Como escolher o teste estatístico

UM GUIA PARA O PESQUISADOR INICIANTE

CLÁUDIA REGINA MOYA

Como escolher o teste estatístico

UM GUIA PARA O PESQUISADOR INICIANTE

CLÁUDIA REGINA MOYA

Atividade apresentada à disciplina de Estágio em Docência do Programa de Mestrado Interdisciplinar em Ciências da Saúde da Universidade Cruzeiro do Sul

SÃO PAULO
2020

Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)

Moya, Claudia Regina

Como escolher o teste estatístico [livro eletrônico] : um guia para o pesquisador iniciante / Claudia Regina Moya. -- São Paulo : Ed. da Autora, 2021.
PDF

Bibliografia.

ISBN 978-65-00-24278-2

1. Amostragem (Estatística) 2. Bioestatística 3. Estatística 4. Pesquisa - Métodos estatísticos 5. Pesquisa - Planejamento 6. Saúde - Pesquisa I. Título.

21-68140 CDD-001.422

Índices para catálogo sistemático:

1. Teste estatístico : Metodologia : Pesquisa 001.422

Cibele Maria Dias - Bibliotecária - CRB-8/9427

“Agradeço à Profa. Dra. Maria Teresa Botti Rodrigues dos Santos e à Profa. Dra. Michele Baffi Diniz por me aceitarem como estagiária em docência na disciplina de Bioestatística; à Profa. Lídia Germana Moya pela revisão do texto desse e-book e ao Prof. Dr. Leonardo Costa pela orientação nessa etapa importante da minha vida.”

Prefácio

Esse e-book foi desenvolvido como atividade avaliativa apresentada à disciplina de Bioestatística, do Programa de Mestrado Interdisciplinar em Ciências da Saúde da Universidade Cruzeiro do Sul. Seu objetivo central é guiar o pesquisador iniciante, principalmente da área da saúde, na escolha do teste estatístico. Além disso, o guia traz uma revisão dos principais aspectos metodológicos, que influenciam o emprego do teste estatístico.

O conteúdo desse e-book foi baseado, principalmente, em três obras, as quais os créditos devem ser atribuídos: livro “Introdução à Bioestatística”, de Mario F. Triola[1], lirvo “Estatística Básica” de Wilton de O. Bussab e Pedro A. Morettin[2] e livro “Como elaborar projetos de pesquisa” de Antônio Carlos Gil[3].

Espera-se que esse guia beneficie o pesquisador que está iniciando na área científica, fornecendo informações que contribuam com a elaboração do projeto de pesquisa e com a escolha do teste estatístico mais adequado para responder à pergunta da pesquisa.

Sumário

1 Introdução

As pesquisas estão presentes em todas as áreas do conhecimento: através delas pode-se estimar quem vencerá uma eleição presidencial, avaliar a efetividade de um tratamento ou investigar o perfil do cliente, auxiliando o empreendedor na implementação de estratégias empresariais. Com a vasta utilização de pesquisas no cotidiano, pode-se imaginar as consequências de uma interpretação equivocada e enviesada de seus resultados, o que pode acontecer caso os testes estatísticos sejam empregados indevidamente.

A estatística é amplamente utilizada em pesquisas na área da saúde, a fim de descrever as características dos pacientes de uma amostra, sumarizar os dados e representá-los de forma organizada, comparar e correlacionar dois ou mais grupos, e fazer predições e inferências para determinada população, a partir de dados provenientes de amostras. A análise estatística existente nas pesquisas científicas permite a avaliação dos dados coletados, para que os resultados e conclusões possam auxiliar na tomada de decisão clínica. Caso os testes estatísticos sejam empregados inadequadamente, as interpretações dos dados são prejudicadas, podendo trazer sérias consequências ao paciente, por exemplo, com a implementação de tratamentos sem respaldo científico.

É fundamental empregar adequadamente o teste estatístico a fim de evitar interpretações equivocadas em pesquisas científicas. Entretanto, nem sempre a decisão sobre qual teste estatístico utilizar é tarefa fácil para o pesquisador, devido à carência de conhecimento sobre estatística básica e à falta de familiaridade com os *softwares* estatísticos. Portanto, o objetivo desse e-book é revisar os conceitos básicos de estatística e servir como guia ao pesquisador iniciante acerca de qual teste estatístico utilizar, principalmente aos profissionais da área da saúde.

2 Planejamento da pesquisa

Como vimos, a escolha correta do teste estatístico é fundamental para confiabilidade dos resultados na pesquisa científica. Mas como decidir qual teste estatístico utilizar? Em primeiro lugar, é necessária a elaboração de um planejamento, apresentado na forma de um projeto de pesquisa, que deve especificar o objetivo, a justificativa, o método de coleta e análise de dados, bem como quais testes estatísticos serão empregados. A elaboração de um projeto de pesquisa permite avaliar a viabilidade de recursos (humanos, financeiros e materiais), determinar prazos para realização das etapas e minimizar a chance de erros durante a execução da pesquisa.

2 – 1 Formulação da Pergunta da Pesquisa

A primeira etapa para elaboração de um projeto de pesquisa é a determinação do tema, com o qual o pesquisador deve, preferencialmente, ter familiaridade, ser relevante para a comunidade científica e ser passível de publicação, contribuindo, assim, com a popularização da ciência. Após a escolha do tema, deve ser realizada uma revisão bibliográfica, a fim de se identificar o estado da arte sobre o tema escolhido e as lacunas de conhecimento, que guiarão na formulação da pergunta da pesquisa. Dessa forma, a problemática que será objeto de pesquisa será facilmente identificada e passível de ser respondida.

Dependendo do delineamento da pesquisa, o pesquisador deverá levantar a hipótese de resposta, chamada de hipótese alternativa (H1) e a hipótese nula (H0), que indica que não há efeito, mudança ou diferença no desfecho. Para comprovar a H1, deve-se assumir que H0 é verdadeira e submetê-la ao teste estatístico (Teste de Hipóteses), para, então, decidir se a rejeita ou se a aceita. Para testar as hipóteses, deve-se determinar o nível de significância α, ou seja, a probabilidade de se cometer o erro de rejeitar a hipótese nula quando ela é verdadeira. Em pesquisas clínicas, normalmente, utiliza-se o valor de α de 0,05 ou de 0,01.

2 – 2 Metodologia da Pesquisa

O próximo passo é detalhar os materiais e métodos que serão utilizados para responder à pergunta da pesquisa. Para isso, inicialmente deve-se especificar o delineamento metodológico, pois indicará quais normas deverão ser seguidas durante o desenvolvimento da pesquisa. Nos materiais e métodos, devem constar: quais as características da população investigada; como será realizada a seleção da amostra; os instrumentos que serão utilizados para coletar os dados; além de como será feita a análise estatística e quais testes estatísticos serão empregados.

É necessário especificar quem será a população a ser investigada, para, assim, selecionar uma amostra que seja representativa, ou seja, que possua as mesmas características da população investigada.

Definições

- **População:** é conjunto integral, que possui a característica a ser investigada.
- **Amostra:** é um subconjunto da população.

Antes da obtenção da amostra, devem-se definir quais serão os critérios de inclusão e de exclusão dos elementos, a fim de criar uma amostra homogênea. Os critérios de inclusão correspondem às características sociodemográficas e clínicas, enquanto os critérios de exclusão eliminam os elementos que seriam elegíveis ao estudo, caso não possuíssem características que pudessem interferir na qualidade dos dados, na aceitabilidade da aleatorização e na interpretação dos resultados.

O método utilizado para selecionar a amostra (amostragem) é de fundamental importância para a pesquisa científica, pois garante o caráter da representatividade e, portanto, permite a generalização dos resultados da pesquisa para a população estudada. Existem dois tipos de amostragem: as probabilísticas (aleatórias) e as não probabilísticas (não aleatórias). As probabilísticas são as amostragens, nas quais cada elemento da população tem a mesma probabilidade de ser escolhido. São os tipos de amostragens mais almejadas pelos pesquisadores, pois permitem melhor representatividade da população (exemplos: amostra aleatória simples, por conglomerados e estratificada). Em contrapartida, as amostragens não probabilísticas estão restritas aos elementos a que se tem acesso e, logo, não representam toda a população investigada.

É importante que amostragem aleatória não seja confundida com outro conceito fundamental em estatística: o da aleatorização.

Definição

- **Aleatorização:** é quando a seleção do grupo de que o paciente fará parte seja feita aleatoriamente.

O princípio da aleatorização permite que cada elemento da amostra tenha a mesma probabilidade de ser distribuído para os grupos da pesquisa (quando houver), aumentando, assim, a confiabilidade e a generalização dos resultados.

Após a seleção da amostra e a distribuição dos pacientes nos diferentes grupos (quando houver), o próximo passo é a coleta de dados. O instrumento utilizado para coletar os dados deve apresentar propriedades de medida adequadas para a população investigada, ou seja, deve ser válido (medir o que ele se propõem a medir), confiável (livre de erro de medida) e responsivo (detectar mudanças ao longo do tempo).

Os dados coletados devem ser tabulados em uma planilha, a fim de facilitar a análise estatística. As linhas da planilha correspondem aos sujeitos da amostra (unidade de investigação) e as colunas correspondem às variáveis, que são as características dos sujeitos a serem avaliadas.

É possível existirem valores ausentes na planilha de dados, que devem ser tratados com seriedade, a fim de não influenciar na análise estatística. Os dados ausentes podem provir de causas aleatórias, como erro de digitalização ou perda de dados da pesquisa, ou de causas não aleatórias, quando, por exemplo, os dados da planilha são omitidos propositalmente. Os dados ausentes não aleatórios podem comprometer os resultados das análises estatísticas, ao contrário dos dados ausentes aleatórios, que podem ser ignorados, por não influenciarem significativamente a análise dos dados

2 – 3 Análise Exploratória dos Dados (estatística descritiva)

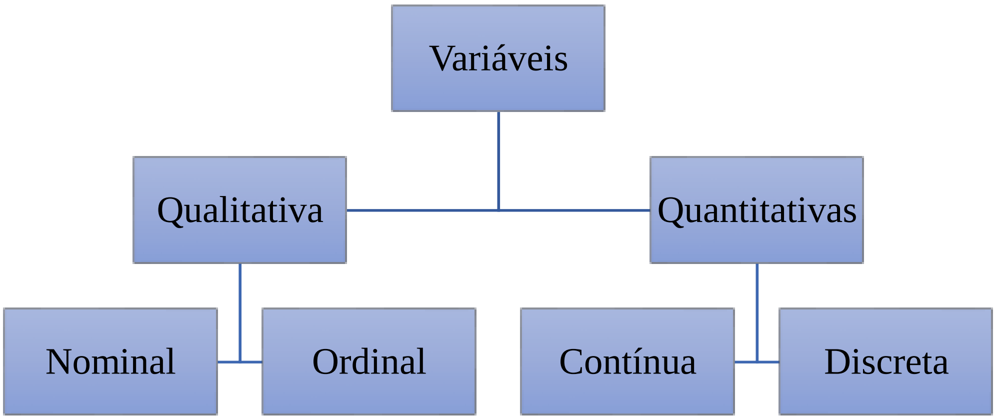
Após a tabulação, os dados estarão prontos para serem descritos, explorados e comparados. Para isso, inicialmente é necessário classificar os tipos de variáveis, pois as classificações influenciarão diretamente na escolha dos testes estatísticos e da representação gráfica dos dados.

As variáveis podem ser classificadas em qualitativas e quantitativas (ou numéricas). As variáveis qualitativas representam qualidades ou atributos dos elementos, como sexo, estado civil e nível de escolaridade. As variáveis quantitativas representam números obtidos a partir de uma mensuração ou contagem, como idade, número de filhos, peso e altura.

As variáveis qualitativas podem ser classificadas em nominais e ordinais. As variáveis serão nominais quando os atributos não apresentarem ordem de importância, como sexo e estado civil, e serão ordinais quando existir ordenação, como nível de escolaridade.

As variáveis quantitativas são divididas em discretas ou contínuas. As variáveis discretas são números finitos resultantes, normalmente, de uma contagem, como número de filhos ou de cáries nos dentes. Já as variáveis contínuas representam infinitas possibilidades numéricas e são obtidas através de mensuração, como peso e nível de glicose no sangue.

Os tipos de variáveis (qualitativas e quantitativas) estão representados a seguir:



Em alguns casos, pode-se atribuir um número à uma variável qualitativa, a fim de viabilizar a análise estatística. Por exemplo, na **tabela 1**, ao sexo feminino foi atribuído o número 1, e ao sexo masculino o número 2. Sugere-se que, ao tabular os dados, já se atribua um número às categorias ou nomes, juntamente com as legendas, conforme demonstrado **na tabela 1**.

Tabela 1: Planilha de tabulação de dados

Número do Paciente	Grupo	Sexo Biológico
1	1	1
2	2	2
3	2	1
4	1	2
5	2	2
6	1	1

Nota: em grupo, o número 1 corresponde ao grupo de intervenção e o número 2 ao grupo controle; já em sexo biológico, 1 indica sexo feminino e 2 o sexo masculino.

As variáveis qualitativas que podem apresentar somente duas possibilidades de resposta, por exemplo, “sim ou não” são chamadas de dicotômicas.

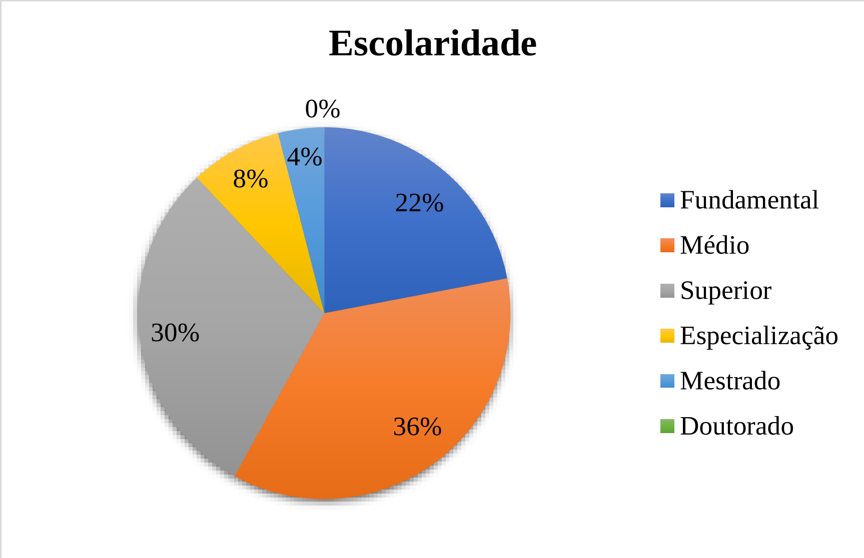
Os dados podem ser organizados, resumidos e representados em tabelas de frequência ou gráficos (por exemplo, de barras, de pizza ou histograma), a fim de se permitir uma mais rápida e fácil compreensão de sua variabilidade, avaliação do tipo de distribuição e de se comparar dois ou mais grupos. Ao método que resume, descreve e organiza os dados de uma amostra de determinada população dá-se o nome de estatística descritiva.

As tabelas de frequência e os gráficos são úteis para avaliar, principalmente, o comportamento das variáveis qualitativas, indicando a ocorrência de cada classe ou categoria da variável. Por exemplo, na **tabela 2** e na **figura 1** estão demonstradas as frequências do nível de escolaridade de uma amostra com 50 participantes.

Tabela 2: Distribuição das frequências absoluta e relativa do nível de escolaridade

Escolaridade	Absoluta	Relativa
Fundamental	11	22%
Médio	18	36%
Superior	15	30%
Especialização	4	8%
Mestrado	2	4%
Doutorado	0	0%

Figura 1: Gráfico de pizza para distribuição de frequência do nível de escolaridade



Para resumir os dados de variáveis quantitativas, além das tabelas de frequências e de gráficos, pode-se utilizar valores que sejam representativos para toda a amostra, por exemplo, as medidas de posição e as medidas de dispersão. As medidas de posição correspondem à média, mediana e moda.

Definições

- **Média:** é a somatória dos valores observados dividido pelo número total da amostra.
- **Mediana:** é o valor central do conjunto de observações, desde que os valores estejam em ordem crescente.
- **Moda:** é o valor mais frequente no conjunto dos valores observados.

A medida de posição mais utilizada é a média, porém seu valor é influenciado pela existência de valores extremos em um conjunto de dados. Nesse caso, é recomendada a utilização da mediana, que é uma medida de posição resistente à presença de valores discrepantes.

A média normalmente vem acompanhada de uma medida de dispersão, por exemplo, do desvio padrão ou da variância, a fim de se indicar a variabilidade dos dados, ou seja, o quão distante cada valor se encontra da média.

Definições

- **Variância:** é a soma dos quadrados da diferença entre cada valor e a média, dividida pela quantidade de elementos da amostra.
- **Desvio Padrão:** é a raiz quadrada positiva da variância.

A avalição do tipo de distribuição dos dados é fundamental para eleição do teste estatístico e pode ser feita através do histograma de frequência.

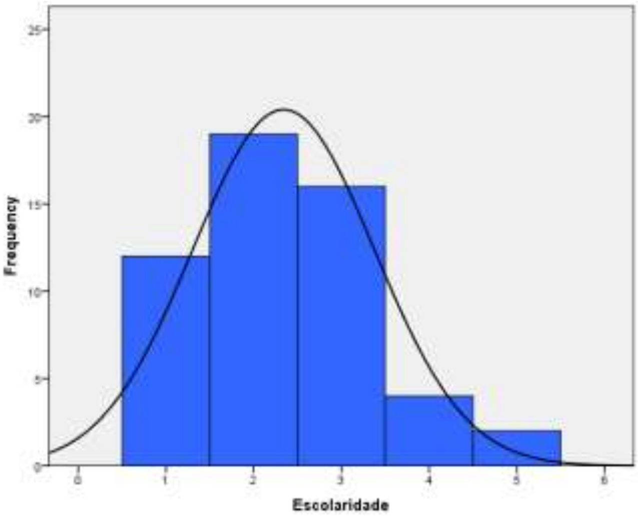
Definição

- **Histograma:** é um gráfico, com barras de mesma largura, posicionadas adjacentes umas às outras. No eixo horizontal estão as classes de valores dos dados quantitativos e no eixo vertical estão as frequências.

Em uma amostra com dados com distribuição normal, o histograma apresenta simetria, ou seja, a metade direita é similar à metade esquerda espelhada, e possui formato de “sino”. Quando a distribuição dos dados da amostra **não** é normal, o histograma será assimétrico e apresentará um lado menor do que o outro, aparentando uma “cauda”.

Na **figura 2,** está representado um histograma com distribuição de dados normal.

Figura 2: Histograma com distribuição normal, do nível de escolaridade de uma amostra.



2 – 4 Análise Confirmatória dos Dados (estatística inferencial)

O próximo passo é a análise estatística dos dados, através de aplicação de testes que permitam responder à pergunta da pesquisa. O nome dado ao método que faz inferências e generalizações sobre os dados de uma amostra representativa de uma população é estatística inferencial.

O uso da tecnologia e dos *softwares* estatísticos facilitou a aplicação dos testes, porém a decisão sobre qual teste estatístico utilizar cabe ao pesquisador e requer o conhecimento de alguns conceitos básicos, que são elucidados neste manual.

2 – 5 Conclusão

Após a aplicação dos testes estatísticos, o pesquisador deve interpretar os resultados e tirar conclusões. As análises terão significância estatística se for muito improvável que os resultados foram obtidos ao acaso. O nível de significância α dos testes é decidido pelo pesquisador e costuma ser de 5% em pesquisas clínicas. Embora algumas pesquisas apresentem resultados estatísticos significantes, pode ser que sua relevância clínica seja irrisória. A significância clínica dos resultados deve ser interpretada pelo pesquisador e fundamentada em bases teóricas e práticas.

3 Escolhendo o teste estatístico

Existem três fatores principais que regem a escolha do teste estatístico. 1) o tipo de distribuição de dados (normal e não normal), 2) a classificação do tipo de dado (qualitativo ou quantitativo) e 3) o tipo de amostras (dependentes ou independentes).

Conforme mencionado anteriormente, a normalidade dos dados amostrais pode ser verificada através do histograma de frequência e da curtose. Em uma amostra com distribuição normal, o histograma se apresentará em forma de "sino", enquanto se apresentará com uma "cauda" em uma distribuição não normal. Outra característica das amostras com distribuição normal é que os valores da média, da mediana e da moda são semelhantes, enquanto se diferem em amostras com distribuição não normal. Por fim, deve-se identificar a presença de valores atípicos (*outliers*), pois podem comprometer os resultados estatísticos. Por isso, no caso da presença de mais de um *outlier*, deve-se considerar que a amostra não possui distribuição normal.

Além da análise da simetria do histograma e da curtose, a simetria dos dados pode ser verificada através do teste de Shapiro-Wilk e do teste de Kolmogorov-Smirnov. O teste de Kolmogorov-Smirnov é utilizado para testar a normalidade em amostras com mais que 51 elementos, enquanto o teste de Shapiro-Wilk é utilizado em amostras com 2 a 51 elementos. Em ambos os testes, a hipótese nula (H0) é que a amostra possua distribuição normal e a hipótese alternativa (H1) é que a amostra não possua distribuição normal. Diferente dos outros testes de hipótese, deseja-se que, nesse caso, o valor de P seja maior que o nível de significância α de 0,05. Assim, a H0 deixará de ser rejeitada, indicando que a amostra possui uma distribuição normal.

As diferenças para a utilização dos testes de Kolmogorov-Smirnov e de Shapiro-Wilk estão demonstradas na **tabela 3.**

Tabela 3: Requisitos para a utilização dos testes de Kolmogorov-Smirnov e Shapiro-Wilk

Teste	Distribuição	Variável	Grupos	Condição
Kolmogorov-Smirnov	Não-paramétrico	Quantitativa	1	Amostras > 51
Shapiro-Wilk	Não-paramétrico	Quantitativa	1	Amostras entre 2 e 51

3 – 1 Testes Paramétricos

Os testes paramétricos são utilizados quando se deseja fazer inferências sobre variáveis qualitativas e que possuam distribuição normal. Caso as amostras não possuam distribuição normal ou se deseja fazer inferências sobre variáveis qualitativas (nominal e ordinal), devem-se utilizar os testes não paramétricos.

A **tabela 4** resume as diferenças no emprego dos testes paramétricos e não paramétricos:

Tabela 4: Requisitos para a utilização dos testes paramétricos e não paramétricos:

Teste	Distribuição	Variável
Paramétricos	Normal	Quantitativa
Não-paramétricos	Não-normal	Quantitativa e qualitativa

Teste T e Teste Z

O Teste T é utilizado para testar uma hipótese sobre uma média populacional. Os requisitos para a utilização do Teste T são: método de seleção amostral aleatória, uma amostra com distribuição normal (paramétrica) e cuja variável a ser analisada é quantitativa. O Teste T é robusto contra um afastamento da normalidade, portanto, deve ser utilizado somente em amostras com distribuição paramétrica.

O Teste Z possui os mesmos parâmetros para a realização do Teste T, porém testa hipóteses a partir do desvio-padrão populacional ou da variância populacional.

Teste T e Teste Z

- **Distribuição**: paramétrica
- **Variável:** quantitativa
- **Número de grupos:** 1

Exemplo 1:

Deseja-se verificar o nível de radiação ionizante (taxa de *kerma*) emitida por 30 equipamentos de cinefluoroscopia, em salas de hemodinâmica hospitalares. O teste de Shapiro-Wilk apresentou valor de P = 0,237 e, portanto, indicou que a amostra possui distribuição normal. Como será testada uma hipótese sobre uma média amostral, deve-se empregar o **Teste T**. Estipulou-se um nível de significância α = 0,05. As hipóteses nulas e alternativas serão as seguintes:

H0: não existe diferença na radiação ionizante das salas de hemodinâmica hospitalares;
H1: existe diferença na radiação ionizante das salas de hemodinâmica hospitalares.

Caso P-valor seja > ou = 0,05: deve-se aceitar H0 e rejeitar H1;
Caso P-valor seja < 0,05: deve-se rejeitar H0 e aceitar H1.

Após empregar o Teste T, observou-se um valor de P = 0,850. Como o valor de P é maior do que o nível de significância α (0,05), deve-se aceitar H0 e rejeitar H1. Portanto, foi observado que não existe diferença estatisticamente significante na radiação ionizante das salas de hemodinâmica hospitalares.

Teste T para 2 amostras, Teste T para 2 amostras com variância diferentes e Teste T pareado

O Teste T e o Teste Z testam hipóteses a partir da média ou da variância provenientes de uma amostra de uma única população. Em contrapartida, os testes T para 2 amostras, T para 2 amostras com variâncias diferentes e T pareado estimam valores de parâmetros populacionais ou testam hipóteses em situações que envolvam duas populações.

Os requisitos para a utilização dos testes T para 2 amostras, T para 2 amostras com variâncias diferentes e T pareado são: amostras provenientes de 2 populações, com distribuição normal de dados e cujas variáveis a serem testadas sejam uma nominal e outra quantitativa. A diferença entre o Teste T pareado e os testes T para 2 amostras e T para 2 amostras com variâncias diferentes é que o Teste T pareado requer 2 amostras dependentes, enquanto os testes T para 2 amostras e T para 2 amostras com variâncias diferentes requerem amostras independentes. Já o Teste T para 2 amostras se difere do Teste T para 2 amostras com variâncias diferentes, por apresentar amostras com variâncias iguais.

Definição

- **Amostras dependentes**: duas amostras são dependentes se os valores são emparelhados, por exemplo, os valores consistem em duas medidas do mesmo sujeito.
- **Amostras independentes**: as amostras são independentes quando os valores de uma amostra não estão relacionados com os valores de outra população.

A **tabela 5** resume as premissas para a utilização do Teste T pareado, Teste T para 2 amostras e Teste T para 2 amostras com variâncias diferentes:

Tabela 5: Requisitos para a utilização do Teste T e do Teste Z:

Teste	Distribuição	Grupos	Dependência	Requisitos
T pareado	Paramétrico	2	Dependentes	Variâncias desconhecidas
T para 2 amostras	Paramétrico	2	Independentes	Variâncias iguais
T para 2 amostras com variâncias diferentes	Paramétrico	2	Independentes	Variâncias diferentes

Exemplo 2:

Deseja-se verificar a eficácia de um novo medicamento para o controle da pressão arterial em uma amostra aleatória simples com 160 indivíduos hipertensos. O grupo experimental fará o uso do novo medicamento, enquanto o grupo controle tomará um medicamento consagrado. O teste de Kolmogorov-Smirnov apresentou valor de P = 0,237 e, portanto, indicou que a amostra possui distribuição normal. Como serão comparadas as médias de duas amostras independentes, deve-se empregar o **Teste T para 2 amostras**. Estipulou-se um nível de significância $\alpha = 0{,}05$. As hipóteses nulas e alternativas serão as seguintes:

H0: não existe diferença entre o uso dos medicamentos novo e consagrado para controlar a pressão arterial de indivíduos hipertensos.
H1: existe diferença entre o uso dos medicamentos novo e consagrado para controlar a pressão arterial de indivíduos hipertensos.

Caso P-valor seja > ou = 0,05: deve-se aceitar H0 e rejeitar H1;
Caso P-valor seja < 0,05: deve-se rejeitar H0 e aceitar H1.

Após empregar o Teste T para 2 amostras, observou-se um valor de P = 0,095. Como o valor de P é maior que o nível de significância α (0,05), deve-se aceitar H0 e rejeitar H1. Portanto, foi observado que não existe diferença estatisticamente significante entre a utilização do medicamento novo e consagrado para controlar a pressão arterial de indivíduos hipertensos.

Exemplo 3:

Deseja-se verificar se o cigarro altera a saturação sanguínea imediatamente após seu consumo. Para isso, será verificada a saturação de 15 sujeitos do sexo masculino, não fumantes, aleatoriamente selecionados, antes e após o consumo de 1 cigarro. Foi utilizado o teste de Shapiro-Wilk para verificar a normalidade dos dados, encontrando um valor de P = 0,566, o que indicou uma distribuição normal. Como serão comparadas as médias de duas amostras pareadas, deve-se empregar o **Teste T pareado**. Estipulou-se um nível de significância α = 0,05. As hipóteses nulas e alternativas serão as seguintes:

H0: não existe diferença na saturação sanguínea de indivíduos do sexo masculino, não fumantes, antes e após o consumo de 1 cigarro;
H1: existe diferença na saturação sanguínea de indivíduos do sexo masculino, não fumantes, antes e após o consumo de 1 cigarro.

Caso P-valor seja > ou = 0,05: deve-se aceitar H0 e rejeitar H1;
Caso P-valor seja < 0,05: deve-se rejeitar H0 e aceitar H1.

Após empregar o Teste pareado, observou-se um valor de P = 0,042. Como o valor de P é menor que o nível de significância (α =0,05), deve-se aceitar H1 e rejeitar H0. Portanto, foi observado que existe diferença na saturação de indivíduos do sexo masculino, não fumantes, antes e após o consumo de 1 cigarro.

ANOVA

Os testes T para duas amostras e T para 2 amostras com variâncias diferentes estimam valores de parâmetros populacionais ou testam hipóteses em situações que envolvam duas populações. Para se testar hipóteses entre três ou mais médias populacionais, utiliza-se o método da análise de variância (ANOVA). Quando se deseja comparar amostras de acordo com apenas uma característica (por exemplo, cor do olho), é utilizada a análise de variância de um fator; por outro lado, quando se deseja comparar duas características amostrais, utiliza-se a análise de variância de dois fatores (por exemplo, cor do olho e cor do cabelo). São requisitos para a utilização do ANOVA: amostras com distribuição paramétrica; amostras aleatórias simples de dados quantitativos; amostras independentes e populações com a mesma variância.

No método ANOVA para um fator, ao utilizar um nível de significância α de 0,05, caso o valor P seja igual ou menor que 0,05, deve-se rejeitar a hipótese nula de que as médias populacionais são iguais e concluir que pelo menos uma das médias populacionais é diferente. Caso o valor do P seja maior do que 0,05, deve-se aceitar a hipótese nula e concluir que não há evidência suficiente de que as médias populacionais não são iguais. No método ANOVA para dois fatores, se o valor de P para interação for igual ou menor do que 0,05, parece haver um efeito de interação, enquanto não haverá efeito de interação se o valor de P for maior que 0,05. Nesse caso, deve-se verificar o efeito dos dois fatores, ou seja, se seus valores são de populações com a mesma média.

Como vimos, o método ANOVA é utilizado para comparar as médias populações entre três ou mais amostras independentes. Em contrapartida, caso se deseje comparar dados provenientes de amostras dependentes, deve-se utilizar o método ANOVA para medidas repetidas. Portanto, é requisito para o método ANOVA para medidas repetidas: distribuição paramétrica de três ou mais amostras dependentes.

A **tabela 6** resume os requisitos para a utilização dos métodos ANOVA e ANOVA para medidas repetidas:

Tabela 6: Requisitos para a utilização dos métodos ANOVA e ANOVA para medidas repetidas:

Teste	Distribuição	Grupos	Dependência
ANOVA	Paramétrico	3 ou mais	Independentes
ANOVA para medidas repetidas	Paramétrico	3 ou mais	Dependentes

Exemplo 4:

Deseja-se verificar a eficácia de um novo medicamento para o controle da pressão arterial em uma amostra aleatória simples com 320 indivíduos hipertensos. Diferente do exemplo 1, agora serão utilizadas 3 amostras: um grupo experimental, que fará o uso do novo medicamento; um grupo controle, que fará o uso do medicamento consagrado e um grupo placebo, que tomará um comprimido sem princípios ativos. O teste de Kolmogorov-Smirnov apresentou valor de P = 0,456 e, portanto, indicou que a amostra possui distribuição normal. Como serão comparadas as médias entre 3 amostras independentes, deve-se empregar o teste **ANOVA.** Estipulou-se um nível de significância $\alpha = 0,05$. As hipóteses nulas e alternativas serão as seguintes:

H0: não existe diferença quanto ao uso dos diferentes medicamentos para controlar a pressão arterial em indivíduos hipertensos.
H1: existe diferença quanto ao uso dos diferentes medicamentos para controlar a pressão arterial em indivíduos hipertensos.

Caso P-valor seja > ou = 0,05: deve-se aceitar H0 e rejeitar H1;
Caso P-valor seja < 0,05: deve-se rejeitar H0 e aceitar H1.

Após empregar o teste ANOVA, observou-se um valor de P = 0,0015. Como o valor de P é menor que o nível de significância (α =0,05), deve-se aceitar H1 e rejeitar H0. Portanto, foi observado que existe diferença estatisticamente significante quanto à utilização dos diferentes medicamentos para controlar a pressão arterial de indivíduos hipertensos.

Exemplo 5:

Deseja-se verificar se um grupo de 10 corredores de rua, em uma prova de 5 Km, leva o mesmo tempo para percorrer cada quilômetro da prova. Foi utilizado o teste de Shapiro-Wilk para verificar a normalidade dos dados, encontrando um valor de $P = 0{,}152$, o que indicou uma distribuição normal. Como serão comparadas as médias de mais de 3 amostras pareadas (5 Km), deve-se utilizar o teste **ANOVA para medidas repetidas.** Estipulou-se um nível de significância $\alpha = 0{,}05$. As hipóteses nulas e alternativas serão as seguintes:

H0: não existe diferença no tempo que os corredores de rua percorrem 1 Km, em uma prova de 5 Km;
H1: existe diferença no tempo que os corredores de rua percorrem 1 Km, em uma prova de 5 Km.

Caso P-valor seja > ou = 0,05: deve-se aceitar H0 e rejeitar H1;
Caso P-valor seja < 0,05: deve-se rejeitar H0 e aceitar H1.

Após empregar o teste ANOVA para medidas repetidas, observou-se um valor de $P = 0{,}0034$. Como o valor de P é menor que o nível de significância ($\alpha = 0{,}05$), deve-se aceitar H1 e rejeitar H0. Portanto, existe diferença estatisticamente significante no tempo em que os corredores de rua percorrem 1 Km, em uma prova de 5 Km.

3 – 2 Testes Não-paramétricos

Os testes descritos anteriormente (paramétricos) exigem que as amostras possuam distribuição de dados normal. Caso as amostras não possuam distribuição normal, devem-se utilizar os testes não paramétricos. Além disso, os testes não paramétricos também são utilizados quando se deseja fazer inferências sobre dados qualitativos (nominais e ordinais). Embora possuam menos requisitos e possam ser utilizados em mais situações, os testes não-paramétricos não são tão eficientes quanto os testes paramétricos, a não ser que essa redução da eficiência seja compensada por um tamanho de amostra maior.

Teste do Sinal

O Teste dos Sinais é um teste não-paramétrico, que converte os valores dos dados em sinais de "mais" ou de "menos", para testar se algum dos sinais ocorre com mais frequência. O Teste dos Sinais é utilizado para testar hipóteses que envolvam dados emparelhados, dados nominais com duas categorias ou hipóteses sobre a mediana de uma única população. O fundamento por trás do Teste dos Sinais é que se dois conjuntos de dados possuem medianas iguais, então o número de sinais positivos é aproximadamente igual ao número de sinais negativos.

Teste de postos com sinais de Wilcoxon

O teste de postos com sinais de Wilcoxon é um teste não paramétrico, que converte os dados amostrais em postos (ordens) em duas situações: 1) para testar a hipótese de que os dados emparelhados de uma população possuem diferenças com mediana igual a zero **(Teste de Wilcoxon pareado)** e 2) ou para testar a hipótese de que uma única população tenha algum valor alegado para a mediana **(Teste de Wilcoxon para 1 amostra)**.

Exemplo 6:

Vamos utilizar o exemplo 3, em que se deseja verificar se o cigarro altera a saturação sanguínea imediatamente após seu consumo, em uma amostra de 15 sujeitos do sexo masculino e não fumantes. Porém, agora o Teste de Shapiro-Wilk encontrou um valor de P = 0,023, o que indicou uma distribuição não normal. Como serão comparadas as médias de duas amostras pareadas, deve-se empregar o **Teste de Wilcoxon pareado**. Estipulou-se um nível de significância α = 0,05. As hipóteses nulas e alternativas serão as seguintes:

H0: não existe diferença na saturação sanguínea de indivíduos do sexo masculino, não fumantes, antes e após o consumo de 1 cigarro;
H1: existe diferença na saturação sanguínea de indivíduos do sexo masculino, não fumantes, antes e após o consumo de 1 cigarro.

Caso P-valor seja > ou = 0,05: deve-se aceitar H0 e rejeitar H1;
Caso P-valor seja < 0,05: deve-se rejeitar H0 e aceitar H1.

Após empregar o Teste pareado, observou-se um valor de P = 0,042. Como o valor de P é menor que o nível de significância (α =0,05), deve-se aceitar H1 e rejeitar H0. Portanto, foi observado que existe diferença na saturação de indivíduos do sexo masculino, não fumantes, antes e após o consumo de 1 cigarro.

Teste da Soma de Postos de Wilcoxon para duas amostras independentes

O Teste da Soma de Postos de Wilcoxon para duas amostras independentes é o equivalente ao teste U de Mann-Whitney e utiliza os postos de dados amostrais de duas populações independentes para testar se elas provêm de populações com medianas iguais. No Teste da Soma de Postos de Wilcoxon para duas amostras independentes, a hipótese nula (H0) é que as duas amostras provêm de populações com medianas iguais, enquanto a hipótese alternativa (H1) é que a mediana da primeira população é diferente (ou maior do que, ou menor do que) da mediana da segunda população.

Os testes de postos com sinais de Wilcoxon e da Soma de Postos de Wilcoxon para duas amostras independentes, além permitirem amostras com distribuição não normal (não-paramétrica), podem ser utilizados tanto com variáveis quantitativas, quanto com qualitativas ordinais.

A **tabela 7** resume os requisitos para a utilização dos testes de postos com sinais de Wilcoxon e da Soma de Postos de Wilcoxon para duas amostras independentes:

Tabela 7: Requisitos para a utilização dos testes de postos com sinais de Wilcoxon e da Soma de Postos de Wilcoxon para duas amostras independentes:

Teste	Distribuição	Grupos	Dependência
Teste de postos com sinais de Wilcoxon	Não-paramétrico	1 ou 2 amostras pareadas	Dependentes
Teste da Soma de Postos de Wilcoxon para duas amostras independentes ou Teste U de Mann-Whitney	Não-paramétrico	2	Independentes

Exemplo 7:

Vamos utilizar o exemplo 2, em que se deseja verificar a eficácia de um novo medicamento para o controle da pressão arterial em uma amostra aleatória simples com 160 indivíduos hipertensos. Porém, agora o teste de Kolmogorov-Smirnov apresentou valor de P = 0,015, indicando que a distribuição não é normal. Como serão comparadas as medianas de duas amostras independentes, deve-se empregar o **Teste de Mann-Whitney** (ou o Teste da Soma de Postos de Wilcoxon para duas amostras independentes). Estipulou-se um nível de significância $\alpha = 0,05$. As hipóteses nulas e alternativas serão as seguintes:

H0: não existe diferença entre o uso dos medicamentos novo e consagrado para controlar a pressão arterial de indivíduos hipertensos.
H1: existe diferença entre o uso dos medicamentos novo e consagrado para controlar a pressão arterial de indivíduos hipertensos.

Caso P-valor seja > ou = 0,05: deve-se aceitar H0 e rejeitar H1;
Caso P-valor seja < 0,05: deve-se rejeitar H0 e aceitar H1.

Após empregar o Teste de Mann-Whitney, observou-se um valor de P = 0,095. Como o valor de P é maior que o nível de significância (α =0,05), deve-se aceitar H0 e rejeitar H1. Portanto, foi observado que não existe diferença estatisticamente significante entre a utilização do medicamento novo e consagrado para controlar a pressão arterial de indivíduos hipertensos.

Teste de Kruskal-Wallis

O Teste de Kruskal-Wallis é um teste não-paramétrico que utiliza postos de dados de três ou mais amostras independentes para testar a hipótese nula (H0) de que as amostras provêm de populações com medianas iguais. Nesse caso, a hipótese alternativa (H1) é que as populações tenham medianas diferentes.

O Teste de Kruskal-Wallis testa a igualdade das **medianas** de três ou mais populações e, por isso, não exige que as populações tenham distribuição normal. Caso as amostras provenham de populações com distribuição normal, sugere-se a utilização do método ANOVA, que compara as **médias** de três ou mais populações e, portanto, é mais robusto.

A **tabela 8** compara os requisitos para a utilização do Teste de Kruskal-Wallis e do teste ANOVA:

Tabela 8: Requisitos para a utilização do Teste de Kruskal-Wallis e do teste ANOVA

Teste	Distribuição	Grupos	Dependência
Teste de Kruskal-Wallis	Paramétrica	3 ou mais	Independentes
ANOVA	Não-paramétrica	3 ou mais	Independentes

Exemplo 8:

Vamos utilizar o exemplo 4, em que se deseja verificar a eficácia de um novo medicamento para o controle da pressão arterial em uma amostra aleatória simples com 320 indivíduos hipertensos. Os pacientes serão distribuídos em 3 grupos: um grupo experimental, que fará o uso do novo medicamento; um grupo controle, que fará o uso do medicamento consagrado e um grupo placebo, que tomará um comprimido sem princípios ativos. Diferente do exemplo 4, agora o teste de Kolmogorov-Smirnov apresentou valor de $P = 0{,}023$, indicando que a amostra não possui distribuição normal. Como serão comparadas as medianas de 3 amostras independentes, deve-se empregar o **Teste de Kruskal-Wallis.** Estipulou-se um nível de significância $\alpha = 0{,}05$. As hipóteses nulas e alternativas serão as seguintes:

H0: não existe diferença quanto ao uso dos diferentes medicamentos para controlar a pressão arterial em indivíduos hipertensos.
H1: existe diferença quanto o uso dos diferentes medicamentos para controlar a pressão arterial em indivíduos hipertensos.

Caso P-valor seja > ou = 0,05: deve-se aceitar H0 e rejeitar H1;
Caso P-valor seja < 0,05: deve-se rejeitar H0 e aceitar H1.

Após empregar o Teste de Kruskal-Wallis, observou-se um valor de $P = 0{,}0015$. Como o valor de P é menor que o nível de significância ($\alpha = 0{,}05$), deve-se aceitar H1 e rejeitar H0. Portanto, foi observado que existe diferença estatisticamente significante quanto à utilização dos diferentes medicamentos para controlar a pressão arterial de indivíduos hipertensos.

Teste de Friedman

O Teste de Friedman é um teste não-paramétrico utilizado para comparar 3 ou mais amostras dependentes, em que se utilizam variáveis qualitativas ordinais ou variáveis quantitativas com distribuição não normal. No Teste de Friedman, é utilizada a ordem ocupada pelos dados (rank), depois de serem organizados de forma crescente.

Para as situações em que sejam utilizadas amostras independentes, então, deve-se utilizar o Teste de Kruskal-Wallis. Por outro lado, caso se deseje testar hipótese entre duas ou mais amostras emparelhadas com distribuição normal, deve-se utilizar o teste ANOVA para medidas repetidas.

As diferenças entre as aplicações do Teste de Friedman, do Teste de Kruskal-Wallis e do ANOVA para medidas repetidas estão descritas **na tabela 9.**

Tabela 9: Requisitos para a aplicação dos testes de Friedman, Kruskal-Wallis e ANOVA para medidas repetidas

Teste	Distribuição	Grupos	Dependência
Teste de Friedman	Não-paramétrica	3 ou mais	Dependentes
Teste de Kruskal-Wallis	Não-paramétrica	3 ou mais	Independentes
ANOVA para medidas repetidas	Paramétrica	3 ou mais	Dependentes

Exemplo 9:

Deseja-se verificar a eficácia de um tratamento fisioterapêutico para dor lombar, em três pontos no tempo. Para isso, os pacientes foram questionados se estavam piores, iguais ou melhores no final do tratamento, após 4 semanas e após 6 meses. Como será avaliada uma variável qualitativa ordinal (pior, igual ou melhor), em três pontos no tempo (pareada), deve-se utilizar o **Teste de Friedman**. Estipulou-se um nível de significância $\alpha = 0,05$. As hipóteses nulas e alternativas serão as seguintes:

H0: não existe diferença nas respostas dos pacientes nos diferentes pontos do tempo;
H1: existe diferença nas respostas dos pacientes nos diferentes pontos do tempo.

Caso P-valor seja > ou = 0,05: deve-se aceitar H0 e rejeitar H1;
Caso P-valor seja < 0,05: deve-se rejeitar H0 e aceitar H1.

Após empregar o Teste de Friedman, observou-se um valor de P = 0,0015. Como o valor de P é menor que o nível de significância (α =0,05), deve-se aceitar H1 e rejeitar H0. Portanto, foi observado que existe diferença estatisticamente significante entre as opções de respostas dos pacientes nos diferentes pontos do tempo, após a realização do tratamento para dor lombar.

3 – 3 Testes de Correlação e Regressão

Os testes de correlação visam verificar a existência ou não de uma correlação entre duas variáveis e se essa correlação é linear. Pode-se afirmar que existe uma correlação quando os valores de uma variável se correlacionam com os de outra variável. Além disso, pode-se afirmar que existe uma correlação linear quando há uma correlação e os pontos referentes aos dados emparelhados se aproximam da forma de uma reta.

Teste de Correlação de Pearson

Para se verificar a existência de uma correlação linear entre duas variáveis quantitativas, provenientes de amostras com distribuição normal, utiliza-se o coeficiente de correlação linear de Pearson. O coeficiente de correlação linear r mede a força de correlação linear entre os valores emparelhados de dados amostrais. O valor de r está entre -1 (negativa) e 1 (positiva) e é sensível à valores atípicos, que devem ser, pois, retirados. A interpretação do coeficiente de correlação r varia de acordo com o autor, porém, normalmente, é considerado ideal valor para o coeficiente de correlação de Pearson acima de 0,7. Caso o valor de r seja igual a 0, então, deve-se concluir que não existe correlação entre as variáveis.

Antes de se interpretar o coeficiente de correlação, deve-se observar o resultado do teste de hipóteses (Teste T), para verificar se a existência de uma correlação entre duas variáveis é estatisticamente significativa. No teste de hipótese para correlação de Pearson, a hipótese nula (H0) é que não há correlação linear e a hipótese alternativa (H1) é que há correlação linear. Caso o valor de P seja maior do que o valor de alfa, com um nível de significância igual a 0,05, então, deve-se aceitar H0 e concluir que não há evidência suficiente de existência de uma correlação linear entres as variáveis.

Vale ressaltar que a existência de correlação não implica em causalidade, ou seja, a existência de uma correlação entre as variáveis não sugere que uma cause ou seja consequência da outra.

Exemplo 10:

Deseja-se verificar se o tempo gasto para 50 corredores de rua percorrer 1 Km (*pace*) é alterado em função da frequência semanal de treinos. Foi realizado o teste de Kolmogorov-Smirnov para verificar a normalidade dos dados, encontrando um valor de P = 0,534, o que indicou uma distribuição normal. Como se deseja verificar se existe correlação entre o *pace* e a frequência semanal de treinos, em uma amostra com distribuição normal, deve-se utilizar o **Teste de Correlação de Pearson**. Estipulou-se um nível de significância $\alpha = 0,05$. As hipóteses nulas e alternativas serão as seguintes:

H0: não existe uma correlação linear entre o *pace* e a frequência de treino semanal;
H1: existe uma correlação linear entre o *pace* e a frequência de treino semanal tempo.

Caso P-valor seja > ou = 0,05: deve-se aceitar H0 e rejeitar H1;
Caso P-valor seja < 0,05: deve-se rejeitar H0 e aceitar H1.

O Teste de Correlação de Pearson apresentou coeficiente de correlação de Pearson r = 0,75 e P valor = 0,075. Como o valor de r é próximo de 1 e o valor de P é menor do que o nível de significância (0,05), então, pode-se concluir que existe uma forte correlação linear positiva entre o *pace* e a frequência de treinos semanal.

Teste de Correlação de Spearman

O Teste de Correlação de postos de Spearman é um teste não paramétrico que utiliza postos de dados amostrais (pares combinados) para verificar a associação entre duas variáveis. Diferente do Teste de Correlação de Pearson, que utiliza dados amostrais em pares, o Teste de Correlação de Spearman utiliza postos de dados em pares para calcular o coeficiente de correlação r. Outra diferença fundamental é que o Teste de Correlação de Pearson requer dados provenientes de amostras com distribuição normal, enquanto o Teste de Correlação de Spearman pode ser utilizado em amostras com distribuição não-paramétrica. Por fim, o Teste de Correlação de Pearson correlaciona dados quantitativos, enquanto o Teste de Correlação de Spearman é utilizado tanto com variáveis quantitativas, quanto com qualitativas ordinais.

As diferenças entre os requisitos para a utilização do Teste de Correlação de Pearson e do Teste de Correlação de Spearman estão demonstradas na **tabela 10**.

Tabela 10: Requisitos para a utilização do Teste de Correlação de Pearson e do Teste de Correlação de Spearman

Teste	Distribuição	Variáveis
Correlação de Pearson	Paramétrica	Quantitativa
Correlação de Spearman	Não-paramétrica	Quantitativa ou qualitativa ordinal

Para calcular o Teste de Correlação de Spearman r e utilizá-lo para verificar a associação entre duas variáveis, deve-se levantar a hipótese nula (H0) e a hipótese alternativa (H1). A H0 é que não existe correlação entre as duas variáveis, enquanto a H1 é que existe correlação. É importante ressaltar que os valores de P da correlação linear não se aplicam para métodos de correlação de postos. Caso o valor da estatística do teste seja maior do que o nível de significância estipulado (por exemplo, de 0,05), deve-se aceitar a hipótese nula e concluir que não há evidência suficiente de existência de uma correlação entres os postos.

Exemplo 11:

Deseja-se verificar se existe uma correlação entre a função social e a presença de dor (com ou sem dor) em uma amostra com 35 pacientes com fibromialgia. Como se deseja verificar se existe correlação entre variáveis quantitativas, deve-se utilizar o **Teste de Correlação de Spearman**. Estipulou-se um nível de significância $\alpha = 0,05$. As hipóteses nulas e alternativas serão as seguintes:

H0: não existe correlação entre a função social e a presença de dor;
H1: existe correlação entre a função social e a presença da dor.

Caso P-valor seja > ou = 0,05: deve-se aceitar H0 e rejeitar H1;
Caso P-valor seja < 0,05: deve-se rejeitar H0 e aceitar H1.

O Teste de Correlação de Spearman apresentou coeficiente de correlação r = -0,55 e P valor = 0,095. Portanto, pode-se concluir que existe uma correlação moderada negativa entre a função social e a presença de dor em pacientes com fibromialgia.

Teste de Regressão Linear Simples

O Teste de Regressão Linear Simples utiliza a equação de regressão para fazer predições para um valor de uma variável, a partir de um valor conhecido de outra variável. A equação de regressão expressa uma relação entre uma variável preditora (ou variável independente) e uma variável resposta (ou variável (dependente). É condição para aplicação do Teste de Regressão Linear Simples: amostra de dados quantitativos emparelhados; distribuição normal dos dados; existência de uma correlação linear entre duas variáveis e ausência de *outliers*, que podem influenciar o resultado do teste.

Pode-se utilizar o Coeficiente de Determinação r^2 para se encontrar a proporção da variação em que uma variável pode ser explicada pela relação linear entre as variáveis, ou seja, a reta de regressão. O Coeficiente de Determinação r^2 é calculado elevando o coeficiente de correlação linear ao quadrado. O valor de r^2 indica em que proporção (porcentagem) a variação total na variável dependente pode ser explicada pela variável independente.

Regressão Múltipla

Diferente da Regressão Linear Simples, que se aplica a uma correlação linear entre duas variáveis, a Regressão Múltipla faz predições que envolvem mais de duas variáveis. Para isso, utiliza-se a equação de regressão múltipla, que expressa uma relação linear entre uma variável resposta (ou dependente) e duas ou mais variáveis preditoras (ou independentes). Portanto, o objetivo da Regressão Múltipla é utilizar a equação de regressão múltipla para fazer predições para as variáveis dependentes, a partir de dados amostrais provenientes de três ou mais variáveis.

Pode-se utilizar o Coeficiente de Determinação Múltipla R^2 para se estimar o ajuste da equação de regressão múltipla aos dados amostrais. Quanto menor o valor de R^2, pior o ajuste da equação de regressão múltipla aos dados, enquanto valores de R^2 altos indicam bons ajustes da equação.

3 – 4 Análises de distribuições de frequências

Os testes a seguir são análises estatísticas de dados categóricos ou qualitativos nominais, através da avaliação das tabelas de contingências (tabelas de frequência).

3 – 4 – 1 Testes Qui-quadrado

Os Testes Qui-quadrado (X^2) avaliam as distribuições de frequências de dados categóricos, a fim de verificar se as distribuições de frequências observadas diferem de uma distribuição alegada, se as variáveis linha e coluna são independentes ou se uma variável ocorre com as mesmas proporções entre duas ou mais populações.

Teste da Qualidade do Ajuste do Qui-quadrado

O Teste da Qualidade do Ajuste avalia se uma distribuição de frequência concorda com alguma distribuição específica. Neste teste, os dados amostrais são contagens de frequências, arranjadas em uma única linha ou coluna (chamada de tabela de frequência de um fator). A frequência esperada para cada categoria é a frequência que ocorreria se os dados tivessem realmente a distribuição alegada. Estima-se que a **frequência esperada** seja de pelo menos 5, enquanto não há requisitos para as **frequências observadas**. Além disso, o Teste de Qualidade do Ajuste pode ser utilizado em populações com distribuição não normal.

No Teste de Qualidade de Ajuste, inicialmente, devem-se identificar as frequências observadas para, assim, determinar as frequências esperadas com o uso da distribuição alegada. A hipótese nula (H0) é que as contagens da frequência concordem com a distribuição alegada, enquanto a hipótese alternativa (H1) é que as contagens de frequência não concordem com a distribuição. Caso a estatística do teste X^2 seja menor do que o valor de P, deve-se aceitar H0 e concluir que há um bom ajuste à distribuição alegada. Caso a estatística do teste X^2 seja maior que o valor de P, então, deve-se rejeitar H0 e concluir que não há um bom ajuste à distribuição alegada.

Teste de Independência do Qui-quadrado

O Teste de Independência é um método para analisar tabelas de contingência, que incluem distribuição de frequência para dados categóricos, dispostos com pelo menos duas linhas e duas colunas. Em um teste de independência, testa-se a hipótese nula (H0) de que em uma tabela de contingência, as variáveis linha e coluna são independentes. Portanto, o objetivo do teste de Independência é testar a hipótese da independência entre as variáveis linha e coluna de uma tabela de contingência. Como requisito, a **frequência esperada** é de no mínimo 5, enquanto não há valor mínimo para a **frequência observada**. Além disso, não há a exigência de que a população tenha uma distribuição normal.

No teste da qualidade do ajuste, a estatística de teste permite medir o grau de discordância entre as frequências observadas e aquelas que seriam teoricamente esperadas quando as duas variáveis são independentes. A hipótese nula (H0) é que as variáveis linha e coluna são independentes, enquanto a hipótese alternativa (H1) é que as variáveis linha e coluna são dependentes. Caso o valor P seja menor do que o nível de significância estipulado (por exemplo, de 0,05), deve-se rejeitar a H0, de independência entre as variáveis; caso o valor de P seja maior do que o nível de significância, deve-se aceitar a H0 e concluir que as variáveis são independentes.

Teste Qui-quadrado de homogeneidade

O Teste Qui-quadrado de homogeneidade visa determinar se duas ou mais populações possuem as mesmas proporções de alguma característica. Através da análise de tabelas de contingência, com distribuição de frequência de dados categóricos, o Teste Qui-quadrado de homogeneidade analisa se as proporções entre diferentes populações são as mesmas, em relação à determinada característica. Como requisito, toda célula deve ter uma frequência esperada de pelo menos 5.

No teste Qui-quadrado de homogeneidade, testa-se a hipótese nula (H0) de que populações diferentes têm as mesmas proporções de alguma característica. Caso o valor P seja menor do que o nível de significância estipulado (por exemplo, de 0,05), deve-se rejeitar a H0 e concluir que as populações diferentes **não** têm as mesmas proporções de alguma característica; caso o valor de P seja maior do que o nível de significância, deve-se aceitar a H0 e concluir que as populações diferentes têm as mesmas proporções de alguma característica.

Exemplo 12:

Deseja-se verificar se, em uma amostra com 100 pacientes com dores musculoesqueléticas crônicas, existe associação entre o uso de medicamentos analgésicos (faz uso ou não faz uso) e o sexo biológico (feminino ou masculino). Como se deseja verificar se existe associação entre duas variáveis quantitativas nominais, deve-se utilizar o **Teste Qui-quadrado de Homogeneidade**. Estipulou-se um nível de significância $\alpha = 0{,}05$. As hipóteses nulas e alternativas serão as seguintes:

H0: não existe associação entre o uso de medicamentos analgésicos e o sexo biológico;
H1: existe associação entre o uso de medicamentos analgésicos e o sexo biológico.

Caso P-valor seja > ou = 0,05: deve-se aceitar H0 e rejeitar H1;
Caso P-valor seja < 0,05: deve-se rejeitar H0 e aceitar H1.

O Teste Qui-quadrado de Homogeneidade apresentou valor de P = 0,085. Como o valor de P é maior do que o nível de significância (0,05), deve-se aceitar H0 e rejeitar H1. Portanto, pode-se concluir que não existe associação estatisticamente significante entre o uso de medicamentos analgésicos e o sexo biológico em pacientes com dores musculoesqueléticas crônicas.

3 – 4 – 2 Teste de Fisher

O teste exato de Fisher é um método para análise de tabelas de contingência que incluem distribuição de frequência para dados categóricos, dispostos em duas linhas e duas colunas. No teste exato de Fisher, pelo menos uma frequência esperada deve apresentar valor menor que 5, o que o diferencia do Teste Qui-quadrado de homogeneidade. Os valores de P no teste de Fischer são exatos e, logo, não exigem uma técnica de aproximação.

Exemplo 13:

Deseja-se verificar se existe associação entre o uso de fio dental (faz uso e não faz uso) e o mau hálito (tem mau hálito e não tem mau hálito) em uma amostra com 4 idosos em um centro de acolhimento comunitário. Como se deseja verificar se existe associação entre duas variáveis quantitativas nominais, em uma amostra menor que 5, deve-se utilizar o **Teste Exato de Fischer.** Estipulou-se um nível de significância $\alpha = 0{,}05$. As hipóteses nulas e alternativas serão as seguintes:

H0: não existe associação entre o uso de fio dental e o mau hálito;
H1: existe associação entre o uso de fio dental e o mau hálito.

Caso P-valor seja > ou = 0,05: deve-se aceitar H0 e rejeitar H1;
Caso P-valor seja < 0,05: deve-se rejeitar H0 e aceitar H1.

O Teste exato de Fisher apresentou valor de P = 0,0015. Como o valor de P é menor do que o nível de significância (0,05), deve-se rejeitar H0 e aceitar H1. Portanto, pode-se concluir que existe uma associação estatisticamente significante entre o uso de fio dental e o mau hálito, em uma amostra de idosos em um centro de acolhimento comunitário.

3 – 4 – 3 Teste de McNemar para Dados Emparelhados

Ao contrário do Teste exato de Fisher, o Teste de Mc Nemar é utilizado para as contagens de frequência de dados emparelhados. Nesse caso, as contagens de frequência dentro de cada par não são independentes. No Teste de Mc Nemar, a hipótese nula (H0) é que as frequências das categorias diferentes ocorrem na mesma proporção, enquanto a hipótese alternativa (H1) é que as frequências das categorias diferentes não ocorrem proporcionalmente. Caso o valor de P seja maior do que o nível de significância estabelecido (por exemplo, de 0,05), deve-se aceitar H0 e concluir que as frequências das categorias diferentes ocorrem na mesma proporção.

Os requisitos para a utilização do teste Qui-quadrado, do Teste exato de Fisher e do Teste de McNemar estão descritos na **tabela 11**.

Tabela 11: Requisitos para a utilização do teste Qui-quadrado, do Teste exato de Fisher e do Teste de McNemar

Teste	Distribuição	Variáveis	Grupos	Dependência	Requisitos
Teste Qui-quadrado	Não-paramétrica	Nominais	2 ou mais	Independentes	f > 5
Teste exato de Fisher	Não-paramétrica	Nominais	2	Independentes	f < 5
Teste de McNemar	Não-paramétrica	Nominais	2 ou mais	Dependentes	f > 5

Exemplo 14:

Deseja-se verificar se existe diferença entre a proporção de pacientes satisfeitos e não satisfeitos com uma sessão de terapia em grupo, em uma amostra de pacientes com Transtorno da Ansiedade Generalizada. Para isso, no final da sessão foi perguntado aos pacientes se eles estavam satisfeitos e foram oferecidas duas opções de resposta: "sim" e "não". Como se deseja testar se há diferença na frequência entre dois dados categóricos ("sim" e "não"), deve-se utilizar o **Teste de McNemar.** Estipulou-se um nível de significância $\alpha = 0{,}05$. As hipóteses nula e alternativa serão as seguintes:

H0: não existe diferença entre o número de pacientes satisfeitos e não satisfeitos com uma sessão de terapia em grupo.
H1: existe diferença entre o número de pacientes satisfeitos e não satisfeitos com uma sessão de terapia em grupo.

Caso P-valor seja > ou = 0,05: deve-se aceitar H0 e rejeitar H1;
Caso P-valor seja < 0,05: deve-se rejeitar H0 e aceitar H1.

Após empregar o **Teste de McNemar**, observou-se um valor de P = 0,550. Como o valor de P é maior que o nível de significância (α =0,05), deve-se aceitar H0 e rejeitar H1. Portanto, não existe diferença estatisticamente significante entre o número de pacientes satisfeitos e não satisfeitos com uma sessão de terapia em grupo.

4 Cálculo Amostral

A estimação do tamanho da amostra é etapa fundamental durante a elaboração do projeto de pesquisa, pois um tamanho de amostra adequado garante a confiabilidade dos resultados. Através do cálculo amostral, pode-se estimar qual o tamanho da amostra necessária para que os resultados da pesquisa sejam reprodutíveis e, assim, poder generalizar os resultados provenientes de uma amostra para toda a população.

São fundamentais tamanhos amostrais apropriados para garantir a confiabilidade e a reprodutibilidade dos resultados da pesquisa. Quanto maior o tamanho da amostra, mais próximos a média e o desvio padrão estarão da população. Por outro lado, quanto menor o tamanho da amostra maior será a variabilidade dos resultados e mais longe a média e o desvio padrão estarão da população.

A utilização de uma amostra com tamanho adequado favorece a economia financeira e de tempo, pois permite que os resultados provenientes de apenas uma amostra sejam generalizados para toda a população. Pesquisas realizadas com amostras maiores do que as necessárias, além de não alterarem os resultados, ainda podem acarretar desperdício de tempo e de recursos financeiros.

Para fazer o cálculo amostral, é necessário estabelecer a pergunta e o delineamento da pesquisa, para que, assim, sejam determinados o desfecho principal (variável dependente), qual população será avaliada, quais testes estatísticos serão empregados e quais serão os parâmetros para a análise estatística (nível de significância, poder estatístico e tamanho do efeito).

O cálculo amostral é realizado em 4 etapas: 1) identificação da variável dependente; 2) seleção do teste estatístico; 3) seleção do nível de significância, do poder estatístico e do tamanho do efeito; e 4) análise da sensibilidade.

A primeira etapa para a determinação do tamanho da amostra é a identificação das variáveis, que conforme visto, podem ser classificadas em qualitativas (nominal ou ordinal) e quantitativas (contínua ou discreta).

A segunda etapa é a escolha do teste estatístico, que depende do tipo de distribuição dos dados (normal ou não normal), do tipo de variável (quantitativa ou qualitativa) e da relação de dependência entre as amostras (dependentes ou independentes).

A terceira etapa consiste na determinação do nível de significância, do poder estatístico e do tamanho do efeito. O nível de significância comumente utilizado em pesquisas clínicas é de 0,05 ou de 0,01; já o poder estatístico normalmente empregado é de 10 a 20%.

O nível de significância α indica a probabilidade de se cometer o erro de rejeitar a hipótese nula quando ela é verdadeira, ou seja, de se cometer o erro tipo 1 (falso positivo). O tamanho da amostra influencia diretamente na probabilidade de se cometer o erro tipo 1: tamanhos amostrais maiores reduzem a probabilidade de se cometer o erro tipo 1, enquanto amostras menores aumentam a probabilidade de se cometer o erro do tipo 1. Portanto, quanto menor o nível de significância α, menor é a probabilidade de se cometer o erro tipo 1 e, consequentemente, será necessária uma amostra de tamanho maior. Por outro lado, quanto maior o nível de significância α, maior a probabilidade de se cometer o erro tipo 1 e, assim, será necessária uma amostra de tamanho menor.

A utilização de amostra com tamanho maior do que o necessário também prejudica a análise estatística, pois aumenta a probabilidade de se cometer o erro tipo 2, que indica a probabilidade de se aceitar a hipótese nula quando ela é falsa (falso negativo). Em contrapartida, amostras com tamanhos menores diminuem a probabilidade de se cometer o erro do tipo 2.

O poder de estudo permite encontrar uma diferença significativa (efeito) que, de fato, aconteça na população. Existe uma relação linear entre o poder do estudo e o tamanho da amostra: quanto maior o poder do estudo, maior o tamanho da amostra necessário, enquanto poderes de estudo menores requerem tamanhos de amostras menores.

O tamanho do efeito indica a menor diferença entre os grupos que é considerada clinicamente relevante (ou biologicamente) e é calculado com base no desvio padrão. A diferença clínica esperada entre grupos deve ser estimada através de resultados de estudos anteriores ou da realização de um estudo piloto. Assim como o nível de significância e o poder do estudo, o tamanho de efeito também possui uma relação proporcional ao tamanho da amostra: quanto menor a diferença clínica a ser detectada, maior o tamanho da amostra necessária, enquanto menores diferenças clínicas requerem amostras de tamanhos menores.

A quarta etapa consiste na realização do cálculo amostral, que pode ser obtido em alguns *softwares* estatísticos. Por fim, a última etapa é a análise de sensibilidade, que se refere ao grau com que desejamos detectar alguma diferença.

Alguns cuidados devem ser tomados durante a realização do cálculo amostral, por exemplo, deve-se considerar a perda de *follow-up*, que alteram o tamanho da amostra e, logo, os resultados estatísticos. O cálculo do tamanho da amostra deve ser feito apenas para o desfecho primário da pesquisa, e não para as variáveis independentes. Por fim, o cálculo amostral deve ser realizado para todos os grupos da pesquisa (por exemplo, grupos de intervenção e controle).

5 Conclusões

Conforme discutido, é essencial a elaboração de um projeto de pesquisa, da escolha correta dos testes estatísticos e do cálculo do tamanho da amostra, para que os resultados sejam confiáveis e que possam ser generalizados para toda a população. Espera-se que esse guia forneça as bases teóricas necessárias para auxiliar o pesquisado iniciante, principalmente da área da saúde, em um dos momentos mais cruciais no desenvolvimento da pesquisa: a escolha do teste estatístico.

6 Referências Bibliográficas

1. TRIOLA, Mario F. **Introdução à Estatística**. Cidade: Editora, 12ª edição. 2017.

2. MORETTIN, Pedro A.; BUSSAB, Wilton de O. **Estatística Básica**. Cidade: Editora, 9ª edição. 2017.

3. GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. Cidade: Editora, 6ª edição, 2017.