



Bioestatística quantitativa aplicada

Edison Capp
Otto Henrique Nienov
Organizadores

Caroline Darski
Charles Francisco Ferreira
Cristiana Palma Kuhl
Fernanda Dapper Machado
Fernanda Vargas Ferreira
Hellen Meiry Grosskopf Werka
Johanna Ovalle Diaz
Marina Petter Rodrigues
Michele Strelow Moreira
Nadine de Souza Ziegler
Paula Barros Terraciano
Pedro Henrique Comerlato
Sinara Santos

Universidade Federal do Rio Grande do Sul
Faculdade de Medicina
Programa de Pós-Graduação em Ciências da Saúde:
Ginecologia e Obstetrícia

Bioestatística Quantitativa Aplicada

Porto Alegre 2020
UFRGS

U58b Universidade Federal do Rio Grande do Sul. Faculdade de Medicina. Programa de Pós-Graduação em Ciências da Saúde: Ginecologia e Obstetrícia Bioestatística quantitativa aplicada/ Universidade Federal do Rio Grande do Sul; organizadores: Edison Capp e Otto Henrique Nienov – Porto Alegre: UFRGS, 2020.

260p.

ISBN: 978-65-86232-43-1

E-Book: 978-65-86232-44-8

1. Epidemiologia e Bioestatística 2. Estatística 3. SPSS I. Capp, Edison, org. II. Nienov, Otto Henrique, org. III Título.

NLM: WA950

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)
(Bibliotecária Shirlei Galarça Salort – CRB10/1929)

Endereço:

PPG em Ciências da Saúde: Ginecologia e Obstetrícia

FAMED – UFRGS

Rua Ramiro Barcellos, 2400/2º andar

CEP 900035-003 – Porto Alegre – RS

Telefone: +55 51 3308 5607

E-mail: ppggo@ufrgs.br

Editoração e diagramação: Edison Capp

Capa: Edison Capp, imagens: www.freepik.com/starline

Edison Capp
Otto Henrique Nienov
Organizadores

Caroline Darski
Charles Francisco Ferreira
Cristiana Palma Kuhl
Fernanda Dapper Machado
Fernanda Vargas Ferreira
Hellen Meiry Grosskopf Werka
Johanna Ovalle Diaz
Marina Petter Rodrigues
Michele Strelow Moreira
Nadine de Souza Ziegler
Paula Barros Terraciano
Pedro Henrique Comerlato
Sinara Santos

1 Conceitos básicos em estatística e epidemiologia

*Caroline Darski
Cristiana Kuhl
Edison Capp
Otto Henrique Nienov*

A pesquisa científica é a aplicação prática de um conjunto de procedimentos objetivos, utilizados por um pesquisador (investigador), para o desenvolvimento de um experimento ou estudo, com a finalidade de produzir um novo conhecimento que pode ser integrado àqueles preexistentes. Constitui-se, portanto, de etapas ordenadamente dispostas, de maneira lógica e racional, as quais o pesquisador deverá conhecer para aplicá-las convenientemente. Estas etapas, de maneira sucinta, incluem desde a escolha do tema a ser pesquisado, o planejamento da investigação, o desenvolvimento do método (que deve ser adequado para responder o problema da pesquisa), a coleta e a tabulação dos dados, a análise dos resultados, a elaboração das conclusões, até a divulgação/publicação do trabalho final.

Portanto, para a realização de uma pesquisa com o rigor científico e metodológico apropriado, pressupõe-se que o pesquisador escolha um tema de sua preferência e seu domínio, defina o problema a ser investigado, elabore um plano de trabalho consistente e, após a execução operacional desse plano, compile e analise os resultados obtidos, e escreva um relatório final, o qual deve ser redigido de forma clara, concisa e responder o objetivo do estudo. Todos estes procedimentos deverão ser executados em etapas.

A definição do tema é o primeiro passo para a elaboração do protocolo de pesquisa. O pesquisador deverá se perguntar: "O que, de fato, quero estudar?". Dentro do tema proposto, o investigador deverá selecionar a questão da pesquisa, a qual corresponde a uma parte delimitada do assunto escolhido. No estabelecimento da questão da pesquisa, o pesquisador deverá partir do assunto geral, o qual deverá ser desmembrado em tópicos específicos, em partes e, então, escolherá uma ou duas dessas partes para elaborar o protocolo de pesquisa.

Uma vez selecionado o tema, a definição do problema é o passo seguinte e de sua correta formulação, dependerá o sucesso da pesquisa. Toda pesquisa nasce de algum problema, elaborado na forma de uma pergunta de pesquisa, que norteará o projeto. É preciso desenvolver uma ideia clara do problema a ser resolvido. Tendo em mente a ideia, você começa a formular hipóteses que explicam a origem desse problema. Portanto, a pergunta de pesquisa estabelecida deve ser estruturada, clara e factível de ser respondida. Lembre-se sempre: todos os procedimentos propostos para a realização da pesquisa deverão ser planejados no sentido de solucionar ou esclarecer o problema proposto. A ordem correta de raciocínio é: “Qual é a questão que necessita de investigação e/ou solução?” “O que ela causa?” “No que a minha pesquisa irá contribuir para solucioná-la”? As características de uma boa questão de pesquisa estão resumidas no quadro 1, no qual estas características básicas são representadas pelo acrônimo FINER: factível, interessante, nova (inovadora), ética e relevante.

Quadro 1 - Características de uma boa questão de pesquisa (adaptado de Hulley e cols., 2008).

Característica	Requisitos
Factível	- Número adequado de sujeitos - Domínio técnico - Viável em tempo e custos - Escopo manejável
Interessante	- Para o investigador
Nova (Inovadora)	- Confirma ou refuta achados anteriores - Expande os achados anteriores
Ética	- Cumpre as normas da Resolução 466/2012 e as demais resoluções do Conselho Nacional de Saúde (CNS).
Relevante	- Para conhecimento científico - Para as diretrizes clínicas de saúde - Para direcionamentos futuros de pesquisa

Portanto, antes do desenvolvimento do projeto de pesquisa, questione se a maneira como a pesquisa está sendo planejada é passível de ser respondida, se você possui domínio suficiente do assunto e experiência para realizá-la e se o tempo e os recursos (sejam financeiros, materiais e/ou humanos) disponíveis serão suficientes. Abaixo, estão descritos detalhadamente os aspectos do acrônimo FINER, que podem ajudar a responder as questões citadas:

Factível: é necessário conhecer não somente seus limites, mas também os recursos disponíveis, antes de se iniciar o estudo. Isso evita o gasto de tempo e de recursos materiais e financeiros. Exemplos de perguntas que devem ser feitas antes de se iniciar um projeto de pesquisa: É possível ser executado? O número de sujeitos é adequado? Domínio técnico e recursos humanos estão disponíveis? Há viabilidade de tempo, custos, espaço físico e recursos materiais?

Interessante: é importante que o problema da pesquisa seja de interesse do pesquisador. O crescimento profissional é um motivo importante e deve ser construído ao longo da sua carreira. Cada pesquisa é um avanço nesta trajetória. No entanto, o conhecimento deve ser a principal motivação do pesquisador.

Inovadora: toda boa pesquisa deve produzir novos conhecimentos, ou, pelo menos, questionar, confirmar ou refutar algum achado anterior, ou, ainda, se os resultados obtidos podem ser replicados ou generalizados em outras populações e cenários. Além disso, uma pesquisa científica não precisa ser necessariamente inédita ou inovadora, porém repetir estudos cujos resultados já estão bem estabelecidos pela comunidade científica, não é recomendado, uma vez que desperdiçam trabalho, tempo e recursos.

Ética: neste aspecto, precisamos considerar se a pesquisa está delineada de modo que não comprometa a integridade, a autonomia e o bem-estar do participante da pesquisa. A Resolução do Conselho Nacional da Saúde Nº 466, de 12 de dezembro de 2012, considera que toda pesquisa envolvendo seres humanos, direta ou indiretamente, envolve risco, podendo o dano eventual ser imediato ou tardio, no indivíduo ou à coletividade. Desta forma, toda e qualquer pesquisa envolvendo seres humanos deve ser submetida

à apreciação de um Comitê de Ética em Pesquisa (CEP). As pesquisas envolvendo seres humanos devem atender às exigências éticas e científicas fundamentais, que implicam consentimento livre e esclarecido dos indivíduos pesquisados, proteção de grupos vulneráveis (como gestantes e comunidades indígenas) e dos legalmente incapazes (como crianças e pessoas com deficiência intelectual), respeitando o princípio da autonomia. Potenciais riscos e benefícios deverão ser ponderados, individuais ou coletivos, comprometendo-se, a pesquisa, com o máximo de benefícios e a redução de danos, seguindo o princípio da beneficência. Por fim, estima-se a relevância social, com vantagens significativas para os sujeitos e minimização do ônus para os vulneráveis, o que garante a igual consideração dos interesses envolvidos sem perder o sentido de sua destinação sócio-humanitária, conforme os princípios da justiça e de equidade. A prática que envolve o uso de animais é regulada pelo Conselho Nacional de Controle de Experimentação Animal (CONCEA), que orienta as Comissões de Ética no Uso de Animais (CEUA), pesquisadores e docentes na utilização de animais em aulas e pesquisas.

Relevante: sem dúvida, este é o requisito mais importante quando se elabora uma questão de pesquisa. Se a pergunta de pesquisa não for relevante, ela pode tornar-se desinteressante para o pesquisador e sua equipe. Sua relevância está no fato de mostrar e justificar como o estudo pretendido pode ser inserido em um contexto mais amplo. Deve-se questionar por que essa questão é importante e como suas respostas poderão contribuir para o conhecimento científico. A questão de pesquisa deve ser cuidadosamente revisada e estruturada e, caso haja necessidade, o projeto deve ser adequado para que possa ser respondida.

Considerando que os critérios do método FINER descrevem os importantes aspectos da questão em geral, um formato útil para usar no desenvolvimento de uma questão de pesquisa específica é o acrônimo PICOT: considera a população (P) de interesse, a intervenção (I) em estudo, a comparação (C) (grupo ou com o que a intervenção está sendo comparada) e o resultado do interesse (O). Também podemos incluir no acrônimo o delineamento do estudo (T), ou seja, o tipo de estudo que melhor responderá a questão de pesquisa. A abordagem PICOT ajuda a gerar uma questão que

auxilia na construção da estrutura do estudo e, posteriormente, no desenvolvimento do protocolo, citando os critérios de inclusão e exclusão e identificando os grupos de pacientes a serem incluídos. Conhecer a população de interesse, a intervenção e o resultado esperado também pode ajudar o pesquisador a identificar uma ferramenta de medição de resultados apropriada.

A questão de pesquisa é o ponto de partida de qualquer projeto de pesquisa, ou seja, é quando o pesquisador define um problema e como procurará abordá-lo. A questão de pesquisa, como mencionado, pode ser construída através do acrônimo PICOT (exemplos no Quadro 2), no qual:

P = população, paciente ou problema/com quem a pergunta se relaciona;

I = intervenção/exposição/pode ser uma terapia, teste diagnóstico, fator prognóstico ou interesse;

C = comparador/não exposição/pode ser outra intervenção, teste diagnóstico, placebo ou cuidado usual padrão;

O = desfecho/especificar claramente no que você está interessado;

T = delineamento/decidir o melhor desenho do estudo que possa responder a sua questão de pesquisa”.

Quadro 2 - Exemplos de PICOT.

		Qual o efeito de antitussígenos na melhora em sinais e sintomas respiratórios e na qualidade de vida de pacientes com tosse?	Qual o desempenho de aprendizagem em crianças com deficiência visual comparado a crianças sem deficiência visual?	Qual o impacto do tabagismo no desenvolvimento do câncer de pulmão em jovens adultos?
Item	Descrição	Exemplo 1	Exemplo 2	Exemplo 3
P	População	Pacientes com tosse	Crianças	Jovens adultos
I	Intervenção/ Exposição	Antitussígenos	Deficiência visual	Hábito de fumar
C	Comparador	Não fazer nada/ placebo	Sem deficiência visual	Sem hábito de fumar
O	Outcome/ Desfecho	Supressão da tosse, qualidade de vida	Aprendizagem	Câncer de pulmão
T	Delineamento	Ensaio clínico randomizado	Estudo de casos e controles	Estudo de coorte

Quanto mais definida a população de interesse e, portanto, quanto melhor definidos os critérios de inclusão e exclusão, maior o efeito na interpretação e subsequente generalização dos resultados da pesquisa. Restringir a população do estudo pode limitar o viés e aumentar a validade interna do estudo. No entanto, essa abordagem limitará a validade externa do estudo e, portanto a generalização dos achados para outros cenários ou populações. Por outro lado, uma população de estudo menos definida pode melhorar a representatividade da amostra, mas aumentar o viés e reduzir a validade interna do estudo. Uma pergunta de pesquisa mal elaborada implicará na escolha equivocada do desenho do estudo e, potencialmente, dificultará a chance de se obter um achado com relevância clínica (positiva ou negativa), o que poderá prejudicar a publicação. Sem dedicar tempo e recursos suficientes para o desenvolvimento da questão de pesquisa, a qualidade do estudo e os resultados subsequentes podem ser comprometidos. Durante o estágios iniciais de qualquer pesquisa, é fundamental formular uma pergunta de pesquisa que seja clinicamente relevante e ética.

Projeto de pesquisa

A elaboração do projeto corresponde à etapa mais importante e de maior complexidade da pesquisa, pois, do correto delineamento (desenho) depende o sucesso na obtenção das respostas esperadas pela questão da pesquisa. Delinear uma pesquisa é, em última análise, planejar a realização de sua parte científica operacional, tanto experimental como observacional; ou seja, é escrever corretamente um projeto onde estarão previstas todas as etapas de sua realização.

Com uma boa questão de pesquisa, se parte para o planejamento e elaboração do projeto de pesquisa para que esse seja submetido nas plataformas necessárias. A parte estatística se encontra inserida dentro da seção Método (também chamada de Material e Métodos ou Metodologia) do seu projeto de pesquisa, no qual devem constar seu cálculo de amostra e quais análises estatísticas são esperadas para responder a questão de pesquisa, conforme observado na figura 1.



Figura 1 - Partes de um projeto de pesquisa.

A compreensão do método estatístico, suas características e, principalmente, suas limitações, é fundamental para uma análise crítica responsável. A análise estatística adequada evita erros nos resultados e, conseqüentemente, na interpretação dos dados.

Delineamento de pesquisa

Os estudos podem ser classificados quanto ao período de tempo durante o qual os dados foram registrados em relação ao tempo no qual o estudo começou. São considerados retrospectivos quando se parte do efeito para a causa, ou seja, os dados sobre os pacientes já existem (ex. em banco de dados ou prontuários eletrônicos). São considerados prospectivos quando se parte da causa e se procura o efeito, com acompanhamento de pacientes ao longo do tempo (*follow up*, estudo com seguimento, sequencial). Estudos prospectivos geralmente são mais caros e sujeitos a perdas de participantes.

O delineamento do projeto precisa ser definido e deve ser apresentado explicitamente como item. Os estudos podem ser divididos em observacionais e experimentais, conforme o esquema da figura 2.

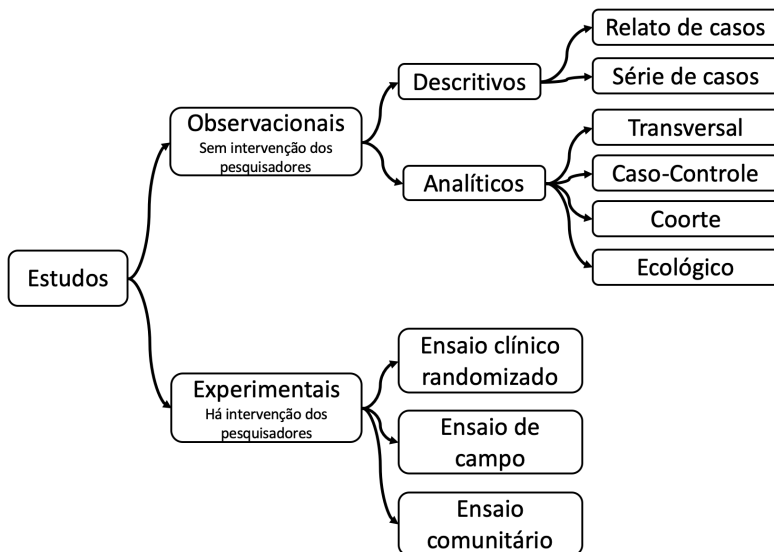


Figura 2 - Delineamentos de pesquisa.

Estudos observacionais

São aqueles em que não há intervenção do pesquisador. O pesquisador simplesmente observa o paciente, as características da doença ou transtorno, e sua evolução, sem interferir ou modificar qualquer aspecto que esteja sob observação. Incluem estudos de coorte, estudos de caso-controle e estudos transversais. Podem ser classificados em descritivos ou analíticos. Os estudos descritivos não testam hipóteses (as hipóteses surgem durante o projeto), portanto não fazem associações. Eles descrevem uma condição de indivíduo ou pequeno grupo de indivíduos. São geralmente usados para assuntos ainda não muito conhecidos como doenças novas, manifestações raras ou associações de doenças, caracterizando seus aspectos semiológicos, etiológicos, fisiopatológicos e epidemiológicos. São considerados estudos descritivos estudos de casos e séries de casos onde a diferença entre eles é o número de indivíduos. O primeiro pode variar entre 1 a 9 indivíduos e o segundo tem um mínimo de 10 participantes.

Os estudos analíticos, por sua vez, testam hipóteses (definidas *a priori*). Estão relacionados a distribuição de eventos na população, estimando incidência e prevalência. O investigador introduz um fator de exposição ou de causalidade, e avalia-o utilizando ferramentas estatísticas. A precisão dos dados está diretamente associada com a representatividade dos dados na população (daí a importância do cálculo amostral). São considerados estudos analíticos os estudos transversal, de coorte, caso-controle e ecológico.

Estudo transversal

É o “retrato” da situação no momento em que o pesquisador observa. A população ou amostra é estudada em um único momento. É utilizado para avaliar se existe relação entre as variáveis da pesquisa e é importante para determinação de prevalência de doenças e acurácia de testes diagnósticos. O estudo transversal é um estudo barato e fácil de ser realizado e gera hipóteses de associação ao risco, porém apresentam desvantagens como a impossibilidade de estabelecer relações causais por não provarem a existência de uma sequência temporal entre a exposição ao fator e o subsequente desenvolvimento da doença (não há possibilidade de determinação do que é exposição e do que é efeito). São exemplos de estudos transversais: estudo ecológico (as medidas são observadas em grandes populações), estudo de prevalência (para determinar a prevalência de uma característica em uma população de interesse é medida a proporção entre o número de pessoas acometidas pela doença e as que permanecem saudáveis), estudo de sensibilidade e especificidade (para avaliar um novo teste diagnóstico é utilizado um padrão de referência, representado por um teste ideal conhecido como “padrão-ouro”).

Estudo de coorte

É um estudo longitudinal, ou seja, envolve acompanhamento do cenário ao longo do tempo (como se fosse um filme). Um grupo de sujeitos com uma característica comum é acompanhado para verificar a ocorrência ou não de determinado desfecho. Os indivíduos são distribuídos em expostos e não expostos a um fator em estudo e são acompanhados durante um determinado período de tempo para verificar a incidência de uma

doença e fatores de risco. Compara-se a proporção dos que ficaram doentes dentre os expostos e a proporção dos que ficaram doentes entre os não expostos. Esse modelo de estudo é do tipo analítico e longitudinal, e pode ser prospectivo ou retrospectivo (coorte histórica). As vantagens desse estudo são: sequência temporal dos acontecimentos, a doença é claramente estabelecida, é ideal para ver incidência, tem alta precisão e pode medir diversos resultados. Como desvantagens, o estudo de coorte é um estudo caro e demorado, não é útil para doenças raras e há perdas durante o seguimento (se espera 20% de perda dos pacientes).

Estudo de caso-control

Também chamado de estudo caso-referência, é um estudo onde se parte de indivíduos com doença (casos) e sem doença (controle) para investigar a associação entre um evento de interesse e alguns preditores, ou seja, o parâmetro a ser estudado é a exposição ou não a um fator (Figura 3). É um estudo importante para analisar doenças raras e situações de surtos ou agravos desconhecidos. É um estudo mais rápido que o Coorte, analisa vários preditores e é um estudo inicial para novas hipóteses, mas suas informações geralmente são incompletas, possui um importante viés de memória e fatores de confusão e não avalia frequência dos eventos. São modelos de estudo analíticos, longitudinais e, geralmente, retrospectivos (casos já tiveram o diagnóstico antes do início do estudo), porém podem ser prospectivos em algumas situações (quando os casos são diagnosticados após o início do estudo).

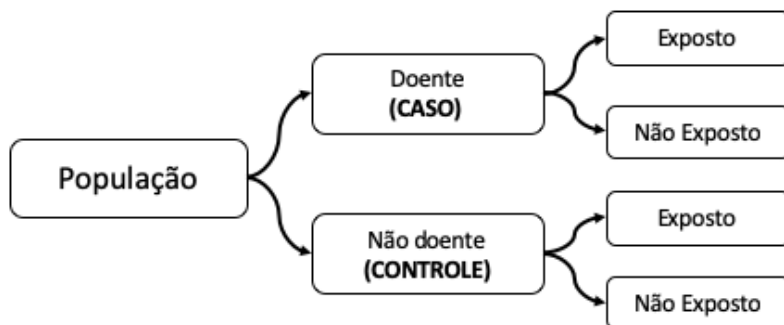


Figura 3 - Estudo do tipo caso-control.

Estudo ecológico

É um estudo onde a unidade de análise é a população ou um grupo de pessoas pertencentes a uma área geográfica definida. São avaliados como os contextos social e ambiental podem afetar a saúde da população de interesse. Esse estudo compara variáveis globais quase sempre por meio de correlação entre indicadores de condições de vida e indicadores de situação de saúde. É um estudo que avalia efeitos contextuais e gera e testa novas hipóteses, porém há dificuldade de controlar fatores de confusão e não há possibilidade de associação entre fator e doença no indivíduo.

Estudos experimentais

Há intervenção do pesquisador. Podem ser classificados como ensaio clínico randomizado, ensaio de campo e ensaio comunitário.

Ensaio clínico randomizado

É um estudo prospectivo usado para comparar determinada intervenção com outra (Figura 5). É considerado o padrão-ouro para teste de eficácia de uma intervenção. As fases de seleção e inclusão dos pacientes são importantes para a alocação aleatória dos grupos. O mascaramento do ensaio clínico randomizado pode ser aberto (não ocorre), cego (apenas a equipe de investigação sabe qual foi o tratamento instituído e os pacientes desconhecem), duplo-cego (a equipe, os investigadores e tampouco os pacientes sabem a que grupos os mesmos pertencem), triplo-cego (equipe de investigação, pacientes e estatístico desconhecem qual é o grupo controle e o grupo experimental) ou quadruplo-cego (além da equipe de investigação, dos pacientes e do estatístico, o investigador que está escrevendo a discussão sobre os resultados, inicialmente de forma genérica quanto à identidade dos grupos, também não sabe qual é o grupo controle e o grupo experimental) (Figura 6). Como vantagens, esse estudo apresenta balanço dos fatores prognósticos entre os grupos, permite coleta de informações detalhadas e o cegamento dos participantes. Como desvantagens, tem-se a generalização externa reduzida por critérios de exclusão, é um tipo de estudo demorado com custo elevado, geralmente necessita de amostras grandes e os indivíduos podem não aderir às intervenções alocadas.

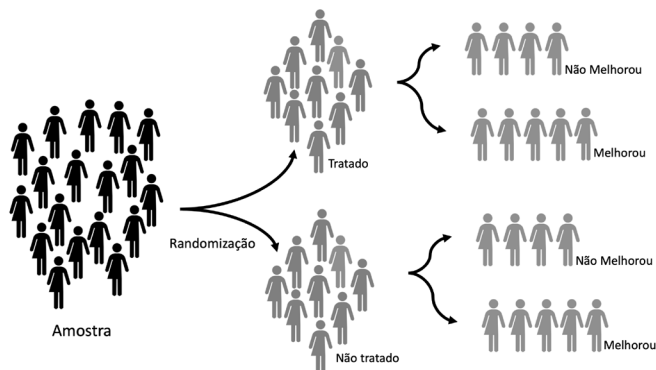


Figura 5 - Ensaio clínico randomizado.

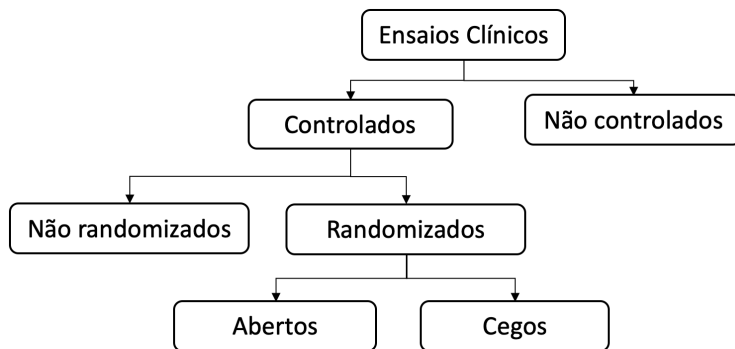


Figura 6 - Tipos de ensaios clínicos randomizados.

Ensaio de campo

É semelhante ao ensaio clínico, mas a população não é de pacientes e sim de indivíduos com risco de doenças. Os dados são coletados na população em geral.

Ensaio comunitário

Envolve a intervenção em nível de comunidades. Avalia a eficácia e efetividade de intervenções que busquem prevenção primária através de modificação de fatores de risco na população. É um estudo conduzido dentro de um contexto socioeconômico de uma população naturalmente formada. Uma importante limitação deste estudo é a dificuldade de isolar uma comunidade (difícil controle de fatores externos).

Estudos pré-clínicos

Estudo *in vivo*

Refere-se à experimentação realizada em animais vivos. São importantes para definição do perfil farmacológico e toxicológico de substâncias, testes de segurança, processos farmacocinéticos e estudos de pré-formulação e formulação.

Estudo *in vitro*

Refere-se a executar um determinado procedimento em um ambiente controlado fora de um organismo vivo. Frequentemente são utilizadas culturas de células. Uma das desvantagens é que dificilmente se consegue replicar as condições celulares de um organismo, levando a resultados que podem não corresponder às condições *in vivo*.

Estudo *in silico*

É feito através de técnicas computacionais (programas/softwares) que simulam eventuais efeitos e/ou interações.

Existe uma hierarquia na força da evidência conforme o delineamento dos estudos, sendo que quanto maior o nível de evidência, mais importante é a informação que o estudo apresenta para a tomada de decisão (Figura 7).

A definição do delineamento de pesquisa a ser utilizado é importante para saber quais testes estatísticos serão usados. Técnicas estatísticas insuficientes ou inadequadas podem ser uma grande fraqueza de um projeto de pesquisa. A qualidade e a adequação do método estatístico utilizado pode implicar no sucesso ou no fracasso de um artigo.

Conceitos básicos em estatística

A estatística é a ciência que se preocupa com a organização e descrição dos dados (estatística descritiva) e com análises e interpretações destes (estatística indutiva ou inferente). Tem como objetivo tirar conclusões sobre populações com base nos resultados observados em amostras extraídas dessas populações.



Figura 7 - Níveis de evidência.

População

É o conjunto de elementos (pessoas, animais,...) com pelo menos uma característica em comum. Esta característica comum deve delimitar claramente quais os elementos que pertencem à população e quais os elementos que não pertencem (através de critérios de inclusão e exclusão). Uma população pode ser finita e pequena, sendo fácil de conhecer todos os seus elementos. Porém, na maioria das vezes, é finita, mas incontável ou mesmo infinita. Nestes dois últimos casos, para conhecer uma população, a estatística lança mão de um recurso que é coletar uma amostra desta população e caracterizar alguma variável da população, ou seja, tirar conclusões sobre a população a partir de resultados obtidos em amostras (inferência estatística).

Amostra

É um subconjunto de uma população, definido segundo algum critério estabelecido pelo observador, sobre o qual se conduz algum estudo com objetivo de se fazer inferências estatísticas sobre toda a população. Esta deve ser representativa de toda a população de interesse.

Indivíduos

São os objetos descritos por um conjunto de dados. Podem ser pessoas, animais ou coisas.

Observação

É o valor de uma variável de um elemento particular de uma amostra ou população.

Variável

É a característica de interesse de elementos individuais de uma população ou amostra que nos interessa averiguar estatisticamente. Geralmente, correspondem às colunas no banco de dados. Podem ser qualitativas ou quantitativas, conforme figura 8.

As **variáveis qualitativas** são medidas originalmente em categorias. A descrição da característica de interesse resulta em valor não numérico. Exemplo: sexo (masculino = 0, feminino = 1), profissão, religião, município, região. São comumente resumidas pelo número de indivíduos em cada grupo (frequência absoluta, "n") e porcentagem de indivíduos em cada grupo (frequência relativa, "n%"). São classificadas em:

- Nominais: não existe ordenação nas possíveis respostas. Ex: sexo, estado civil;
- Ordinais: existe uma certa ordem nas possíveis respostas, ou seja, podem ser arranjados em alguma ordem. Ex: grau de escolaridade, hierarquias, classe de renda familiar.

As variáveis qualitativas são avaliadas por distribuição das frequências (descrição numérica), isto é, conhecer qual o comportamento dessa variável, analisando a ocorrência de suas possíveis realizações. Como mencionado anteriormente, são expressas em frequência absoluta e frequência relativa (tabela 1). Não se pode calcular média, variância ou desvio-padrão de variáveis qualitativas ou categóricas. A representação gráfica pode ser feita em gráficos em barras e de composição de setores (retângulos ou pizza).

Tabela 1 - Exemplo de variáveis qualitativas.

Grau de instrução	Frequência	Proporção	Porcentagem
Fundamental	12	0,333	33,33%
Médio	18	0,500	50,00%
Superior	6	0,167	16,67%
Total	36	1,000	100,00%

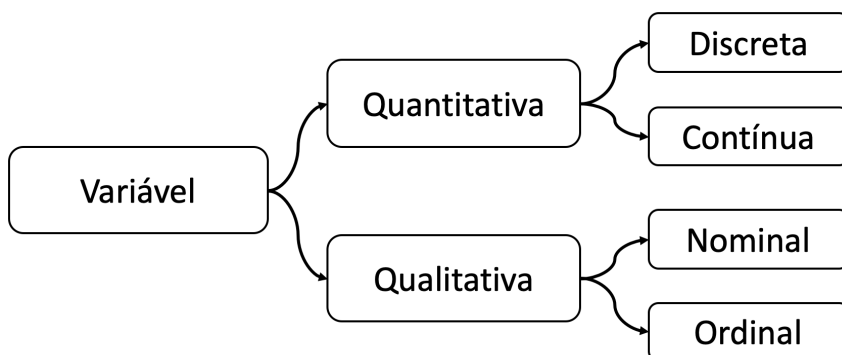


Figura 8 - Variáveis quantitativas e qualitativas.

As **variáveis quantitativas** podem ser mensuradas através de escalas, isto é, através de unidades de medida. Exemplos: idade (medida em anos, ou meses), massa (em kg), estatura (em cm). São expressas por medidas de tendência central (média, mediana e moda), de dispersão (variância, desvio padrão, amplitude), e em certos casos, a relação entre duas variáveis (massa e estatura). São classificadas em:

- Discretas: quando assumem apenas valores pertencentes a um conjunto finito ou enumerável de números inteiros. São obtidas mediante alguma forma de contagem. Ex: número de filhos;
- Contínuas: quando os possíveis valores estão dentro de um intervalo, aberto ou fechado, dos números reais. Ex: massa corporal, estatura, triglicerídeo sérico.

As variáveis quantitativas são expressas usando medidas descritivas (ex: média, desvio padrão, valor mínimo, valor máximo) e demonstradas em gráficos (ex: histograma, gráficos de dispersão e *box plot*).

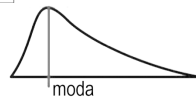
Medidas de tendência central

Servem para verificar a distribuição dos dados brutos (ou frequências) sobre o eixo de variação da variável em questão.

- Média: é o resultado do somatório de todos os valores da amostra dividido pelo número de valores (n);
- Moda (ou modas): é definida como o valor (ou valores) de máxima frequência. Uma amostra pode ser amodal (não ter moda), unimodal (apenas um valor com repetição expressiva), bimodal (dois valores tem repetições expressivas);
- Mediana: é o valor que está em posição intermediária na distribuição de valores, isto é, tem tantos elementos acima quanto abaixo de sua posição. Para determinar a mediana, primeiramente deve-se ordenar os "n" valores em ordem crescente. Quando "n" é ímpar, a mediana é um valor real da amostra. Quando "n" é par, a mediana é um valor fictício entre dois valores reais da amostra.

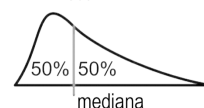
ID	P1	P2	P3	P4	P5	P6
Idade (em anos)	25	22	35	28	35	33

MODA: 35 anos de idade



MEDIANA:

22 25 28 33 35 35
 $28 + 33 = 61$
 $61/2 = 30,5$ anos de idade



MÉDIA:

$25 + 22 + 35 + 28 + 35 + 33 = 178$
 $178/6 = 29,67$ anos de idade

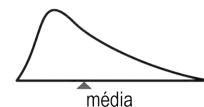


Figura 9 - Exemplos de mediana, moda e média.

Em distribuições normais (chamadas de simétricas ou paramétricas) a média e a mediana são próximas e os dados são expressos em média e desvio padrão. Em distribuições não normais (chamadas assimétricas ou não paramétricas) a média e a mediana são distantes e os dados são expressos em mediana e quartis ou valores mínimo e máximo.

Medidas de distribuição

Servem para indicar como os dados se distribuem em torno do valor central (média ou mediana). Portanto, caracterizam o grau de variação existente em um conjunto de valores.

- Amplitude: é definida como a diferença entre o maior e o menor valores do conjunto de dados;
- Variância: é definida como a medida do espalhamento dos valores em torno da média;
- Desvio padrão: é a raiz quadrada positiva da variância. É expressa na mesma unidade da variável, por isso é de maior interesse que a variância nas aplicações práticas, sendo mais realístico para efeito de comparação de dispersões. O desvio padrão é uma medida que indica a dispersão dos dados dentro de uma amostra com relação à média. Assim, quando se calcula o desvio padrão juntamente com a média de diferentes grupos, obtêm-se mais informações para avaliar e diferenciar seus comportamentos;
- Erro padrão: é uma medida de variação de uma média amostral em relação à média da população. Sendo assim, é uma medida que ajuda a verificar a confiabilidade da média amostral calculada. Através do erro padrão, pode-se estimar um intervalo de confiança (IC95%) para a média populacional a partir da média amostral calculada. A média das amostras \bar{x} encontra-se no centro desse intervalo, e o intervalo é $\bar{x} \pm$ intervalo de confiança.

O desvio padrão trata de um índice de dispersão da amostra em relação à média, enquanto o erro padrão é uma medida que ajuda a avaliar a confiabilidade da média calculada.

- Coeficiente de variação: é definido como o quociente entre o desvio padrão e a média, sendo frequentemente expresso em porcentagem;
- Percentil: são os valores de uma variável quantitativa que dividem os dados ordenados em grupos de modo que uma determinada porcentagem esteja acima e outra porcentagem esteja abaixo;
- Quartis: são valores da variável que dividem a distribuição em quatro partes iguais, cada uma delas com 25% dos dados totais ordenados. Mediana divide as observações em duas partes iguais. São utilizados em amostras com distribuições não normais (Figura 10).

Medidas de assimetria

São importantes para descrição visual quando a distribuição dos dados é normal ou não normal. Em uma distribuição estatística, a assimetria (*skewness*) é o quanto sua curva de frequência se desvia ou se afasta da posição simétrica.

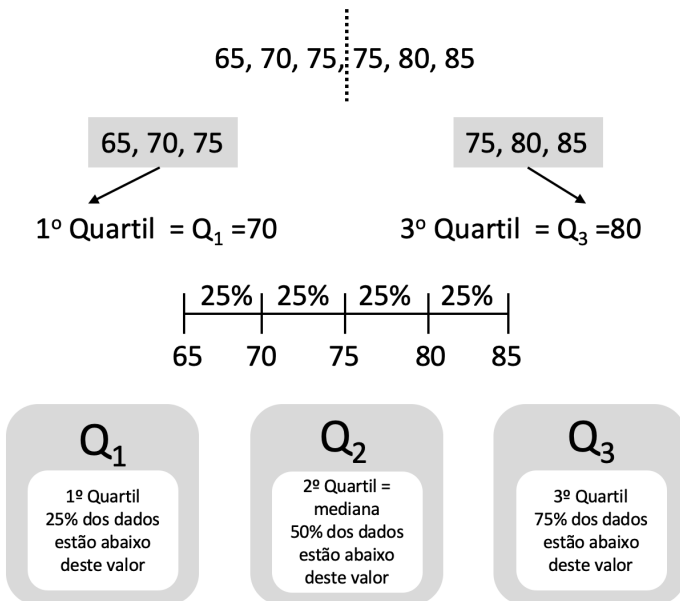


Figura 10 - Intervalos interquartis.

Curtose

É o menor ou maior grau de "achatamento" da distribuição ou curva de frequência considerada em relação a uma curva normal representativa da distribuição.

Referências

Bonita R, Beaglehole R, Kjellström T. Epidemiologia básica. 2. ed. São Paulo: Santos, 2010. 213 p.

Callegari-Jacques SM. Bioestatística: princípios e aplicações. Porto Alegre: ArtMed, 2011. 255p.

Farrugia P, Petrisor BA, Farrokhyar F, Bhandari M. Practical tips for surgical research: Research questions, hypotheses and objectives. Canadian Journal of surgery. Journal Canadien de Chirurgie. 2010;53(4):278-281.

Guimarães LSP, Hirakata VN, Camey AS, Nunes LN, Mancuso ACB. Os principais delineamentos na epidemiologia: ensaios clínicos (Parte II). Revista HCPA. 2013;33(3/4):295-302

Hulley SB, Cumming SR, Browner WS, Grady DG, Hearst NB, Newman TB. Delineando a pesquisa clínica: uma abordagem epidemiológica. Artmed, Porto Alegre; 2008. 384 p.

Mancuso ACB, Camey AS, Nunes LN, Hirakata VN, Guimarães LSP. Os principais delineamentos na epidemiologia: ensaios clínicos (Parte I). Revista HCPA. 2013;33(3/4):286-294.

Nunes LN, Camey AS, Guimarães LSP, Mancuso ACB, Hirakata VN. Os principais delineamentos na epidemiologia. Revista HCPA. 2013;33(2):178-183.

Regatieri CV, Alves AV, Rocha E. Medical acumen and scientific approach meet the FINER method. Arq. Bras. Oftalmol. 2019;82(4): 1-5.

Riva JJ, Malik KM, Burnie SJ, Endicott AR, Busse JW. What is your research question? An introduction to the PICOT format for clinicians. J Can Chiropr Assoc. 2012 Sep;56(3):167-71.

Exercícios sugeridos

Escolha a alternativa correta:

1. Para realizar um estudo epidemiológico, um investigador selecionou 33 mulheres com câncer de endométrio e 33 mulheres sem a doença. As mulheres foram pareadas por idade e paridade. Todas as mulheres foram questionadas sobre o uso de estrogênio e progesterona em algum momento de sua vida. O desenho de estudo utilizado foi:

- a) Estudo de coorte.
- b) Estudo caso-controle.
- c) Estudo ecológico.
- d) Estudo transversal.
- e) Ensaio clínico randomizado.

2. Um pesquisador está tentando determinar os fatores de risco para o câncer de mama. A amostra selecionada foi de 48.000 mulheres, acompanhados ao longo de 5 anos, para ver quais desenvolveriam a doença. Qual tipo de estudo é esse?

- a) Estudo de coorte.
- b) Estudo de caso-controle.
- c) Ensaio clínico randomizado.
- d) Estudo de caso.
- e) Estudo ecológico.

3. Um grupo de 500 mulheres é observado em um dado momento e mede-se a prevalência de Síndrome do Ovário Policístico (SOP). Que tipo de estudo é esse?

- a) Estudo caso-controle.
- b) Revisão sistemática e meta-análise.
- c) Estudo de coorte.
- d) Estudo ecológico.
- e) Estudo transversal.

4. Um pesquisador atende uma paciente com Síndrome de Rokitsky, uma doença muito rara que provoca alterações no útero e na vagina, e escreve um artigo sobre esta paciente, especialmente sobre a sua história, aspectos físicos e curso clínico. Que tipo de estudo é esse?

- (a) Estudo caso-controle.
- (b) Série de casos.
- (c) Estudo de coorte.
- (d) Estudo de caso.
- (e) Estudo transversal.

5. Para avaliar a relação entre alterações no percentil do Índice de Massa Corporal (IMC), através da curva de Atalah, e resultados perinatais, foi realizado um estudo com 1.475 mulheres. Os dados referentes ao peso gestacional, características sociodemográficas e resultados perinatais foram coletados de forma retrospectiva, por meio de prontuários, cartão pré-natal e entrevistas realizadas, em um momento do período pós-parto. As mulheres foram classificadas, de acordo com a curva de Atalah, em: baixo peso, peso adequado, sobrepeso e obesidade. Isto é um:

- a) Estudo transversal.
- b) Estudo de coorte.
- c) Ensaio clínico randomizado.
- d) Estudo caso-controle.
- e) Série de casos.

6. Um estudo envolvendo 32 parturientes, realizado em um hospital universitário, buscou determinar o desfecho materno e neonatal de acordo com a aplicação ou não da estimulação elétrica nervosa transcutânea (TENS) para alívio da dor do parto antes da analgesia combinada raqui-peridural (ACRP). Essas pacientes foram aleatoriamente designadas para receber ou não a TENS antes da ACRP. A análise dos resultados incluiu a intensidade da dor medida pela escala visual analógica (EVA), o intervalo entre a avaliação inicial e a instalação da ACRP, a duração do trabalho de parto, as taxas de parto por cesárea e por parto vaginal instrumental (fórceps) e, frequência de hipóxia neonatal. Isto representa um:

- a) Estudo transversal.
- b) Estudo de coorte.
- c) Ensaio clínico randomizado.
- d) Estudo caso-controle.
- e) Série de casos.

7. Um estudo teve como objetivo avaliar a frequência de deleção homozigótica dos genes GSTM1 e GSTT1 e suas combinações entre pacientes com câncer de mama e indivíduos saudáveis, associando-os à suscetibilidade da doença. Assim, 39 mulheres diagnosticadas com câncer de mama confirmadas por exame patológico e 39 mulheres saudáveis sem evidência de câncer e sem histórico familiar prévio de câncer de mama foram convidadas a participar. Todas responderam a um questionário e foram submetidos à coleta de amostras de sangue. O DNA foi extraído do sangue e a genotipagem foi realizada por reação em cadeia da polimerase (PCR). Isto caracteriza um:

- a) Estudo transversal.
- b) Estudo de coorte.
- c) Ensaio clínico randomizado.
- d) Série de casos.
- e) Estudo de caso-controle.

8. Um estudo avaliou as alterações de HDL-Colesterol induzidas pela cirurgia bariátrica em obesos não diabéticos com níveis baixos ou normais de HDL-Colesterol antes da cirurgia. Neste estudo de coorte prospectivo, 90 indivíduos foram avaliados quanto ao sexo (masculino ou feminino), idade (em anos), técnica cirúrgica (bypass gástrico ou gastrectomia em manga), tempo de cirurgia (em meses), grau de obesidade (sobrepeso, obesidade graus I, II e III), perda de peso (em kg), circunferência da cintura (em cm), níveis séricos de glicose (em mg/dL), triglicerídeos (em mg/dL), HDL-Colesterol (em mg/dL), não HDL-Colesterol (em mg/dL), LDL-Colesterol (em mg/dL), além do gasto energético estimado (em METs). Responda às seguintes questões:

a) Estruture a questão de pesquisa no acrônimo PICOS;

b) Das variáveis avaliadas, quais são consideradas quantitativas e quais são consideradas qualitativas? Classifique-as em nominal ou ordinal se qualitativas e em contínua ou discreta se quantitativas.

9. Neste exercício, estruture a questão de pesquisa do seu projeto de pesquisa no formato PICOT (*patients, interventions, comparisons, outcomes, study design*). Avalie se sua questão está bem definida, se é factível, interessante, inovadora, ética e relevante. Também descreva as possíveis variáveis do estudo e classifique-as quanto à sua natureza, se qualitativa (nominal ou ordinal) ou quantitativa (contínua ou discreta). Este é um bom exercício prático para aplicar os conceitos vistos neste capítulo.