

Lista 1 - MAE0127 - Prof Julio Singer

Alexandre Félix, Eduardo Janotti, Emanuel Hark, Emerson Aragão, Pedro Gallo, William

26/03/2020

Capítulo 2

Exercício 4

A Tabela abaixo contém uma planilha encaminhada pelos investigadores responsáveis por um estudo sobre AIDS para análise estatística. Organize-a de forma a permitir sua análise por meio de um pacote computacional como o R.

Grupo I registro	Tempo de Diagnostico	DST	MAC	Ganho de Peso por Semana
2847111D	pre natal	nao	Pilula	11Kg em 37 semanas
3034048F	6 meses	nao	pilula	?
3244701J	1 ano	nao	Condon	?
2943791B	pre natal	nao	nao	8 Kg em 39 semanas
3000327F	4 anos	condiloma/ sífilis	nao	9Kg em 39 semanas
3232893D	1 ano	nao	DIU	3Kg em 39 semanas
3028772E	3 anos	nao	nao	3 kg em 38 semanas
3240047G	pre natal	nao	pilula	9 Kg em 38 semanas
3017222G		HPV	CONDON	falta exame clinico
3015834J	2 anos	nao	condon	14 Kg em 40 semanas
Grupo I registro	Tempo de Diagnostico	DST	MAC	Ganho de Peso por Semana
3173611E	3 meses	abcesso ovariano	condon	15 Kg em 40 semanas
3296159D	pre natal	nao	condon	0 Kg em ? semanas
3147820D1	2 anos	nao	sem dados	4 Kg em 37 semanas
3274750K	3 anos	nao	condon	8 Kg em 38 semanas
3274447H	pre natal	sífilis com 3 meses	condon	
2960066D	5 anos	nao	?	13 Kg em 36 semanas
3235727J	7 anos	nao	Condon	(-) 2 Kg em 38 semanas
3264897E		condiloma	condon	nenhum Kg
3044120J	5 anos	HPV		3 Kg em 39 semanas 1

Resolução

grupo	reg	tempo	dst	mac	gansem	gantot
1	2847111D	Pré natal	Não	Pílula	0.297	11
1	3034048F	6 meses	Não	Pílula	NA	NA
1	3244701J	12 meses	Não	Condon	NA	NA
1	2943791B	Pré natal	Não	Não	0.205	8
1	3000327F	48 meses	Sim	Não	0.231	9

grupo	reg	tempo	dst	mac	gansem	gantot
1	3232893D	12 meses	Não	DIU	0.077	3
1	3028772E	36 meses	Não	Não	0.079	3
1	3240047G	Pré natal	Não	Pílula	0.237	9
1	3017222G	NA	Sim	Condon	NA	NA
1	3015834J	24 meses	Não	Condon	0.35	14
2	3173611E	3 meses	Sim	Condon	0.375	15
2	3296159D	Pré natal	Não	Condon	0	0
2	3147820D1	24 meses	Não	NA	0.108	4
2	3274750K	3 meses	Não	Condon	0.211	8
2	3274447H	Pré natal	Sim	Condon	NA	NA
2	2960066D	60 meses	Não	NA	0.361	13
2	3235727J	72 meses	Não	Condon	-0.053	-2
2	3264897E	NA	Sim	Condon	0	0
2	3044120J	60 meses	Sim	NA	0.077	3

Exercício 5

Num estudo planejado para avaliar o consumo médio de combustível de veículos em diferentes velocidades foram utilizados 4 automóveis da marca A e 3 automóveis da marca B selecionados ao acaso das respectivas linhas de produção. O consumo (em L/km) de cada um dos 7 automóveis foi observado em 3 velocidades diferentes (40 km/h, 80 km/h e 110 km/h). Construa uma tabela apropriada para a coleta e análise estatística dos dados, rotulando-a adequadamente.

Exercício 6

A planilha apresentada na Figura 2.7 contém dados de um estudo em que o limiar auditivo foi avaliado nas orelhas direita (OD) e esquerda (OE) de 13 pacientes em 3 ocasiões (Limiar, Teste 1 e Teste 2). Reformate-a segundo as recomendações da Seção 2.3

Resolução

Capítulo 3

Exercício 1

O arquivo *rehabcardio* contém informações sobre um estudo de reabilitação de pacientes cardíacos. Elabore um relatório indicando possíveis inconsistências na matriz de dados e faça uma análise descritiva de todas as variáveis do estudo.

#Resolução

Exercício 6

Considere do arquivo *vento*. Observe o valor atípico 61,1, que na realidade ocorreu devido a forte tempestade no dia 2 de dezembro. Calcule as medidas de posição e dispersão dadas na Seção 3.3. Comente os resultados.

Resolução

Exercício 7

Construa gráficos ramo-e-folhas e *boxplot* para os dados do Exercício 6.

Resolução

Exercício 12

Construa gráficos de quantis e simetria para os dados de manchas solares disponíveis no arquivo *manchas*.

Resolução

Exercício 17

Os dados apresentados na figura abaixo referem-se aos instantes nos quais o centro de controle operacional de estradas rodoviárias recebeu chamados solicitando algum tipo de auxílio em duas estradas num determinado dia.

Estadra 1	12:07:00 AM	12:58:00 AM	01:24:00 AM	01:35:00 AM	02:05:00 AM
	03:14:00 AM	03:25:00 AM	03:46:00 AM	05:44:00 AM	05:56:00 AM
	06:36:00 AM	07:26:00 AM	07:48:00 AM	09:13:00 AM	12:05:00 PM
	12:48:00 PM	01:21:00 PM	02:22:00 PM	05:30:00 PM	06:00:00 PM
	07:53:00 PM	09:15:00 PM	09:49:00 PM	09:59:00 PM	10:53:00 PM
	11:27:00 PM	11:49:00 PM	11:57:00 PM		
Estrada 2	12:03:00 AM	01:18:00 AM	04:35:00 AM	06:13:00 AM	06:59:00 AM
	08:03:00 AM	10:07:00 AM	12:24:00 AM	01:45:00 PM	02:07:00 PM
	03:23:00 PM	06:34:00 PM	07:19:00 PM	09:44:00 PM	10:27:00 PM
	10:52:00 PM	11:19:00 PM	11:29:00 PM	11:44:00 PM	

- Construa um histograma para a distribuição de frequências de chamados em cada uma das estradas.
- Calcule os intervalos de tempo entre as sucessivas chamadas e descreva-os, para cada uma das estradas, utilizando medidas resumo gráficos do tipo *boxplot*. Existe alguma relação entre o tipo de estrada e o intervalo de tempo entre as chamadas ?
- Por intermédio de um gráfico do tipo QQ, verifique se a distribuição da variável “Intervalo de tempo entre as chamadas” em cada estrada é compatível com um modelo Normal. Faça o mesmo para um modelo exponencial. Compare as distribuições de frequências correspondentes às duas estradas.

Resolução

Exercício 18

As notas finais de um curso de Estatística foram: 7,5,4,5,6,3,8,4,5,4,6,4,5,6,4,6,3,8,4,5,4,5,5 e 6.

- Determine a mediana, os quartis e a média.

- b) Separe o conjunto de dados em dois grupos denominados **aprovados**, com nota pelo menos igual a 5, e **reprovados**. Compare a variância desses dois grupos.

Resolução

Exercício 25

Na tabela abaixo estão indicadas as durações de 335 lâmpadas.

Duração (horas)	Número de lâmpadas
0 – 100	82
100 – 200	71
200 – 300	68
300 – 400	56
400 – 500	43
500 – 800	15

- a) Esboce o histograma correspondente.
- b) Calcule os quantis de ordem $p=0,1, 0,3, 0,5, 0,7$ e $0,9$.

Resolução

Exercício 28

Os dados encontrados no arquivo *esforco* são provenientes de um estudo sobre teste de esforço cardiopulmonar em pacientes com insuficiência cardíaca. As variáveis medidas durante a realização do teste foram observadas em quatro momentos distintos: repouso (REP), limiar anaeróbio (LAN), ponto de compensação respiratório (PCR) e pico (PICO). As demais variáveis são referentes às características demográficas e clínicas dos pacientes e foram registradas uma única vez.

- a) Descreva a distribuição da variável consumo de oxigênio (VO2) em cada um dos quatro momentos de avaliação utilizando medidas resumo (mínimo, máximo, quartis, mediana, média, desvio padrão, etc), *boxplots* e histogramas. Você identifica algum paciente com valores de consumo de oxigênio discrepantes? Interprete os resultados.
- b) Descreva a distribuição da classe funcional NYHA por meio de uma tabela de frequências. Utilize um método gráfico para representar essa tabela.

Resolução

Exercício 32

Considere uma amostra aleatória simples X_1, \dots, X_n de uma variável X que assume o valor 1 com probabilidade $0 < p < 1$ e o valor 0 com probabilidade $1 - p$. Seja $\hat{p} = n^{-1} \sum_{i=1}^n X_i$. Mostre que

- i) $E(X_i) = p$ e $\text{Var}(X_i) = p(1-p)$.
 ii) $E(\hat{p}) = p$ e $\text{Var}(\hat{p}) = p(1-p)/n$.
 iii) $0 < \text{Var}(X_i) < 0,25$.

Com base nesses resultados, utilize o Teorema Limite Central [ver Sen et al. (2009), por exemplo] para construir um intervalo de confiança aproximado conservador (i.e. com a maior amplitude possível) para p . Utilize o Teorema de Sverdrup [ver Sen et al. (2009), por exemplo] para construir um intervalo de confiança aproximado para p com amplitude menor que a do intervalo mencionado acima.