

1ª Parte: Analise cada variável para cada companhia (A e B).

1.i) Classifique as variáveis por tipo

- Companhia A

dados_grupo_3_comp_A.csv		dados_grupo_3_Incremento_salario_comp_A.csv	
Variável	Tipo	Variável	Tipo
Sexo	Qualitativa Nominal	Indicador incremento salário real	Qualitativa Nominal
Educ	Qualitativa Ordinal		
cargo	Qualitativa Nominal		
Sal	Quantitativa Contínua		

- Companhia B

dados_grupo_3_comp_B.csv		dados_grupo_3_Incremento_salario_comp_B.csv	
Variável	Tipo	Variável	Tipo
Sexo	Qualitativa Nominal	Indicador incremento salário real	Qualitativa Nominal
Educ	Qualitativa Ordinal		
cargo	Qualitativa Nominal		
Sal	Quantitativa Contínua		

1.ii) Apresente uma tabela de distribuição de frequências para cada variável.

- Companhia A

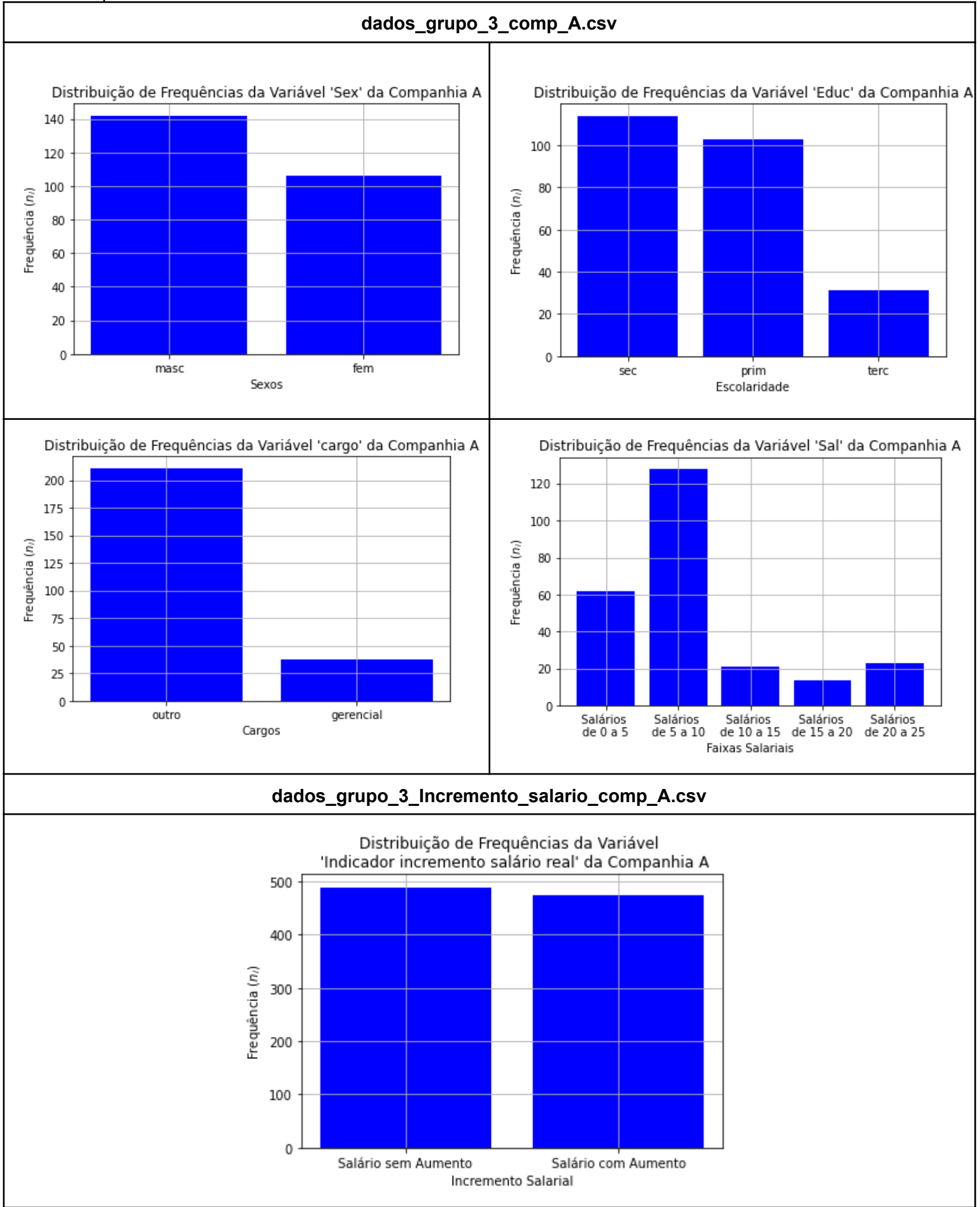
dados_grupo_3_comp_A.csv					dados_grupo_3_Incremento_salario_comp_A.csv				
Variável		Freq. (n _i)	Proporção (f _i)	%(100×f _i)	Variável		Freq. (n _i)	Proporção (f _i)	%(100×f _i)
Sexo	masc	142	0.572581	57.26	Indicador incremento salário real	0	489	0.507788	50.78
	fem	106	0.427419	42.74					
Educ	prim	103	0.415323	41.53					
	sec	114	0.459677	45.97					
	terc	31	0.125000	12.50					
cargo	outro	211	0.850806	85.08					
	gerencial	37	0.149194	14.92					
Sal	Salários de 0 a 5	62	0.250000	25.0000		1	474	0.492212	49.22
	Salários de 5 a 10	128	0.516129	51.6129					
	Salários de 10 a 15	21	0.084677	8.4677					
	Salários de 15 a 20	14	0.056452	5.6452					
	Salários de 20 a 25	23	0.092742	9.2742					

- Companhia B

dados_grupo_3_comp_B.csv					dados_grupo_3_Incremento_salario_comp_B.csv				
Variável		Freq. (n _i)	Proporção (f _i)	%(100×f _i)	Variável		Freq. (n _i)	Proporção (f _i)	%(100×f _i)
Sexo	masc	114	0.44358	44.36	Indicador incremento salário real	0	463	0.4739	47.39
	fem	143	0.55642	55.64					
Educ	prim	97	0.377432	37.74					
	sec	127	0.494163	49.42					
	terc	33	0.128405	12.84					
cargo	outro	220	0.856031	85.6					
	gerencial	37	0.143969	14.4					
Sal	Salários de 0 a 5	92	0.357977	35.7977		1	514	0.5261	52.61
	Salários de 5 a 10	106	0.412451	41.2451					
	Salários de 10 a 15	47	0.182879	18.2879					
	Salários de 15 a 20	8	0.031128	3.1128					
	Salários de 20 a 25	4	0.015564	1.5564					

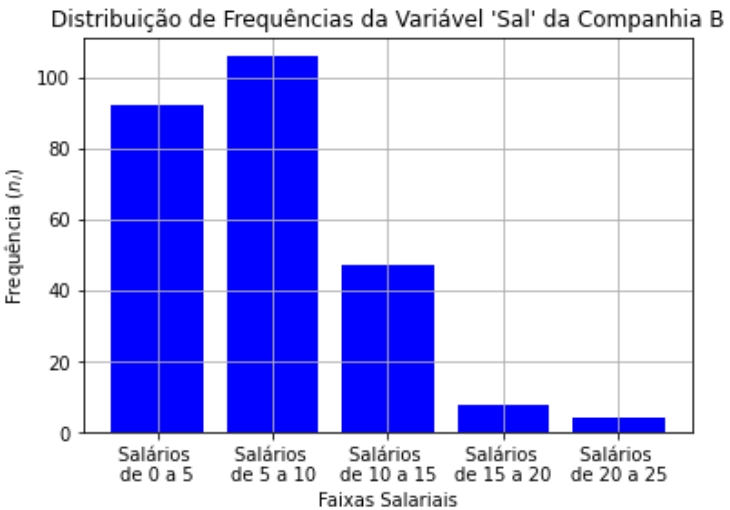
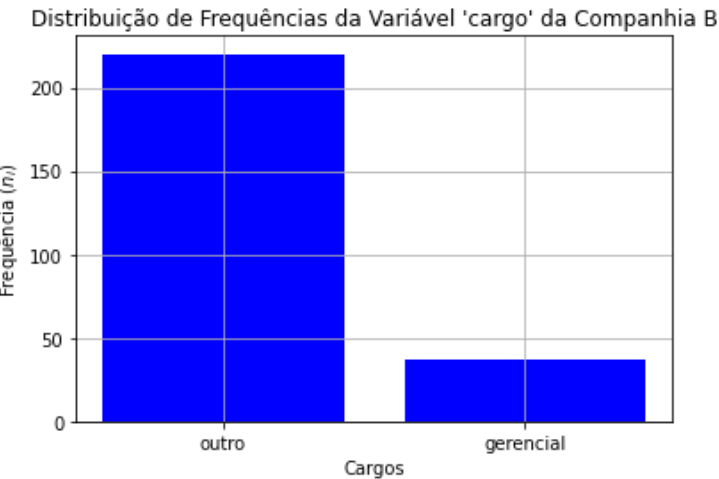
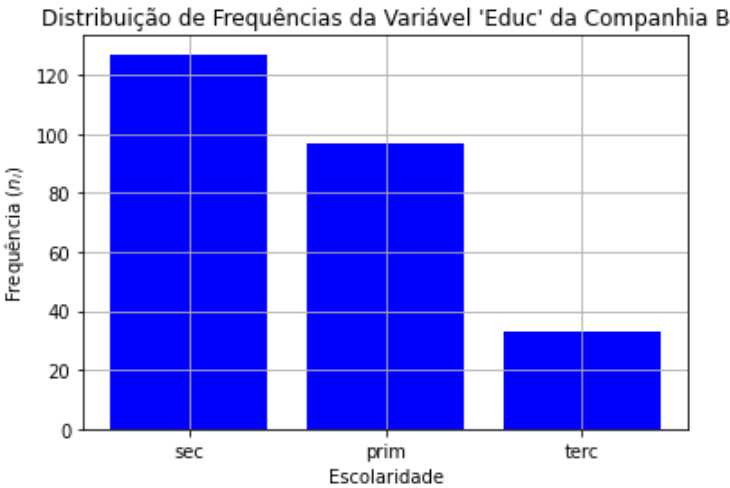
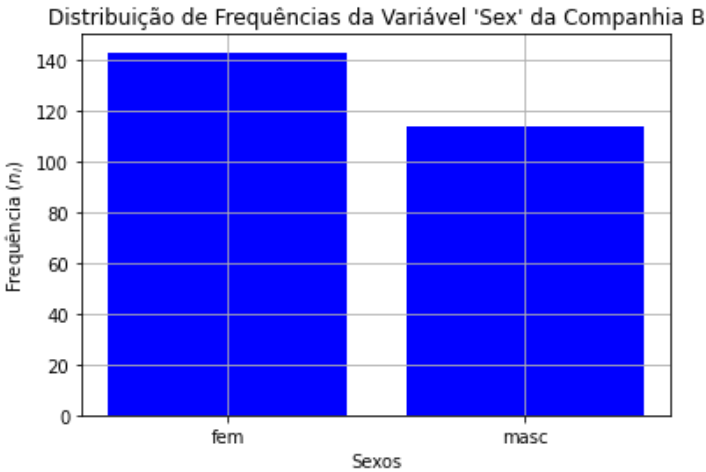
1.iii) Apresente um gráfico de distribuição de frequências para cada variável.

- Companhia A

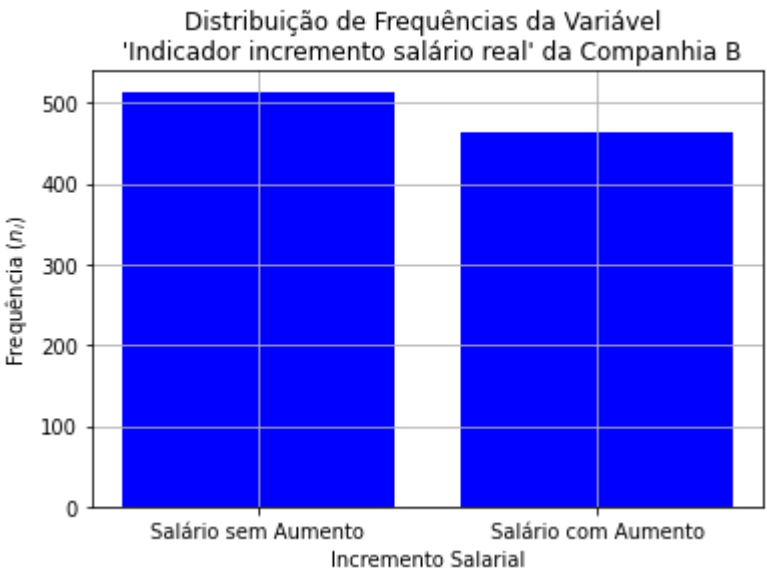


• Companhia B

dados_grupo_3_comp_B.csv

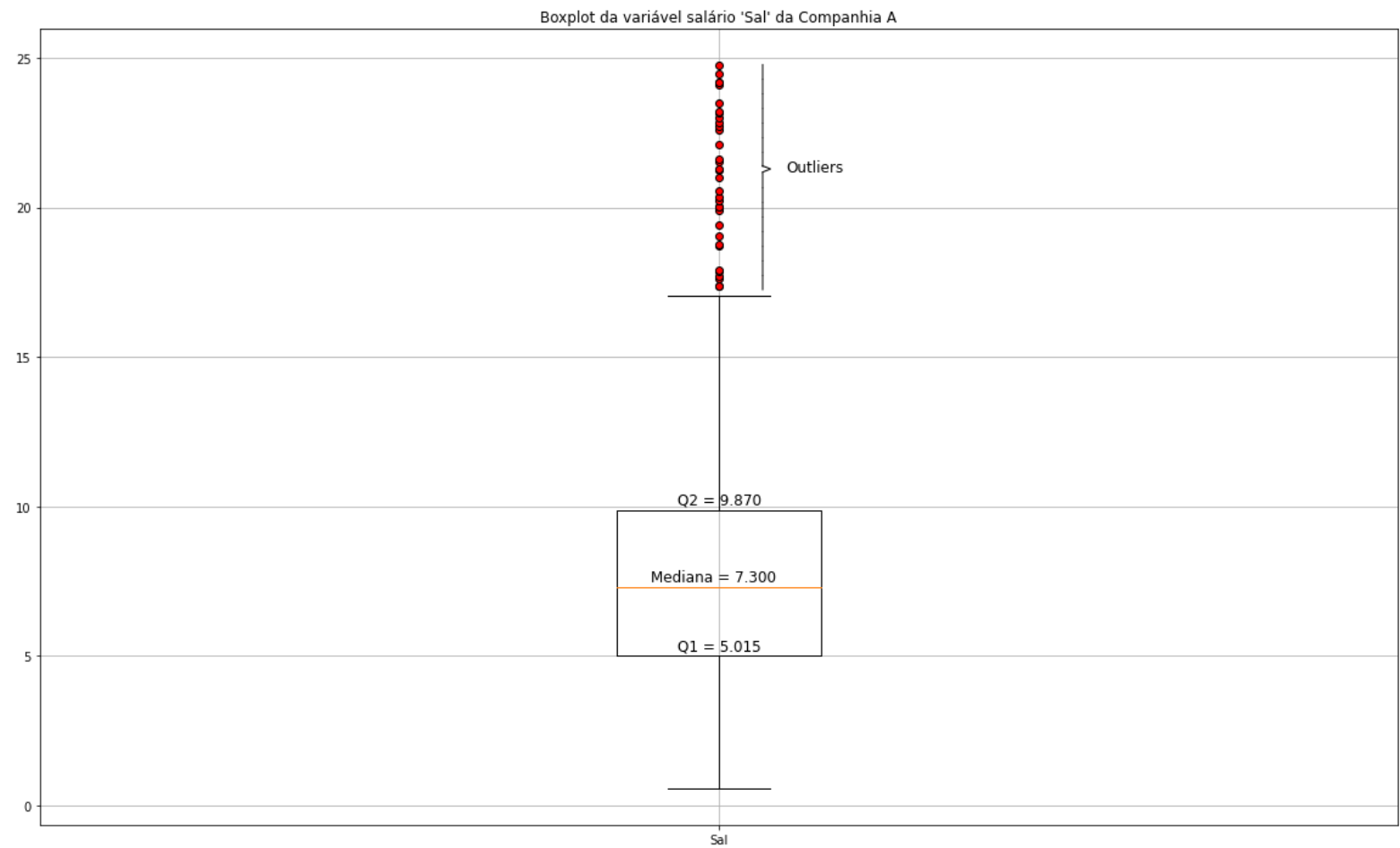


dados_grupo_3_Incremento_salario_comp_B.csv

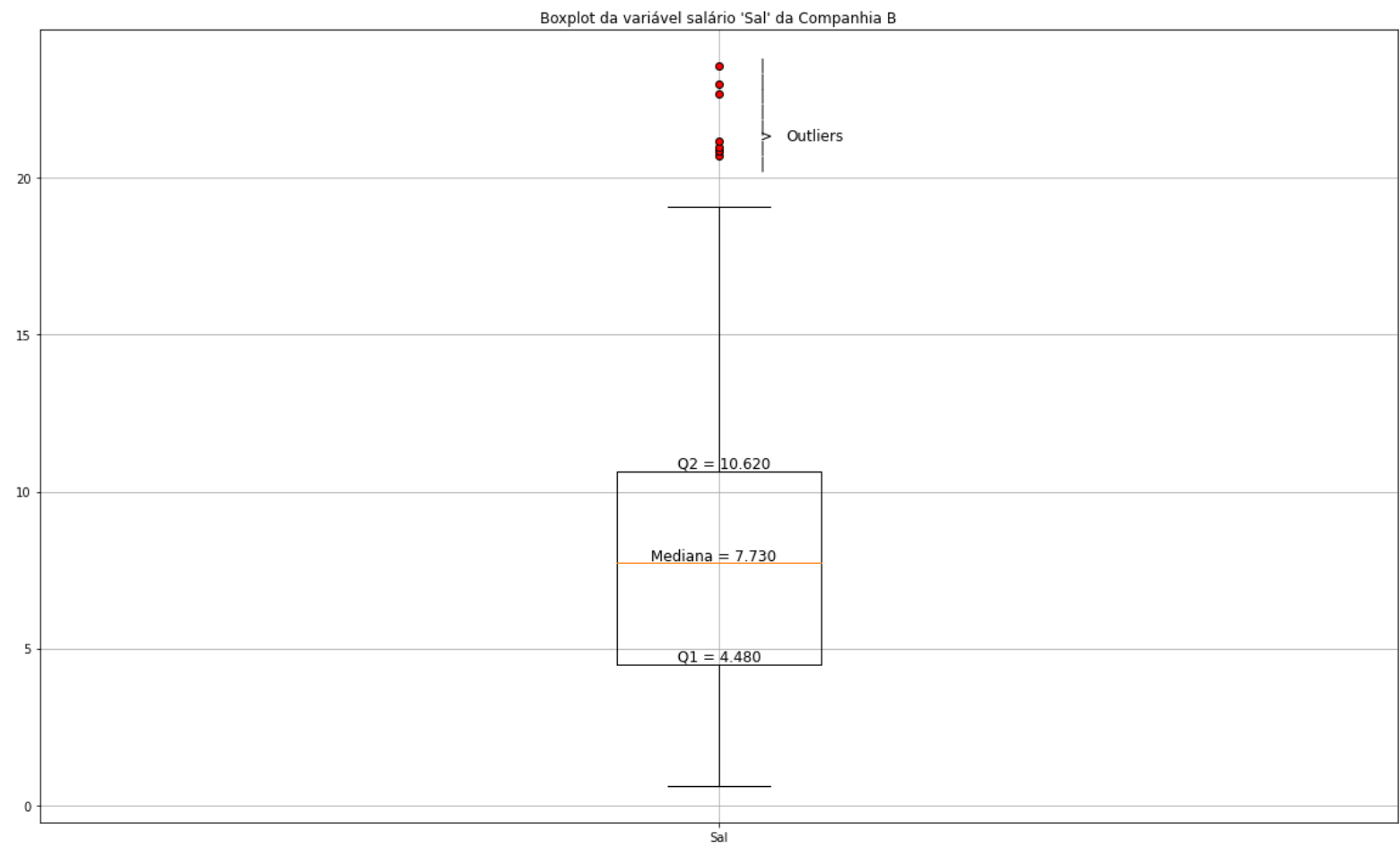


1.iv) Apresente um boxplot para a variável salário.

- Companhia A



- Companhia B



1.vi) Escreva as conclusões da análise de cada variável.

A variável qualitativa nominal “Sexo” possui, para o gênero masculino, maior frequência na Companhia A se comparado com a Companhia B; já para o gênero feminino, ocorre o oposto. Para a variável qualitativa ordinal “Educ”, temos para a categoria ‘prim’ maior quantidade na Companhia A do que na Companhia B; para ambas as categorias ‘sec’ e ‘terc’, maior quantidade na Companhia B do que na Companhia A. A variável qualitativa nominal “cargo” possui maior quantidade da categoria ‘outro’ em ambas as Companhias A e B, sendo que a categoria ‘gerencial’ possui uma frequência muito menor, com menos de 15% de presença. Para a variável “Sal”, foram criados cinco intervalos com base no menor salário (Companhia A: 0.56, Companhia B: 0.61) e no maior salário (Companhia A: 24.76, Companhia B: 23.56): de 0 até 5, estritamente maiores que 5 até 10, estritamente maiores que 10 até 15, estritamente maiores que 15 até 20, e estritamente maiores que 20 até 25; observamos que as frequências dessas divisões em grupos de salários são muito semelhantes em ambas as Companhias, então foi uma boa decisão do grupo fazer esta divisão em cinco intervalos. No que diz respeito à variável “Indicador incremento salário real”, podemos dizer que a Companhia A possui uma divisão mais equilibrada, isto é, mais próxima das duas categorias ‘0’ e ‘1’ possuírem 50% de frequência que a Companhia B.

Em termos gerais do salário na Companhia A e na Companhia B, observamos através dos boxplots que há uma quantidade significativamente grande de outliers na Companhia A se comparada à B; notamos também que média e a mediana das duas companhias são relativamente próximas, assim como os quartis Q1 e Q2.

2ª Parte: Para cada companhia, analise a dependência entre as variáveis.

Para cada item e para cada companhia (A e B), foram realizados os seguintes passos: primeiramente foi feita uma tabela de Distribuição Conjunta. Para ter uma melhor ideia da distribuição dos dados fizemos a Distribuição Relativa às Colunas e também a Distribuição Relativa às Linhas. Com essas porcentagens já poderíamos ter uma ideia de como os dados se comportam em relação a dependência de variáveis.

Por fim, fizemos o cálculo do Qui-quadrado de Pearson que nos diz que se a hipótese de não-associação for verdadeira, o valor calculado deve estar próximo de zero, mas se as variáveis forem associadas, o valor deve ser grande. Para tanto, fizemos uma tabela com os valores esperados para cada posição de acordo com a porcentagem da distribuição das frequências relativas ao total por linhas e, a partir disso, fizemos outra tabela com os desvios relativos entre observados e esperados $((oi - ei)^2/ei)$. A soma desta última tabela nos dá o Qui-quadrado de Pearson.

Para o item 2.i), cada tabela usada está indicada neste relatório. Para os demais itens, mostramos apenas a tabela de Distribuição Conjunta, Distribuição das frequências relativas ao total por linhas e o valor do Qui-quadrado de Pearson. Os códigos e tabelas completas se encontram no arquivo jupyter.

2.i) Analise a dependência entre as variáveis sexo e educação.

- Companhia A

Distribuição Conjunta			
Educação	Masculino	Feminino	Total
prim	50	53	103
sec	71	43	114
terc	21	10	31
Total	142	106	248

Distribuição das frequências relativas ao total por coluna				Valores esperados para cada posição na tabela de acordo com a porcentagem do total por coluna			Distribuição das frequências relativas ao total por linhas			
Educ	Masc (%)	Fem (%)	Total (%)	Educ	Masc	Fem	Educ	Masc (%)	Fem (%)	Total (%)
prim	35.2	50.0	41.5	prim	59.0	44.0	prim	48.5	51.5	100.0
sec	50.0	40.6	46.0	sec	65.3	48.7	sec	62.3	37.7	100.0
terc	14.8	9.4	12.5	terc	17.8	13.2	terc	67.7	32.3	100.0
Total	100.0	100.0	100.0				Total	57.3	42.7	100.0

Qui-quadrado de Pearson (X_quadrado): 5.763436970424404		
Educação	Masculino	Feminino
prim	1.366070	1.830019
sec	0.502264	0.672844
terc	0.595070	0.797170

- Companhia B

Distribuição Conjunta					
Educação	Masculino	%(100×freq_masc)	Feminino	%(100×freq_fem)	Total
prim	42	43.298969	55	56.701031	97
sec	58	45.669291	69	54.330709	127
terc	14	42.424242	19	57.575758	33
Total	143	55.642023	114	44.357977	257

Distribuição das frequências relativas ao total por coluna				Valores esperados para cada posição na tabela de acordo com a porcentagem do total por coluna			Distribuição das frequências relativas ao total por linhas			
Educ	Masc (%)	Fem (%)	Total (%)	Educ	Masc	Fem	Educ	Masc (%)	Fem (%)	Total (%)
prim	35.2	50.0	41.5	prim	59.0	44.0	prim	48.5	51.5	100.0
sec	50.0	40.6	46.0	sec	65.3	48.7	sec	62.3	37.7	100.0
terc	14.8	9.4	12.5	terc	17.8	13.2	terc	67.7	32.3	100.0
Total	100.0	100.0	100.0				Total	57.3	42.7	100.0

Qui-quadrado de Pearson (X_quadrado): 5.763436970424404		
Educação	Masculino	Feminino
prim	1.366070	1.830019
sec	0.502264	0.672844
terc	0.595070	0.797170

2.ii) Analise a dependência entre as variáveis sexo e salário.

Companhia A: Distribuição Conjunta

Faixas Salariais	Masculino	Feminino	Total
Salários de 0 a 5	45	17	62
Salários de 5 a 10	66	62	128
Salários de 10 a 15	6	15	21
Salários de 15 a 20	6	8	14
Salários de 20 a 25	19	4	23
Total	142	106	248

Companhia A: Distribuição das frequências relativas ao total por linhas

Faixas Salariais	Masculino (%)	Feminino (%)	Total (%)
Salários de 0 a 5	72.6	27.4	100.0
Salários de 5 a 10	51.6	48.4	100.0
Salários de 10 a 15	28.6	71.4	100.0
Salários de 15 a 20	42.9	57.1	100.0
Salários de 20 a 25	82.6	17.4	100.0
Total	57.3	42.7	100.0

Companhia A - Qui-quadrado de Pearson: **21.931966697642785**

Companhia B: Distribuição Conjunta

Faixas Salariais	Masculino	Feminino	Total
Salários de 0 a 5	12	65	77
Salários de 5 a 10	45	53	98
Salários de 10 a 15	43	23	66
Salários de 15 a 20	6	2	8
Salários de 20 a 25	8	0	8
Total	114	143	257

Companhia B: Distribuição das frequências relativas ao total por linhas

Faixas Salariais	Masculino (%)	Feminino (%)	Total (%)
Salários de 0 a 5	15.6	84.4	100.0
Salários de 5 a 10	45.9	54.1	100.0
Salários de 10 a 15	65.2	34.8	100.0
Salários de 15 a 20	75.0	25.0	100.0
Salários de 20 a 25	100.0	0.0	100.0
Total	44.4	55.6	100.0

Companhia B: Qui-quadrado de Pearson: **50.56566436146067**

2.iii) Analise a dependência entre as variáveis educação e salário.

Companhia A: Distribuição Conjunta

Faixas Salariais	prim	sec	terc	Total
Salários de 0 a 5	60	2	0	62
Salários de 5 a 10	40	87	1	128
Salários de 10 a 15	2	15	4	21
Salários de 15 a 20	1	8	5	14
Salários de 20 a 25	0	2	21	23
Total	103	114	31	248

Companhia A: Distribuição das frequências relativas ao total por linhas

Faixas Salariais	prim (%)	sec (%)	terc (%)	Total (%)
Salários de 0 a 5	96.8	3.2	0.0	100.0
Salários de 5 a 10	31.2	68.0	0.8	100.0
Salários de 10 a 15	9.5	71.4	19.0	100.0
Salários de 15 a 20	7.1	57.1	35.7	100.0
Salários de 20 a 25	0.0	8.7	91.3	100.0
Total	41.5	46.0	12.5	100.0

Companhia A: Qui-quadrado de Pearson: **258.77989847852007**

Companhia B: Distribuição Conjunta

Faixas Salariais	prim	sec	terc	Total
Salários de 0 a 5	61	16	0	77
Salários de 5 a 10	31	63	4	98
Salários de 10 a 15	5	47	14	66
Salários de 15 a 20	0	1	7	8
Salários de 20 a 25	0	0	8	8
Total	97	127	33	257

Companhia B: Distribuição das frequências relativas ao total por linhas

Faixas Salariais	prim (%)	sec (%)	terc (%)	Total (%)
Salários de 0 a 5	79.2	20.8	0.0	100.0
Salários de 5 a 10	31.6	64.3	4.1	100.0
Salários de 10 a 15	7.6	71.2	21.2	100.0
Salários de 15 a 20	0.0	12.5	87.5	100.0
Salários de 20 a 25	0.0	0.0	100.0	100.0
Total	37.7	49.4	12.8	100.0

Companhia B: Qui-quadrado de Pearson: **189.09070938729698**

2.iv) Analise a dependência entre as variáveis sexo e cargo.

Companhia A: Distribuição Conjunta

Cargo	Masculino	Feminino	Total
Outro	121	90	211
Gerencial	21	16	37
Total	142	106	248

Companhia A: Distribuição das frequências relativas ao total por linhas

Cargo	Masculino (%)	Feminino (%)	Total (%)
Outro	57.3	42.7	100.0
Gerencial	56.8	43.2	100.0
Total	57.3	42.7	100.0

Companhia A: Qui-quadrado de Pearson: **0.004465693941545747**

Companhia B: Distribuição Conjunta

Cargo	Masculino	Feminino	Total
Outro	100	120	220
Gerencial	14	23	37
Total	143	114	257

Companhia B: Distribuição das frequências relativas ao total por linhas

Cargo	Masculino (%)	Feminino (%)	Total (%)
Outro	45.5	54.5	100.0
Gerencial	37.8	62.2	100.0
Total	55.6	44.4	100.0

Companhia B: Qui-quadrado de Pearson: **14.002788732453812**

2.v) Analise a dependência entre as variáveis cargo e salário.

Companhia A: Distribuição Conjunta

Faixas Salariais	Outro	Gerencial	Total
Salários de 0 a 5	62	0	62
Salários de 5 a 10	128	0	128
Salários de 10 a 15	16	5	21
Salários de 15 a 20	4	10	14
Salários de 20 a 25	1	22	23
Total	211	37	248

Companhia A: Distribuição das frequências relativas ao total por linhas

Faixas Salariais	Outro (%)	Gerencial (%)	Total (%)
Salários de 0 a 5	100.0	0.0	100.0
Salários de 5 a 10	100.0	0.0	100.0
Salários de 10 a 15	76.2	23.8	100.0
Salários de 15 a 20	28.6	71.4	100.0
Salários de 20 a 25	4.3	95.7	100.0
Total	85.1	14.9	100.0

Companhia A: Qui-quadrado de Pearson: **5213.73120689614**

Companhia B: Distribuição Conjunta

Faixas Salariais	Outro	Gerencial	Total
Salários de 0 a 5	77	0	77
Salários de 5 a 10	92	6	98
Salários de 10 a 15	49	17	66
Salários de 15 a 20	2	6	8
Salários de 20 a 25	0	8	8
Total	220	37	257

Companhia B: Distribuição das frequências relativas ao total por linhas

Faixas Salariais	Outro (%)	Gerencial (%)	Total (%)
Salários de 0 a 5	100.0	0.0	100.0
Salários de 5 a 10	93.9	6.1	100.0
Salários de 10 a 15	74.2	25.8	100.0
Salários de 15 a 20	25.0	75.0	100.0
Salários de 20 a 25	0.0	100.0	100.0
Total	85.6	14.4	100.0

Companhia B: Qui-quadrado de Pearson: **96.71461407435436**

2.vi) Escreva as conclusões da análise de dependência.

- **i) Sexo e Educação:**
Para a companhia A: a partir da tabela de distribuição conjunta relativa ao total por colunas podemos perceber que 41,5% do total das pessoas completou o ensino primário, 46% o secundário e 12,5% o ensino superior ('terc'). Dado isso, para que não haja dependência esta proporção deveria ser seguida para os dois sexos, porém percebemos que há uma pequena divergência principalmente na porcentagem referente ao ensino primário que 35,2% dos homens fizeram e 41,5% das mulheres fizeram. Observando também o Qui-quadrado percebemos que há uma pequena dependência entre as variáveis.
Para a companhia B: Da mesma forma, observando a tabela de tabela de distribuição conjunta relativa ao total por colunas percebemos uma pequena discrepância na porcentagem. Como o Qui-quadrado é um pouco maior, podemos dizer que as variáveis são mais dependentes do que para a companhia A.
- **ii) Sexo e Salário:** Aqui percebemos que conforme o salário aumenta a porcentagem de mulheres com esses salários diminui, tanto para companhia A quanto para B, principalmente para a faixa "Salários de 20 a 25". Logo, há uma maior dependência entre as variáveis.
- **iii) Educação e Salário:** Aqui podemos observar a maior dependência entre as variáveis, pois claramente as faixas de salários maiores são dominadas por quem possui ensino superior ('terc') tanto para a companhia A quanto para a B. Isso é visto também no alto valor de Qui-quadrado.
- **iv) Sexo e Cargo:** Com o cruzamento dos dados apresentado neste item, em relação à companhia A, podemos observar que, de acordo com a distribuição das frequências relativas ao total por linhas, a relação entre sexo e cargo é praticamente irrelevante, inexistente. Isso é indicado pela similaridade de cada frequência em ambas as colunas "Masculino" e "Feminino" com o seu respectivo total. Sintetizando este fato, observa-se o valor de Qui-quadrado de Pearson, que para este caso é bem próximo de 0, indicando a baixíssima relação entre as variáveis em questão: sexo e cargo. Para a companhia B, por uma lógica similar, observa-se uma maior

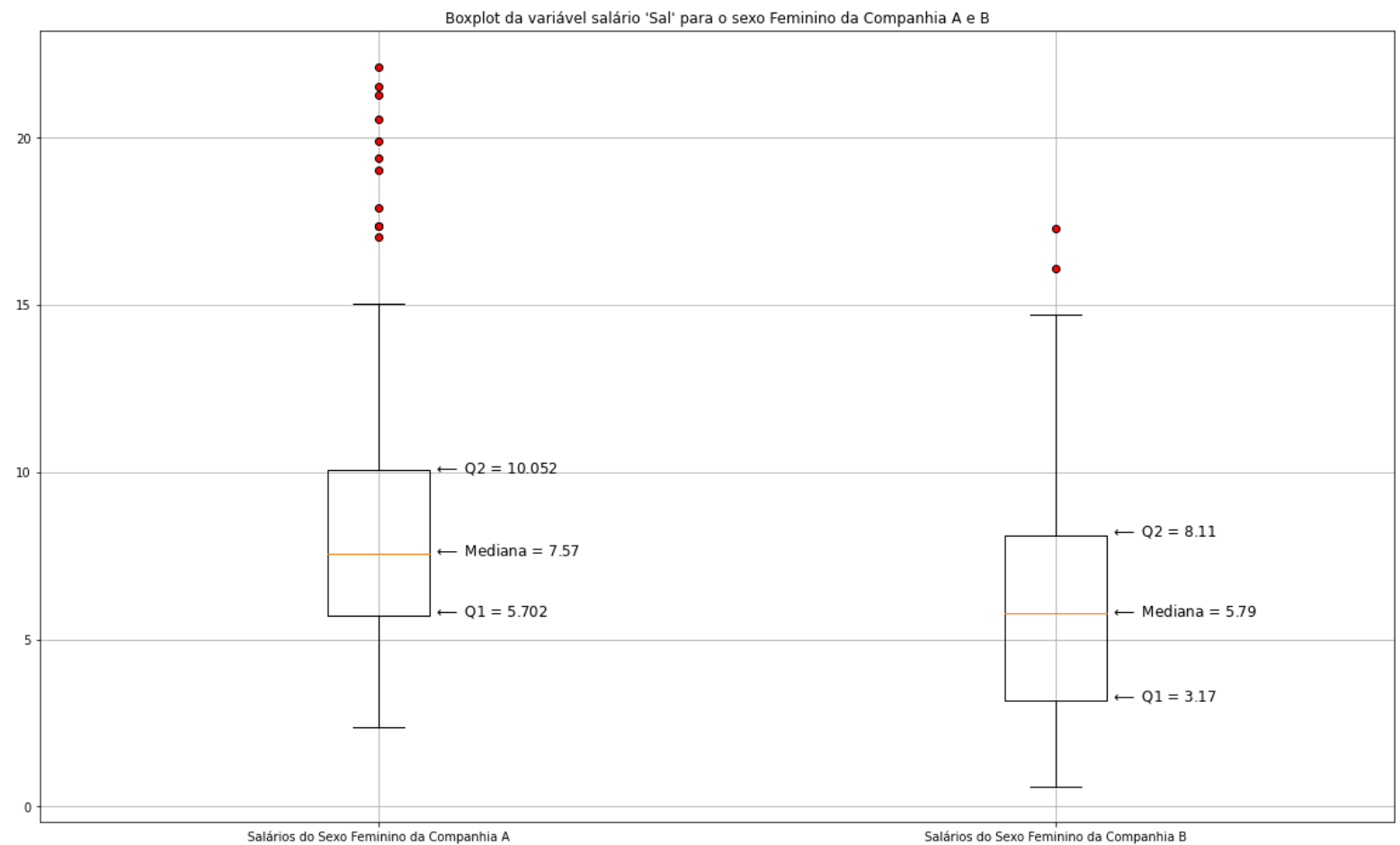
discrepância de cada frequência em ambas as colunas, o que revela uma maior dependência entre as variáveis sexo e cargo em relação aos dados da companhia A, com um Qui-quadrado de aproximadamente 14.00.

v) **Cargo e Salário:** Em relação à companhia A, observemos que a distribuição das frequências relativas ao total por linhas é consideravelmente discrepante para cada faixa de salário, de modo que a maioria dos funcionários com os maiores salários são de cargos gerenciais, enquanto todos os funcionários com salários de 0 a 10 são de outros cargos. A altíssima dependência entre as variáveis, para o caso desta companhia, é evidenciada pelo valor de Qui-quadrado de Pearson de 5213.73. Já em relação à companhia B, podemos observar uma menor discrepância em relação ao total de cada coluna ao compararmos tais dados com os da companhia A. Na companhia B, a menor faixa de salários é composta em sua totalidade por funcionários de cargos não-gerenciais, enquanto a maior faixa é totalmente composta de funcionários de cargos gerenciais. O valor de Pearson valendo 96.71 nos ajuda a entender ainda mais a magnitude da altíssima dependência entre as variáveis salário e cargo observada no caso da companhia A em relação à B.

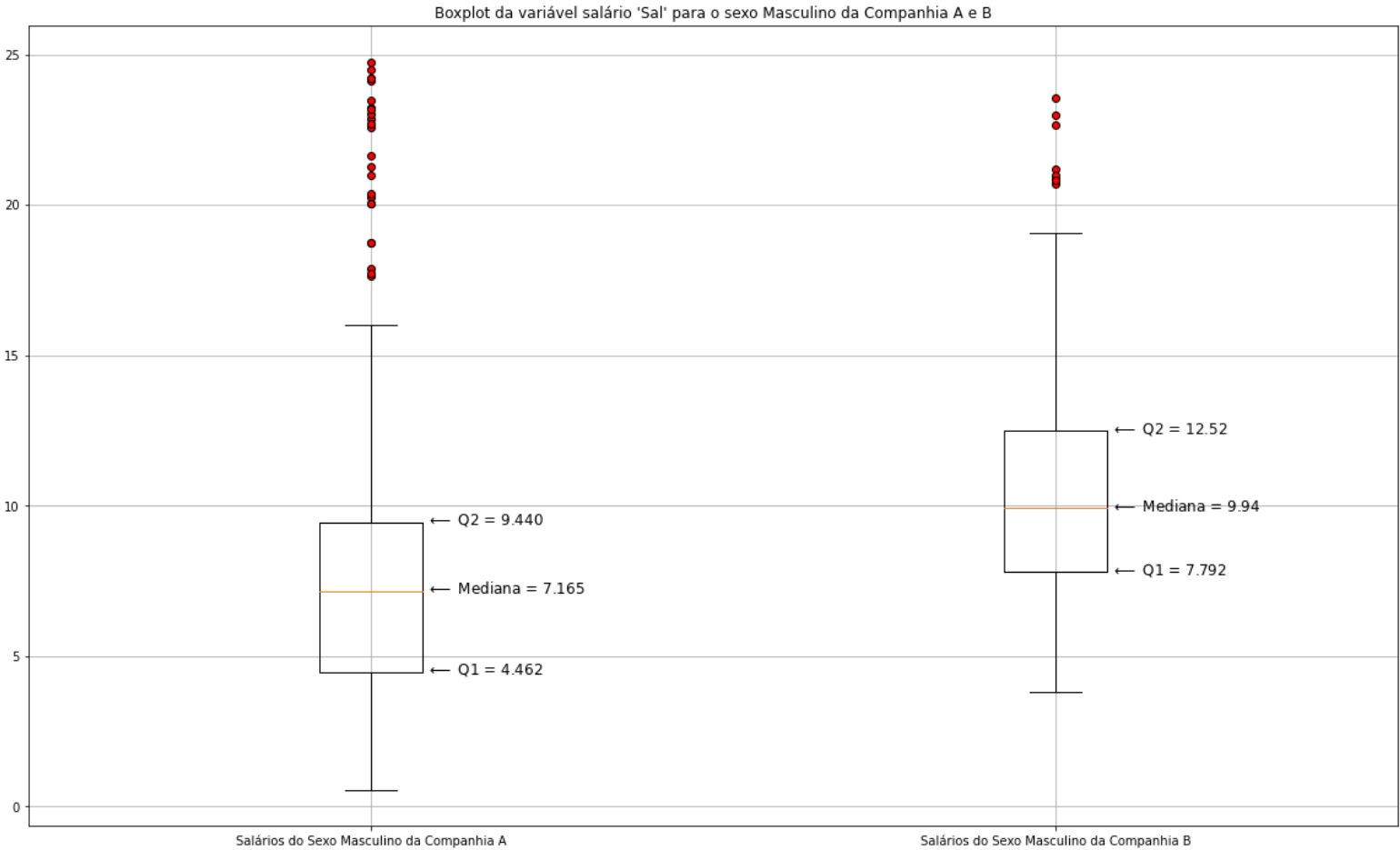
3ª Parte: Comparações entre as companhias.

3.i) Compare os salários das duas companhias para as diferentes categorias de sexo e educação (use boxplots).

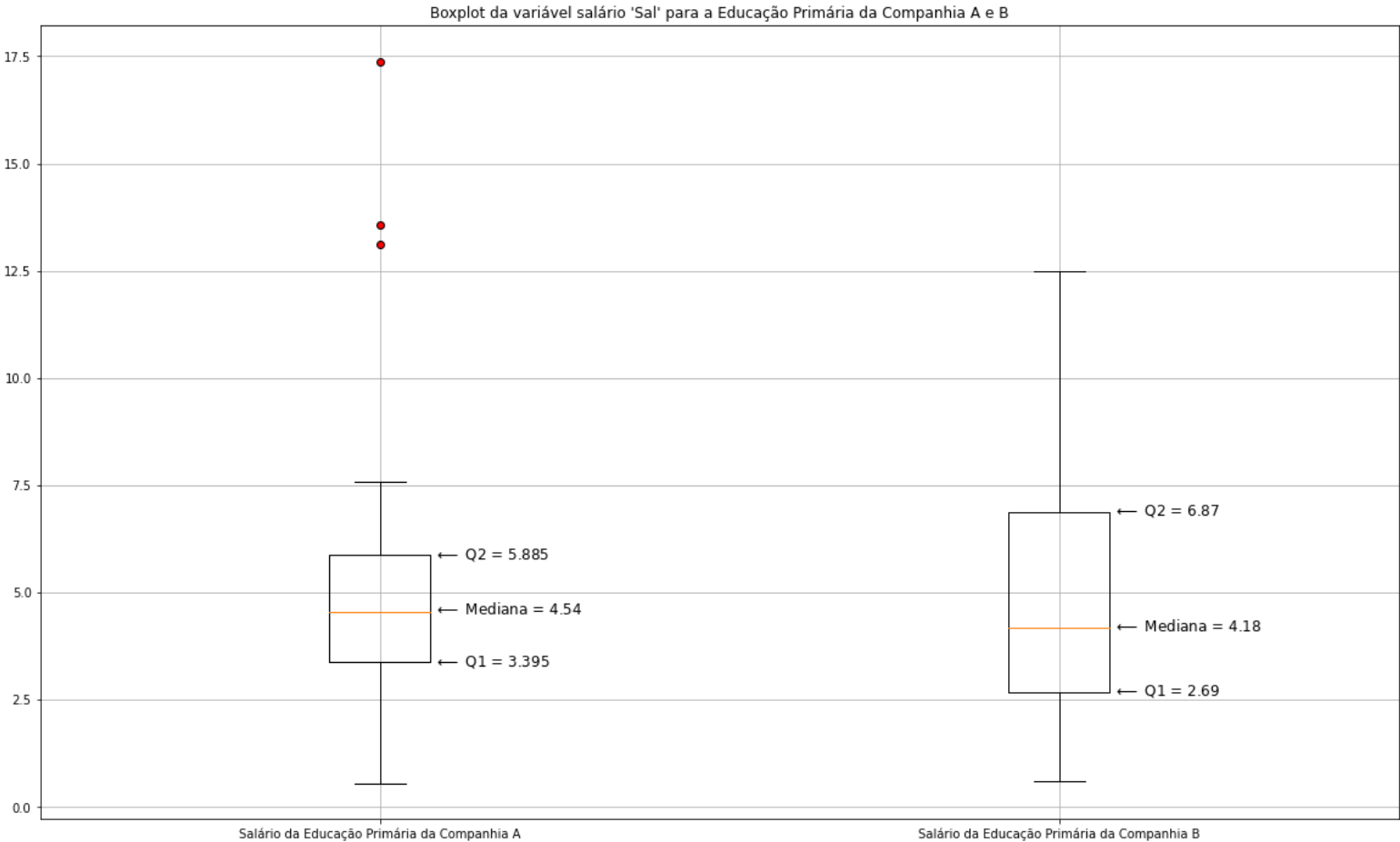
- Salário das Mulheres da Companhia A e das Mulheres da Companhia B



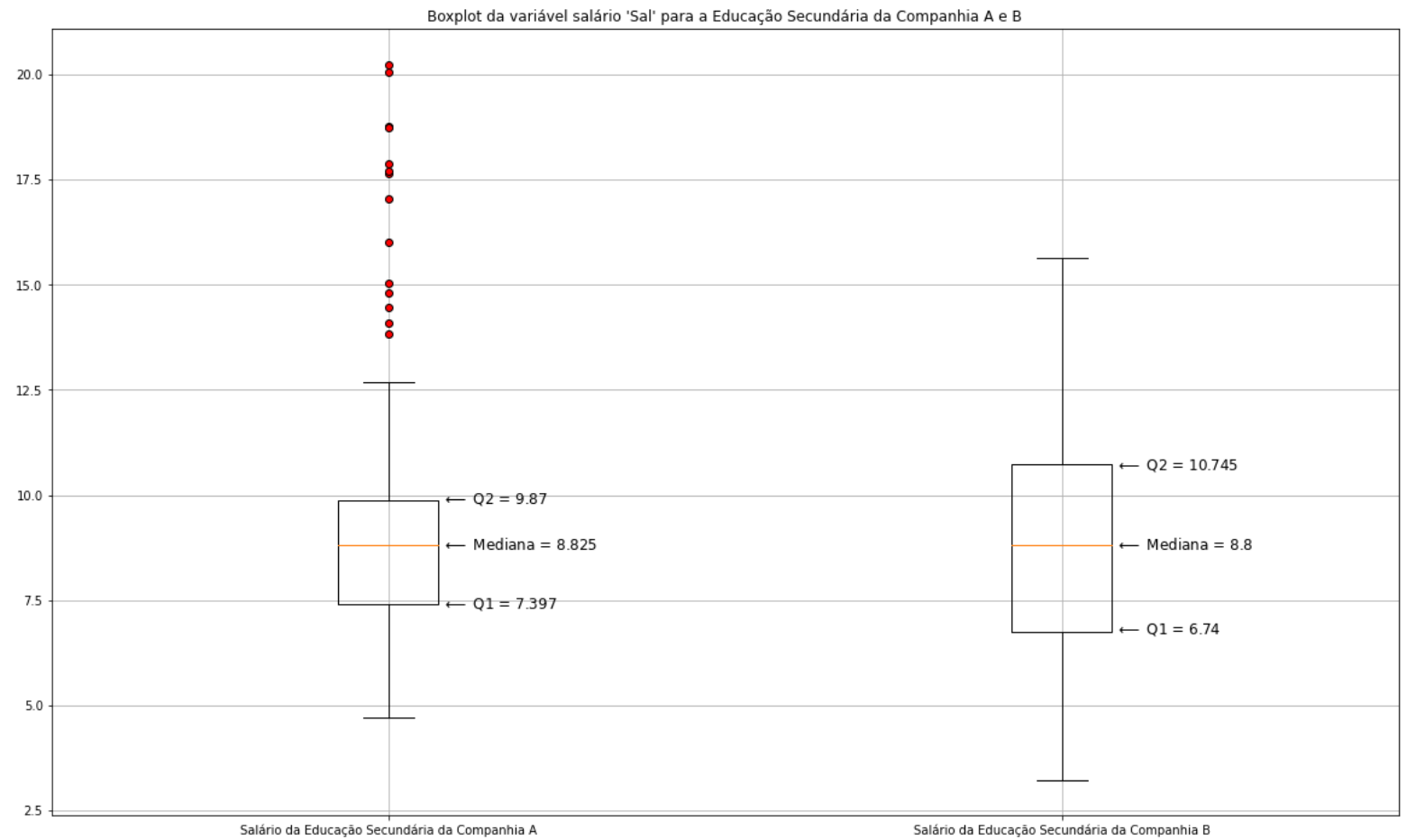
● Salário dos Homens da Companhia A e dos Homens da Companhia B



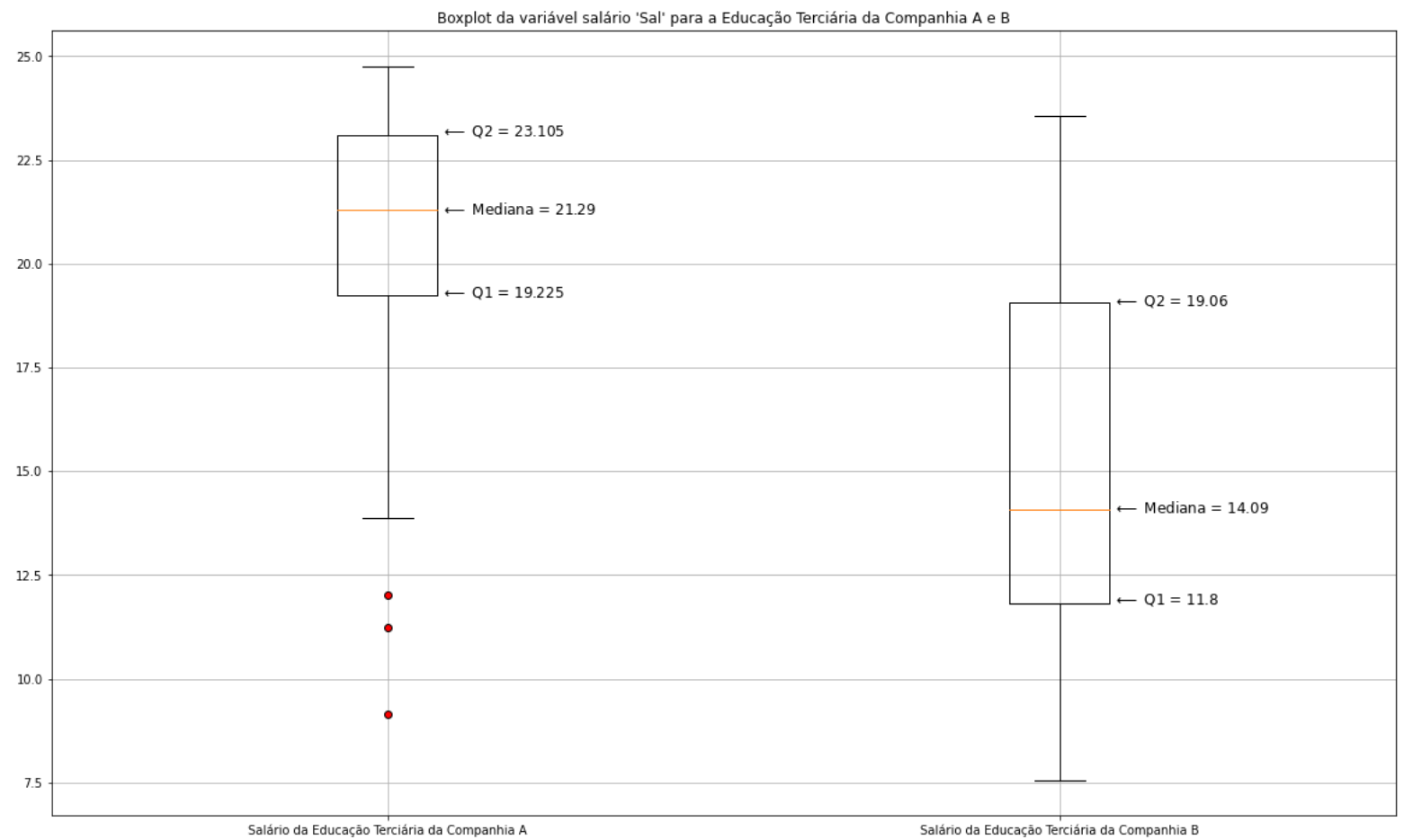
● Salário da Educação Primária da Companhia A e Educação Primária Companhia B



● Salário e Educação Secundária Companhia A e Educação Secundária Companhia B



● Salário e Educação Terciária Companhia A e Educação Terciária Companhia B

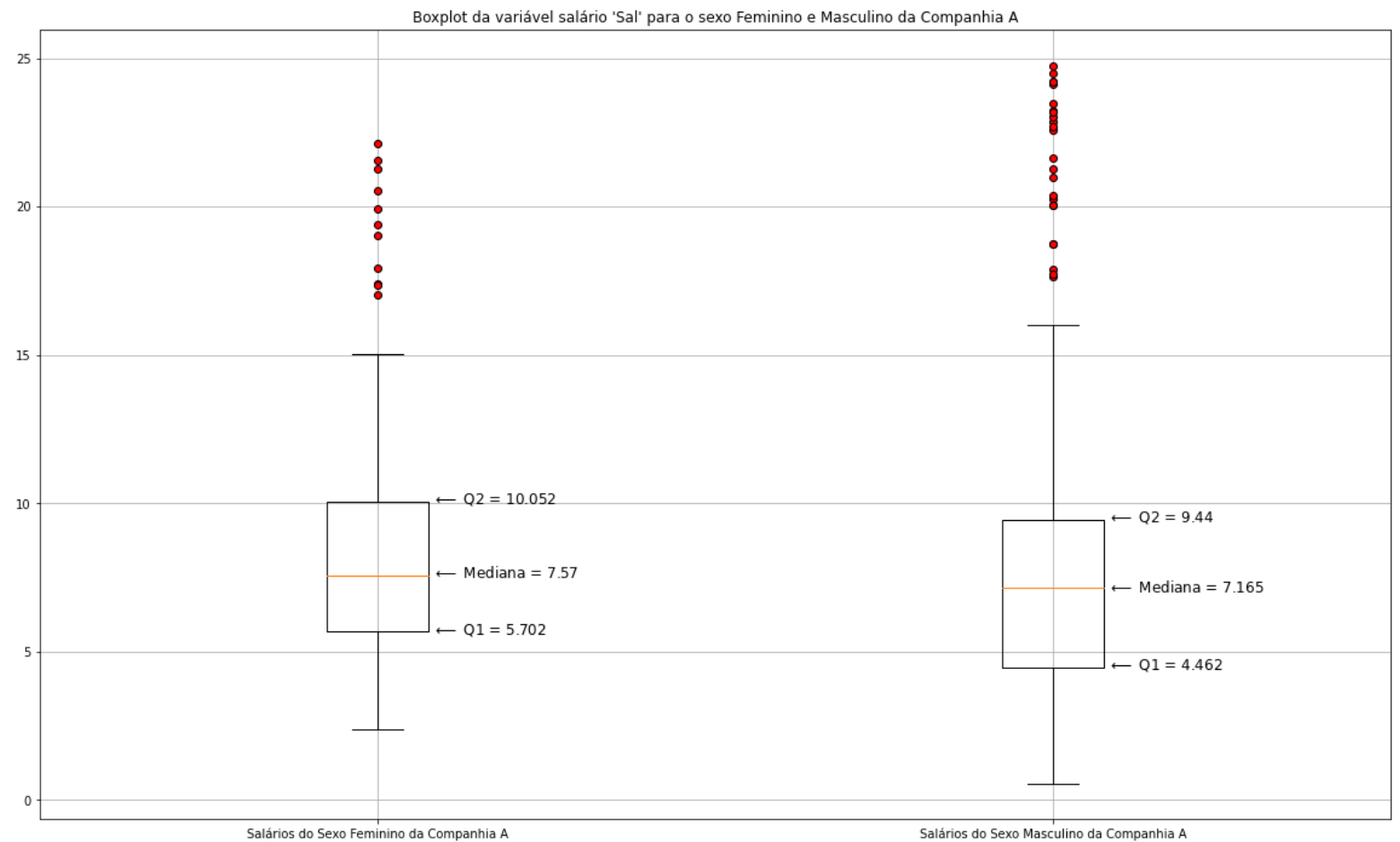


3.ii) Utilize intervalos de confiança para comparar os salários das duas companhias para as diferentes categorias de sexo e educação.

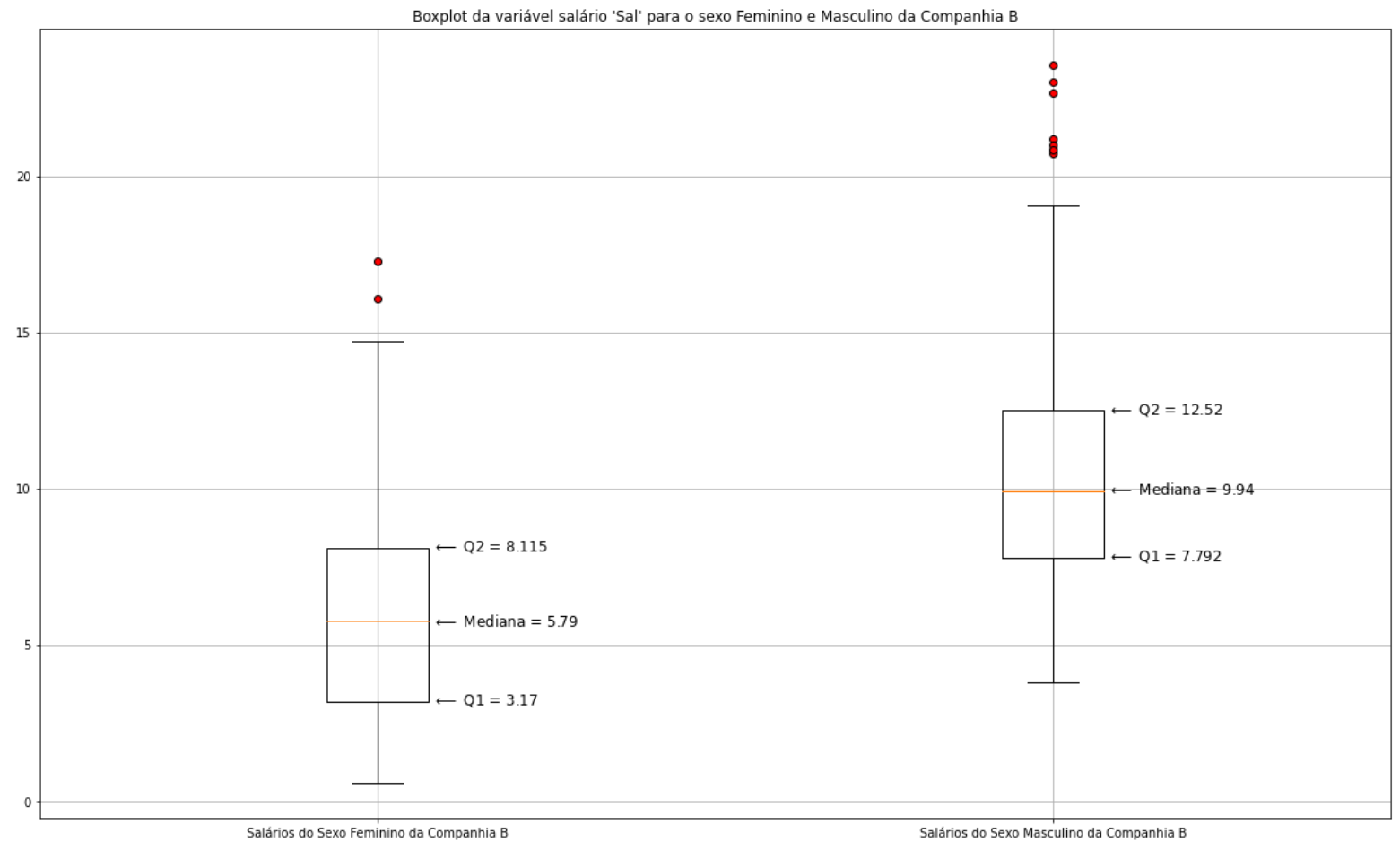
Intervalos de Confiança para os Salários de Funcionários do Sexo Masculino:	
Companhia A	[7.779675478273895, 9.921451282289478]
Companhia B	[9.730159129244326, 11.367560169001289]
Intervalos de Confiança para os Salários de Funcionários do Sexo Feminino:	
Companhia A	[7.909019158286598, 9.679094049260572]
Companhia B	[5.514971220859486, 6.700133674245416]
Intervalos de Confiança para os Salários de Funcionários com Escolaridade Primária:	
Companhia A	[4.291024795084898, 5.248004331128694]
Companhia B	[4.191560484595262, 5.345965288600614]
Intervalos de Confiança para os Salários de Funcionários com Escolaridade Secundária:	
Companhia A	[8.78336172139881, 9.99558564702224]
Companhia B	[8.279874559345817, 9.293038826480949]
Intervalos de Confiança para os Salários de Funcionários com Escolaridade Terciária:	
Companhia A	[18.72383193404335, 21.746490646601817]
Companhia B	[13.419939637185301, 16.731575514329844]

3.iv) Compare a diferença salarial entre homens e mulheres em cada companhia (use boxplots).

- Salário das Mulheres da Companhia A e dos Homens da Companhia A



- Salário das Mulheres da Companhia B e dos Homens da Companhia B



3.v) Compare a diferença salarial entre homens e mulheres em cada companhia com teste de hipóteses.

- Companhia A
 - H0: todos os salários, independente do sexo
 - H1: salário de homens

Teste de Hipótese	U - estatístico	p-valor
U de Mann-Whitney	18329.0	0.501187 > 0.05
Ranksum	0.673081	0.500896 > 0.05
Kolmogorov-Smirnov	-	0.7661683804306618
Conclusão:	Não há significância estatística para rejeitar a hipótese nula	

- H0: todos os salários, independente do sexo
- H1: salário de mulheres

Teste de Hipótese	U - estatístico	p-valor
U de Mann-Whitney	12423.0	0.413913 > 0.05
Ranksum	-0.817585	0.413594 > 0.05
Kolmogorov-Smirnov	-	0.5326906011770652
Conclusão:	Não há significância estatística para rejeitar a hipótese nula	

- Companhia B
 - H0: todos os salários, independente do sexo
 - H1: salário de homens

Teste de Hipótese	U - estatístico	p-valor
U de Mann-Whitney	9890.0	0.000001 < 0.05
Ranksum	-4.993624	0.000001 < 0.05
Kolmogorov-Smirnov	-	1.2974658015973972e-05 < 0.05
Conclusão:	Possível rejeitar a hipótese nula de acordo com: U de Mann- Whitney, Ranksum e Kolmogorov-Smirnov	

- H0: todos os salários, independente do sexo
- H1: salário de mulheres

Teste de Hipótese	U - estatístico	p-valor
U de Mann-Whitney	23134.5	0.000018 < 0.05
Ranksum	4.294371	0.000018 < 0.05
Kolmogorov-Smirnov	-	0.00030348525039980423 < 0.05
Conclusão:	Possível rejeitar a hipótese nula de acordo com: U de Mann- Whitney, Ranksum e Kolmogorov-Smirnov	

3.vi) Compare a participação de mulheres em cargos gerenciais nas duas companhias usando intervalos de confiança.

Intervalos de Confiança para a Participação de Funcionários do Sexo Feminino em Cargos Gerenciais:	
Companhia A	[0.033727871161186244, 0.09530438690332987]
Companhia B	[0.05436038456476303, 0.124627942283486]

3.vii) Compare a participação de mulheres em cargos gerenciais nas duas companhias usando teste de hipótese.

- Companhia A
 - H0: todos os cargos gerenciais, independente do sexo
 - H1: cargos gerenciais de mulheres

Teste de Hipótese	U - estatístico	p-valor
U de Mann-Whitney	Todos os números são idênticos no teste U de Mann-Whitney, então o desconsideramos	
Ranksum	0.00	1.0 > 0.05
Kolmogorov-Smirnov	-	1.0 > 0.05
Conclusão:	Não há significância estatística para rejeitar a hipótese nula	

- Companhia B
 - H0: todos os cargos gerenciais, independente do sexo
 - H1: cargos gerenciais de mulheres

Teste de Hipótese	U - estatístico	p-valor
U de Mann-Whitney	Todos os números são idênticos no teste U de Mann-Whitney, então o desconsideramos	
Ranksum	0.00	1.0 > 0.05
Kolmogorov-Smirnov	-	1.0 > 0.05
Conclusão:	Não há significância estatística para rejeitar a hipótese nula	

3.viii) Escreva as conclusões da análise da possível diferença salarial entre homens e mulheres em cada companhia.

A partir dos boxplots de comparação salarial entre as mulheres da Companhia A e mulheres da Companhia B, pode-se afirmar que, apesar da Companhia A apresentar mais outliers que a B, as mulheres da Companhia A ganham mais que a da B; tal fato pode ser observado também pelos intervalos de confiança calculados: a Companhia A apresenta um intervalo de confiança de 95% para o salário médio das mulheres de [7.9, 9.7] enquanto que na Companhia B temos [5.5, 6.7]. Já, ao analisar os boxplots de comparação salarial entre os homens de ambas as companhias, vemos uma situação inversa à das mulheres: os homens da Companhia B ganham mais que os homens da Companhia A. Assim como nas mulheres, os intervalos de confiança também ressaltam isso: a Companhia A apresenta um intervalo de confiança de 95% para o salário médio dos homens de [7.8, 9.9] e para a Companhia B [9.7, 11.4].

Quando comparado as mulheres e os homens de uma mesma companhia, temos para a Companhia A um boxplot que nos diz que os salários para ambos os sexos são muito similares; sendo que isso pode ser confirmado pelo teste de hipótese, onde não podemos afirmar que há uma diferença salarial discriminada pelo sexo. Já na Companhia B temos uma situação bem diferente: é possível afirmar, a partir do boxplot plotado, que os homens ganham mais que as mulheres; além disso, os testes de hipótese U de Mann-Whitney, Ranksum e Kolmogorov-Smirnov também nos comprovam que há a discriminação por sexo na Companhia B.

4ª Parte: Conclusões.

4.i) Conclusões sobre as diferenças salariais entre níveis de educação em cada companhia.

No que diz respeito aos salários da Educação Primária em ambas as companhias, a partir do boxplot criado no item 3.i, observamos que a Companhia A possui outliers enquanto que a Companhia B não, podemos observar também que as pessoas com escolaridade primária da Companhia A, possuem um salário relativamente maior se comparado com a B. O primeiro quartil e a mediana das duas companhias são muito próximos, sendo que o que difere bem uma da outra é o percentil 75.

Sobre os salários da Educação Secundária das duas companhias, observamos novamente nos boxplots que as medianas permanecem muito próximas e o percentil 25 e o 75 diferem de pouco, podendo, então, nos levar a concluir que os salários para as pessoas com escolaridade de nível secundário são parecidos em ambas as companhias.

Para a Educação Terciária, observamos que para ambas as companhias, o salário aumenta significativamente. No entanto, a partir do boxplot, podemos ver que, mesmo com esse aumento, a Companhia A paga muito melhor seus funcionários que a Companhia B: o percentil 25 da Companhia A está para o percentil 75 da Companhia B. Nota-se, então, uma grande discrepância de valores entre elas, apesar das semelhanças para as educações de nível primário e secundário.

4.ii) Conclusões sobre a participação de mulheres em cargos gerenciais nas duas companhias.

A partir do intervalo de confiança de 95% calculado no item 3.iv, observamos que para a Companhia A ele é [0.03, 0.09] e para a Companhia B é [0.05, 0.12], ou seja, nota-se uma participação feminina muito baixa em ambas as companhias. Tal fato pode ser certificado com o teste de hipótese utilizado no item 3.vii, pois não há dados suficientes que justifiquem a presença feminina, isso apenas ressalta a baixíssima presença de mulheres no cargo de gerente para as duas companhias.

4.iii) Conclusões da análise da possível diferença salarial entre homens e mulheres em cada companhia.

Como dito no item 3.viii, a partir dos boxplots de comparação salarial entre as mulheres da Companhia A e mulheres da Companhia B, pode-se afirmar que, apesar da Companhia A apresentar mais outliers que a B, as mulheres da Companhia A ganham mais que a da B; tal fato pode ser observado também pelos intervalos de confiança calculados: a Companhia A apresenta um intervalo de confiança de 95% para o salário médio das mulheres de [7.9, 9.7] enquanto que na Companhia B temos [5.5, 6.7]. Já, ao analisar os boxplots de comparação salarial entre os homens de ambas as companhias, vemos uma situação inversa à das mulheres: os homens da Companhia B ganham mais que os homens da Companhia A. Assim como nas mulheres, os intervalos de confiança também ressaltam isso: a Companhia A apresenta um intervalo de confiança de 95% para o salário médio dos homens de [7.8, 9.9] e para a Companhia B [9.7, 11.4].

Quando comparado as mulheres e os homens de uma mesma companhia, temos para a Companhia A um boxplot que nos diz que os salários para ambos os sexos são muito similares; sendo que isso pode ser confirmado pelo teste de hipótese, onde não podemos afirmar que há uma diferença salarial discriminada pelo sexo. Já na Companhia B temos uma situação bem diferente: é possível afirmar, a partir do boxplot plotado, que os homens ganham mais que as mulheres; além disso, os testes de hipótese U de Mann-Whitney, Ranksum e Kolmogorov-Smirnov também nos comprovam que há a discriminação por sexo na Companhia B.

4.v) Conclusões sobre qual companhia paga melhores salários.

Como podemos observar ao longo deste trabalho, há muitas variáveis que interferem o salário de um funcionário: o sexo, o nível educacional de cada indivíduo e o tipo de cargo que a pessoa empregada ocupa.

Analisando as tabelas e gráficos acima expostos, concluímos que para a questão de sexo, a Companhia A paga as mulheres melhor que a Companhia B, mas a Companhia B paga melhor os homens que a Companhia A. No que diz respeito à escolaridade primária e secundária, temos um bom equilíbrio entre as duas companhias; sendo bem divergentes à questão da escolaridade terciária, a Companhia A paga muito melhor seus funcionários de nível terciário do que a Companhia B. Para o tipo de cargo que os funcionários ocupam, observamos nas tabelas de distribuição conjunta, fornecidas na 2ª Parte, que os gerentes nas duas companhias ganham mais que os outros cargos, no entanto, os gerentes são em menor quantidade de funcionários.

Dessa forma, a partir de nossa análise, é possível inferir que o menor salário na Companhia A será para um homem com ensino primário ou secundário que ocupa um cargo que não seja gerencial; já para a Companhia B, seria uma mulher com ensino primário ou secundário que também ocupa um cargo não gerencial. Em termos de melhores salários, para a Companhia A teríamos uma mulher com ensino terciário e com cargo de gerente; para a Companhia B seria um homem com ensino terciário e com cargo de gerente. Isso nos leva a concluir que há uma discriminação salarial forte em ambas as companhias no que diz sentido ao nível escolar e ao sexo dos funcionários.