

Análise
Quantitativa
em
Pesquisa em
Educação

**Fábio José da Costa
Alves**

**Admilson Alcântara
da Silva**

**Fabício Martins da
Costa**



UNIVERSIDADE DO ESTADO DO PARÁ
CENTRO DE CIÊNCIAS SOCIAIS E EDUCAÇÃO

Clay Anderson Nunes Chagas
REITOR

Ilma Pastana Ferreira
VICE-REITOR

Anderson Madson Oliveira Maia
DIRETORA DO CENTRO DE CIÊNCIAS SOCIAIS E EDUCAÇÃO

Fabrcio Martins da Costa
CHEFE DO DEPARTAMENTO DE MATEMÁTICA, ESTATÍSTICA E INFORMÁTICA.

Carlos Alberto de Miranda Pinheiro
COORDENADOR DO CURSO DE MATEMÁTICA

Pedro Franco de Sá
PROGRAMA DE PÓS-GRADUAÇÃO EM ENSINO DE MATEMÁTICA

Marta Genu Soares
PROGRAMA DE PÓS-GRADUAÇÃO EM EDUCAÇÃO



GPEMT
GRUPO DE PESQUISA EM ENSINO DA
MATEMÁTICA E TECNOLOGIAS

ALVES, Fábio José Costa da; SILVA, Admilson Alcântara da; COSTA, Fabrício Martins da. Análise Quantitativa em Pesquisa em Educação. Grupo de Pesquisa em Ensino da Matemática e Tecnologias, Universidade do Estado do Pará (UEPA), Belém-Pa, 2023.

ISBN: 978-65-84998-58-2

Análise Quantitativa; Pesquisa em Educação; Método Estatístico.

SUMÁRIO

	APRESENTAÇÃO	05
1.	CONFIABILIDADE E O ALFA DE CRONBACH	06
2.	COEFICIENTE DE CORRELAÇÃO DE PEARSON	14
3.	INTERVALO DE CONFIANÇA	23
4.	TESTE DE HIPÓTESE	35
5	ANOVA	44
6.	QUI-QUADRADO	58
7.	ANEXOS	72

APRESENTAÇÃO

Prezado(a) Leitor(a)

Esta publicação se constitui em um importante recurso didático, posto à disposição dos alunos para a análise de dados relacionados as pesquisa em educação, tendo como finalidade apresentar, de forma didática e objetiva, como determinar a confiabilidade de formulário e métodos estatísticos para analisar e comparar por meio desempenho em grupo de alunos e unidades escolares.

O material está organizado em seis capítulos. O primeiro capítulo apresenta o Alfa de Cronbach, usado na avaliação de confiabilidade de um questionário aplicado em uma pesquisa. O segundo capítulo apresenta coeficiente de correlação de Pearson a fim de medida do grau de relação linear entre duas variáveis quantitativas.

No terceiro capítulo trataremos do intervalo de confiança, onde a média de um parâmetro de uma amostra tem uma dada probabilidade de ocorrer ou o intervalo onde há 95% de probabilidade da média verdadeira da população inteira ocorrer. Já no quarto capítulo apresentaremos o Teste de Hipótese para analisar suposição sobre desempenho de turmas e unidades escolares.

O quinto capítulo apresenta a ANOVA ou Análise de Variância que é um procedimento usado para comparar a distribuição de três ou mais grupos em amostras independentes. Já no sexto capítulo tratamos do método Qui-Quadrado que é um teste de hipóteses, não paramétrico, pois não depende de parâmetros populacionais (média e variância) que se destina a encontrar um valor da dispersão para duas variáveis categóricas nominais e avaliar a associação existente entre variáveis qualitativas.

Na abordagem de cada método, é apresentada exemplos práticos com o detalhamento da solução de cada problema e sua respectiva análise,

Os Autores

1. CONFIABILIDADE E O ALFA DE CRONBACH

Nas pesquisa em educação é muito comum o uso de questionários, como instrumento de coleta de dados, por vários motivos entre eles podemos citar o baixo custo, a garantia do anonimato dos informantes, possibilitar um maior tempo para os informantes pensarem sobre as respostas, por facilitar a padronização que garante a uniformidade das questões que poder ser objetivas e de fácil pontuação, o que facilita a conversão dos dados para seu processamento.

Devido sua importância devemos ter alguns cuidados na construção do questionário, começando com a sequência lógica das perguntas, nos tipos e opções de respostas. A disposição das opções devem apresentar um padrão visual único em todo o documento, seja na construção do cabeçalho, na apresentação dos conteúdos, na sinalização das respostas, no espaçamento do documento e em outros pontos para que posteriormente não causem confusão na interpretação dos dados e permitam a codificação correta dos dados;

A construção do formulário se baseia nos objetivos da pesquisa, e devem ter perguntas que ajudem a encontrar cada uma das informações necessárias. Passamos a dar algumas dicas para a elaboração de perguntas:

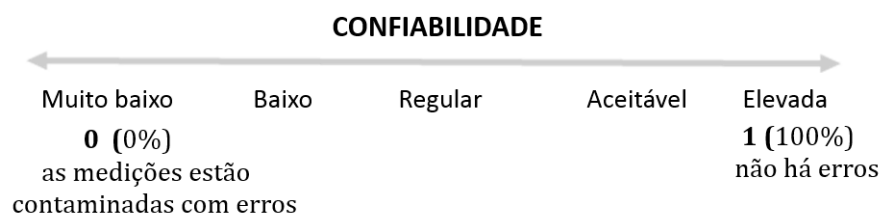
- Pergunte de forma direta, seja objetivo e sem rodeios.
- Use o português formal, sem gírias, para construir um texto claro.
- Se possível, use exemplos, imagens e até vídeos, já que estamos falando de pesquisas online.
- Empregue perguntas de forma a encerrar o questionário caso o informante não se enquadre no perfil, por exemplo: você usa esmaltes? Se o questionário for entregue a um homem.
- Antes de iniciar o questionário, contextualize o informante sobre do que se trata a pesquisa.
- Varie os tipos de perguntas, conforme seus objetivos.
- Quando as respostas avaliam o grau de contentamento ou descontentamento de algum item e julga opiniões de produtos ou serviços. O mais comum são: “concordo totalmente”, “concordo em parte”, “discordo em parte” e “discordo totalmente”.

As respostas de um formulário com quatro respostas, pode ter a seguinte parametrização:

- Discordo totalmente → 1
 Discordo em parte → 2
 Concordo em parte → 3
 Concordo totalmente → 4

Já as escalas de 0 a 10 também são bastante utilizadas, como na pesquisa do grau de satisfação do cliente.

Após construído e aplicado o questionário é importante avaliarmos a confiabilidade que pode ser definida como a estabilidade ou consistência dos resultados obtidos. Isto é, refere-se à medida em que a aplicação repetida do instrumento, ao mesmo sujeito ou objeto, produz os mesmos resultados.



Para medir a confiabilidade dos dados de um questionário utilizaremos o coeficiente alfa de Cronbach que foi apresentado por Lee J. Cronbach, em 1951, como uma forma de estimar a confiabilidade de um questionário aplicado em uma pesquisa. O Alfa é um coeficiente de correlação quadrado que mede a homogeneidade das perguntas por meio da média todas as correlações entre os itens para ver que de fato parecem. Este índice de consistência interna assume valores entre 0 e 1.

Trata-se de uma correlação média entre perguntas, dado que todos os itens de um questionário utilizam a mesma escala de medição, o coeficiente α é calculado a partir da variância dos itens individuais e da variância da soma dos itens de cada avaliador através da seguinte equação:

$$\alpha = \left(\frac{k}{k-1} \right) \times \left(1 - \frac{\sum_{i=1}^k S_i^2}{S_t^2} \right)$$

Onde:

k corresponde ao número de itens do questionário;

S_i^2 corresponde a variância de cada item;

S_t^2 corresponde a variância total do questionário, determinada como a soma de todas as variâncias.

A Tabela a seguir ilustra a aplicação passo a passo do coeficiente, onde cada coluna indica um item, cada linha indica um avaliador, e o encontro entre um item e um avaliador (X_{nk}) indica a resposta deste avaliador para este item, dentro da escala.

Tabela: Tabulação dos dados de questionário para cálculo do alfa de Cronbach.

Avaliadores	Itens						Total
	1	2	...	i	...	k	
1	X_{11}	X_{12}	...	X_{1i}	...	X_{1k}	X_1
2	X_{21}	X_{22}	...	X_{2i}	...	X_{2k}	X_2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
p	X_{p1}	X_{p2}	...	X_{pi}	...	X_{pk}	X_p
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	X_{n1}	X_{n2}	...	X_{ni}	...	X_{nk}	X_n
	S^2_1	S^2_2	...	S^2_n	...	S^2_k	S^2_t

Fonte: Adaptado de Cronbach, 2004.

Cada item, a priori, deve abordar uma única ideia de cada vez, isto é, os itens devem ser independentes. A aplicação do alfa de Cronbach contempla alguns pressupostos, tais como:

- O questionário deve estar dividido e agrupado em dimensões, ou seja, questões que tratam de um mesmo aspecto;
- O questionário deve ser aplicado a uma amostra significativa e heterogênea: Quando se elabora um questionário para especialistas, a confiabilidade não pode ser aferida internamente, pois se subentende que avaliadores especialistas tendem a ter a mesma opinião sobre o assunto em discussão, diminuindo a variabilidade total do questionário e diminuindo assim o alfa;

O modelo para estimação Alfa de Cronbach é válido para

$$\{\alpha \in \mathfrak{R} \mid -\infty < \alpha < 1 \wedge \alpha \neq 0\}$$

mas deve ser interpretado no intervalo entre 0 e 1, onde os valores negativos do alfa devem ser considerados como escalas sem confiança (ou seja, zero).

Para exemplificarmos a aplicação do Alfa de Cronbach, usaremos o dado que está No Anexo I, neste Anexo I estão as respostas de 96 alunos de um curso da área de biológica de uma universidade pública do Pará, que responderam ao questionário com

45 questões, sendo 15 referentes ao ensino, 15 relacionado a aprendizagem e 15 voltadas a avaliação. O referido questionário foi utilizado em um projeto de investigação intitulado “Avaliação, Ensino e Aprendizagem no Ensino Superior em Portugal e no Brasil: realidades e perspectivas”, parte do pressuposto de que os conhecimentos, concepções e experiências dos professores se relacionam com as formas como organizam o ensino e a avaliação do processo ensino-aprendizagem, envolvendo universidades portuguesas (Universidade de Lisboa – Instituição proponente; Universidade de Évora; Universidade do Minho e Universidade de Coimbra) e brasileiras (UEPA; USP e UNAMA)

A seguir apresentaremos um dos questionários utilizado nessa pesquisa, será objeto de nosso estudo.

Ensino

Pedimos que você reflita sobre o ENSINO. Recorra à escala apresentada, manifeste, por favor, a sua posição com base no que considera que acontece mais frequentemente no curso em que estuda. Lembrando: 1 – Discordo Totalmente. 2 – Discordo. 3 – Concordo. 4 – Concordo Totalmente.

Questões:

PE1. No início de cada ano/semestre/bloco, a maioria dos professores apresenta aos estudantes os programas de cada disciplina.

PE2. Em geral, os programas das disciplinas são claros e compreensíveis para a maioria dos estudantes.

PE3. Na maioria dos casos, os professores utilizam metodologias de ensino variadas, favoráveis ao desenvolvimento das aprendizagens previstas nos programas.

PE4. A maioria dos professores domina os conteúdos que ensina.

PE5. Em geral, os professores utilizam uma diversidade de recursos para apoiarem o desenvolvimento das aulas (ex. transparências, power point, plataforma on-line, sites da internet, modelos físicos, material bibliográfico, vídeos e programas informáticos).

PE6. A maioria dos professores proporciona oportunidades para que, nas aulas, os estudantes participem numa variedade de atividades de aprendizagem (ex. discussões, análise de textos, resolução de problemas, pesquisas, entre outros).

PE7. Os professores convidam, frequentemente, os estudantes para participarem na organização e dinamização de atividades a desenvolver nas salas de aula.

PE8. Os professores desenvolvem a maioria das aulas recorrendo a diferentes dinâmicas de trabalho (ex. trabalho em grupo, grande grupo, pequenos grupos, pares, individual).

PE9. Em geral, nas aulas, os estudantes têm oportunidades para esclarecer dúvidas, verificando se aprenderam os conteúdos previstos nos programas.

PE10. Os professores estão, normalmente, disponíveis para ajudar os estudantes a superarem dificuldades de aprendizagem para além do horário das aulas.

PE11. A maioria dos professores organiza e desenvolve o ensino de forma a que todos os estudantes tenham oportunidades para aprender o que está previsto nos programas.

PE12. Os professores utilizam, sistematicamente, os resultados da avaliação das aprendizagens dos estudantes para alterar as metodologias de ensino.

PE13. Em geral, os professores incentivam os estudantes a relacionar e a integrar conhecimentos para resolver uma diversidade de problemas.

PE14. A organização da maioria das aulas favorece o desenvolvimento das aprendizagens previstas nos programas.

PE15. Na maioria das disciplinas, as aulas desenvolvem-se com os professores expondo os conteúdos previstos nos programas enquanto os alunos vão ouvindo e/ou tomando notas.

Aprendizagem

Pedimos que você reflita sobre a APRENDIZAGEM. Recorra à escala apresentada, manifeste, por favor, a sua posição com base no que considera que acontece mais frequentemente no curso em que estuda. Lembrando: 1 – Discordo Totalmente. 2 – Discordo. 3 – Concordo. 4 – Concordo Totalmente.

Questões:

PA1. A maioria dos estudantes aprende os conteúdos constantes nos programas porque tem que estudar para as provas e/ou exames finais.

PA2. Estudar para as provas e/ou exames finais é, em geral, suficiente para que os estudantes aprendam o que está previsto nos programas.

PA3. A maioria dos estudantes aprende melhor quando os professores lhes propõem, regularmente, tarefas para serem realizadas ao longo do ano/semestre/bloco.

PA4. A maioria dos estudantes aprende melhor nas disciplinas em que professores e estudantes estabelecem relações que facilitam a cooperação no processo de aprendizagem.

PA5. A maioria dos estudantes aprende melhor quando os professores utilizam uma diversidade de dinâmicas de sala de aula (ex. trabalho em grande grupo, trabalho em pequenos grupos, trabalho em pares, trabalho individual)

PA6. A maioria dos estudantes aprende melhor quando os professores utilizam uma diversidade de recursos para ensinar (ex. materiais bibliográficos, transparências, power point, plataforma on-line, sites da internet, modelos físicos, programas informáticos).

PA7. A maioria dos professores orienta e apoia os estudantes a aprenderem de forma autônoma, para além das aulas.

PA8. Os estudantes, em geral, são adequadamente acompanhados e orientados pelos professores ao longo do processo de aprendizagem.

PA9. Em geral, os estudantes utilizam as informações decorrentes das avaliações para a melhoria das suas aprendizagens.

PA10. Para a maioria dos professores e dos estudantes, aprender significa alcançar os objetivos previstos nos programas de cada disciplina.

PA11. De modo geral, os estudantes aprendem melhor nas disciplinas em que a avaliação é um processo que se desenvolve de forma contínua ao longo do ano/semestre/bloco.

PA12. Os estudantes, em geral, envolvem-se ativamente nas tarefas de aprendizagem que lhes são propostas pelos professores.

PA13. A maioria das aulas são estruturadas de forma a criar oportunidades para os estudantes verificarem se aprenderam os assuntos previstos nos programas.

PA14. Os estudantes aprendem melhor nas disciplinas em que são, sistematicamente, incentivados pelos professores a participar em todas as atividades.

PA15. O clima existente na maioria das aulas é favorável ao desenvolvimento das aprendizagens por parte dos estudantes.

Avaliação

Pedimos que você reflita sobre a AVALIAÇÃO. Recorra à escala apresentada, manifeste, por favor, a sua posição com base no que considera que acontece mais frequentemente no curso em que estuda. Lembrando: 1 – Discordo Totalmente. 2 – Discordo. 3 – Concordo. 4 – Concordo Totalmente.

Questões:

PAv1. Na maioria das disciplinas, a avaliação desenvolve-se de maneira contínua ao longo do ano/semestre/bloco.

PAv2. Os professores, ao longo do ano/semestre/bloco, debatem com os estudantes, frequentemente, a organização e o desenvolvimento do processo de avaliação.

PAv3. Os critérios de avaliação utilizados em cada disciplina são conhecidos pela maioria dos estudantes.

PAv4. No processo de avaliação utiliza-se, com frequência, uma diversidade de tarefas e/ou instrumentos (ex. provas, testes, trabalhos, relatórios, apresentações, pesquisas).

PAv5. Na maioria das disciplinas, a avaliação é utilizada para atribuir as notas/classificações aos estudantes.

PAv6. A avaliação é utilizada, com frequência, para orientar os estudantes a aprenderem melhor.

PAv7. Os professores utilizam, sistematicamente, a avaliação para que os estudantes se conscientizem das suas dificuldades e/ou dos seus progressos na aprendizagem.

PAv8. Em geral, a avaliação realizada é consistente com o ensino e com as aprendizagens.

PAv9. Os estudantes são, frequentemente, envolvidos em situações em que têm que avaliar o seu próprio trabalho e/ou o dos seus colegas.

PAv10. Os professores, em geral, criam oportunidades para que os estudantes entendam e/ou questionem os juízos formulados acerca das suas aprendizagens.

PAv11. Na maioria das disciplinas, os resultados obtidos pelos estudantes nas provas e/ou exames finais são os elementos mais determinantes para a atribuição das suas classificações.

PAv12. Na maioria das disciplinas, as provas e/ou exames finais são a melhor forma de conhecer o que os estudantes sabem e são capazes de fazer.

PAv13. Os professores incentivam, com frequência, os estudantes a analisar e/ou refletir acerca das aprendizagens realizadas.

PAv14. De modo geral, as informações resultantes das avaliações são utilizadas pelos estudantes para orientarem e/ou reorientarem as suas formas de estudar.

PAv15. Para a maioria dos estudantes, a avaliação é um processo de aplicação de provas e/ou exames destinado a atribuir classificações.

A variância dos dados, contidos na planilha disponibilizada no Anexo I são:

Questão	Variância	Questão	Variância	Questão	Variância
PE1	0,51	PA1	0,64	PAv1	0,82
PE2	0,40	PA2	0,47	PAv2	0,51
PE3	0,43	PA3	0,65	PAv3	0,54
PE4	0,28	PA4	0,44	PAv4	0,71
PE5	0,57	PA5	0,88	PAv5	0,49
PE6	0,41	PA6	0,49	PAv6	0,47
PE7	0,36	PA7	0,60	PAv7	0,62
PE8	0,47	PA8	0,41	PAv8	0,48
PE9	0,40	PA9	0,55	PAv9	0,44
PE10	0,61	PA10	0,61	PAv10	0,47

PE11	0,48	PA11	0,41	PAv11	0,59
PE12	0,47	PA12	0,50	PAv12	0,94
PE13	0,44	PA13	0,38	PAv13	0,44
PE14	0,35	PA14	0,40	PAv14	0,56
PE15	0,40	PA15	0,46	PAv15	0,55
VT	196,55				

A soma de todas as variâncias é:

$$\sum S_i^2 = 23,11$$

A variância da soma dos itens é:

$$S_T^2 = 196,55$$

Se $k = 45$ o número de itens, podemos calcular o Alfa de Cronbach

$$\alpha = \left(\frac{k}{k-1} \right) \times \left(1 - \frac{\sum_{i=1}^k S_i^2}{S_T^2} \right)$$

Substituindo os valores temos

$$\alpha = \left(\frac{45}{45-1} \right) \times \left(1 - \frac{23,11}{196,55} \right)$$

de onde obtermos o valor:

$$\alpha = 0,90$$

Que segundo a escala a seguir:

Consistência interna do questionário segundo o valor de alfa

Valor de alfa	Consistência interna
Maior do que 0,80	Quase perfeito
De 0,80 a 0,61	Substancial
De 0,60 a 0,41	Moderado
De 0,40 a 0,21	Razoável
Menor do que 0,21	Pequeno

Fonte: Landis, J.R., Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*. 33:159.

Podemos afirmar que os dados obtidos com o questionário são confiáveis.

ATIVIDADE

Avalia a confiabilidade dos dados que foram obtidos com esse mesmo formulário, apresentado nesse capítulo, em que responderam ao questionário com 45 questões, sendo 15 referentes ao ensino, 15 relacionado a aprendizagem e 15 voltadas a avaliação aplicado em uma turma com engenharia de universidade pública do Estado do Pará, com 23 alunos. Os dados se encontram no Anexo II.

Os dados estão com a seguinte parametrização:

Discordo totalmente → 1

Discordo em parte → 2

Concordo em parte → 3

Concordo totalmente → 4

A variância dos dados, contidos na planilha disponibilizada no Anexo II, objeto dessa atividade, são:

Questão	Variância	Questão	Variância	Questão	Variância
PE1	0,25	PA1	0,30	PAv1	0,42
PE2	0,30	PA2	0,49	PAv2	0,31
PE3	0,25	PA3	0,95	PAv3	0,39
PE4	0,76	PA4	0,68	PAv4	0,67
PE5	0,59	PA5	0,84	PAv5	0,68
PE6	0,68	PA6	0,93	PAv6	0,33
PE7	0,40	PA7	0,54	PAv7	0,25
PE8	0,59	PA8	0,29	PAv8	0,23
PE9	0,77	PA9	0,42	PAv9	0,43
PE10	0,64	PA10	0,54	PAv10	0,49
PE11	0,33	PA11	0,93	PAv11	0,57
PE12	0,40	PA12	0,45	PAv12	0,55
PE13	0,42	PA13	0,24	PAv13	0,37
PE14	0,42	PA14	0,90	PAv14	0,28
PE15	0,60	PA15	0,56	PAv15	0,67
VT	183,53				

2. COEFICIENTE DE CORRELAÇÃO DE PEARSON

Em pesquisa em educação, quando se aplica questionários, nos deparamos com várias variáveis ou fatores, que podem ou não ter relações entre si, sendo necessário, a identificação dessas relações para que se estabeleça conclusões confiáveis. Uma das ferramentas estatística que nos ajuda nesta tarefa é o coeficiente de correlação de Pearson que é uma medida do grau de relação linear entre duas variáveis quantitativas.

O coeficiente de correlação de Pearson é um índice adimensional com valores situados ente -1,0 e 1.0 inclusive, que reflete a intensidade de uma relação linear entre dois conjuntos de dados. Este coeficiente, normalmente representado pela letra " ρ " assume apenas valores entre -1 e 1.

$\rho = 1 \Rightarrow$ Significa uma correlação perfeita positiva entre as duas variáveis.

$\rho = -1 \Rightarrow$ Significa uma correlação negativa perfeita entre as duas variáveis - Isto é, se uma aumenta, a outra sempre diminui.

$\rho = 0 \Rightarrow$ Significa que as duas variáveis não dependem linearmente uma da outra. No entanto, pode existir uma outra dependência que seja "não linear". Assim, o resultado $\rho = 0$ deve ser investigado por outros meios

Calcula-se o coeficiente de correlação de Pearson segundo a seguinte fórmula:

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

onde $x_1, x_2, x_3, \dots, x_n$ e $y_1, y_2, y_3, \dots, y_n$ são valores medidos de ambas as variáveis. Para isso, temos que $\bar{x} = \frac{1}{n} \sqrt{\sum_{i=1}^n x_i}$ e $\bar{y} = \frac{1}{n} \sqrt{\sum_{i=1}^n y_i}$ são as médias aritméticas de ambas as variáveis

Cabe observar que, como o coeficiente é concebido a partir do ajuste linear, então a fórmula não contém informações do ajuste, ou seja, é composta apenas dos dados. De onde podemos ter a seguinte interpretação desse coeficiente:

- 0.9 para mais ou para menos indica uma correlação muito forte.
- 0.7 a 0.9 positivo ou negativo indica uma correlação forte.
- 0.5 a 0.7 positivo ou negativo indica uma correlação moderada.
- 0.3 a 0.5 positivo ou negativo indica uma correlação fraca.
- 0 a 0.3 positivo ou negativo indica uma correlação desprezível.

Os Diagramas de dispersão ou gráficos de dispersão são representações de duas ou mais variáveis que são organizadas em um gráfico, uma em função da outra. Ela permite verificar a existência ou não de relação entre duas variáveis de natureza quantitativa.

Os gráficos podem se relacionar e serem interpretados como:

- Correlação positiva: quando um aumento de x acarreta em um aumento em y , assim se controlarmos x , y será também controlado.
- Correlação negativa: quando um aumento de x acarreta em uma diminuição em y , assim se controlarmos x , y será também controlado.

Exemplo: Em uma pesquisa para descobrir se quem sabe inglês, em que disciplina tem melhor desempenho (português ou matemática), foram escolhidos alunos do 1º ano do ensino médio de uma certa escola, sendo selecionado os quinze alunos com melhor desempenho (nota acima de 7) em inglês. Veja na tabela abaixo as notas em inglês, português e matemática destes quinze alunos.

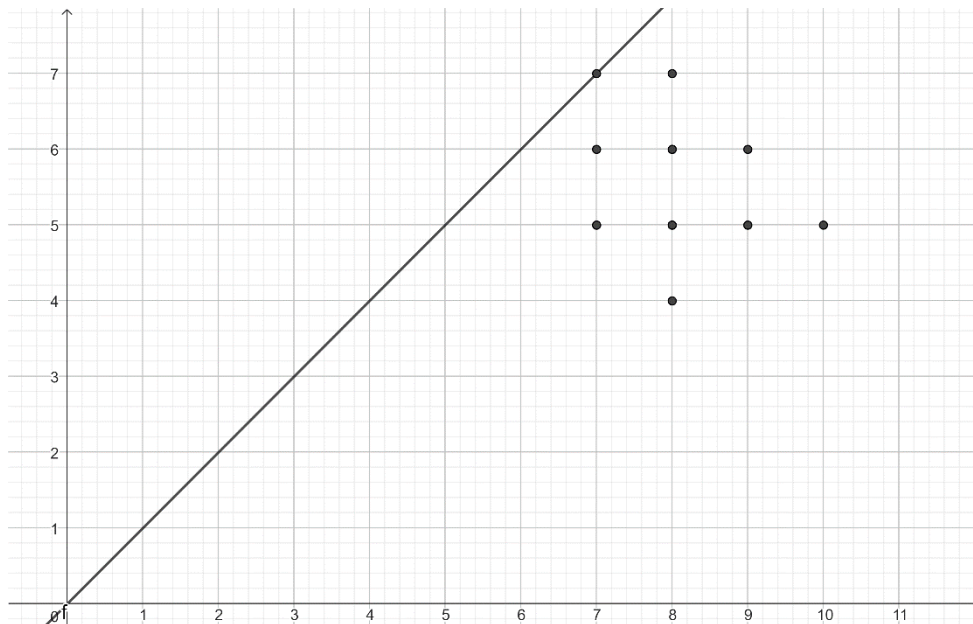
Alunos	Notas		
	Inglês	Português	Matemática
1	7,0	8,0	5,0
2	7,0	6,0	6,0
3	7,0	8,0	7,0
4	7,0	9,0	5,0
5	8,0	8,0	5,0
6	8,0	6,0	5,0
7	8,0	9,0	6,0
8	8,0	7,0	4,0
9	8,0	7,0	7,0
10	8,0	6,0	6,0
11	8,0	7,0	5,0
12	9,0	8,0	5,0
13	9,0	9,0	6,0
14	9,0	8,0	5,0
15	10,0	8,0	5,0

Para uma análise inicial da associação entre duas variáveis quantitativas, o diagrama de dispersão é a ferramenta indicada, e que consiste em exibir no plano cartesiano os pares de valores observados para duas variáveis quantitativas.

Figura 1 – Diagrama de dispersão entre notas de inglês e português



Figura 2 – A dispersão da nota de matemática em relação a de inglês



Na Figura 1 está apresentado um diagrama de dispersão entre as notas de inglês e as notas de português dos 15 alunos, sendo a linha contínua ($y = x$) o comportamento teórico se a associação entre as duas notas fosse perfeita, ou seja, se os alunos tivessem obtido as mesmas notas em português e inglês. A reta é referência para avaliar o

comportamento dos alunos. Aqueles que tiveram desempenho melhor em português do que inglês estão acima da reta e os que tiveram melhor desempenho em inglês, estão abaixo da reta. Na Figura 2 está o diagrama de dispersão das notas de inglês e matemática, onde se verifica a tendência dos alunos com bom desempenho em inglês apresentarem mau desempenho em matemática.

A seguir apresentaremos o cálculo dos coeficientes de correlação de Pearson dos dados que geraram os gráficos que estão na Figura 1 e 2, para que possa observar a relação deste coeficiente com o comportamento gráfico dos dados.

N	Inglês(x)	Português(y)	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	Inglês(x)	Matemática(y)	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
1	7	8	-0,427	1,138	0,160	7	5	0,498	1,138	0,218
2	7	6	1,707	1,138	2,560	7	6	-0,569	1,138	0,284
3	7	8	-0,427	1,138	0,160	7	7	-1,636	1,138	2,351
4	7	9	-1,493	1,138	1,960	7	5	0,498	1,138	0,218
5	8	8	-0,027	0,004	0,160	8	5	0,031	0,004	0,218
6	8	6	0,107	0,004	2,560	8	5	0,031	0,004	0,218
7	8	9	-0,093	0,004	1,960	8	6	-0,036	0,004	0,284
8	8	7	0,040	0,004	0,360	8	4	0,098	0,004	2,151
9	8	7	0,040	0,004	0,360	8	7	-0,102	0,004	2,351
10	8	6	0,107	0,004	2,560	8	6	-0,036	0,004	0,284
11	8	7	0,040	0,004	0,360	8	5	0,031	0,004	0,218
12	9	8	0,373	0,871	0,160	9	5	-0,436	0,871	0,218
13	9	9	1,307	0,871	1,960	9	6	0,498	0,871	0,284
14	9	8	0,373	0,871	0,160	9	5	-0,436	0,871	0,218
15	10	8	0,773	3,738	0,160	10	5	-0,902	3,738	0,218
			2,400	10,933	15,600			-2,467	10,933	9,733

$$\bar{x} = \frac{1}{n} \sqrt{\sum_{i=1}^n x_i} = 8,067$$

$$\bar{x} = \frac{1}{n} \sqrt{\sum_{i=1}^n x_i} = 8,067$$

$$\bar{y} = \frac{1}{n} \sqrt{\sum_{i=1}^n y_i} = 7,600$$

$$\bar{y} = \frac{1}{n} \sqrt{\sum_{i=1}^n y_i} = 5,467$$

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

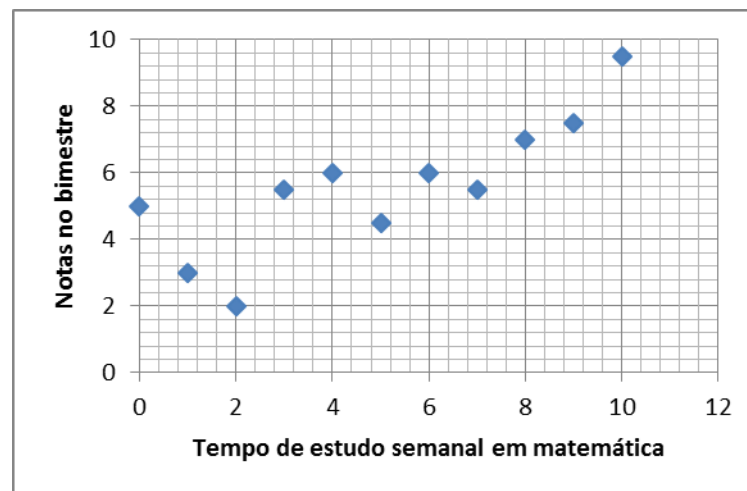
$$\rho = 0,184$$

$$\rho = -0,239$$

Exemplo: Em uma pesquisa investigou a influência do tempo de estudo semanal destinado de matemática com as notas que o aluno tira no bimestre. Os dados levantados em um grupo de alunos de ensino médio estão no quadro a seguir.

Tempo de estudo semanal em matemática	Média das notas de matemática no bimestre
0	5,0
1	3,0
2	2,0
3	5,5
4	6,0
5	4,5
6	6,0
7	5,5
8	7,0
9	7,5
10	9,5

Figura: Gráfico das médias das notas dos alunos em função do tempo gasto semanalmente e a nota que os alunos toraram na disciplina matemática no bimestre.



Quando vemos uma relação em um diagrama de dispersão, podemos usar uma reta para resumir essa relação nos dados, e usa-la para fazer previsões a partir dos dados, processo que é chamado de regressão linear.

É bom destacarmos que a regressão, em geral, tem como objetivo tratar de um valor que não se consegue estimar inicialmente, para isso determinamos uma equação teórica (equação da reta $y = a x + b$) que representa o dado, a partir da qual modelos fazer previsões.

No modelo teórico $y = a x + b$, os coeficientes a e b , são determinados da seguinte forma a seguir. Dados os dados tabelados:

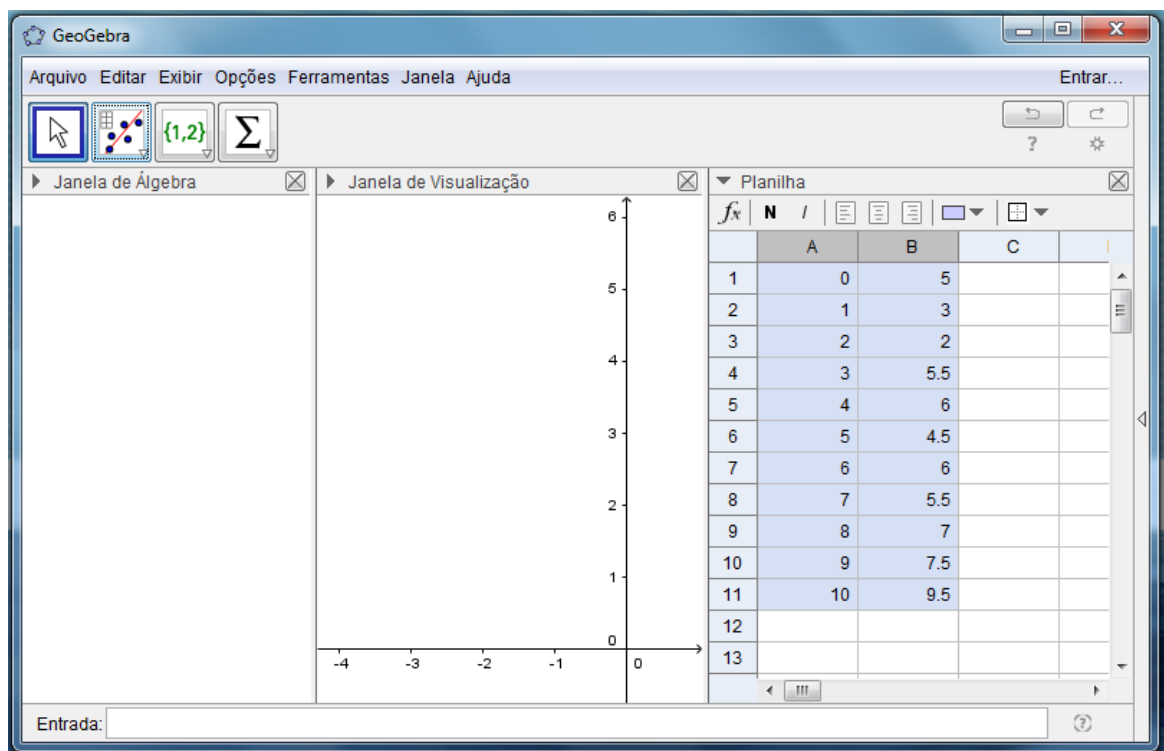
X	Y	X^2	XY
x_1	y_1	x_1^2	$x_1 y_1$
x_2	y_2	x_2^2	$x_2 y_2$
x_3	y_3	x_3^2	$x_3 y_3$
\vdots	\vdots	\vdots	\vdots
x_n	y_n	x_n^2	$x_n y_n$
ΣX	ΣY	ΣX^2	$\Sigma(XY)$

Se $y = a x + b$, determinamos os coeficientes pelas fórmulas:

$$a = \frac{\Sigma X^2 \Sigma Y - \Sigma(XY) \Sigma X}{n \Sigma X^2 - (\Sigma X)^2} \quad e \quad b = \frac{n \Sigma(XY) - \Sigma X \Sigma Y}{n \Sigma X^2 - (\Sigma X)^2}$$

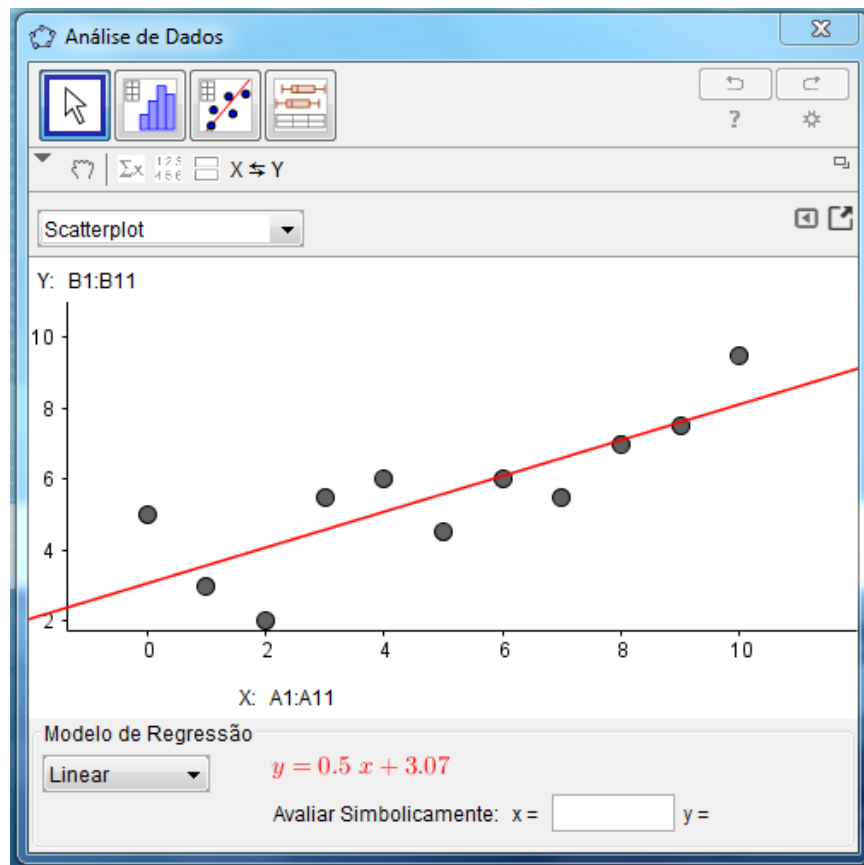
Para exemplificamos vamos fazer a regressão linear dos dados apresentados na tabela anterior, de forma que obtemos a seguinte equação com auxílio do Geogebra. Para isso, dentro do Geogebra, pedimos para exibir a “Planilha”, onde inserimos os dados, tendo o cuidado de usar ponto para separar a parte inteira da decimal, como vemos na figura a seguir.

Figura 3 – O Geogebra, onde está com o ambiente de planilha onde foram inseridos os dados a serem ajustados.



Após a inserção dos dados, deve ser acionado o botão “Análise Bivariada”, aparecendo a tela que está na figura a seguir, onde se faz a determinação da equação da reta por regressão linear.

Figura – Tela que aparece após acessar o botão “Análise Bivariada”, onde é feita a regressão linear.



Vamos agora calcular o coeficiente de correlação de Pearson, para podermos compreendermos melhor a relação entre essas duas variáveis.

n	Tempo de estudo semanal em matemática (x)	Média das notas de matemática no bimestre (y)	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
1	0	5	2,955	25,00	0,349
2	1	3	10,364	16,00	6,713
3	2	2	10,773	9,00	12,895
4	3	5,5	0,182	4,00	0,008
5	4	6	-0,409	1,00	0,167
6	5	4,5	0,000	0,00	1,190
7	6	6	0,409	1,00	0,167
8	7	5,5	-0,182	4,00	0,008
9	8	7	4,227	9,00	1,986
10	9	7,5	7,636	16,00	3,645
11	10	9,5	19,545	25,00	15,281
			55,500	110,000	42,409

$$\bar{x} = 5,00$$

$$\bar{y} = 5,59$$

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} = 0,81$$

É importante destacarmos que o coeficiente de correlação de Pearson dessas duas variáveis é 0,81 que significa que as variáveis tem uma correlação forte, e como é um valor positivo, indica que quando uma variável cresce a outra cresce também, isso é, se um aluno passar a gastar mais tempo, por semana estudando matemática, á uma forte tendência de sua nota ser maior na próxima avaliação.

ATIVIDADE

(01) Os dados a baixo mostra o tempo que os alunos, de duas turmas, reservam para estudar matemática e o desempenho destes no bimestre. Faça um estudo comparativo do comportamento dessas duas turmas em função do tempo gato no estudo de matemática, fazendo uso do coeficiente de correlação de Pearson, de representação gráfica e regressão linear.

Tempo de estudo semanal em matemática	Notas Turma A	Notas Turma B
0	5,0	4,5
1	3,0	5,0
2	2,0	3,5
3	5,5	4,5
4	6,0	5,0
5	4,5	4,5
6	6,0	4,5
7	5,5	5,0
8	7,0	5,5
9	7,5	6,0
10	9,5	6,5

3. INTERVALO DE CONFIANÇA

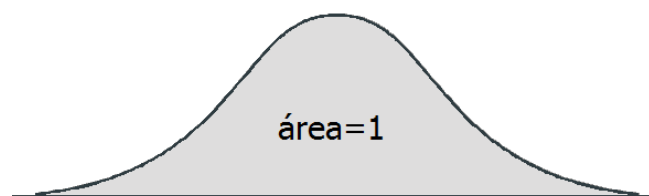
Em estatística, intervalo de confiança é um tipo de estimativa por intervalo de um parâmetro populacional desconhecido. Introduzido na estatística por Jerzy Neyman em 1937, é um intervalo observado (calculado a partir de observações), pode variar de amostra para amostra e que com dada frequência (nível de confiança) inclui o parâmetro de interesse real não observável. Desta forma, precisamos primeiro compreender o que distribuição normal que é a mais importante das distribuições estatísticas, tanto na teoria como na prática, por representar a distribuição de frequência de muitos fenômenos naturais.

Uma distribuição estatística que é uma função que define uma curva, e a área sob essa curva determina a probabilidade de ocorrer o evento por ela correlacionado. Nesse contexto, podemos dizer uma distribuição normal, também conhecida como distribuição gaussiana, é uma curva simétrica em torno do seu ponto médio, que tem seu famoso formato de sino. A curva normal representa o comportamento de diversos processos nas empresas e muitos fenômenos comuns, como por exemplo, altura ou peso de uma população, a pressão sanguínea de um grupo de pessoas, o tempo que um grupo de estudantes gasta para realizar uma prova.

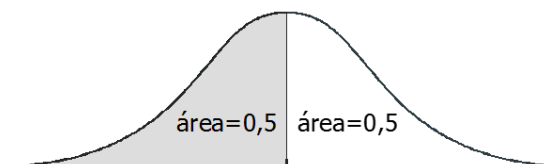
Para determinar se uma determinada variável aleatória segue uma distribuição normal, basta verificar se essa segue a função densidade de probabilidade, dada por:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

cuja representação gráfica é uma função cujo gráfico é semelhante a um sino, cm área igual a uma unidade.



A curva normal é simétrica em relação ao seu ponto central, como podemos ver na figura a seguir.



O valor da área a baixo da curva normal é igual a um, e corresponde a uma probabilidade de 100% para a ocorrência de um fenômeno. Isso quer dizer que a probabilidade de uma observação assumir um valor entre dois pontos quaisquer é igual à área compreendida entre esses dois pontos no gráfico.

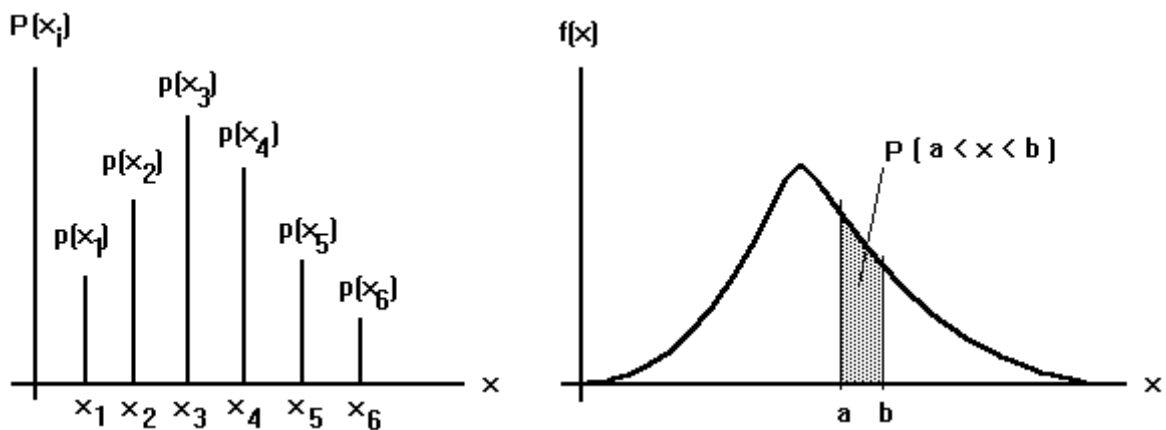
No caso de distribuições discretas, a probabilidade de que a variável X assumira um valor específico x_0 é dada por:

$$P(X = x_0) = P(x_0)$$

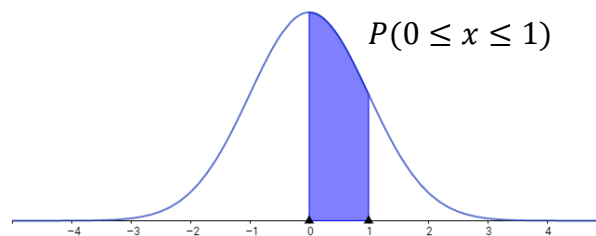
No caso de variáveis contínuas, as probabilidades são especificadas em termos de intervalos, pois a probabilidade associada a um número específico é zero.

$$P(a \leq x \leq b) = \int_a^b f(x)dx$$

Cuja representação gráfica é apresentada a seguir.



Para exemplificar apresentaremos, na figura seguir, a área preenchida no intervalo $0 \leq x \leq 1$

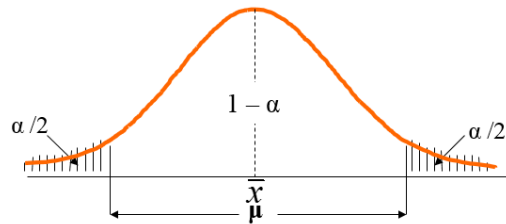


que representa a probabilidade de um evento acontecer no intervalo de $[0, 1]$, isso é

$$P(0 \leq x \leq 1) = \int_0^1 f(x)dx$$

Em estatística, um intervalo de confiança é o intervalo estimado onde a média de um parâmetro de uma amostra tem uma dada probabilidade de ocorrer. Comumente

define-se como o intervalo onde há 95% de probabilidade da média verdadeira da população inteira ocorrer. Para estabelecermos um intervalo de confiança (μ), precisamos da média amostral (\bar{x}), com base no número de amostra (n), sendo que o limite do intervalo de confiança é determinado por: $\mu = \bar{x} \pm E$. Onde $E = z_c \cdot \sigma_x$ é o erro máximo de estimativa, e $\sigma_x = \frac{\sigma}{\sqrt{n}}$, em que σ é o desvio padrão da população e z_c é o z crítico. O grau de confiança é dado por $1 - \alpha$, onde α é a significância.



onde temos que: $P(\bar{x} - E \leq \mu \leq \bar{x} + E) = 1 - \alpha$

A escolha da distribuição de probabilidade depende do número de elementos que possui a amostra, deforma que:

Se $n > 30$ se utiliza modelo normal, isto é, curva Gaussiana

Se $n \leq 30$ se utiliza Curva de Student, com grau de liberdade igual a $GL = n - 1$.

EXEMPLO:

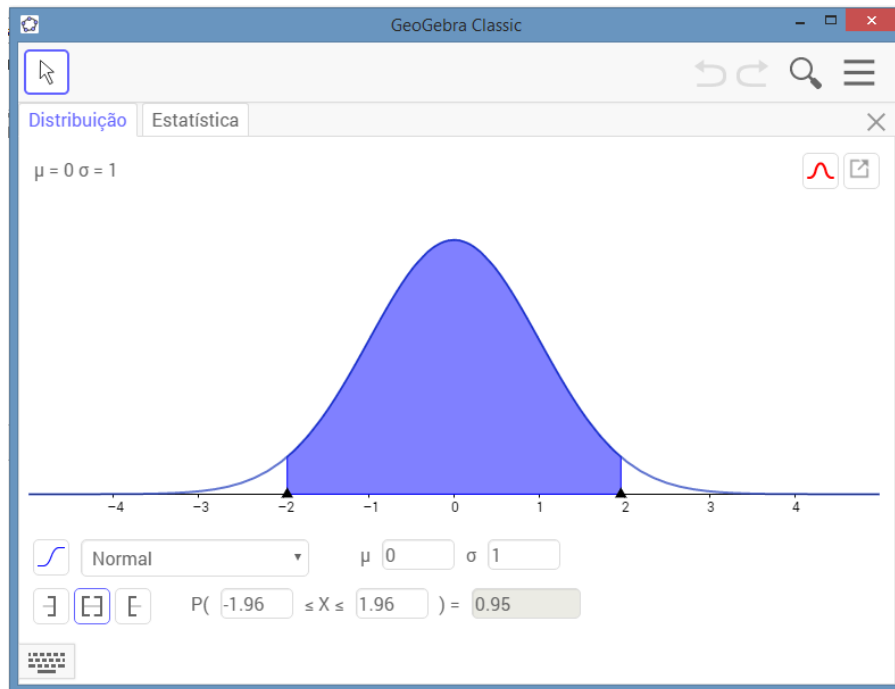
(01) Mediu-se a estatura de 40 alunos do curso de Ecologia da UFG. A média e desvio padrão estimados foram: $\bar{x} = 170 \text{ cm}$ e $\sigma = 5$. Qual o intervalo de confiança da média a 95%?

Solução:

Primeiro vamos calcular o σ_x , pela fórmula $\sigma_x = \frac{\sigma}{\sqrt{n}}$

$$\sigma_x = \frac{5}{\sqrt{40}} \Rightarrow \sigma_x = 0,79$$

Para se determinar o z_c usaremos a calculadora de probabilidade do Geogebra, onde entraremos com o valor $\frac{\alpha}{2} = 0,025$, que corresponde a uma significância $\alpha = 5\%$.



de onde temos que o z_c crítico é $z_c = 1,96$

A partir do z_c crítico devemos calcular o valor de $E = z_c \cdot \sigma_x$

$$E = 1,96 \times 0,79 \Rightarrow E = 1,55$$

Com isso determinamos o intervalo de confiança $\bar{x} - E \leq \mu \leq \bar{x} + E$

$$170 - 1,55 \leq \mu \leq 170 + 1,55$$

$$168,45 \leq \mu \leq 171,55$$

De forma que temos que:

$$P(168,45 \leq \mu \leq 171,55) = 95\%$$

P que significa que 95% dos alunos tem altura que está dentro do intervalo $168,45 \leq \mu \leq 171,55$.

(02) Os dados a seguir correspondem as notas, numa escala de zero a 10, de 30 alunos de uma escola.

5,5	7,5	4,0	5,0	3,0	4,5	6,0	8,5	9,0	2,5
4,5	7,0	6,0	8,5	3,0	7,0	1,5	4,5	9,5	5,5
7,0	5,5	3,5	6,5	5,0	8,5	9,0	6,5	6,0	7,5

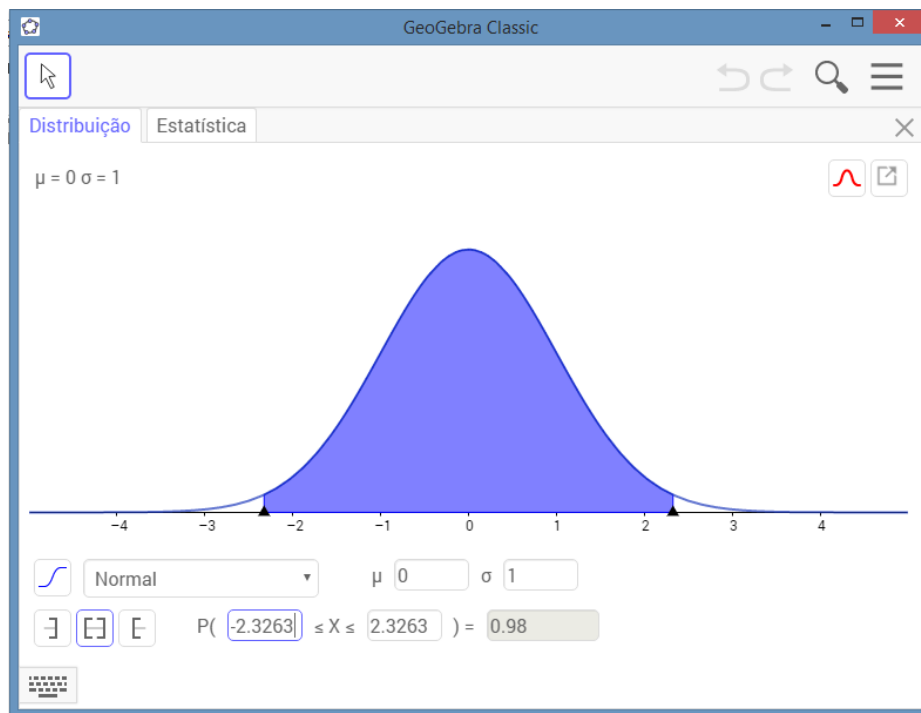
A escola considera que um aluno excelente tem nota no intervalo de 8,5 a 10,0 dentro os 30 alunos dessa turma. Determine um intervalo de confiança de 98% para a verdadeira proporção dos alunos ditos excelentes para os demais alunos dessa classe.

Solução:

Organizando os dados em ordem crescente, temos:

1,5	2,5	3,0	3,0	3,5	4,0	4,5	4,5	4,5	5,0
5,0	5,5	5,5	5,5	6,0	6,0	6,0	6,5	6,5	7,0
7,0	7,0	7,5	7,5	8,5	8,5	8,5	9,0	9,0	9,5

Para $1 - \alpha = 0,98$



De onde temos $Z_{0,98} = 2,3263 \approx 2,33$

Há 6 alunos que estão dentro do intervalo de 8,5 a 10,0, considerados como alunos excelentes e a proporção amostral dos alunos excelentes com os demais alunos é

$$p = \frac{n_{amostra}}{n_{população}}, \quad p = \frac{6}{30} = 0,2$$

$$\text{Daí, } d = z_{0,98} \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} = 2,33 \sqrt{\frac{0,2 \times (1 - 0,2)}{30}} = 0,17$$

Portanto, os limites de confiança pedidos são:

$$L_I = p - d = 0,2 - 0,17 = 0,03$$

$$L_S = p + d = 0,2 + 0,17 = 0,37$$

Com o nível de confiança de 98%, espera-se que a proporção dos alunos considerados excelentes esteja entre 3% e 37%.

Para calcular o Intervalo de confiança para a diferença de duas médias de populações normais independentes, temos que:

Caracterização da situação			Caracterização do procedimento		
Tamanho das amostras	σ_x^2 e σ_y^2 conhecidos?	$\sigma_x^2 = \sigma_y^2$?	Quartil a ser usado	Graus de liberdade	Erro Padrão $\bar{x} - \bar{y}$
Quaisquer	Sim	Não importa	$z_{1-\frac{\alpha}{2}}$	Não se aplica	$\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}$
Grandes	Não	Não importa	$z_{1-\frac{\alpha}{2}}$	Não se aplica	$\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}$
Pequenas	Não	Sim	$t_{1-\frac{\alpha}{2}}$	$n + m - 2$	$\sqrt{\frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2} \left(\frac{1}{n} + \frac{1}{m}\right)}$

Intervalo de confiança para a diferença de $(1 - \alpha)$ para $(\mu_x - \mu_y)$

$$(\bar{x} - \bar{y}) - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}} \leq \mu_x - \mu_y \leq (\bar{x} - \bar{y}) + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}$$

Populações normais independentes, $\sigma_x^2 \neq \sigma_y^2$, porém desconhecidas

$$(\bar{x} - \bar{y}) - t_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}} \leq \mu_x - \mu_y \leq (\bar{x} - \bar{y}) + t_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}$$

Populações normais independentes, $\sigma_x^2 = \sigma_y^2 = \sigma^2$, desconhecidas

$$(\bar{x} - \bar{y}) - t_{1-\frac{\alpha}{2}} \cdot s_c \cdot \sqrt{\frac{1}{n} + \frac{1}{m}} \leq \mu_x - \mu_y \leq (\bar{x} - \bar{y}) + t_{1-\frac{\alpha}{2}} \cdot s_c \cdot \sqrt{\frac{1}{n} + \frac{1}{m}}$$

Onde $s_c = \sqrt{\frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}}$

(03) Os 36 alunos de uma turma são divididos ao acaso em dois grupos de 18. Para o primeiro grupo o ensino de Matemática é feito usando elementos de multimídia.

Enquanto isso, no segundo grupo o ensino é feito pelo método tradicional (quadro negro e giz). No final do período é aplicado um teste, comum aos dois grupos, com os seguintes resultados:

Grupo1:	7,3	8,2	6,0	7,7	8,0	6,1	5,6	5,3	5,9
	5,8	5,8	7,1	5,1	8,0	7,6	8,3	4,9	6,5
Grupo2:	7,5	6,2	5,7	4,4	4,7	5,8	5,0	6,0	6,5
	5,8	4,5	5,1	5,5	6,0	5,8	5,8	5,7	7,5

Considerando os dois grupos como amostras aleatórias de duas populações independentes e normalmente distribuídas, determine um intervalo de confiança de 95% para a verdadeira diferença das médias populacionais dos dois grupos.

Solução:

Sejam X e Y as variáveis aleatórias representando as notas nos grupos 1 e 2, respectivamente. Denotemos as correspondentes médias populacionais por μ_X e μ_Y .

Então temos: $n = m = 18$

$\bar{x} = 6,622$; $s_X = 1,151$

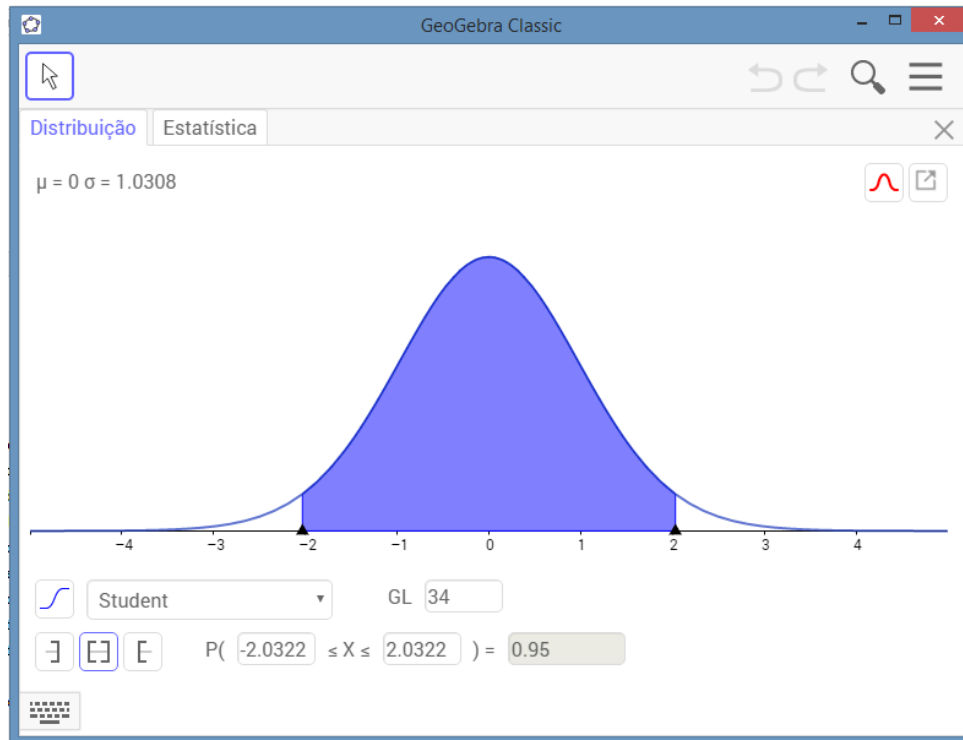
$\bar{y} = 5,750$; $s_Y = 0,860$

Estamos supondo normalidade para as distribuições de notas nos dois grupos. Como os desvios padrões são desconhecidos e os tamanhos amostrais são pequenos usaremos a distribuição t de Student. Os desvios padrões amostrais são relativamente próximos. Por

isso usamos a estimativa combinada: $s_c = \sqrt{\frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2}}$

$$s_c = \sqrt{\frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2}} = \sqrt{\frac{(18-1)1,151^2 + (18-1)0,86^2}{18+18-2}} = 1,016$$

Para $1 - \alpha = 0,95$ e $g.l. = 36 - 2 = 34$ graus de liberdade



$$t_{1-\frac{\alpha}{2}} = t_{0,975} = 2,032$$

Daí temos

$$E = t_{1-\frac{\alpha}{2}} \cdot s_c \cdot \sqrt{\frac{1}{n} + \frac{1}{m}} = 2,032 \times 1,016 \times \sqrt{\frac{1}{18} + \frac{1}{18}} = 0,688$$

Com os dados:

$$n = m = 18 ; \bar{x} = 6,622 ; s_x = 1,151 ; \bar{y} = 5,750 ; s_y = 0,860 ; E = 0,688$$

Desta maneira, os limites do intervalo de confiança de 95% para

$\mu_X - \mu_Y$ são:

$$L_{inf} = (\bar{X} - \bar{Y}) - E = (6,622 - 5,750) - 0,688 = 0,184 \cong 0,18$$

$$L_{sup} = (\bar{X} - \bar{Y}) + E = (6,622 - 5,750) + 0,688 = 1,560 \cong 1,56$$

Notemos que o intervalo inclui apenas valores positivos, indicando uma tendência no sentido de que a média do primeiro grupo é maior do que a média do segundo grupo para um nível de confiança de 95%. Então podemos afirmar que o método que usa multimídia tende a produzir notas superiores às do método tradicional, ao nível de confiança dado.

(04) A tabela a seguir contém as notas na 1ª prova (x) e na 2ª prova (y) de uma amostra de alunos de uma determinada disciplina. Deseja-se construir um intervalo de confiança a 95% para a diferenças entre as médias populacionais das notas da 1ª e da 2ª provas. Os dados são os seguintes:

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17
1ª Prova (x)	6,3	1,5	5,9	6,4	5,5	5,4	5,4	8,0	5,9	8,0	6,5	2,0	3,6	6,0	9,8	6,8	5,3
2ª Prova (y)	3,6	3,8	3,0	6,0	4,3	4,6	6,4	5,5	6,0	4,3	4,3	5,2	3,4	2,8	8,3	7,1	5,5
	A18	A19	A20	A21	A22	A23	A24	A25	A26	A27	A28	A29	A30	A31	A32	A33	A34
1ª Prova (x)	8,7	6,5	6,4	7,7	8,5	5,3	6,9	8,0	8,2	7,1	8,4	6,0	5,5	7,2	6,4	5,5	6,4
2ª Prova (y)	8,2	3,8	5,5	6,7	6,7	4,4	3,4	5,9	6,0	5,9	6,8	5,0	6,2	5,4	4,7	3,6	5,2

- (a) Que suposições deveriam ser feitas? O método estatístico a ser usado neste caso deve considerar amostras pareadas ou amostras independentes? Por que?
- (b) Use o método que você considera o mais adequado para obter o Intervalo de Confiança pedido.
- (c) Pode-se concluir com base nessa análise que, em média, o desempenho dos alunos piorou da 1ª prova para a 2ª prova?

Solução:

(a) Devemos considerar os 34 alunos como sendo uma amostra aleatória da população de todos os alunos que poderiam fazer essas duas provas. Além disso, devemos considerar as notas de ambas as provas como normalmente distribuídas.

Como o desvio padrão populacional das diferenças das duas notas é desconhecido, e como a amostra contém apenas 30 elementos, usaremos a distribuição t de Student.

Dado que, neste caso, as amostras não são independentes, porque se trata de observar um mesmo aluno em 2 momentos diferentes, claramente devemos usar o método para amostras pareadas.

(b) $\Delta = \text{Nota na 1ª Prova} - \text{Nota na 2ª Prova}$.

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17
1ª Pr (x)	6,3	1,5	5,9	6,4	5,5	5,4	5,4	8,0	5,9	8,0	6,5	2,0	3,6	6,0	9,8	6,8	5,3
2ª Pr (y)	3,6	3,8	3,0	6,0	4,3	4,6	6,4	5,5	6,0	4,3	4,3	5,2	3,4	2,8	8,3	7,1	5,5
	A18	A19	A20	A21	A22	A23	A24	A25	A26	A27	A28	A29	A30	A31	A32	A33	A34
1ª Pr (x)	8,7	6,5	6,4	7,7	8,5	5,3	6,9	8,0	8,2	7,1	8,4	6,0	5,5	7,2	6,4	5,5	6,4
2ª Pr (y)	8,2	3,8	5,5	6,7	6,7	4,4	3,4	5,9	6,0	5,9	6,8	5,0	6,2	5,4	4,7	3,6	5,2

x - y	2,7	-2,3	2,9	0,4	1,2	0,8	-1,0	2,5	-0,1	3,7	2,2	-3,2	0,2	3,2	1,5	-0,3	-0,2
x - y	0,5	2,7	0,9	1,0	1,8	0,9	3,5	2,1	2,2	1,2	1,6	1,0	-0,7	1,8	1,7	1,9	1,2

Obs.: Para simplificar os cálculos são fornecidos os seguintes valores:

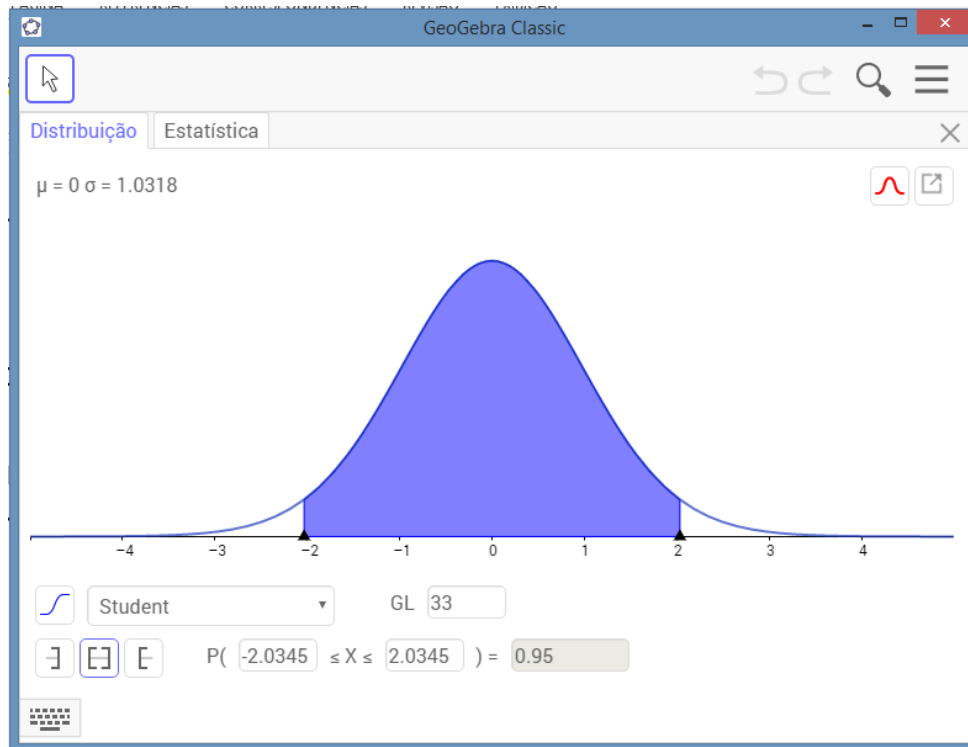
$$x \quad y \quad \Delta = x - y$$

Média 6,37 5,21 1,16

Desvio Padrão 1,73 1,40 1,54

Temos então $n = 34$ pares de dados, implicando que:

Graus de liberdade = $34 - 1 = 33$



$\bar{\Delta}_{obs} = 1,16$ e $(S_{\Delta})_{obs} = 1,54$

Com esses dados calculamos o valor de $E = t_{1-\frac{\alpha}{2}} \cdot \frac{S_{\Delta}}{\sqrt{n}}$

$$E = t_{1-\frac{\alpha}{2}} \cdot \frac{S_{\Delta}}{\sqrt{n}} = 2,0345 \cdot \frac{1,54}{\sqrt{34}} = 0,54$$

Assim, os limites do intervalo de confiança são:

$$L_{sup} = \bar{\Delta}_{obs} + E = 1,16 + 0,54 = 1,70 \quad e$$

$$L_{inf} = \bar{\Delta}_{obs} - E = 1,16 - 0,54 = 0,62$$

Para verificar se a premissa de normalidade está sendo atendida aqui, você pode construir, por exemplo, um gráfico de probabilidade normal.

Como o tamanho da amostra é relativamente grande, uma solução alternativa (aproximada) seria trabalharmos com a própria distribuição Normal padrão (já que, para valores altos do no de graus de liberdade, a t de Student praticamente se confunde

com a Normal padrão). Neste caso, o desvio padrão amostral $S_{\Delta} = 1,54$ seria usado como se fosse igual ao desvio padrão populacional σ_{Δ} . Como $z_{0,975} = 1,96$, resulta em:

$$E = z_{0,975} \frac{\sigma_{\Delta}}{\sqrt{n}} = 1,96 \times \frac{1,54}{\sqrt{34}} = 0,52$$

Assim, os limites do intervalo de confiança são:

$$L_{sup} = \bar{\Delta}_{obs} + E = 1,16 + 0,52 = 1,68 \text{ e}$$

$$L_{inf} = \bar{\Delta}_{obs} - E = 1,16 - 0,52 = 0,64$$

Observe que os limites obtidos são muito semelhantes em ambos os casos.

(c) Como o intervalo de confiança obtido não inclui o zero no seu interior, tudo leva a crer que em média o desempenho dos alunos de fato piorou da 1ª para a 2ª prova.

ATIVIDADE

(01) Um professor aplicou provas de Estatística a duas turmas no mesmo dia. Os resultados obtidos foram os seguintes:

Turma da Manhã: número de alunos = 19

2,5 3,5 4,5 5,0 5,5 5,5 6,0 6,5 7,0 7,5

7,5 7,5 7,5 8,0 8,0 8,5 9,0 9,0 10,0

Turma da Tarde: número de alunos = 20

3,5 3,5 4,0 4,5 5,0 5,0 5,5 5,5 6,0 6,5

6,5 7,0 7,0 7,0 7,0 7,5 7,5 8,0 9,0 9,5

O professor era o mesmo e foi usado o mesmo material didático, em ambos os casos. Além disso, supõe-se que em média os alunos de uma turma estudaram tanto quanto os da outra. Assim sendo, as médias populacionais podem ser vistas como representativas do nível de facilidade de cada uma das duas provas. O que podemos concluir a partir de um estudo de intervalo de confiança.

(02) Os 36 alunos de uma turma são divididos ao acaso em dois grupos de 18. Para o primeiro grupo o ensino de Matemática é feito usando elementos de multimídia. Enquanto isso, no segundo grupo o ensino é feito pelo método tradicional (quadro negro e giz). No final do período é aplicado um teste, comum aos dois grupos, com os seguintes resultados:

Grupo 1: 7,0 8,0 6,0 7,5 8,0 6,0 5,5 5,0 5,5

5,0 5,5 7,0 4,5 8,0 7,5 7,5 4,5 6,5

Grupo 2: 7,5 7,0 6,0 4,5 5,0 6,0 5,0 7,0 6,5

6,0 4,5 5,5 5,5 6,0 6,0 6,0 6,0 7,5

Considerando os dois grupos como amostras aleatórias de duas populações independentes e Normalmente distribuídas, determine um intervalo de confiança de 95% para a verdadeira diferença das médias populacionais dos dois grupos.

(03) A tabela a seguir contém as notas na 1ª prova (x) e na 2ª prova (y) de uma amostra de alunos de uma determinada disciplina. Deseja-se construir um intervalo de confiança a 95% para a diferença entre as médias populacionais das notas da 1ª e da 2ª provas. Os dados são os seguintes:

Tabela - Notas da 1ª e 2ª provas para um grupo de alunos

1ª Prova(x)	6,5	2,0	6,0	6,5	5,5	5,5	6,0	8,0	6,0	8,0	6,5	3,0	4,0
	7,0	10,0	7,0	5,5									
2ª Prova(x)	8,5	6,5	6,5	7,5	8,5	5,5	78,0	8,0	7,0	8,0	6,0	5,5	7,5
	6,5	5,5	6,0	4,0									
1ª Prova(x)	4,0	4,0	4,4	6,0	5,0	5,0	7,0	6,0	5,5	6,0	5,0	5,0	5,5
	4,0	3,0	9,0	7,5									
2ª Prova(x)	8,0	4,0	5,5	6,5	6,5	4,0	3,0	6,0	6,0	5,5	7,0	5,0	6,0
	5,5	5,5	4,0	5,0									

(a) Que suposições deveriam ser feitas? O método estatístico a ser usado neste caso deve considerar amostras pareadas ou amostras independentes? Por que?

(b) Use o método que você considera o mais adequado para obter o Intervalo de Confiança.

(c) Pode-se concluir com base nessa análise que, em média, o desempenho dos alunos piorou da 1ª prova para a 2ª prova?

4. TESTE DE HIPÓTESE

Um teste de hipótese é um método de inferência estatística usando dados de um estudo científico, correspondendo a uma suposição quanto ao valor de um parâmetro populacional, ou uma afirmação quanto a natureza da população. Ele corresponde a uma regra decisória que nos permite rejeitar ou não rejeitar uma hipótese estatística com base nos resultados de uma amostra.

Ao aplicarmos um teste de hipótese, temos que tomar uma decisão com base em informações apenas parciais (as que estão na amostra) sobre a realidade. Há dois tipos de erros de decisão que podem ser cometidos, que devem ser evitados:

Erro I: Rejeitar H_0 , quando H_0 é verdadeiro

Erro II: Aceitar H_0 , quando H_0 é falso

Probabilidade de ocorrência:

$\alpha = P[\text{Erro I}] = P[\text{"Rejeitar " } H_0], \text{ se } H_0 \text{ é verdadeiro}$

$\beta = P[\text{Erro II}] = P[\text{"Aceitar " } H_0], \text{ se } H_0 \text{ é falso}$

Quando aplicamos o teste de hipótese numa amostra, seguimos as seguintes fases:

1ª Fase: Seja μ_0 uma constante conhecida, para especificar H_0 e H_1 , usaremos uma dentre as três possibilidades a seguir:

$H_0: \mu = \mu_0$ versus $H_1: \mu \neq \mu_0$ (teste bilateral)

$H_0: \mu \leq \mu_0$ versus $H_1: \mu > \mu_0$ (teste unilateral)

$H_0: \mu \geq \mu_0$ versus $H_1: \mu < \mu_0$ (teste unilateral)

2ª Fase: Em seguida devemos escolher o nível de significância α

3ª Fase: Usar como estatística de teste a expressão: 2ª Fase usando a expressão: $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$

4ª Fase: A distribuição de probabilidade de Z correspondente H_0 é $N(0; 1)$

5ª Fase: Obter a região de rejeição de R (ou seja, o conjunto de valores de Z que nos levarão a rejeitar H_0), levando em conta os passos anteriores.

6ª Fase: Coletar os dados x_1, x_2, \dots, x_n , calcular o valor de $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$ e decidir pela rejeição ou pela aceitação de H_0 , conforme o critério específico no 5º passo.

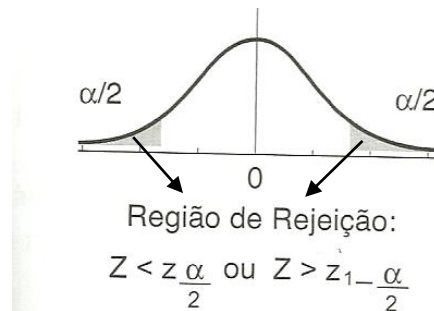
$$\bar{x} = \mu_0 + z \frac{\sigma}{\sqrt{n}}$$

No teste de hipótese podemos ter as seguintes situações:

1ª Situação: Teste bilateral

$H_0: \mu = \mu_0$ versus $H_1: \mu \neq \mu_0$

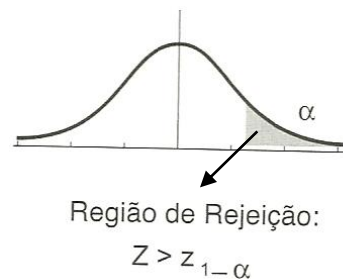
Nesta situação temos a seguinte interpretação gráfica:



2ª Situação: Teste unilateral

$H_0: \mu \leq \mu_0$ versus $H_1: \mu > \mu_0$

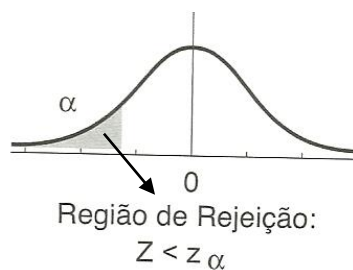
Nesta situação temos a seguinte interpretação gráfica:



3ª Situação: Teste unilateral

$H_0: \mu \geq \mu_0$ versus $H_1: \mu < \mu_0$

Nesta situação temos a seguinte interpretação gráfica:



Quando há duas populações normais com médias e variâncias desconhecidas, as hipóteses, para testar se as médias são iguais são as seguintes: $H_0: \mu_1 = \mu_2$. O procedimento de teste irá depender de que $\sigma_1^2 = \sigma_2^2$. Se essa suposição for razoável, então calcula-se a variância combinada. A seguir segue a estatística de teste:

$$t_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{e} \quad S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

EXEMPLO

(01) Um professor aplicou provas de Estatística a duas turmas no mesmo dia. Os resultados obtidos foram os seguintes:

Turma da Manhã: número de alunos = 19

2,5 3,5 4,5 5,0 5,5 5,5 6,0 6,5 7,0 7,5
7,5 7,5 7,5 8,0 8,0 8,5 9,0 9,0 10,0

Turma da Tarde: número de alunos = 20

3,5 3,5 4,0 4,5 5,0 5,0 5,5 5,5 6,0 6,5
6,5 7,0 7,0 7,0 7,0 7,5 7,5 8,0 9,0 9,5

O professor era o mesmo e foi usado o mesmo material didático, em ambos os casos. Além disso, supõe-se que em média os alunos de uma turma estudaram tanto quanto os da outra. Assim sendo, as médias populacionais podem ser vistas como representativas do nível de facilidade de cada uma das duas provas.

(a) Caso a hipótese de variâncias iguais tenha sido aceita no item (a), teste a hipótese nula de médias populacionais iguais contra a hipótese alternativa de que a média da Turma da Manhã é maior que a da Turma da Tarde, ao nível de significância de 5%.

(b) O que se pode afirmar sobre o p-valor no caso do teste de médias

O teste a ser aplicado aqui é:

$$H_0: \mu_M = \mu_T \text{ versus } H_1: \mu_M \neq \mu_T,$$

onde:

μ_M = média populacional das notas na prova da Manhã e

μ_T = média populacional das notas na prova da Tarde.

Podemos aplicar aos dados o teste t de Student para amostras não pareadas.

A estatística de teste é:

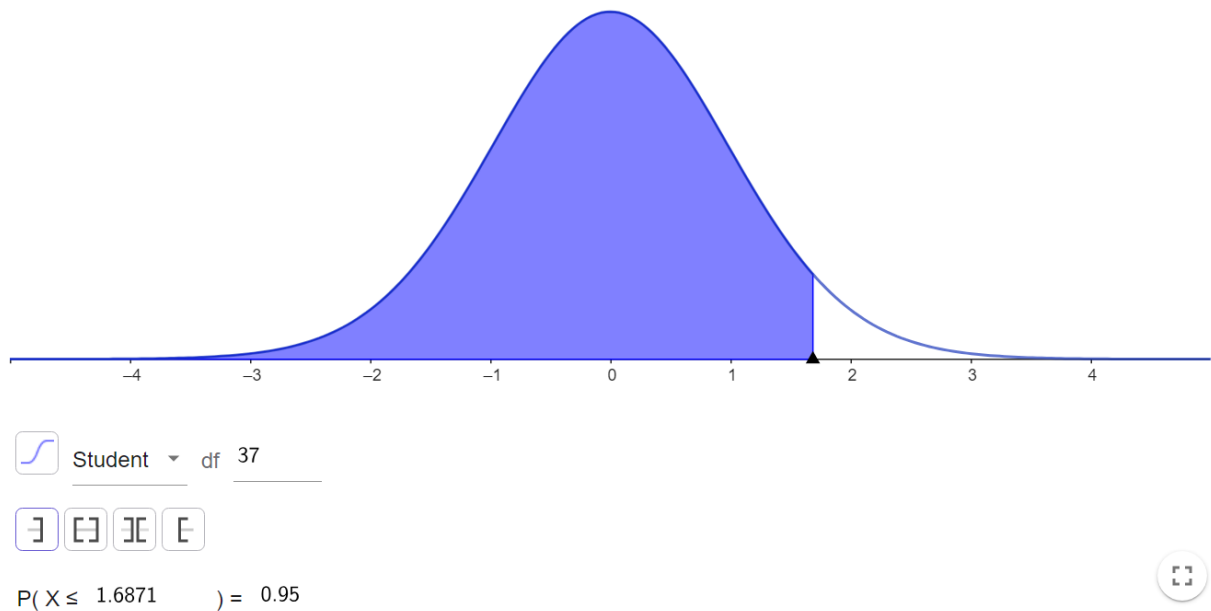
$$T = \frac{\bar{x}_M - \bar{x}_T}{\sqrt{S_c^2 \left(\frac{1}{n_M} + \frac{1}{n_T} \right)}} \quad \text{onde} \quad S_c^2 = \frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{(n_x + n_y - 2)}$$

Inserindo os valores temos:

$$S_c^2 = \frac{(19 - 1) \times 1,97^2 + (20 - 1) \times 1,69^2}{(19 + 20 - 2)} = 3,35$$

$$T = \frac{x_1 - x_2}{\sqrt{S_c^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{6,76 - 6,25}{\sqrt{3,35 \left(\frac{1}{19} + \frac{1}{20} \right)}} = 0,875$$

Utilizando a calculadora de probabilidade do Geogebra, temos



Se trata de um teste unilateral, o critério de decisão é:

Rejeitar H_0 se $T_{obs} > t_{1-\alpha}$. Caso contrário, aceita H_0 .

O número de grau de liberdade é $19 + 20 - 2 = 37$, portanto, para $\alpha = 5\%$ temos $t_{1-\alpha} = 1,687$.

Como $T_{obs} = 0,875 < 1,687 = t_{0,95}$. A hipótese H_0 não deve ser rejeitada

Ao trabalharmos ao nível de significância de 5%, com base nos dados não há evidências de que a média μ_M seja superior à média μ_T . A análise aqui apresentada nos leva a concluir, com base nos dados, pela plausibilidade da hipótese nula de que o nível de dificuldade foi o mesmo nas duas provas.

Como H_0 foi aceitável ao nível $\alpha = 5\%$, é claro que o valor p-valor é maior ou igual a 5%.

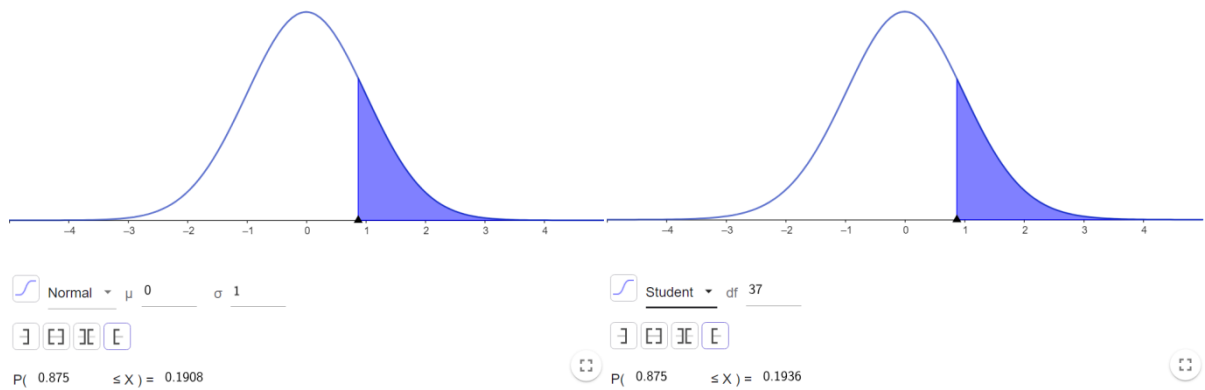
$$P[T > 0,875] = 5\%$$

Se no cálculo tivéssemos usado a curva normal o resultada seria semelhante, como podemos observar, pois:

$$P[Z > 0,875] = 0,1908 \text{ ou } 19,08\%$$

Na distributividade t de Student com 37 graus de liberdade

$$P[T > 0,875] = 0,1936 \text{ ou } 19,36\%$$



(02) Os registros dos últimos anos de uma faculdade atestam, para calouros admitidos, uma nota média 115, na prova de seleção. Para testar a hipótese de que a média de uma nova turma é a mesma das turmas anteriores, retirou-se, ao acaso, uma amostra de 20 notas, obtendo-se média 118 e desvio padrão 20. Analise o resultado com uma confiança de 95%, o que significa usar $\alpha = 0,05$

Supondo que as notas dos novos calouros têm distribuição normal com média μ e desvio padrão σ .

(i) As hipóteses de interesse são:

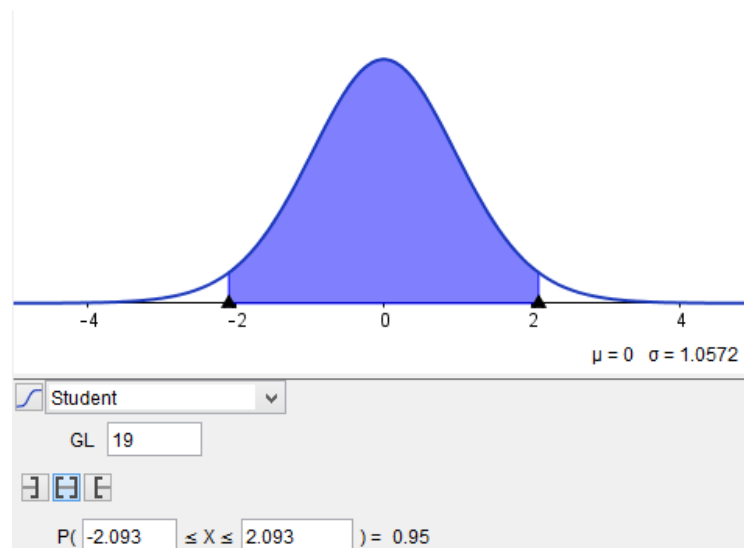
$$H_0: \mu = 115$$

$$H_1: \mu \neq 115$$

(ii) Estatística de teste $t(n - 1)$

$$T = \frac{\bar{x} - 115}{S/\sqrt{n}}$$

(iii) A região crítica para um nível de significância $\alpha = 0,05$ fixado



Região crítica = $\{z \in T / |T| \geq 2,093\}$

(iv) Do enunciado temos

$$T = \frac{\bar{x} - 115}{S/\sqrt{n}} = \frac{118 - 115}{20/\sqrt{20}} = 0.67$$

Como $T = 0.67$ está fora da região crítica, não rejeita-se H_0 ao nível de 5% de significância

(03) A tabela a seguir contém as notas na 1ª prova (x) e na 2ª prova (y) de uma amostra de alunos de uma determinada disciplina. Deseja-se fazer uma análise utilizando o teste de hipótese, com confiança a 95% para verificar se o desempenho dos alunos melhorou da 1ª prova para a 2ª prova.

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17
1ª Pr (x)	6,3	1,5	5,9	6,4	5,5	5,4	5,4	8,0	5,9	8,0	6,5	2,0	3,6	6,0	9,8	6,8	5,3
2ª Pr (y)	3,6	3,8	3,0	6,0	4,3	4,6	6,4	5,5	6,0	4,3	4,3	5,2	3,4	2,8	8,3	7,1	5,5
	A18	A19	A20	A21	A22	A23	A24	A25	A26	A27	A28	A29	A30	A31	A32	A33	A34
1ª Pr (x)	8,7	6,5	6,4	7,7	8,5	5,3	6,9	8,0	8,2	7,1	8,4	6,0	5,5	7,2	6,4	5,5	6,4
2ª Pr (y)	8,2	3,8	5,5	6,7	6,7	4,4	3,4	5,9	6,0	5,9	6,8	5,0	6,2	5,4	4,7	3,6	5,2

Solução:

Considerando que há 34 alunos como sendo uma amostra e que as notas de ambas as provas são consideradas como normalmente distribuídas. Como o desvio padrão populacional das diferenças das duas notas é desconhecido, e como a amostra contém apenas 34 elementos, usaremos a distribuição t de Student.

Dado que, neste caso, as amostras não são independentes, porque se trata de observar um mesmo aluno em 2 momentos diferentes, claramente devemos usar o método para amostras pareadas.

(i) As hipóteses de interesse são:

$$H_0: \mu_{antes} < \mu_{depois}$$

$$H_1: \mu_{antes} \geq \mu_{depois}$$

Δ = Nota na 2ª Prova – Nota na 1ª Prova.

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17
1ª Pr (x)	6,3	1,5	5,9	6,4	5,5	5,4	5,4	8,0	5,9	8,0	6,5	2,0	3,6	6,0	9,8	6,8	5,3
2ª Pr (y)	3,6	3,8	3,0	6,0	4,3	4,6	6,4	5,5	6,0	4,3	4,3	5,2	3,4	2,8	8,3	7,1	5,5
	A18	A19	A20	A21	A22	A23	A24	A25	A26	A27	A28	A29	A30	A31	A32	A33	A34
1ª Pr (x)	8,7	6,5	6,4	7,7	8,5	5,3	6,9	8,0	8,2	7,1	8,4	6,0	5,5	7,2	6,4	5,5	6,4
2ª Pr (y)	8,2	3,8	5,5	6,7	6,7	4,4	3,4	5,9	6,0	5,9	6,8	5,0	6,2	5,4	4,7	3,6	5,2

y - x	-2,7	2,3	-2,9	-0,4	-1,2	-0,8	1,0	-2,5	0,1	-3,7	-2,2	3,2	-0,2	-3,2	-1,5	0,3	0,2
y - x	-0,5	-2,7	-0,9	-1,0	-1,8	-0,9	-3,5	-2,1	-2,2	-1,2	-1,6	-1,0	0,7	-1,8	-1,7	-1,9	-1,2

Obs.: Para simplificar os cálculos são fornecidos os seguintes valores:

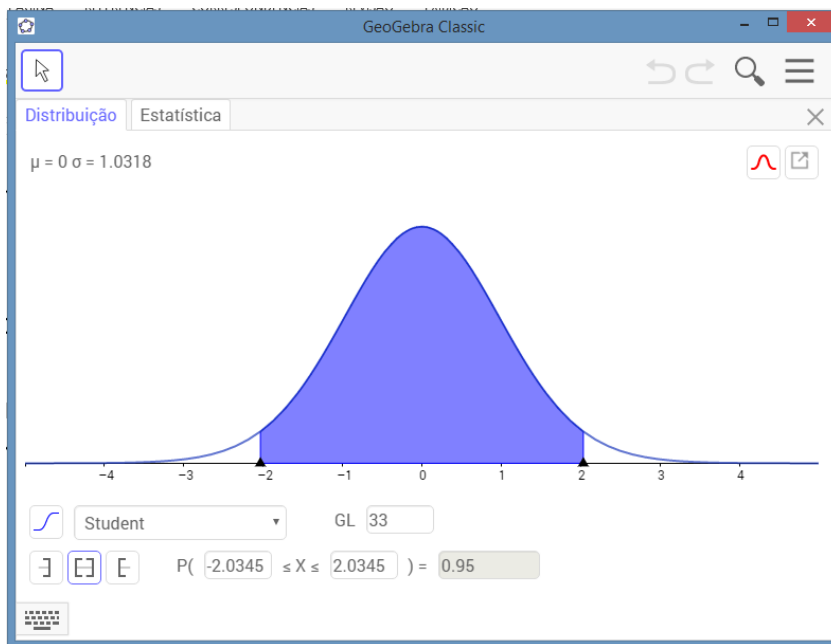
$$x \quad y \quad \Delta = y - x$$

Média 6,37 5,21 -1,16

Desvio Padrão 1,73 1,40 1,53

Temos então $n = 34$ pares de dados, implicando que:

Graus de liberdade = $34 - 1 = 33$



$$\bar{\Delta}_{obs} = -1,16 \text{ e } (S_{\Delta})_{obs} = 1,53$$

$$\text{Média amostral das diferenças: } \bar{D} = \frac{\sum D}{n}$$

onde $\sum D$ é a soma das diferenças e n é o número de elementos da amostra. Então:

$$\bar{D} = \frac{\sum D}{n} = \frac{-39,5}{34} = -1,16$$

$$\text{Desvio padrão amostral das diferenças: } S_D = \sqrt{\frac{\sum D^2 - n \bar{D}^2}{n-1}}$$

y - x	-2,7	2,3	-2,9	-0,4	-1,2	-0,8	1,0	-2,5	0,1	-3,7	-2,2	3,2	-0,2	-3,2	-1,5	0,3	0,2
y - x	-0,5	-2,7	-0,9	-1,0	-1,8	-0,9	-3,5	-2,1	-2,2	-1,2	-1,6	-1,0	0,7	-1,8	-1,7	-1,9	-1,2

(y - x)²	7,3	5,3	8,4	0,2	1,4	0,6	1,0	6,3	0,0	13,7	4,8	10,2	0,0	10,2	2,3	0,1	0,0
(y - x)²	0,3	7,3	0,8	1,0	3,2	0,8	12,3	4,4	4,8	1,4	2,6	1,0	0,5	3,2	2,9	3,6	1,4

Então: $\sum D^2 = 123,5$

$$S_D = \sqrt{\frac{\sum D^2 - n \bar{D}^2}{n-1}} = \sqrt{\frac{123,5 - 34 \cdot (-1,16)^2}{34 - 1}} = 1,535$$

$$\text{Teste estatístico: } t = \frac{\bar{D} \cdot \sqrt{n}}{S_D}$$

Onde:

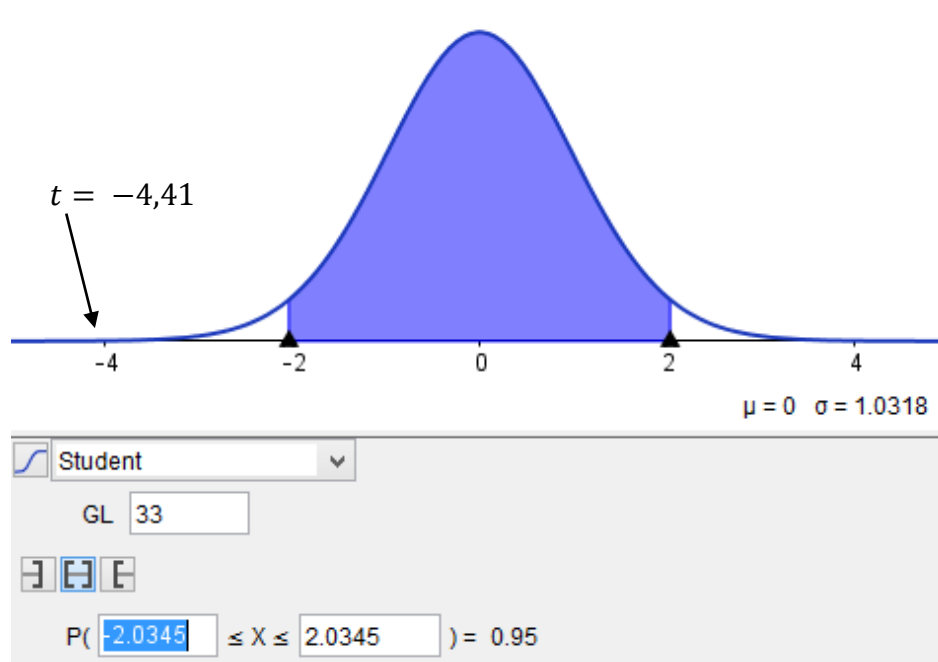
n é o número de alunos (antes, depois) observados

\bar{D} é a média das diferenças observadas

S_D é o desvio padrão das diferenças observadas

$$t = \frac{\bar{D} \cdot \sqrt{n}}{S_D} = \frac{-1,16 \cdot \sqrt{34}}{1,535} = -4,41$$

Utilizando a calculadora de probabilidade do Geogebra temos:



Temos que $t < -2,0345$ o que implica que devemos rejeitar H_0 , isso significa que o desempenho dos alunos piorou da 1ª prova para a 2ª prova.

ATIVIDADE

(01) Um professor aplicou provas de Matemática a duas turmas no mesmo dia. Os resultados obtidos foram os seguintes:

Turma da Manhã: número de alunos = 19

2,0 3,5 4,5 5,0 5,0 5,5 6,0 6,0 6,5 7,5
8,5 7,5 7,5 8,0 8,0 8,5 9,0 9,0 10,0

Turma da Tarde: número de alunos = 20

3,0 3,5 4,0 4,5 5,0 5,0 5,5 5,5 6,0 7,5
6,5 7,0 7,0 7,5 7,5 7,5 8,0 8,5 9,0 9,5

O professor era o mesmo e foi usado o mesmo material didático, em ambos os casos.

Além disso, supõe-se que em média os alunos de uma turma estudaram tanto quanto os da outra. Assim sendo, as médias populacionais podem ser vistas como representativas do nível de facilidade de cada uma das duas provas.

(a) Caso a hipótese de variâncias iguais tenha sido aceita no item (a), teste a hipótese nula e médias populacionais iguais contra a hipótese alternativa de que a média da Turma da Manhã é maior que a da Turma da Tarde, ao nível de significância de 5%.

(b) O que se pode afirmar sobre o p-valor no caso do teste de médias

(02) A tabela a seguir contém as notas na 1ª prova (x) e na 2ª prova (y) de uma amostra de alunos de uma determinada disciplina. Deseja-se construir um intervalo de confiança a 95% para a diferença entre as médias populacionais das notas da 1ª e da 2ª provas. Verifique a hipótese que a média de desempenho dos alunos piorou da 1ª prova para a 2ª prova?

Os dados são os seguintes:

Tabela - Notas da 1ª e 2ª provas para um grupo de alunos																	
1ª Prova	6,5	2,0	6,0	6,5	5,5	5,5	6,0	8,0	6,0	8,0	6,5	3,0	4,0	7,0	10,0	7,0	5,5
1ª Prova	4,0	4,0	4,4	6,0	5,0	5,0	7,0	6,0	5,5	6,0	5,0	5,5	4,0	3,0	9,0	7,5	
2ª Prova	8,5	6,5	6,5	7,5	8,5	5,5	8,0	8,0	7,0	8,0	6,0	5,5	7,5	6,5	5,5	6,0	4,0
2ª Prova	8,0	4,0	5,5	6,5	6,5	4,0	3,0	6,0	6,0	5,5	7,0	5,0	6,0	5,5	5,5	4,0	5,0

5. ANOVA

A ANOVA ou Análise de Variância é um procedimento usado para comparar a distribuição de três ou mais grupos em amostras independentes.

ANOVA = **A**nalise **O**f **V**ariance

A variância (S^2) é uma medida de dispersão que indica o quanto que os valores se distanciam da média.

Geralmente se pergunta: Quando usar a Anova?

- Quando fazemos comparação envolvendo mais de dois tratamentos ou situações experimentais (Y contínuo e X categóricos com mais de 2 níveis)

E o que a Anova faz?

- Compara variação dentro de cada tratamento ou situação experimental e entre os tratamentos ou situações experimentais.

Para facilitarmos o entendimento da Anova, vamos estudá-la a partir da resolução de um problema.

Problema: Um professor de matemática notou que quanto mais longe os alunos sentavam, mais provável que percam sua atenção na explicação, e façam uso de redes sociais e mensagens instantâneas durante as aulas. Ele deseja saber se os alunos que sentam mais longe foram pior nos exames. Para isso dividiu a sala de aula em três partes: frente (F), meio (M), e fundo (B).

Mas será que a média é suficiente para afirmar que o desempenho dos alunos de cada grupo é realmente diferente? E se um dos grupos tiver um aluno que não se preparou e errou quase todas as questões? Esse aluno não seria responsável por ter diminuído a média do grupo?

Para verificar então se realmente o desempenho dos alunos variou de acordo com o grupo (frente (F), meio (M), e fundo (B)), se faz necessário a utilização de teste estatístico, que além de considerar a média das notas, leva também em conta a variação das notas dentro de cada grupo.

Na Anova ou análise de variância, o objetivo é verificar se há diferença entre distribuição de uma medida entre três ou mais grupos. Isto é, a ANOVA não testa que uma média é menor que outra, apenas se elas são todas iguais ou ao menos uma é diferente. Sendo consideradas as seguintes hipóteses:

$H_0: \mu_F = \mu_M = \mu_B$ (todas as médias são iguais)

$H_1: \mu_i \neq \mu_j$, para algum $i \neq j$

Em nosso exemplo, em outras palavras podemos definir as hipóteses do teste como:

H_0 : Não existe diferença entre o desempenho das notas dos alunos de cada grupo.

H_1 : Há pelo menos um grupo com alunos com desempenho diferente.

Uma amostra aleatória de alunos de cada parte da sala foi selecionada, e as notas para os alunos na segunda prova foram anotadas:

Frente (F): 82, 83, 97, 93, 55, 67, 53

Meio (M): 83, 78, 68, 61, 77, 54, 69, 51, 63

Fundão (B): 38, 59, 55, 66, 45, 52, 52, 61

Ressaltamos que a nota máxima na avaliação é 100. As medidas resumo para as partes da sala são apresentadas na tabela abaixo:

Parte	Frente(F)	Meio (M)	Fundão (B)
N	7	9	8
Média	75,71	67,11	53,50
Desvio Padrão	17,63	10,95	8,96
Variância	310,90	119,86	80,29

Para calcularmos a Anova como um fator precisamos seguir os seguintes passos:

1º passo: Calcular a Média Geral ($\bar{\bar{x}}$)

É a média aritmética de todos os valores quando o fator é ignorado, ou ainda, a média ponderada das médias amostrais de cada grupo pelo tamanho de cada amostra.

$$\bar{\bar{x}} = \frac{\sum_{i=1}^k n_i \bar{x}_i}{\sum_{i=1}^k n_i} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2 + \dots + n_k \bar{x}_k}{n_1 + n_2 + \dots + n_k}$$

Usando os valores temos:

$$\bar{\bar{x}} = \frac{7 \times 75,71 + 9 \times 67,11 + 8 \times 53,50}{7 + 9 + 8}$$

$$\bar{\bar{x}} = \frac{1562}{24} \Rightarrow \bar{\bar{x}} = 65,08$$

2º passo: Variação Entre Grupos, SQ(E)

A SQ(E) é a variação entre cada média amostral, de cada grupo, e a média geral. A variação de cada grupo é ponderada pelo tamanho amostral.

$$SQ(E) = \sum_{i=1}^k n_i (\bar{x}_i - \bar{\bar{x}})^2$$

$$SQ(E) = n_1(\bar{x}_1 - \bar{\bar{x}})^2 + n_2(\bar{x}_2 - \bar{\bar{x}})^2 + \dots + n_k(\bar{x}_k - \bar{\bar{x}})^2$$

Substituindo os valores temos:

$$SQ(E) = 7(75,71 - 65,08)^2 + 9(67,11 - 65,08)^2 + 8(53,50 - 65,08)^2$$

$$SQ(E) = 1900,8376 \approx 1901$$

3º passo: Variação Dentro dos Grupos, SQ(D)

A SQ(D) é o total das variações individuais ponderadas. A ponderação é realizada com os graus de liberdade, gl , da amostra de cada grupo que é o tamanho amostral menos uma unidade.

$$SQ(D) = \sum_{i=1}^k gl_i s_i^2$$

$$SQ(D) = gl_1 s_1^2 + gl_2 s_2^2 + \dots + gl_k s_k^2$$

Considerando os dados do nosso exemplo, temos:

$$SQ(D) = 6 \times 310,90 + 8 \times 119,86 + 7 \times 80,29$$

$$SQ(D) = 3386,31 \approx 3386$$

Com os valores que calculamos até esse momento podemos iniciar o preenchimento do quadro a seguir.

Fonte da Variação	SQ	gl	QM	F	p
Entre	1901				
Dentro	3386				
Total	5287				

4º passo: Grau de Liberdade

No cálculo do grau de liberdade (gl) usaremos a seguinte regra:

- Os gl entre os grupos é o número de grupos menos 1. Temos três grupos, logo $gl(E) = 2$
- Os gl dentro dos grupos é a soma dos gl individuais de cada grupo. Tamanho das amostras são 7, 9 e 8, $gl(D) = 6 + 8 + 7 = 21$
- Os gl total é o tamanho da amostra geral menos 1. $gl(Total) = (7 + 9 + 8) - 1 = 23$

Com isso passamos a acrescentar as seguintes informações:

Fonte da Variação	SQ	gl	MQ	F	P
Entre	1901	2			
Dentro	3386	21			
Total	5287	23			

5º passo: Quadrado médios dos desvios (QM)

As variâncias são chamadas Quadrados Médios abreviadamente QM , acompanhadas da identificação da variável a que se refere $QM(E)$ ou $QM(D)$. São os quadrados médios dos desvios das médias, encontrados dividindo a variação (SQ) pelos seus respectivos gl .

$$QM = \frac{SQ}{gl}$$

A partir dos dados do exemplo, temos:

$$QM(E) = \frac{1901}{2} = 950,5$$

$$QM(D) = \frac{3386}{21} = 161,2$$

$$QM(T) = \frac{5287}{23} = 229,9$$

Obs.: Note que o $QM(Total)$ não é a soma dos $QM(E)$ e $QM(D)$. Isso vale para as Somas de Quadrados $SS(Total) = SQ(E) + SQ(D)$, mas não para os quadrados médios. O $QM(Total)$ não costuma ser usado.

Com os novos dados, vamos atualizar nosso quadro.

Fonte da Variação	SQ	gl	QM	F	p
Entre	1901	2	950,5		
Dentro	3386	21	161,2		
Total	5287	23	229,9		

6º passo: Estatística F

Uma estatística teste F é a razão de duas variâncias amostrais. O $QM(E)$ e $QM(D)$ são as variâncias amostrais que divididas uma pela outra fornecem o valor de F.

$$F = \frac{QM(E)}{QM(D)}$$

A partir dos dados do exemplo, temos:

$$F = \frac{950,5}{161,2} = 5,9$$

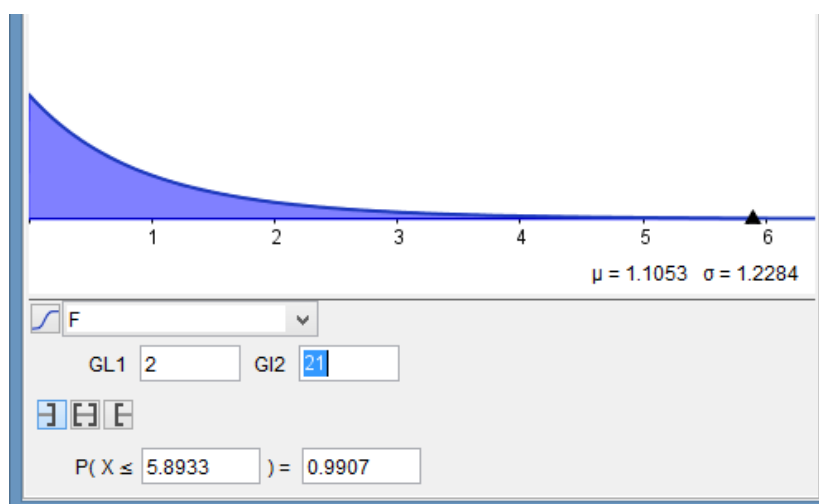
Atualizando a tabela, temos:

Fonte da Variação	SQ	gl	QM	F	p
Entre	1901	2	950,5	5,9	
Dentro	3386	21	161,2		
Total	5287	23	229,9		

A estatística F é unicaudal a direita, como podemos observar na calculadora de probabilidade do Geogebra, onde a estatística F vem da distribuição F com $gl(E)$ no numerador e $gl(D)$ no denominador. O valor-p é a área a direita do valor da estatística F

$$P(F > 5,9) = 0,0093 \approx 0,009$$

Como podemos ver na figura a seguir.



Completando o quadro, temos:

Fonte da Variação	SQ	gl	MQ	F	P
Entre	1901	2	950,5	5,9	0,009
Dentro	3386	21	161,2		
Total	5287	23	229,9		

O valor $-p = 0,009$ é menor que o nível de significância de 0,05, então rejeitamos a hipótese nula. Isto é, rejeitamos que as médias das três partes da sala foram as mesmas. Então ao menos uma das partes da sala possui média diferente.

A ANOVA não diz qual média é diferente, para identificar quem teve o melhor desempenho é necessário fazer verificações que são conhecidas como Testes de Comparações Múltiplas.

A comparação múltipla pode ser feita dois a dois como teste de hipótese:

$$t_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{e} \quad S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

1º caso: $H_0: \mu_F = \mu_M$ (Teste bilateral)

$$H_1: \mu_F \neq \mu_M$$

Obs: frente (F) e meio (M)

Significância: $\alpha = 5\%$

$$n_F = 7 \quad \text{e} \quad n_M = 9$$

$$\bar{x}_F = 75,71 \quad \text{e} \quad \bar{x}_M = 67,11$$

$$S_F = 17,63 \quad \text{e} \quad S_M = 10,95$$

$$S_p^2 = \frac{(n_F - 1)S_F^2 + (n_M - 1)S_M^2}{n_F + n_M - 2}$$

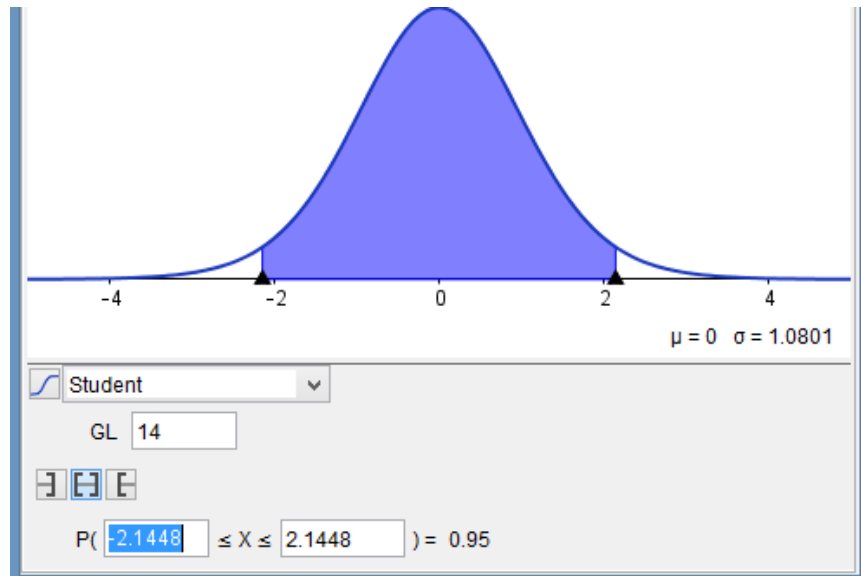
$$S_p^2 = \frac{(7 - 1)17,63^2 + (9 - 1)10,95^2}{7 + 9 - 2} = 201,72$$

$$S_p = \sqrt{201,72} = 14,2$$

$$t_{cal} = \frac{\bar{x}_F - \bar{x}_M}{S_p \sqrt{\frac{1}{n_F} + \frac{1}{n_M}}}$$

$$t_{cal} = \frac{75,71 - 67,11}{14,2 \sqrt{\frac{1}{7} + \frac{1}{9}}} = 1,20$$

$$gl = 7 + 9 - 2 = 14$$



Como $t_{cal} < 2,145$ implica que não devemos rejeitar H_0

Comparaçã ão		H_0	H_1	α	t_{cal}	$t_{critico}$	Resultado
F	M	$\mu_F = \mu_M$	$\mu_F \neq \mu_M$	5%	1,20	2,145	Não rejeitar H_0

2º caso: $H_0: \mu_F = \mu_B$ (Teste bilateral)

$H_1: \mu_F \neq \mu_B$

Obs: frente (F) e fundão (B)

Significância: $\alpha = 5\%$

$n_F = 7$ e $n_B = 8$

$\bar{x}_F = 75,71$ e $\bar{x}_B = 53,50$

$S_F = 17,63$ e $S_B = 8,96$

$$S_p^2 = \frac{(n_F - 1)S_F^2 + (n_B - 1)S_B^2}{n_F + n_B - 2}$$

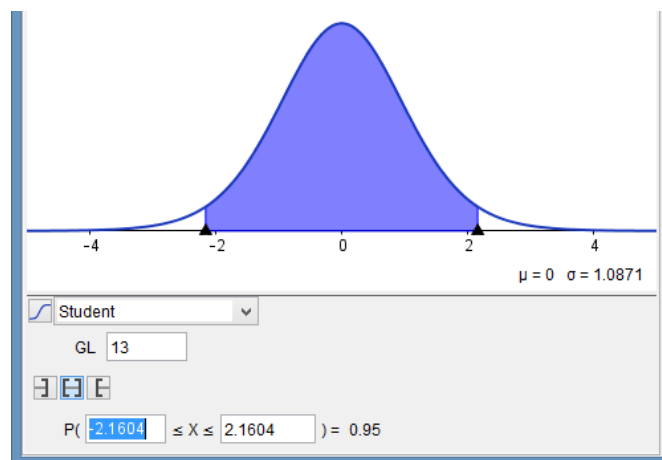
$$S_p^2 = \frac{(7 - 1)17,63^2 + (8 - 1)8,96^2}{7 + 8 - 2} = 186,68$$

$$S_p = \sqrt{186,68} = 13,66$$

$$t_{cal} = \frac{\bar{x}_F - \bar{x}_B}{S_p \sqrt{\frac{1}{n_F} + \frac{1}{n_B}}}$$

$$t_{cal} = \frac{75,71 - 53,50}{13,66 \sqrt{\frac{1}{7} + \frac{1}{8}}} = 3,14$$

$$gl = 7 + 8 - 2 = 13$$



Como $t_{cal} > 2,16$ implica que devemos rejeitar H_0

Comparaçã ão		H_0	H_1	α	t_{cal}	$t_{critico}$	Resultado
F	M	$\mu_F = \mu_M$	$\mu_F \neq \mu_M$	5%	1,20	2,145	Não rejeitar H_0
F	B	$\mu_F = \mu_B$	$\mu_F \neq \mu_B$	5%	3,14	2,160	Rejeitar H_0

3º caso: $H_0: \mu_F \leq \mu_B$ (Teste unilateral)

$$H_1: \mu_F > \mu_B$$

Obs: frente (F) e fundão (B)

Significância: $\alpha = 5\%$

$$n_F = 7 \text{ e } n_B = 8$$

$$\bar{x}_F = 75,71 \text{ e } \bar{x}_B = 53,50$$

$$S_F = 17,63 \text{ e } S_M = 8,96$$

$$S_p^2 = \frac{(n_F - 1)S_F^2 + (n_B - 1)S_B^2}{n_F + n_B - 2}$$

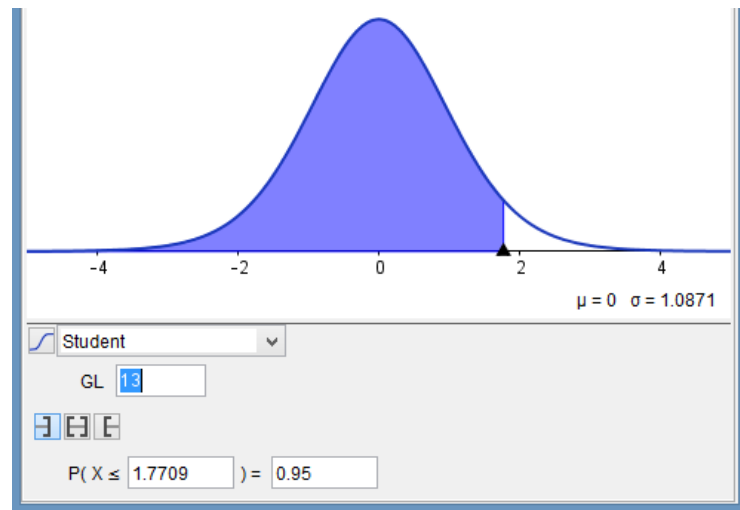
$$S_p^2 = \frac{(7 - 1)17,63^2 + (8 - 1)8,96^2}{7 + 8 - 2} = 186,68$$

$$S_p = \sqrt{186,68} = 13,66$$

$$t_{cal} = \frac{\bar{x}_F - \bar{x}_B}{S_p \sqrt{\frac{1}{n_F} + \frac{1}{n_B}}}$$

$$t_{cal} = \frac{75,71 - 53,50}{13,66 \sqrt{\frac{1}{7} + \frac{1}{8}}} = 3,14$$

$$gl = 7 + 8 - 2 = 13$$



Como $t_{cal} > 1,771$ implica que devemos rejeitar H_0 , o que significa $\mu_F > \mu_B$, a média dos alunos da frente é maior dos que sentam no fundo da sala.

Atualizando o quadro, temos:

Comparação		H_0	H_1	α	t_{cal}	$t_{crítico}$	Resultado
F	M	$\mu_F = \mu_M$	$\mu_F \neq \mu_M$	5%	1,20	2,145	Não rejeitar H_0
F	B	$\mu_F = \mu_B$	$\mu_F \neq \mu_B$	5%	3,14	2,160	Rejeitar H_0
F	B	$\mu_F \leq \mu_B$	$\mu_F > \mu_B$	5%	3,14	1,771	Rejeitar H_0

4º caso: $H_0: \mu_M = \mu_B$ (Teste bilateral)

$$H_1: \mu_M \neq \mu_B$$

Obs: meio (M) e fundo (B)

Significância: $\alpha = 5\%$

$$n_M = 9 \text{ e } n_B = 8$$

$$\bar{x}_M = 67,11 \text{ e } \bar{x}_B = 53,50$$

$$S_M = 10,95 \text{ e } S_B = 8,96$$

$$S_p^2 = \frac{(n_M - 1)S_F^2 + (n_B - 1)S_B^2}{n_M + n_B - 2}$$

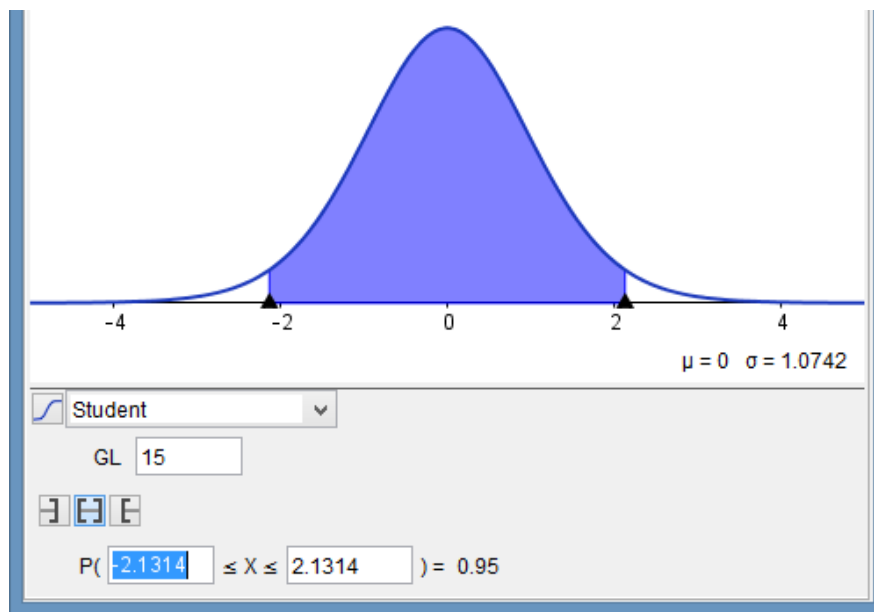
$$S_p^2 = \frac{(9 - 1)10,95^2 + (8 - 1)8,96^2}{9 + 8 - 2} = 101,41$$

$$S_p = \sqrt{101,41} = 10,07$$

$$t_{cal} = \frac{\bar{x}_M - \bar{x}_B}{S_p \sqrt{\frac{1}{n_M} + \frac{1}{n_B}}}$$

$$t_{cal} = \frac{67,11 - 53,50}{10,07 \sqrt{\frac{1}{9} + \frac{1}{8}}} = 2,78$$

$$gl = 9 + 8 - 2 = 15$$



Como $t_{cal} > 2,131$ implica que devemos rejeitar H_0 , o que significa $\mu_M \neq \mu_B$.

Atualizando o quadro, temos:

Comparação	H_0	H_1	α	t_{cal}	$t_{crítico}$	Resultado
F M	$\mu_F = \mu_M$	$\mu_F \neq \mu_M$	5%	1,20	2,145	Não rejeitar H_0
F B	$\mu_F = \mu_B$	$\mu_F \neq \mu_B$	5%	3,14	2,160	Rejeitar H_0
F B	$\mu_F \leq \mu_B$	$\mu_F > \mu_B$	5%	3,14	1,771	Rejeitar H_0
M B	$\mu_M = \mu_B$	$\mu_M \neq \mu_B$	5%	2,78	2,131	Rejeitar H_0

5º caso: $H_0: \mu_M \leq \mu_B$ (Teste unilateral)

$$H_1: \mu_M > \mu_B$$

Obs: meio (M) e fundão (B)

Significância: $\alpha = 5\%$

$n_M = 9$ e $n_B = 8$

$\bar{x}_M = 67,11$ e $\bar{x}_B = 53,50$

$S_M = 10,95$ e $S_B = 8,96$

$$S_p^2 = \frac{(n_M - 1)S_M^2 + (n_B - 1)S_B^2}{n_M + n_B - 2}$$

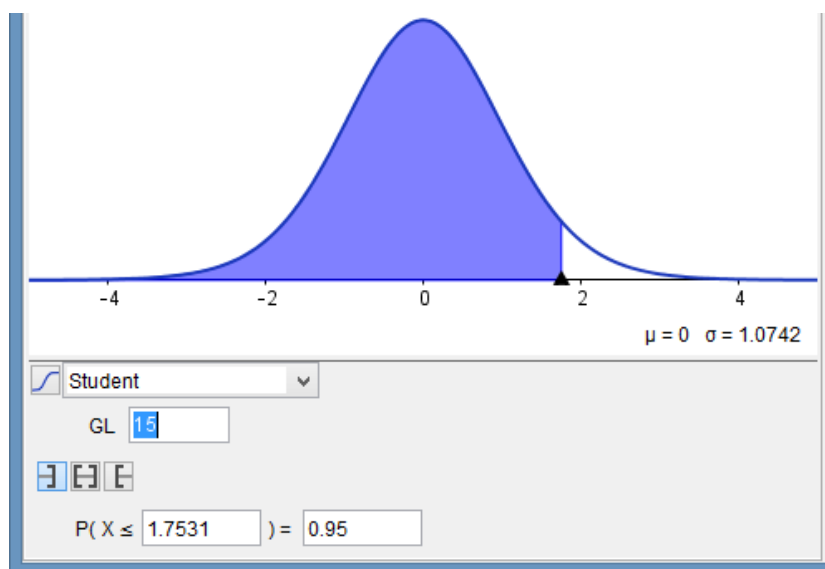
$$S_p^2 = \frac{(9 - 1)10,95^2 + (8 - 1)8,96^2}{9 + 8 - 2} = 101,41$$

$$S_p = \sqrt{101,41} = 10,07$$

$$t_{cal} = \frac{\bar{x}_M - \bar{x}_B}{S_p \sqrt{\frac{1}{n_M} + \frac{1}{n_B}}}$$

$$t_{cal} = \frac{67,11 - 53,50}{10,07 \sqrt{\frac{1}{9} + \frac{1}{8}}} = 2,78$$

$gl = 9 + 8 - 2 = 15$



Como $t_{cal} > 1,753$ implica que devemos rejeitar H_0 , o que significa $\mu_M > \mu_B$, a média dos alunos do meio é maior dos que sentam no fundão da sala.

Atualizando o quadro, temos:

Comparação	H_0	H_1	α	t_{cal}	$t_{crítico}$	Resultado
F M	$\mu_F = \mu_M$	$\mu_F \neq \mu_M$	5%	1,20	2,145	Não rejeitar H_0
F B	$\mu_F = \mu_B$	$\mu_F \neq \mu_B$	5%	3,14	2,160	Rejeitar H_0
F B	$\mu_F \leq \mu_B$	$\mu_F > \mu_B$	5%	3,14	1,771	Rejeitar H_0
M B	$\mu_M = \mu_B$	$\mu_M \neq \mu_B$	5%	2,78	2,131	Rejeitar H_0

M	B	$\mu_M \leq \mu_B$	$\mu_M > \mu_B$	5%	2,78	1,753	Rejeitar H_0
-----	-----	--------------------	-----------------	----	------	-------	----------------

Analisando os dados quadro podemos afirmar que os alunos que sentam na frente(F) e no meio(M) tem melhor desempenho dos alunos que sentam no fundam(B) da sala.

ATIVIDADE

(01) Um especialista em educação quer avaliar o nível de ensino de matemática para alunos do ensino fundamental das escolas de sua cidade. Para isto, foi realizado uma amostragem entre os alunos das escolas, no qual foi aplicado uma prova. A partir dos dados coletados na tabela, o que podemos dizer a respeito da uniformidade do ensino entre as escolas?

Nota	Escola		
	A (%)	B (%)	C (%)
1	7,88	8,46	10,00
2	8,90	11,38	6,10
3	8,86	10,74	5,84
4	12,42	9,00	6,76
5	8,40	7,56	10,82
6	11,18	12,38	10,94
7	12,60	10,86	11,48
8	7,78	10,60	13,06
9	11,68	11,02	15,20
10	10,30	8,00	9,80
Total	100,0	100,0	100,0

Tabela: Notas das provas de matemática e o percentual dos alunos, em cada escola do ensino fundamental que tirou a nota.

(02) Em um curso de extensão universitária pesquisaram-se os salários mensais (em unidades de referência) e a área de formação acadêmica dos estudantes, com base em uma amostra aleatória. Após eliminar-se os dados excessivamente destoantes, obteve-se o resultado abaixo. Podemos considerar que os salários de cada área são iguais?

	N	Média	Desvio-Padrão ²
Professores	21	4635,00	2880,00

Engenheiros	15	5130,00	4230,00
Biólogos	7	5715,00	3345,00

(03) Admitindo-se que as notas em Estatística, para cada turma, distribuem-se normalmente com mesma variância, quer-se saber se as médias obtidas nas provas de aproveitamento em cada uma das turmas são iguais, com $\alpha = 5\%$. Para tal, sortearam-se ao acaso alunos em cada uma das turmas e verificaram-se as suas notas, obtendo-se os seguintes resultados:

	ADM Diu.	ADM Not.	ECO Not.	ECO Diu.
	2,5	1,0	9,5	3,5
	6,5	0,5	2,0	5,0
	3,5	0,5	1,0	2,0
	4,0	8,0	5,0	7,0
	5,5	3,0	2,0	5,0
	5,5	0,5	4,5	4,0
	4,5	3,0	9,0	4,5
	4,0	7,0	5,5	2,0
	10,0	6,5	3,0	8,5
	5,5	5,5	4,5	4,0
	2,5	0,5	7,0	1,5
	4,0	6,5	3,5	2,5
	3,5	8,5	9,0	9,5
			8,0	3,0
			1,0	8,0
			5,5	6,5
			8,5	1,5
N	13	13	17	17
$\sum x$	61,5	51,0	88,5	78,0
$\sum x^2$	338,25	320,00	594,75	460,00

(04) Digamos que temos 6 métodos de ensino aplicados a 30 crianças cada e gostaríamos de fazer uma comparação entre os métodos. Na tabela a seguir apresenta-se os métodos de ensino A, B, C, D, E e F. Apresenta-se também, a média, o desvio padrão, o nº de crianças em cada método e o respectivo grau de liberdade - $gl = \delta = N - 1$.

	A	B	C	D	E	F
\bar{x}	75	72	76	79	82	72
S^2	173,2	168,7	170,1	169,8	172,0	167,6

<i>N</i>	30	30	30	30	30	30
<i>gl</i>	29	29	29	29	29	29

6. QUI-QUADRADO

O teste Qui Quadrado é um teste de hipóteses que se destina a encontrar um valor da dispersão para duas variáveis categóricas nominais e avaliar a associação existente entre variáveis qualitativas. É um teste não paramétrico, pois não depende de parâmetros populacionais (média e variância), sendo o princípio básico a comparação de proporções, ou seja, verificação de possíveis divergências entre as frequências observadas e esperadas para um certo evento.

O teste é utilizado para verificar se a frequência com que um determinado acontecimento observado em uma amostra se desvia significativamente ou não da frequência com que ele é esperado. Para isso, comparamos a distribuição de diversos acontecimentos em diferentes amostras, a fim de avaliar se as proporções observadas destes eventos mostram ou não diferenças significativas ou se as amostras diferem significativamente quanto às proporções desses acontecimentos.

Para o uso desse teste é considerado que os grupos devem ser independentes, os itens de cada grupo são selecionados aleatoriamente, e que as observações devem ser frequências ou contagens, e que cada observação pertence a uma e somente uma categoria além de que a amostra deve ser relativamente grande (pelo menos 5 observações em cada célula e, no caso de poucos grupos, pelo menos 10. Exemplo: em tabelas 2x 2)

EXEMPLO

(01) Com objetivo de se verificar se há ou não dependência entre o rendimento escolar com o turno de estudo dos alunos foi feito um levantamento com um amostra de 800.000 alunos dos ensinos fundamental e médio, em que foi selecionado, ao acaso, 500.000 alunos do turno matutino e 300.000 alunos do turno vespertino, e foi verificado o rendimento escolar da cada um deles, sendo classificado como: ótimo, regular e péssimo, estando os dados na tabela a seguir.

	Rendimento Ótimo	Rendimento Regular	Rendimento Péssimo	Totais
Turno Matutino	80.000	300.000	120.000	500.000
Turno Vespertino	150.000	60.000	90.000	300.000
Totais	230.000	360.000	210.000	800.000

Observa-se na amostra total que o Rendimento ótimo é: $230.000/800.000 = 0,2875$

isto é, 28,75% dos alunos pesquisados apresentam rendimento ótimo.

Cálculo do Resultado Esperado:

Resultado Observado	Proporção	Resultado Esperado
80.000	$230.000/800.000 = 0,2875$	$0,2875 \times 500.000 = 143.750$
300.000	$360.000/800,00 = 0,4500$	$0,4500 \times 500.000 = 225.000$
120.000	$210.000/800,00 = 0,2625$	$0,2625 \times 500.000 = 131.250$
150.000	$230.000/800.000 = 0,2875$	$0,2875 \times 300.000 = 86.250$
60.000	$360.000/800.000 = 0,4500$	$0,4500 \times 300.000 = 135.000$
90.000	$210.000/800.000 = 0,2625$	$0,2625 \times 300.000 = 78.750$

Cálculo do χ_0^2 :

Resultado Observado	Resultado Esperado	Diferença entre o Resultado Observado e o Resultado Esperado	Diferença elevado ao quadrado	Quociente entre a diferença ao quadrado e o Resultado Esperado
80.000	143.750	$80.000 - 143.750 = - 63.750$	4.064.062.500	28.271,7391
300.000	225.000	$300.000 - 225.000 = 75.000$	5.625.000.000	25.000,0000
120.000	131.250	$120.000 - 131.250 = - 11.250$	126.562.500	964,2857
150.000	86.250	$150.000 - 86.250 = 63.750$	4.064.062.500	47.119,5652
60.000	135.000	$60.000 - 135.000 = - 75.000$	5.625.000.000	41.666,6667
90.000	78.750	$90.000 - 78.750 = 11.250$	126.562.500	1.607,1429

A soma de todos os valores da última coluna é: $\chi_0^2 = 144.629,3996$

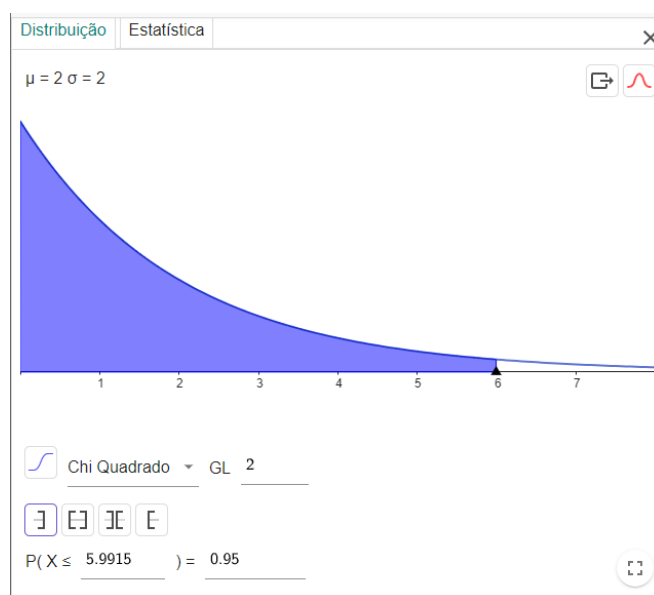
Cálculo do Grau de liberdade: O número de graus de liberdade para o teste da independência é calculado pelo produto entre o número de linhas da tabela de dados subtraído de um e o número de colunas da tabela de dados, subtraído de um:

$$gl = (2 - 1) * (3 - 1) = 1 * 2 = 2$$

Tabela: Qui-Quadrado

g.l	α									
	0,995	0,990	0,975	0,950	0,900	0,100	0,050	0,025	0,010	0,005
1			0,001	0,004	0,016	2,706	3,841	5,024	6,635	7,879
2	0,010	0,020	0,051	0,103	0,211	4,605	5,991	7,378	9,210	10,597
3	0,072	0,115	0,216	0,352	0,584	6,251	7,815	9,348	11,345	12,838
4	0,207	0,297	0,484	0,711	1,064	7,779	9,488	11,143	13,277	14,860
5	0,412	0,554	0,831	1,145	1,610	9,236	11,071	12,833	15,086	16,750
6	0,676	0,872	1,237	1,635	2,204	10,645	12,592	14,449	16,812	18,548
7	0,989	1,239	1,690	2,167	2,833	12,017	14,067	16,013	18,475	20,278
8	1,344	1,646	2,180	2,733	3,490	13,362	15,507	17,535	20,090	21,955
9	1,735	2,088	2,700	3,325	4,168	14,684	16,919	19,023	21,666	23,589
10	2,156	2,558	3,247	3,940	4,865	15,987	18,307	20,483	23,209	25,188
11	2,603	3,053	3,816	4,575	5,578	17,275	19,675	21,920	24,725	26,757
12	3,074	3,571	4,404	5,226	6,304	18,549	21,026	23,337	26,217	28,299
13	3,565	4,107	5,009	5,892	7,042	19,812	22,362	24,736	27,688	29,819
14	4,075	4,660	5,629	6,571	7,790	21,064	23,685	26,119	29,141	31,139
15	4,601	5,229	6,262	7,261	8,547	22,307	24,996	27,488	30,578	32,801
16	5,142	5,812	6,908	7,962	9,312	23,542	26,296	28,845	32,000	34,267
17	5,697	6,408	7,564	8,672	10,085	24,769	27,587	30,191	33,409	35,718
18	6,265	7,015	8,231	9,390	10,865	25,989	28,869	31,526	34,805	37,156
19	6,844	7,633	8,907	10,117	11,651	27,204	30,144	32,852	36,191	38,582
20	7,434	8,260	9,591	10,851	12,443	28,412	31,410	34,170	37,566	39,997

Na tabela do Qui-Quadrado devemos localizar o parâmetro x_t^2 correspondente a $\alpha = 0,05$ (ou $\alpha = 5\%$) e ao valor gl calculado. Para o exemplo temos: $x_t^2 = 5,991$



Conclusão: Como x_0^2 (144.629,3996) $>$ x_t^2 (5,991), a hipótese H_0 é rejeitada, ou seja a variável “turno de estudo dos alunos dos ensinos fundamental e médio” e a variável “rendimento escolar” são independentes.

(02) Deseja-se saber se o número de pacientes que chega a um hospital é aproximadamente o mesmo em todos os dias da semana. Esta informação é importante para o administrador do hospital, que pode considerar a realocação de funcionários, caso, em dias específicos da semana, o hospital receba mais pacientes que em outros. Em uma determinada semana, registrou-se o número de pacientes que procurou a assistência do hospital, por dia da semana. A tabela a baixo apresenta esse registro:

Dias da Semana	Dom	Seg	Ter	Quar	Quin	Sex	Sab	Total
N paciente	36	20	17	22	21	26	33	175

Há evidências de que alguns dias receba mais pacientes que outros?

H_0 : a probabilidade de um paciente procurar o hospital é a mesma para todos os dias da semana

H_A : a probabilidade de um paciente procurar o hospital não é a mesma para todos os dias da semana

Seja $\pi_j = p$ um paciente chegar ao hospital no dia $f_j^d = \frac{1}{7}$ para $j = 1, 2, \dots, 7$

Então a hipótese nula pode ser reescrita na forma:

$H_0: \pi_j = \pi = 1/7 \Rightarrow$ hipótese de homogeneidade

Se H_0 é verdadeiro, então $1/7$ de 175 pacientes seriam esperados em cada dia da semana $\frac{175}{7} = 25$ pacientes por dia. (frequência esperada)

Para comparara os dados observados com o que se esperaria sob a hipótese nula, para cada dia da semana calculamos

$$\frac{(\text{frequencia observada} - \text{frequencia esperada})^2}{\text{frequencia esperada}}$$

frequencia esperada $\neq 0$

$$\frac{(36 - 25)^2}{25} + \frac{(20 - 25)^2}{25} + \frac{(17 - 25)^2}{25} + \frac{(22 - 25)^2}{25} + \frac{(21 - 25)^2}{25} + \frac{(26 - 25)^2}{25} + \frac{(33 - 25)^2}{25} = 12,0$$

Se os dados compatíveis com H_0 , então as diferenças não seriam muito grandes. Deseja-se uma estatística que nos permita avaliara o que é diferença suficientemente grande.

\Rightarrow Estatística Qui-Quadrado, χ^2

Podem-se classificar os resultados de um evento em k categorias e as frequências observadas serão O_1, O_2, \dots, O_k em que O_i representa a contagem registrada para a i -ésima categoria. O total de casos observados é fixo e representado por

$$N = O_1 + O_2 + \dots + O_k$$

Vamos representar por $(\pi_1, \pi_2, \dots, \pi_k)$ a distribuição de probabilidade esperada em cada uma das categorias $j | H_0, j = 1, \dots, k$

As frequências esperadas definidas na H_0 são representados por:

e_1, e_2, \dots, e_k em que $e_j = N\pi_j$ é a frequência esperada para j -ésima categoria.

Note que: $e_1 + e_2 + \dots + e_k = N$

Então, a estatística qui-quadrado é dada por:

$$\chi^2 = \frac{(o_1 - e_1)^2}{e_1} + \frac{(o_2 - e_2)^2}{e_2} + \dots + \frac{(o_k - e_k)^2}{e_k}$$

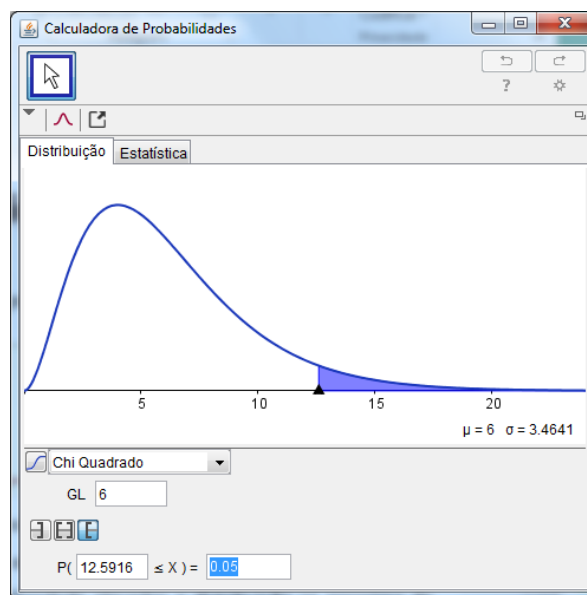
Se o_i é muito próximo de e_i então χ^2 será baixo.

Se o_i é muito diferente de e_i então χ^2 será alto.

Se $\chi^2 > \chi_c^2$, rejeita-se H_0 , caso contrário aceita-se H_0

Sob a hipótese nula de homogeneidade a estatística χ^2 tem distribuição aproximadamente qui-quadrada com $(k - 1)$ graus de liberdade.

A distribuição qui-quadrado apresenta densidade de probabilidade dependente do número de liberdade $d.f.$



- 1) Qui-quadrado assume somente valores positivos
- 2) A distribuição é assimétrica

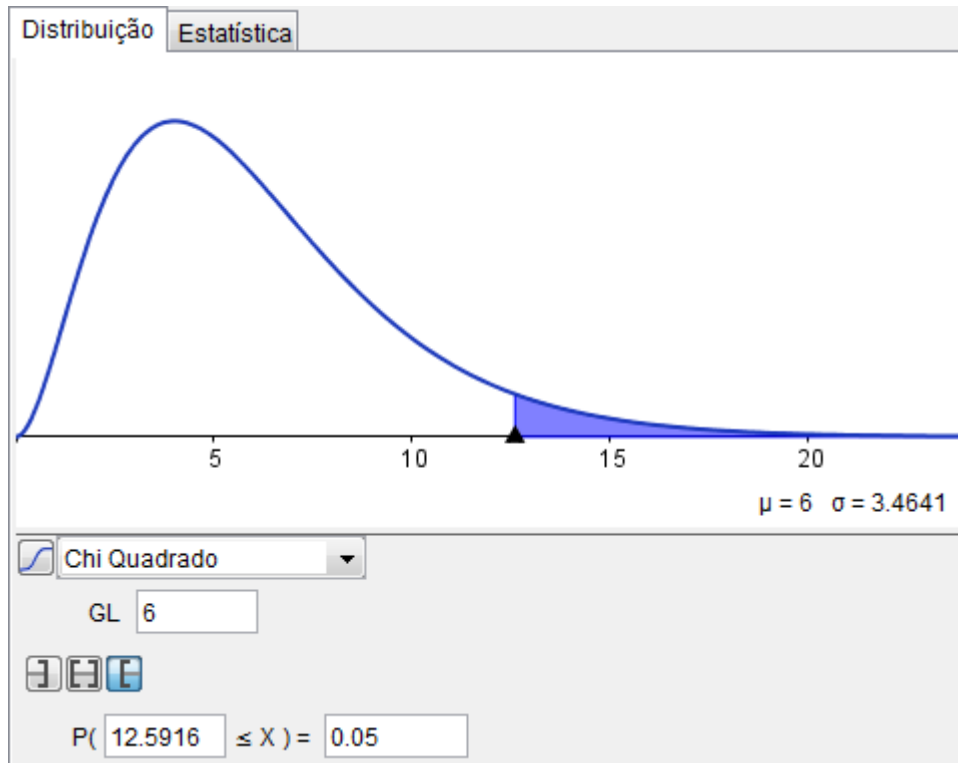
3) Para graus de liberdade elevados a distribuição se aproxima da normal

$$\chi^2 = 12,0$$

$$\text{g.l.} = 7 - 1 = 6$$

$$\alpha = 0,05$$

$$\chi_c^2 = 12,5916$$



Como $\chi^2 < \chi_c^2$, aceitamos H_0

Obs: Apesar de considerarmos apenas a área superior, o teste é bicaudal

(03) Considere a tabela 2X2 a baixo, na qual estão os resultados de um estudo que investiga a efetividade dos capacetes de segurança de bicicleta na prevenção de lesões na cabeça. Os dados consistem de uma amostra aleatória de 793 indivíduos envolvidos em acidentes ciclísticos durante um certo período. Deseja-se testar, com $\alpha = 0,05$ Se o uso de capacetes tem funcionado, efetivamente, como um fator de proteção.

Lesão na cabeça	Uso de capacete		Total
	Sim	Não	
Sim	17	218	235
Não	130	428	558
Total	147	646	793

H_0 : a proporção de pessoa que sofrem lesões na cabeça na população de indivíduos de capacete no momento do acidente é igual a proporção de pessoas que sofreram lesão na cabeça entre os não usuários de capacete.

H_A : a proporção de pessoa que sofrem lesões na cabeça é diferente nos dois grupos. Usuários e não usuários de capacete.

Primeira etapa: Calcular as frequências esperadas para cada caso, para H_0 verdadeiro:

$$p\{\text{sofrer lesão}\} = \frac{235}{793} = 0,296 \text{ ou } 29,6\%$$

$$p\{\text{não sofrer lesão}\} = \frac{558}{793} = 0,704 \text{ ou } 70,4\%$$

Lesão na cabeça	Uso de capacete		Total
	Sim	Não	
Sim	17	218	235
Não	130	428	558
Total	147	646	793

$147 \cdot \left(\frac{235}{793}\right) = 43,6$

$646 \cdot \left(\frac{235}{793}\right) = 191,4$

$147 \cdot \left(\frac{558}{793}\right) = 103,4$

$646 \cdot \left(\frac{558}{793}\right) = 454,6$

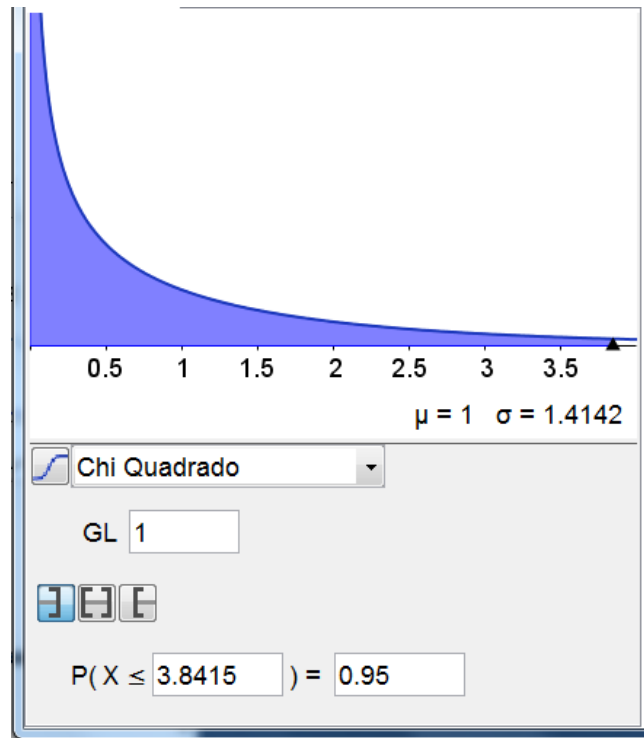
$$\text{grau de liberdade} = (2 - 1)(2 - 1) = 1$$

Então, a estatística qui-quadrado é dada por:

$$\chi^2 = \frac{(17 - 43,6)^2}{43,6} + \frac{(130 - 103,4)^2}{103,4} + \frac{(218 - 191,4)^2}{191,4} + \frac{(428 - 454,6)^2}{454,6}$$

$$\chi^2 = 28,32$$

$$\text{Como } \chi_c^2 = 3,8415$$



Para $d.f. = 1$, temos que a tabela χ^2 que o valor crítico para α de 5% é 3,8415 logo $\chi^2 > \chi_c^2$ logo rejeitamos H_0 e concluímos que as proporções de indivíduos que sofrem lesão na cabeça não são iguais nos dois grupos. Portanto, o uso de capacete funciona como fator de proteção.

ANEXO I - Respostas de 96 alunos de um curso da área de biológica de uma universidade pública do Pará, que responderam ao questionário com 45 questões, sendo 15 referentes ao ensino, 15 relacionado a aprendizagem e 15 voltadas a avaliação.

PE01	PE02	PE03	PE04	PE05	PE06	PE07	PE08	PE09	PE10	PE11	PE12	PE13	PE14	PE15	PA01	PA02	PA03	PA04	PA05	PA06	PA07	PA08	PA09	PA10	PA11	PA12	PA13	PA14	PA15	PAV01	PAV02	PAV03	PAV04	PAV05	PAV06	PAV07	PAV08	PAV09	PAV10	PAV11	PAV12	PAV13	PAV14	PAV15		
3	3	2	3	1	1	2	3	4	2	2	1	2	2	2	4	2	3	4	2	3	2	1	3	3	4	2	2	3	2	2	1	3	3	3	3	1	1	2	2	1	3	1	1	4	4	
3	3	2	3	3	3	2	3	4	3	3	2	3	2	3	3	2	3	3	3	3	2	1	2	2	2	3	2	3	3	2	2	4	4	4	4	2	1	3	1	3	4	1	3	2	4	
3	3	1	3	2	2	3	2	3	1	2	2	2	3	4	4	1	3	4	4	4	2	3	3	3	3	2	2	4	3	1	1	3	1	4	3	3	3	2	2	4	3	2	3	4		
3	2	1	3	2	2	2	3	3	1	1	1	3	1	3	3	1	1	3	1	1	3	2	2	2	3	2	1	3	1	1	2	3	4	4	2	2	2	2	2	3	2	2	2	3		
3	3	2	3	2	2	2	2	3	2	3	2	3	3	3	3	2	3	2	2	3	3	2	2	1	2	3	2	3	3	3	2	3	3	3	3	2	2	1	2	2	3	3	2	3	3	
3	3	3	3	4	3	2	3	3	3	3	2	4	4	4	4	2	3	3	2	3	3	3	4	3	4	3	2	4	4	3	2	3	3	4	3	3	3	2	3	4	4	3	3	3	3	
2	3	2	3	2	3	2	3	4	2	3	1	3	3	4	2	1	3	4	3	3	2	2	3	3	3	3	2	3	3	1	2	3	3	4	1	2	3	1	2	4	1	2	3	3		
4	2	2	3	4	2	3	3	3	1	3	1	3	2	3	2	1	2	4	2	3	2	1	3	1	3	3	2	3	2	3	1	3	3	1	2	2	3	1	3	1	1	3	2	1		
3	2	1	2	3	2	2	2	3	1	2	2	3	2	4	3	2	3	3	3	3	2	2	3	3	3	2	3	3	2	3	3	3	3	3	2	1	2	2	3	2	3	2	3	2	3	
3	3	2	4	3	3	2	3	3	1	2	1	2	3	4	3	2	3	4	2	3	4	1	3	3	3	2	1	3	3	2	2	3	2	4	2	2	2	2	2	2	4	2	2	2	3	
2	2	2	3	2	2	2	2	3	1	1	2	3	2	3	2	1	4	3	3	3	2	2	2	1	3	2	2	3	2	3	1	1	3	3	1	2	1	3	3	3	3	2	2	3		
4	3	2	3	2	1	1	2	3	1	2	1	3	2	4	4	2	4	4	4	4	2	2	3	3	4	2	1	3	4	1	1	1	1	4	2	1	1	2	1	4	2	1	1	4		
3	3	2	3	3	2	2	2	1	1	2	1	2	3	3	4	1	3	2	2	4	2	2	2	3	2	3	2	2	2	2	2	2	3	3	4	2	2	1	2	2	4	1	2	2	4	
4	3	2	3	3	2	2	3	3	2	3	2	3	3	3	3	2	3	3	2	4	3	2	3	3	4	3	3	4	3	4	3	3	4	4	3	4	3	2	3	4	3	3	4	4		
3	1	2	3	3	2	2	2	3	2	1	1	2	2	3	1	1	1	4	2	2	2	2	3	3	3	2	2	4	2	4	3	2	4	4	2	3	2	4	2	4	3	2	2	4		
1	2	1	3	2	1	3	3	3	2	2	1	1	2	4	3	1	4	4	1	3	2	2	2	3	4	2	1	3	2	2	1	4	3	3	2	2	2	2	2	2	3	1	2	3	3	
3	3	2	3	3	2	3	3	2	2	3	1	3	2	3	4	1	3	3	2	3	3	2	2	3	3	2	2	3	2	2	1	2	2	3	1	2	2	2	2	3	2	2	2	3		
4	3	2	3	3	2	2	3	3	2	2	2	2	3	3	2	1	3	4	3	3	4	2	3	3	4	3	2	4	2	2	2	3	3	4	2	3	3	1	2	4	2	2	3	3		
2	3	3	3	3	3	2	2	4	2	3	2	2	3	4	3	2	4	4	4	4	2	2	3	3	3	3	2	4	2	4	2	3	2	4	3	3	3	2	2	3	2	2	3	3		
2	2	2	3	3	2	2	3	3	1	1	2	2	2	4	3	1	3	4	4	3	2	1	3	2	3	3	2	3	2	1	1	3	2	4	2	3	2	1	2	4	2	1	3	3		
3	2	2	2	3	3	1	2	3	1	1	1	2	2	3	4	2	4	4	4	4	2	2	3	4	3	3	2	2	3	3	2	4	4	4	4	2	2	1	2	3	3	1	2	4		
3	3	3	3	4	2	2	2	3	1	2	2	2	2	4	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	2	2	2	3	1	3	3	3	3	3	3	3	3	3	3	
4	3	3	3	2	2	2	3	3	2	3	1	2	3	4	4	2	2	4	1	2	1	2	3	3	3	2	2	3	3	3	2	4	4	3	3	3	3	2	2	3	3	2	3	2	3	4
3	2	2	2	2	2	2	3	3	3	3	2	2	2	2	2	3	4	4	4	4	2	2	4	3	3	2	2	4	4	3	2	1	3	3	3	2	2	2	3	3	2	3	2	3	2	3
3	3	2	3	1	3	1	2	3	2	3	1	2	3	4	4	2	2	4	1	2	3	2	2	3	3	1	2	3	2	2	1	3	2	3	1	1	3	1	1	3	2	2	2	4		
4	4	3	4	3	3	3	3	4	3	3	2	3	2	4	3	2	3	3	3	3	3	3	3	3	3	4	3	3	3	3	3	3	3	3	3	3	3	3	3	2	3	3	3	2	3	3
4	3	2	3	4	2	1	1	2	1	2	2	2	2	4	4	2	4	4	2	4	3	2	3	4	4	4	3	4	2	1	2	3	1	4	2	2	3	2	2	4	2	2	4	4	4	
3	2	2	3	2	1	2	2	3	2	3	2	3	2	4	3	1	2	4	2	4	4	2	3	4	3	1	2	4	2	1	1	3	2	4	2	1	2	2	1	4	1	1	3	4		
2	3	1	3	2	1	1	1	3	1	3	2	2	2	4	3	1	3	4	4	4	1	1	1	3	4	2	2	3	2	2	1	1	1	4	1	1	3	2	1	4	4	1	1	4		

4	3	3	3	3	2	2	3	3	2	3	2	2	3	4	4	2	2	3	3	3	3	2	3	3	4	3	3	3	3	2	3	3	3	3	3	2	2	3	4	2	3	4				
3	3	3	2	3	2	2	3	2	4	3	3	3	2	3	3	2	2	2	2	3	2	2	1	2	2	2	2	2	3	3	2	2	2	1	2	2	2	2	2	2	2	2	2	3	2	
2	3	2	3	2	2	2	2	3	1	3	1	2	3	3	3	1	4	4	2	3	3	2	3	3	3	3	2	3	3	3	2	3	3	3	2	3	2	2	3	1	1	1	3	2		
3	3	3	3	3	3	2	3	3	3	3	3	3	3	3	3	1	3	3	3	3	3	3	3	3	3	3	3	3	3	2	3	3	3	3	3	2	2	2	2	3	2	2	3	2		
3	3	2	3	3	2	2	2	3	2	3	2	2	2	4	3	1	4	3	3	3	2	2	3	3	4	2	2	4	3	3	2	2	2	3	2	1	3	3	2	2	1	1	3	2		
4	3	3	4	3	2	2	3	4	2	3	2	3	3	4	3	3	2	3	3	4	4	3	3	4	3	4	3	4	3	3	4	3	4	4	4	3	4	3	3	3	3	3	3	4		
2	2	2	3	1	1	2	1	3	2	2	1	2	3	4	2	2	3	4	4	3	2	1	3	1	4	2	1	4	3	2	2	3	2	3	2	3	2	2	3	2	2	3	3	2	2	4
3	2	2	2	3	2	2	3	4	3	2	2	2	3	3	2	1	4	4	4	4	2	2	3	2	3	3	2	4	3	2	2	3	3	3	3	3	2	2	1	1	4	1	1	2	2	
3	3	2	3	2	2	2	3	1	3	2	2	2	3	3	2	3	2	3	3	2	2	2	4	4	3	2	1	4	3	1	2	4	2	4	3	1	1	1	3	4	2	2	2	3		
2	2	2	3	2	2	2	2	3	2	2	2	2	2	3	3	2	3	3	3	3	2	2	2	3	3	2	2	3	2	2	1	2	2	3	2	2	2	2	2	2	2	3	1	2	2	3
3	2	2	2	2	2	2	2	2	1	2	1	1	2	2	3	1	3	3	3	3	1	1	3	3	4	2	2	3	2	1	1	1	1	1	3	1	1	1	1	1	4	1	1	1	4	
2	2	3	3	3	3	2	2	3	2	2	2	2	3	4	4	1	3	4	3	4	1	2	2	3	3	2	3	3	3	1	2	2	2	3	2	2	2	2	2	1	4	4	2	2	3	
3	4	2	3	4	2	2	3	4	3	3	2	3	3	3	3	2	3	4	2	4	2	3	3	3	3	3	3	3	4	2	2	4	4	4	3	3	3	2	2	4	1	3	4	4		
3	3	3	3	3	2	2	2	3	3	3	3	3	3	3	3	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2	3	3	3	3	3	
3	3	2	3	2	2	2	3	3	3	3	2	2	2	4	3	2	3	4	2	3	2	3	2	3	4	3	2	4	3	3	1	2	1	3	2	1	3	2	2	3	2	2	2	2	2	
2	3	2	2	2	2	3	2	3	3	3	1	2	2	4	4	2	3	3	2	2	3	2	2	3	4	1	3	3	2	1	1	2	2	4	2	2	3	2	1	3	3	2	3	3		
4	3	3	4	3	2	2	1	4	2	2	1	2	2	4	3	1	4	3	4	3	3	2	2	3	3	3	2	3	3	1	1	3	1	4	2	3	2	1	2	3	1	2	2	3		
3	3	2	4	3	3	1	2	3	2	3	2	2	3	4	3	1	3	3	2	3	3	2	2	2	3	2	2	4	1	3	3	4	3	3	2	2	2	3	2	2	2	3	2	2	3	3
3	3	2	3	3	2	2	3	3	2	2	1	2	2	3	2	1	3	3	2	3	2	2	2	3	3	2	2	4	2	2	2	3	3	2	2	4	2	2	2	3	3	2	2	2	3	
2	2	1	3	2	1	2	1	2	1	3	1	1	2	4	3	1	3	4	4	4	3	2	2	2	3	1	1	4	2	2	1	2	2	2	3	3	2	2	2	1	1	1	1	1		
4	3	2	3	3	2	2	2	3	2	3	2	3	3	3	3	3	2	4	3	3	3	2	4	3	3	3	3	3	2	2	3	2	3	3	3	3	2	2	3	3	3	2	3	3	3	
2	2	1	3	2	1	1	1	3	2	2	1	2	1	4	3	3	2	4	2	4	2	2	3	4	3	4	1	4	2	2	1	3	2	4	2	2	3	1	2	4	1	2	3	4		
4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	
3	3	2	3	3	2	2	2	3	2	3	2	3	3	3	2	2	3	3	3	3	3	2	3	3	3	3	2	3	3	3	1	3	1	4	2	1	2	1	2	3	3	2	3	3		
3	3	2	3	3	2	1	3	3	1	3	1	3	2	4	4	2	4	3	3	3	2	2	3	2	3	2	1	3	4	3	2	3	3	3	2	2	3	1	1	3	1	2	2	3		
3	3	3	3	2	2	2	3	3	2	2	2	2	2	3	2	1	3	4	3	3	3	2	2	3	3	3	2	3	2	2	2	3	2	3	2	2	2	2	2	2	2	3	1	2	2	3
3	2	2	2	3	2	2	2	3	2	2	2	2	2	4	4	2	3	4	2	3	2	2	2	2	3	3	2	3	3	2	2	2	3	2	2	2	2	2	2	2	2	3	2	2	2	3
4	4	2	3	3	3	1	2	3	1	1	2	2	3	3	1	1	3	3	2	3	3	2	3	3	3	2	1	4	3	3	1	3	2	4	2	3	2	1	1	4	1	1	4	2		
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
2	2	2	3	2	1	2	3	3	3	2	1	2	2	4	4	2	4	4	4	4	3	1	3	3	3	3	2	4	3	3	1	2	3	3	2	2	2	3	3	4	1	2	3	4		
3	3	3	3	3	3	3	3	3	2	3	3	3	3	3	3	2	3	3	3	3	2	2	3	3	3	2	3	3	3	2	2	3	3	3	2	2	3	3	2	2	3	2	2	2	3	
3	2	1	3	2	2	2	2	3	1	2	1	2	2	3	4	1	3	4	4	4	1	1	3	3	3	2	2	4	3	2	1	2	2	4	2	1	3	1	2	3	2	2	4	4		
3	3	1	4	1	2	2	2	4	1	3	2	2	2	4	3	2	4	4	3	4	2	2	3	3	4	2	1	4	3	1	1	3	2	2	2	2	3	1	1	4	3	2	3	3		



FÁBIO JOSÉ DA COSTA ALVES - Licenciatura em Matemática pela União das Escolas Superiores do Pará, Licenciatura em Ciências de 1º Grau pela União das Escolas Superiores do Pará, Graduação em Engenharia Civil pela Universidade Federal do Pará. Possui Mestrado e Doutorado em Geofísica pela Universidade Federal do Pará e Pós-Doutorado pelo Programa de Pós-Graduação em Ensino de Ciências e Matemática da Universidade Federal do Rio Grande do Norte. Professor da Universidade do Estado do Pará. Docente do Programa de Pós-Graduação em Educação/UEPA e Docente do Programa de Pós-Graduação em Ensino de Matemática/UEPA. Entre os anos de 2019 a 2023 coordenou o Programa de Pós-Graduação em Ensino de Matemática – Curso de Mestrado. Líder do Grupo de Pesquisa em Ensino de Matemática e Tecnologias. Experiência em desenvolvimento de software educativo para o ensino de matemática.



ADMILSON ALCANTARA DA SILVA - Graduado em Estatística pela Universidade Federal do Pará, possui Especialização em Estatística pela Universidade Federal do Pará, Mestrado em Estatística pela Universidade Federal do Pará e Doutor em Engenharia de Produção pela Universidade Federal de São Carlos/SP. Professor da Universidade do Estado do Pará e Universidade Federal Rural da Amazônia. Tem experiência na área de Probabilidade, Estatística e Pesquisa Operacional, com ênfase em Probabilidade, Estatística e Otimização, atuando principalmente nos seguintes temas: turismo, educação, economia sócio, agronomia, ambiental e de política social. É membro da Associação Brasileira de Pesquisa Operacional - ABEPRO, pesquisador do Grupo de Pesquisa Operacional do Programa de Pós-Graduação da Universidade Federal de São Carlos - GPO, e membro do Grupo de Pesquisa em Estatística Aplicada e Ensino de Matemática da Universidade



Possui graduação em Estatística pela Universidade Federal do Pará, Mestre em Estatística pela Universidade Federal do Pará e Doutor em Engenharia de produção pela Universidade Federal de São Carlos. Professor Assistente da Universidade do Estado do Pará e Membro do Conselho Universitário da UEPA. Tem experiência na área de Probabilidade e Estatística, com ênfase em Estatística aplicada. Atua como vice-líder no grupo de pesquisa Estatística Aplicada e Ensino de Matemática (GPEAEM/UEPA). Entre os anos de 2011 a 2015 foi coordenador do curso de licenciatura em Matemática plano nacional de formação de docentes da educação básica - Parfor, pela Universidade do Estado do Pará. Chefe do Departamento de Matemática Estatística e Informática.