

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA**

**CADERNOS DE MATEMÁTICA E ESTATÍSTICA
SÉRIE H: CADERNOS DE EXERCÍCIOS**

MAT02254 - AMOSTRAGEM II LISTA DE EXERCÍCIOS E TRABALHOS

Elsa Cristina de Mundstock

Digitação Parcial: Bolsistas Anderson do Couto Soares
Rudimar José Schreiber Júnior

UFRGS - SISTEMA DE BIBLIOTECAS
BIBLIOTECA SETORIAL DE MATEMÁTICA
SEÇÃO DE PERIÓDICOS

**SÉRIE H, Nº 2
Porto Alegre, agosto de 2000**

LISTA DE EXERCÍCIOS

1. AMOSTRAGEM SISTEMÁTICA

1.1) Pretende-se estimar a opinião dos arquitetos registrados no CREA sobre a construção de um aeroporto num local atualmente ocupado por uma reserva florestal. Conseguiu-se a lista dos 9000 arquitetos sócios do CREA, estando os nomes ordenados segundo a data de admissão ao quadro da entidade. Decidiu-se por uma amostra de 300 pessoas, usando o processo de 3 sub-amostras sistemáticas da seguinte maneira:

- dividiu-se a população em 100 zonas consecutivas de 90 arquitetos cada,
- sortearam-se 3 números entre 01 e 90 (17, 56, 81),
- tomou-se então a opinião dos arquitetos ocupando as seguintes posições da lista:

Amostra 1: 17, 107, 197, ...

Amostra 2: 56, 146, 236, ...

Amostra 3: 81, 171, 261, ...

- o número de pessoas contra o projeto em cada sub-amostra foi respectivamente, 70, 60 e 50

a) Por que você acha que foi proposto este esquema amostral ?

b) Ache um IC 95% para a proporção de arquitetos contra o projeto, especificando as fórmulas usadas.

1.2) Seja a população hipotética

1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

a) Selecione todas as amostras possíveis com $k = 4$

b) Calcule $V(\bar{x}_{\text{sis}})$ e ρ_I

Suponha a população anterior arranjada assim:

1, 4, 7, 10, 2, 5, 8, 11, 3, 6, 9, 12

c) Calcule $V(\bar{x}_{\text{sis}})$ e ρ_I com $k = 4$

d) Compare os resultados obtidos em ambos os casos. Comente.

1.3) Num cadastro de 13 casas, numa rua, as pessoas estão relacionadas na seguinte ordem M = masculino adulto, F = feminino adulto, m = masculino criança e f = feminino criança.

1	2	3	4	5	6	7	8	9	10	11	12	13
M	M	M	M	M	M	M	M	M	M	M	M	M
F	F	F	F	F	F	F	F	F	F	F	F	F
F	f	m		m	m	f	m	m	m	f	f	
M	m	f		m	m	f	f		f	m		
F	f			f		m						

Tomando como unidades amostrais as pessoas, compare as variâncias de uma amostra sistemática de $k = 5$ e uma amostra aleatória simples de 20% para estimar:

a) Proporção do sexo masculino.

b) Proporção de crianças

c) Proporção de pessoas morando em casa alugada (casas 1, 2, 3, 12 e 13 são alugadas)

Para a amostra sistemática, numere verticalmente cada coluna, passe ao topo da coluna seguinte.

1.4) Suponha uma população de tamanho 9 cujo cadastro está relacionado a seguir:

2 4 5 10 15 18 25 30 32

Selecione uma amostra aleatória de tamanho 3 e ache um IC95% para a média dessa população baseado na amostra selecionada. Interprete o resultado obtido.

2. AMOSTRAGEM POR CONGLOMERADOS

2.1) A seguir está relacionada uma população hipotética de 3 conglomerados, cada um com um determinado número de sub-conglomerados. Os valores dentro dos parênteses indicam a quantidade de unidades em cada caso.

POPULAÇÃO

Conglomerado	Sub-conglomerado
A (15)	A1 (6) A2 (4) A3 (5)
B (20)	B1 (5) B2 (4) B3 (6) B4 (5)
C (20)	C1 (5) C2 (5) C3 (10)

Suponha que foi selecionada uma amostra para estimar a média dessa população. A seguir estão listados os elementos selecionados na amostra. Os números entre parênteses mostram os valores observados da variável X em cada caso.

AMOSTRA

Conglomerado	Sub-conglomerado
A	A1 (10 12 14) A3 (15 11)
B	B2 (15 13) B3 (12 14)

- Estime a média da população em base à amostra observada e mostre as fórmulas utilizadas.
- Suponha que a média da população seja $\mu = 13$, teça seus comentários em relação ao possível valor de ρ_i na população.

2.2) Uma pesquisa foi realizada para estimar o peso de porcos em uma criação. Os porcos se encontravam em 50 baias de 25 porcos cada uma. Foram selecionadas aleatoriamente 3 baias e em cada baia foram sorteados n_i porcos. Os pesos observados foram:

Baia	Peso dos porcos
1	110 130 120
2	90 110 130 150
3	90 150 160 150

- Ache um IC 95% para o peso médio da população de porcos em estudo.
- Explique para o dono da criação o que informa esse intervalo.

2.3) Suponha a população hipotética:

Conglomerado	x_{ij}
1	1, 2, 3
2	5, 5
3	9, 9, 10

Utilizando uma amostra de tamanho $m = 2$ unidades primárias selecionadas aleatoriamente, e $\bar{n} = 2$ unidades secundárias. Calcule \bar{x} e $\hat{\sigma}_{\bar{\mu}}^2$. Ache um IC 95% para μ .

2.4) Suponha a população seguinte, classificada em conglomerados .

UP	1	2	3	4	5
N_i	70	50	80	40	60

Suponha que foram selecionadas 2 unidades primárias. Em cada UP foram sorteadas 5 unidades secundárias. Os valores obtidos foram:

Conglomerado						
3	x_{1j}	25	20	20	20	25
5	x_{2j}	20	25	25	30	30
2	x_{3j}	40	25	35	20	30

- a) Estime a média da população.
b) Ache IC de amplitude 2σ para μ .

2.5) No cadastro do exercício 1.3, considerar cada casa um conglomerado. Selecionar $m=5$ casas e incluir na amostra todos os moradores das casas selecionadas.

- a) Estimar a proporção de pessoas do sexo masculino e a proporção de crianças.
b) Comparar a precisão deste método com a amostragem sistemática.

3. AMOSTRAGEM COM PROBABILIDADES VARIÁVEIS

3.1) Suponha a população seguinte, classificada em conglomerados:

UP	1	2	3	4
N_i	50	40	50	60

Suponha que foram selecionadas 2 unidades primárias com probabilidade proporcional ao tamanho. Em cada UP foram selecionadas 3 unidades secundárias. Os valores obtidos foram:

UP	x_{ij}		
3	10	12	14
2	8	12	10

Estime a média da população com um IC 95%.

3.2) Um sindicato pretende estimar a proporção de operários a favor de uma determinada mudança na política salarial da categoria. Há ao todo, 15 empresas nesse setor e foi selecionada uma amostra casual de 4 empresas com PPT. Em cada empresa selecionada foram sorteados 30 operários obtendo-se a seguinte quantidade a favor da proposta:

Empresa selecionada	1	2	3	4
Operários a favor	28	25	25	17

- a) Estime a proporção de operários a favor da proposta na população com um IC 95%
b) Como você explicaria aos líderes sindicais o que representa esse intervalo?

3.3) Repita o exercício 2.3, considerando que as $m=2$ unidades primárias foram selecionadas com probabilidade proporcional ao tamanho e com reposição.

3.4) Repita o exercício 2.4, considerando que as $m=2$ unidades primárias foram selecionadas com probabilidade proporcional ao tamanho e com reposição.

4. ESTIMAÇÃO PELO MÉTODO DA RAZÃO

4.1) Uma fábrica de suco de laranja quer estimar quanto um caminhão com 1000 kg de laranja produzirá de suco natural. Para isso selecionou 10 laranjas e obteve os seguintes resultados para cada laranja:

Laranja	1	2	3	4	5	6	7	8	9	10
Peso (g)	150	130	140	120	160	160	130	170	140	150
Suco (ml)	60	55	50	40	70	60	45	65	55	65

Chamando $P = \text{Peso}$ e $S = \text{Suco}$ foram calculados:

$$\sum P = 1450$$

$$\sum S = 565$$

$$\sum (P - \bar{P})^2 = 2250$$

$$\sum (S - \bar{S})^2 = 802,5$$

$$\sum (P - \bar{P})(S - \bar{S}) = 1175$$

Ache um IC 95% para o total de suco que será obtido do caminhão em questão.

4.2) Utilizando o cadastro de 13 casas do exercício 1.3, selecionar uma amostra aleatória simples de 6 casas.

- Estime a quantidade de pessoas do sexo feminino utilizando o estimador da razão com $Y = \text{total de pessoas da família}$.
- Compare a precisão deste método de amostragem com a aleatória simples.

4.3) Suponha que queremos estimar a proporção de apartamentos vazios em edifícios de apartamentos para alugar. Uma amostra de $n=20$ edifícios foi selecionada e produziu os seguintes dados:

Edifício	Total de aptos.	Aptos vazios	Edifício	Total de aptos.	Aptos vazios
1	20	4	11	60	15
2	10	2	12	50	8
3	30	6	13	30	5
4	40	8	14	80	15
5	20	5	15	50	12
6	15	3	16	30	7
7	25	5	17	40	8
8	30	6	18	90	20
9	20	3	19	60	12
10	20	5	20	70	8

a) Calcule r

b) Estime σ_r^2 supondo $(N-n)/N=1$

c) Com os dados anteriores, ache o tamanho de amostra necessário para obter uma precisão de $\pm 0,005$ com uma confiança de 95%

4.4) Suponha que queremos estimar o número de vacas em 1996 em certo estado. O número de vacas em 1994 era $Y=5000$. A unidade amostral era a propriedade agrícola e supõe-se que o número de propriedades é $N=500$. Uma amostra de $n=20$ propriedades foi selecionada e os dados estão a seguir. Estime o número de vacas para 1996, com um IC95%.

Propriedade	1994 Y_i	1996 X_i	Propriedade	1994 Y_i	1996 X_i
1	12	14	11	11	14
2	22	25	12	17	19
3	38	37	13	12	12
4	15	18	14	22	23
5	18	20	15	14	16
6	31	30	16	26	28
7	15	15	17	8	9
8	20	21	18	16	15
9	10	12	19	13	15
10	25	28	20	19	20

4.5) Os dados a seguir representam a disponibilidade de gado bovino nas 319 propriedades de uma região, por faixa de tamanho da propriedade.

Tamanho da propriedade	0-100	101-200	201-300	301-400	401-600	601-1000	Todas as prop.
Número de propriedades	11	48	84	60	77	39	319
Área agrícola por prop. (\bar{y})	63,7	155,3	245,7	344,4	491,6	767,5	367,5
Cabeças de gado por prop. (\bar{x})	25,4	50,1	76,0	99,2	150,8	244,4	113,4
ρ	0,17	0,11	0,05	0,15	0,24	0,26	0,68
$\sigma_{\bar{y}}^2$	454	765	885	707	2775	8928	39528
$\sigma_{\bar{x}}^2$	641	1201	6174	4055	5912	11119	8292

a) Examine a possibilidade de aplicar o método da razão para estimar o total de gado bovino nessa população. Justifique

b) Suponha que se pretende selecionar uma amostra aleatória simples de 64 propriedades para estimar o número total de cabeças de gado no ano posterior ao do levantamento do cadastro. Estime a variância da média no caso de fazer a estimação;

b.1) Com base na média aritmética simples

b.2) Utilizando o método da razão.

c) Calcule a eficiência relativa do método da razão em relação ao método da média aritmética simples. Interprete o resultado obtido.

TRABALHO 1

AMOSTRAGEM SISTEMÁTICA

- 1.1) O arquivo soja.sav contém os dados da produção de soja, por município, dos 7 estados com maior produção no país, estimada pelo IBGE a partir da pesquisa de Produção Agrícola Municipal. Os estados incluídos são: Goiás, Mato Grosso, Mato Grosso do Sul, Minas Gerais, Paraná, Rio Grande do Sul e São Paulo. O arquivo contém as seguintes variáveis: Área colhida (ha), Quantidade produzida (toneladas), Rendimento (kg/ha) e Valor da produção (mil reais).
- Calcule μ e σ^2 para área, quantidade, rendimento e valor.
 - Calcule P, a proporção de municípios com produção acima da média geral dos 7 estados.
 - Ache a matriz de correlações entre área, quantidade, rendimento e valor. Analise a relação existente entre essas variáveis.
 - Crie uma variável com distribuição Uniforme (0,1). Ordene o cadastro por essa variável (“ordem aleatória”), dentro de cada estado.
 - Calcule ρ_1 para quantidade produzida, com $n = 141$, $k=9$.
 - Selecione uma amostra sistemática de $n=141$ municípios.
 - Ache um IC 95% para μ de área, quantidade, rendimento e valor. Calcule o erro relativo em cada caso. Comente.
 - Ache um IC 95% para a proporção de municípios com produção acima da média geral dos 7 estados. Calcule o erro relativo. Comente.
 - Ordene o cadastro pela área colhida, dentro de cada estado (“tendência linear”).
 - Calcule ρ_1 para área colhida, com $n = 141$, $k=9$.
 - Selecione uma amostra sistemática de $n=141$ municípios.
 - Ache um IC 95% para μ de área, quantidade, rendimento e valor. Calcule o erro relativo em cada caso. Comente.
 - Ache um IC 95% para a proporção de municípios com produção acima da média geral dos 7 estados.
 - Compare os resultados em f e i. Explique.

COMANDOS UTILIZADOS NO SPSS:

SUGESTÃO: CRIE UM ARQUIVO COM OS COMANDOS UTILIZADOS PARA REPETIR O PROCEDIMENTO COM O ARQUIVO ORDENADO DE PELA ÁREA. PARA ISSO, APERTE O BOTÃO “PASTE” ANTES DE EXECUTAR O COMANDO.

As letras que identificam cada item (a, b, etc) neste roteiro correspondem ao mesmo item no exercício.

a) STATISTICS / SUMMARIZE / DESCRIPTIVES / VARIABLES / área, quantidade, rendimento, valor.

b)Primeiro, cria-se uma variável indicadora dos municípios com produção acima da média do estado, seja essa variável Prod1:

TRANSFORM/RECODE/INTO DIFFERENT VARIABLE/

Numeric variable: Produção; Name: Prod1; Change;

Old and new values: Old value: Range/ lowest through 1000/New value:0/Add

Old value: All other values/New value:1/Add/Continue/OK

Depois, calcula-se a frequência da nova variável

STATISTICS / SUMMARIZE / FREQUENCIES / PROD1

c) STATISTICS / CORRELATE / BIVARIATE / VARIABLES / área, quantidade, rendimento, valor

d) TRANSFORM/COMPUTE/aleat = UNIFORM(1)
DATA / SORT CASES / SORT BY / estado aleat

e) Calcular ρ_I pela fórmula:

$$\rho_I = \frac{\sum_i N_i^2 (\mu_i - U)^2 - \sum_i \sum_j (x_{ij} - U)^2}{\sum_i (N_i - 1) \sum_j (x_{ij} - U)^2}$$

onde $U = \frac{1}{G} \sum_i (N_i - 1) \sum_j x_{ij}$; $G = \sum_i N_i (N_i - 1)$;

$N_i = 141$ (tamanho dos grupos ou amostras possíveis).

Portanto,

$$G = \sum_i (141)(140) = 5(141)(140);$$

$$U = \frac{1}{5(141)(140)} \sum_i (140) \sum_j x_{ij} = \frac{1}{705} \sum_i \sum_j x_{ij};$$

$$\rho_I = \frac{\sum_i N_i^2 (\mu_i - U)^2 - \sum_i \sum_j (x_{ij} - U)^2}{\sum_i (N_i - 1) \sum_j (x_{ij} - U)^2} = \frac{(141)^2 \sum_i (\mu_i - U)^2 - \sum_i \sum_j (x_{ij} - U)^2}{140 \sum_i \sum_j (x_{ij} - U)^2}$$

e.1) Identificar os grupos (k=9 amostras possíveis)

TRANSFORM / COMPUTE / GRUPO = \$CASENUM - TRUNC ((\$CASENUM - 1) / 9) * 9

e.2) Para calcular

$$\sum_i \sum_j x_{ij}$$

STATISTICS / SUMMARIZE / DESCRIPTIVES / Area / OPTIONS:

e.3) Calcular U como coluna na planilha:

TRANSFORM / COMPUTE

e.4) Calcular $DIF = (x_{ij} - U)^2$

TRANSFORM / COMPUTE /

e.5) Criar um arquivo AGGR.sav com 9 linhas (uma para cada grupo).

Cada linha do novo arquivo (AGGR.SAV) terá a média de Quant e a soma de DIF.

Observe que nos botões (Name&Label) e (Function) define-se o nome da variável resultante e a estatística que será calculada.

DATA / AGGREGATE / BREAK VARIABLE = Grupo;

AGGREGATE VARIABLE = Quant / (Name&Label) MEDIA; (Function) MEAN(Quant);

AGGREGATE VARIABLE = DIF / (Name&Label) SUMDIF; (Function) SUM (DIF));

FILE = A:\AGGR.sav

e.6) Abrir AGGR.sav

e.7) Calcular $(\mu_i - U)^2 = \text{Desvio}$.

TRANSFORM / COMPUTE

Depois, calcular a soma de Desvio

STATISTICS / SUMMARIZE / DESCRIPTIVES

e.8) Para calcular

$$\sum_i \sum_j (x_{ij} - U)^2$$

abrir novamente o arquivo SOJA.SAV

Calcular $\text{Desvio} = (x_{ij} - U)^2$:

TRANSFORM/COMPUTE

Depois, fazer a soma dessa variável

STATISTICS / SUMMARIZE / DESCRIPTIVES / Desvio / OPTIONS=SUM

e.9) Cálculo de ρ_1 :

TRANSFORM / COMPUTE / ROI...

f) Selecione aleatoriamente um número i entre 1 e 9 ($k=9$) e selecione a amostra (seja $i = 3$)

DATA / SELECT CASES / IF CONDITION IS SATISFIED / IF / GRUPO = 3

Calcular média e variância da amostra selecionada

STATISTICS / SUMMARIZE / DESCRIPTIVES /

Calcular proporção de empresas exportadoras:

STATISTICS / SUMMARIZE / FREQUENCIES /

Para as etapas posteriores, o procedimento consiste em repetir o que foi feito até agora, com o arquivo ordenado por Área Colhida. Isto pode ser feito com os comandos que foram salvos usando o "PASTE".

OBS: Conserve todos os arquivos deste exercício (dados e comandos) pois eles serão utilizados no trabalho de Amostragem por Conglomerados.

TRABALHO 2

AMOSTRAGEM POR CONGLOMERADOS

- 1.2) O arquivo soja.sav contém os dados da produção de soja, por município, dos 7 estados com maior produção no país, estimada pelo IBGE a partir da pesquisa de Produção Agrícola Municipal. Os estados incluídos são: Goiás, Mato Grosso, Mato Grosso do Sul, Minas Gerais, Paraná, Rio Grande do Sul e São Paulo. O arquivo contém as seguintes variáveis: Área colhida (ha), Quantidade produzida (toneladas), Rendimento (kg/ha) e Valor da produção (mil reais).
- a) Calcule μ e σ^2 para área, quantidade, rendimento e valor.
 - b) Calcule P, a proporção de municípios com produção acima da média geral dos 7 estados.
 - c) Ache a matriz de correlações entre área, quantidade, rendimento e valor. Analise a relação existente entre essas variáveis.
 - d) Utilize o cadastro ordenado “aleatoriamente” dentro de cada estado.
 - d.1) Agrupar municípios consecutivos dentro do estado em conglomerados de tamanho $\bar{N}=5$ (aproximadamente). Esses grupos de municípios são conglomerados.
 - d.2) Calcule μ e σ^2 para área colhida, quantidade, rendimento e valor em cada conglomerado.
 - d.3) Calcule ρ_i e σ_x^2 para área colhida.
 - d.4) Selecione uma amostra aleatória de $n=140$ municípios em m conglomerados.
 - d.5) Ache um IC99% para a média das variáveis área, quantidade, rendimento e valor, na população, e para a proporção de municípios com produção acima da média geral dos 7 estados. Mostre o erro relativo em cada caso.
 - e) Utilize o cadastro ordenado pela Área Colhida dentro do estado. Repita o procedimento em d).
 - f) Compare os resultados obtidos em d) e e). Explique.

TRABALHO 3

PESQUISA APLICADA

MODELO DE RELATÓRIO

I.- PLANEJAMENTO AMOSTRAL

- 1) Identificação
- 2) Objetivos da pesquisa
- 3) Descrição da população
- 4) Descrição do cadastro
- 5) Tamanho da amostra: cálculo de n , erro esperado
- 6) Delineamento amostral
- 7) Seleção das unidades amostrais
- 8) Procedimentos de substituição em caso de não resposta
- 9) Bibliografia e software

II).- RELATÓRIO DE CAMPO

- 1) A amostra que foi planejada e a amostra que foi observada
- 2) Problemas enfrentados (não resposta, deficiências do cadastro, etc)

III) RESULTADOS

- 1) Estimativas por ponto e por intervalo
- 2) Erros de amostragem