

LISTA DE EXERCÍCIOS 1 – INE 7001 – PROF. MARCELO MENEZES REIS

ANÁLISE EXPLORATÓRIA DE DADOS

- 1) Identificar nas pesquisas a seguir quais são as variáveis independentes e as dependentes. JUSTIFIQUE suas respostas.
 - a) Seja uma pesquisa de opinião eleitoral, em que há interesse na preferência declarada de homens e mulheres por candidatos a presidência da república.
 - b) Levantamento do perfil dos alunos dos diferentes cursos do CSE, indagando o curso a que pertence, sexo, idade, se trabalha ou não.
 - c) Pesquisa de mercado sobre um novo refrigerante, procurando identificar a preferência de consumidores de 5 faixas etárias, público masculino e feminino.
 - d) Uma metalúrgica está realizando um experimento, variando a temperatura, percentual de ferro e tempo de fundição para observar a dureza do aço fabricado.
 - e) Avaliação dos salários mensais de professores de universidades públicas e privadas.
 - f) Há uma disciplina de Estatística oferecida para 5 cursos diferentes do CTC da UFSC, e os respectivos colegiados querem saber como os alunos estão se saindo. Há um arquivo de dados com as notas finais dos alunos de cada curso.
 - g) A ONU coletou informações sobre o PIB per capita de 155 países, com o objetivo de obter uma “radiografia da pobreza”. Os países podem ser enquadrados em seis regiões econômico-geográficas.

- 2) Identificar qual é o grau de mensuração das variáveis descritas a seguir. JUSTIFIQUE suas respostas.
 - a) Curso a que pertencem os alunos regularmente matriculados na UFSC em 2019.2.
 - b) Preferência declarada de um eleitor por um candidato à presidência da República.
 - c) Consumo de combustível de diferentes modelos de automóveis em km/l.
 - d) Número de filhos por família em diferentes países.
 - e) Número de residentes nos domicílios da grande Florianópolis.
 - f) Classificação da redação dos candidatos de um concurso público em “excelente”, “ótima”, “boa”, “regular” e “insuficiente”.
 - g) Classificação de pisos cerâmicos produzidos como “perfeitos”, “aceitáveis” e “inaceitáveis”.
 - h) Temperatura medida em graus Celsius.
 - i) Nível de instrução dos funcionários de uma empresa, medido pelo número de anos completos na escola (excluindo repetências).
 - j) Velocidade de automóveis em uma rodovia federal, declarada “legal” ou “em excesso”.

- 3) Para os casos a seguir indique qual é o método estatístico (tabela, gráfico, medida de síntese) mais apropriado para resumir e interpretar os dados. JUSTIFIQUE suas respostas.
 - a) A diretoria de uma empresa especializada em entrega expressa quer ter uma ideia do peso dos pacotes das encomendas, para decidir ou não a adoção de um tipo maior de caixa. Há um arquivo com os pesos em kg (com até 3 casas decimais) das últimas 3000 encomendas. Você tem 5 minutos para fazer uma apresentação para a diretoria que é IGNORANTE em Estatística.
 - b) Uma corretora de valores tem 450 clientes, desde pequenos a mega-investidores. A diretoria quer ter uma ideia dos valores investidos por estes clientes, identificando inclusive os “discrepantes”. Você tem 3 minutos para apresentar o relatório, e a diretoria é versada em Estatística.
 - c) Foram registrados os sexos (masculino e feminino) dos 851 alunos de um curso de graduação da UFSC. O objetivo é apresentar a distribuição da variável sexo para o público em geral.
 - d) Estamos avaliando a qualidade de dois processos produtivos de cerâmica, que costumam apresentar comportamentos diferentes. Um aspecto crucial é a granulometria da massa (medida em g/m^3). Queremos avaliar qual dos dois processos apresenta maior homogeneidade (ou seja, apresenta menor dispersão), dispomos de 40 medidas de cada um deles, e nós conhecemos análise exploratória de dados.
 - e) O Ministério da Saúde está conduzindo um estudo sobre o IDH em 200 municípios de Santa

Catarina. Há interesse em fazer uma análise a mais completa possível do IDH, identificando valores típicos e discrepantes. Exige-se uma apresentação gráfica para ser divulgada em um relatório que será lido por técnicos experientes em Estatística.

f) Uma empresa avícola quer avaliar os pesos (medidos por balanças de precisão) das embalagens de filé de frango, procurando identificar eventuais problemas. Você deve preparar um relatório com os dados de uma amostra de produção, 800 embalagens, que possibilite uma rápida ideia dos pesos, para apresentação aos operadores, “analfabetos em Estatística”.

g) A direção da montadora quer ter uma ideia, o mais completa possível, da distribuição da renda mensal dos respondentes em função da preferência por modelo. Todos os integrantes da diretoria têm conhecimentos avançados de Estatística, mas exigem que a apresentação seja rápida.

h) Você precisa apresentar um relatório sobre a dimensão de uma peça. Dispõe de uma amostra de apenas 50 peças, e a diretoria exige uma apresentação rápida e objetiva (sem margem de dúvida se uma disposição diferente poderia alterar as conclusões). A diretoria conhece Estatística.

i) Alguns pesquisadores precisam apresentar um relatório para a diretoria de um banco, mostrando o relacionamento entre as variáveis opinião sobre o banco atual e renda (mostrando a distribuição da renda para cada tipo de opinião) de um grupo de clientes. Há apenas 5 minutos disponíveis para a apresentação, há menos de 100 observações de renda para cada valor de opinião, e a diretoria é versada em Análise Exploratória de Dados.

j) Alguns pesquisadores precisam apresentar um relatório para a diretoria de um banco, mostrando o relacionamento entre as variáveis opinião sobre o banco atual e renda (mostrando a distribuição da renda para cada tipo de opinião) de um grupo de clientes. Há apenas 5 minutos disponíveis para a apresentação, há menos de 100 observações de renda para cada valor de opinião, e a diretoria é ignorante em Estatística.

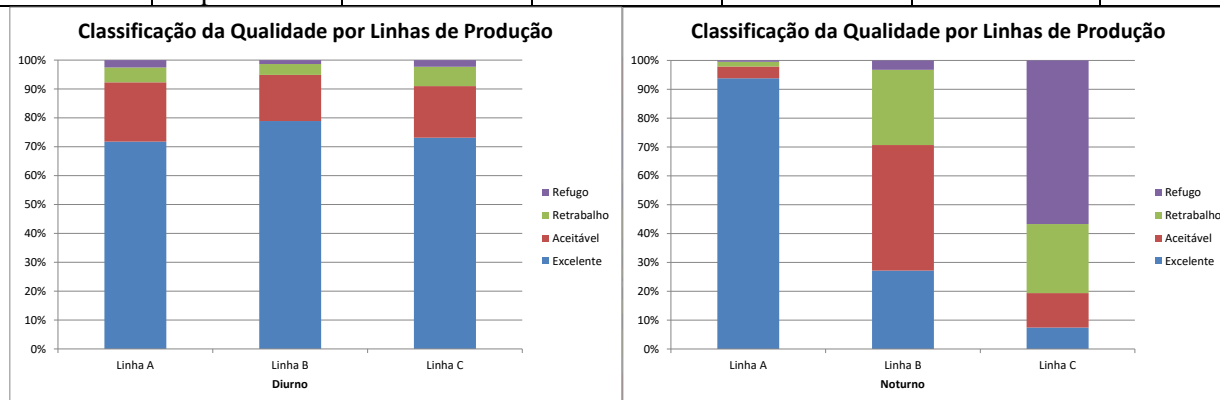
k) Alguns pesquisadores precisam apresentar um relatório para a diretoria de um banco, mostrando o relacionamento entre as variáveis opinião sobre o banco atual e renda (mostrando a distribuição da renda para cada tipo de opinião) de um grupo de clientes. Há apenas 5 minutos para a apresentação, há mais de 100 observações de renda para cada valor de opinião, e a diretoria é versada em Análise Exploratória de Dados.

l) Alguns pesquisadores precisam apresentar um relatório para a diretoria de um banco, mostrando o relacionamento entre as variáveis opinião sobre o banco atual e renda (mostrando a distribuição da renda para cada tipo de opinião) de um grupo de clientes. Há apenas 5 minutos para a apresentação, há mais de 100 observações de renda para cada valor de opinião, e a diretoria é ignorante em Estatística.

4) Certo fabricante de ferramentas dispõe de três linhas de produção (A, B e C), que são operadas em dois turnos de 8 horas (diurno e noturno). Houve reclamações sobre a qualidade dos produtos, então a direção resolveu intensificar a vigilância avaliando a qualidade (classificada como excelente, aceitável, retrabalho ou refugo) das peças produzidas nas três linhas nos dois turnos. Os resultados estão nas tabelas e gráficos a seguir.

Diurno		Qualidade				Total
Linha		Excelente	Aceitável	Retrabalho	Refugo	
A	Frequência	420	120	30	15	585
	% por linha	71,79%	20,51%	5,13%	2,56%	100,00%
	% por coluna	31,44%	37,50%	33,71%	41,67%	32,85%
B	Frequência	568	115	27	10	720
	% por linha	78,89%	15,97%	3,75%	1,39%	100,00%
	% por coluna	42,51%	35,94%	30,34%	27,78%	40,43%
C	Frequência	348	85	32	11	476
	% por linha	73,11%	17,86%	6,72%	2,31%	100,00%
	% por coluna	26,05%	26,56%	35,96%	30,56%	26,73%
Total	Frequência	1336	320	89	36	1781
	% por linha	75,01%	17,97%	5,00%	2,02%	100,00%
	% por coluna	100%	100%	100%	100%	100%

Noturno		Qualidade				Total
Linha		Excelente	Aceitável	Retrabalho	Refugo	
A	Frequência	575	25	10	3	613
	% por linha	93,80%	4,08%	1,63%	0,49%	100,00%
	% por coluna	79,31%	9,43%	4,76%	1,44%	43,54%
B	Frequência	125	200	120	15	460
	% por linha	27,17%	43,48%	26,09%	3,26%	100,00%
	% por coluna	17,24%	75,47%	57,14%	7,21%	32,67%
C	Frequência	25	40	80	190	335
	% por linha	7,46%	11,94%	23,88%	56,72%	100,00%
	% por coluna	3,45%	15,09%	38,10%	91,35%	23,79%
Total	Frequência	725	265	210	208	1408
	% por linha	51,49%	18,82%	14,91%	14,77%	100,00%
	% por coluna	100%	100%	100%	100%	100%



- a) Qual é a qualidade predominante no turno diurno? E no noturno? JUSTIFIQUE.
- b) O ideal era a produção total distribuir-se igualmente entre as três linhas de produção. Isso ocorre no diurno? E no noturno? JUSTIFIQUE.
- c) Existe associação entre a qualidade das peças e a linha de produção no diurno? E no noturno? JUSTIFIQUE.

5) Você é o responsável pela confecção de um relatório sobre o desempenho de vendas das 22 filiais da empresa em que trabalha. A alta direção investiu muito em uma série de promoções para aumentar as vendas, e quer verificar se isso realmente ocorreu: havia sido estabelecido um valor *mínimo aceitável* de 10%, em relação ao mesmo período do ano passado. Os valores de crescimento percentual das vendas, estão abaixo:

4.30 6.34 7.38 3.51 10.90 0.28 7.96 8.12 1.58 5.53 4.11 8.43 2.95 8.54
 18.76 7.34 0.41 8.31 1.54 4.49 4.65 4.19 -1.28 5.67 -0.25 12.00 4.32 1.82
 -3.15 9.03 2.31 0.17

- a) Construa a tabela de frequências que julgar apropriada para resumir os dados. Com base na tabela o que você conclui acerca da eficácia das promoções?
- b) Que ferramenta gráfica você utilizaria para resumir os dados? Por quê?
- c) Construa o gráfico escolhido no item b, e com base nele, o que você conclui acerca da eficácia das promoções?
- d) Que medida(s) de síntese você utilizaria para resumir os dados? Por quê?
- e) Calcule a(s) medida(s) escolhida(s) no item d, e com base nela(s), o que você conclui acerca da eficácia das promoções?

6) A padaria que o seu bisavô fundou ao chegar de Portugal, apesar de ainda próspera e com 5 filiais, tem um número talvez excessivo de produtos, o que está aumentando em demasia os custos. Após acalorados debates, sua família decidiu que cinco tipos de pães (aqui chamados de A, B, C, D e E) poderiam ser retirados de produção: mas apenas um deles. Seria arriscado retirar todos, pois a empresa poderia perder a variedade que é sua grande vantagem competitiva. Há interesse na

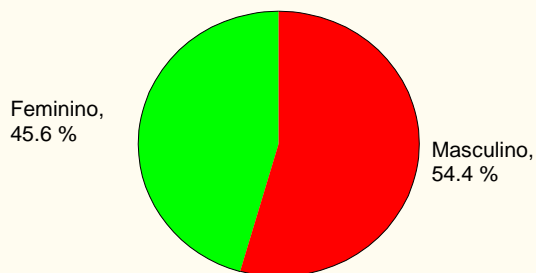
racionalização do processo, mantendo os produtos que têm demanda mais homogênea ao longo do ano. Como você está estudando Estatística neste momento, decidiram que você fará a análise dos dados, para orientar a escolha. Na tabela abaixo você tem as vendas, em milhares de reais, de cada produto nos últimos doze meses.

Pão	Mês 1	Mês 2	Mês 3	Mês 4	Mês 5	Mês 6	Mês 7	Mês 8	Mês 9	Mês10	Mês11	Mês12
A	19.7	15.4	19.3	17.9	19.7	19.5	21.7	19.9	18.6	20.0	17.4	20.5
B	18.7	17.8	18.3	19.4	19.6	16.3	18.4	19.5	19.2	19.7	21.4	20.9
C	20.2	19.8	18.2	19.6	21.3	10.3	12.6	5.5	18.5	22.2	18.7	20.6
D	5.0	4.5	6.7	20.1	19.9	22.4	20.9	19.6	19.6	18.3	8.0	7.0
E	21.2	18.9	19.2	19.6	17.2	19.3	21.9	15.7	20.0	19.3	20.8	18.9

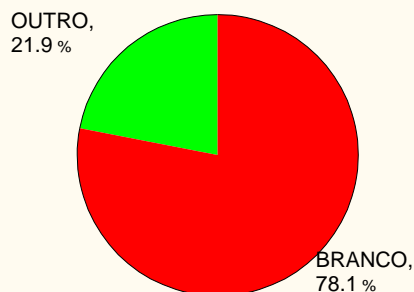
- Qual a ferramenta gráfica que você utilizaria para descrever os dados? Por quê?
- Utilizando a ferramenta escolhida no item a, que produto você recomenda retirar de produção?
- Que medida(s) de síntese você utilizaria para resumir os dados? Por quê?
- Calcule a(s) medida(s) escolhida(s) no item c, e com base nela(s) indique qual produto você recomenda retirar de produção.

7) Você agora faz parte do corpo funcional de uma multinacional, que decidiu por sua transferência para uma unidade no sul dos EUA. Lideranças de direitos civis locais acusam a empresa de não proporcionar as mesmas oportunidades à mulheres e minorias raciais (a região tem 50 % de mulheres e 30% de minorias). A direção da unidade nega o preconceito, mas admite que não tem uma ideia completa do problema. Você recebeu a tarefa de não somente coletar e organizar os dados, mas também interpretá-los. Os gráficos abaixo foram elaborados:

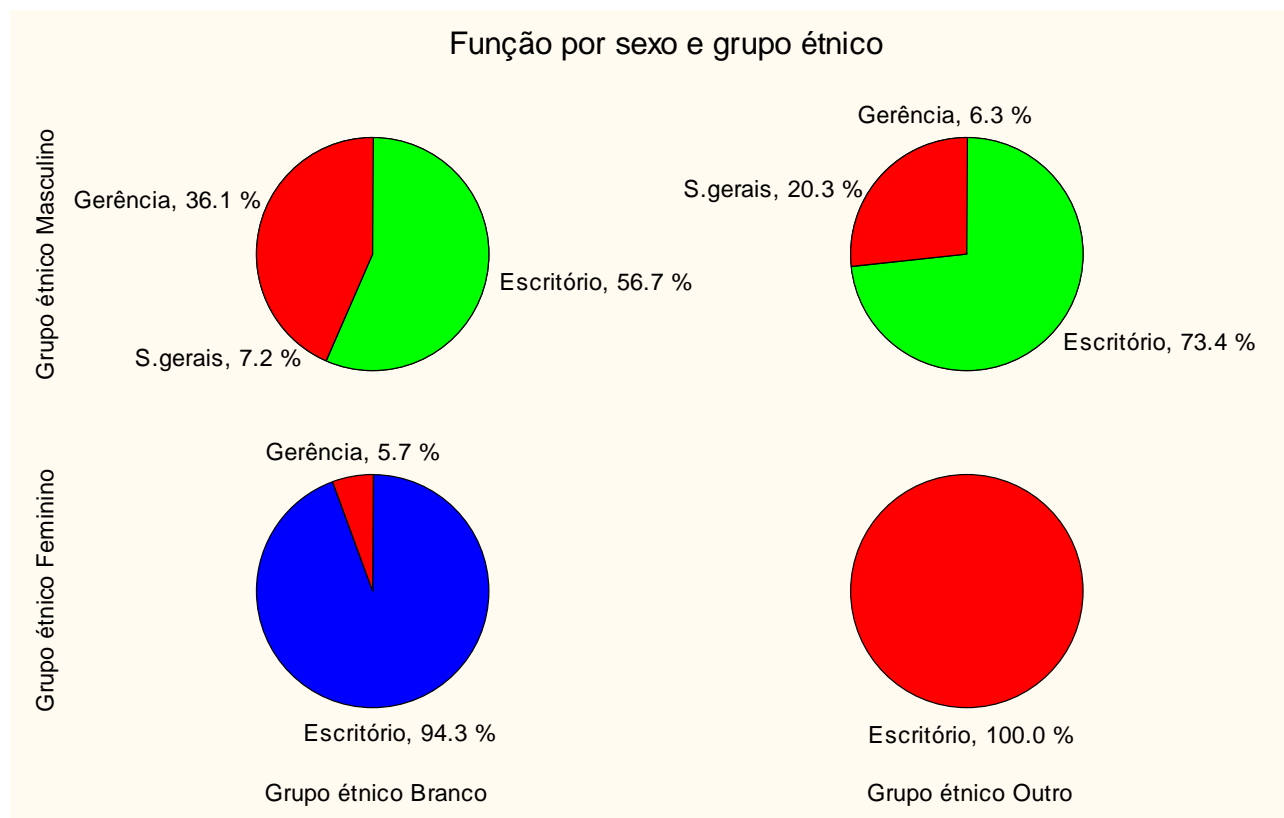
Distribuição dos funcionários por Sexo



Distribuição de funcionários por grupo étnico



- Com base nos gráficos há discriminação quanto ao sexo, baseando-se unicamente no percentual de funcionários em cada categoria? Por quê?
- Com base nos gráficos há discriminação devido ao grupo étnico, baseando-se unicamente no percentual de funcionários em cada categoria? Por quê?
- Homens e mulheres têm as mesmas oportunidades na organização? Por quê?
- Pessoas de diferentes grupos étnicos têm as mesmas oportunidades na organização? Por quê?
- O que você recomenda que a empresa faça?



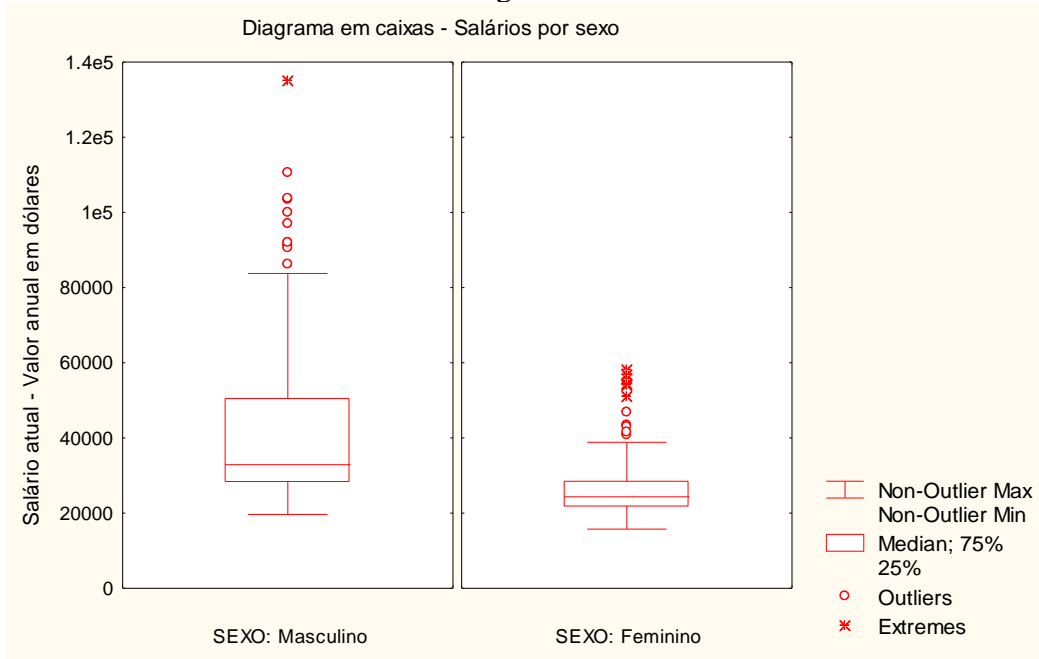
8) As tabelas a seguir apresentam: distribuição de frequências agrupada em classes dos gastos diários de turistas (em R\$) por 3 balneários brasileiros (A, B e C); medidas de síntese dos gastos diários de turistas (em R\$) para os mesmos dados da distribuição de frequências. Com base nestas tabelas responda os itens a seguir, JUSTIFICANDO suas respostas.

Gastos diários (R\$)	Balneário						Total	
	A		B		C			
	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Classes								
92 -114	0	0,00%	0	0,00%	2	0,80%	2	0,27%
114 -136	0	0,00%	0	0,00%	19	7,60%	19	2,53%
136 -158	0	0,00%	0	0,00%	35	14,00%	35	4,67%
158 -180	1	0,40%	6	2,40%	64	25,60%	71	9,47%
180 -202	6	2,40%	131	52,40%	66	26,40%	203	27,07%
202 -224	44	17,60%	112	44,80%	48	19,20%	204	27,20%
224 -246	97	38,80%	1	0,40%	11	4,40%	109	14,53%
246 -268	82	32,80%	0	0,00%	5	2,00%	87	11,60%
268 -290	19	7,60%	0	0,00%	0	0,00%	19	2,53%
290 -312	1	0,40%	0	0,00%	0	0,00%	1	0,13%
Total	250	100,0%	250	100,0%	250	100,0%	750	100,0%

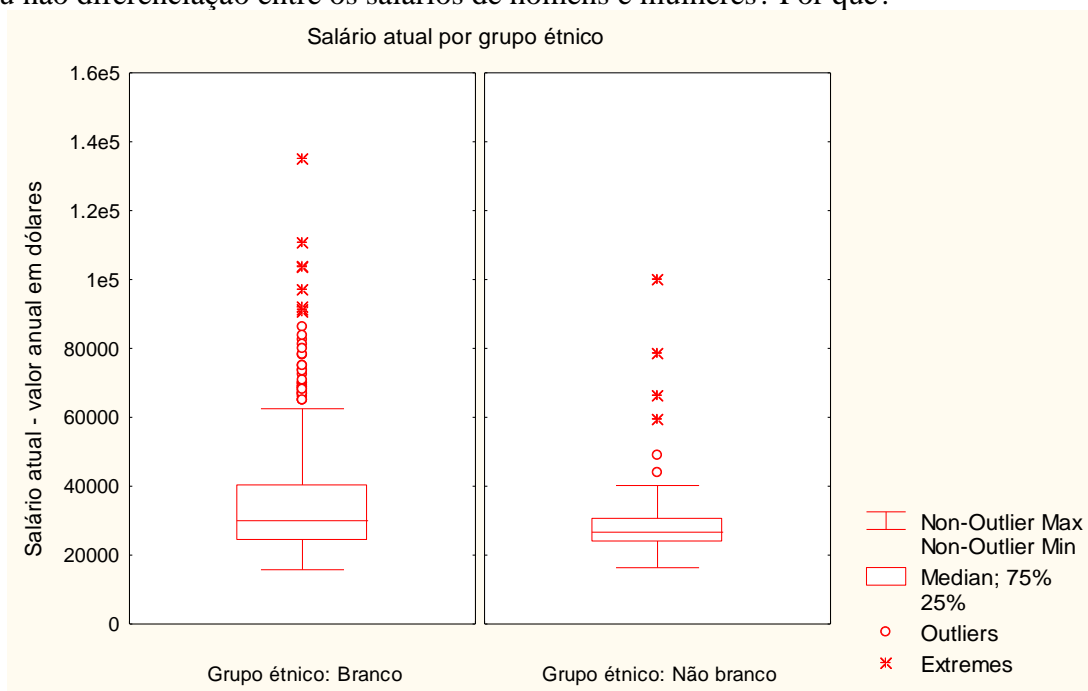
Medidas de Gastos (R\$)	Balneário			Total
	A	B	C	
Média	240,37	200,12	181,32	207,27
Mediana	238,67	200,48	181,86	204,95
Desvio padrão	20,07	9,71	30,63	32,93
CV%	8,41%	4,84%	16,84%	16,07%
Quartil inferior	226,65	193,95	162,25	190,43
Quartil superior	254,61	206,44	203,19	228,36

- a) APENAS PELA DISTRIBUIÇÃO AGRUPADA EM CLASSES, há evidência de relação entre as variáveis Gasto diário e Balneário? Use percentuais na justificativa.
- b) APENAS PELAS MEDIDAS DE SÍNTESE, analise as medidas média, mediana e quartis para o TOTAL de balneários. Caracterize a tendência central dos gastos diários.
- c) APENAS PELAS MEDIDAS DE SÍNTESE, em qual dos balneários os gastos diários são mais homogêneos em torno da média?

9) Ainda o caso da pesquisa do problema 7. As lideranças de direitos civis locais também se queixam que os salários pagos a homens e mulheres, brancos e não brancos, são diferentes, mesmo exercendo a mesma função (os dos homens e dos brancos seriam maiores). Novamente a direção da unidade nega o preconceito, e você recebeu a tarefa de avaliar os dados. Você tem os gráficos e as tabelas com as medidas de síntese descritas a seguir.



- a) Analise o diagrama em caixas múltiplo acima em termos de valor típico, assimetria, dispersão e valores discrepantes.
- b) Há ou não diferenciação entre os salários de homens e mulheres? Por quê?



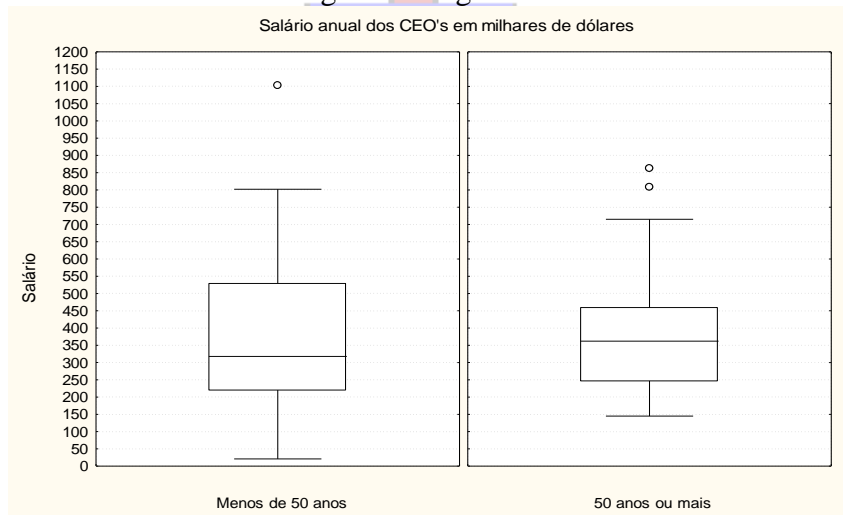
- c) Analise o diagrama em caixas múltiplo acima em termos de valor típico, assimetria, dispersão e valores discrepantes.
 d) Há ou não diferenciação entre os salários devido ao grupo étnico? Por quê?

10) Os dados a seguir referem-se aos salários anuais (em dólares) dos diretores executivos de 60 “pequenas” empresas americanas (com vendas anuais entre 5 e 350 milhões de dólares). Os executivos foram classificados em dois grupos de acordo com a sua idade: menos de 50 anos, 50 anos ou mais. Responda as questões apresentadas, JUSTIFICANDO suas respostas

Idades	Medidas de síntese para salários anuais (milhares de dólares)		
	Número	Média	D. Padrão
Menos de 50	35	382,2	219,98
50 ou mais	25	398,61	197,92

- a) Com base apenas na tabela acima, qual das faixas de idade tem os maiores salários?
 b) Com base apenas na tabela acima, e em outras medidas calculadas a partir dela, qual das faixas de idade tem os salários mais homogêneos?

11) Ainda os dados do Exercício 10. A seguir os diagramas em caixa dos salários dos executivos.



- a) Analise os diagramas em caixa quanto ao valor típico, assimetria, dispersão e valores discrepantes.
 b) Com base no diagrama em caixas qual das faixas de idade está ganhando melhor?