

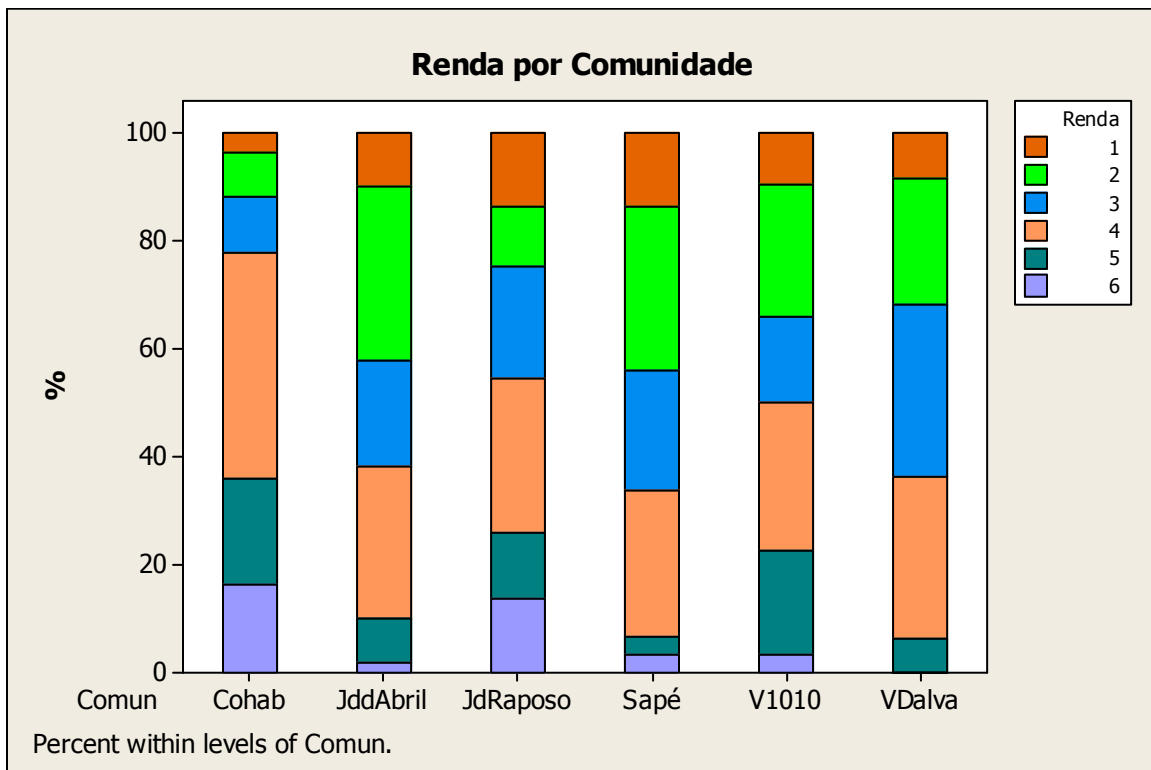
Exploração Banco de dados: aeusp

Questão 1):

- Não estão disponíveis os últimos valores das variáveis Ttrab e ltrab;
- Há valor não relacionado para a variável Trab;
- etc.

recomendado cruzar valores das variáveis Trab e Ttrab, Serief com Idade

Questão 2)



Tabulated statistics: Comun; Renda

Rows: Comun	Columns: Renda						
	1	2	3	4	5	6	All
Cohab	3,49	8,14	10,47	41,86	19,77	16,28	100,00
JddAbril	5,00	16,00	20,00	28,00	8,00	2,00	100,00
JdRaposo	13,58	11,11	20,99	28,40	12,35	13,58	100,00
Sapé	13,56	30,51	22,03	27,12	3,39	3,39	100,00
V1010	6	15	10	17	12	2	62

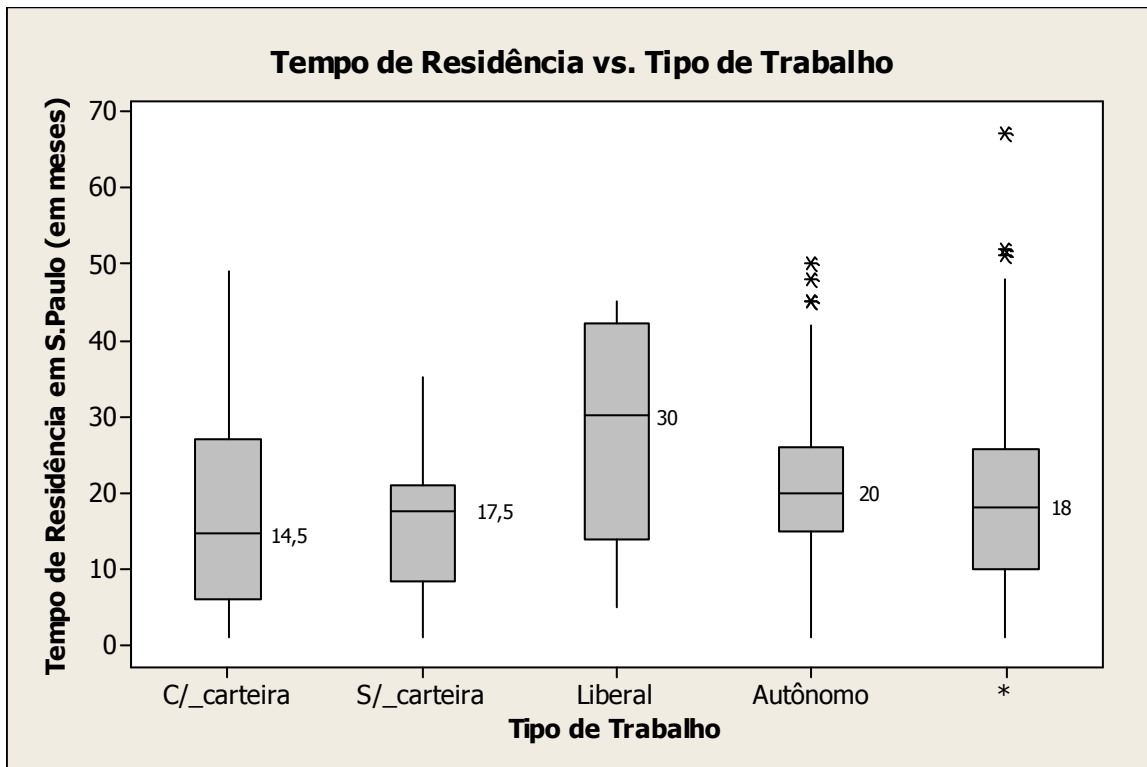
	9,68	24,19	16,13	27,42	19,35	3,23	100,00
VDalva	4	11	15	14	3	0	47
	8,51	23,40	31,91	29,79	6,38	0,00	100,00
All	37	76	74	120	48	30	385
	9,61	19,74	19,22	31,17	12,47	7,79	100,00

Cell Contents: Count
 % of Row

Comentário:

Cohab: 77,91 % nas classes superiores de renda (4, 5 e 6)
 Jardim d'Abril: 62,00 % nas classes inferiores de renda (1, 2 e 3)
 Os dados amostrais mostram evidências de diferenças de renda entre os moradores das duas comunidades

Questão 3)



Descriptive Statistics: Temposp

Variable	Ttrab	N	N*	Mean	StDev	Minimum	Q1	Median	Q3
Temposp	1	78	0	17,42	13,32	1,00	6,00	14,50	27,00
	2	56	0	15,77	8,88	1,00	8,25	17,50	21,00
	3	12	0	28,08	13,69	5,00	13,75	30,00	42,25
	4	95	0	21,01	10,85	1,00	15,00	20,00	26,00
	*	144	0	19,056	11,743	1,000	10,000	18,000	25,750

Variable	Ttrab	Maximum
Temposp	1	49,00

2	35,00
3	45,00
4	50,00
*	67,000

Comentários:

Verifica-se que o tempo mediano do profissional liberal é bastante diferenciado dos demais, embora apresente maior dispersão

As demais profissões apresentam tempos medianos semelhantes com mesmo nível de dispersão, com exceção dos sem carteira, que apresentam uma maior concentração. Pelo coeficiente de variação podemos avaliar melhor esta questão:

Descriptive Statistics: Temposp

Variable	Ttrab	CoefVar
Temposp	1	76,48
	2	56,30
	3	48,76
	4	51,65
	*	61,63

e os níveis de variação mostram bastante semelhantes, com exceção do grupo com carteira assinada (1) que tem uma dispersão maior em comparação com sua média. Salienta-se entretanto que há muitos dados perdidos (144), representando 37,4 % do total.

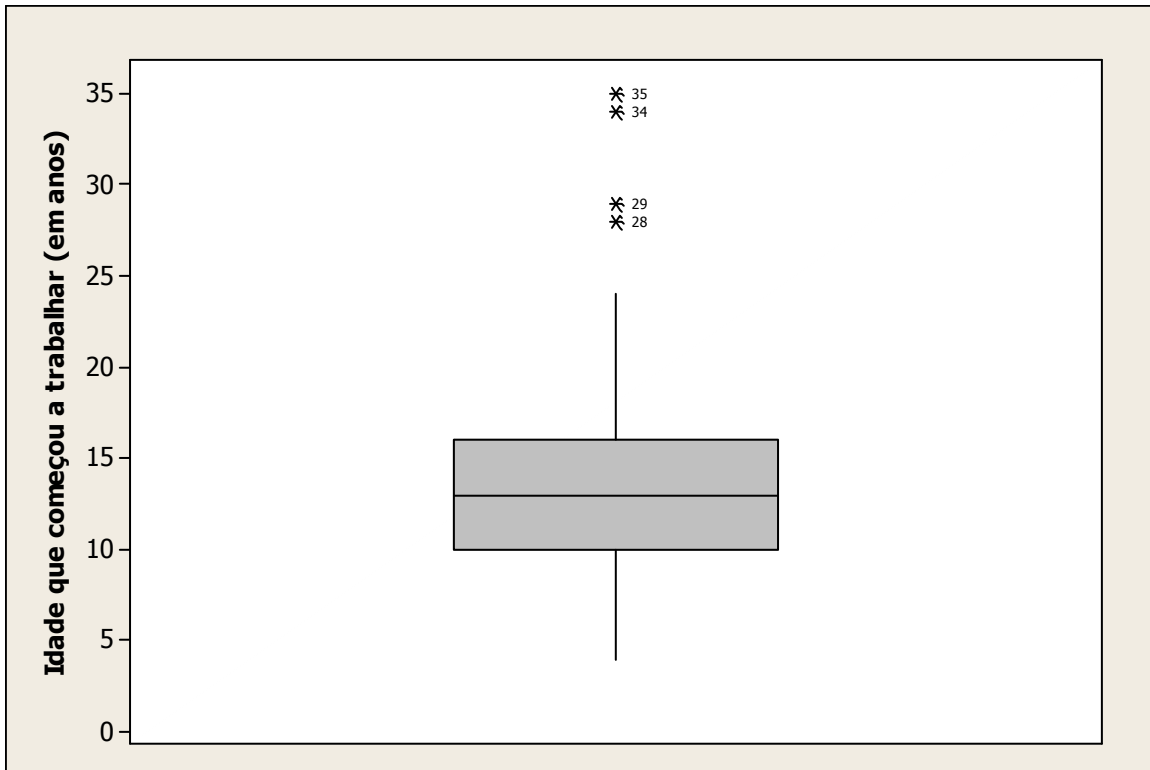
Isso pode distorcer a análise.

Por outro lado, verificando as variabilidades de Temposp por grupos e total:

Level	N	Mean	StDev
1	78	17,42	13,32
2	56	15,77	8,88
3	12	28,08	13,69
4	95	21,01	10,85
Total	385	19,01	11,75

podemos perceber que elas estão muito próximas, indicando que a variável Temposp não é bem explicada pela variável Ttrab

Questão 4)

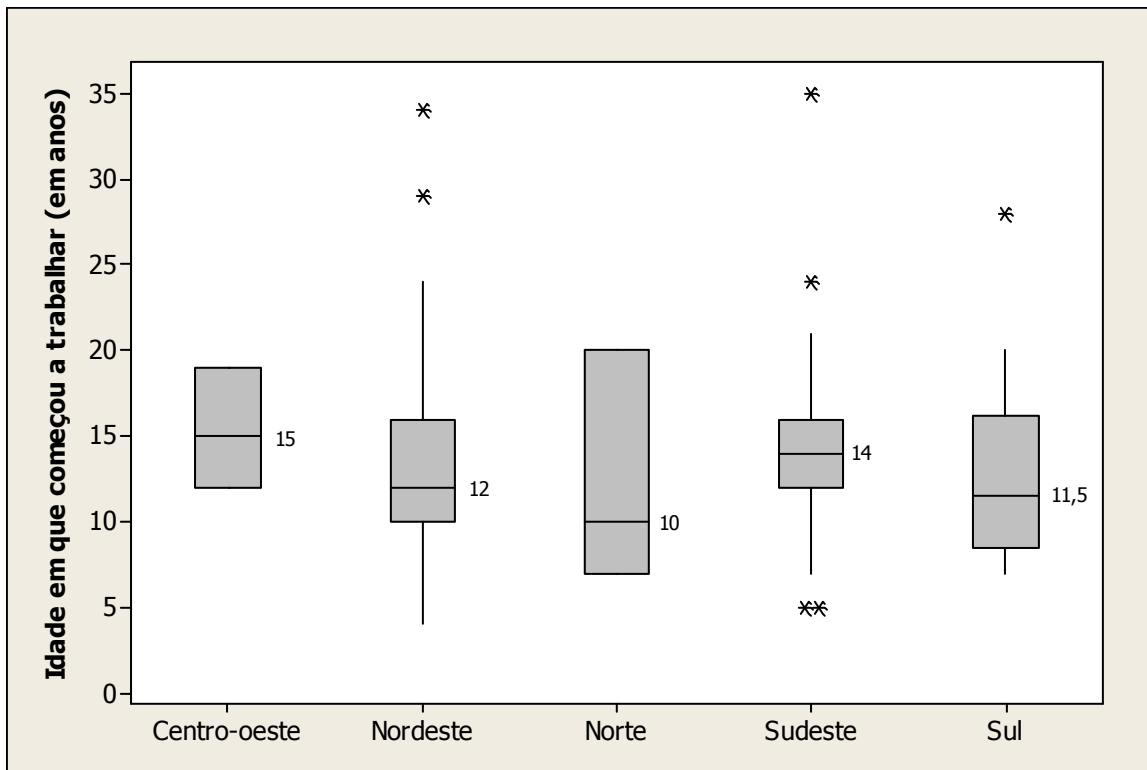


Descriptive Statistics: Itrab

Variable	Minimum	Q1	Median	Q3	Maximum
Itrab	4,000	10,000	13,000	16,000	35,000

Há outliers com idade de início de trabalhar acima de 28 anos

Um exame da distribuição através do resumo do 5 números remete a indagações sobre a consistência dos dados, pois 25 % dos indivíduos da amostra começaram a trabalhar antes dos 10, com a menor idade sendo de 4 anos. Há 33 dados perdidos (8,6% dos dados).



Analisando as variabilidades têm-se:

Level	N	Mean	StDev
Centro-oeste	3	15,333	3,512
Nordeste	180	13,000	4,519
Norte	3	12,333	6,807
Sudeste	156	13,853	3,794
Sul	10	13,100	6,506
Total	352	13,395	4,291

A região de procedência aparenta explicar pouco a idade de início de trabalho, embora haja diferenças nos padrões dos primeiros quartis em todas as regiões.

Questão (5)

Descriptive Statistics: Itrab

Variable	Mean	StDev	CoefVar	Minimum	Q1	Median	Q3	Maximum
Itrab	13,395	4,291	32,03	4,000	10,000	13,000	16,000	35,000

Descriptive Statistics: Itrab

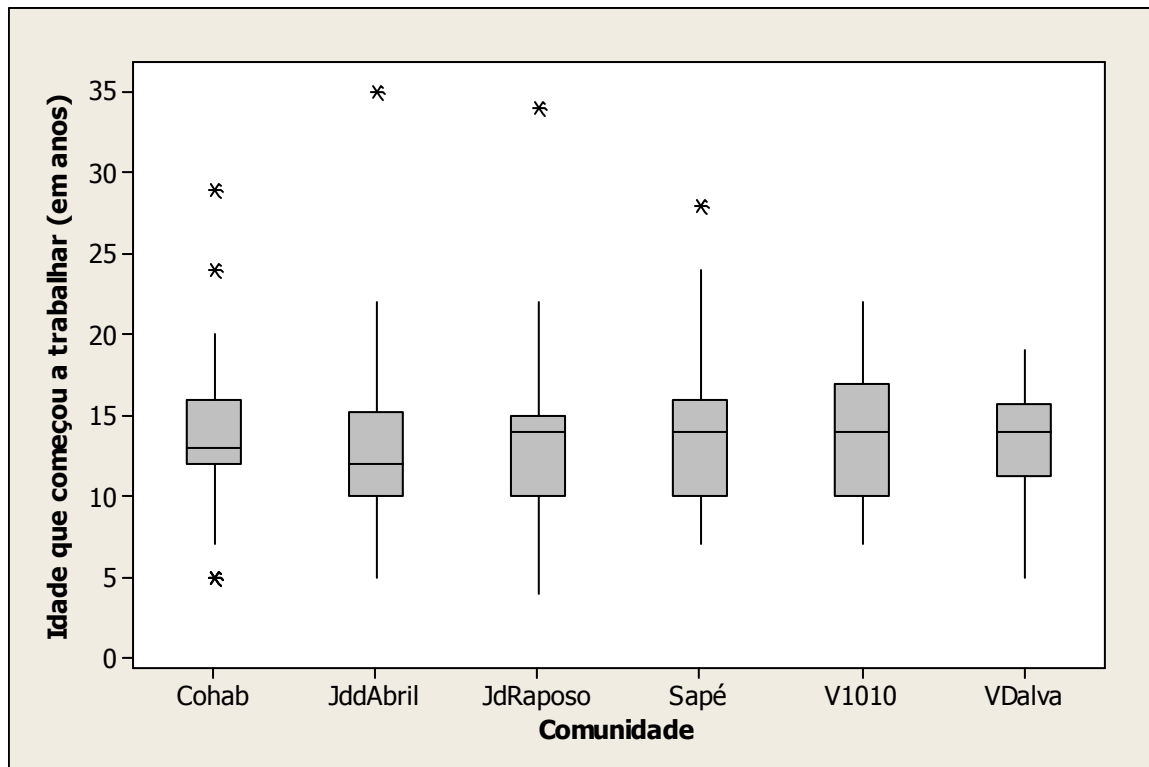
Variable	Comun	Mean	StDev	CoefVar	Minimum	Q1	Median	Q3
Itrab	Cohab	13,649	4,220	30,92	5,000	12,000	13,000	16,000
	JddAbril	13,022	4,851	37,25	5,000	10,000	12,000	15,250
	JdRaposo	13,274	4,565	34,39	4,000	10,000	14,000	15,000
	Sapé	13,474	4,214	31,27	7,000	10,000	14,000	16,000
	V1010	13,339	4,326	32,43	7,000	10,000	14,000	17,000
	VDalva	13,525	3,419	25,28	5,000	11,250	14,000	15,750

Variable	Comun	Maximum
Itrab	Cohab	16,000
	JddAbril	15,250
	JdRaposo	15,000
	Sapé	16,000
	V1010	17,000
	VDalva	15,750

Itrab	Cohab	29,000
	JddAbril	35,000
	JdRaposo	34,000
	Sapé	28,000
	V1010	22,000
	VDalva	19,000

As distribuições são razoavelmente homogêneas em termos de médias medianas e dispersão (quartis também). por outro lado, analisando as variabilidades em cada comunidade, verificamos que essas (comunidades) não ajudam a explicar a variabilidade de Itrab.

Um box-plot facilita evidenciar a homogeneidade das situações da idade de início de trabalho.



Não podemos calcular diretamente as estatísticas descritivas para renda, já que a mesma se encontra categorizada, mas já verificamos as diferenças na resposta à questão (2). Podemos usar uma aproximação das rendas médias:

Tabulated statistics: Renda; Comun

Rows: Renda Columns: Comun

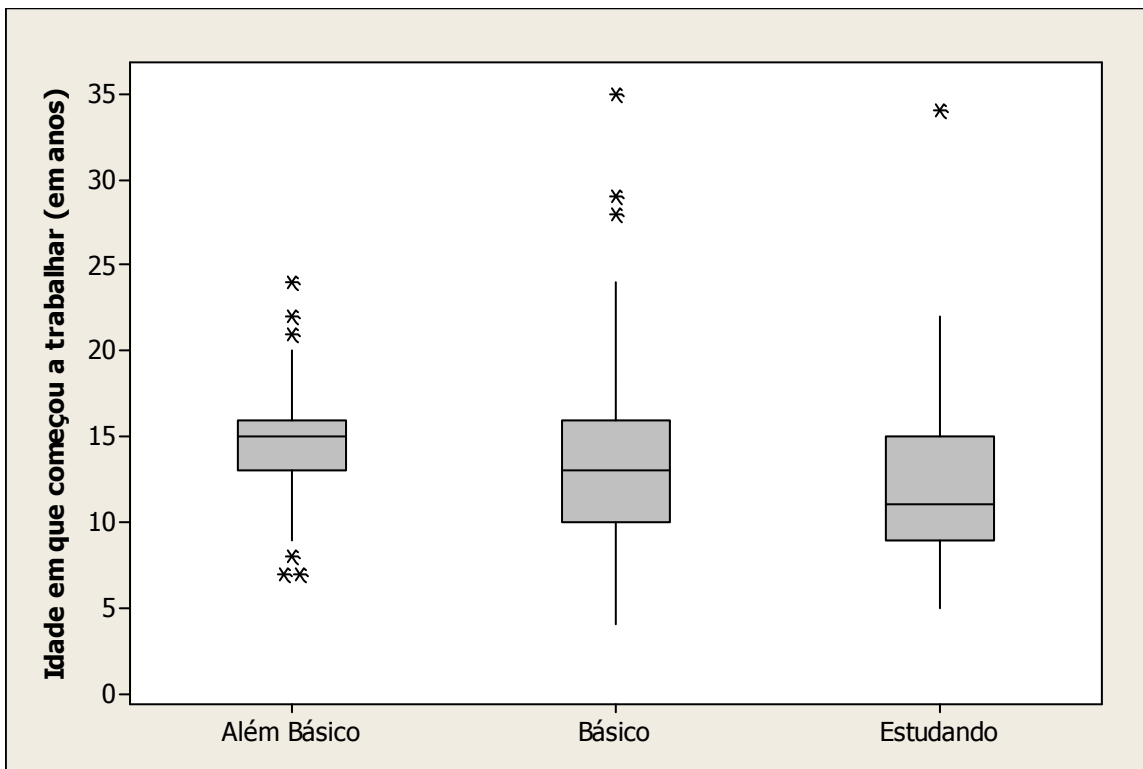
	Cohab	JddAbril	JdRaposo	Sapé	V1010	VDalva	All
1	3	5	11	8	6	4	37
2	7	16	9	18	15	11	76
3	9	10	17	13	10	15	74
4	36	14	23	16	17	14	120
5	17	4	10	2	12	3	48
6	14	1	11	2	2	0	30

Considerando-se o ponto médio para cada classe de renda, exceto para a última classe em que se considera o menor valor (resultado conservativo):

Classes de Renda	Pto Médio	Cohab	JddAbril	JdRaposo	Sapé	V1010	VDalva	All
1	75	3	5	11	8	6	4	3
2	225	7	16	9	18	15	11	7
3	375	9	10	17	13	10	15	7
4	675	36	14	23	16	17	14	12
5	1200	17	4	10	2	12	3	4
6	1500	14	1	11	2	2	0	3
Total		86	50	81	59	62	47	38
Total médio		824,13	469,50	657,41	436,02	587,90	456,38	600,5

Há diferenças em termos de rendas médias entre as comunidades.

Questão (6)



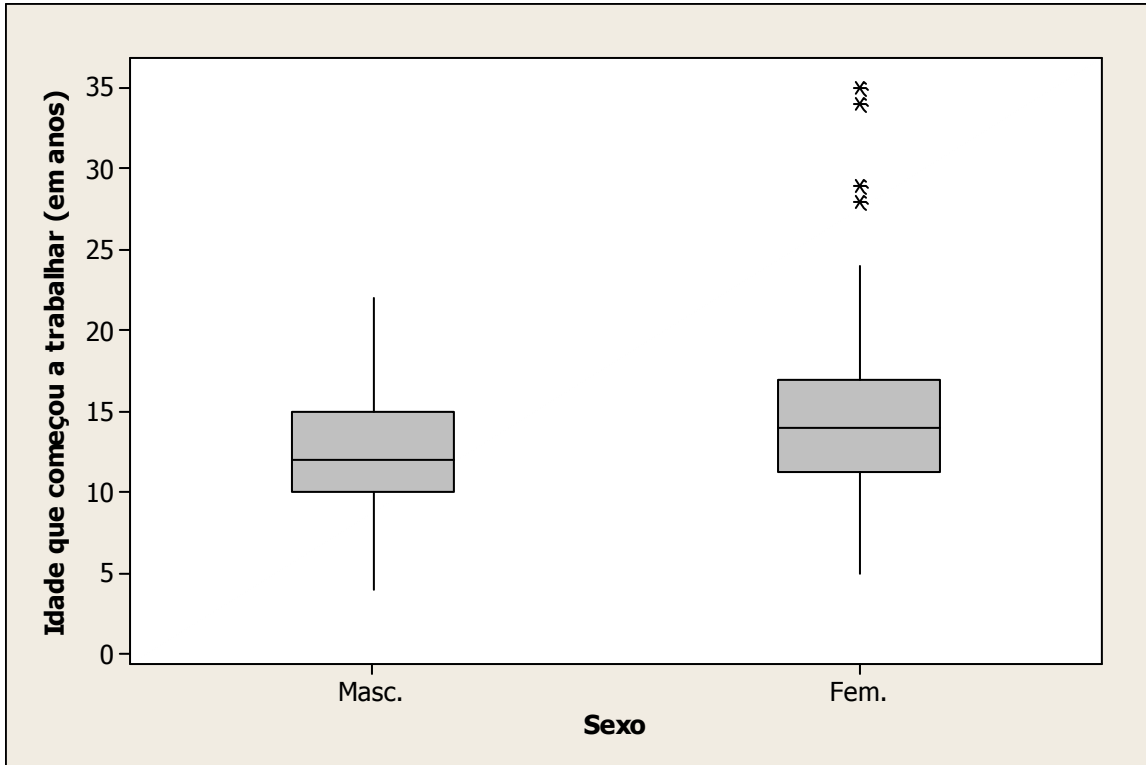
Descriptive Statistics: Itrab

Variable	Cat_serie	N	Mean	StDev	Median	IQR
Itrab	Além_Basico	75	14,773	3,315	15,000	3,000
	Basico	225	13,160	4,297	13,000	6,000
	Estudando	52	12,423	5,066	11,000	6,000

Comentário:

As medianas estão próximas. A variabilidade da categoria além básico, apresenta menor variabilidade (característica acentuada quando se verifica a distância interquartílica), embora todos apresentem variabilidade muito próxima da total (4,291).

Questão (7)



Descriptive Statistics: Itrab

Variable	Sexo	Mean	StDev	CoefVar	Minimum	Q1	Median	Q3
Itrab	1	12,500	3,725	29,80	4,000	10,000	12,000	15,000
	2	14,107	4,578	32,45	5,000	11,250	14,000	17,000

Variable	Sexo	Maximum
Itrab	1	22,000
	2	35,000

Sim, há evidências de que começam a trabalhar mais cedo. Há um deslocamento de aproximadamente 2 anos nas medidas (quartis, média e mediana). Não há muita diferença nas dispersões que são pequenas, ou seja, começam a trabalhar muito próximo à média. No caso dos homens, 50% começou a trabalhar entre os 10 e 15 anos. Há ligeira assimetria à direita (maiores valores), mais acentuada no caso das mulheres. Entretanto, considerada a variabilidade nos grupos, verifica-se que a categorização por sexo pouco explica a variabilidade total.

Questão (8)

A tabela de dupla entrada entre Comunidade e Renda encontra-se na resposta à questão

(ii). Mesmo sem identificar sua intensidade, aparenta haver uma associação entre as duas variáveis, embora não tenhamos identificado.

Tabulated statistics: Renda; Reproce

Rows: Renda Columns: Reproce

	Centro-oeste	Nordeste	Norte	Sudeste	Sul	All
1	0	20	0	16	1	37
2	1	46	0	28	1	76
3	0	38	1	33	2	74
4	1	52	0	63	4	120
5	0	22	1	24	1	48
6	1	12	1	14	2	30
All	3	190	3	178	11	385

Cell Contents: Count

Tabulated statistics: Renda; Reproce

Rows: Renda Columns: Reproce

	Centro-oeste	Nordeste	Norte	Sudeste	Sul	All
1	0,00	10,53	0,00	8,99	9,09	9,61
2	33,33	24,21	0,00	15,73	9,09	19,74
3	0,00	20,00	33,33	18,54	18,18	19,22
4	33,33	27,37	0,00	35,39	36,36	31,17
5	0,00	11,58	33,33	13,48	9,09	12,47
6	33,33	6,32	33,33	7,87	18,18	7,79
All	100,00	100,00	100,00	100,00	100,00	100,00

Cell Contents: % of Column

Poderíamos calcular a renda média pro região de procedência através do mesmo procedimento efetuado anteriormente. Considerando a participação dos maiores níveis de renda (5 e 6), verificamos o seguinte:

	Centro-oeste	Nordeste	Norte	Sudeste	Sul	All
1 a 4	66,64	82,10	66,64	78,65	72,73	79,74
5 e 6	33,33	17,90	33,33	21,35	27,27	20,26
All	100,00	100,00	100,00	100,00	100,00	100,00

Considerados os maiores tamanhos amostrais (NE e SE) verifica-se muito pouca diferença na distribuição através desse corte (sem critério técnico). Aparentemente a região de procedência não explica bem a renda.

Tabulated statistics: Renda; Trab

Rows: Renda Columns: Trab

	1	2	3	4	All
1	10	21	6	0	37
2	43	29	4	0	76
3	48	24	1	1	74
4	77	41	2	0	120

5	40	6	2	0	48
6	20	10	0	0	30
All	238	131	15	1	385

Cell Contents: Count

Tabulated statistics: Renda; Trab

Rows: Renda Columns: Trab

	1	2	3	4	All
1	4,20	16,03	40,00	0,00	9,61
2	18,07	22,14	26,67	0,00	19,74
3	20,17	18,32	6,67	100,00	19,22
4	32,35	31,30	13,33	0,00	31,17
5	16,81	4,58	13,33	0,00	12,47
6	8,40	7,63	0,00	0,00	7,79
All	100,00	100,00	100,00	100,00	100,00

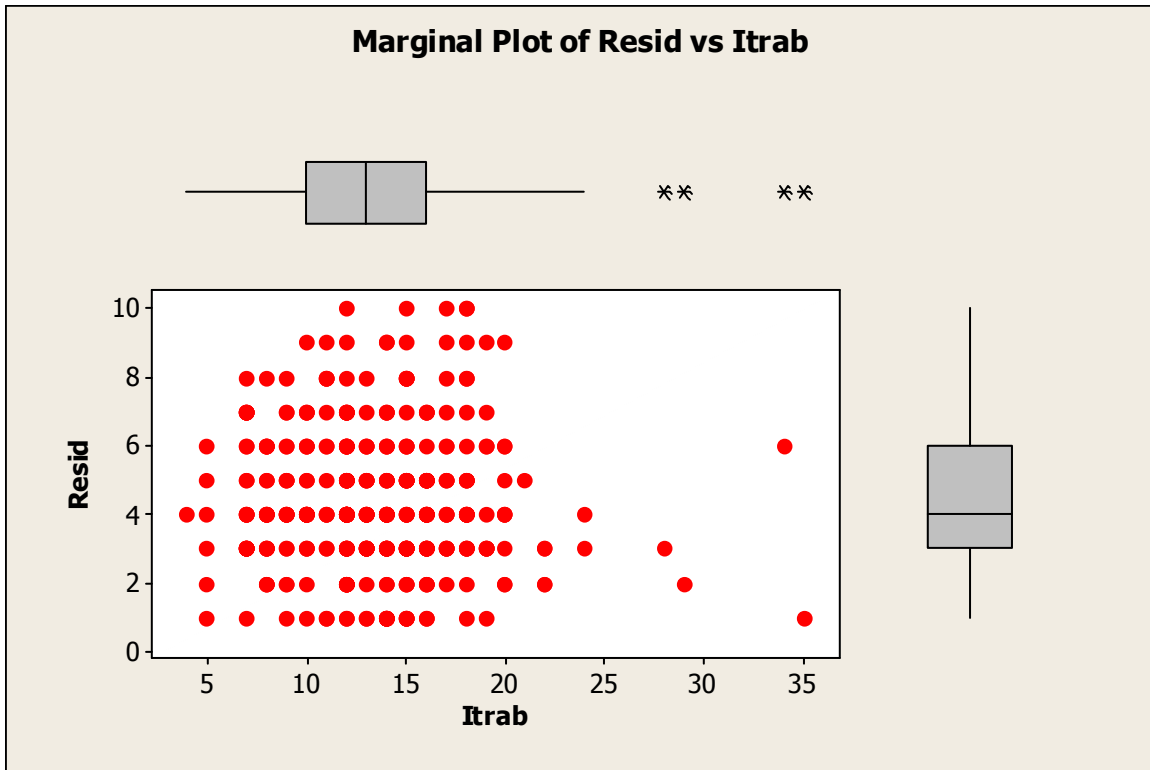
Cell Contents: % of Column

Poderíamos calcular a renda média pro região de procedência através do mesmo procedimento efetuado anteriormente. Usando a mesma consideração para aferição da participação dos maiores níveis de renda (5 e 6), verificamos o seguinte:

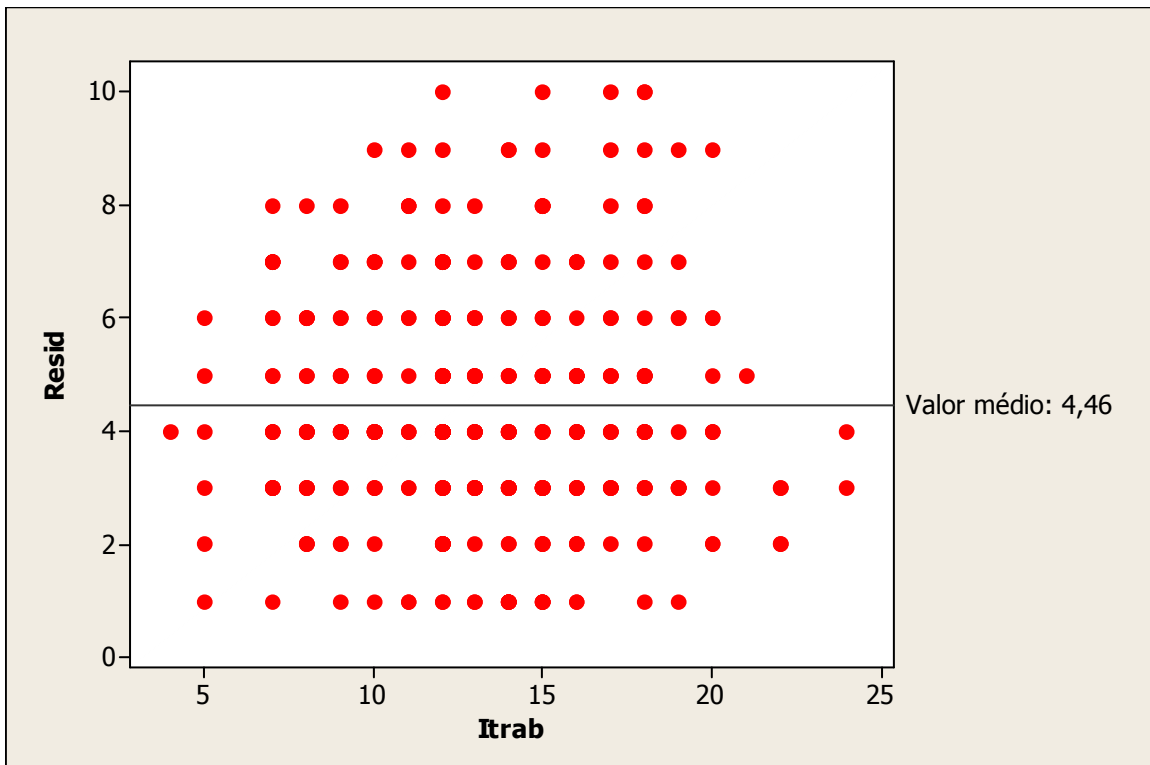
	1	2	3	4	All
1 a 4	74,79	87,79	86,67	100,00	79,74
5 e 6	25,21	12,21	13,33	0,00	20,26
All	100,00	100,00	100,00	100,00	100,00

Considerados os maiores tamanhos amostrais (com e sem trabalho), como esperado, verifica-se alguma diferença na distribuição dos valores através desse corte (sem critério técnico). Há pouca diferença no padrão da distribuição de rendas entre os sem-trabalho e os aposentados. Salienta-se que o valor 4 é inconsistente com o informado.

Questão (9)

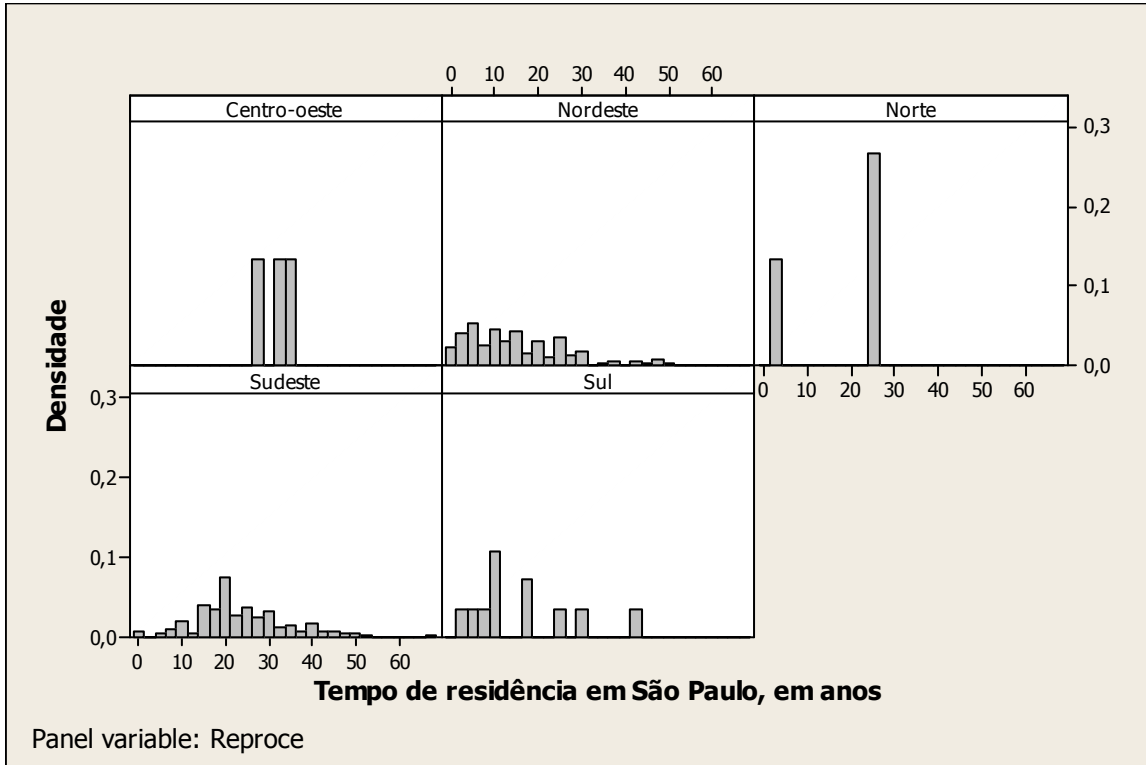


A variável Itrab tem uma distribuição aparentemente simétrica e a distribuição de Resid mostra-se com aproximadamente o mesmo padrão para quase todos os valores de Itrab. Há 4 pontos atípicos identificados como os indivíduos das linhas: 24, 150, 196 e 345. Abaixo o plot com a exclusão desses pontos:



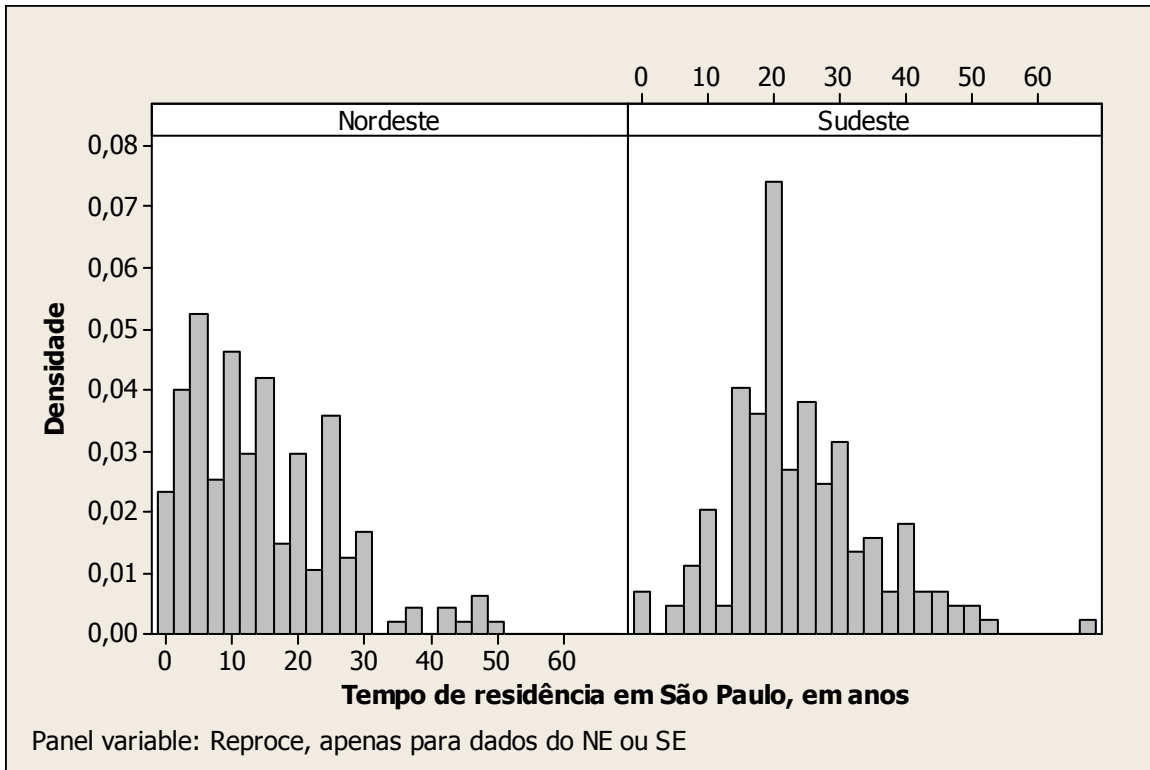
Considerado Itrab, Resid está aproximadamente distribuído em torno de sua média, ou seja, Resid não aparenta estar associado com Itrab.

Questão (10)

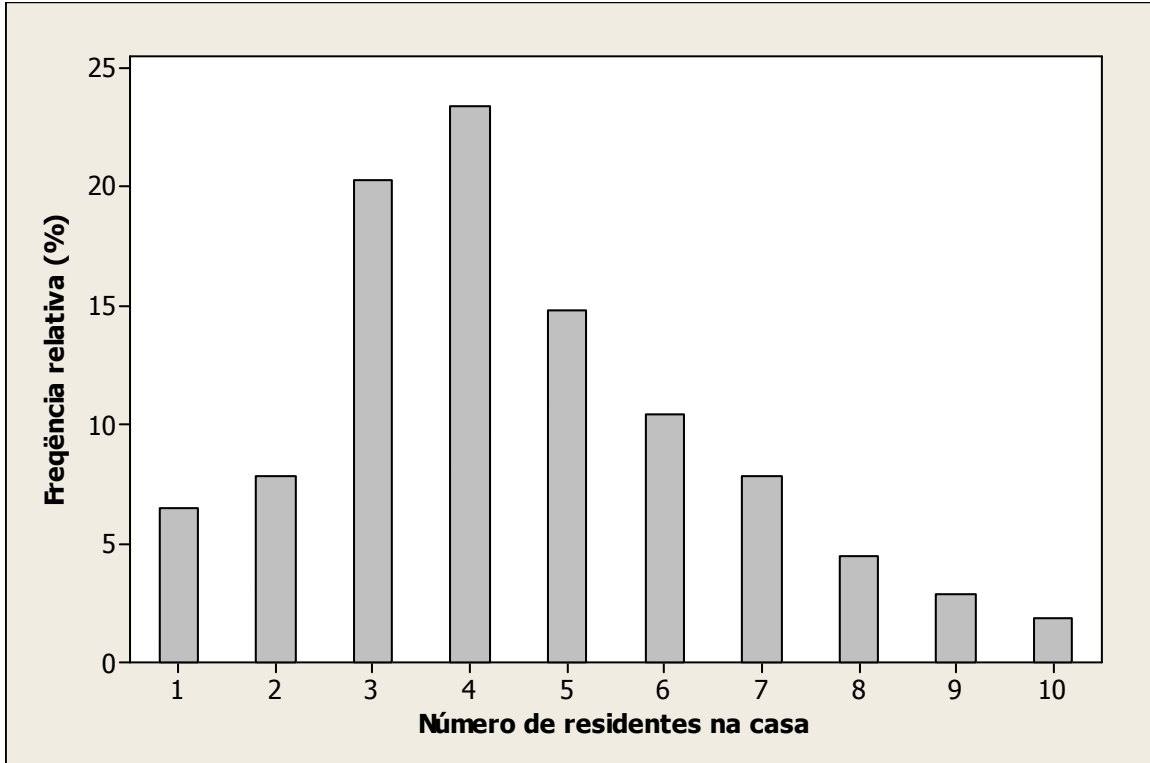


Comentário:

Considerados as duas regiões de procedência com maiores quantidade de observações, verificamos que há uma diferenciação bastante razoável: a região NE apresenta uma maior concentração de freqüência no menores valores de Temposp (até 20 anos, por exemplo), em comparação com os oriundos da região SE, por outro lado, a distribuição para o NE é marcadamente assimétrica, enquanto que para o SE é razoavelmente simétrica. Pode-se verificar melhor esse fato analisando-se apenas as observações para NE ou SE.



Questão (11)



Questão (12)

Indivíduos com idade inferior a 35 anos:	60,5195 %
Mulheres com idade inferior a 35 anos:	33,5065 %
Moradores do Jardim Raposo com acesso a computador:	3,6364 %
Mulheres, vindas do Nordeste, e que estejam trabalhando:	13,5065 %
Mulheres, vindas do Nordeste, e que estejam trabalhando com carteira assinada:	30,7692
% das mulheres vindas do NE e que trabalham (4,15584 % do total)	