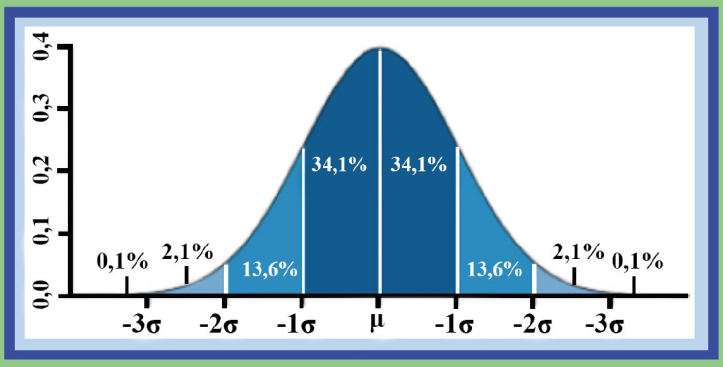
AMOSTRAGEM APLICADA EM PSICOLOGIA

Com introdução no SPSS e R



Edwirde Luiz Silva Camêlo
Dalila Camêlo Aguiar
Ivan Olier
Ramón Gutiérrez Sánchez

30 anos
eduepb

Edwirde Luiz Silva Camêlo
Dalila Camêlo Aguiar
Ivan Olier
Ramón Gutiérrez Sánchez

Amostragem aplicada em Psicologia

Com introdução no SPSS e R

eduepb
Campina Grande-PB | 2025

Universidade Estadual da Paraíba
Profª. Célia Regina Diniz | *Reitora*
Profª. Ivonildes da Silva Fonseca | *Vice-Reitora*

Editora da Universidade Estadual da Paraíba
Cidoval Morais de Sousa | *Diretor*

Editora indexada no SciELO desde 2012

Editora filiada a ABEU

EDITORA DA UNIVERSIDADE ESTADUAL DA PARAÍBA
Rua Baraúnas, 351 - Bairro Universitário - Campina Grande-PB - CEP 58429-500
Fone: (83) 3315-3381 - http://eduepb.uepb.edu.br - email: eduepb@uepb.edu.br

Editora da Universidade Estadual da Paraíba
Cidoval Morais de Sousa (*Diretor*)

Expediente EDUEPB

Design Gráfico e Editoração
Erick Ferreira Cabral
Jefferson Ricardo Lima A. Nunes
Leonardo Ramos Araujo

Revisão Linguística e Normalização
Antonio de Brito Freire
Elizete Amaral de Medeiros

Assessoria Editorial
Eli Brandão da Silva

Assessoria Técnica
Thaise Cabral Arruda

Divulgação
Danielle Correia Gomes

Comunicação
Efigênio Moura

Depósito legal na Câmara Brasileira do Livro - CDL

A525 Amostragem aplicada em Psicologia [recurso eletrônico] : com introdução no SPSS e R / Edwirde Luiz Silva Camêlo … [et al.]. – Campina Grande : EDUEPB, 2025.
113 p. : il. color. ; 15 x 21 cm.

ISBN: 978-65-5221-099-9 (Impresso)
ISBN: 978-65-5221-098-2 (3.100 KB - PDF)
ISBN: 978-65-5221-101-9 (Epub)

1. Estatística Descritiva. 2. Estatística Aplicada à Psicologia. 3. Programa SPSS. 4. Programa R. 5. Estatística Básica. I. Camêlo, Edwirde Luiz Silva. II. Aguiar, Dalila Camêlo. III. Olier, Ivan. IV. Sánchez, Ramón Gutiérrez. V. Título.

21. ed. CDD 519.53

Ficha catalográfica elaborada por Fernanda Mirelle de Almeida Silva - CRB - 15/483

AGRADECIMENTOS

Gostaríamos de agradecer à EDUEPB, que muito tem apoiado e estimulado a divulgação dos trabalhos dos professores da UEPB. Devemos também agradecer ao trabalho profissional do Prof. Cidoval??, que pacientemente abrilhantou este trabalho realizando as devidas correções gramaticais. Finalmente, nós autores não poderíamos deixar de agradecer aos alunos que nos permitiram a experiência para elaboração deste livro, aos colegas que contribuíram para a realização desta compilação e em especial ao **Eterno, Criador dos céus e da terra** que nos outorgou vida para concluir esta obra.

Lista de Tabelas

Lista de Figuras

Sumário

Prefácio

O propósito deste livro é apresentar as técnicas de amostragem estatística em duas facetas: teoria e prática. Seu contenido está enfocado a docentes e discentes universitários de todos níveis que utiliza a amostragem estatística, assim como aos profissionais dos setores en que se aplica a técnica de amostragem (economia, transporte, medicina, psicologia da saúde, matemática, comércio, controle estatístico de qualiade, etc.).

O livro inicia apresentando as ferramentas básicas para a amostragem estatística explicando os passos para sua utilização em psicologia. Sabendo que a teoria da probabilidade é o fundamento dos dois métodos de amostragens observado neste livro. Um conhecimento dos métodos gerais de estatística e da teoria básica das estimações do ponto de vista estatístico é essencial para um entendimento adequado do desenvolvimento rigoroso da teoria de amostragem.

NO liro se oferece algumas demonstrações que incorporam instrumentos metamáticos avançados mesclados com a utilização prática dos resultados e cuja finalidade é a de dirigir a atenção do estudante para utilidade dos resultados obtidos.

O capítulos 1 apresenta uma breve introdução ao SPSS. No capítulo 2 apresenta uma introdução as técnicas de amostragem probabilística: amostragem aleatória simples, seleção do tamanho, tamanho amostral mínimo para a estimação populaciona, a amostragem aleatória estratificada, procedimentos e aplicações em psicologia.

Queremos ainda mostrar aos graduandos e pós-graduandos como incorparar os resultados das suas análises e como interpretar os resultados nos artigos. Tentamos simplificar conceitos complexos, e,

algumas vezes, bastantes complexos. Entretanto, ao facilitar existe uma perda de acurácia. Nos exercícios propostos após exemplos de cálculos ajudarão a resolver os problemas de intervalo de confiança. Os conjuntos de dados para exemplos e exercícios constam no próprio texto, de onde podem ser lidos para qualquer programa estatístico além do SPSS, como por exemplo, o programa R.

Sobre o autor

Edwirde Luiz Silva Camêlo é professor associado da Universidade Estadual da Paraíba, campus I em Campina Grande. Possui graduação em Matemática pela Universidade Rural de Pernambuco, mestrado em Biometria e doutorado (2007) em Estadística e Investigación Operativa pela Universidad de Granada (UGR) - España. Tem experiência na área de estatística univariada e multivariada aplicado a psicologia da saúde, experiência em Redes neurais artificias. Possui vários trabalhos publicados na área. Estou hebraico bíblico e hebraico moderno na UGR.

Capítulo 1

Uma breve introdução ao SPSS

Este guia fornece um conjunto de tutoriais para realizar uma análise útil do seu dados em psicologia. Você pode trabalhar com os tutoriais sequencialmente ou pular para os tópicos sobre os quais deseja informações adicionais. Este capítulo apresenta as funções básicas e mostra uma sessão típica. Considere o arquivo INTRODUCAO.sav de dados do IBM SPSS, vamos gerar um resumo estatístico simples e um gráfico.

Os capítulos a seguir incluirão muitos tópicos do programa SPSS. Esperamos poder fornecer a você uma estrutura básica para entender as principais ferramentas para auxiliar no conteúdo programático da disciplina Estatística e Psicologia.

O programa SPSS é um pacote estatístico, composto de diferentes módulos, desenvolvido também na área de psicologia. Está baseado no ambiente Windows, sendo de fácil operação e muito abrangente, pois permite realizar uma grande amplitude de análises

estatísticas e gráficas (análises descritivas de posição, dispersão e forma; teste de hipóteses, intervalos dde confiança, análises multivariadas, módulos gráficos, entre outras).

1.1 Janela do SPSS

Com o SPSS é possível criar, definir e modificar variáveis quantitativas e qualitativas; realizar cruzamentos de variáveis; gerar diversos tipos de gráficos; verificar a existências de associações o verificar a existências de correlações, etc.

O SPSS mostra a janela de digitação (ou input) de dados SPSS Data Editor, na qual os bancos de dados são gerados e analisados. Na janela do programa SPSS Data Editor as linhas são relativas aos indivíduos (casos), participantes ou respondentes e as colunas relativas as variáveis investigadas que podem ser quantitativa ou qualitativa.

Esta janela gera um dos três tipos de arquivos associados ao SPSS. Esse arquivo tem terminação .sav e armazena todas as informações relativas ao banco de dados, como definição de variáveis e os dados digitados.

A janela acima apresentada possui várias colunas relativas aos principais parâmetros de cada uma das variáveis da planilha. São 11 colunas:

1. Name

   Refere-se ao nome atribuído a variável, composto de até oito caracteres, que será colocado nas colunas da janela de input de dados. Deve-se clicar em qualquer cela dessa coluna para que se possa digitar o nome da variável;

2. Type

   Refere-se ao tipo de variável, ou seja, sua característica de notação (numérica: contínua ou discreta; qualitativa: nominal ou ordinal). Clicando-se na cela desta coluna abre-se a caixa de diálogo. Observe que a caixa de diálogo Variable Type permite a escolha de vários tipos de variáveis. A definição de cada uma pode ser acessada clicando-se no botão direito do mouse colocado em cima da opção. Esta mesma caixa de diálogo também permite a escolha de outras características da variável (width (comprimento) e número de casas decimais).

3. Width

   É o número de caracteres da variável nomeada. Pode ser definido diretamente na cela ou através da caixa de diálogo Variable Type.

4. Decimals

   É o número de casas decimais, a direita da vírgula, que serão apresentadas tanto para um número categorizado como para variáveis métricas.

5. Label

   Define o nome atribuído a uma variável e não possui restrição de número de caracteres (e.g., idade de um idoso estressado, escolaridade de uma pessoa ansiosa, descrição do item de um questionário para análise psicológica, entre outras);

6. Value

   São os valores que os labels podem assumir. Neste parâmetro o pesquisador deve definir todos os possíveis valores que uma

variável pode assumir. Um tipo comum em psicologia é a escala Likert entre outras em psicologia.

7. Missing(ausente)

   Define o tipo de tratamento para os indivíduos ausentes que o pesquisador deseja considerar. O Default do programa geralmente é usado, mas outros valores podem ser definidos como 999 ou 99.

8. Columns(colunas)

   Indicar o tamanho da coluna da variável, que será apresentada na janela de input de dados;

9. Align

   Define o alinhamento dentro de cada célula da planilha de dados (esquerda, centralizado ou direita).

10. Measure

    Aqui informa o tipo de variável (contínua ou discreta). Esta definição é fundamental, pois o SPSS habilitará o uso das variáveis em certos procedimentos a partir do tipo de medida (measure) selecionada.

A maioria dos exemplos fornecidos usa o arquivo de dados no formato .sav (do SPSS). Esses dados é um estudo fictício de 16 pessoas que contém informações básicas sobre idade (em anos) e nível de estresse variando de 1 a 10. O arquivo DadosIdadeEstresse.sav será uma amostra representativa do arquivo de dados original, reduzido considerando apenas 16 casos (indivíduos).

Na Tabela 1.1 mostra alguns dados para inserir no SPSS.

Tabela 1.1: Algumas técnicas estatística apropriada

Idades	Níveis de estresse	Idades	Níveis de estresse
21	4	23	4
45	9	15	6
32	8	33	7
54	6	44	6
21	4	25	4
35	8	31	7
32	9	71	8
51	7	19	5

Passos para inserir esses dados no SPSS:

1. Digitar os dados nas duas primeiras colunas no SPSS

   Antes de executar qualquer análise, precisa-se fornecer os dados ao SPSS. Observa-se a planilha, formada por células, é semelhante ao Excel, que são o encontro das linhas (indivíduos) e colunas (variáveis).

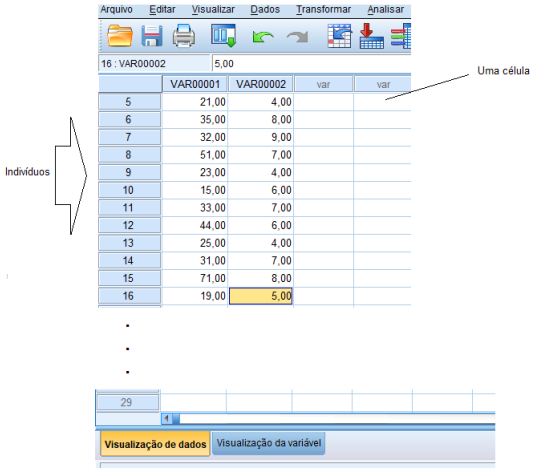


Figura 1.1: Inserindo as duas colunas no SPSS

2. Nomeando as colunas. Click em Visualização das variáveis como se observa na Figura 1.1 e visualizando as linhas.

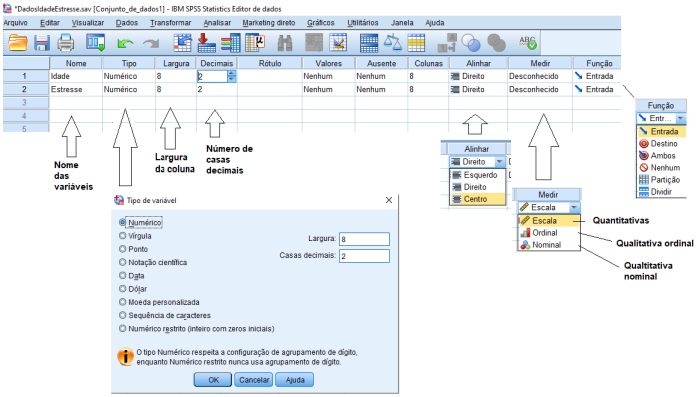


Figura 1.2: Nomeando variáveis

Na tela de visualização de variáveis (variable View), as colunas representam variáveis, e as linhas os indivíduos. Precisa-se fornecer o nome de cada variável. Click em Visualização de variáveis. Em outras palavras, o editor de dados tem dois painéis (views): o de dados e o de variáveis. Nas colunas de dados ficam por exemplo os dois grupos e a quantidade de clínicas em Campina Grande e João Pessoa. No editor de dados entram os dados obviamente dados e o editor de variáveis permite que definamos várias características das variáveis do editor de dados.

3. Estudando as funções: rótulos, valores e ausente. Considere mais uma variável que chamaremos de gênero:

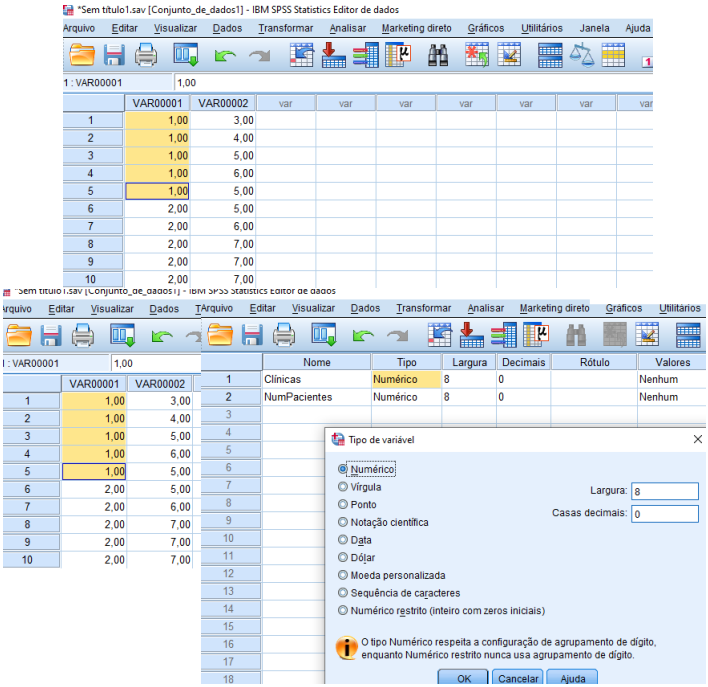


Figura 1.3: Número de decimais, tipos de variáveis, rótulos e valores

4. Valores

   Em Valores click em Nenhum, inserir 1 para Campina Grande 2 para João Pessoa.

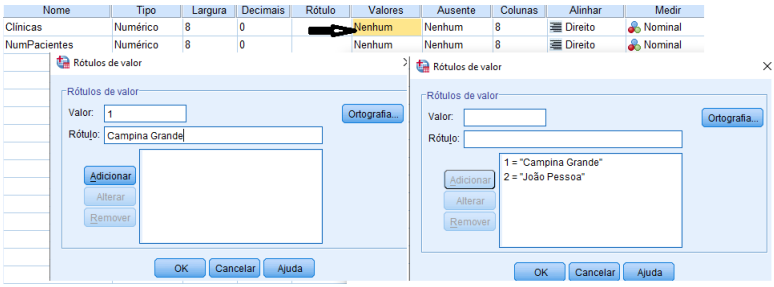


Figura 1.4: Nomes dos valores 1 para CG e 2 para JP

Tudo que temos que fazer é clicar com o mouse: Data View w Variable View. E criar variáveis codificadas. Uma variável codificadora é uma variável que consiste em uma série de números representada em níveis de uma variável de tratamento ou que descreve diferentes números de grupos, no caso, 1 para Campina Grande e 2 para João Pessoa.

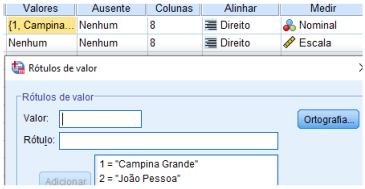


Figura 1.5: Nomes dos valores 1 para CG e 2 para JP

Finalmente, definimos as variáveis de códigos e seus valores no SPSS e os tipos de variáveis utilizadas, quantidade de decimais, alinhamento dos números e nomes dentro das células e a dimesão das colunas.

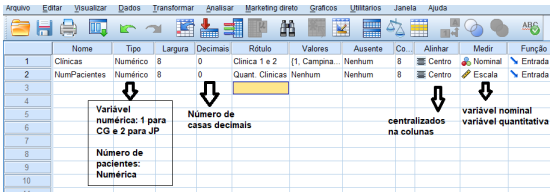


Figura 1.6: Visualizando as variáveis

5. Valores que faltam (missing)

   Muitas vezes ocorre que não temos dados que não podem, por algum causa, ser obtidos, dados que faltam ou são desconhecidos. As vezes, o intrevistado tem vergonha de falar, esquecer-se de responder a algumas questões, etc. Mais adiante veremos como preencher essa lacuna.

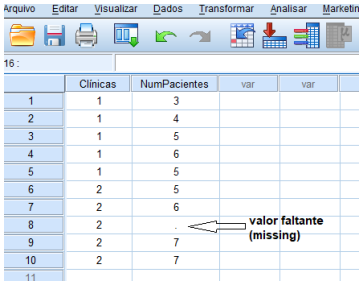


Figura 1.7: Elementos faltante na variável

1.2 Resumo de casos

Vamos iniciar as análise preliminares usando Resumo de Casos, como se observa na Figura 1.8.

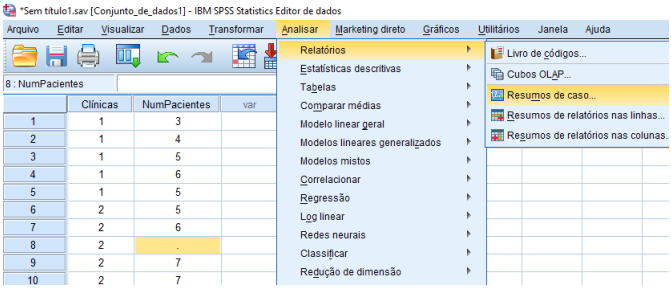


Figura 1.8: Resumos de casos no SPSS

Clicando em Ok, temos:

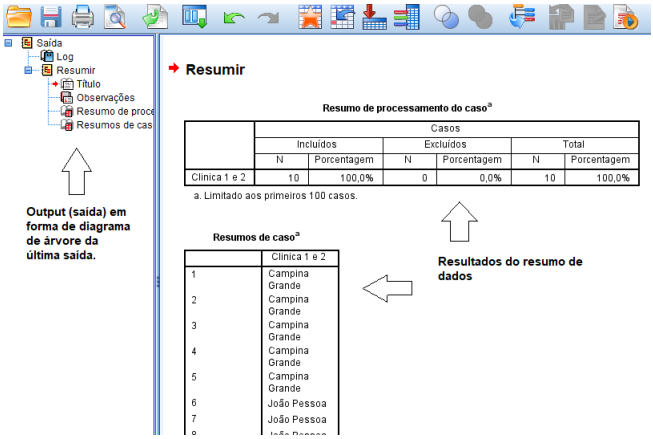


Figura 1.9: Resultado do resumos de casos

Observe que temos 100% dos dados válidos, nenhum caso foi excluído. Na última coluna temos Válido + excluído = total.

Este livro não tem a pretensão de substituir a aprendizagem séria e comprometida do pensamento estatístico. Muito pelo contrário, este manual é apenas uma leitura complementar para uso do SPSS. Portanto, foi concebido como uma ferramenta extra para

servir de apoio ao ensino de estatística, feito com, no mínimo, o uso de um livro texto de estatística voltado para cientistas sociais. Isto significa que o manual não pretende substituir a leitura de um bom livro introdutório ou de análise multivariada de dados, muito menos a realização de exercícios de aplicação de conhecimentos estatísticos, mas sim auxiliar o aluno na realização de procedimentos estatísticos via SPSS. Considera que o desenvolvimento destas competências é parte essencial da formação de pesquisa em estatística e psicologia.

1.3 Menus de comando

Na janela SPSS Data Editor o programa possui uma série de menus de comandos, que possibilitam a manipulação e análise dos dados, bem como os procedimentos do windows para trabalhar com arquivos.

O primeiro desses menus é denominado File e tem como principais funções abrir, salvar e importar, além de outras funções comuns aos diferentes programas da mesma plataforma operacional ou específicos do SPSS.

O segundo menu é o Editar, que possui as seguintes funções: voltar a última ação, copiar, colar, cortar, procurar (find) e opções. O menu opções define-se diferentes parâmetros no SPSS, como formato de apresentação dos dados digitados na SPSS Data Editor, o formato de geração das tabelas na janela de output (saída) de dados, entre outras funções.

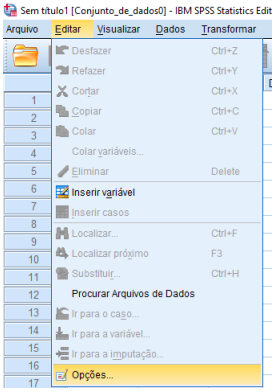


Figura 1.10: Menu de opções no SPSS

A Figura 1.11 apresenta a janela de diálogo do menu editar em opções. Este menu de opções é importante para adaptar o padrão de apresentação do programa às características ou preferências de trabalho do pesquisador. Idioma que o pesquisador deseja na saída do SPSS; o tipo de fonte das tabelas, configurações de gráficos, $\cdots$.

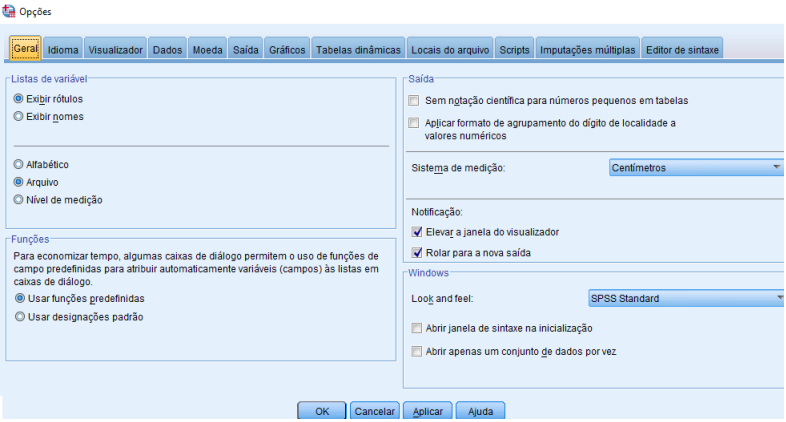


Figura 1.11: Geral no Menu de opções no SPSS

O próximo menu da barra de comandos é denominada Data. Esse menu possibilita manipular o arquivo de dados de diferentes maneiras. A Figura 1.3 apresenta a tela do SPSS com este menu aberto.

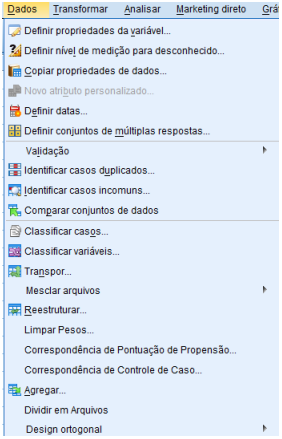


Figura 1.12: Menu de dados no SPSS

Vamos estudar alguns desses. Como pode ser observado na Figura 1.13 o menu Data possui vários comandos. Alguns serão

apresentados em tópicos a seguir. Para isso considere a seguinte amostra:

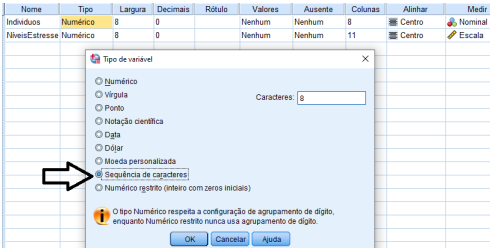
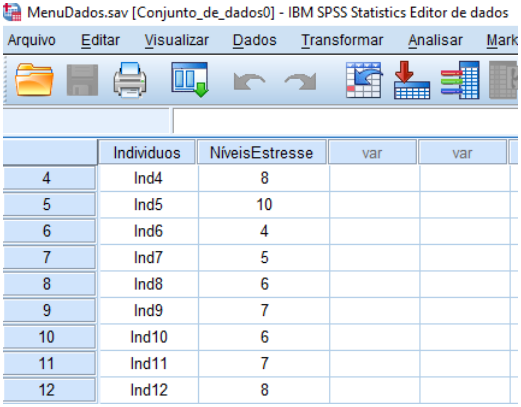


Figura 1.13: Tipos de variáveis no SPSS

Observe que em indivíduos o tipo de evariável considerada foi: Sequência de caracteres.



	Individuos	NíveisEstresse	var	var
4	Ind4	8		
5	Ind5	10		
6	Ind6	4		
7	Ind7	5		
8	Ind8	6		
9	Ind9	7		
10	Ind10	6		
11	Ind11	7		
12	Ind12	8		

Figura 1.14: Inserindo as variáveis indivíduos e níveis de estresse

O comando select cases possibilita a seleção de um grupo de casos em um mesmo arquivo ou a criação de um outro arquivo a partir de um grupo de casos inicial.

- Selecionando indivíduos

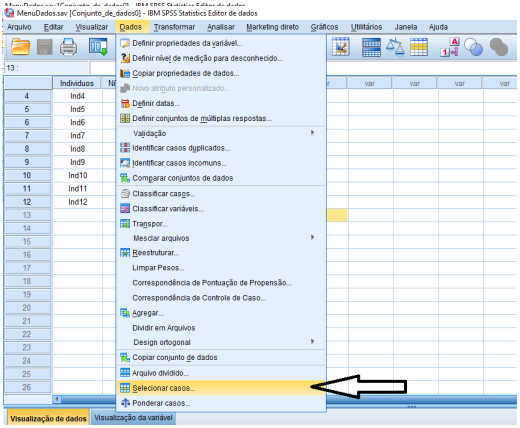


Figura 1.15: Seleção de casos

Seleciona casos: se a condição for cumprida. Continuar e Ok

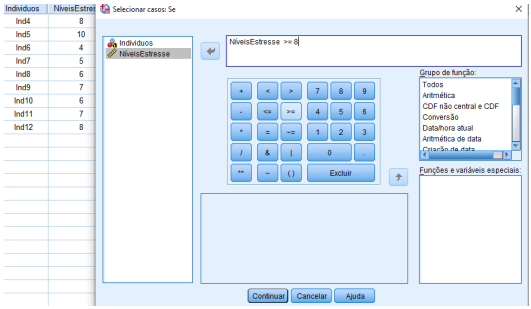


Figura 1.16: Seleção de casos para nível de estresse maior ou igual a 8

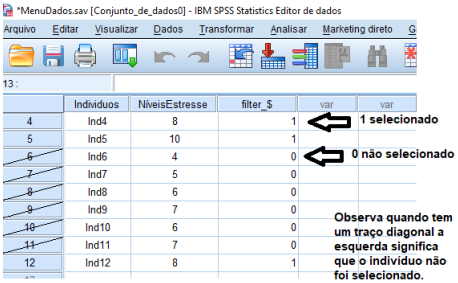


Figura 1.17: Os casos considerados

- Selecionando indivíduos aleatórios. Dados - Selecionar casos. Amostra aleatória de casos, como se observa na Figura 1.18. Clicar em continuar e ok.

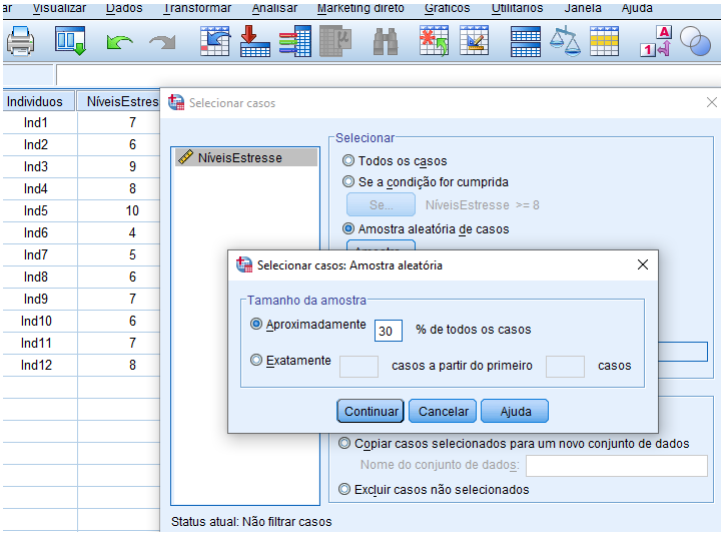


Figura 1.18: Seleção de apenas 30% dos casos

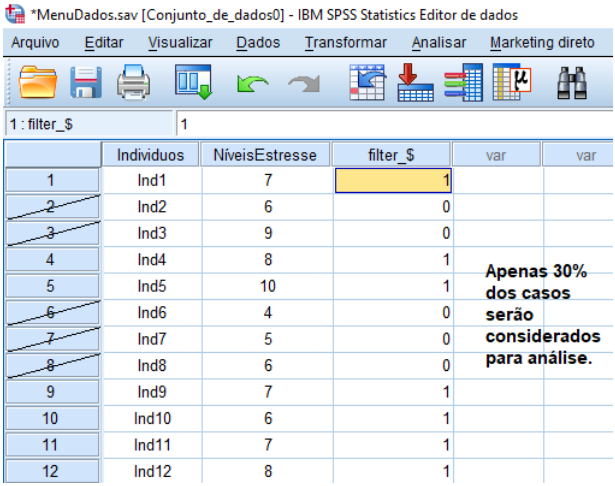


Figura 1.19: Seleção de apenas 30% dos casos

- Seleção considerando intervalo de casos. No caso tivemos uma mostra aleatória de 5 indivíduos e selecionou apenas 3 indivíduos aleatoriamente dos cinco.

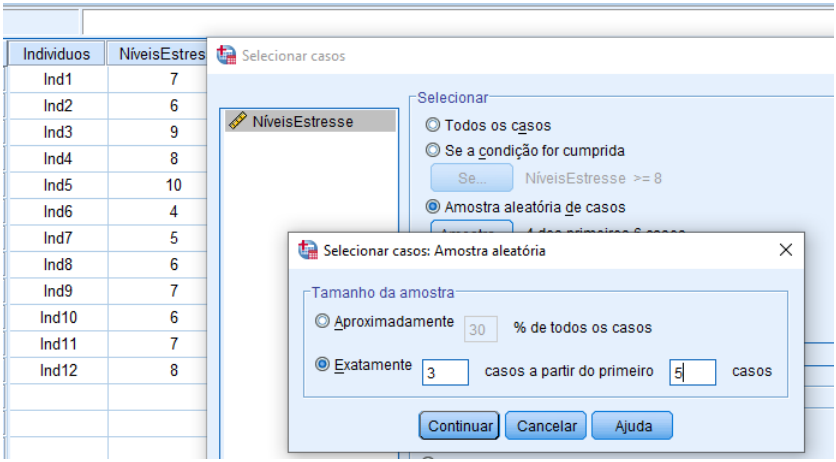


Figura 1.20: Seleção de apenas 3 de 5 indivíduos

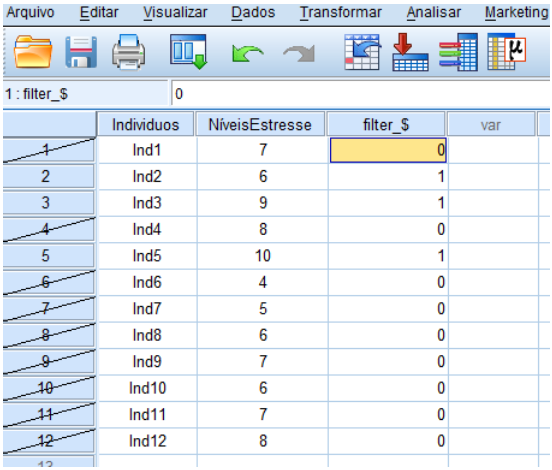


Figura 1.21: Seleção de apenas 3 de considerando 5 indivíduos

Capítulo 2

Amostragem

A representatividade de uma amostra permite extrapolar e, portanto, generalizar os resultados observados em isso, para a população acessível (conjunto de sujeitos que pertencem à população-alvo, que estão disponíveis para a investigação); e de lá, para a população, assim uma amostra será representativa ou não; apenas se fosse selecionado ao acaso, ou seja, que todos os sujeitos da população branca e acessível tivessem a mesma possibilidade ser selecionados nesta amostra e, portanto, ser incluído no estudo (técnica de amostragem probabilística); e por outro lado, o número de sujeitos selecionados representa numericamente a população que o originou em relação à da distribuição da variável em estudo na população, ou seja, a estimativa ou cálculo do tamanho da amostra **??**.

É assim que a análise de uma amostra permite fazer inferências, extrapolar ou generalizar conclusões para a população com um alto grau de certeza (Dieterich, 1996); de forma que uma amostra seja considerada representativa da população-alvo, quando a distribuição e o valor das diversas variáveis podem ser reproduzidos com

margens de erro calculáveis.

Uma amostra pode ser obtida de dois tipos: probabilística e não probabilística.

Definição 2.0.1. *Amostragem é um conjunto de técnicas estatísticos para o desenho, dimensionamento e seleção da amostra. A escolha do método mais conveniente para selecionar uma amostra é condicionada por vários fatores, como a existência ou não de uma base de sondagem completa e atualizada da população, grau de homogeneidade da população, dispersão territorial da população alvo, método de recolha da informação, tempo e custo disponível para a obtenção e tratamento dos dados da amostra e meios materiais disponíveis para o estudo.*

As principais vantagens de uma amostragem probabilísticos são as seguintes: não utilizam critérios subjetivos na escolha dos elementos da amostra, permitem avaliar a precisão das estimativas obtidas e determinar a priori a dimensão da amostra que garante a precisão pretendida para os resultados, pelo que são os esquemas de amostragem utilizados em estudos definitivos.

A desvantagens do uso dessa amostragem pode ser o custos elevados e são menos expeditos na obtenção da amostra do que os esquemas de amostragem empíricos. A dificuldade na obtenção de uma base de sondagem atualizada leva muitas vezes a optar por um esquema de amostragem empírico, onde os elementos da amostra são escolhidos segundo determinado critério de conveniência.

Exemplos de esquemas de amostragem probabilísticos: amostragem aleatória simples, amostragem aleatória estratificada), amostragem aleatória multi-etápica, amostragem aleatória por cachos e amostragem sistemática.

Definição 2.0.2. *O que é uma amostragem probabilística? Cada item ou pessoa na população estudada têm uma probabilidade (não nula) conhecida de ser incluída na amostra. Já as técnicas de amostragem não probabilística, a seleção de objeto de estudo dependerá de certas características, critérios, etc que o pesquisador considera naquele momento para que possam ser válidos e confiáveis ou reproduzíveis. Esses tipos de amostras não estão em conformidade com uma base probabilística, ou seja, não dão certeza de que cada sujeito em estudo representa a população.*

Uma amostra será ou não representativa, caso tenha sido selecionada ao acaso, ou seja, todos os sujeitos da população-alvo têm a mesma possibilidade de serem selecionados na amostra. A população acessível é o conjunto de indivíduos que pertencem à população-alvo e que estão disponíveis para investigação.

No processo de pesquisa científica em psiclogia da saúde, as diferentes etapas para um estudo estatístico são as seguintes:

2.1 Introdução as técnicas de amostragem probabilistica

2.1.1 Amostragem aleatória simples

Aleatório simples: garante que todos os indivíduos que compõem a população-alvo tem chances iguais de ser incluída na amostra. Isso significa que a probabilidade de seleção de um indivíduo a estudar "x" é independente da probabilidade de que os demais indivíduos que o compõem fazer parte da população-alvo.

Uma amostra escolhida de tal forma que cada item ou pessoa na população tem a mesma probabilidade de ser incluída.

Se a população tem um tamanho N, cada pessoa desta população tem a mesma probabilidade igual a $1/N$ de entrar na amostra. Utilizamos uma tabela de números aleatórios para sortear (com mesma probabilidade) os elementos da amostra. Também pode ser utilizada uma função randômica:

Exemplo 2.1.1. *Qual é a amostra necessária para estabelecer que os jovens numa escola estejam estressados na segunda feira? Uma amostragem aleatória simples se aplicaria da seguinte forma: entre todos os indivíduos estressados, selecione aleatoriamente um subgrupo que representam na Figura 2.1. Na amostragem aleatória simples o número de indivíduo (n) necessário para completar o estudo é selecionado aleatoriamente.*

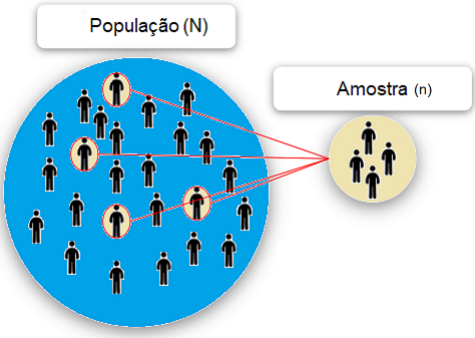


Figura 2.1: Amostra aleatória simples

Definição 2.1.1. *Amostragem aleatória simples é o esquema de amostragem probabilístico mais simples, adequado quando a população em estudo é bastante homogénea e a dispersão territorial é reduzida (especialmente se for necessário realizar entrevistas diretas ou presenciais para obter a informação). Segundo este esquema*

de amostragem todos os elementos da população têm a mesma probabilidade de serem escolhidos para integrar a amostra, e assim, todas as amostras possíveis de uma determinada dimensão dessa população têm a mesma hipótese de seleção. A probabilidade de um qualquer elemento da população pertencer à amostra é dada por nf N; sendo N a dimensão da população e n a dimensão da amostra. Para construir uma amostra aleatória simples de dimensão n a partir de uma população de dimensão N, procedemos do seguinte modo:

Uma amostra escolhida de tal forma que cada item ou pessoa na população tem a mesma probabilidade de ser incluída. Se a população tem um tamanho N, cada pessoa desta população tem a mesma probabilidade igual a 1/N de entrar na amostra. Utilizamos uma tabela de números aleatórios para sortear (com mesma probabilidade) os elementos da amostra. Também pode ser utilizada uma função randômica: No Excel, por exemplo, temos a função ALEATÓRIO ENTRE.

Passos para realizar uma amostra aleatória simples:

1. numeramos os elementos da população de 1 a N;

2. geramos, por exemplo em R, n números inteiros compreendidos entre 1 e N. Numa amostra aleatória simples sem reposição os números repetidos não servem, i.e., têm de ser substituidos por outros (pois o mesmo indivíduo da população não pode ser escolhido mais do que uma vez); numa amostra aleatória simples com reposição os números repetidos servem (o mesmo indivíduo da população pode ser escolhido mais do que uma vez). É de notar que na prática as amostras são quase sempre obtidas através de amostragem sem reposição,

mas em termos teóricos convém comparar os dois esquemas, pois em geral eles diferem pouco (atendendo a que a amostra é sempre de dimensão muito menor do que a dimensão da população) e os cálculos considerando amostragem com reposição são mais simples;

3. a amostra é constituída pelos elementos da população correspondentes aos números escolhidos.

Exemplo 2.1.2. *Qual é a amostra necessária para estabelecer que os jovens numa escola estejam estressados na segunda feira? Uma amostragem aleatória simples se aplicaria da seguinte forma: entre todos os indivíduos estressados, selecione aleatoriamente um subgrupo que representam.*

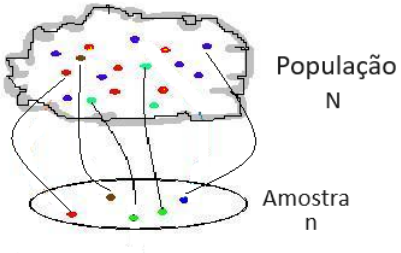


Figura 2.2: Exemplo de uma amostragem aleatória simples

Da população de estressados, o número de indivíduos necessários para completar o estudo é selecionado aleatoriamente.

2.1.2 Estimadores da média, população total e proporções

Seja uma amostra $A = u_1, u_2, \cdots, u_n$, formada por n unidades dentro de uma população finita de tamanho N obtida mediante um

procedimento de amostragem dado, e considere os valores $(x_1, x_2, \cdots, x_n)$ que toma a característica X em dita amostra.

Para estimar diversas características populacionais de dira variável X, como são a média (μ), o total populacional (τ) e a proporção (p), ou seja, respectivamente.

1. Média amostral:
$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

2. População total:
$$\tau = \sum_{i=1}^{n} \frac{x_i}{n/N} = N\bar{x}$$

Como a média populacional está relacionada com o total por $\tau/N = \mu$, a média amostral será um estimador não viesado da média da média popilacional. A média amostral será um estimador não viesado da média populacional.

Para conseguir isso é necessário a variância do estimador, $V(\bar{x})$; para uma amostra simples sem substituição de uma população de tamanho N.

$$V(\bar{x}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

Sabendo que a variância amostral é dado por:

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

Para encontrar $E(s^2)$ temos que:

$$
\begin{aligned}
E(s^2) &= E\left[\left(\frac{1}{n-1}\right)\sum_{i=1}^{n}(x_i-\bar{x})^2\right] \\
&\left(\frac{1}{n-1}\right)E\left\{\sum_{i=1}^{n}(x_i-\mu)-(\bar{x}-\mu)^2]\right\} \\
&= \left(\frac{1}{n-1}\right)E\left[\sum_{i=1}^{n}(x_i-\mu)^2-n(\bar{x}-\mu)^2\right] \\
&= \frac{1}{n-1}\left[\sum_{i=1}^{n}E(x_i-\mu^2)-nE(\bar{x}-\mu)^2\right] \\
&= \frac{1}{n-1}[n\sigma^2-nV(\bar{x})] \\
&= \frac{1}{n-1}\left[n\sigma^2-n\left(\frac{N-n}{N}\right)\left(\frac{N}{N-1}\right)\right] \\
&= \frac{\sigma^2}{n-1}\left(n-\frac{N-n}{N-1}\right) \\
&= \frac{N}{N-1}\sigma^2
\end{aligned}
$$

Em efeito, a partir da fórmula acima pelas propriedades do valor esperado e substituindo $E[s^2]$ pelo seu valor obtido acima, tem-se:

$$E\left[\frac{N-1}{N}s^2\right] = \frac{N-1}{N}E[s^2] = \frac{N-1}{N}\sigma^2\frac{N}{N-1} = \sigma^2$$

Portanto,

$$
\begin{aligned}
E\left[\left(\frac{N-n}{N}\right)\left(\frac{s^2}{n}\right)\right] = \left(\frac{N-n}{N}\right)\left(\frac{1}{n}\right)\left(\frac{N}{N-1}\sigma^2\right) = \\
\left(\frac{N-n}{N-1}\right)\left(\frac{\sigma^2}{n}\right) = V(\bar{x}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)
\end{aligned}
$$

O paramétrico valor da variância populacional da média amostral será:

$$V_{\bar{x}} = \frac{s^2}{n}\frac{N-n}{N}$$

O erro padrão do paramétrico valor será $V_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$, e o erro padrão do estimador baseado na amostra será: $S_{\bar{x}} = \frac{S_x}{\sqrt{n}}$.

O limite do erro de estimação será:

$$2\sqrt{\hat{V}(\bar{x})} = 2\sqrt{\frac{s^2}{n}\left(\frac{N-n}{N}\right)}$$

A quantidade $(N-n)/N$ é denominada de fator de correlação de população finaita (fcpf).

Exemplo 2.1.3. *Considere uma única amostra de n =10 da população (N=30) de pessoas com ansiedade (com nível de ansiedade variando de 1 a 12). Suponha que os seguintes elementos foram selecionados aleatoriamente. Os 10 indivíduos selecionados foram:*

8 6 10 6 10 9 12 5 5 9

- Média amostral: $\bar{x} = 8$
- Variância amostral $V(\bar{x}) = s^2 = \hat{\sigma}^2 = 5,77$
- Estimativa do total: O estimador do total é $\tau = N\bar{x} = 240$
- Estimativa da variância do total estimado: $V(\tau) = N^2 V(\bar{x}) = 30^2.0,38 = 346,,66$

  Contudo, podemos fazer uma afirmação probabilística sobre o intervalo em que esperamos que o valor verdadeiro esteja; que é o intervalo de confiança. Para a média estimada, o

intervalo de confiança é calculado a partir do erro padrão estimado e de uma suposição sobre a distribuição das médias amostrais que se reflete no valor da distribuição t. Aceitando uma probabilidade de erro de $\alpha = 5\%$ de que nossa afirmação está errada, a largura de um lado do intervalo de confiança é:

- Intervalo de confiança: $t_{(\alpha;gl=n-1)} = 0.859$.

  Assim, o I.C. será:

$$I.C. = [8 - 0,859; 8 + 0,859] = [7,14; 8,86]$$

  Isso diz: a probabilidade de que a verdadeira média paramétrica esteja no intervalo entre 7,14 e8,86 é 0,95; pode, contudo, acontecer que a verdadeira média paramétrica seja menor ou maior; esta é a probabilidade de erro $\alpha = 5\%$. O valor t de Student fornecido pode ser lido em tabelas ou calculado a partir de funções que normalmente todo software estatístico incorpora. Os valores reais dependem dos graus de liberdade e da probabilidade de erro escolhida.

Exemplo 2.1.4. *Suponha-se que seleciona uma amostra aleatória n = 200 (cada amostra tem a mesma probabilidade de ser escolhida) do total de N = 1000. A média amostral foi,* $\bar{x} = 94,2$ *e a variância amostral,* $s^2 = 445,2$. *Estime* μ *e estabeleça um limite para o erro de estimação.*

Utiliza-se $\bar{x} = 94,2$ para estimar μ. O limite para o erro de estimação será:

$$
\begin{aligned}
2\sqrt{\hat{V}(\bar{x})} &= 2\sqrt{\frac{s^2}{n}\left(\frac{N-n}{N}\right)} \\
&= 2\sqrt{\frac{445,2}{200}\left(\frac{1000-200}{1000}\right)} \\
&= 2,67
\end{aligned}
$$

Portanto, como n é grande, a média amostral terá uma distribuição aproximadamente normal, pelo que $94,2 \pm 2,67$ é um intervalo de confiança para a média populacional de aproximadamente 95%.

2.2 Seleção do tamanho da amostra para estimar as médias e totais da população

O número de observações necessária para estimar uma média populacional μ com um limite para o erro de estimação de magnitude (M) encontra-se ao estabelecer dois desvios padrões do estimador, $\bar{x}$, igual a M, resolvendo para n. Ou seja, deve-se resolver: $2\sqrt{V(\bar{x})} = M$ para n. Lembre-se que a variância estimada de $\bar{x}$, $V(\bar{x})$, foram dadas por:

$$
\hat{V}(\bar{x}) = \frac{s^2}{n}\left(\frac{N-n}{N}\right) \qquad \hat{V}(\mu) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)
$$

O tamanho amostra necessário agora pode se encontrar deixando n em evidência da seguinte equação:

$$
2\sqrt{V(\bar{x})} = 2\sqrt{\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)}
$$

Assim, o tamanho necessário para estimar μ como limite para o erro de estimação M será:

$$n = \frac{N\sigma^2}{(N-1)\left(\frac{M}{2}\right)^2 + \sigma^2}$$

Exemplo 2.2.1. *Deseja-se estimar a quantidade de jovens que tiveram depressão no período da Covid 19. Ainda não se considerar dados anteriores para estimar a variância populacional. Sabe-se que a maioria tiveram depressão causada pelo isolamento, ou seja, um amplitude de variação de 100. Considere que a população é de N = 1000 jovens. Encontrar o tamanho da amostra necessária para estimar μ com um limite para o erro de estimação de M = 3.*

É necessário uma estimação para σ^2. Como a amplitude da variação é aproximadamente igual a 4 desvio padrão (4σ), um quarto de tal amplitude proporcionará um valor aproximado de σ. Portanto,

$$\frac{Amplitude}{4} = \frac{100}{25} = 4 \qquad \sigma \approx 25^2 = 625$$

$$n = \frac{1000.625}{(1000-1)\left(\frac{3}{2}\right)^2 + 625} = 217,5$$

Assim, são necessários aproximadamente 217 observações para estimar μ, a média de jovens que tiveram depressão após Covid, com um limite para o erro de estimação de 3.

De igual maneira, pode-se determinar o número de observações necessárias para estimar um total populacional τ, com um limite de erro de magnitude M.

O tamanho amostral necessário se encontra ao estabelecer que M é igual a dois desvios padrões da variância do estimador será:

$$2\sqrt{N^2V(\bar{x})} = M \qquad 2N\sqrt{V(\bar{x})} = M$$

O tamanho da amostra necessária para estimar τ com um limite para o erro de M será:

$$n \geq \frac{N\sigma^2}{(N-1)\left(\frac{M}{2}\right)^2 + \sigma^2}$$

Exemplo 2.2.2. *Deseja-se conhecer, com uma cota de erro de 12 indivíduos, a quantidade média de pessoas com depressão . Qual o tamanho amostral de pessoas com depressão? Supoha que* $\sigma^2 = 60^2$*, o tamanho populacional seja N = 6500 pessoas e M = 12 (a cota de erro).*

O número da amostra será:

$$n \geq \frac{6500.60^2}{(6500-1)\left(\frac{12}{2}\right)^2 + 60^2} = 98,49$$

```
S2<-60^2
N<-6500
M<- 12
n <- (S2*N)/(S2 + (N-1)*(M^2/4)); n
```

Para que o erro de estimação da média populacional, com uma confiança de 95% não supere a quantidade de 12 pessoas, o tamnaho amostral será no mínimo 99 pessoas com depressão .

Exercício 2.2.1.

Exemplo 2.2.3. *Deseja-se conhecer, com uma cota de erro de 9 indivíduos, a quantidade média de pessoas com ansiedade. Qual*

o tamanho amostral de pessoas que tem ansiedade? Supoha que $\sigma^2 = 50^2$*, o tamanho populacional seja N = 5770 pessoas e M = 9 (a cota de erro).*

2.3 Tamanho amostral mínimo para que a estimação da população total

O tamanho amostral mínimo para que a estimação da população total não supere uma cota de erro M, com probabilidade $1-\alpha$ fixada antecipadamente. Isso se procede de forma análoga no caso da média populacional. O propósito é que M seja o erro máximo que se cometa al estimar a população total mediante $\hat{\tau}$. Sabe-se que o erro é o dobro da raiz quadrada da variância do estimador, ou seja:

$$M \geq 2\sqrt{V[\hat{\tau}]}$$

Aos valores que toma a variável em cada um dos elementos da população é deminado por $u_1, u_2, \cdots, n_N$. O estimador e sua variância serão:

$$\hat{\tau} = N\frac{\sum_{i=1}^{n} X_i}{n} \qquad V(\hat{\tau}) = N^2\frac{\sigma^2}{n}\frac{N-n}{N-1}$$

Quando σ^2 é desconhecido, usa-se a seguinte fórmula:

$$V(\hat{\tau}) = N^2\frac{s^2}{n}\frac{N-n}{N}$$

A cota do erro de estimação será:

$$M = 2\sqrt{V(\hat{\tau})} = 2\sqrt{N^2V(\bar{X})} = N2\sqrt{V(\bar{X})}$$

Assi, temos que:

$$M \geq \sqrt{N^2 \frac{\sigma^2}{n} \frac{N-n}{N-1}}$$

Colocando N em evidência e fora da raiz quadrada temos:

$$\frac{M}{2N} \geq \sqrt{\frac{\sigma^2}{n} \frac{N-n}{N-1}}$$

Elevando ambos os membros ao quadrado temos:

$$\frac{M^2}{4N^2} \geq \frac{\sigma^2}{n} \frac{N-n}{N-1}$$

Deixando fixo $\frac{M^2}{4N^2}$ e colocando n em evidência temos:

$$\left(\sigma^2 + (N-1)\frac{M^2}{4N^2}\right) n \geq \sigma^2 N$$

Sadendo que $\sigma^2 + (N-1)\frac{M^2}{4N^2}$ é positivo, pode-se dividir ambos membros pela dita desigualdade, então:

$$n \geq \frac{N\sigma^2}{(N-1)\frac{M^2}{4N^2} + \sigma^2}$$

Exemplo 2.3.1. *Suponha que o erro de estimação do total da população de grande hospital não ultrapasse 100, com um nível de confiança de 95%. Qual o tamanho mínimo da amostra? Dado* $\sigma^2 = 17$*, N = 800 e M = 200.*

$$n \geq \frac{800\sigma^2}{(800-1)\frac{200^2}{4.800^2} + 17} = 461,26$$

```
S2<-17
N<-800
```

```
M <- 200
n <- (S2*N)/(S2 + (N-1)*(M^2/(4*N^2))); n
```

A amostra deve ter no mínimo 462 indivíduos.

Exercício 2.3.1. *Suponha que o erro de estimação do total da população de grande hospital não ultrapasse 80, com um nível de confiança de 95%. Qual o tamanho mínimo da amostra? Dado* $\sigma^2 = 12$*,* $N = 400$ *e* $M = 80$*.*

2.4 Tamanho amostral mínimo para que a estimação da proporção

Encontraremos o tamanho amostral mínimo para que a estimação da proporção populacional não supere uma cota de erro M, fixada antecipadamente.

Da mesma forma que o caso anterior, M tem que ser maior ou igual ao dobro da raiaz quadrada da variância do estimador.

$$M \geq 2\sqrt{V[\hat{p}]}$$

Sabendo que:

$$V(\hat{p}) = \frac{p(1-q)}{n}\frac{N-n}{N-1}$$

A variável discreta do tipo Bernouilli toma apenas dois valores: um (1) se o elemento i-ésimo possui a caracterísrica objeto de estudo ou zero (0) se não possui.

$$E[X_i] = p \qquad V[X_i] = p(1-p) \quad \forall i$$

A proporção amostral, definida como média amostral das variáveis que formam a amostra tem a seguinte distribuição:

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim Binomial(1, p)$$

É um estimador enviesado da proporção populacional.

A esperança matemática coincide com o parâmetro p: $E[X_i] = p, \forall i$, pois,

$$E[\hat{p}] = \frac{1}{n}\sum_{i=1}^{n} p = \frac{1}{n}np = p$$

Já a variância do estimador proporção amostral será:

$$V(\hat{p}) = \frac{p(p-1)}{n}\frac{N-n}{N} - 1$$

Sabe-se por inferência que n pela variância amostral é igual a (n-1) é:

$$n\hat{p}(1-\hat{p}) = (n-1)s^2$$

Resulta que:

$$s^2 = \frac{n\hat{p}(1-\hat{p})}{n-1}$$

Sabendo que:

$$\hat{V}(\bar{X}) = \frac{(N-1)s^2}{Nn}\frac{N-n}{N-1}$$

Então,

$$\hat{V}(\hat{p}) = \frac{\hat{p}(1-\hat{p})}{n-1}\frac{N-n}{N}$$

Exemplo 2.4.1. *Seleconando de maneira aleatória, 100 dos 1100 jovens com crise de ansiedade que uma grande cidade, 81 deles estavam de acordo com um período de descanso no trabalho por causa da crise de ansiedade. Estime a proporção de jovens com crise de ansiedade que estão de acordo com o período e estabeleça o limite para o erro de estimação.*

```
N<-1100
n<-100
p<-81/100
Vp<-(p*(1-p))*(N-n)/((n-1)*N)
Vp
B<- 2*sqrt(Vp);B
LImErr<-c(p-B, p+B); LImErr
```

O estimador pontual da proporção populacional é a proporção amostral: $\hat{p} = 81/100 = 0,81$

A correspondente cota de erro vale:

$$\hat{V}(\hat{p}) = \frac{0,81(1-0,81)}{100-1}\frac{1100-100}{1100} = 0,0014$$

A cota máxima de erro de estimação, como sempre, igual ao dobro da raiz quadrada da variância do estimador:

$$M = 2\sqrt{V(\hat{\hat{p}})} = 2\sqrt{0,0014} = 0,0751$$

Supõe-se que a proporção populacional está compreendida entre:

$$I.C. = [0,0751 - M; 0,0751 + M] = [0,7348; 0,8851]$$

Como mínimo o 73,48% dos jovens e no máximo 88,51% concordam com descanso. Também é dito que 11,49% (100-88,51) dos jovens não se manisfestaram e no máximo 26,52% (100-73,48) dos jovens deixaram de manifestar que concordam.

Exercício 2.4.1. *Seleconando de maneira aleatória, 100 dos 1350 jovens com depressão que uma grande cidade, 84 deles estavam de acordo com um período de descanso no trabalho por causa da depressão. Estime a proporção de jovens com depressão que estão de acordo com o período e estabeleça o limite para o erro de estimação.*

2.4.1 Tamanho amostral mínimo pra que a estimação da proporção populacional não supere a cota de erro, fixado anteriormente

Como vimos M tem que ser maior ou igual ao dobro da raiz quadrada da variância do estimador:

$$M \geq 2\sqrt{V[\hat{p}]}$$

Substituindo a variância do estimador temos:

$$\frac{M^2}{4} \geq \frac{p(1-p)}{n}\frac{N-n}{N-1}$$

Deixando o termino $M^2/4$ juntos, com os mesmos passos dos casos anteriores, temos:

$$n(N-1)\frac{M^2}{4} \geq p(1-p)N - p(1-p)n$$

Passando $p(1-p)n$ para primiero membro e colocando n em evidência, temos:

$$\left(p(1-p)+(N-1)\frac{M^2}{4}\right) \geq p(1-p)N$$

Como $p(1-p)+(N-1)\frac{M^2}{4}$ é positivo, pode passar dividindo mudando o segundo membro sem alterar o sentido da desigualdade.

$$n \geq \frac{p(1-p)N}{p(1-p)+(N-1)\frac{M^2}{4}}$$

Exemplo 2.4.2. *A cota do erro de estimação foi M = 7,51%. Para estimar a proporção de jovens que está de acordo com um período de descanso no trabalho por causa da crise de ansiedade não supere 3%. Qual deve ser o tamanho da amostra?*

Posto que se dispõe de uma amostra previa, pode-se usar essa informação para encontrar o tamanho da amostra. Temos: p = 0,81, N = 1100 e M = 3%. Substituindo na fórmula temos:

$$n \geq \frac{p(1-p)N}{p(1-p)+(N-1)\frac{M^2}{4}} = \frac{0,81(0,19)N}{0,81(0,19)+(1100-1)\frac{0,03^2}{4}} = 421,98$$

No R seria:

```
p<-0.81
N<-1100
B<-0.03
n<-(p*(1-p)*N)/(p*(1-p) + (N-1)*(B^2/4));n
```

O tamano da amostra necessário para que o erro de estimação não supere 3% será n = 422 jovens que esteja de acordo.

Exemplo 2.4.3. *Para entrevistar uma população de 2000 jovens é muito trabalhoso. Determinar o tamanho da amostra necessário*

para estimar p com um limite de erro de estimação de magnitude M = 5/100. Suponha que não existe informação disponível para estimar p.

Quando não se dispõe de informação previa, podem-se aproximar os tamanhos da amostra necessária, estabelecendo p = 0,05. Assim, temos:

$$\left(\frac{M}{2}\right)^2 = \frac{0,05}{4} = \frac{0,0025}{4} = 0,00062$$

$$n = \frac{2000.0,5(1-0,5)}{(2000-1)(0,000062)^2 + 0,5(1-0,5)} = 333,6$$

Ou seja, é necessário entrevistar 333 jovens para estimar a proporção de estudante, com um limite de estimação M.

Exercício 2.4.2. *A cota do erro de estimação foi M = 7,51%. Para estimar a proporção de jovens que está de acordo com um período de descanso no trabalho por causa da crise de ansiedade não supere 1%. Qual deve ser o tamanho da amostra?*

Exercício 2.4.3. *A cota do erro de estimação foi M = 7,51%. Para estimar a proporção de jovens que está de acordo com um período de descanso no trabalho por causa da crise de ansiedade não supere 5%. Qual deve ser o tamanho da amostra?*

Exercício 2.4.4. *Para entrevistar uma população de 1890 jovens é muito trabalhoso. Determinar o tamanho da amostra necessário para estimar p com um limite de erro de estimação de magnitude M = 5/100. Suponha que não existe informação disponível para estimar p.*

2.4.2 Estimação do error amostral para uma amostra aleatória simples: Teorema de Chebyshev

Os estimadores da média, o total, a proporção e a variância populacional que se tem visto, são estimadoraes que cumprir as propriedades desejadas de um estimador pontual para estimar um valor de um parâmetro. Entretanto, na prática em psicologia, o mais usual é obter um posto de possíveis valores entre os que se encontrar o verdadeiro valor do parâmetro, ou seja, estimar intervalos de confiança.

$$P(\aleph < \Theta < \beth) = 1 - \alpha$$

($\aleph$ lê-se álef e $\beth$, lê-se beit).

Em que $\aleph$ e $\beth$ são os respectivo estimadores extremos do intervalo, $\aleph$ é o parâmetro a estimar e α é o nível de significação. O teorema de Chebyshev permite obter a probabilidade de que uma variável aleatória tome determinados valores dentro de um intervalo de amplitude c vezes o desvio padrão σ.

$$P[\bar{x} - c\sigma < X < \bar{x} = c\sigma] \geq 1 - \frac{1}{k^2}$$

Exemplo 2.4.4. *Dada a seguinte informação sobre a quantidade de pacientes estressados em 10 bairros em pequena cidade. Pergunta-se: Qual o intervalo contém a quantidade de estressados com uma probabilidade de 75%?*

Cidades	1	2	3	4	5	6	7	8	9	10
Estressados	20	21	18	14	20	19	17	18	16	14

Segundo o teorema de Chebyshev, para uma probabilidade de 75% se cumprir:

$$0,75 = 1 - \frac{1}{c^2} \Rightarrow c^2 = 4 \Rightarrow c = 2$$

A média e o desvio padrão dos valores observados na tabela vale:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{10} = 177/10 = 17,0$$

e o desvio padrão

$$\hat{s}^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})}{n-1} = \frac{54,1}{9} = 6 \Rightarrow \hat{s}^2 = \sqrt{6} = 2,45$$

Aplicando o teorema de Chebysshev:

$$P[17,7 - 2(2,45) < X < 17,7 + 2(2,45)] \geq 0,75$$

$$P[12,8 < X < 22,6)] \geq 0,75$$

Ao menos 75% dos estressados estarão entre os 12,8 e 22,6. Quando a proporção é normal, o intervalo formado por duas vezes o desvio padrão contém em torno 95% dos dados.

Exercício 2.4.5. *Dada a seguinte informação sobre a quantidade de pacientes com ansiedade em 12 cidades no interior de Pernambuco. Pergunta-se: Qual o intervalo contém a quantidade de ansiosos com uma probabilidade de 85%?*

Cidades	1	2	3	4	5	6	7	8	9	10	11	12
Ansiosos	18	22	17	15	18	17	19	16	18	12	12	13

Exercício 2.4.6. *Considere a quantidade de jovens que fazem terapia de resolução de problemas em 8 clínicas. Pergunta-se: Qual*

o intervalo contém a quantidade desses jovens com uma probabilidade de 80%?

Clinicas	A	B	C	D	E	F	G	H
Quantidades	10	22	15	15	20	16	15	18

Exemplo 2.4.5. *Para uma população de clínicas psicológicas deseja-se estimar a quantidade de pessoas com crise de ansiedade em 24 grandes capitais ou municípios. Se a clínica localiza na capital o valor será (1), se no interior (0).*

Nomes	Ansiosos	Localidade	Nomes	Ansiosos	Localidade
C1	33056	1	C13	23740	0
C2	27046	1	C14	15686	1
C3	34261	0	C15	30231	1
C4	30051	1	C16	7212	1
C5	33657	1	C17	36061	1
C6	37864	0	C18	48081	1
C7	21035	1	C19	6010	0
C8	25243	1	C20	6010	0
C9	37563	1	C21	6010	0
C10	15025	0	C22	13823	0
C11	21396	0	C23	6010	1
C12	15266	0	C24	7212	1

Pede-se:

1. Qual é o valor médio e total da quantidade de pessoas com crise de ansiedade?

2. Qual o tamanho deve ser a amostra e que custo teria para estimar o número de pessoas com ansiedade e que o erro devido a amostra não seja superior a 5%.

3. Qual a proporção de clínicas na capital (1) ou interior (0)? Qual o tamanho amostral deve ter a mostra para que qo estimar a proporção do erro devido a amostra não seja superior a 5%?

1)

$$\bar{x} = \frac{\sum_{i=1}^{24} x_i}{10} = 537549/24 = 22397,87$$

e o desvio padrão

$$\hat{s}^2 = \frac{\sum_{i=1}^{24}(x_i - \bar{x})}{n-1} = \frac{3610145707}{24-1} = 156962857,00 \Rightarrow \sqrt{\hat{s}^2} = 12528,48$$

.

$$V(\bar{x}) = \frac{\hat{s}^2}{n}\left(\frac{N-n}{N}\right) = \frac{156962857}{24}\left(\frac{15000-24}{15000}\right) = 6529655$$

Cabe resaltar que o fator de correção para população finita seria:

$$\left(\frac{N-n}{N}\right) \geq 0,95$$

.

Para calcular que a quantidade de pessoas com crise de ansiedade, basta multiplicar o valor médio da quantidade:

$$\hat{x} = N\bar{x} = 15000(22398) = 335968125$$

.

2) Supondo normalidade, e dado que a amostra é pequena, o erro amostral é de 5286,95, o que se obtem a seguinte expressão:

$$E = t_{5\%/2;23}\sqrt{V(\bar{x})} = 2,069\sqrt{}6529655 = 5268,04$$

. Que, em termos relativo de média, dito erro é:

$$E(\%) = \frac{E}{\bar{x}}.100 = \frac{5286,04}{22398}(100) = 23,60\%$$

Esse erro de 23,6% é muito elevado e no estudo exige-se que o dito erro não supere a 5% do valor do estimador. Multiplicandso o valor do estimador (média amostral) pela porcentagem do erro permitido, e no caso de 5%, obtem-se que o valor do erro máximo permitido para a média é: E(5%) = 0,05.22398 = 1119,90.

$$n = \frac{N\sigma^2}{\frac{NE^2}{Z^2_{\alpha/2}} + \sigma^2} = \frac{15000(156962857)}{\frac{15000(1119,90)^2}{2,069}^2} + 156962857 = 517,19$$

$Assim, ocustoser = 517.15 = 7755,25.$

3) Para estimar a proporção da quantidade de ansiosos calculamos a expressão:

A continuação, apenas tem que substituir os respectivos valores na expressão do tamanho da amostra para um erro desejado e estimar o tamanho apropriado:

$$\hat{p} = \frac{\sum_{i=1}^{n} a_i}{N} = \frac{\sum_{i=1}^{24} a_i}{24} = \frac{13}{24} = 0,54$$

É dizer, estima-se que o valor 54,2% quantidade de pessoas com crise de ansiedade. Para encontrar o erro amostral dessa estimação precisa-se calcular a variância que se obtem com a seguinte forma:

$$var(\hat{p}) = \frac{\hat{p}\hat{q}}{n}\left(\frac{N-n}{N}\right) = \frac{0,54.0,46}{24}\left(\frac{15000-24}{15000}\right) = 0,0103$$

Sabendo que $V(\hat{p}) = 13/24 = 0,54$, pois $p = 13$ sucessos (1) e $\hat{q} = 1 - 0,54 = 0,46$, pois $q = 11$ fracassos (0). Assim, o erro amostral é igual a:

$$E = Z_{0,02/2}\sqrt{V(\hat{p})} = 1,96\sqrt{0,0103} = 0,20$$

Dado que as proposições vem expressa em termos relativos, multiplicando por 100 tem-se o erro em porcentagem:

$$E(\%) = E.100 = 0,20.100 = 20\%$$

Mas, a pergunta especifica que o erro não deve ser superior a 5% (E=0,05). Assim,tem que determinar o tamanho amostral apropriado para dito error. Para ele, encontra-se usando a expressão:

$$n = \frac{N\hat{p}\hat{q}}{\frac{NE^2}{Z^2_{\alpha/2}} + \hat{p}\hat{q}} = \frac{15000(0,54)(0,46)}{\frac{15000.(0,05)^2}{(1,96)^2} + (0,54)(0,46)} = 372,02$$

Resume-se, para estimar que proporção de quantidade de pessoas com crise de ansiedade com um erro devido a amostra que não seja superior a 5%, são necessários 372 pessoas com crise de ansiedade aproximadamente.

Exercício 2.4.7. *Suponha que tenha uma população de 15000 indi-*

víduos com ansiedade. Pede-se estimar a quantidade de indivíduos com ansiedade em 24 grandes municípios (M). Sabendo também que (1) é capital e (0) a zona rural.

Nomes	Ansiosos	Localidade	Nomes	Ansiosos	Localidade
M1	32054	1	M13	24731	0
M2	17026	1	M14	14681	1
M3	44320	0	M15	34237	1
M4	40091	1	M16	7111	1
M5	32659	1	M17	56031	1
M6	35814	0	M18	38091	1
M7	27031	1	M19	8017	0
M8	31240	1	M20	4010	0
M9	40573	1	M21	5010	0
M10	17045	0	M22	23823	0
M11	41391	0	M23	6710	1
M12	17261	0	M24	8212	1

Pede-se:

1. O valor médio e total da quantidade de pessoas com ansiedade?

2. O tamanho deve ser a amostra e que custo teria para estimar o número de pessoas com ansiedade e que o erro amostral não seja superior a 10%.

3. A proporção de clínicas na capital (1) ou interior (0)? Qual o tamanho amostral deve ter a mostra para que qo estimar a proporção do erro devido a amostra não seja superior a 5%?

2.4.3 Amostra aleatória simples no SPSS

Esta é uma amostra feita a partir de populações definidas em termos de unidades simples com probabilidade semelhante de ser selecionado. Para fazer isso, você pode seguir a seguinte sequência de menus e sub menus: Dados- Selecionar casos- Amostra caso aleatório. Assim que esta última opção for marcada, o botão deve ser selecionado imediatamente abaixo, que em algumas versões do SPSS vem como um exemplo (que é uma má tradução do termo em inglês Sample). Em seguida, uma janela irá aparecer, conforme mostrado na Figura 2.3, onde duas opções são oferecidas para a escolha da amostra. O primeiro é selecionar aleatoriamente uma porcentagem aproximada do total de casos incluídos no arquivo de dados. Outra possibilidade é selecionar aleatoriamente uma quantidade exata de casos, a partir do número de casos que especificamos no arquivo. Para isso, é necessário determinar quantos serão os "primeiros caso" considerado para a amostra. Se um número igual ao número de casos contidos no (em nosso exemplo são 100 indivíduos), isso implicará que a amostra será selecionada aleatoriamente entre todos os casos. Lembre-se cada indivíduo tem a mesma chance de ser selecionado.

Seleciona Random sample of cases (Random = aleatória, sample – amostra e cases – indivíduos.)

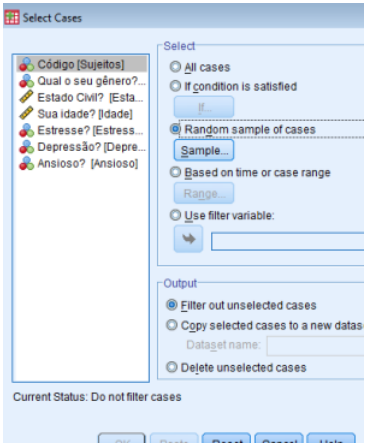


Figura 2.3: Amostra aleatória simples no SPSS

Queremos uma amostragem aleatória simples de 10 indivíduos (n=10) dos 100 indivíduos (N=100).

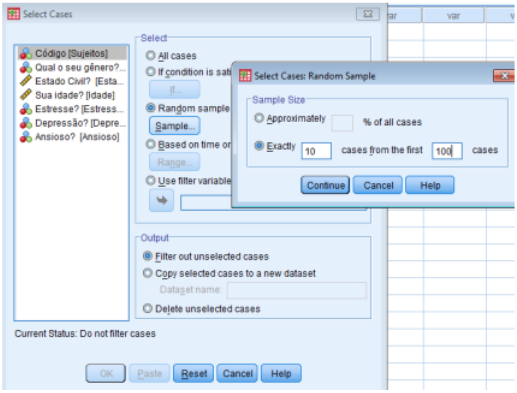


Figura 2.4: Selecionado uma amostra aleatória simples no SPSS

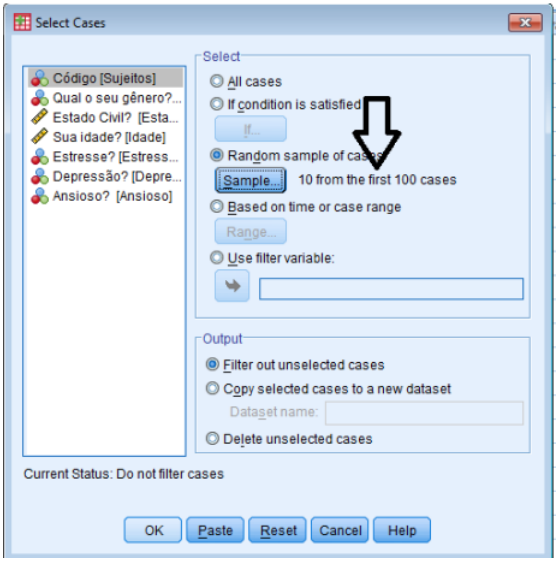
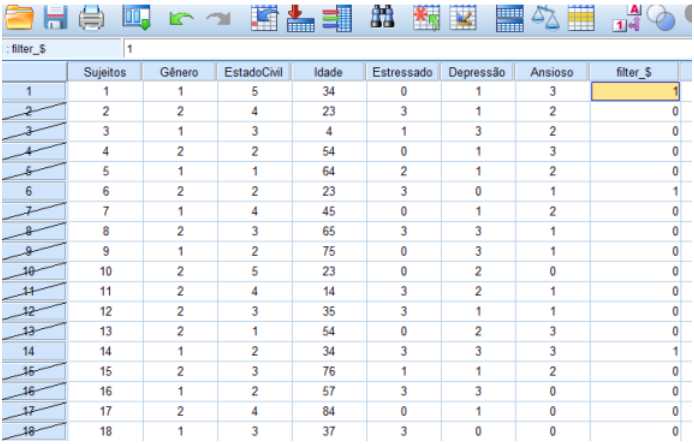


Figura 2.5: Selecionado uma amostra aleatória simples no SPSS de tamanho 100



: filter_$ 1

	Sujeitos	Gênero	EstadoCivil	Idade	Estressado	Depressão	Ansioso	filter_$
1	1	1	5	34	0	1	3	1
2	2	2	4	23	3	1	2	0
3	3	1	3	4	1	3	2	0
4	4	2	2	54	0	1	3	0
5	5	1	1	64	2	1	2	0
6	6	2	2	23	3	0	1	1
7	7	1	4	45	0	1	2	0
8	8	2	3	65	3	3	1	0
9	9	1	2	75	0	3	1	0
10	10	2	5	23	0	2	0	0
11	11	2	4	14	3	2	1	0
12	12	2	3	35	3	1	1	0
13	13	2	1	54	0	2	3	0
14	14	1	2	34	3	3	3	1
15	15	2	3	76	1	1	2	0
16	16	1	2	57	3	3	0	0
17	17	2	4	84	0	1	0	0
18	18	1	3	37	3	0	0	0

Figura 2.6: Output de uma amostra aleatória simples no SPSS selecionada

Será que foram selecionados 10 indivíduos aleatoriamente? Passos no SPSS: Analyse – Descriptive Statistics - Frequencies.

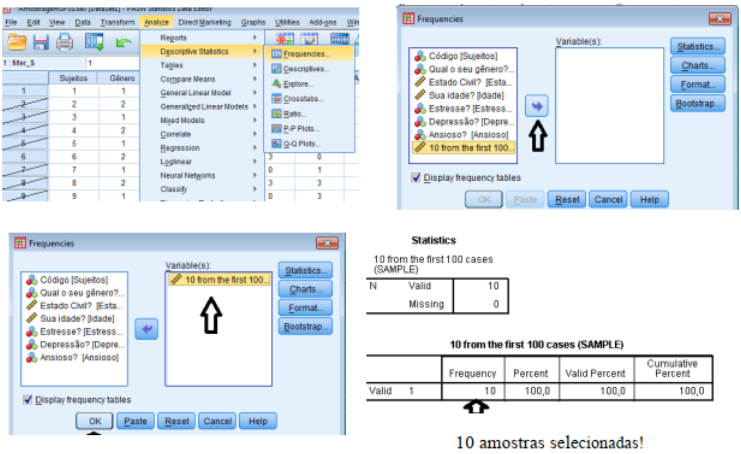


Figura 2.7: Selecionada uma amostra aleatória

2.5 Amostragem aleatória estratificada

A amostra aleatória estratificada consiste em dividir a população em grupos, chamados estratos, que representam homgeneidade interna em relação a alguma caracterísrica, mas que são heterogêneos entre si em relação a dita caracterísrica. Pode-se estratificar:

- Segundo gênero de ansiosos
- Estado civil de pessoas deprimidas
- Lugar de residência etc

Cada estrato funciona independentemente, logo pode-se aplicar dentro de cada estrato uma amostra aleatória simples, por eemplo.

Em algumas ocasiões, esse tipo de amostragem apresenta dificuldades pois exige um conhecimento da população (tamanho geográfico, sexo, idades, escolaridade, $\cdots$).

A distribuição da amostra entre os distintos estratos se faz mediante um procedimento chamado fixação que pode ser de três tipos:

1. Igual ou uniforme - quando se reparte a amostra uniformemente nos estratos

2. Porporcional - coniste em repetir a amostra de forma diretamente proporcional ao número de elemento de cada estrato

3. Óptima - quando além do número de elemento nos estratos, tem-se em conta sua variabilidade

Exemplo 2.5.1. *De uma população de 80 pacientes, pretende-se retirar uma amostra de 20 pacientes controlada pelo nível de estresse. Sabendo que 246 deles tem um nível de estresse 1, 26 nível de estresse 2, 14 nível de estresse 3 e 14, nível de estresse 4. Qual o tamanho a amostra em cada estrato quando se utiliza uma fixação porporcional?*

Divide-se a população N = 80 em 4 estratos (4 níveis de estresse) e divide-se a amostra entre os estratos proporcionalmente ao número de elementos de cada um. Assim,

- Para os estressados nível 1 terá $\frac{26}{80} = 0,32$, pelo que selecionamos $0,32 \times 20 = 6,4$ estressados

- Para os estressados nível 2 terá $\frac{26}{80} = 0,32$, pelo que selecionamos $0,32 \times 20 = 6,4$ estressados

- Para os estressados nível 2 terá $\frac{14}{80} = 0,175$, pelo que selecionamos $0,175 \times 20 = 3,5$ estressados

	Nível 1	Nível 2	Nível 3	Nível 4	Total
N_i	26	26	14	14	N = 80
$f_i = \frac{N_i}{N}$	0,32	0,32	0,175	0,175	
$n_i = nf_i = 20f_i$	6.4	6,4	3,5	3,5	$n \approx 20$

- Para os estressados nível 2 terá $\frac{14}{80} = 0,175$, pelo que selecionamos $0,175 \times 20 = 3,5$ estressados

O somando o número em cada estrato temos: $6,4+6,4+3,4+3,4 \approx 20$.

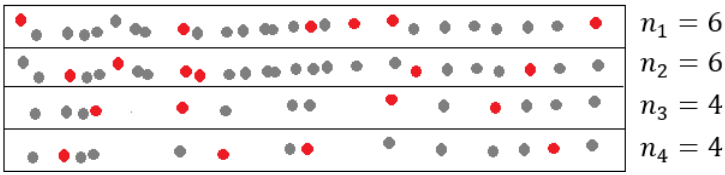


Figura 2.8: Exemplo de amostragem estratificada

Sendo N_i tamanho populacional, $f_h = N_h/N$ é a proporção sobre o total e $n_h = bf_h = 20f_h$.

Esta técnica pertence à família de amostras probabilísticas e consiste em dividir toda a população ou o 'objeto de estudo' em diferentes subgrupos ou estratos diferentes, de maneira que um indivíduo pode fazer parte apenas de um único estrato ou camada. Após as camadas serem definidas, para criar uma amostra, selecionam-se indivíduos utilizando qualquer técnica de amostragem em cada um dos estratos de forma separada. Por exemplo, se usamos uma amostragem aleatória simples em cada estrato, estamos falando de amostragem aleatória estratificada. Podemos usar outras técnicas de amostragem em cada estrato (amostragem sistemática, aleatória, com reposição, etc).

Depois que um arquivo de dados é criado, às vezes pode ser necessário selecionar certos casos através de algum procedimento de amostragem. O programa SPSS permite selecionar casos de acordo com diferentes procedimentos de amostragem (aleatórios ou não acaso; por unidades simples ou compostas). A seguir iremos descrever alguns exemplos de como realizar com o programa SPSS diferentes tipos de procedimentos de amostragem. As definições e critérios dessas amostras podem ser visto em (Martínez e Moreno, 2014) assim como sua implicação para validez.

Como resultado dos diferentes procedimentos de amostragem com SPSS pode novas variáveis irão aparecer no arquivo de dados, que irá registrar os casos selecionados. Normalmente, essas variáveis usam 1 ou um valor decimal diferente de 0 para indicar os casos selecionados, e um 0 ou um dado ausente para os casos não selecionado. Algumas dessas variáveis são temporárias fazemos uma nova amostra, as informações sobre os casos podem ser perdidas selecionado. Por este motivo, é recomendado em todos os casos copiar estes variáveis temporárias em novas variáveis de seleção que podem ser retidas.

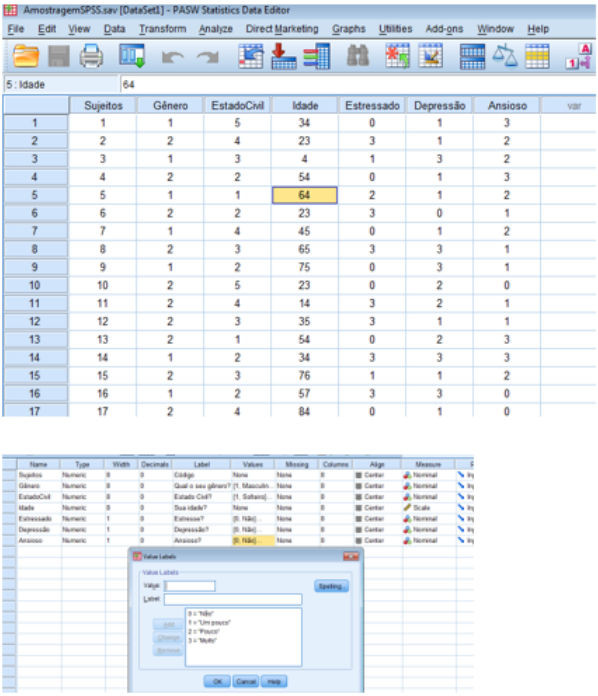
Quando usamos a função de seleção de dados com o programa com um dos essas variáveis de seleção (ou filtro), os casos não selecionados serão riscados com uma linha diagonal.

Por padrão, o programa não exclui os arquivos do arquivo casos não selecionados, mas apenas os descarta para análises subsequentes. Está opção geralmente é a mais conveniente, pois permite reutilizar as descartadas se estiver interessado mais tarde.

Para exemplificar use o Arquivo AmostragemSPSS.sav salvo na plataforma da disciplina.

- Sujeito: 1 a 100 indivíduos

- Gênero: 1 Masculino e 2 Feminino
- Estado Civil: 1 = Solteiro, 2 = Casado, 3 = Separado, 4 = Divorciado e 5 = Viúvo.
- Idade: Variável quantitativa
- Ansioso: 1 = Não, 2 = Um pouco, 3 = Pouco e 4 = Muito.



Selecionando indivíduos (casos) no SPSS

Figura 2.9: Identificando as variáveis da amostra

Selecionando indivíduos (casos) no SPSS. Data – Select Cases – If condition is satisfied

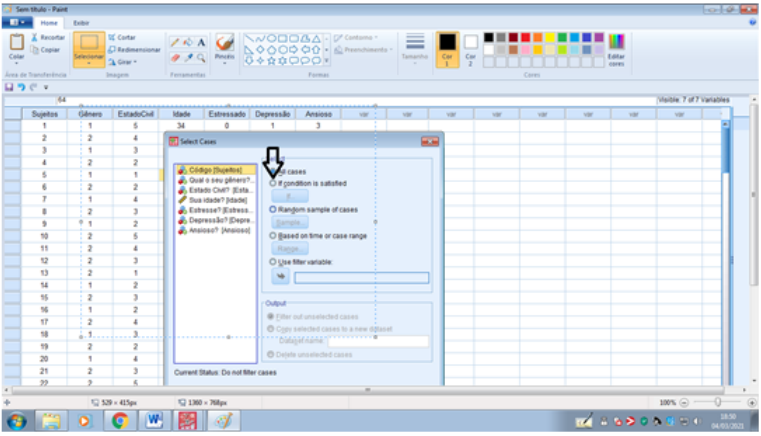


Figura 2.10: Selecionando casos

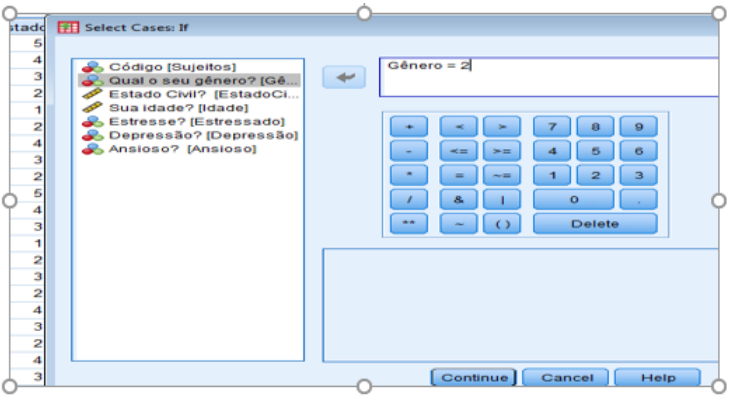
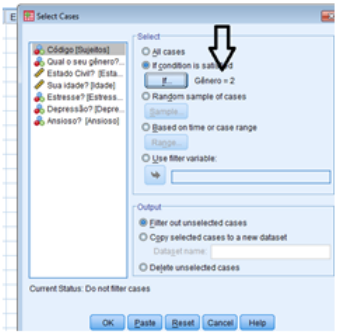


Figura 2.11: Selecionando casos no SPSS



Observa-se que apenas o gênero feminino (2) foi selecionado.

Figura 2.12: Selecionando o gênero feminino

Observa-se na Figura abaixo que os casos não selecionados foram riscados, como se observa à esquerda da figura.

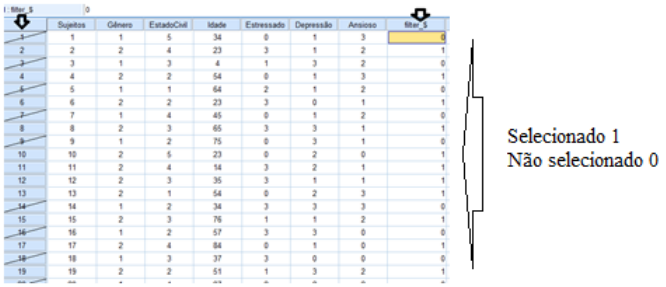


Figura 2.13: Identificação dos selecionado e não selecionados

Esse tipo de amostragem a população em estudo não é homogénea no que respeita à caraterísticas a observar. Neste caso é conveniente agrupar os indivíduo da população em subgrupos, chamados estratos, diferenciados entre si mais possível, e de modo que cada um deles constitua uma subpopulação mais homogénea (i.e., de modo que os elementos de cada estrato sejam o mais parecidos

possível entre si).

A definição dos estratos é feita com base em informação auxiliar relevant fornecida, por exemplo, por dados do recenseamento da população, por estudos piloto, por opiniões de conhecedores da população, por intuiçã do investigador, ou ainda recorrendo a métodos de análise estatística multivariada, tais como, métodos classificatórios e abordagens fatoriais. E, geral considera-se poucos estratos, sendo estes de dimensão elevada.

A amostra (global) estratificada da população será constituida por algun elementos de cada um dos estratos da população, mais precisamente, ser constituída pelo conjunto das sub-amostras aleatórias simples escolhida em cada estrato.

No que respeita à determinação da dimensão das várias subamostras, repartição proporcional (taxa de amostragem igual em todos os estratos) uma das afetações mais usadas, principalmente quando não se consegue estimar a variabilidade de cada estrato, conduzindo ao denominado esquema de amostragem estratificada proporcional. No entanto a repartição proporcional só é ótima (no sentido de conduzir a estimadores com ameno variância) no caso em que todos os estratos têm variabilidade idêntica, o que geralmente não acontece.

Para determinar a afetação ótima, i.e., a que minimiza a variância dos estimadores a utilizar, com determinadas restrições (por exemplo, quando o custo unitário de sondagem não é igual em todos os estratos e existe uma restrição orçamental fixa, ou quando a dimensão global da amostra não pode exceder um valor previamente fixado por restrições de tempo na recolha e tratamento dos dados), recorre-se ao critério de Neyman, o qual reforça as sub-amostras em estratos de maior variabilidade tendo ainda em atenção outras

restrições pertinentes.

Para construir uma amostragem aleatória estratificada de dimensão global n a partir de uma população de dimensão N, dividida em L estratos de dimensão $N_h, i = 1, \cdots, L$, sendo a dimensão das sub-amostras de cada estrato n_h, $h = 1, \cdots, L$, procedemos do seguinte modo:

1. Para cada estrato, digamos para o h-ésimo, numeramos os elementos; do estrato de 1 a N_h

2. Gera-se n_h, números inteiros compreendidos entre 1 e N_h. Numa amostragem aleatória estratificada sem reposição os números repetidos não servem, i.e., têm de ser substituidos por outros; numa amostragem aleatória estratificada com reposição os números repetidos servem;

3. A sub-amostra de elementos do estrato i é constituída pelos elementos do estrato i correspondentes aos números escolhidos;

4. Repetindo os passos anteriores para os C estratos, uma amostragem aleatória estratificada de dimensão global $n = n_1 + \cdots + n_L$ é constituida pelas sub-amostras retiradas dos L estratos pelo processo atrás referido.

Nesse tipo de amostragem, a população é dividida em diferentes camadas ou estratos com base em critérios geográficos, econômicos, sociais, psicológicos etc. Para aplicar isso tipo de amostragem é necessário que os estratos estejam bem definidos para que não há sobreposições e, se possível, conheça os tamanhos dos estratos. A amostragem aleatória estratificada é obtida selecionando indivíduos

aleatoriamente duplas de uma população em cada um dos estratos, respeitando seus pesos e tamanhos.

Entre as vantagens da amostragem estratificada sobre a amostragem aleatória simples, toca a precisão; se a estratificação for adequada, o erro de estimação será menor do que em uma amostragem aleatória simples do mesmo tamanho. Essa ganância na precisão da estimativa se deve ao fato de que os indivíduos dentro dos estratos são homogêneas e diferentes dos indivíduos que pertencem ao restante estratos.

2.5.1 Procedimento de amostragem estratificada

As unidades da população, N, de acordo com um determinado critério, se dividem em subpopulações ou estratos:

$$N = N_1 + N_2 + \cdots + N_L = \sum_{h=1}^{L} N_h$$

Sendo L o número de estratos. Uma vez que os tamanhos dos estratos foram $(N_1, N_2, \cdots, N_L)$ é selecionado, dentro de cada um deles, aleatoriamente, uma amostra de tamanho n_h, $h = 1, 2, 3, \cdots, L$, e os tamanhos dessas amostras independentes denotam-se por $n_1, \cdots, n_L$. Com este procedimento se completa a amostra definitiva n, formada pela soma de cada estrato:

$$n = n_1 + n_2 + \cdots + n_L = \sum_{h=1}^{L} n_h$$

2.5.2 Os estimadores da média, o total e a proporção numa amostra estratificada

Para estimar a média de uma população usa-se seu estimador ou média amostral, que na amostra estratificada é igual a seguinte expressão:

$$\bar{x}_{cd} = \frac{1}{N}\sum_{h=1}^{L} N_h \bar{x}_h = \frac{1}{N}\sum_{h=1}^{L} N_h \bar{x}_h$$

A variância da média obtem-se a partir das somas ponderadas das variâncias $V(\bar{x}_{cd})$ dos estratos será então:

$$V(\bar{x}_{cd}) = \frac{1}{N^2}\sum_{i=1}^{L} N_h^2 var(\bar{x}_h) =$$

$$\frac{1}{N^2}\sum_{i=1}^{L} N_h^2 (1-f_h)\left(\frac{N_h - n_h}{N_h}\right)\left(\frac{\hat{s}_h^2}{n_h}\right) =$$

$$\frac{1}{N^2}\sum_{i=1}^{L} N_h^2 (1-f_h)\left(\frac{\hat{s}_h^2}{n_h}\right) =$$

$$\sum_{i=1}^{L} W_h^2 (1-f_h)\left(\frac{\hat{s}_h^2}{n_h}\right)$$

Sendo que

$$V(\bar{x}_h) = \left(\frac{N_h - n_h}{N_h}\right)\left(\frac{\hat{s}_h^2}{n_h}\right)$$

e $V\hat{s}_h^2$ a variância amostral de cada estrato, sendo dada por:

$$\hat{s}_h^2 = \frac{\sum_{h=1}^{n_k}(x_{ih} - \bar{x}_h)^2}{n_h - 1}$$

É o mesmo que:

$$\hat{s}_h^2 = \frac{\sum_{h=1}^{n_h} x_{ih}^2 - n_h \bar{x}_h^2}{n_h - 1}$$

A expressão

$$\left(\frac{N_h - n_h}{N_h}\right)$$

é o fator de correção para população finita para cada estrato. Finalmente, a partir do intervalo para a média, o valor para o erro de estimação da média vem dado pela seguinte expressão:

$$E = t_{\alpha/2;n-1}\sqrt{V(\bar{x}_{cd})} = t_{\alpha/2;n-1}\sqrt{\frac{1}{N^2}\sum_{h=1}^{L} N_h^2(1-f_h)\frac{\hat{s}_h^2}{n_h}}$$

.

2.5.3 Determinação do tamanho da amostra (n) para o estimador da média

Na amostra estratificada, para fizar o tamanho da amostra, tem que prefixar o erro de estimação que se está disposto a aceitar e, a demais eleger um critério de fixação.

Partindo da extressão para a média da população estratificada, assumindo normalidade e informação populacional, temos que o erro amostral apresenta a seguinte formulação:

$$E = t_{\alpha/2}\sqrt{V(\bar{x}_{cd})} = t_{\alpha/2}\sqrt{\frac{1}{N^2}\sum_{h=1}^{L} N_h^2(1-f_h)\frac{\sigma_h^2}{n_h}}$$

Temos que:

$$\frac{E^2}{Z_{\alpha/2}^2} = \frac{1}{N^2}\sum_{h=1}^{L} N_h^2(1-f_h)\frac{\sigma_h^2}{n_h}$$

ou, o mesmo que:

$$\frac{E^2}{Z^2_{\alpha/2}} = \sum_{h=1}^{L} N_h^2 \frac{N_h - n_h}{N_h} \frac{\sigma_h^2}{n_h}$$

Que equivale a:

$$\frac{E^2}{Z^2_{\alpha/2}} = \sum_{h=1}^{L} N_h^2 \frac{\sigma_h^2}{n_h} - \sum_{h=1}^{L} N_h \sigma_h^2$$

Dessa forma, temos:

$$\frac{N^2 E^2}{Z^2_{\alpha/2}} + \sum_{h=1}^{L} N_h \sigma_h^2 = \sum_{h=1}^{L} \frac{N_h^2 \sigma_h^2}{n_h}$$

A incognita n não aparece na fórmula anterior. Para superar esta limitação utiliza-se a expressão do peso da amostra de cada estrato em relação a toda a amostra.

$$V_h = \frac{n_h}{n} \Rightarrow n_h = nV_h$$

E, a continuação, na variância dos estratos, quando substitue n_h por nv_h. Desta forma, fica:

$$\frac{N^2 E^2}{Z^2_{\alpha/2}} + \sum_{h=1}^{L} N_h \sigma_h^2 = \sum_{h=1}^{L} \frac{N_h^2 \sigma_h^2}{nv_h}$$

Deixando n em evidência temos:

$$n = \frac{\sum_{h=1}^{L} \frac{N_h^2 \sigma_h^2}{v_h}}{\frac{N^2 E^2}{Z^2_{\alpha/2}} + \sum_{h=1}^{L} N_h^2 \sigma_h^2}$$

2.5.4 Estimador da população total

É obtido multiplicando a média amostral pelo número total de indivíduos (N), ou seja,

$$\hat{\tau}_{cd} = N\bar{x}_{cd} = N\left(\frac{\sum_{h=1}^{L} N_h\bar{x}_h}{N}\right) = \sum_{h=1}^{L} N_h\bar{x}_h$$

Aplicando a variância o estimador do total populacional e levando em conta a propriedade da variância do produto de uma constante por uma variável, $V(kX) = k^2V(X)$, a variância do estimador do total fica da seguinte forma:

$$V(\hat{\tau}) = V[N\hat{\mu}] = N^2V[\hat{\mu}]$$

Sabendo que

$$V[\hat{\mu}] = V\left[\frac{1}{N}\sum_{h=1}^{L}\bar{X}_hN_h\right] = \frac{1}{N^2}\sum_{h=1}^{L}\bar{X}_hN_hV[\bar{X}_h]$$

Então:

$$V[\hat{\tau}] = N^2\frac{1}{N^2}\sum_{h=1}^{L}N_h^2\frac{s_h^2}{n_h}\frac{N_h-n_h}{N_h} = \sum_{h=1}^{L}N_h^2\frac{\sigma_h^2}{n_h}\frac{N_h-n_h}{N_h}$$

Quando não tem σ_h^2 não se pode calcular, então se busca um estimador de $V[\hat{\tau}]$, assim:

$$\hat{V}[\hat{\tau}] = N^2\hat{V}[\hat{\tau}]$$

Assim, temos agora:

$$V[\hat{\tau}] = N^2 \frac{1}{N^2} \sum_{h=1}^{L} N_h^2 \frac{s_h^2}{n_h} \frac{N_h - n_h}{N_h} = \sum_{h=1}^{L} N_h^2 \frac{s_h^2}{n_h} \frac{N_h - n_h}{N_h}$$

Como se conhecem as estimações das variâncias para cada um dos estratos, pode-se escrever então:

$$\hat{V}[\hat{\tau}] = \sum_{h=1}^{L} N_h^2 \hat{V}[\bar{X}_h]$$

O erro cometido ao estimar o total da população, com probabilidade $1 - \alpha$, é menor ou igual ao dobro da raiz quadrada da raiz da variância, portanto,

$$M = 2\sqrt{\hat{V}[\hat{\tau}]} = 2\sqrt{N^2 \hat{V}[\hat{\mu}]} = N2\sqrt{\hat{V}[\hat{\mu}]}$$

Assim, $M = Ncota[\hat{\mu}]$, ou seja, a cota para o erro na estimação do total é igual al número de elementos da população multiplicado pela cota do erro na estimação da média.

Exemplo 2.5.2. *Considere 700 discentes que tem problema psicológico de uma grande universidade. Selecionamos de forma aleatória 12 dos 140 deprimidos, 25 dos 500 anciosos e 8 dos 60 estressados. Estime a quantidade média de discentes que tem algum problema psicológico e obtenha a cota do erro de estimação. Sendo deprimido (D), Ansiosos (A) e Estressados (E). Os dados estão nas duas tabelas abaixo.*

D	23	22	14	20	22	32	34	8	15	16	26
A	18	29	16	32	17	14	11	14	20	18	31
E	10	7	20	14	19	8	14	24			

D	20													
A	15	24	22	16	14	12	0	18	12	16	30	22	20	14
E														

Estime a quantidade média de deprimidos, ansiosos e estressados. Obtenha a cota do erro M de estimação.

Claro que é um caso de amostragem estratificada, pois se tem dividido a população em três estratos. Depois para cada estrato se tem tomado uma amostra aleatória.

Vamos simplificar os cálculos usando a seguinte tabela:

Estrato	N_j	n_h	$\sum_{h=1}^{n_h} X_{hi}$	$\bar{X}_h$	$N_h\bar{X}_h$	S_h^2	
D	140	12	252	5894	21	2940	54,72
A	500	25	455	9517	18,2	9100	51,5
E	60	8	116	1942	14,5	870	37,14
Soma	700	45				12910	

Sendo 23 + 22 + 14 + 20 + 22 + 32 + 34 + 8 + 15 + 16 + 2 6 + 20 = 252. Elevando todos os elementos de D ao quadrado temos: $23^2 + 22^2 + 14^2 + \cdots + 16^2 + 26^2 + 20^2 = 5894$.

Na coluna $\sum_{h=1}^{n_h} X_{hi}$ da tabela temos:

$$\sum_{h=1}^{n_h} X_{hi} = \frac{1}{12}252 = 21$$

A média populacional estimada é

$$\mu = \bar{X} = \frac{1}{N}\sum_{h=1}^{n_h} N_h\bar{X}_h = \frac{1}{700}12910 = 18,44$$

Assim, temos 18,44 pessoais que tem problemas psicológicos. Já a variância amostral será:

$$S_1^2 = \frac{1}{n_1 - 1}\left(\sum_{h=1}^{n_h} X_{1h}^2 - n\bar{X}_h^2\right) = \frac{1}{12-1}(5895 - 12.21^2) = 54,72$$

Na seguinte tabela temos o estimador da variância em cada estrato e de toda população.

Estrato	N_j	n_h	S_h^2	$\frac{S_h^2}{n_h}$	$\frac{N_h - n_h}{N_h}$	$\hat{V}[\bar{X}_h]$	$N_h^2\hat{V}[\bar{X}_h]$
D	41	38	39	47	40	35	
A	16	19	15				
E	16	19	15				

Assim, a variância do primeiro estrato será:

$$\hat{V}[\bar{X}_1] = \frac{S_1^2}{n_1}\frac{N_1 - n_1}{N_1} = 4,56.0,91 = 4,16$$

Assim, a variância do segunda estrato será:

$$\hat{V}[\bar{X}_2] = \frac{S_2^2}{n_2}\frac{N_2 - n_2}{N_2} = 2,06.0,95 = 1,95$$

A variância do terceiro estrato será:

$$\hat{V}[\bar{X}_3] = \frac{S_3^2}{n_3}\frac{N_3 - n_3}{N_3} = 4,64.0,86 = 4,02$$

Para estimar a variância da média populacional estimada usa-se a seguinte fórmula:

$$\hat{V}[\hat{\mu}] = \frac{1}{N^2}N_h^2\hat{V}[\bar{X}_h] = \frac{1}{700^2}585462,52 = 1,19$$

A quantidade da média estimada de D, A e E serão:

	quantidade média estimada $\bar{X}_h$
D	21
A	18,2
E	14,5

As cotas de erro de estimação de calculam, como sempre, igual ao dobro da raiz quadrada da correspondente variância estimada. AS cota de erro dos deprimido (D), dos ansiosos (A) e dos estressados(E) e do total estão na tabela abaixo.

	cota de erro $M_h = 2\sqrt{\hat{V}(\bar{X}_h)}$
D	4,08
A	2,79
E	4,01
Total pessoas	$M = 2\sqrt{\hat{V}(\hat{mu})}$ M = 2,18

Exercício 2.5.1. *Considere 700 estudantes que tem problema psicológico de uma grande universidade. Selecionamos de forma aleatória 12 dos 140 deprimidos , 25 dos 500 anciosos e 8 dos 60 estressados. Estime a quantidade média de discentes que tem algum problema psicológico e obtenha a cota do erro de estimação. Sendo deprimido(D), Ansiosos (A) e Estressados (E). Os dados estão nas tabelas abaixo.*

D	20	21	16	24	20	30	36	10	11	17	23
A	16	20	18	34	19	15	13	18	24	20	32
E	12	9	22	12	21	10	16	26			

D	19													
A	9	28	24	15	12	0	12	20	14	18	28	21	20	15
E														

2.5.5 Tamanho da amostra

Vamos encontrar o tamanho amostral para que, com uma probabilidade do 95%, o erro de estimação da média populacional não supere a cota M.

$$M \geq 2\sqrt{V[\hat{\mu}]}$$

$$\frac{M^2}{4} \geq V[\hat{\mu}]$$

Se o número de elementos de cada estrato é grande, existe pouca diferença entre dividir por $N_h - 1$ ou por N_h. Toma-se a decisão de substituir $N_h - 1$ por N_h, então:

$$\frac{M^2}{4} \geq \frac{1}{N^2} \sum_{h=1}^{L} N_h^2 \frac{\sigma_h^2}{n_h} \frac{N_h - n_h}{N_h - 1}$$

$$N^2 \frac{M^2}{4} \geq \sum_{h=1}^{L} N_h^2 \frac{\sigma_h^2}{n_h} \frac{N_h - n_h}{N_h - 1}$$

Conhecendo $V[\hat{miu}]$, $n_h = nw_h$,e passando N^2 para primeiro membro, temos:

$$N^2\frac{M^2}{4} \geq \sum_{h=1}^{L} N_h^2 \frac{\sigma_h^2}{nw_h}\frac{N_h - nw_h}{N_h}$$
$$N^2\frac{M^2}{4} \geq \sum_{h=1}^{L} N_h^2 \frac{\sigma_h^2}{nw_h}\frac{N_h}{N_h} - \sum_{h=1}^{L} N_h^2 \frac{\sigma_h^2}{nw_h}\frac{nw_h}{N_h}$$
$$N^2\frac{M^2}{4} \geq \frac{1}{n}\sum_{h=1}^{L} N_h^2 \frac{\sigma_h^2}{w_h} - \sum_{h=1}^{L} N_h\sigma_h^2$$

Passando $-n\sum_{h=1}^{L} N_h\sigma_h^2$ para o primeiro membro e colocando n em evidência temos:

$$\left(\sum_{h=1}^{L} N_h\sigma_h^2 + N^2\right) n \geq \sigma_h^2 N_j^2 \frac{\sigma_h^2}{w_h}$$
$$n \geq \frac{\sum_{h=1}^{L} \frac{N_h^2\sigma_h^2}{w_h}}{\sum_{h=1}^{L} N_h\sigma_h^2 + N^2\frac{M^2}{4}}$$

Exemplo 2.5.3. *Considerando os dados do exemplo passado. Vamos estimar a média de pessoas que tem D,A e E com uma cora de erro* $\frac{M^2}{4} = \frac{1^2}{4} = 1/4$.

É habitual utilizar alocação proporcional ao produto do tamanho pelo desvio padrão de cada estrato, ou seja, utilizar a alocação de Neyman. Vamos utilizar a tabela abaixo para facilitar os calculos.

Estrato	N_h	S_h^2	$\hat{\sigma}_h = S_h$	$\hat{\sigma}_h N_h$	w_h
D	140	54,72	7,39	1035,69	0,20
A	500	51,5	7,17	3588,12	0,71
E	60	37,14	6,09	365,67	0,07
Soma				4989,53	1,00

A segunda e terceira coluna já foram calculados, a quarta, quinta e sexta se calculam com esses dois resultados, a saber, N_h e S_h^2. Assim, temos:

$$w_1 = \frac{\hat{\sigma}_1 N_1}{Total} = \frac{1035,7}{4989,53} = 0,20$$
$$w_2 = \frac{\hat{\sigma}_2 N_2}{Total} = \frac{3588,8}{4989,53} = 0,72$$
$$w_3 = \frac{\hat{\sigma}_3 N_3}{Total} = \frac{363,7}{4989,53} = 0,07$$

Ou seja, apoximadamente $w_1 = 20\%$ dos elementos são indivíduos com depressão, $w_2 = 72\%$ são de ansiosos e, $w_3 = 7\%$ estressados.

Agora, vamos encontrar o valor do tamanho amostral, n.

Estrato	N_h	$\hat{\sigma}_h^2 = S_h^2$	$N_h \hat{\sigma}_h^2$	$N_h^2 \hat{\sigma}_h^2$	w_h	$\frac{N_h^2 \hat{\sigma}_h^2}{w_h}$
D	140	54,72	7661,82	1076255,08	0,10	9813861,6
A	500	51,5	25750	12785000	0,75	17005679
E	60	37,14	2228,57	133714,44	0,13	1000856,6
Soma			35640,39	4989,53	1,00	27820397

$$n \geq \frac{\sum_{h=1}^{L} \frac{N_h^2 \sigma_h^2}{w_h}}{\sum_{h=1}^{L} N_h \sigma_h^2 + N^2 \frac{M^2}{4}}$$
$$n \geq \frac{27820397,8}{35640,39 + 700^2 \frac{1}{4}} = 175,92$$

A amostra completa tem 176 elementos, repartidos da seguinte forma:

Estratos	w_h	$n_h = 175,92.w_h$	Tamano	Tamanho aprox.
D	0,10	19,22	20	19
A	0,75	133,19	134	133
E	0,13	23,5	24	24
Total	1		178	176

Assim, cada estrato terá 20 deprimidos, 134 ansiosos e 24 estressados.

Exercício 2.5.2. *Considerando os mesmos dados. Vamos estimar a média de pessoas que tem D,A e E com uma cora de erro 0,15.*

2.5.6 Estimador da proporção populacional p

Agora, a variável de estudo separa os indivíduos em duas classes mutuamente excludentes, C e $C^{'}$ (acerto, erro). Denominamos A_h ao número de unidades do estrato h que pertence a classe C, então $P_h = \frac{A_h}{N_h}$ é a proporção de unidade da classe C no estrato h (porporção de aceros na estrato h). O estimador da proporção populacional é a proporção amostral $p_h = \frac{a_h}{n_h}$ ou porporção de unidades da amostra desse estrato que pertence a classe C. Para estimar a proporção populacional leva-se em conta as proporções de cada um dos estratos ponderados pelos seus respectivos tamanhos, de tal maneira que:

$$\hat{p}_{cd} = \frac{1}{N^2}(N_1\hat{p}_1 + \cdots + N_L\hat{p}_h) = \sum_{h=1}^{L} N_h\hat{p}_h$$

Assim,

$$\hat{p}_h = \frac{\sum_{h=1}^{n_h} a_{ih}}{n_h}$$

´é o estimador das respectivas proporções em cada estrato.

2.5.7 Estimador da variância estimada de $\hat{p}_{cd}$

Por outra lado, a variância da proporção populacional se obtém a partir da soma ponderada das respectivas variâncias dos estratos, ou seja,

$$\begin{aligned} V(\hat{p}_{cd}) &= \sum_{h=1}^{L} V_h^2 V(\hat{p}_h) \\ &= \frac{1}{N^2} \sum_{h=1}^{L} N_h^2 \hat{V}(\hat{p}_h) \\ &= \frac{1}{N^2} \sum_{h=1}^{L} N_h^2 \left(\frac{N_h - n_h}{N_h} \right) \left(\frac{\hat{p}_h \hat{q}_h}{n_h - 1} \right) \end{aligned} \tag{2.1}$$

Exemplo 2.5.4. *Deseja-se estimar a proporção de casas numa região em que tem indivíduo com algum problema psicológico. A região se divide em três estratos: região A, região B e região C (área rual). Os respectivos estratos são* $N_1 = 150$*,* $N_2 = 62$ *e* $N_3 = 98$ *casas. Seleciona-se uma amostra aleatória estratificada de n = 40 casas com fixação proporcional. Em outras palavras, toma-se uma amostra aleatória simples em cada estrato; os tamanhos das amostras são:* $n_1 = 18$*,* $n_2 = 10$ *e* $n_3 = 12$ *casas. Estime a proporção de casas e fixe um limite para o erro de estimação. Na tabela abaixo apresenta-se alguns resultados.*

Estratos	n	Número de casas	$\hat{p}_h$
A	$n_1 = 18$	14	0,77
B	$n_2 = 10$	4	0,40
C	$n_3 = 12$	6	0,50

O valor estimado de $\hat{p}_{cd}$ será:

```
N1=150; N2=62; N3=98
```

```
N=N1+N2+N3;N
n1=18; n2=10; n3=12
p1<-0.77; p2<-0.25; p3<-0.50
pcd<-(1/N)*(N1*p1+N2*p2+N3*p3); pcd
q1<-1-p1; q2<-1-p2;q3<-1-p3
Vp1<-((N1-n1)/N1)*((p1*q1/(n1-1))); Vp1
Vp2<-((N2-n2)/N2)*((p2*q2/(n2-1))); Vp2
Vp3<-((N3-n3)/N3)*((p3*q3/(n3-1))); Vp3
Vpcd<- (1/N^2)*(N1^2*Vp1+ N2^2*Vp2 +N3^2*Vp3)
Vpcd
pcd + 2*sqrt(Vpcd)
pcd - 2*sqrt(Vpcd)
```

$$\hat{V}(\hat{p}_1) = \left(\frac{N_1 - n_1}{N_1}\right)\left(\frac{\hat{p}_1\hat{q}_1}{n_1 - 1}\right) = \left(\frac{150-18}{150}\right)\left(\frac{(0.77)(0.22)}{18-1}\right) = 0,009$$

De forma análoga temos: $\hat{V}(\hat{p}_2) = 0,022$ e $\hat{V}(\hat{p}_2) = 0,020$.

Assim, o valor de $\hat{V}(\hat{p}cd)$ será:

$$\hat{V}(\hat{p}_c d) = \frac{1}{N^2}\sum_{h=1}^{L=3} N_h^2 \hat{V}(\hat{p}_h)$$

A variância $\hat{V}(\hat{p}_{cd})$ será:

$$\hat{V}(\hat{p}_{cd}) = \frac{1}{310^2}(150^2(0,007) + (62)^2(0,017) + 98^2(0,019)) = 0,0049$$

Então, o valor estimado da proporção de casas na região, com um erro limite de estimação será:

$$\hat{p}_{cd} \pm 2 * \sqrt{\hat{V}(\hat{p}cd)} = 0,614 \pm \sqrt{0,0049} = 0,614 \pm 0.07.$$

Assim, o limite de erro de estimação é grande. Podemos reduzir esse limite para fazer o estimador mais preciso aumentando o tamanho da amostra.

Exercício 2.5.3. *Considerando as informações da tabela baixo encontre o tamanho amostral dos estratos A, B e C.*

Estratos	n	N_h	Número de casas	$\hat{p}_h$
A	$n_1 = 18$	$N_1 = 150$	14	14/20 = 0,77
B	$n_2 = 10$	$N_2 = 62$	4	0,40
C	$n_3 = 12$	$N_3 = 94$	6	0,50

Se seleciona uma amostra aleatória estratificada de n = 40 casas com fixação proporcional. Em outras palavras, toma-se uma amostra aleatória simples em cada estrato; os tamanhos das amostras são: $n_1 = 18$, $n_2 = 10$ e $n_3 = 12$. Estimar a proporção de casas e fixe um limite para o erro de estimação.

Os valores estimados das proporções de cada casa $\hat{p}_{cd}$ serão:

$$\hat{p}_{cd} = \frac{1}{300}(150.(0,77) + 62.(0,40) + 94.(0,50)) = 0,6144$$

$$\hat{V}(\hat{p}_1) = \left(\frac{N_1 - n_1}{N_1}\right)\left(\frac{\hat{p}_1\hat{q}_1}{n_1 - 1}\right) = \left(\frac{150 - 18}{150}\right)\left(\frac{(0,77)(0,27)}{18 - 1}\right) = 0,008$$

$$\hat{V}(\hat{p}_2) = \left(\frac{N_2 - n_2}{N_2}\right)\left(\frac{\hat{p}_2\hat{q}_2}{n_2 - 1}\right) = \left(\frac{62 - 10}{62}\right)\left(\frac{(0,40)(0,60)}{18 - 1}\right) = 0,022$$

$$\hat{V}(\hat{p}_3) = \left(\frac{N_3 - n_3}{N_3}\right)\left(\frac{\hat{p}_3\hat{q}_3}{n_3 - 1}\right) = \left(\frac{94-12}{94}\right)\left(\frac{(0,50)(0,50)}{12-1}\right) = 0,020$$

Assim, o valor da variância estimada será:

$$\begin{aligned}\hat{V}(\hat{p}cd) &= \frac{1}{N^2}\sum_{h=1}^{3} N_h^2 \hat{V}(\hat{p}_h) \\ &= \frac{1}{310^2}[(150)^2(0,008) + (62)^2(0,022) + (98)^2(0,020)] \\ &= 0,004\end{aligned}$$

Então, o valor estimado da proporção de casas, com um limite para o erro de estimação, será dado por:

$$\begin{aligned}\hat{V}(\hat{p}_{cd}) \pm \sqrt{\hat{V}(\hat{p}_{cd})} &= 0,6144 \pm 2\sqrt{0,004} = 0,6144 \pm 0.07058 \\ &= [0,4723; 0,7555]\end{aligned}$$

Para auxiliar nos cálculos podemos fazer no R:

```
N1=150; N2=62; N3=98
N=N1+N2+N3;N
n1=18; n2=10; n3=12
n<- n1+n2+n3; n
np1<-14; np2<-4; np3<-6
p1<-np1/n1; p2<-np2/n2; p3<-np3/n3
p1;p2;p3
pcd<-(1/N)*(N1*p1+N2*p2+N3*p3); pcd
q1<-1-p1; q2<-1-p2;q3<-1-p3
Vp1<-((N1-n1)/N1)*((p1*q1/(n1-1))); Vp1
```

```
Vp2<-((N2-n2)/N2)*((p2*q2/(n2-1))); Vp2
Vp3<-((N3-n3)/N3)*((p3*q3/(n3-1))); Vp3
Vpcd<- (1/N^2)*(N1^2*Vp1+ N2^2*Vp2 +N3^2*Vp3)
pcd + 2*sqrt(Vpcd)
pcd - 2*sqrt(Vpcd)
```

Exercício 2.5.4. *Considerando os dados da tabela. Estimar a proporção de casas e fixe um limite para o erro de estimação.*

Estratos	n	N_h	Número de casas	$\hat{p}_h$
E1	$n_1 = 32$	$N_1 = 160$	10	10/32
E2	$n_2 = 12$	$N_2 = 82$	9	9/12
E3	$n_3 = 14$	$N_3 = 94$	7	7/14

Observou-se que o limite para o erro de estimação é bastante grande. Poderíamos reduzir esse limite e fazer o estimador mais preciso incrementando o tamanho da amostra. Esse problema que será visto na próxima secão.

2.5.8 Seleção do tamanho da amostra e fixação para a estimação de proporção

Para estimar uma proporção populacional, primeiro indica-se o limte de erro e o tamanho da amostra. A fórmula para o tamanho da amostra n (para um limite M do erro de estimação) será:

$$n = \frac{\sum_{h=1}^{3} N_h^2 \frac{\sigma_h^2}{v_1}}{N^2 \frac{M^2}{4} + \sum_{h=1}^{3} N_h \sigma_h^2}$$

Sendo v_h a fração de observações fixadas ao estrato h, σ_h^2 é a variância populacional para o estrato h. No caso σ_h^2 vem dado por $p_h q_h$.

O tamanho da amostra aproximado para estimar p com um limite M para o erro de estimação será:

$$n = \frac{\sum_{h=1}^{L} N_h^2 \frac{p_h q_h}{v_1}}{N^2 \frac{M^2}{4} + \sum_{h=1}^{L} N_h p_h q_h}$$

A fórmula para a fixação aproximada ao custo para um valor fixo de $V(\hat{p}_{cd})$, ou mínimizar $V(\hat{p}_{cd})$ para um custo fixo será:

$$n_h = n \left(\frac{N_h \sqrt{\frac{p_h q_h}{c_h}}}{N_1 \sqrt{\frac{p_h q_h}{c_h}} + N_2 \sqrt{\frac{p_h q_h}{c_h}} + \cdots + N_H \sqrt{\frac{p_L q_L}{c_L}}} \right)$$

Aplicando o operador somatório, teremos então:

$$n_h = n \left(\frac{N_h \sqrt{\frac{p_h q_h}{c_h}}}{\sum_{h=1}^{L} N_h \sqrt{\frac{p_h q_h}{c_h}}} \right)$$

Sendo N_h o tamanho do h-ésimo estrato e p_h denota a proporção populacional para o h-ésimo estrato e c_h é o costo de obter uma única observação do h-ésimo estrato.

Exemplo 2.5.5. *Considerando os seguintes dados da tabela abaixo. Encontre:*

Estratos	n	N_h	Número de casas	$\hat{p}_h$
A	$n_1 = 20$	$N_1 = 160$	17	17/24= 0,71
B	$n_2 = 8$	$N_2 = 82$	3	3/7= 0,43
C	$n_3 = 12$	$N_3 = 94$	7	7/12 =0,58

a) $\sum_{h=1}^{3} N_h \sqrt{\frac{p_h q_h}{c_h}}$

b) v_1, v_2, v_3, $v_1 a$, $v_1 b$ e $v_1 c$

c) $\sum_{h=1}^{3} \frac{N_h^2 \hat{p}_h \hat{q}_h}{v_h a}$

d) $\sum_{h=1}^{3} N_h \hat{p}_h \hat{q}_h$

e) $V(\hat{p}_{cd})$

f) O tamanho amostral de cada estrato.

Primeiro usa-se Equação 2.5.8 para encontrar as frações de fixação c_h. Utiliza-se $\hat{p}_h$ para aproximar p_h temos que:

$$\sum_{h=1}^{L} N_h \sqrt{\frac{p_h q_h}{c_h}} = N_1 \sqrt{\frac{p_h q_h}{c_h}} + N_2 \sqrt{\frac{p_h q_h}{c_h}} + N_3 \sqrt{\frac{p_h q_h}{c_h}}$$

Aplicando os dados na fórmula acima temos:

$$\begin{aligned}
\sum_{h=1}^{L} N_h \sqrt{\frac{\hat{p}_h \hat{q}_h}{c_h}} &= N_1 \sqrt{\frac{\hat{p}_h \hat{q}_h}{c_h}} + N_2 \sqrt{\frac{\hat{p}_h \hat{q}_h}{c_h}} + N_3 \sqrt{\frac{\hat{p}_h \hat{q}_h}{c_h}} \\
&= 160 \sqrt{\frac{(0,71)(029)}{8}} + 160 \sqrt{\frac{(0,43)(0,57)}{8}} + \\
&+ 94 \sqrt{(\frac{(0,58)(0,42)}{8})} \\
&= 46,25
\end{aligned}$$

Para auxiliar nos cálculos façamos v_1, V_2 e v_3 respectivamente, igual a:

$$v_1 = N_1 \sqrt{\frac{p_1 q_1}{c_1}} = 160 \sqrt{\frac{(0,80).(0,20)}{8}} = 22.627$$

$$v_2 = N_2\sqrt{\frac{p_2q_2}{c_2}} = 82\sqrt{\frac{(0,25)(0,75)}{8}} = 12.553$$

$$v_3 = N_3\sqrt{\frac{p_3q_3}{c_3}} = 94\sqrt{\frac{(0,50)(0,50)}{18}} = 11.078$$

Assim, $v_1a = 22.627/46.259 = 0.49$, $v_2a = 12.553/46.259 = 0,27$, e $v_3a = 11.078/46.259 = 0.24$,

Antes de encontrar o valor do tamanho da amostra precisamos calcular:

$$\begin{aligned}\sum_{h=1}^{3}\frac{N_h^2\hat{p}_h\hat{q}_h}{v_ha} =& \frac{N_1^2\hat{p}_1\hat{q}_1}{v_1a} + \frac{N_2^2\hat{p}_2\hat{q}_2}{v_2a} + \frac{N_3^2\hat{p}_3\hat{q}_3}{v_3a}\\ =& \frac{160^2(0,80)(0,20)}{0,49} + \frac{82^2(0,25)(0,27)}{0,27} + \frac{94^2(0,50)(0,50)}{0,24}\\ =& 22243,79\end{aligned}$$

$$\begin{aligned}\sum_{h=1}^{3}N_h\hat{p}_h\hat{q}_h &= N_1\hat{p}_1\hat{q}_1 + N_2\hat{p}_2\hat{q}_2 + N_3\hat{p}_3\hat{q}_3\\ &= (160)(0,80)(0,20) + (82)(0,25)(0,75) + (94)(0,50)(0,50)\\ &= 64.475\end{aligned}$$

Fazendo o limite de erro de estimação igual 10%, temos então: $2\sqrt{V(\hat{p}_{cd})} = 10\%$, assim:

$$V(\hat{p}_{cd}) = \frac{0,10}{4} = 0,0025 = M$$

Logo, $N^2M = 336^2(0,0025) = 282,24$. Assim, o valor de n será:

$$ne = \frac{\sum_{h=1}^{3} N_h^2 \frac{\hat{p}_h \hat{q}_h}{v_h}}{N^2 M \sum_{h=1}^{3} \hat{p}_h \hat{q}_h} = \frac{22243.79}{282,24 + 64.475} = 64$$

Sendo $v_1a = 22.627/46.259 = 0.49$, $v_2a = 12.553/46.259 = 0,27$, e $v_3a = 11.078/46.259 = 0.24$. O número amostral de cada estrato será:

$$n_1 = ne.v_1a = 64.0,49 = 31$$
$$n_2 = ne.v_2a = 64.0,27 = 17$$
$$n_3 = ne.v_3a = 64.0,24 = 15$$

```
N1=160; N2=82; N3=94
N=N1+N2+N3;N
n1=20; n2=8; n3=12
n<- n1+n2+n3; n
p1<-16/20; p2<-2/8; p3<-6/12

pcd<-(1/N)*(N1*p1+N2*p2+N3*p3); pcd
q1<-1-p1; q2<-1-p2;q3<-1-p3
c1<-8; c2<-8; c3<-18

S3<-N1*sqrt(p1*q1/c1)+ N2*sqrt(p2*q2/c2)
+N3*sqrt(p3*q3/c3); S3
v1<-N1*sqrt(p1*q1/c1);v1
v2<-N2*sqrt(p2*q2/c2);v2
v3<-N3*sqrt(p3*q3/c3);v3

n1<- (N1*sqrt(p1*q1/c1)/S3); n1
```

```
n2<- (N2*sqrt(p2*q2/c2)/S3); n2
n3<- (N3*sqrt(p3*q3/c3)/S3); n3

a1<-n1; a2<-n2; a3<-n3
S3a<-N1^2*p1*q1/a1 + N2^2*p2*q2/a2 +
N3^2*p3*q3/a3; S3a
SNpq<-N1*p1*q1+N2*p2*q2+N3*p3*q3; SNpq
D<- 0.1^2/4; D
N^2*D
n <- S3a/(N^2*D + SNpq); n
n1<-n*a1;n1
n2<-n*a2;n2
n3<-n*a3;n3
```

Exercício 2.5.5. *Considerando os seguintes dados da tabela abaixo. Encontre:*

Estratos	n	N_h	Número de casas	$\hat{p}_h$
E_1	$n_1 = 22$	$N_1 = 180$	17	17/22= 0,717
E_2	$n_2 = 9$	$N_2 = 84$	3	3/9= 0,33
E_3	$n_3 = 12$	$N_3 = 96$	7	7/12 =0,58

a) $\sum_{h=1}^{3} N_h \sqrt{\frac{p_h q_h}{c_h}}$

b) v_1, v_2, v_3, $v_1 a$, $v_1 b$ e $v_1 c$

c) $\sum_{h=1}^{3} \frac{N_h^2 \hat{p}_h \hat{q}_h}{v_h a}$

d) $\sum_{h=1}^{3} N_h \hat{p}_h \hat{q}_h$

e) $V(\hat{p}_{cd})$

f) O tamanho amostral em cada estrato.

Exemplo 2.5.6. *A amostra piloto de uma população está dividida em três estratos* (E_1, E_2, E_3). *O primeiro estrato se caracteriza por seis indivíduos estressados leves. O segundo está composto por 5 indivíduos que tem nível de estresse médio e o terceiro formado por 4 indivíduos que tem estresse alevado.*

Estrato	n_1	Estrato	n_2	Estrato	n_3
E1	20	E2	156	E3	393
E1	13	E2	174	E3	388
E1	24	E2	169	E3	395
E1	31	E2	175	E3	400
E1	19	E2	160		
E1	21				

a) Mediante uma amostragem aleatória simples, estimar a média do número de estrassado populacional e um erro de amostragem. Qual deve ser o tamanho da amostra para que o erro devido a amostragem não seja superior a 10%.

b) Mediante amostragem aleatória simples estratificada. Estimar as quantidades médias de estressados usando os níveis de estresses como estratos.

O desenho amostral serão:

- População objetivo: 2500 indivíduos
- Unidade amostral: indivíduos
- Parâmetros: número de jovens
- Estimadores: média amostral

- Método de selação amostral: amostragem aleatória estratificada
- Critério de estratificação: niveis de estresse

Considere os seguintes estratos:

	Nível 1	Nível 2	Nível 3	N
N_h	1100	800	600	2500

O estimador para calcular a média da quantidade de estressados é igual a:

$$\bar{x} = \frac{\sum_{h=1}^{L} x_i}{n} = \frac{2538}{15} = 154,43$$

A variância amostral é igual a:

$$\hat{s}^{10} = \frac{1}{n-1} = \left(\sum_{i=1}^{10} x_i^2 - n\bar{x}^2 \right) = \frac{763324 - 15(154,43)^2}{15-1} = 23612,5$$

O estimador da variância da média da quantidade de estressados é igual a:

$$V(\bar{x}) = \frac{\hat{s}^2}{n} \left(\frac{N-n}{N} \right) = \frac{23612,5}{15} \left(\frac{2500-15}{2500} \right) = 1367,27$$

O erro de amostragem vale:

$$E = T_{0,05/2;gl-1} \sqrt{V(\bar{x})} = T_{0,0025;14} \sqrt{1564,72} = 85,07$$

Sendo em porcentagem, temos:

$$E(\%) = \frac{E}{\bar{x}} * 100 = \frac{85,07}{169,20} * 100 = 50,28\%$$

Sabendo que o erro amostral vale 50,28%, o que é muito elevado, para reduzir devemos aumentar a amostra. Por exemplo, vamos fazer o erro devido a amostragem ser menor que 15, neste caso, o tamanho amostral será:

$$n = \frac{N\hat{s}^2}{\frac{Ne^2}{z^2_{0,02/2}} + \hat{s}^2} = \frac{2500(23849,6)}{\frac{2500(15)^2}{1,96^2} + 23849,6} = 350,16$$

Assim, precisa-se de uma amostra de 350 indivíduos.

Agora, vamos fazer mediante amostragem aleatória simples estratificada:

Considere a seguinte tabela que resumem as informações necessárias para realizar os cálcuos da média amostral da amostragem estratificada, sua variância, o erro amostral e a sua porcentagem.

Informações	Nível 1	Nível 2	Nível 3
$N_h =$	1000	850	650
$n_h =$	6	5	4
$\frac{N_h - n_h}{N_h} =$	0,99	0,99	0,99
$V_h = \frac{N_h}{N} =$	0,4	0,34	0,26
$\bar{x}_h = \frac{\sum_{h=1}^{2} x_{ih}}{n_h} =$	21,33	166,8	394
$V_h \bar{x}_h =$	8,53	56,71	102,44
$\hat{s}_h^2 = \frac{1}{n_h - 1}\left(\sum_{h=1}^{2} x_{ih}^2 - n_h \bar{x}_h^2\right) =$	35,46	71,7	24,66
$V_h^2 \frac{\hat{s}^2}{n}\left(\frac{N-n}{N}\right) =$	9,01	3,94	1,65

A partir das infromações acima, a média e a variância valem:

$$\bar{x}_{cd} = \sum_{h=1}^{3} V_x \bar{x}_h = 156,42$$

$$V(\bar{x}_{cd}) = \sum_{h=1}^{3} V_h^2 \frac{\hat{s}^2}{n} \left(\frac{N-n}{N} \right) = 9,01 + 3,94 + 1,65 = 14,60$$

Enquanto que o erro e porcentagem são iguais, respectivamente, aos seguintes valores:

$$E = t_{0,05/2;15-1} \sqrt{V(\bar{x}_{cd})} = 2,14.\sqrt{14,60} = 8,18$$

$$E(\%) = \frac{E}{\bar{x}_{cd}}.100 = \frac{8,18}{156,42}.100 = 5,22\%$$

. Esse é erro por estratificar a amostra.

2.5.9 Critérios de alocação

Nas equações para determinar o tamanho de n para um dado erro, o resultado sempre depende do critério de distribuição (v_h) que é aplicado.

A afixação é a distribuição, atribuição, destacamento ou a distribuição do tamanho amostral n entre os diferentes tamanhos da amostra nos estratos, n_h. A soma dos estratos são: $n_1 + n_2 + \cdots + n_L = n$, sendo n o tamanho total considerando todos os estrados. Portanto, as afixações explicam de que modo se repete as n unidades amostrais entre os estratos considerados, ou seja, $n_h = nv_h$. Existem quatro critérios para fazer alocação (afixação): afixação uniforme, afixação proporcional, afixação de Neyman ou de variância mínima e afixação ótima. Vamos exemplificar algumas.

Afixação uniforme: (menos usada). Consiste em distribuir o tamanho da amostra n artes iguais, ou seja, $n_= n_L$. Em todos os

estratos as amostras do mesmo tamanho seriam coletadas.

Afixação proporcional

Consiste simplesmente em distribuir o tamanho total n proporcionalmente ao tamanho de cada estrato. Desta forma, que a amostra a ser retirada do estrato L seria do tamanho:

$$n_h = n \cdot \frac{N_h}{N} = n \cdot C_h \tag{2.2}$$

Exemplo 2.5.7. *Uma população de 1200 pessoas com ansiedade está dividida em 3 estratos para os tais se conhecem o desvio padrão populacional e o peso de cada estrato.*

Desvio padrão populacional (σ)	Peso (w) de cada estrato
$\sigma_1 = 6$	$W_1 = 0,5$
$\sigma_2 = 14$	$W_2 = 0,3$
$\sigma_3 = 80$	$W_3 = 0,2$

Pede-se:

1. Considerando a variância do estimador da média igual a 5. Encontre as respectivas afixação proporcional com e sem reposição, afixação de variância mínimo com e sem reposição e afixação ótima com e sem reposição.

- $w_1 = \frac{N_1}{N} \Rightarrow N_1 = N.w_1 = 1200.0,5 = 600$
- $w_2 = \frac{N_2}{N} \Rightarrow N_2 = N.w_2 = 1200.0,3 = 360$
- $w_3 = \frac{N_3}{N} \Rightarrow N_3 = N.w_3 = 1200.0,2 = 240$
- $\sigma_1^2 = \frac{(N_1-1)}{s_1^2} \Rightarrow s_1 = \sqrt{\frac{36.600}{600-1}} = 6,005$

- $\sigma_2^2 = \frac{(N_2-1)}{s_2^2} \Rightarrow s_1 = \sqrt{\frac{196.360}{360-1}} = 14,02$
- $\sigma_3^2 = \frac{(N_3-1)}{s_3^2} \Rightarrow s_3 = \sqrt{\frac{6400.240}{240-1}} = 80,17$

a) Na afixação proporcional sem reposição

$$e^2 = \left(\frac{1}{n} \sum_{h=1}^{L} W_h S_h^2 - \frac{1}{N} \sum_{h=1}^{L} W_h S_h^2 \right)$$

Deixando n em evidência tem-se:

$$n = \frac{\sum_{h=1}^{L} W_h S_h^2}{e^2 + \frac{1}{N} \sum_{h=1}^{L} W_h S_h^2}$$

Substituindo os valores w_1, w_2, w_3, n, N, S_1^2, S_2^2 e S_3^2 na Equação 2.3 tem-se:

$$n = \frac{\sum_{h=1}^{3} W_h S_h^2}{e^2 + \frac{1}{1000} \sum_{h=1}^{3} W_h S_h^2} = 222,05$$

Agora, encontra-se o tamanho amostral para realizar a afixação. O valor C será: $C = \frac{n}{N} = \frac{222,05}{1200} = 0,1850$. Assim, tem-se:

$$n_1 = cN_1 = 0,1850.600 = 111,02$$
$$n_2 = cN_2 = 0,1850.360 = 66,61$$
$$n_3 = cN_3 = 0,1850.240 = 44,41$$

b) Na afixação proporcional com reposição

Observa-se que o tamanho amostral para obter o mesmo erro da afixação sem reposição agora é superior.

$$n = \frac{\sum_{h=1}^{3} W_h \sigma_h^2}{e^2} = 271,36$$

A afixação proporcional será então:

$$c = n/N = 0,2261$$
$$n_1 = cN_1 = 135,68$$
$$n_2 = cN_2 = 81,408$$
$$n_3 = cN_3 = 54,27$$

No R seria:

```
SomaLsg<- W1*sg1^2+W2*sg2^2+W3*sg3^2; SomaLsg
e<-sqrt(5)
n <- (SomaLsg)/e^2; n
```

```
k<-n/N; k
n1<-k*N1; n1
n2<-k*N2; n2
n3<-k*N3; n3
```

a) Na afixação de mínima variância sem reposição

Nesse caso, tem-se:

$$n = \frac{\left(\sum_{h=1}^{L} W_h S_h^2\right)^2}{e^2 + \frac{1}{N}\sum_{h=1}^{L} W_h S_h^2} = 88,04$$

Considerando os mesmos valores, tem-se no R:

```
SomaL<- W1*S1+W2*S2+W3*S3; SomaL
SomaLS2<- W1*S1^2+W2*S2^2+W3*S3^2; SomaLS2
e<-sqrt(5)
n <- (SomaL)^2/(e^2+(1/N)*(SomaLS2)); n
% [1] 34.62723
```

A afixação de mínima variância sem reposição será:

- $n_h = n.\frac{N_h S_h}{\sum_{h=1}^{L} N_h S_h}$
- $n_1 = n.\frac{N_1 S_1}{\sum_{h=1}^{3} N_h S_h} = \frac{N_1 S_1}{N_1 S_1 + N_2 S_2 + N_3 S_3} = 11,37$
- $n_2 = n.\frac{N_2 S_2}{\sum_{h=1}^{3} N_h S_h} = \frac{N_2 S_2}{N_1 S_1 + N_2 S_2 + N_3 S_3} = 15,93$
- $n_3 = n.\frac{N_3 S_3}{\sum_{h=1}^{3} N_h S_h} = \frac{N_3 S_3}{N_1 S_1 + N_2 S_2 + N_3 S_3} = 60,73$

```
SomaNS<- N1*S1+N2*S2+N3*S3; SomaNS
n1<-n*(N1*S1)/SomaNS;n1
n2<-n*(N2*S2)/SomaNS;n2
n3<-n*(N3*S3)/SomaNS;n3
```

b) Na afixação de mínima variância com reposição

$$n = \frac{\left(\sum_{h=1}^{L} W_h \sigma_h\right)^2}{e^2} = 107,64$$

- $n_h = n.\frac{N_h \sigma_h}{\sum_{h=1}^{L} N_h \sigma_h}$
- $n_1 = n.\frac{N_1 \sigma_1}{\sum_{h=1}^{3} N_h \sigma_h} = \frac{N_1 \sigma_1}{N_1 \sigma_1 + N_2 \sigma_2 + N_3 \sigma_3} = 4,52$
- $n_2 = n.\frac{N_2 \sigma_2}{\sum_{h=1}^{3} N_h \sigma_h} = \frac{N_2 \sigma_2}{N_1 \sigma_1 + N_2 \sigma_2 + N_3 \sigma_3} = 6,33$
- $n_3 = n.\frac{N_3 \sigma_3}{\sum_{h=1}^{3} N_h \sigma_h} = \frac{N_3 \sigma_3}{N_1 \sigma_1 + N_2 \sigma_2 + N_3 \sigma_3} = 24,13$

O cálculo de n no R seria:

```
e<-sqrt(5)
n<- ((W1*sg1+W2*sg2+W3*sg3)^2)/e^2; n
SomaLsg<- N1*sg1+N2*sg2+N3*sg3; SomaLsg
n1<-35*(N1*sg1)/SomaLsg;n1
n2<-35*(N2*sg2)/SomaLsg;n2
n3<-35*(N3*sg3)/SomaLsg;n3
```

A afixação ótima consiste em distribuir o tamanho global n proporcionalmente à variabilidade em cada estrato. Assim, estratos com pouca variância (estratos mais homogêneos) exigirão um tamanho de amostra menor, enquanto estratos mais heterogêneos

exigirão uma amostra maior. Existe a afixação ótima com e sem reposição.

a) Na afixação ótima sem reposição

Na afixação ótima sem reposição o valor de n será:

$$n = \frac{\left(\frac{N_h S_h}{\sqrt{C_h}}\right)\left(N_h S_h \sqrt{C_h}\right)}{e^2 + \frac{1}{N}\sum_{h=1}^{L} W_h S_h^2}$$

Encontrado o valor n a afixação ótima de cada um dos três estrato será dado por:

$$n_h = n \frac{\frac{N_h S_h}{\sqrt{C_h}}}{\frac{\sum_{h=1}^{L} N_h S_h}{\sqrt{C_h}}}$$

Assim, n_1, n_2 e n_3 será calculado, respectivamente, por:

$$n_1 = n\frac{N_1 S_1/\sqrt{C_1}}{N_1 S_1/\sqrt{C_1} + N_2 S_2/\sqrt{C_2} + N_3 S_3/\sqrt{C_3}} = 14,63$$
$$n_2 = n\frac{N_2 S_2/\sqrt{C_2}}{N_1 S_1/\sqrt{C_1} + N_2 S_2/\sqrt{C_2} + N_3 S_3/\sqrt{C_3}} = 18,16$$
$$n_3 = n\frac{N_3 S_3/\sqrt{C_3}}{N_1 S_1/\sqrt{C_1} + N_2 S_2/\sqrt{C_2} + N_3 S_3/\sqrt{C_3}} = 56,55$$

```
C1<-1100; C2<-1400; C3<-2100
```

```
SC<- C1+C2+C3; SC
SomaLdC<- W1*S1/sqrt(C1) +W2*S2/sqrt(C2)+
W3*S3/sqrt(C3); SomaLdC
SomaLmC<- W1*S1*sqrt(C1) +W2*S2*sqrt(C2)+
W3*S3*sqrt(C3); SomaLmC
SomaLS2<- W1*S1^2+W2*S2^2+W3*S3^2; SomaLS2
n<- (SomaLdC*SomaLmC)/(e^2 + (1/N)*SomaLS2)
SomaNdC<- N1*S1/sqrt(C1)+N2*S2/sqrt(C2)+
N3*S3/sqrt(C3); SomaNdC
n1<-n*(N1*S1/sqrt(C1))/SomaNdC; n1
n2<-n*(N2*S2/sqrt(C2))/SomaNdC; n2
n3<-n*(N3*S3/sqrt(C3))/SomaNdC; n3
```

a) Na afixação ótima com reposição

Observa-se que o número amostral necessario para cometer o mesmo erro sem reposição é agora superior. Encontrado o número da amostra (n) realiza-se a afixação ótima da seguinte maneira:

$$n_h = n\frac{\frac{N_h\sigma_h}{\sqrt{C_h}}}{\frac{\sum_{h=1}^{L} N_h\sigma_h}{\sqrt{C_h}}} = 109,25$$

Assim, n_1, n_2 e n_3 será calculado, respectivamente, por:

$$n_1 = n\frac{N_1\sigma_1/\sqrt{C_1}}{N_1\sigma_1/\sqrt{C_1}+N_2\sigma_2/\sqrt{C_2}+N_3\sigma_3/\sqrt{C_3}} = 17,90$$
$$n_2 = n\frac{N_2\sigma_2/\sqrt{C_2}}{N_1\sigma_1/\sqrt{C_1}+N_2\sigma_2/\sqrt{C_2}+N_3\sigma_3/\sqrt{C_3}} = 22,22$$
$$n_3 = n\frac{N_3\sigma_3/\sqrt{C_3}}{N_1\sigma_1/\sqrt{C_1}+N_2\sigma_2/\sqrt{C_2}+N_3\sigma_3/\sqrt{C_3}} = 69,12$$

```
C1<-1100; C2<-1400; C3<-2100
SC<- C1+C2+C3; SC
SomaLdC<- W1*sg1/sqrt(C1)+W2*sg2/sqrt(C2)+
W3*sg3/sqrt(C3); SomaLdC
SomaLmC<- W1*sg1*sqrt(C1) +W2*sg2*sqrt(C2)+
W3*sg3*sqrt(C3); SomaLmC
SomaLS2<- W1*sg1^2+W2*sg2^2+W3*sg3^2
SomaLS2
n<- (SomaLdC*SomaLmC)/e^2;n
SomaNdC<- N1*sg1/sqrt(C1)+N2*sg2/sqrt(C2)+
N3*sg3/sqrt(C3); SomaNdC
n1<-n*(N1*sg1/sqrt(C1))/SomaNdC; n1
n2<-n*(N2*sg2/sqrt(C2))/SomaNdC; n2
n3<-n*(N3*sg3/sqrt(C3))/SomaNdC; n3
```

Exemplo 2.5.8. *Uma população de 1400 pessoas com ansiedade está dividida em 3 estratos para os tais se conhecem o desvio padrão populacional e o peso de cada estrato. Considerando a variância do estimador da média igual a 5.*

Desvio padrão populacional (σ)	Peso (w) de cada estrato
$\sigma_1 = 7$	$W_1 = 0,6$
$\sigma_2 = 15$	$W_2 = 0,3$
$\sigma_3 = 90$	$W_3 = 0,1$

Pede-se:

Pede-se:

1. A afixação proporcional com e sem reposição.
2. A afixação de variância mínima com e sem reposição.
3. A afixação ótima com e sem reposição.

Exemplo 2.5.9. *Uma população de 1000 pessoas transtorno afetivo bipolar está dividida em 3 estratos para os tais se conhecem a variância populacional e o peso para cada estrato. Considerando a variância do estimador da média igual a 5.*

Desvio padrão populacional (σ)	Peso (w) de cada estrato
$\sigma_1^2 = 100$	$W_1 = 0,6$
$\sigma_2^2 = 324$	$W_2 = 0,3$
$\sigma_3^2 = 9025$	$W_3 = 0,1$

Pede-se:

1. A afixação proporcional com e sem reposição.
2. A afixação de variância mínima com e sem reposição.
3. A afixação ótima com e sem reposição.

Índice Remissivo

Referências Bibliográficas

Ávila, M. J. del M., García, J. M. T. *Técnicas Estadísticas Aplicadas*. Grupo Editorial Universitário, 2006.

Ávila, M. J. del M. *Estadística Matemática*. Grupo Editorial Universitário, 2006.

Paradis, E. *R for Beginners*. Institut des Sciences de l'Évolution, Université Montpellier II, France, 2005. https://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf

Pérex, César. *Técnicas de muestreo estatístico*. Instituto de Estudios Fiscales (IEF). Iniversidad de Complutense de Madrid. Madrid: Garceta grupo editoria, 2009.

R Core Team (2024). *R: A Language and Environment for Statistical Computing* Vienna: Austria, 2015. http://www.R-project.org/.

Rmetrics Core Team., Wuertz, D., Setz, T., Chalabi, Y. *fBasics: Rmetrics - Markets and Basic Statistics*. R package version 3011.87, 2014, http://CRAN.R-project.org/package=fBasics.

Sánchez, J., M, C.; Pérez, C. G.; Galícia, L.F.R.; Sanz, A.I.Z. *Problemas de estadística. Descriptiva, probabilidad e inferencia*. Madrid: Ediciones Pirâmides, 1998.

Scheaffer, R.L.; III, W.M.; Ott, R.L. *Elementos de Muestreo*. 6ed., Madrid: Thomson, 2007.

Sarkar, D., Andrews, F. *latticeExtra: Extra Graphical Utilities Based on Lattice*. R package version 0.6-28, 2016. http://CRAN.R-project.org/package=latticeExtra

T. Lumley. *survey: analysis of complex survey samples*. R package version 4.4, 2014.

T. Lumley. *Analysis of complex survey samples. Journal of Statistical Software 9(1): 1-19*. R package version 4.4, 2014

T. Lumley. *Complex Surveys: A Guide to Analysis Using R*. R package version 4.4, 2010.

Wickham H, François R, Henry L, Müller K, Vaughan D. *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4, <https://CRAN.R-project.org/package=dplyr>.

Organizador e Autores

Dr Edwirde Luiz Silva Camêlo (Brasil) - (Organizador) Professor Associado da Universidade Estadual da Paraíba (UEPB). Pós-doutorado em *Estadística Aplicada* (2016) e Doutor em *Estadística e Investigación Operativa* (2007) pela *Universidad de Granada*. Mestrado em Biometria e Estatística Aplicada (2001) pela Universidade Federal Rural de Pernambuco (UFRPE). Atualmente, é professor associado da Universidade Estadual da Paraíba, com atuação nos Departamentos de Estatística e Psicologia e no Programa de Pós-Graduação em Psicologia da Saúde (PPGPS), sendo membro do grupo de pesquisa de Psicologia da Saúde (CNPq/UEPB). E-mail: edwirde@servidor.uepb.edu.br

Dra Dalila Camêlo Aguiar (Brasil) - Doutora em *Estadística Matemática y Aplicada* pela *Universidad de Granada* (UGR), Mestra em *Estadística Aplicada* (2016) pela UGR, Especialista em Estatística Aplicada (2011) pela Fundação de Apoio, Pesquisa e Extensão (FURNE) e Bacharela em Estatística (2010) pela Universidade Estadual da Paraíba (UEPB). Pesquisadora na área de estatís-

tica multivariada aplicada. E-mail: dalilacamelo@correo.ugr.es

Dr Ramón Gutiérrez Sánchez (España) - Professor Titular da *Universidad de Granada* (UGR), Diretor do *Departamento de Estadística e Investigación Operativa*, Coordenador do Mestrado em *Estadística Aplicada*, *CATEDRÁTICO DE UNIVERSIDAD Departamento de Estadística e Investigación Operativa*.
E-mail: ramongs@ugr.es

Dr Ivan Olier (England) His is a *Reader in Artificial Intelligence and Data Science*. My research interests lie in algorithms for Artificial Intelligence, with a specific focus on Causal AI, Digital Twins, and the modelling of large-scale, highly structured, and/or relational data (including multivariate time series, graphs, networks, etc.). My research finds its applications in various domains such as cardiovascular science, drug development, bioinformatics, healthcare, astrophysics, engineering, etc. Additionally, I hold a senior membership position at the Liverpool Centre for Cardiovascular Science (LCCS). E-mail:
I.A.OlierCaparroso@ljmu.ac.uk.

O propósito deste livro é apresentar as técnicas de amostragem estatística em duas facetas: teoria e prática. Seu contenido está enfocado a docentes e discentes universitários de todos níveis que utiliza a amostragem estatística, assim como aos profissionais dos setores em que se aplica a técnica de amostragem (economia, transporte, medicina, psicologia da saúde, matemática, comércio, controle estatístico de qualidade,etc.). O livro apresenta as ferramentas básicas para a amostragem estatística explicando os passos para sua utilização em psicologia. Sabendo que a teoria da probabilidade é o fundamento dos dois métodos de amostragens observado neste livro. Um conhecimento dos métodos gerais de estatística e da teoria básica das estimações do ponto de vista estatístico é essencial para um entendimento adequado do desenvolvimento rigoroso da teoria de amostragem.

30 anos
eduepb