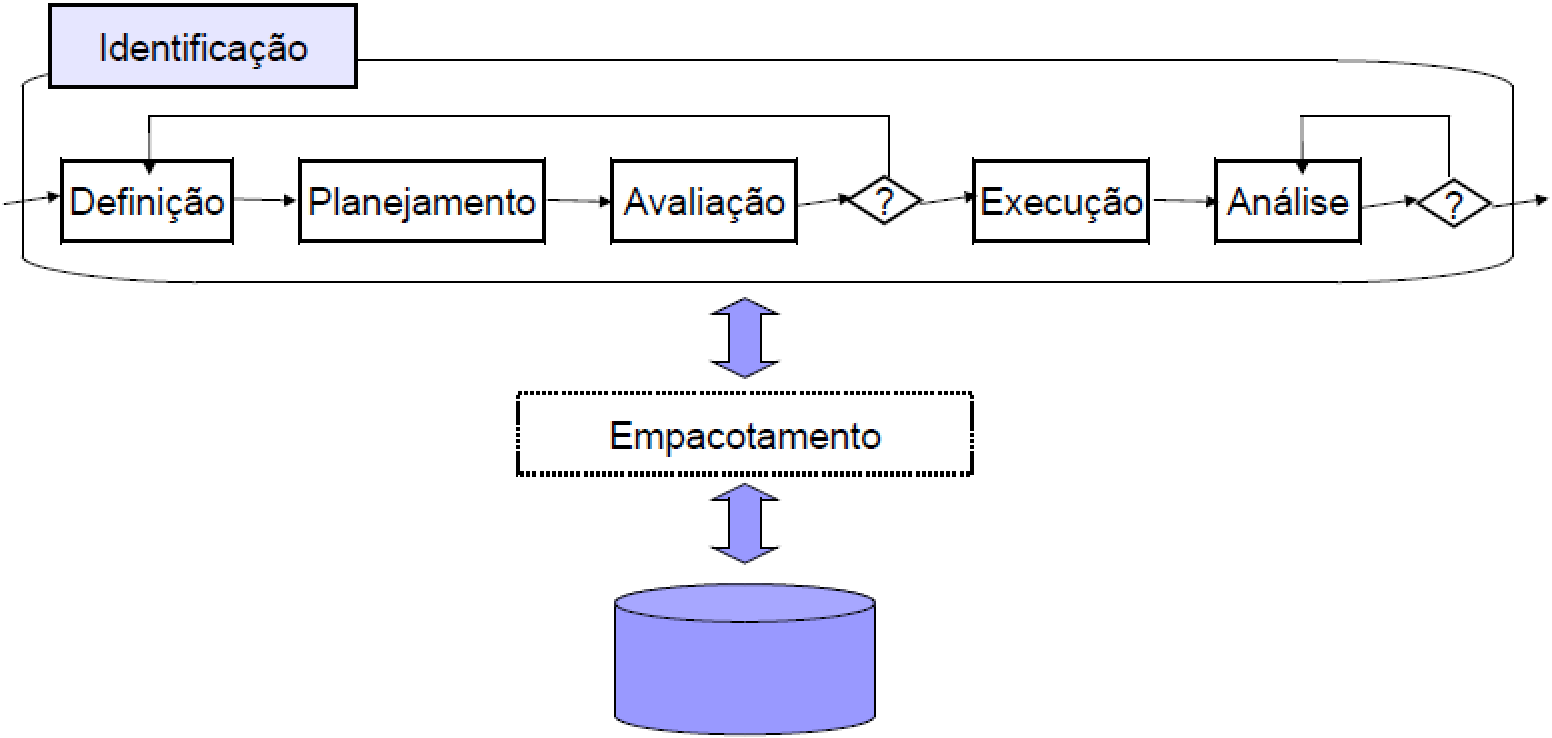
INTRODUÇÃO À ANALISE DE DADOS ESE

Prof. Paulo C. Masiero

1º. 2017

Processo de Experimentação



Análise dos Resultados

- Se possível, entreviste os participantes para obter feedback:
  - Sobre os artefatos
  - Sobre o processo experimental
  - Para capturar sua impressão sobre os resultados
- Revise os dados coletados para verificar se eles são úteis e válidos
- Organize os dados em conjuntos para análise de validade, exploração e teste das hipóteses
- Analise os dados com base em princípios estatísticos válidos
- Verifique se as hipóteses são aceitas ou rejeitadas
- O processo de análise pode ser iterativo.

Análise e interpretação dos resultados: três atividades principais

- Redução do conjunto de dados,
- Análise descritiva dos dados,
- Análise das hipóteses,

Redução do conjunto de dados

- Todos os métodos estatísticos dependem da qualidade dos dados usados.
- Se os dados não representam o que nós pensamos que eles representam, então a conclusão que obtemos dos resultados dos tratamentos não são corretas
- Os erros podem ocorrer de forma sistemática ou como outliers.
- Redução do conjunto de dados é relacionado com validação dos dados.

Outliers

- Diagramas de espalhamento são efetivos para identificar outliers, assim como box plots.

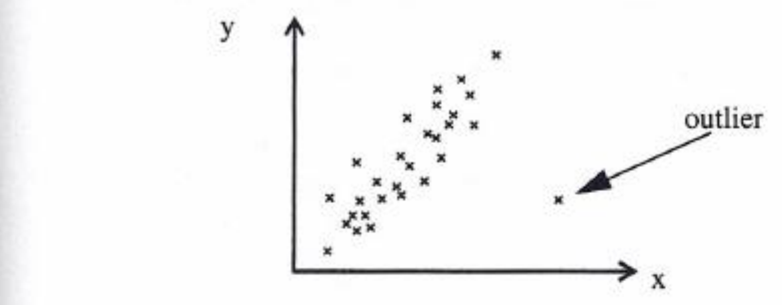


Figure 22. *An outlier detected in a scatter plot.*

Outliers (Cont.)

- Os outliers devem ser identificados com base na execução do experimento na forma dos dados coletados, do conjunto dos dados e da análise descritiva.
- Quando outliers são identificados, o importante é decidir o que fazer com eles, analisando também porque eles ocorreram.

Outliers: diretrizes

- Um evento estranho ou raro que não deve ocorrer novamente : exclua o dado da amostra.
  - Ex. Dado não entendido, ou errado
- Um evento raro que pode ocorrer novamente; é mais sensato não excluir, pois o outlier tem uma informação.
  - Ex. Resultado de sujeito inexperiente.
  - Se essa variável não foi considerada antes (ex. experiência, pode-se também dividir a amostra com base nela e fazer duas (ou mais) análises. Isso deve ser feito caso a caso.

Outliers: diretrizes

- Não são apenas dados inválidos que podem ser retirados da amostra. A redundância dos dados também pode ser analisada.
- Muitas vez não é efetivo analisar dados redundantes se forem muitos.
  - Técnicas para identificar redundância são a análise de componentes e identificação de fatores ortogonais (não tratadas no livro)

Análise Descritiva

- Estatística Descritiva
  - Medidas de Tendência Central (média, mediana, moda)
  - Medidas de dispersão (desvio padrão, variância)
  - Correlações (Pearson, Spearman)
- Análise Gráfica
  - Diagramas de dispersões
  - Histogramas e Gráficos de Pizza
  - Box Plots

Metas da Análise Descritiva

– Identificar tendências centrais das variáveis e seus tratamentos

– Identificar o grau de dispersão

– Identificar pontos fora da curva (outliers)

– Identificar Correlações

Estatísticas relevantes por tipo de escala

Tipo escala	Med. Tend. Central	Dispersão	Dependência
Nominal	Moda	Frequência	
Ordinal	Média, percentil	Intervalo de variação	Coefs. de correção de Spearman e de Kendall
Intervalar	Média	Desvio padrão, variância, range	Coef. de correção de Pearson
Razão	Média geométrica	Coeficiente de variação	

Medidas de Tendência Central

- Média aritmética:

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$$

  – Ex. para o conjunto de dados (1,1,2, 4), X=2
  – A média corresponde ao percentil 50%, indicando que 50% das amostras estão abaixo da média.

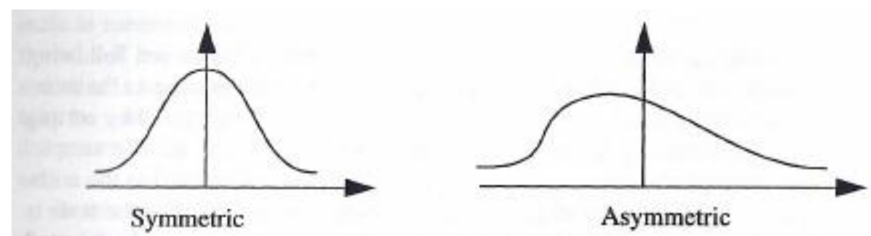
- Moda
  – Representa a amostra que ocorre mais vezes.
  – Ex. A moda para o conjunto de dados (1,1,2,4) é 1

Medidas de Tendência Central

- Mediana
  - Representa o valor médio do conjunto de dados
  - Se n é impar, pega-se a amostra média do conjunto ordenado
  - Se n é par, pode-se calcular a média das duas amostras centrais.
  - Ex. a mediana de (1,1,2,4) é 1,5. A mediana de (1,1,2,4,6) é 2

Medidas de Tendência Central

- A média e a mediana são a mesma se a distribuição é simétrica



- Média Geométrica

$$\sqrt[n]{\prod_{i=1}^{n} x_i}$$

Medidas de Dispersão

- A variância é definida como: $$s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$$
- O desvio padrão s é definido como a raiz quadrada da variância.
- Ele é geralmente preferido em relação à variância porque tem a mesma unidade de medida que os valores da amostra.
- O *range* de um conjunto de dados é a distância entre os valores máximos e mínimos do conjunto de dados.
  – Range = Xmax - Xmin

Medidas de Dispersão (Cont.)

- Intervalo de variação (Xmin,Xmax)
- Coeficiente de variação: é expresso como uma porcentagem da média:

$$100 \cdot \frac{s}{\bar{x}}$$

  – Não tem dimensão e tem significado para a escala razão.
- Uma visão geral da dispersão é dada pelo frequência de cada valor.

Medidas de Dispersão (Cont.)

- A frequência relativa é calculada dividindo-se cada frequência pelo número total da amostra.
  - Ex (1,1,1,2,2,3,4,4,4,5,6,6,7), com tamanho 13. A frequência relativa dos valores é
    - 1, 23%
    - 2,15%
    - 3, 8%
    - etc.

Medidas de Dependência: Regressão

- Se o conjunto de dados contém variáveis estocásticas X e Y em pares (X_i,Y_i) e suspeitamos que há uma função $y= f(x)$ que relaciona os pares x e y.
- Regressão significa ajustar os pontos de dados a uma curva.

Regressão Linear

- Se y= α+βx então dizemos que a regressão é linear.
- Somatórias :

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \left(\sum_{i=1}^{n} x_i y_i\right) - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)$$

- The sums can be used to compute the regression line $y = \bar{y} + \beta(x - \bar{x})$ where the slope of the line is:

$$\beta = \frac{S_{xy}}{S_{xx}}$$

and the line crosses the y-axis at $\alpha = \bar{y} - \beta\bar{x}$.

Regressão não linear

- Para dependências não lineares, é possível conseguir uma transformação para transformá-la em linear.
- Por exemplo, uma relação exponencial pode ser transformada em uma logarítmica e depois usa-se regressão linear para calcular os parâmentros da linha.

Covariância

- Um número único que quantifica o quanto dois conjuntos de dados Xi e Yi variam juntos é a covariância: Cxy=Sxy/(n-1)
- É útil para escalas intervalar e razão.
- A covariância é dependente da variância de cada variável e pode ser normalizada para permitir comparar a dependência entre diferentes variáveis relacionadas.
- Esse valor é chamado de coeficiente de correlação de Pearson (r).

Pearson

$$r = \frac{c_{xy}}{s_x \cdot s_y} = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}} = \frac{\left(n \cdot \sum_{i=1}^{n} x_i y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{\left(n \cdot \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right) \cdot \left(n \cdot \sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right)}}$$

- O valor de r fica entre -1 e +1
- Se não há correlação r=0
- O Contrário não é verdadeiro. Pode haver uma correlação não linear mesmo se r=0.

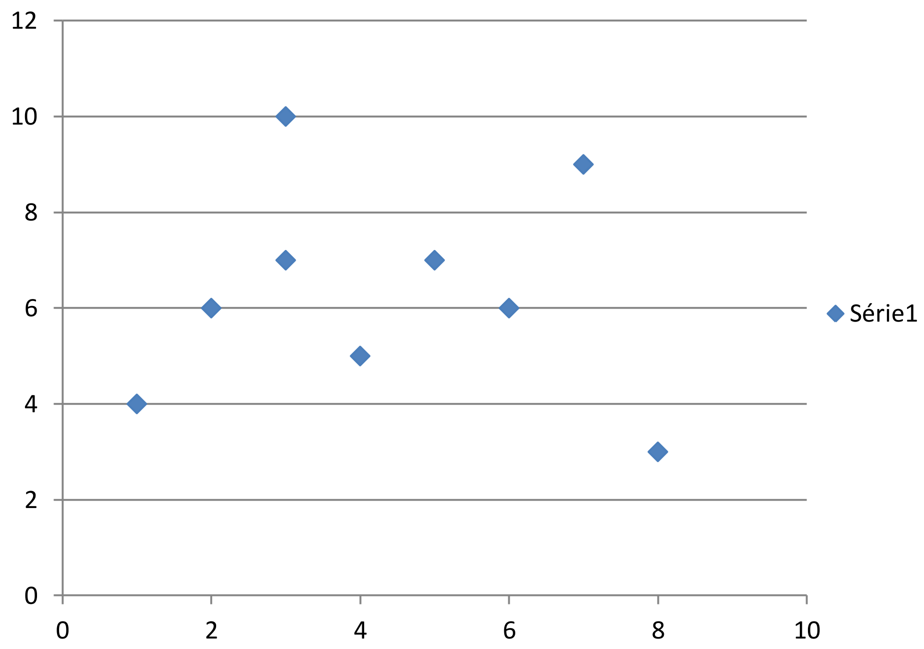
Regressão Linear

- Se a escalar é ordinal ou se o conjunto de dados não tem uma distribuição normal, o coeficiente de correlação de Spearman pode ser usado (Rs)
- O cálculo é feito da mesma forma mas os ranks (isto é, os números ordinais da amostra ordenada) é que são usados, ao invés dos valores da amostra

Visualização gráfica

- As descrições de conjuntos de dados, medidas quantitativas de tendência central, dispersão e dependência podem (devem) ser combinadas com técnicas de visualização gráfica.
- Gráficos são muito ilustrativos e mostram uma boa visão geral do conjunto de dados.

Gráfico de Dispersão



Bom para visualizar dependências entre variáveis

Visualização gráfica (Histrograma)

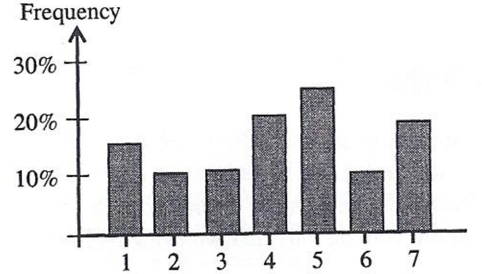
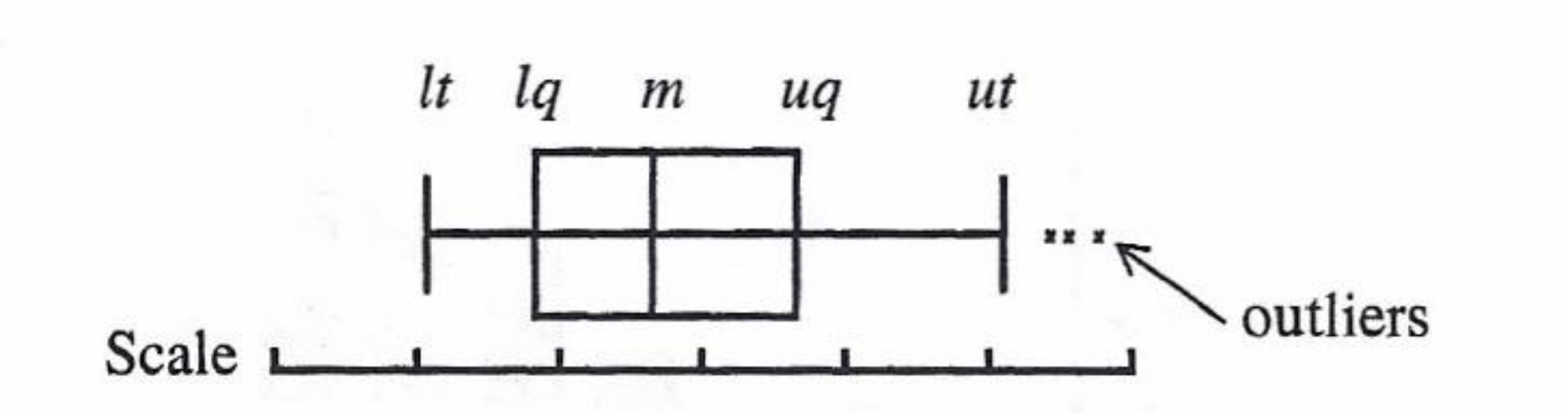


Figure 19. *A histogram.*

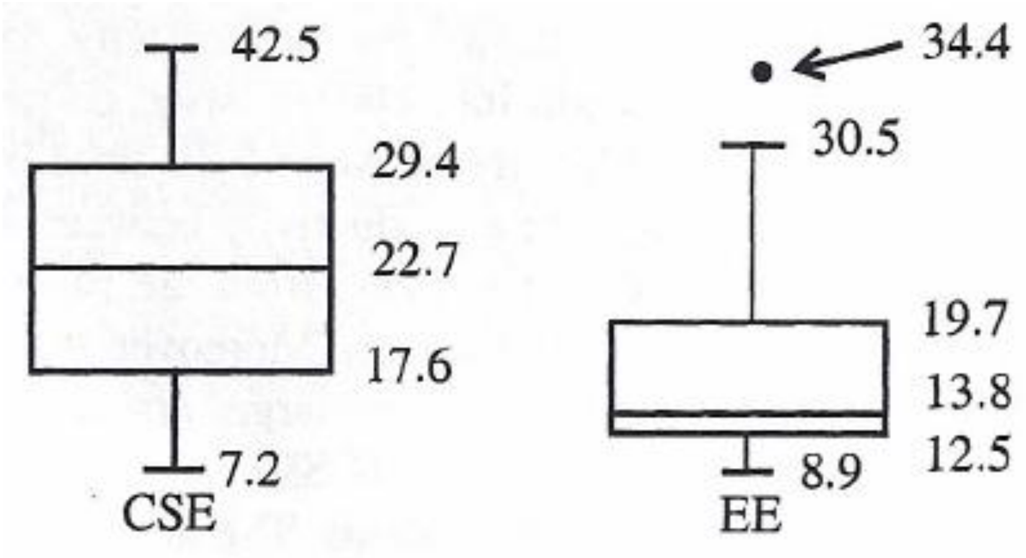
Mostra a densidade da distribuição
para uma variável

Box Plot

Bom para visualizar a dispersão
e a assimetria da amostra



m = mediana
lq= 25% (média dos valores menores que m)
uq = 75% (média dos valores maiores que m)
O tamanho da caixa é d = lq – uq
ut= uq+1.5d lt= lq – 1.5d (truncados para o ponto de dado mais próximo, se necessário)

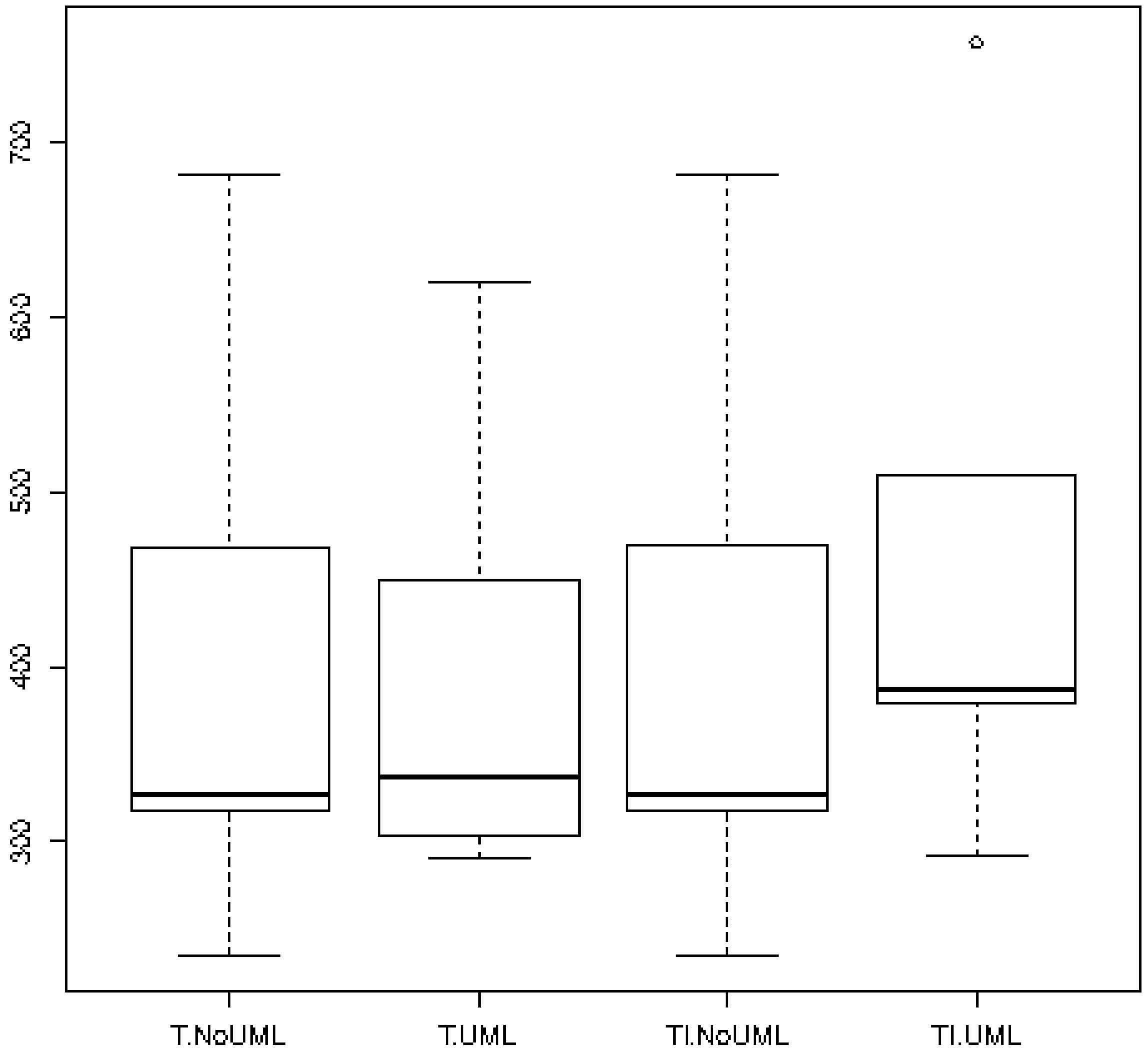


d= 29.4-17.6=11.8
lq= 17.6-1.5*11.8=-0,1
ut= 29.4+1.5*11.8=47..1
lt e uq são truncados para os pontos de dados mais próximos.

d=19.7-12.5=7.2
lq=12.5-1.5*7.2=1.7
ut=19,7+1.5*7.2=30.5
lt truncado para o ponto mais próximo

Task	Var	Treat	Mean	Std Dev	Min	Lower Quart	Med	Upper Quart	Max	% diff	Coh-en's d	*t* test	Wilc-oxon
All	T	No UML	2030.3	921.6	950.0	1318.0	1913.0	2759.0	3458	1.4%	0.04	0.47	0.46
		UML	2001.6	602.8	1112.0	1618.8	2074.1	2429.0	2838				
	T'	No UML	2030.3	921.6	950.0	1318.0	1913.0	2759.0	3458	-14.5%	-0.36	0.43	0.35
		UML	2325.4	697.1	1410.0	1855.0	2314.0	2849.0	3610				
1	T	No UML	327.1	145.0	145.0	183.0	345.0	355.0	565	6.8%	0.16	0.37	0.41
		UML	304.7	142.1	154.0	198.5	298.8	355.0	622				
	T'	No UML	327.1	145.0	145.0	183.0	345.0	355.0	565	-16.1%	-0.36	0.44	0.64
		UML	379.7	151.2	194.0	265.0	331.5	480.0	685				
2	T	No UML	234.9	117.8	130.0	140.0	172.0	332.0	445	-23.3%	-0.39	0.81	0.88
		UML	289.7	161.9	173.5	210.0	220.0	315.0	721				
	T'	No UML	234.9	117.8	130.0	140.0	172.0	332.0	445	-24.0%	-0.40	0.39	0.23
		UML	291.2	162.3	180.0	210.0	222.5	315.0	725				
3	T	No UML	681.4	417.9	323.0	410.0	533.5	765.0	1471	9.0%	0.19	0.34	0.29
		UML	620.4	196.3	337.5	479.0	646.0	722.0	894				
	T'	No UML	681.4	417.9	323.0	410.0	533.5	765.0	1471	-11.1%	-0.22	0.62	0.25
		UML	757.3	237.6	435.0	535.0	800.0	910.0	1125				
4	T	No UML	318.0	157.8	135.0	155.0	325.5	477.0	552	-6.0%	-0.13	0.61	0.63
		UML	337.1	132.1	184.0	230.0	303.0	440.0	542				
	T'	No UML	318.0	157.8	135.0	155.0	325.5	477.0	552	-21.7%	-0.45	0.33	0.35
		UML	387.0	151.0	220.0	270.0	342.5	540.0	640				
5	T	No UML	468.9	295.0	197.0	220.0	377.5	639.0	1119	4.1%	0.08	0.43	0.49
		UML	449.8	200.6	140.0	301.0	454.8	550.3	746				
	T'	No UML	468.9	295.0	197.0	220.0	377.5	639.0	1119	-8.8%	-0.17	0.71	0.49
		UML	510.2	191.8	220.0	355.0	535.0	590.0	812				

Box Plot da tabela 7 de Dzidek: T e T’



Exemplo de Análise Descritiva

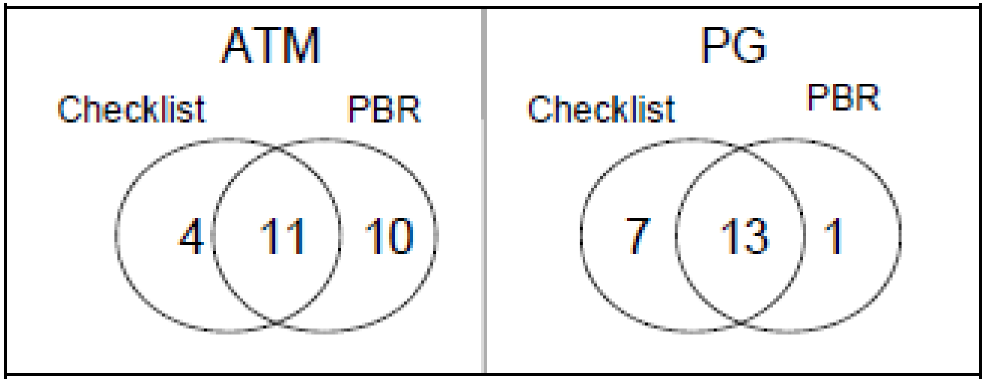
➢ **O3') Does the reviewer's experience affect his or her effectiveness?**

PBR effectiveness versus readers' role experience

Analysis

- We used a questionnaire to measure the subject's experience in their assigned perspective. The relationship between experience and effectiveness is weak
- Reviewers with more experience do not perform better than reviewers with less experience
- This conclusion is supported by the results of the Spearman's and Pearson's correlation tests that showed numbers smaller than 14%, far from indicating a high degree of correlation

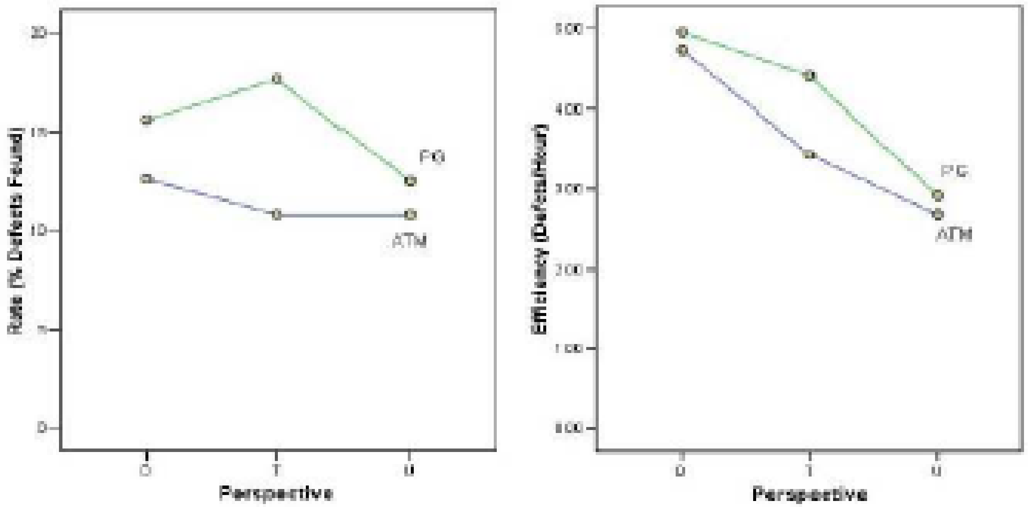
- **R1) Do individual reviewers using PBR and Checklist find different defects?**



Analysis

- **ATM**: the two techniques appear to be complementary in that users of each technique found defects that were not found by the other technique
- **PG**: the techniques do not appear to be complementary, because the PBR users only found 1 defect not found by the checklist users

R2) Do the PBR perspectives have the same effectiveness and efficiency?

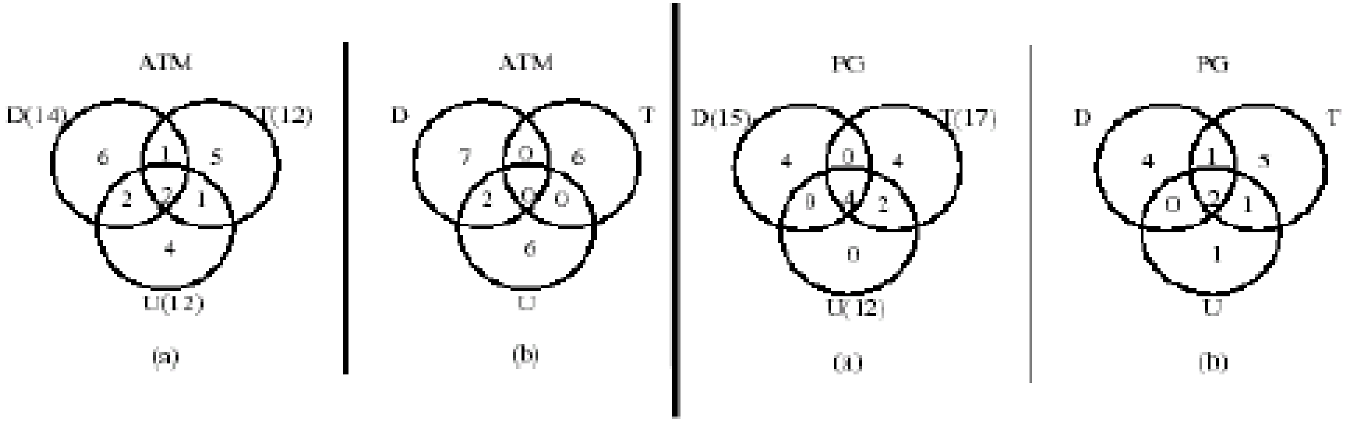


Rate and Efficiency by Perspective

Analysis

- Each point represents the mean of the 3 reviewers composing the group.
- ATM: Designer were the most effective and efficient
- PG: Tester were the most effective and Designer were the most efficient
- The perspectives had no significant effect on either effectiveness (p=.654) or efficiency (p=.182)

➢ R3) Do the PBR perspectives find different defects?



Analysis

- ATM:
  - each perspective identified unique defects with little overlap
  - the three perspectives were more likely to find different defects
  - The perspectives identified a similar number of occurrences overall
- PG:
  - the Designer and Tester perspectives appear to be complementary, but the User perspective does not provide much added benefit
  - The perspectives identified a similar number of occurrences

Testes de Hipóteses

- Hipóteses avaliadas por testes estatísticos definidos por pesquisadores da estatística inferencial
- Normalmente são definidas duas hipóteses:
  - **Hipótese nula (H0): descreve propriedades da amostra** que o experimentador deseja rejeitar com a maior significância possível;
  - **Hipótese alternativa (H1): é a hipótese inversa à hipótese nula, que** será aceita caso a hipótese nula seja rejeitada.
- Os testes estatísticos verificam se é possível rejeitar a hipótese nula, de acordo com um conjunto de dados observados e suas propriedades estatísticas. Portanto, ela deve ser formulada de forma negativa.

Testes de Hipótese

- Os testes comparam médias entre grupos de participantes realizando tratamentos diferentes

"Utilizando a técnica XYZ, os desenvolvedores concluem a atividade de projeto em menos tempo do que utilizando a técnica ABC"

Hipótese Nula: $\mu\ (Tempo_{XYZ}) = \mu\ (Tempo_{ABC})$

Hipótese Alternativa: $\mu\ (Tempo_{XYZ}) \neq \mu\ (Tempo_{ABC})$

?????

Teste Estatístico

- Calculados fundamentalmente a partir de uma função de teste que considera três valores:
  - Diferença entre os valores “médios” das estatísticas para os tratamentos
  - “Dispersão” dos valores da estatística
  - Número de amostras
- A função de teste, F(m,σ,N), depende do:
  - tipo de distribuição dos dados, e.x., normalidade e homocedasticidade.
  - Número de fatores e tratamentos.

Homogeneidade da variância

Exemplo

- Quer-se testar a Hipótese “homens são mais altos que mulheres”
  - Determina-se uma amostra da população utilizando um fator e dois tratamentos

Idade ?

Qual é ?

  - A certeza depende de:
    - Número de pessoas amostradas,
    - Diferença entre a altura média nos tratamentos,
    - Dispersão da altura nos tratamentos.

Tipos de Erros

- A verificação das hipóteses sempre lida com o risco de um erro de análise acontecer
  - O erro do tipo I (a) acontece quando o teste estatístico indica um relacionamento entre causa e efeito e o relacionamento real não existe
  - O erro do tipo II (b) acontece quando o teste estatístico não indica o relacionamento entre causa e efeito, mas existe este relacionamento

α = P (erro-tipo-I) = P (H_{NULA} é rejeitada | H_{NULA} é verdadeira)

β = P (erro-tipo-II) = P (H_{NULA} não é rejeitada | H_{NULA} é falsa)

Nível de Significância

- Indica a probabilidade de se cometer um erro do tipo I:
  - Os níveis de significância (α) mais comumente utilizados são 10%, 5%, 1% e 0.1%,
  - Chama-se de *p-value o menor nível de significância com que se* pode rejeitar a hipótese nula,
  - Dizemos que há significância estatística quando o *p-value é menor* que o nível de significância adotado.

Área crítica

- Para testar Ho, uma unidade de teste (t) é definida e uma área crítica C é definida.
- E também é definida em que parte da área t varia.
- Então: Se $t \in C$, rejeite Ho; Se $t \sim \in C$, não rejeite Ho.
- Exemplo:
  - Intervalo one-sided: $t \leq a$ ou $t \geq b$
  - Intervalo two-sided: $(t \leq a, t \geq b)$ com $a < b$

Procedimento para o Teste de Hipótese

- Fixar o nível de significância do teste (α).
- Obter uma estatística (estimador do parâmetro que se está testando) que tenha distribuição conhecida sob Ho.
- A estatística de teste e o nível de significância constroem a região crítica pela qual o teste passa.
- Usando as informações amostrais, obter o valor da estatística (estimativa do parâmetro).
- Se valor da estatística pertencer à região crítica, rejeita-se a hipótese nula, aceitando-se a hipótese alternativa.
- Caso contrário, não se rejeita a hipótese nula e nada se pode dizer a respeito da hipótese alternativa.

Teste de Hipótese na Prática

- Escolhe-se a estatística de teste,
- Escolhe-se o valor de p (p-value) (significância),
- Usa-se uma ferramenta estatística para aplicar o teste e verificar (estimar) o valor de p.

Teste de Hipótese: Tipos

- A escolha do teste depende da determinação do tipo de distribuição dos dados e de quantos fatores e tratamentos vão ser analisados no teste
- <u>Testes paramétricos (TP)</u>: assumem uma distribuição e são mais poderosos
- <u>Testes não paramétricos</u>: não assumem uma distribuição e têm uma aplicação mais abrangente.

Teste de Hipótese: Escolha

- Há dois fatores a considerar
- Aplicabilidade: É muito importante que as suposições relativas à distribuição dos parâmetros sejam realísticas. Geralmente a suposição é que a distribuição seja normal.
- Potência: TP é geralmente mais potente, isto é, requer menos pontos de dado se a suposição for verdadeira.

Alguns Tipos de Teste

Table 10.3 Overview of parametric/non-parametric tests for different designs

Design	Parametric	Non-parametric
One factor, one treatment		Chi-2, Binomial test
One factor, two treatments, completely randomized design	t-test, F-test	Mann-Whitney, Chi-2
One factor, two treatments, paired comparison	Paired t-test	Wilcoxon, Sign test
One factor, more than two treatments	ANOVA	Kruskal-Wallis, Chi-2
More than one factor	ANOVA[a]	

[a] This test is not described in this book. Refer instead to, for example, Marascuilo and Serlin [119] and Montgomery [125]

Teste t (Student)

- Usado para comparar duas amostras <u>independentes</u> (um fator com dois tratamentos).
- Pode ser usado com diferentes suposições.
- No exemplo usa-se uma suposição comum, que ocorre com alta frequência

t-test	
Input	Two independent samples: $x_1, x_2, \ldots x_n$ and $y_1, y_2, \ldots y_m$.
H_0	$\mu_x = \mu_y$, i.e. the expected mean values are the same.
Calculations	Calculate $t_0 = \dfrac{\bar{x} - \bar{y}}{S_p\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}}$, where $S_p = \sqrt{\dfrac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}}$, and, S_x^2 and S_y^2 are the individual sample variances.
Criterion	Two sided (H_1: $\mu_x \neq \mu_y$): reject H_0 if $\lvert t_0 \rvert > t_{\alpha/2,\, n+m-2}$. Here, $t_{\alpha,\, f}$ is the upper α percentage point of the t distribution with f degrees of freedom, which is equal to $n+m-2$. The distribution is tabulated in, for example, Table A1 and [Montgomery97, Marascuilo88]. One sided (H_1: $\mu_x > \mu_y$): reject H_0 if $t_0 > t_{\alpha,\, n+m-2}$.

Example of t-test. The defect densities in different programs have been compared in two projects. In one of the projects the result is $x = \{3.42, 2.71, 2.84, 1.85, 3.22, 3.48, 2.68, 4.30, 2.49, 1.54\}$ and in the other project the result is $y = \{3.44, 4.97, 4.76, 4.96, 4.10, 3.05, 4.09, 3.69, 4.21, 4.40, 3.49\}$. The null hypothesis is that the defect density is the same in both projects, and the alternative hypothesis that it is not. Based on the data it can be seen that $n = 10$ and $m = 11$. The mean values are $\bar{x} = 2.853$ and $\bar{y} = 4.1055$.

It can be found that $S_x^2 = 0.6506$, $S_y^2 = 0.4112$, $S_p = 0.7243$ and $t_0 = -3.96$.

The number of degrees of freedom is $f = n+m-2 = 10+11-2 = 19$. In Table A1, it can be seen that $t_{0.025,\ 19} = 2.093$. Since $|t_0| > t_{0.025,\ 19}$ it is possible to reject the null hypothesis with a two tailed test at the 0.05 level.

Table A1. *Critical values two-tailed t-test (5%), see Section 8.3.4 and 8.3.7.*

Degrees of freedom	t-value
1	12.706
2	4.303
3	3.182
4	2.776
5	2.571
6	2.447
7	2.365
8	2.306
9	2.262
10	2.228
11	2.201
12	2.179
13	2.160
14	2.145
15	2.131
16	2.120
17	2.110
18	2.101
19	2.093
20	2.086
21	2.080
22	2.074
23	2.069
24	2.064
25	2.060

t teste pareado

- É usado quando duas amostras obtidas com <u>medidas repetidas </u>são comparadas.
- Isso significa que, por exemplo, as medidas foram obtidas de um mesmo sujeito mais do que uma vez.
- Exemplo: Dez programadores desenvolveram independentemente dois programas diferentes.
  - Ho: o tempo requerido para desenvolver P1 é o mesmo que para P2.

t teste pareado

Table 10.8 Paired t-test

Item	Description
Input	Paired samples: $(x_1, y_1), (x_2, y_2) \ldots (x_n, y_n)$
H_0	$\mu_d = 0$, where $d_i = x_i - y_i$, i.e. the expected mean of the differences is 0
Calculations	Calculate $t_0 = \dfrac{\bar{d}}{S_d/(\sqrt{n})}$, where $S_d = \sqrt{\dfrac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n-1}}$
Criterion	Two sided ($H_1 : \mu_d \neq 0$): reject H_0 if $\lvert t_0 \rvert > t_{\alpha/2,n-1}$. Here, $t_{\alpha,f}$ is the upper α percentage point of the t distribution with f degrees of freedom. The distribution is tabulated, for example, in Table B.1 and by Montgomery [125], and Marascuilo and Serlin [119]
	One sided ($H_1 : \mu_d > 0$): reject H_0 if $\lvert t_0 \rvert > t_{\alpha,n-1}$

Table 10.9 Required effort

Programmer	1	2	3	4	5	6	7	8	9	10
Program 1	105	137	124	111	151	150	168	159	104	102
Program 2	86.1	115	175	94.9	174	120	153	178	71.3	110

d={18.9, 22, -51,16.1,23, 30,15,19,32,7,9}

Sd=27.358

to=0.39

Table B.1 Critical values two-tailed t-test (5%), see Sects. 10.3.4 and 10.3.7

Degrees of freedom	t-value
1	12.706
2	4.303
3	3.182
4	2.776
5	2.571
6	2.447
7	2.365
8	2.306
9	2.262
10	2.228
11	2.201
12	2.179
13	2.160
14	2.145
15	2.131
16	2.120
17	2.110
18	2.101
19	2.093
20	2.086
21	2.080
22	2.074
23	2.069
24	2.064
25	2.060
26	2.056
27	2.052
28	2.048
29	2.045
30	2.042
40	2.021
60	2.000
120	1.980
∞	1.960

$f=n-1=10-1=9$

Na tabela B1, $t_{0.025,9} = 2.262$

Portanto, como $t_0=0{,}39$ é menor que 2.262, não é possível rejeitar a hipótese nula.

ANOVA (ANalysis Of VAriance)

- Usado para avaliar experimentos com várias quantidades de projetos.
- Baseado na análise da variabilidade total dos dados e da variabilidade de uma partição de acordo com diferentes componentes.
- Na sua forma mais simples compara a variabilidade devida ao tratamento com a variabilidade devida a erros randômicos.

Este exemplo mostra a forma mais simples: compara se as amostras têm o mesmo valor médio, isto é, o projeto tem um fator com mais de dois tratamentos.

ANOVA, one factor, more than two treatments	
Input	*a* samples: $x_{11}, x_{12}, \ldots x_{1n_1}; x_{21}, x_{22}, \ldots x_{2n_2}; \ldots; x_{a1}, x_{a2}, \ldots x_{an_a}$
H_0	$\mu_{x_1} = \mu_{x_2} = \ldots = \mu_{x_a}$, i.e. all expected means are equal
Calculations	Calculate: $SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n_i} x_{ij}^2 - \frac{x_{..}^2}{N}$ $SS_{Treatment} = \sum_{i=1}^{a} \frac{x_{i.}^2}{n_i} - \frac{x_{..}^2}{N}$ $SS_{Error} = SS_T - SS_{Treatment}$ $MS_{Treatment} = SS_{Treatment}/(a-1)$ $MS_{Error} = SS_{Error}/(N-a)$ $F_0 = MS_{Treatment}/MS_{Error}$ where *N* is the total number of measurements and a dot index denotes a summation over the dotted index, e.g. $x_{i.} = \sum_j x_{ij}$
Criterion	Reject H_0 if $F_0 > F_{\alpha, a-1, N-a}$. Here, F_{α, f_1, f_2} is the upper α percentage point of the F distribution with f_1 and f_2 degrees of freedom, which is tabulated in, for example, Table A5.1, Table A5.2 and [Montgomery97, Marascuilo88].

a = no. de tratamentos

Pode ser usado para outras formas de projeto: com diferentes fatores, blocagem, etc.

Table 22. *ANOVA table for the ANOVA test described above.*

Source of variation	**Sum of squares**	**Degrees of freedom**	**Mean square**	$\mathbf{F_0}$
Between treatments	$SS_{Treatment}$	a–1	$MS_{Treatment}$	$F_0 = \frac{MS_{Treatment}}{MS_{Error}}$
Error[a]	SS_{Error}	N–a	MS_{Error}	
Total	SS_T	N–1		

a. This is sometimes denoted within treatments.

Exemplo

Example of ANOVA. The module sizes in three different programs have been measured. The result is:

Program 1: 221, 159, 191, 194, 156, 238, 220, 197, 197, 194
Program 2: 173, 171, 168, 286, 206, 140, 226, 248, 189, 208, 213
Program 3: 234, 188, 181, 207, 266, 153, 190, 195, 181, 238, 191, 260

The null hypothesis is that the mean module size is the same in all three programs. The alternative hypothesis is that it is not. Based on the data above the ANOVA table in Table 23 can be calculated.

Table 23. *ANOVA table.*

Source of variation	Sum of squares	Degrees of freedom	Mean square	F_0
Between treatment	579.0515	2	289.5258	0.24
Error	36151	30	1205	
Total	36730	32		

The number of degrees of freedom are $f_1 = a-1 = 3-1 = 2$ and $f_2 = N-a = 33-3 = 30$. In Table A5.1, it can be seen that $F_{0.025,\,2,\,30} = 4.18$. Since $F_0 < F_{0.025,\,2,\,30}$ it is impossible to reject the null hypothesis at the 0.025 level.

Table A5.1. *Critical values two-tailed F-test (5%), see Section 8.3.6. For ANOVA, this is equivalent to a significance level of 2.5%, see Section 8.3.10.*

Note

f_2 \ f_1	1	2	3	4	5	6	7	8	9	10
1	648	800	864	900	922	937	948	957	963	969
2	38.5	39.0	39.2	39.2	39.3	39.3	39.4	39.4	39.4	39.4
3	17.4	16.0	15.4	15.1	14.9	14.7	14.6	14.5	14.5	14.4
4	12.2	10.6	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84
5	10.0	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37
15	6.20	4.76	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05

Teste Mann-Whitney

- Alternativa não-paramétrica para o teste t, quando a suposição usada no teste t não parece correta (incerta).
- É baseado em ranks (ordenações) – Não é mostrado de forma completa no livro.

Resumo do teste

Mann-Whitney	
Input	Two independent samples: $x_1, x_2, \ldots x_n$ and $y_1, y_2, \ldots y_m$.
H_0	The two samples come from the same distribution.
Calculations	Rank all samples and calculate $U = N_A N_B + \frac{N_A(N_A+1)}{2} - T$ and $U' = N_A N_B - U$, where $N_A = \min(n, m)$, $N_B = \max(n, m)$, and T is the sum of the ranks of the smallest sample.
Criterion	Tables providing criterion for rejection of the null hypothesis based on the calculations are available in, for example, Table A3 and [Marascuilo88.]. Reject H_0 if $\min(U, U')$ is less than or equal to the value in Table A3.

Exemplo

- Mesmo exemplo do teste t: n=10 e m=11
  - x=3.42,2.71,2.84,1.85,3.22,3.48,2.68,4.30,2.49,1.54
  - y=3.44,4.97,4.76,4,96,4.10,3.05,4.09,3.69,4.21,4.40,3.49

- Na=min(10,11)=10 e Nb=max(10,11)=11

- Os ranks da amostra menor (x) são: {9,5,6,2,8,11,4,17,3,1} e da maior (y) são: {10,21,19,20,15,7,14,13,16,18,12}

  (juntar as duas amostras e ordenar)

Exemplo (Cont.)

- T=66, U=99 e U’=11
- O mínimo entre U e U’ é 11
- Como é menor que 26 (ver tabela A3) → é possível rejeitar a hipótese nula com nível de significância 0.05

Table A3. *Critical values two-tailed Mann-Whitney (5%), see Section 8.3.5.*

N_B / N_A	5	6	7	8	9	10	11	12
3	0	1	1	2	2	3	3	4
4	1	2	3	4	4	5	6	7
5	2	3	5	6	7	8	9	11
6		5	6	8	10	11	13	14
7			8	10	12	14	16	18
8				13	15	17	19	22
9					17	20	23	26
10						23	26	29
11							30	33
12								37

Sugestão de exercício

- Instale ou use o software estatístico R. (Lab. De Estatística). Aprenda a usar as funções básicas.
- Considere os dois conjuntos de amostras do exemplo do t-test
  - Faça um box plot das duas sequências. Descubra como é calculado o “bigode” pelo R, ou se é opcional.
  - Repita o teste como no exemplo e comprove os resultados
  - Retire três amostras de cada conjunto e repita o t-test. Conclua se é possível continuar a rejeitar To com nível de significância (alfa) igual 5%.