\ & = \frac{(2r-2)!}{(r-1)!(r-1)!} - \frac{(2r-2)!}{r!(r-2)!} = \frac{(2r-2)!}{(r-1)!(r-1)!} \left[1 - \frac{r-1}{r} \right] = \\ & = \frac{(2r-2)!}{(r-1)!(r-1)!r} = \frac{(2r-2)!}{r!(r-1)!} = \frac{(2r)(2r-1)(2r-2)!}{(2r)(2r-1)r!(r-1)!} = \frac{1}{2(2r-1)} \frac{(2r)!}{r!r!} = \\ & = \frac{1}{2(2r-1)} \frac{(2r)(2r-1)(2r-2)!}{r(r-1)!r(r-1)!} = \frac{1}{r} \frac{(2r-2)!}{(r-1)!(r-1)!} = \frac{1}{2(2r-1)} \binom{2r}{r} \end{aligned}$$

Portanto, o número de caminhos estritamente positivos de $(0, 0)$ até $(2r, 0)$ é:

$$\frac{1}{r} \binom{2r-2}{r-1}$$

Segue então que o número de caminhos com $2k$ lados acima do eixo Ox de $(0, 0)$ até $(2n, 0)$ e que são caminhos estritamente acima do eixo Ox de $(0, 0)$ até $(2r, 0)$, Figura 7, é dado por:

$$\frac{1}{r} \binom{2r-2}{r-1} \frac{1}{n-r+1} \binom{2n-2r}{n-r} \quad 1 \leq r \leq k$$

A outra possibilidade são os caminhos tais que de $(0, 0)$ até $(2r, 0)$ que estão estritamente abaixo do eixo Ox , conforme Figura 8.

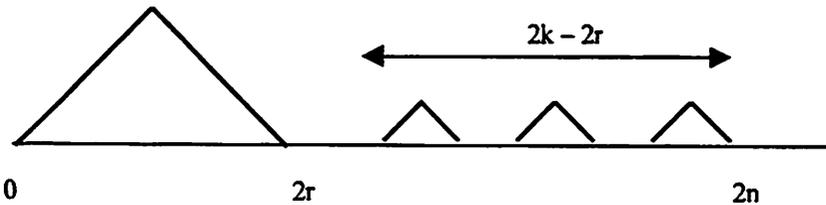


FIGURA 7. Caminhos estritamente acima do eixo Ox de $(0, 0)$ até $(2r, 0)$

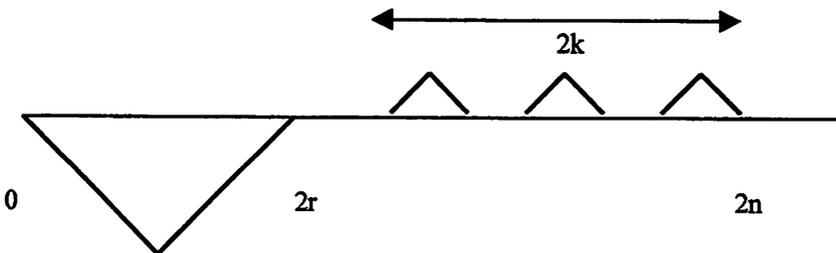


FIGURA 8 Caminhos estritamente abaixo do eixo Ox de $(0, 0)$ até $(2r, 0)$

Nesse caso, a ocorrência dos $2k$ acima do eixo, se dá no intervalo $[2r, 2n]$, o que implica $n - r \geq k$. Pela hipótese de indução, os caminhos de $(2r, 0)$ até $(2n, 0)$ com $2k$ lados acima do eixo é:

$$(I) \quad \frac{1}{n-r+1} \binom{2n-2r}{n-r} \quad n-k < r$$

E, portanto para esses caminhos também vale a fórmula:

$$(II) \quad \frac{1}{r} \frac{1}{n-r+1} \binom{2r-2}{r-1} \binom{2n-2r}{n-r} \quad 1 \leq r \leq n-k$$

O número total de caminhos de $(0, 0)$ até $(2n, 0)$ com $2k$ lados acima do eixo Ox , $1 \leq k \leq n-1$ é obtido então somando-se em r (I) e (II). Tem-se que:

$$\sum_{r=1}^{n-k} \frac{1}{r(n-r+1)} \binom{2r-2}{r-1} \binom{2n-2r}{n-r}$$

fazendo $\rho = n - r + 1 \Rightarrow r = n + 1 - \rho$ e $1 \leq r \leq n - k \Rightarrow k + 1 \leq \rho \leq n$

$$\begin{aligned} & \sum_{\rho=k+1}^n \frac{1}{(n-\rho+1)\rho} \binom{2(n+1-\rho)-2}{n+1-\rho-1} \binom{2n-2(n+1-\rho)}{n-(n+1-\rho)} = \\ & = \sum_{\rho=k+1}^n \frac{1}{(n-\rho+1)\rho} \binom{2\rho-2}{\rho-1} \binom{2n-2\rho}{n-\rho} \end{aligned}$$

De onde segue que a soma em r de (I) e (II) é da forma:

$$\sum_{r=1}^n \frac{1}{r(n-r+1)} \binom{2r-2}{r-1} \binom{2n-2r}{n-r}$$

Como o somatório não depende de k, segue que o número de caminhos de (0, 0) até (2n, 0) com 2k, 1 ≤ k ≤ n-1, acima do eixo Ox é o mesmo para cada k. Combinando com o caso de k = 0 e k = n segue que o número de tais caminhos para cada k é:

$$\frac{1}{n+1} \binom{2n}{n}$$

Assim,

$$P\{S=i\} = \frac{\text{n}^\circ \text{ de seqüências tais que } b_k > a_k \text{ i vezes}}{\binom{2n}{n}} = \frac{\frac{1}{n+1} \binom{2n}{n}}{\binom{2n}{n}} = \frac{1}{n+1}$$

Tem-se então um resultado surpreendentemente não intuitivo que enganou até a Galton.

4.2 Teste do sinal com postos

Para obtenção do teste do sinal com postos é necessário definir o valor médio de postos do conjunto Y em relação a seu posto em W por:

$$\bar{P}_Y = \frac{\sum_{i=1}^N (i \times I(Y)_{(i)})}{n}$$

O valor esperado e a variância de \bar{P}_Y sob $H_0: \mu_Y = \mu_X$ são:

$$E(\bar{P}_Y) = \frac{N+1}{2} \quad \text{e} \quad V(\bar{P}_Y) = \frac{m(N+1)}{12n}$$

Os valores mínimo e máximo de \bar{P}_Y são:

$$\bar{P}_{Y(\min)} = \frac{n+1}{2} \quad \text{e} \quad \bar{P}_{Y(\max)} = \frac{2N-n+1}{2}$$

Para definir o teste do sinal, cuja principal idéia baseia-se na contagem do número de variáveis do conjunto X que são iguais ou inferiores ao valor \bar{P}_Y , é necessário definir uma variável para indicar se o posto i de $w_{(i)}$ é inferior a \bar{P}_Y por:

$$I(\bar{P})_{(i)} = \begin{cases} 1 & \text{se } i \leq \bar{P}_Y \\ 0 & \text{se } i > \bar{P}_Y \end{cases}$$

Com estas definições estabelecidas é possível definir o teste do sinal baseado em postos. Para isso definiu-se a estatística S como o número de variáveis $w_{(i)}$ pertencentes à amostra X cujo posto correspondente em W é menor ou igual a \bar{P}_Y . A estatística S é formalmente dada por:

$$S = \sum_{i=1}^N [I(X)_{(i)} I(\bar{P})_{(i)}]$$

A distribuição exata de S , para valores pequenos de m e n , não foi conseguida analiticamente, foi determinada computacionalmente gerando a distribuição nula de S . Foram computadas as médias e as variâncias e a proporção média esperada de sucessos, ou seja, da ocorrência de valores pertencentes a X na região determinada pela média

de postos de $Y(\bar{P}_Y)$. Esses valores para diferentes combinações de m e n estão resumidos na Tabela 1, considerando ainda pequenos valores amostrais.

Observa-se que as médias encontradas se aproximam dos valores da média da distribuição binomial com $p = 1/2$, dada por np , para praticamente todas as combinações de valores de m e n . As maiores divergências foram encontradas para pequenos valores de n ($n \leq 6$) e de m . A partir de $n = 7$, a média tende rapidamente para a metade do valor de m .

Já as variâncias tendem a se aproximar da variância de uma binomial com $p = 1/2$, dada por $np(1 - p)$, para valores maiores de m e n . Observa-se que, para $n = 1$, os valores da proporção média esperada de sucessos são todos iguais a $1/2$, tendo portanto uma distribuição uniforme.

Os valores da proporção média esperada de sucessos da ocorrência de valores pertencentes a X na região determinada pela média de postos de Y em função do tamanho amostral de $X_{(m)}$ fixado o tamanho amostral de Y , para alguns casos, estão representados nas Figuras 9 e 10. Analisando a Figura 9, quando foram fixados valores pares de n ($n = 2$, $n = 8$ e $n = 14$), verifica-se que valores da proporção são inicialmente altos para m pequeno e decrescem rapidamente tendendo assintoticamente a $1/2$. Na Figura 10, que representa os valores da proporção quando foram fixados n ímpares ($n = 3$, $n = 9$ e $n = 15$), verifica-se que esses valores iniciam em $1/2$, crescem e depois tendem a $1/2$ novamente. Observa-se portanto, que à medida que m e n crescem a proporção média esperada de sucessos aproxima-se muito fortemente de $1/2$ em ambos os casos.

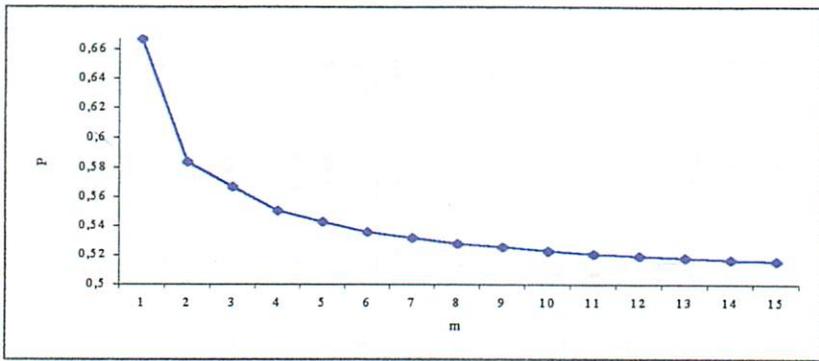
TABELA 1. Médias, variâncias e proporção média esperada de sucessos da distribuição exata de S

n		m														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	μ_s	0,50	1,00	1,50	2,00	2,50	3,00	3,50	4,00	4,50	5,00	5,50	6,00	6,50	7,00	7,50
	σ_s^2	0,25	0,67	1,25	2,00	2,92	4,00	5,25	6,67	8,25	10,00	11,92	14,00	16,25	18,67	21,25
	P_s	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50
2	μ_s	0,67	1,17	1,70	2,20	2,71	3,21	3,72	4,22	4,73	5,23	5,73	6,23	6,73	7,23	7,73
	σ_s^2	0,22	0,47	0,81	1,23	1,73	2,31	2,98	3,73	4,56	5,48	6,48	7,56	8,73	9,98	11,31
	P_s	0,67	0,58	0,57	0,55	0,54	0,54	0,53	0,53	0,53	0,53	0,52	0,52	0,52	0,52	0,52
3	μ_s	0,50	1,00	1,55	2,06	2,57	3,08	3,59	4,10	4,60	5,11	5,61	6,11	6,62	7,12	7,62
	σ_s^2	0,25	0,60	0,95	1,31	1,74	2,21	2,74	3,23	3,96	4,64	5,39	6,18	7,03	7,94	8,90
	P_s	0,50	0,50	0,52	0,51	0,51	0,51	0,51	0,51	0,51	0,51	0,51	0,51	0,51	0,51	0,51
4	μ_s	0,60	1,07	1,57	2,07	2,57	3,07	3,58	4,08	4,58	5,09	5,59	6,09	6,59	7,09	7,60
	σ_s^2	0,24	0,46	0,76	1,09	1,48	1,90	2,35	2,84	3,37	3,94	5,55	5,20	5,89	6,62	7,40
	P_s	0,60	0,53	0,52	0,52	0,51	0,51	0,51	0,51	0,51	0,51	0,51	0,51	0,51	0,51	0,51
5	μ_s	0,50	1,00	0,52	2,03	2,54	3,05	3,55	4,05	4,55	5,06	5,56	6,06	6,56	7,07	7,57
	σ_s^2	0,50	0,57	0,86	0,16	0,49	1,85	2,24	2,67	3,13	3,62	4,14	4,70	5,27	5,90	6,55
	P_s	0,25	0,50	0,51	0,51	0,51	0,51	0,51	0,51	0,51	0,51	0,51	0,51	0,50	0,50	0,50
6	μ_s	0,57	1,03	1,54	2,03	2,54	3,04	3,54	4,04	4,55	5,05	5,55	6,05	6,55	7,05	7,55
	σ_s^2	0,24	0,46	0,75	1,06	0,39	1,74	2,11	2,51	2,92	3,36	3,83	4,33	4,83	5,40	5,97
	P_s	0,57	0,52	0,51	0,51	0,51	0,51	0,51	0,51	0,51	0,50	0,50	0,50	0,50	0,50	0,50
7	μ_s	0,50	1,00	1,52	2,02	2,53	3,03	3,53	4,03	4,53	5,03	5,53	6,04	6,54	7,04	7,54
	σ_s^2	0,25	0,56	0,82	1,09	1,39	1,71	2,05	2,42	2,81	3,20	3,61	4,08	4,56	5,06	5,57
	P_s	0,50	0,50	0,51	0,51	0,51	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50

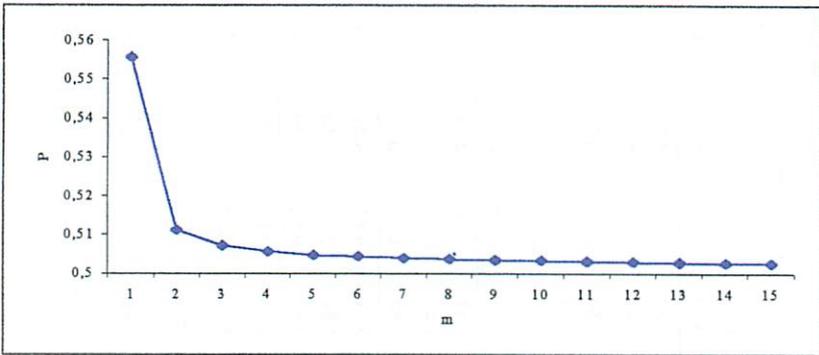
... cont. ...

TABELA 1, Cont.

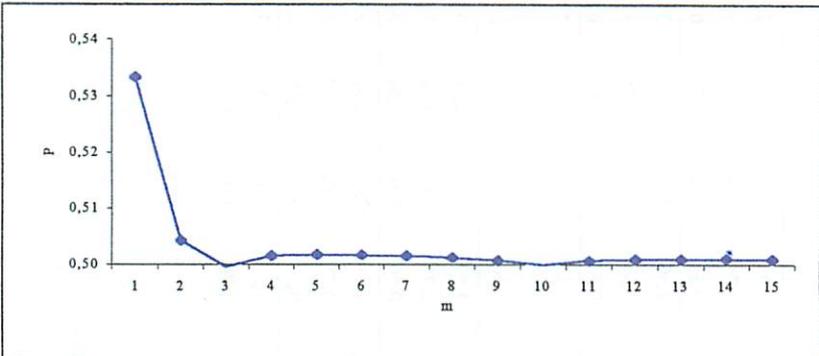
8	μ_s	0,55	1,02	1,52	2,02	2,52	3,03	3,53	4,03	4,53	5,03	5,53	6,03	6,54	7,04	7,54
	σ_s^2	0,25	0,47	0,75	1,04	1,34	1,66	1,99	2,33	2,70	3,08	3,49	3,91	4,35	4,81	5,29
	P_s	0,56	0,51	0,51	0,51	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50
9	μ_s	0,50	1,00	1,50	2,01	2,51	3,02	3,52	4,02	4,52	5,03	5,53	6,03	6,53	7,03	7,52
	σ_s^2	0,25	0,55	0,80	1,06	1,35	1,65	1,96	2,29	2,64	3,00	3,38	3,77	4,18	4,61	5,06
	P_s	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50
10	μ_s	0,55	1,00	1,51	2,01	2,51	3,01	3,51	4,01	4,52	5,02	5,52	6,02	6,5	7,02	7,52
	σ_s^2	0,25	0,45	0,75	1,03	1,31	1,61	1,91	2,24	2,58	2,93	3,28	3,67	4,06	4,46	4,88
	P_s	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50
11	μ_s	0,50	1,00	1,50	2,00	2,51	3,01	3,51	4,00	4,51	5,02	5,52	6,02	6,51	7,02	7,52
	σ_s^2	0,25	0,54	0,79	1,05	1,32	1,59	1,90	2,20	2,52	2,87	3,22	3,58	3,93	4,33	4,74
	P_s	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50
12	μ_s	0,54	1,01	1,51	2,01	2,51	3,01	3,51	4,01	4,51	5,01	5,51	6,01	6,52	7,01	7,52
	σ_s^2	0,25	0,47	0,75	1,02	1,30	1,58	1,88	2,18	2,50	2,82	3,16	3,51	3,88	4,25	4,64
	P_s	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50
13	μ_s	0,50	1,00	1,50	2,01	2,51	3,01	3,51	4,01	4,50	5,01	5,51	6,01	6,52	7,01	7,51
	σ_s^2	0,25	0,53	0,78	1,04	1,31	1,58	1,87	2,16	2,46	2,79	3,12	3,46	3,81	4,17	4,54
	P_s	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50
14	μ_s	0,53	1,01	1,50	2,01	2,51	3,01	3,51	4,01	4,51	5,00	5,51	6,01	6,51	7,01	7,51
	σ_s^2	0,25	0,47	0,74	1,02	1,30	1,57	1,85	2,15	2,44	2,75	3,07	3,41	3,74	4,10	4,46
	P_s	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50
15	μ_s	0,50	1,00	1,50	2,00	2,50	3,01	3,51	4,01	4,51	5,00	5,51	6,01	6,51	7,01	7,52
	σ_s^2	0,25	0,53	0,77	1,03	1,28	1,56	1,84	2,13	2,43	2,73	3,05	3,37	3,70	4,04	4,39
	P_s	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50



(a)

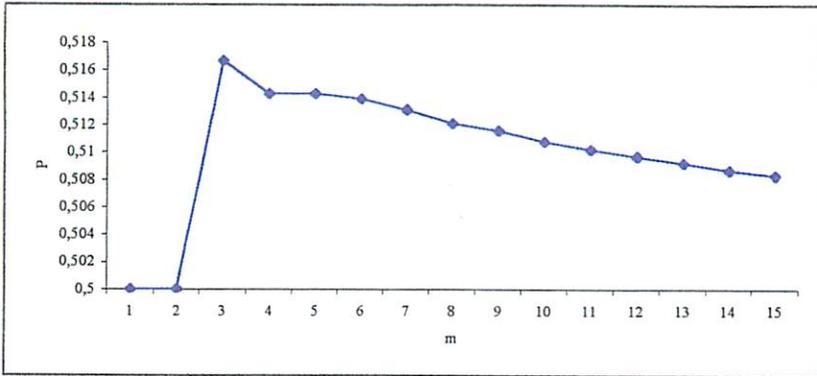


(b)

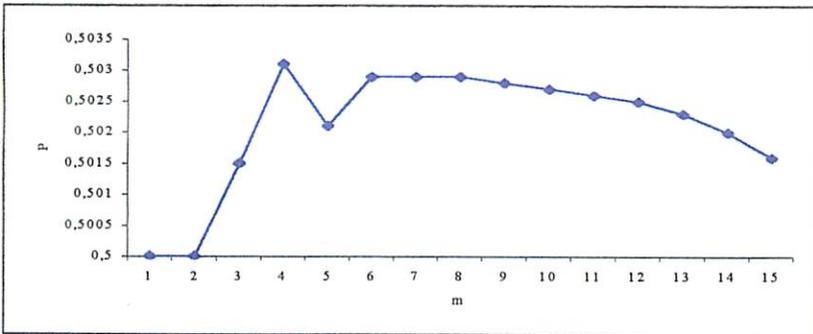


(c)

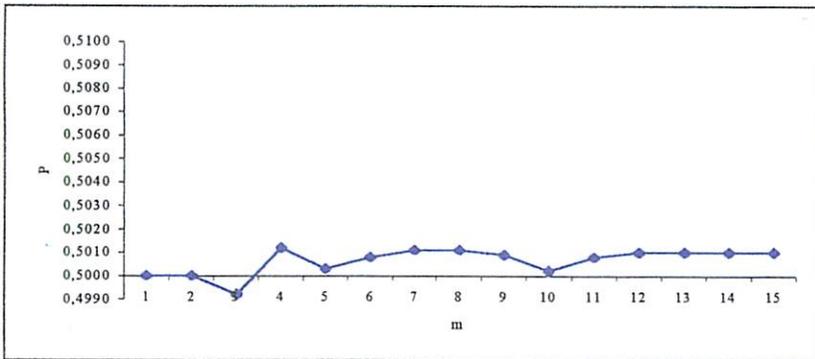
FIGURA 9. Valores da proporção média esperada de sucessos da ocorrência de valores pertencentes a X na região determinada pela média de postos de Y (\bar{P}_Y) em função do tamanho amostral de X (m) fixado o tamanho amostral de Y em $n = 2$ (a), $n = 8$ (b) e $n = 14$ (c).



(a)



(b)



(c)

FIGURA 10. Valores da proporção média esperada de sucessos da ocorrência de valores pertencentes a X na região determinada pela média de postos de Y (\bar{P}_Y) em função do tamanho amostral de X (m) fixado o tamanho amostral de Y em $n = 3$ (a), $n = 9$ (b) e $n = 15$ (c).

Através da distribuição nula de S foram obtidos os valores críticos da estatística S para pequenos valores de m e n (no caso, até 15), considerando significâncias de 1%, 5% e 10%, em um teste bilateral, conforme mostrado nas Tabelas 2, 3 e 4. Ao se aplicar o teste, a hipótese nula deverá ser rejeitada se os dados amostrais conduzem a uma estatística do teste que está na região crítica – isto é, a estatística do teste é no máximo igual ao(s) valor(es) crítico(s) no caso de 0,5%, 2,5% e 5% e no mínimo igual ao(s) valor(es) crítico(s) no caso de 99,5%, 97,5% e 95% . Em caso contrário, a hipótese nula não deve ser rejeitada. As lacunas em branco nas tabelas indicam que não é possível obter um valor na região crítica.

Foi feita uma comparação da distribuição nula de S para alguns valores de m e n com a binomial correspondente. Na Figura 11 apresenta-se a comparação da distribuição nula de S para os valores de $m = 15$ e $n = 5$ com a distribuição binomial com $p = 1/2$ e $n = 5$. Observa-se que as médias são bastante próximas e as variâncias diferem entre si, sendo que a distribuição nula apresenta uma variância bem maior que a binomial. A Figura 12 representa a comparação da distribuição nula de S para os valores de $m = 15$ e $n = 15$ com a distribuição binomial com $p = 1/2$ e $n = 15$. Verifica-se que as médias e as variâncias são bastante próximas e que a distribuição nula continua apresentando uma variância um pouco superior à binomial. Verifica-se portanto a adequação do modelo binomial à distribuição da estatística S , principalmente para maiores valores de m e n .

Para valores elevados de m e n (15 e 15), a distribuição nula de S pode ser satisfatoriamente aproximada pela binomial. Essa estratégia foi realizada e a sua acurácia foi avaliada por simulação.

TABELA 2. Valores Críticos de 1% da distribuição nula de S

		Valores Críticos de $S_{0,005}$														
n	m	3	4	5	6	7	8	9	10	11	12	13	14	15		
8						0	0	0	0	0	0	0	0	0		
9					0	0	0	0	0	0	0	0	0	0		
10				0	0	0	0	0	0	0	0	0	0	0		
11			0	0	0	0	0	0	0	0	0	0	0	0		
12			0	0	0	0	0	0	0	0	0	0	0	1		
13			0	0	0	0	0	1	1	1	1	1	1	1		
14		0	0	0	0	1	1	1	1	1	1	1	1	1		
15		0	0	0	1	1	1	1	1	1	1	1	1	1		

		Valores Críticos de $S_{0,995}$														
n	m	3	4	5	6	7	8	9	10	11	12	13	14	15		
9							9	9	9	9	9	9	9	9		
10					10	10	10	10	10	10	10	10	10	10		
11				11	11	11	11	11	11	11	11	11	11	11		
12			12	12	12	12	12	12	12	12	12	12	12	12		
13			13	13	13	13	13	13	13	13	13	13	13	13		
14			14	14	14	14	14	14	14	14	14	14	14	14		
15		15	15	15	15	15	15	15	15	15	15	15	15	15		

TABELA 3. Valores Críticos de 5% da distribuição nula de S

		Valores Críticos de $S_{0,025}$														
n	m	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
6				0	0	0	0	0	0	0	0	0	0	0	0	
7			0	0	0	0	0	0	0	0	0	0	0	0	0	
8	0		0	0	0	0	0	0	0	0	0	0	0	0	0	
9	0	0		0	0	0	0	0	0	0	0	1	1	1	1	
10	0	0	0		0	1	1	1	1	1	1	1	1	1	1	
11	0	0	0	1		1	1	1	1	1	1	1	1	1	1	
12	0	0	0	1	1		1	1	1	1	1	1	1	1	1	
13	0	1	1	1	1	1		2	2	2	2	2	2	2	2	
14	0	1	1	1	1	2	2		2	2	2	2	2	2	2	
15	0	1	1	2	2	2	2	2		2	2	2	2	2	2	

		Valores Críticos de $S_{0,975}$														
n	m	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
6																
7			7	7	7	7	7	7	7	7	7	7	7	7	7	7
8			8	8	8	8	8	8	8	8	8	8	8	8	8	8
9			9	9	9	9	9	9	9	9	9	9	9	9	9	9
10			10	10	10	10	10	10	10	10	10	10	10	10	10	10
11			11	11	11	11	11	11	11	11	11	11	11	11	11	11
12	12		12	12	12	12	12	12	12	12	12	12	12	12	12	12
13	13		13	13	13	13	13	13	13	13	13	13	13	13	13	13
14	14		14	14	14	14	14	14	14	14	14	14	14	14	14	14
15	15		15	15	15	15	15	15	15	15	15	15	15	15	15	15

TABELA 4. Valores Críticos de 10% da distribuição nula de S

		Valores Críticos de $S_{0,05}$														
$n \backslash m$	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
5			0	0	0	0	0	0	0	0	0	0	0	0		
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
8	0	0	0	0	1	1	1	1	1	1	1	1	1	1		
9	0	0	1	1	1	1	1	1	1	1	1	1	1	1		
10	0	1	1	1	1	1	1	1	1	1	1	1	1	1		
11	0	1	1	1	1	1	2	2	2	2	2	2	2	2		
12	1	1	1	2	2	2	2	2	2	2	2	2	2	2		
13	1	1	2	2	2	2	2	2	2	2	2	2	2	2		
14	1	2	2	2	2	2	3	3	3	3	3	3	3	3		
15	1	2	2	2	3	3	3	3	3	3	3	3	3	3		

		Valores Críticos de $S_{0,95}$														
$n \backslash m$	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
5				5	5	5	5	5	5	5	5	5	5	5		
6			6	6	6	6	6	6	6	6	6	6	6	6		
7		6	7	7	7	7	7	7	7	7	7	7	7	7		
8		6	8	8	8	8	8	8	8	8	8	8	8	8		
9	8	8	9	9	9	9	9	9	9	9	9	9	9	9		
10	9	9	10	10	10	10	10	10	10	10	10	10	10	10		
11	10	10	11	11	11	11	11	11	11	11	11	11	11	11		
12	11	11	12	12	12	12	12	12	12	12	12	12	12	12		
13	12	12	13	13	13	13	13	13	13	13	13	13	13	13		
14	13	13	14	14	14	14	14	14	14	14	14	14	14	14		
15	14	14	15	15	15	15	15	15	15	15	15	15	15	15		

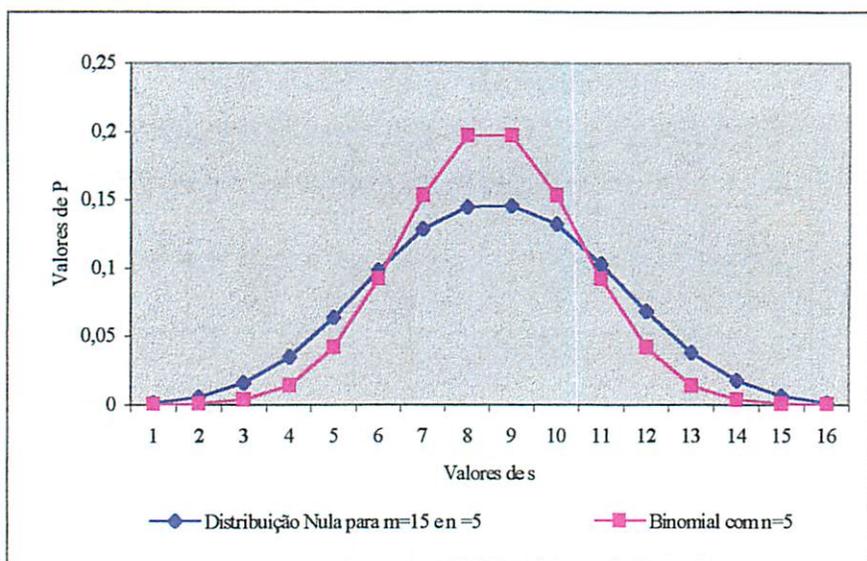


FIGURA 11. Comparação da Distribuição Nula de S para $m = 15$ e $n = 5$ com a Binomial com $p = 1/2$ e $n = 5$

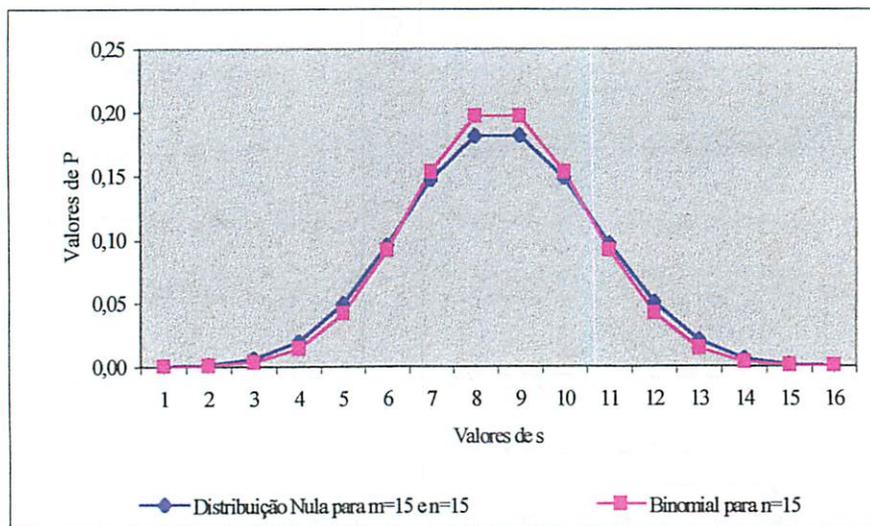


FIGURA 12. Comparação da Distribuição Nula de S para $m = 15$ e $n = 15$ com a Binomial com $p = 1/2$ e $n = 15$

4.3 Simulação

Os resultados das taxas de erro tipo I, para o valor nominal de 5%, estão resumidos em gráficos, para tamanhos de amostras idênticos e em tabelas, para tamanhos de amostras diferentes, para amostras provenientes de distribuição normal e exponencial. Foi usado intervalo de confiança exato para proporções de Leemis e Trivedi (1996). O intervalo de confiança para a taxa empírica de erro tipo I de 99% para o valor nominal de 5%, em porcentagem, foi [3,83; 6,39].

Na Figura 13 apresenta-se a comparação das taxas de erro tipo I entre os testes de Mann-Whitney, Galton, t de Student e o teste proposto de sinais com postos, para dados provenientes de duas amostras independentes de mesmo tamanho de uma distribuição normal. Observa-se que os testes de Mann-Whitney e t não diferem muito entre si, oscilando em torno do valor de significância de 5%. O teste de Galton só apresenta significância a partir de amostras de tamanho 20, sendo que de 20 e 40 a significância empírica é menor que 5% (teste rigoroso) e a partir desse tamanho de amostral oscila em torno do valor nominal de 5%. O teste de sinal com postos apresenta taxa de erro tipo I maior que o valor nominal de 5% até tamanho de amostras 35. A partir desse valor de m e n , as taxas são consideradas iguais à nominal, indicando que a aproximação binomial é adequada para essa situação.

Na Figura 14 apresenta-se a comparação das taxas de erro tipo I para os mesmos testes, considerando dados provenientes de duas amostras independentes de mesmo tamanho de uma distribuição exponencial. Os resultados foram bastante parecidos com os anteriores. Eles comprovam a robustez do teste t. Já os testes de Mann-Whitney e Galton, por serem não-paramétricos, era esperado que a mudança de distribuição não

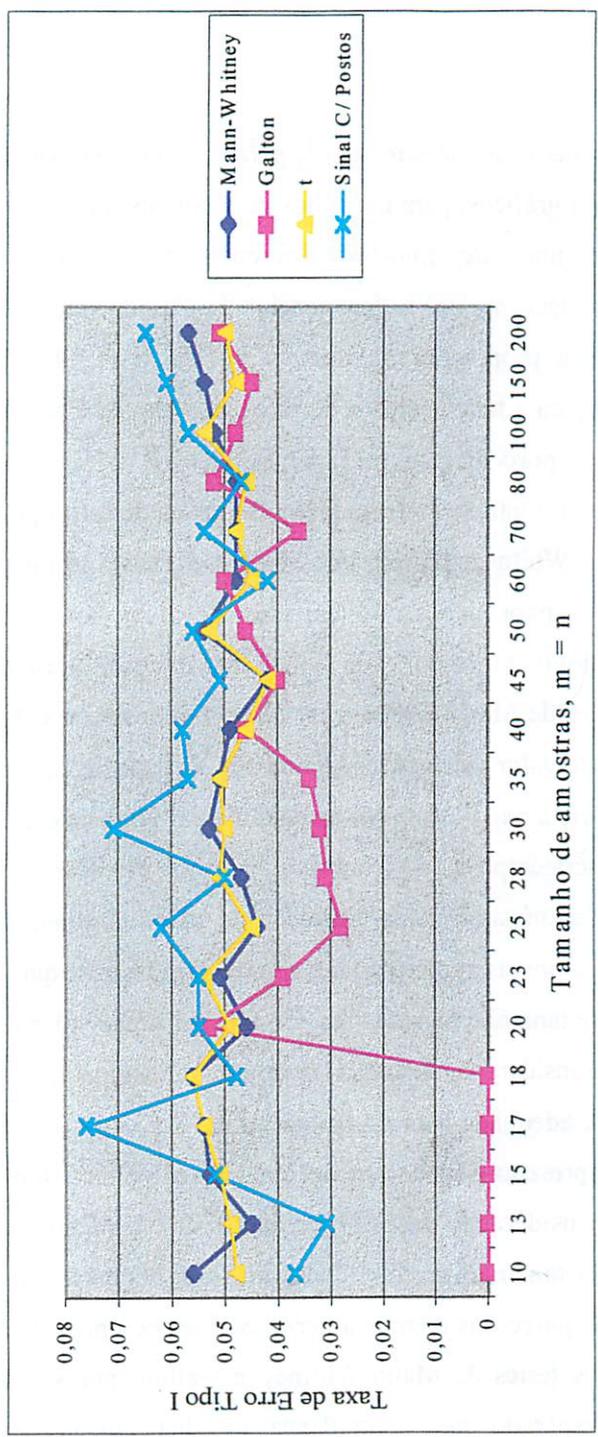


FIGURA 13. Comparação das taxas de erro tipo I para dados provenientes de duas amostras independentes, de mesmo tamanho, de uma distribuição normal.

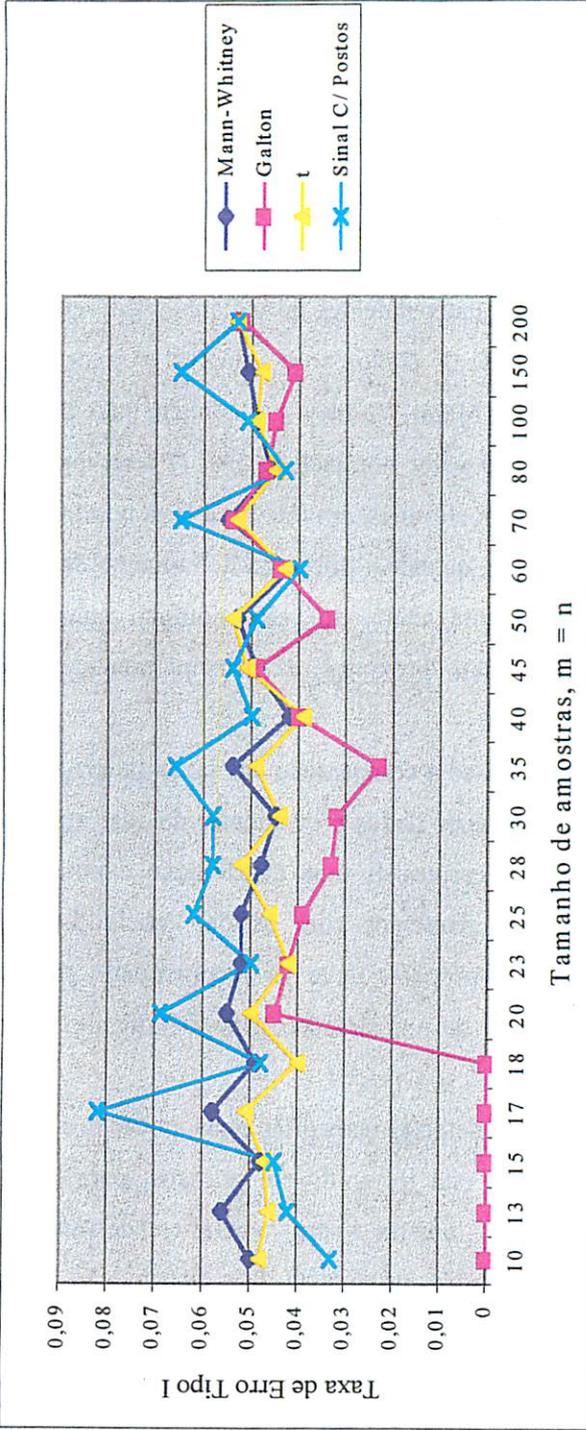


FIGURA 14. Comparação das taxas de erro tipo I para dados provenientes de duas amostras independentes, de mesmo tamanho, de uma distribuição exponencial.

afetasse. Da mesma forma o teste proposto manteve suas características por ser também um teste não-paramétrico.

Na Tabela 5 apresenta-se a comparação das taxas de erro tipo I dos testes de Mann-Whitney, t de Student e sinal com postos para dados provenientes de duas amostras independentes de tamanhos diferentes de uma distribuição normal. O teste de Galton só é aplicável a amostras de mesmo tamanho. Os testes de Mann-Whitney e t também não diferem entre si, oscilando em torno do valor nominal de 5%. O teste proposto apresenta taxas de erro bem mais elevadas, em relação à taxa nominal e às situações em que m é igual a n , quando a diferença de tamanho das duas amostras é maior e, ligeiramente acima do valor nominal quando os tamanhos de amostras são mais próximos e, principalmente, se os tamanhos amostrais são grandes.

Na Tabela 6 apresenta-se a comparação das taxas de erro tipo I para os mesmos testes, considerando dados provenientes de duas amostras independentes de tamanhos diferentes de uma distribuição exponencial. Os resultados foram bastante parecidos com os da Tabela 5. Esses resultados mostram que o teste proposto deve ser recomendado para ser aplicado somente em situações de amostras de mesmo tamanho ou de tamanhos próximos.

A probabilidade de se detectar reais diferenças, poder dos testes, para o valor nominal de 5%, está resumido em gráficos, para mesmo tamanho de amostras e em tabelas, para tamanhos de amostras diferentes, tanto para duas amostras provenientes de distribuição normal quanto de exponencial.

A comparação dos testes quanto ao poder foi realizada considerando diferenças específicas entre médias populacionais, fixadas

TABELA 5. Comparação das taxas de erro tipo I para dados provenientes de duas amostras independentes, de tamanhos diferentes, de uma distribuição normal.

Tamanho de Amostras		Testes		
m	n	Mann-Whitney	t-Student	Sinal com Posto
24	13	0,049	0,048	0,058
27	19	0,048	0,048	0,039
50	5	0,052	0,052	0,301
50	10	0,044	0,047	0,169
50	15	0,052	0,046	0,115
50	20	0,051	0,054	0,101
50	25	0,050	0,049	0,074
50	30	0,048	0,050	0,067
50	35	0,047	0,050	0,057
50	40	0,053	0,053	0,057
100	5	0,051	0,047	0,456
100	10	0,037	0,043	0,276
100	20	0,058	0,060	0,184
100	30	0,049	0,048	0,117
100	40	0,051	0,055	0,101
100	50	0,050	0,048	0,083
100	60	0,052	0,053	0,074
100	70	0,061	0,062	0,070
100	80	0,054	0,053	0,061
100	90	0,051	0,046	0,055

parametricamente em função do erro padrão da diferença entre médias, em $k = 1/2, 1, 2$ e 4 .

Na Figura 15 apresenta-se a comparação do poder dos testes de Mann-Whitney, Galton, t de Student e o teste de sinais com postos proposto, para dados provenientes de duas amostras independentes de mesmo tamanho de uma distribuição normal para $k = 1/2, 1, 2$ e 4 erros padrão da diferença entre médias. Observa-se que os testes de Mann-Whitney e t não diferem muito entre si, mantendo o valor de

TABELA 6. Comparação das taxas de erro tipo I para dados provenientes de duas amostras independentes, de tamanhos diferentes, de uma distribuição exponencial.

Tamanho de Amostras		Testes		
m	n	Mann-Whitney	t-Student	Sinal com Posto
24	13	0,042	0,043	0,052
27	19	0,042	0,048	0,034
50	5	0,054	0,049	0,314
50	10	0,050	0,041	0,159
50	15	0,049	0,046	0,125
50	20	0,053	0,046	0,097
50	25	0,052	0,050	0,078
50	30	0,054	0,051	0,069
50	35	0,051	0,061	0,062
50	40	0,052	0,052	0,054
100	5	0,045	0,039	0,446
100	10	0,049	0,051	0,308
100	20	0,047	0,039	0,166
100	30	0,045	0,049	0,113
100	40	0,048	0,050	0,093
100	50	0,056	0,049	0,092
100	60	0,048	0,052	0,070
100	70	0,056	0,052	0,072
100	80	0,050	0,048	0,056
100	90	0,047	0,047	0,062

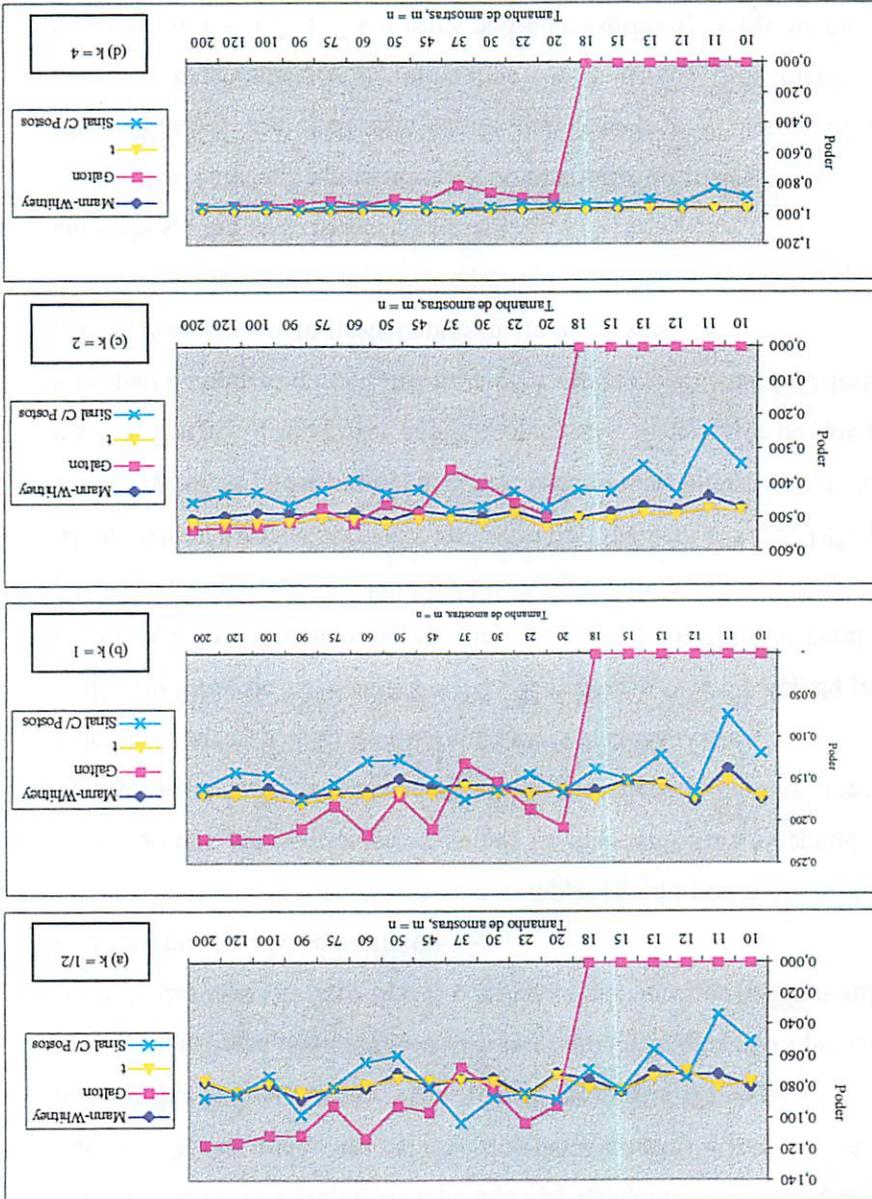
8% para $k = 1/2$ erro padrão da diferença entre médias, 17% para $k = 1$, 50% para $k = 2$ e 98% para $k = 4$. O teste de Galton só detecta diferenças a partir de amostras de tamanho 20, e seu poder cresce a partir desse tamanho de amostra, sendo um bom teste, pois além de controlar as taxas de erro tipo I, mantém um poder elevado. Por ele ser mais rigoroso, era esperado um poder menor. Ocorreu o contrário para $n \geq 35$ e, principalmente, nas situações mais críticas (menores valores de k). O poder fica menor que o de Mann-Whitney e t para $k = 4$ erros padrão de diferença entre médias. O teste de sinal com postos apresenta poder

variando mais para pequenos tamanhos de amostras e $k = 1/2$ e $k = 1$. Para $k = 2$, seu poder fica abaixo dos demais testes e para $k = 4$, seu poder é ligeiramente menor. Para o teste proposto era esperado, para $n < 40$, um poder maior que os demais, uma vez que não apresentou controle das taxas de erro tipo I, com taxas inferiores ao valor nominal. Isso, no entanto, não se verificou.

Na Figura 16 apresenta-se a comparação do poder dos mesmos testes, só que para dados provenientes de duas amostras independentes de mesmo tamanho de uma distribuição exponencial com os mesmos valores de erros padrão de diferença entre médias. O teste t apresenta poder em torno de 12% para $k = 1/2$, 15% para $k = 1$, e valor de poder crescente com o tamanho das amostras para $k = 2$ e $k = 4$. Em geral, seu poder para exponencial é bem menor do que para a normal. O teste de Mann-Whitney apresenta poder em torno de 11% para $k = 1/2$, 14% para $k = 1$, e valor de poder crescente com o tamanho das amostras para $k = 2$ e $k = 4$. Em geral, seu poder para exponencial é bem menor do que para a normal e também é ligeiramente menor do que o do teste t. O teste de Galton e do sinal com postos apresentam as mesmas características para dados provenientes de distribuição normal e exponencial, sendo o poder, em ambos, um pouco menor na última.

Na Tabela 7 apresenta-se a comparação do poder dos testes de Mann-Whitney, t e sinal com postos para dados provenientes de duas amostras independentes de tamanhos diferentes de uma distribuição normal com $k = 1/2, 1, 2$ e 4 erros padrão da diferença entre médias. Os testes de Mann-Whitney e t de Student também não diferem entre si, com o segundo tendo um poder ligeiramente superior. O teste do sinal com posto apresenta poder mais elevado quando a diferença de tamanho da duas amostras é maior, principalmente para $k = 1$ e também

FIGURA 15. Comparação do poder dos testes para dados provenientes de duas amostras independentes, de mesmo tamanho, de uma distribuição normal com (a) $1/2$, (b) 1 , (c) 2 e (d) 4 erros padrão da diferença entre médias, respectivamente.



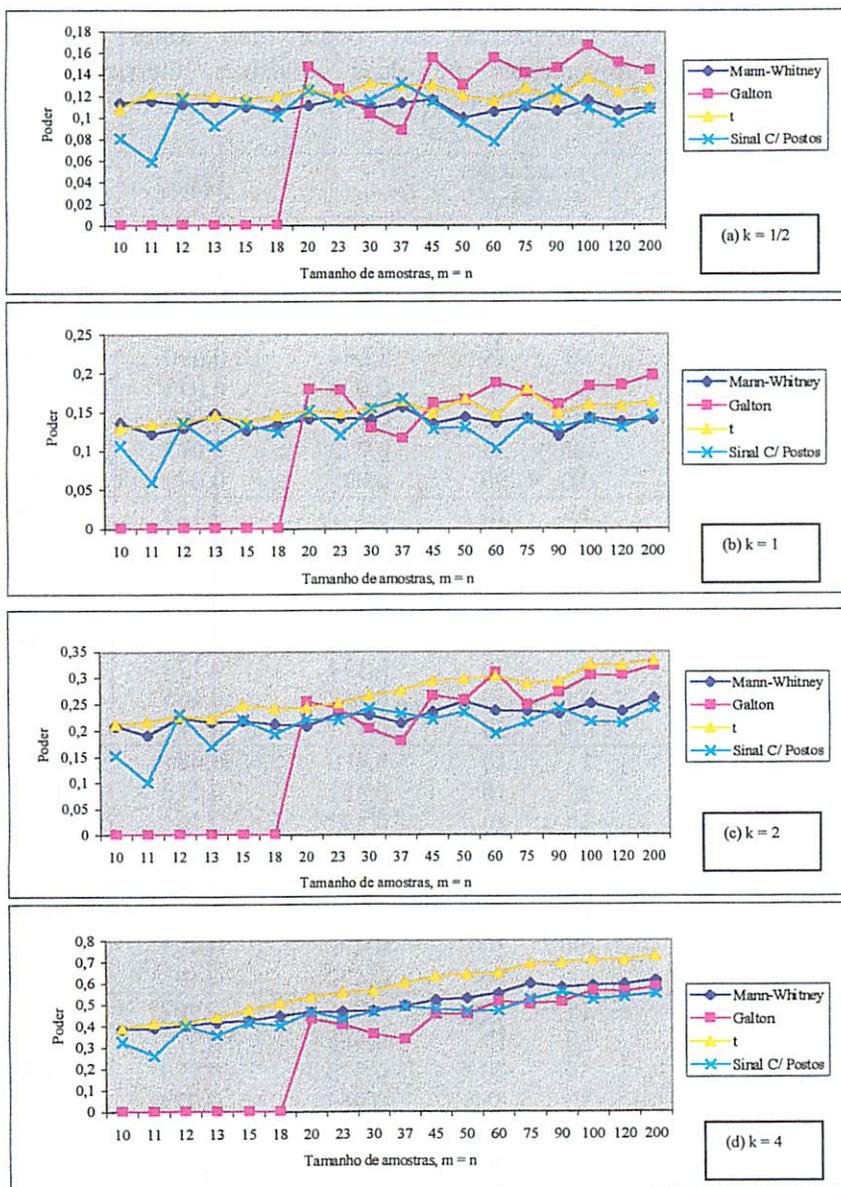


FIGURA 16. Comparação do poder dos testes para dados provenientes de duas amostras independentes, de mesmo tamanho, de uma distribuição exponencial com (a) $1/2$, (b) 1 , (c) 2 e (d) 4 erros padrão da diferença entre médias, respectivamente.

TABELA 7. Comparação do poder dos testes para dados provenientes de duas amostras independentes, de tamanhos diferentes, de uma distribuição normal.

Diferença em Erros Padrão	Tamanho de Amostras		Testes			
	k	m	n	Mann-Whitney	t-Student	Sinal c/ Postos
0,5		15	10	0,075	0,081	0,093
		30	10	0,073	0,079	0,170
		50	20	0,083	0,079	0,130
		50	35	0,084	0,090	0,102
		80	50	0,076	0,073	0,091
		80	70	0,069	0,067	0,074
		100	30	0,072	0,077	0,165
		100	50	0,077	0,078	0,110
1		15	10	0,142	0,153	0,168
		30	10	0,161	0,166	0,291
		50	20	0,164	0,173	0,238
		50	35	0,163	0,169	0,173
		80	50	0,173	0,172	0,187
		80	70	0,174	0,179	0,152
		100	30	0,157	0,165	0,271
		100	50	0,175	0,181	0,226
2		15	10	0,460	0,484	0,474
		30	10	0,481	0,512	0,646
		50	20	0,477	0,504	0,560
		50	35	0,486	0,501	0,464
		80	50	0,499	0,518	0,497
		80	70	0,507	0,514	0,447
		100	30	0,472	0,493	0,626
		100	50	0,472	0,493	0,626
4		15	10	0,955	0,967	0,950
		30	10	0,966	0,975	0,989
		50	20	0,973	0,980	0,982
		50	35	0,969	0,977	0,955
		80	50	0,972	0,976	0,962
		80	70	0,974	0,978	0,948
		100	30	0,971	0,978	0,983
		100	50	0,978	0,983	0,976

para as situações de maior diferença nos tamanhos amostrais. Mas não deve ser recomendado indiscriminadamente, pois não controla bem as taxas de erro tipo I, sob diferenciação nos tamanhos amostrais.

Na Tabela 8 apresenta-se a comparação do poder dos testes para os mesmos testes, só que para uma distribuição exponencial com os mesmos k erros padrão de diferença entre médias. Os resultados foram bastante parecidos com os da Tabela 7.

TABELA 8. Comparação do poder dos testes para dados provenientes de duas amostras independentes, de tamanhos diferentes, de uma distribuição exponencial.

Diferença em Erros Padrão	Tamanho de Amostras		Testes			
	k	m	n	Mann-Whitney	t-Student	Sinal c/ Postos
0,5		15	10	0,117	0,146	0,150
		30	10	0,143	0,215	0,264
		50	20	0,136	0,201	0,204
		50	35	0,104	0,141	0,118
		80	50	0,129	0,158	0,137
		80	70	0,109	0,133	0,110
		100	30	0,135	0,200	0,245
		100	50	0,133	0,158	0,170
1		15	10	0,141	0,184	0,191
		30	10	0,167	0,250	0,293
		50	20	0,162	0,221	0,224
		50	35	0,139	0,174	0,151
		80	50	0,148	0,187	0,165
		80	70	0,132	0,175	0,125
		100	30	0,148	0,218	0,262
		100	50	0,154	0,182	0,197
2		15	10	0,209	0,269	0,261
		30	10	0,220	0,333	0,339
		50	20	0,227	0,317	0,312
		50	35	0,227	0,284	0,238
		80	50	0,218	0,285	0,229
		80	70	0,238	0,295	0,218
		100	30	0,202	0,283	0,339
		100	50	0,223	0,302	0,271
4		15	10	0,196	0,260	0,253
		30	10	0,342	0,498	0,513
		50	20	0,372	0,525	0,456
		50	35	0,188	0,244	0,205
		80	50	0,474	0,608	0,479
		80	70	0,546	0,657	0,502
		100	30	0,377	0,518	0,522
		100	50	0,441	0,600	0,501

4.4 Exemplo de aplicação

A teoria desenvolvida para o teste do sinal com postos foi aplicada a um exemplo real, de duas cultivares de batata (Achat e Monalisa). As hipóteses a serem consideradas são:

$$H_0: \mu_{\text{Achat}} = \mu_{\text{Monalisa}}$$

$$H_1: \mu_{\text{Achat}} \neq \mu_{\text{Monalisa}}$$

Foram retiradas duas amostras independentes de tamanho 25 e aplicou-se os testes de Mann-Whitney, Galton, t de Student e sinal com postos. Os dados estão resumidos na Tabela 9.

Para aplicar o teste de Mann-Whitney, dispõem-se ambas as seqüências em ordem crescente e atribuem-se os postos como se todos os dados formassem uma única seqüência, aos empates atribui-se o posto médio. A seguir, somam-se os postos para cada seqüência, conforme Tabela 10.

Calcula-se então a soma esperada dos postos para qualquer uma das seqüências e compara-a com a soma observada dos postos para aquela seqüência. Considerando a variedade Achat, tem-se:

$$E(R) = \frac{m(N+1)}{2} = \frac{25(50+1)}{2} = 637,50$$

$$\sigma = \sqrt{\frac{mn(m+n+1)}{12}} = \sqrt{\frac{25 \cdot 25 \cdot (25+25+1)}{12}} = 51,54$$

$$Z = \frac{R - E(R)}{\sigma} = \frac{773 - 637,50}{51,54} = 2,63$$

Tabela 9. Pesos dos tubérculos de uma planta para 2 cultivares de batata

Amostra	Achat	Monalisa
1	500	560
2	580	130
3	520	690
4	610	300
5	760	370
6	910	670
7	940	350
8	730	370
9	1.070	1.260
10	920	710
11	680	490
12	990	530
13	1.060	390
14	560	390
15	710	550
16	1.010	380
17	540	1.020
18	1.010	500
19	960	750
20	1.100	560
21	1.160	900
22	870	940
23	1.450	1.390
24	1.300	1.050
25	770	1.190

Achat (posto)	Monalisa (posto)
500 (10,5)	130 (1)
520 (12)	300 (2)
540 (14)	350 (3)
560 (16,5)	370 (4,5)
580 (19)	370 (4,5)
610 (20)	380 (6)
680 (22)	390 (7,5)
710 (24,5)	390 (7,5)
730 (26)	490 (9)
760 (28)	500 (10,5)
770 (29)	530 (13)
870 (30)	550 (15)
910 (32)	560 (16,5)
920 (33)	570 (18)
940 (34,5)	670 (21)
960 (36)	690 (23)
990 (37)	710 (24,5)
1.010 (38,5)	750 (27)
1.010 (38,5)	900 (31)
1.060 (42)	940 (34,5)
1.070 (43)	1.020 (40)
1.100 (44)	1.050 (41)
1.160 (45)	1.190 (46)
1.300 (48)	1.260 (47)
1.450 (50)	1.390 (49)
Soma dos postos = 773	Soma dos postos = 502

Tabela 10. Pesos, em gramas, dos tubérculos de uma planta e seus postos

O teste de hipótese de igualdade de médias é baseado na distribuição normal e como $Z \geq z_{\alpha/2}$, é significativo a 1% (o p-valor foi 0,0086 no Minitab). Portanto, rejeita-se H_0 , ou seja, as médias são diferentes.

Para aplicar o teste de Galton, combinam-se as duas seqüências formando uma única, crescente de $2n$ (50) termos: (130, 300, ... , 1390, 1450). Contam-se quantos elementos da segunda seqüência (Monalisa) são maiores que os da primeira (Achat) de mesmo posto. No caso, é apenas 1. Portanto, a probabilidade de a ocorrência ser devida somente ao acaso é $1/(n + 1) = 1/(25 + 1) = 0,0385 = 3,85\%$.

Para aplicar o teste t de Student, devem-se calcular as médias e variâncias das variedades Achat e Monalisa. Os resultados são: 868,40 e 657,60 para as médias e 63.511,33 e 108.560,67 para as variâncias, respectivamente. Aplicando o teste t de igualdade de médias, encontra-se $t_c = 4,17$, rejeitando-se H_0 para o valor nominal de 5% de significância. Já a 1%, não há evidências para rejeitar H_0 , pois como $t_c = 2,54$ a probabilidade p-valor foi igual a 0,014348.

Para aplicar o teste do sinal com posto combinam-se as duas seqüências formando uma única, crescente de 50 termos: (130, 300, ... , 1390, 1450). A média dos postos da segunda seqüência (Monalisa) e o número de valores de Achat que são iguais ou inferiores ao valor do posto médio são, respectivamente:

$$\bar{P}_Y = \frac{\sum_{i=1}^N (i \times I(Y)_{(i)})}{n} = 20 \qquad S = \sum_{i=1}^N [I(X)_{(i)} - I(\bar{P})_{(i)}] = 6$$

Como os tamanhos amostrais (25) excedem os limites dos tamanhos amostrais tabelados de S, faz-se a aproximação para binomial. Assim, tem-se:

$$2 \cdot P(x \leq 6) = 2 \cdot \sum_{x=0}^6 \binom{25}{x} \cdot \left(\frac{1}{2}\right)^{25} = 0,014633$$

Portanto, rejeita-se H_0 para o valor nominal de 5% de significância. Já a 1%, não há evidências para rejeitar H_0 , visto que $P = 0,014633$. Observa-se que essa probabilidade é bem próxima da encontrada pelo teste t.

5 CONCLUSÕES

- O teste dos postos ordenados de Galton é baseado em uma distribuição uniforme pois, todos os resultados são equiprováveis ao se fazer uma amostra conjunta ordenada de duas seqüências.
- A distribuição exata da estatística S para o teste do sinal com postos, para pequenos valores de m e n foram determinadas computacionalmente e gerada uma tabela de valores críticos mediante sua distribuição nula. A distribuição de S para grandes amostras (m e n grandes) aproximou-se da binomial com $p = 1/2$.
- O teste de Galton controla as taxas de erro tipo I, sendo considerado rigoroso e mantém o poder elevado; deve ser recomendado para $n \geq 35$.
- O teste de sinal com postos, usando a aproximação binomial, apresenta controle da taxa de erro tipo I com amostras de mesmo tamanho e poder igual aos demais testes para $n \geq 40$.
- Para amostras de tamanhos diferentes não se recomenda a aplicação do novo teste.

6 REFERÊNCIAS BIBLIOGRÁFICAS

- BANERJEE, S. On confidence interval for two-mean problem based on separate estimates of variances and tabulated values of t-table **Sankhyā: Indian Journal Statistical**, Calcutta, n.10, p.359-378, 1961.
- BEHRENS, W. V. Ein beitrage zur fehlerberechnung bei weinige beobachtungen. **Landwirtsch Jahrbücher**, Buchverlag, n.68, p.807-837, 1929.
- BEARZOTI, E. ; PINTO, C. A.B. P. Dimensionamento de parcela em experimentos de seleção em batata. **Ciência e Agrotecnologia**. Lavras: v.23, p.390-403, 1999.
- BORGES , L. C.; FERREIRA, F. D. Comparação de duas aproximações do teste t com variância heterogêneas através de simulação. **Ciência e Agrotecnologia**. Lavras: v.23, n.2, p.390-430, 1999.
- BUSSAB, W. O.; MORETTIN, P. A. **Estatística Básica**. 4. ed. São Paulo: Atual, 1987.
- CAMPOS, H. **Estatística Experimental Não-Paramétrica**. Piracicaba: ESALQ, 1983. 349p.
- COCHRAN, W. G. Approximate significance levels of the Behrens-Fisher test. **Biometrics**, Washington, v.20, p.191-195, 1964.
- COCHRAN, W. G.; COX, G. M. **Experimental Designs**. 2. ed. New York: John Wiley 1957. 617p.
- CONOVER, W. J. On the methods of handling ties in the Wilcoxon signed-rank test. **Journal American Statistician Association**. Washington, n.68, p. 985-988, 1973.
- CONOVER, W. J.; IMAN, R. L. On the some alternative procedures using ranks for the analysis of experimental designs. **Communications Statistics - Theory and Methods**, New York, A5: p. 1349-1368, 1976.

- DACHS, J. N. **Estatística computacional: Uma introdução em turbo pascal.** Rio de Janeiro: Livros Técnicos e Científicos, 1988. 236p.
- DIXON, W. J.; MASSEY, F. J. Jr. **Introduction to Statistical Analysis.** 3. ed. New York: McGraw-Hill, 1969. 638p.
- FELLER, W., **An Introduction to Probability Theory and Its Applications.** New York: John Willey, 1968. v1, 460p.
- FISHER, R. A. The fiducial argument in statistical inference. **Annals of Eugenics**, London, n.6, p.391-398, 1935.
- FISHER, R. A. The comparison of samples with possibly unequal variances. **Annals of Eugenics**, London, n.9, p.174-180, 1939.
- FISHER, R. A.; YATES, F. **Statistical Tables for Biological, Agricultural, and Medical Research Workers.** 6.ed. New York: Hafner, 1963. 146p.
- FIX, E.; HODGES, J. L. Jr. Significance probabilities of the Wilcoxon test. **Annals Mathematical Statistics**, Hayward, n.26, p. 301-312, 1955.
- GILL, J. L. Analysis of data with heterogeneous data: A review. **Journal Dairy Scientist**, n.54, p.369-373, 1971.
- HODGES, J. L. Jr. Galton's rank-order test. **Biometrika**, London, v.42, p.261-262, 1955.
- HODGES, J. L. Jr.; LEHMANN, E. L. The efficiency of some nonparametric competitors of the t-test. **Annals Mathematical Statistics**, Hayward, n.27, p. 324-335, 1956.
- HODGES, J. L. Jr.. Improved significance probabilities of the Wilcoxon test. **Journal Educational Statistics**, Washington, n.15, p. 249-265, 1990.
- HOLLANDER, M. ; WOLFE, D. **A Nonparametrical Statistical Methods.** New York: John Wiley, 1973.503p.
- IMAN, R. L. Use of a t-statistic as an approximation to the exact distribution of the Wilcoxon signed ranks test statistic.

Communications Statistics – Theory and Methods, New York, A3: p. 795-806, 1974.

JEFFREYS, H. Notes on the Behrens-Fisher formula. **Annals of Eugenics**, Londres, n.10, p.48–51, 1940.

LEE, A. F. S.; FINEBERG, N. S. A fitted test for the Behrens-Fisher problem. **Communication Statistical – Theory and Methods**, New York, n.20, p.653-666, 1991.

LEE, A. F. S.; GURLAND, J. Size and power of tests for equality of means of two normal populations with unequal variances. **Journal of the American Statistical Association**, Washington, v.70, p.933-941, 1975.

LEEMIS, L. ; TRIVEDI, K. S. A comparison of approximation interval estimators of the Bernoulli parameter. **The American Statistician**, Alexandria, v. 50, n. 1, p. 63-68, Feb. 1996.

LEHMER, D. H. In **proceedings of the second symposium on large-scale digital computing machinery**. Cambridge: Harvard University Press, , 1951.

LEONEL, A C. **Combinação de probabilidade na comparação em análise conjunta de experimentos**. Lavras: UFLA-DEX, 2000, 81p. (Dissertação – Mestrado em Estatística e Experimentação Agropecuária).

MANN, H. B.; WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. **Annual Mathematics and Statistical**, Hayward, v.18, p.50–60, 1947.

MEYER, P. L. **Probabilidade: Aplicações à Estatística**. Tradução do prof. Ruy de C. B. Lourenço Filho. 2.ed. Rio de Janeiro: LTC, 1984. 426p.

MCCULLOUGH, R. S. , GURLAND, J. ; ROSENBERG, L. Small sample behaviour of certain tests of the hypothesis of equal means under variance heterogeneity. **Biometrika**, London, v.47, p.345–353, 1960.

- MILLER, I. ; MILLER, M. **John E. Freund's Mathematical Statistics.** 6^o ed. New Jersey: Prentice Hall, 1999. 624 p.
- MOOD, A M.; GRAYBILL, F.A; BOES, D. C. **Introduction to theory of statistics.** 3. ed. New York: Wiley & Sons, 1974. 564 p.
- MOOD, A M On the asymptotic efficiency of certain non-parametric two-sample tests. . **Annual Mathematics and Statistical**, Hayward, n.25, p. 514-522, 1954.
- NAYLOR, T. H.; BALINTFY. J.L.; BURDICK. D.S.; CHU, K. **Técnicas de simulação em computadores.** Petrópolis: Vozes, 1971, 402 p.
- NEYMAN, J.; PEARSON, E. S. On the use and interpretation of certain test criteria for purposes of statistical inference. **Biometrika.** London, v.20A, p.175-240 e p.263-294, 1928.
- NOETHER, G. E. Nonparametrics: the early years – impressions and recollection. **The American Statistician**, Alexandria, n.38, p. 173-178, 1984.
- PESSOA, D. **Estatística Não-Paramétrica.** Poços de Caldas: IMPA:, 1977, 195 p. Colóquio Brasileiro de Matemática, 11.
- PERECIN, D.; BARBOSA, J. C. Uma avaliação de seis procedimentos para comparações múltiplas. **Revista de Matemática e Estatística.** Marília – SP, v.6, p.95-103, 1988.
- PITMAN, E. J. G. **Notes on non-parametric statistical inference.** New York: Columbia University, 1949.
- PRATT, J. W. Remarks on zeroes and ties in the Wilcoxon signed rank procedures. **Journal of the American Statistical Association**, Washington, 54, p. 655-667, 1959.
- SACHS, L. **Applied Statistics: a handbook of techniques.** New York: Springer-Verlag, 1982.
- SANTOS, C. **Novas alternativas de testes de agrupamento avaliadas por meio de simulação de Monte Carlo.** 2000, 85p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) Universidade Federal de Lavras, Lavras.

- SATTERTHWAITE, F. E. Na approximate distribution of estimates of variance components. **Biometric Bulletin**, London, v.2, p.110-114, 1946.
- SAVAGE, L. J. The foundations of statistics reconsidered. **Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability**, Berkeley : University of California Press, 1961.
- SCHEFFÉ, H. Statistical inference in the nonparametric case, **Annual Mathematics and Statistical**, Hayward, v.10, p.129–137, 1949.
- SCHEFFÉ, H. On solutions of the Behrens-Fisher problem based on t-distribution, **Annual Mathematics and Statistical**, Hayward, v.14, p.305–332, 1943.
- SIEGEL, S. **Estatística não-paramétrica para ciências do comportamento**. São Paulo: McGraw-Hill do Brasil, 1975. 350 p.
- SILVA, E. C. **Avaliação do poder e taxa de erro tipo I do teste de Scott-Knott por meio do Método de Monte Carlo**. 1998. 54 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária)-Universidade Federal de Lavras, Lavras.
- TONHASCA, Jr. A. The three “capital sins” of statistics used in biology. **Ciência e cultura**, São Paulo, v. 43, n.6, p. 417-422, 1991.
- WELCH, B. L. The generalization of Student’s problem when several populations are involved, **Biometrika**, London, v.34, p.28–35, 1947.
- WILCOXON, F. Individual comparisons by ranking method. **Biometrics Bulletin**, London, n.1, p.80-83, 1945.
- WILCOXON, F. Some Rapid Approximate Statistical Procedures. **American Cyanamid Co. Stamford Research Laboratories**, 1949.
- ZAR, J. H. **Biostatistical Analysis**. 4. ed. New Jersey: Prentice Hall, 1999, 663 p.
- ZIMMERMANN, F. P., **Introdução aos Processos Não Paramétricos**, 34ª RBRAS, ESAL, 1989, 103p.