

1) Gráficos e Tabelas

a) 15 pontos) Elaborar os gráficos box-plot e histograma das variáveis “age” (idade da esposa) e “husage” (idade do marido) e comparar os resultados.

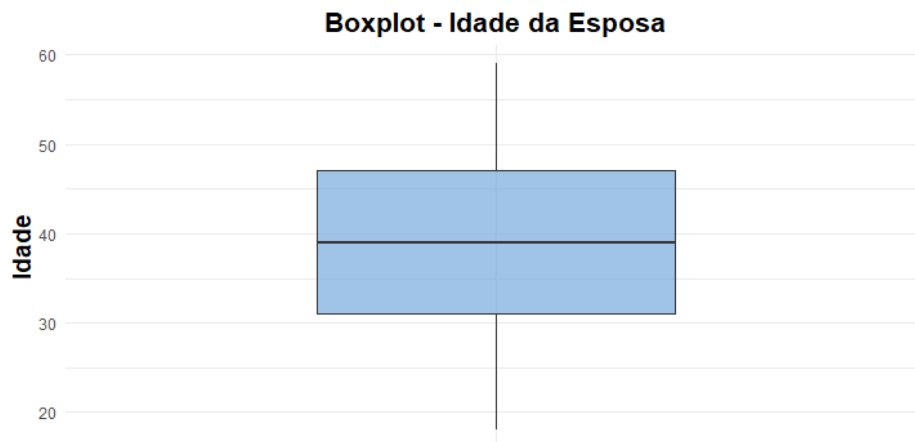


Figura 1 - Distribuição de idades das esposas usando box-plot

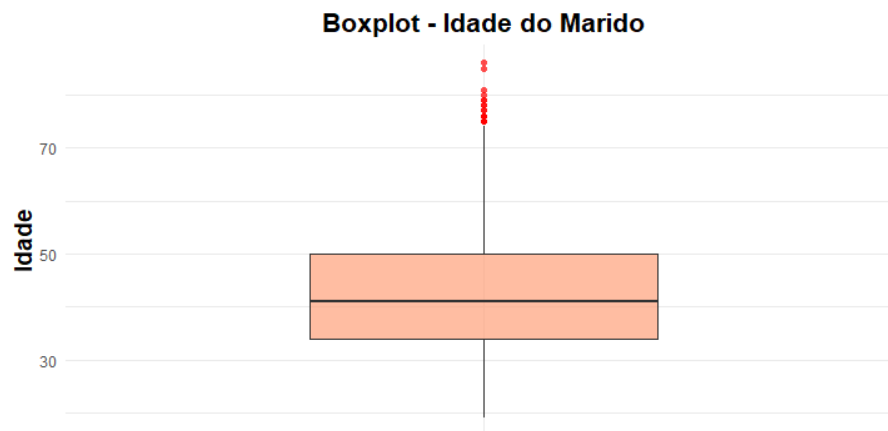


Figura 2 - Distribuição de idades dos maridos usando box-plot

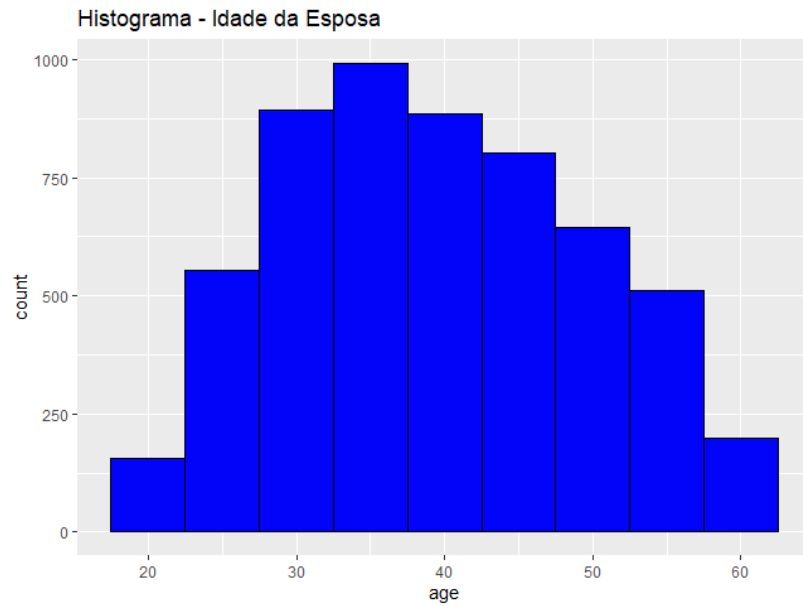


Figura 3 - Histograma idade das esposas

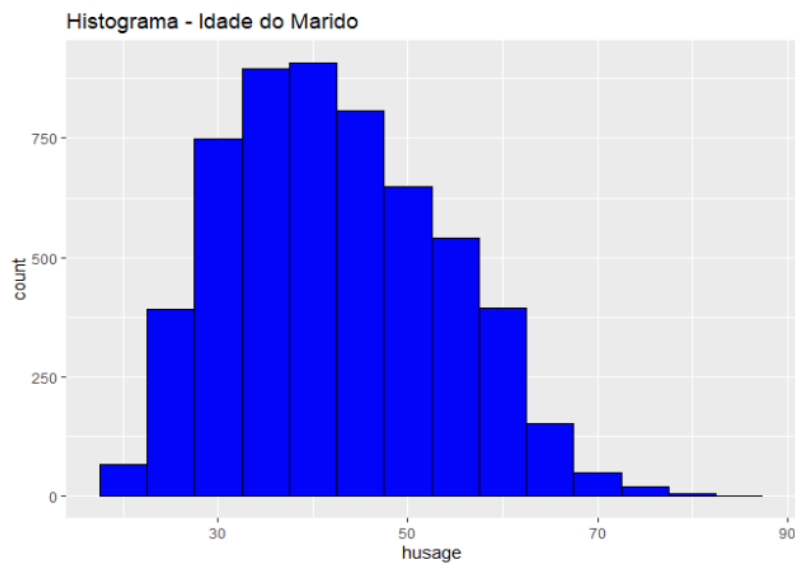


Figura 4 - Histograma idade dos maridos

Análise: Analisando os gráficos é possível notar que os maridos possuem uma idade maior em relação as esposas analisando pela mediana, o mesmo comportamento é observado no primeiro e terceiro quartil.

b) (15 pontos) Elaborar a tabela de frequências das variáveis “age” (idade da esposa) e “husage” (idade do marido) e comparar os resultados

Tabela de frequência idade marido

[1] "Tabela de Frequência da Idade do Marido:"				32	50	127	2.25417110
> print(df_husage)				33	51	112	1.98793042
	Idade	Frequência	Percentual	34	52	128	2.27192048
1	19	5	0.08874689	35	53	102	1.81043663
2	20	6	0.10649627	36	54	104	1.84593539
3	21	25	0.44373447	37	55	126	2.23642173
4	22	31	0.55023074	38	56	103	1.82818601
5	23	35	0.62122826	39	57	106	1.88143415
6	24	71	1.26020589	40	58	86	1.52644657
7	25	59	1.04721335	41	59	91	1.61519347
8	26	115	2.04117856	42	60	81	1.43769968
9	27	113	2.00567980	43	61	74	1.31345403
10	28	108	1.91693291	44	62	62	1.10046148
11	29	137	2.43166489	45	63	45	0.79872204
12	30	157	2.78665247	46	64	45	0.79872204
13	31	172	3.05289315	47	65	30	0.53248136
14	32	174	3.08839191	48	66	18	0.31948882
15	33	169	2.99964501	49	67	14	0.24849130
16	34	196	3.47887824	50	68	14	0.24849130
17	35	161	2.85764998	51	69	13	0.23074192
18	36	174	3.08839191	52	70	8	0.14199503
19	37	195	3.46112886	53	71	7	0.12424565
20	38	169	2.99964501	54	72	9	0.15974441
21	39	200	3.54987575	55	73	6	0.10649627
22	40	175	3.10614129	56	74	2	0.03549876
23	41	192	3.40788072	57	75	7	0.12424565
24	42	172	3.05289315	58	76	4	0.07099752
25	43	178	3.15938942	59	77	2	0.03549876
26	44	201	3.56762513	60	78	2	0.03549876
27	45	149	2.64465744	61	79	2	0.03549876
28	46	149	2.64465744	62	80	1	0.01774938
29	47	130	2.30741924	63	81	1	0.01774938
30	48	147	2.60915868	64	85	1	0.01774938
31	49	135	2.39616613	65	86	1	0.01774938
--	--	---	---				

Tabela de frequência idade esposas

	Idade	Frequência	Percentual				
1	18	12	0.2129925				
2	19	18	0.3194888				
3	20	31	0.5502307				
4	21	47	0.8342208				
5	22	47	0.8342208				
6	23	67	1.1892084				
7	24	84	1.4909478				
8	25	114	2.0234292				
9	26	114	2.0234292				
10	27	174	3.0883919				
11	28	161	2.8576500				
12	29	170	3.0173944				
13	30	184	3.2658857				
14	31	191	3.3901313				
15	32	187	3.3191338				
16	33	180	3.1948882				
17	34	204	3.6208733				
18	35	187	3.3191338				
19	36	205	3.6386226				
20	37	217	3.8516152				
21	38	202	3.5853745				
22	39	171	3.0351438				
23	40	181	3.2126376				
24	41	158	2.8044018				
25	42	172	3.0528931	34	51	117	2.0766773
26	43	185	3.2836351	35	52	138	2.4494143
27	44	185	3.2836351	36	53	103	1.8281860
28	45	140	2.4849130	37	54	109	1.9346823
29	46	164	2.9108981	38	55	104	1.8459354
30	47	128	2.2719205	39	56	91	1.6151935
31	48	131	2.3251686	40	57	104	1.8459354
32	49	125	2.2186723	41	58	104	1.8459354
33	50	133	2.3606674	42	59	95	1.6861910

Observações Principais:

1. Idade: A distribuição etária segue uma curva em forma de sino, com o pico ocorrendo entre os 20 e 30 anos e diminuindo à medida que a idade avança.

Isso é típico para uma população trabalhadora.

2. Salário Familiar: A distribuição dos salários familiares é assimétrica à direita, indicando que há mais famílias com salários baixos e menos com salários muito altos.

3. Tendências Centrais:

- **Idade:** A faixa etária mais comum (moda) está entre 20 e 30 anos.
- **Salário Familiar:** A moda (categoria de salário mais frequente)

provavelmente está na parte inferior da faixa.

4. Variabilidade:

- **Idade:** Existe uma boa variedade de idades, sugerindo que o conjunto de dados inclui pessoas em diferentes estágios de carreira.
- **Salário Familiar:** A variabilidade nos salários é provavelmente maior do que a da idade, como é comum em distribuições de renda desiguais.

2) Medidas de posição e dispersão

a) 15 pontos) Calcular a média, mediana e moda das variáveis “age” (idade da esposa) e “husage” (idade do marido) e comparar os resultados

Medidas “age” – esposa

Média: 39.42758

Mediana: 39

Moda: 37

Medidas “husage” – Marido

Média: 42.45296

Mediana: 41

Moda: 44

Observações Gerais:

Distribuição de Idade: A média e a mediana para "idade" são relativamente próximas, sugerindo uma distribuição bastante simétrica.

Idade do Chefe de Família: A média para "husage" é ligeiramente superior à mediana, indicando que a distribuição pode estar levemente inclinada para a direita (algumas famílias com chefes de família mais velhos estão elevando a média).

b) (15 pontos) Calcular a variância, desvio padrão e coeficiente de variação das variáveis “age” (idade da esposa) e “husage” (idade do marido) e comparar os resultados

Medidas “age” - esposa

Variância: 99.75234

Desvio padrão: 9.98761

Coeficiente de variação: 25.33153

Medidas “husage” - marido

Variância: 126.0717

Desvio padrão: 11.22817

Coeficiente de variação: 26.44849

Variância:

- **Idade:** A variância da idade é 99,75. Isso significa que há um certo grau de dispersão nas idades dentro do seu conjunto de dados.
- **Idade Chefe Família:** A variância da idade do chefe da família é 126,07. Isso indica uma dispersão ligeiramente maior das idades dos chefes de família em comparação com a distribuição geral de idade.

Desvio Padrão:

- **Idade:** O desvio padrão da idade é 9,99. Em média, as idades individuais em seu conjunto de dados desviam-se da idade média em cerca de 9,99 anos.
- **Idade Chefe Família:** O desvio padrão da idade do chefe da família é 11,23. Em média, as idades dos chefes de família desviam-se da média da idade do chefe da família em cerca de 11,23 anos.

Coeficiente de Variação (CV):

- **Idade:** O CV para idade é de 25,33%, indicando um nível moderado de variabilidade relativa nos dados de idade.
- **Idade Chefe Família:** O CV para a idade do chefe da família é de 26,45%, indicando também um nível moderado de variabilidade relativa. A variabilidade da idade dos chefes de família é ligeiramente superior à variabilidade geral da idade.

3) Testes paramétricos e não paramétricos

a) (40 pontos) Testar se as médias (se você escolher o teste paramétrico) ou as medianas (se você escolher o teste não paramétrico) das variáveis “age” (idade da esposa) e “husage” (idade do marido) são iguais, construir os intervalos de confiança e comparar os resultados.

1) Testada normalidade com Shapiro-Wilk, utilizando uma subamostra e também com Kolmogorov-Smirnov utilizando a amostra completa.

Ambos obtendo o resultado de ***p-value < 2.2e-16***

2) Utilizado o teste Levene com centro na mediana, foi obtido o seguinte resultado:

```
> leveneTest(salarios$age, salarios$husage, center = "median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value    Pr(>F)
group  64  4.0244 < 2.2e-16 ***
      5569
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

3) Devido a este resultado de normalidade, e o resultado do teste de Levene onde apontou variâncias desiguais, ou seja, o valor p extremamente pequeno ($< 2,2e-16$) no teste de Levene é uma forte indicação de que as variâncias dos grupos que em comparação não são iguais. Significando assim que há uma diferença significativa na distribuição dos dados entre os grupos. Por isso foi escolhido o *Welch Two Sample t-test*.

```
> welch_test_result <- t.test(salarios$age, salarios$husage, var.equal = FALSE)
> print(welch_test_result)
```

Welch Two Sample t-test

```
data: salarios$age and salarios$husage
t = -15.111, df = 11115, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.417821 -2.632943
sample estimates:
mean of x mean of y
 39.42758  42.45296
```