18-04-2021

Trabalho 1 Métodos Estatísticos

Insurance

Gonçalo Pratas: 180221094

João Santos: 170221029

Margarida Maunu: 190221070

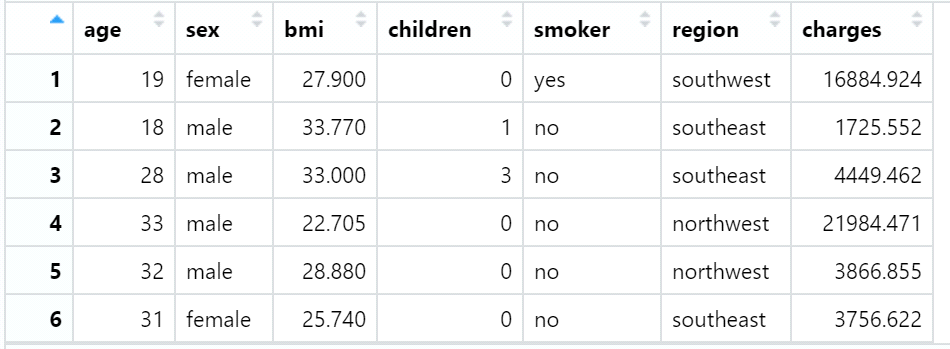
Descrição de Dados

Enquadramento

Este trabalho foi realizado no âmbito da disciplina de **Métodos Estatísticos** na qual foi nos fornecido um conjunto de dados referente a uma companhia de Seguro “X” onde tinha como objetivo fazer uma análise descritiva com recurso ao Software R.

Ao longo deste relatório será apresentado os resultados e conclusões dessa mesma análise.

Análise Superficial

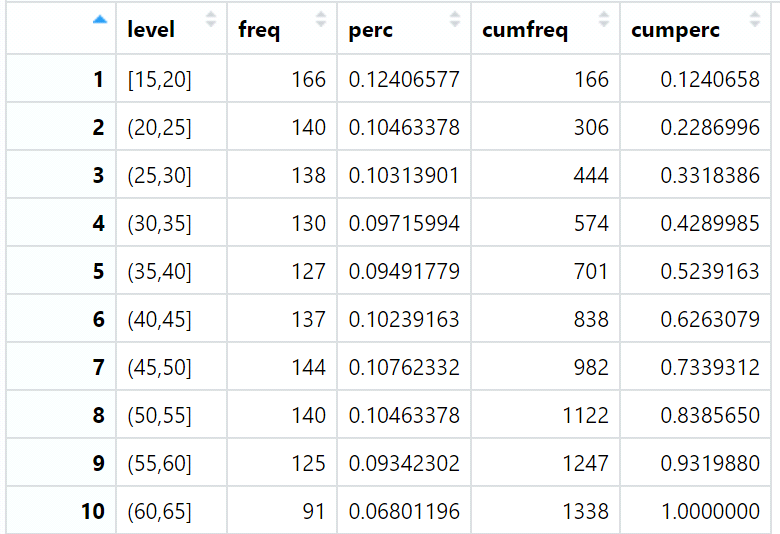


*Figura 1 - Base de Dados Insurance*

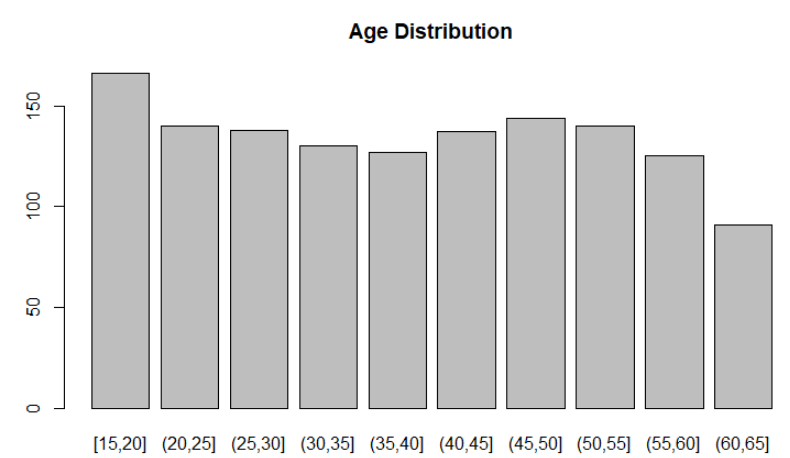
À primeira vista podemos constatar uma base de dados relativa a uma possível empresa de seguros, que possui cerca de 1338 registos de “clientes”.

Cada cliente possui: uma idade (age), um género (sex), um índice de massa corporal (bmi), números de filhos (children), se fuma ou não (smoker), a sua região (region) e finalmente quanto paga de seguro (charges).

Análise das Idades



*Figura 2- Tabela de Frequências Idades*

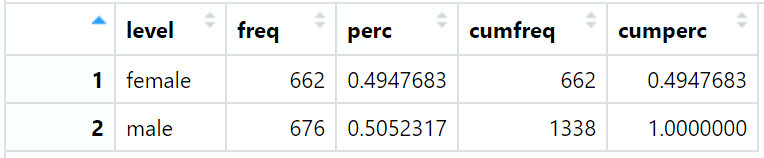


*Figura 3 -Grafico de barras com a distribuição de idades*

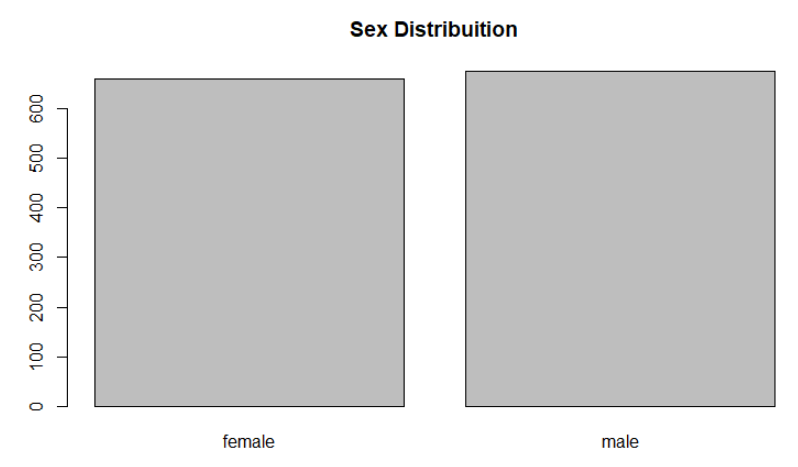
Como podemos ver esta amostra possuí um intervalo de idades que vai desde os **15 até aos 65** anos de idade. Onde **12.41% (frequência relativa)** representa as **pessoas dos 15 aos 20 anos** com uma **frequência absoluta de 166 indivíduos**, **10.46% (frequência relativa)** representa as **pessoas de 20 aos 25 anos** com uma **frequência absoluta de 140 indivíduos**, **10.31% ( frequência relativa)** representa as **pessoas de 25 aos 30 anos** com uma **frequência absoluta de 138 indivíduos**, **9.72% (frequência relativa)** representa as **pessoas de 30 aos 35 anos** com uma **frequência absoluta de 130 indivíduos**, **9.49% (frequência relativa)** representa as **pessoas de 35 aos 40 anos** com **uma frequência absoluta de 127 indivíduos**, **10.24% (frequência relativa)** representa as **pessoas de 40 aos 45 anos** com **uma frequência absoluta de 137 indivíduos**, **10.76% (frequência relativa)** representa as **pessoas de 45 a 50** **anos** com uma **frequência absoluta de 144 indivíduos**, **10.46% (frequência relativa)** representa as **pessoas de 50 a 55 anos** com uma **frequência absoluta de 140 indivíduos**, **9.34% (frequência relativa)** representa as **pessoas de 55 a 60 anos** com uma **frequência absoluta de 125 indivíduos**, **6.80% (frequência relativa)** representa as **pessoas de 60 a 65 anos** com uma **frequência absoluta de 91 indivíduos**.

A amostra possui uma **média de idades igual a 39.20703**, **mediana igual a 39**, **moda igual a 18** e um **desvio padrão de 14.04996** concluindo que está bem distribuída em relação à média.

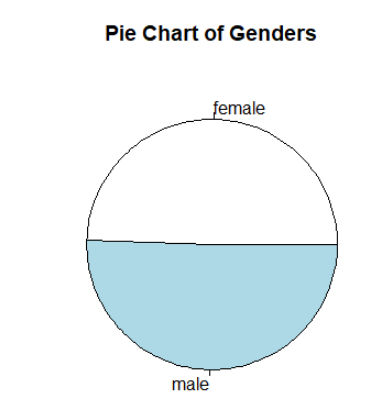
Análise do género



*Figura 4 - Tabela de Frequências Género*



*Figura 5 - Grafico de barras da distribuição de sexos*

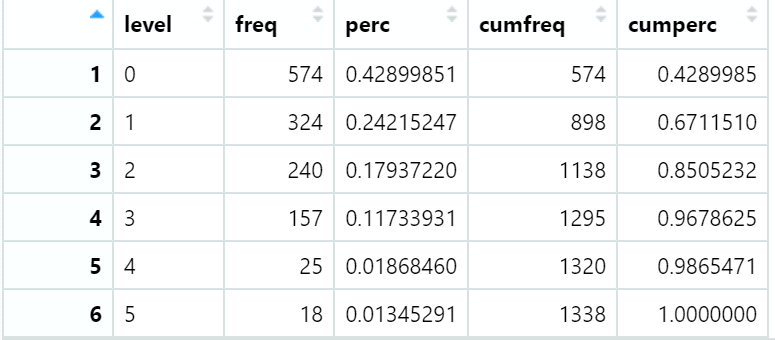


*Figura 6 - Pie Chart da distribuição de sexos*

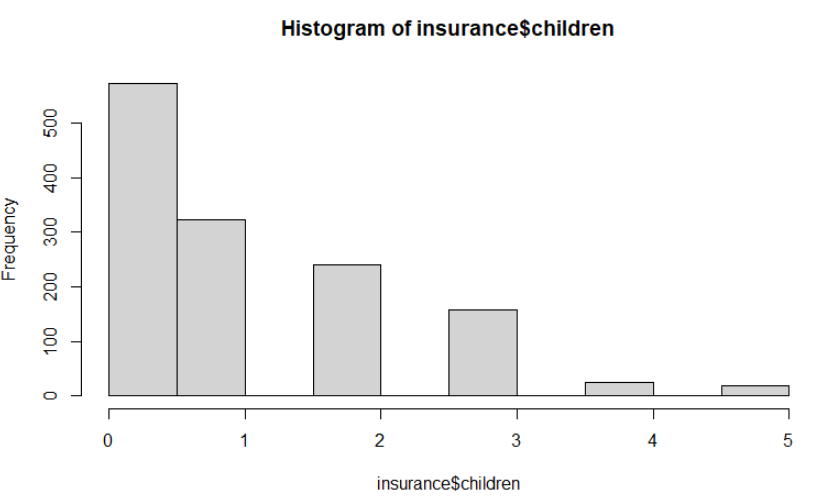
Como podemos ver temos **49.48% (frequência relativa)** **de** **mulheres com 662 indivíduos (frequência absoluta)** e **50.52% (frequência relativa) de homens** **com 676 indivíduos (frequência absoluta).**

A amostra esta bastante **uniforme** em termos de género no entanto temos cerca de **mais 14 homens do que mulheres tornando eles a maioria neste caso.**

Análise do número de filhos



*Figura 7 - Tabela de Frequência Número de Filhos*

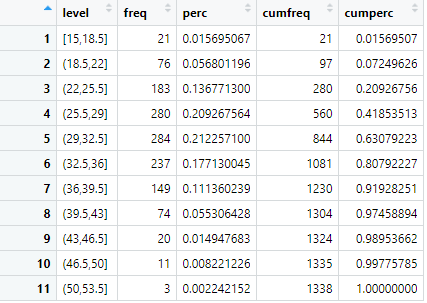


*Figura 8 - Historial do Número de Filhos por individuo*

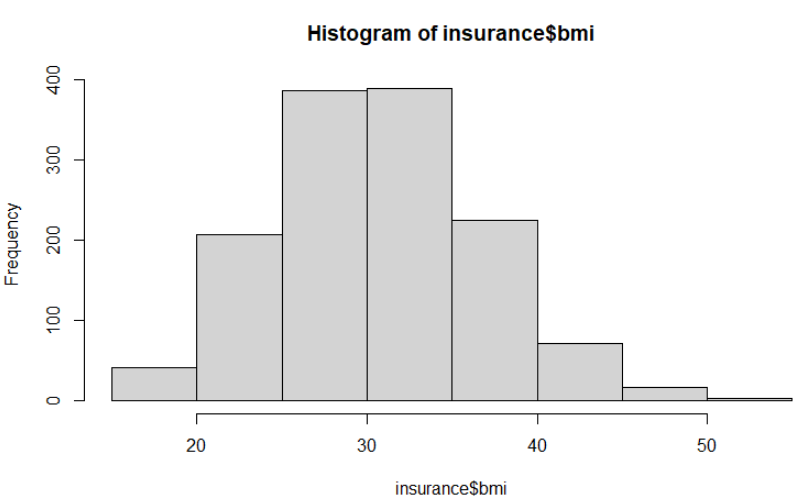
Como podemos ver temos **42.90% (frequência relativa)** **de** **pessoas sem filhos** e com uma **frequência absoluta de 574 indivíduos**, **24,22% ( frequência relativa)** **de** **pessoas com 1 filho** e com uma **frequência absoluta de 324 de indivíduos**, **17,94% ( frequência relativa) de pessoas com 2 filhos** e com uma **frequência absoluta de 240 indivíduos**, **11,73% (frequência relativa) de pessoas com 3 filhos** com uma **frequência absoluta de 157 indivíduos**, **1.87% (frequência relativa) de pessoas com 4 filhos** com uma **frequência absoluta de 25** e **1.35% (frequência relativa) de pessoas com 5 filhos** com uma **frequência absoluta de 18 indivíduos**.

A amostra possui uma **média de filhos igual a 1.094918**, **mediana igual 1**, **moda igual a 0** e um **desvio padrão de 1.205493** concluindo que temos os dados muito agrupados nos 0 e 1.

Análise BMI



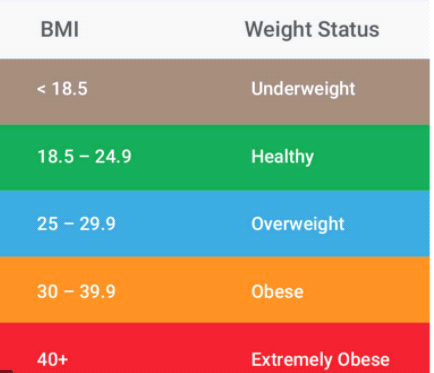
*Figura 9 - Tabela de frequências BMI*



*Figura 10 -Histograma do BMI*

Como podemos ver temos **1.56% ( frequência relativa) de pessoas com um nível de BMI dos 15 aos 18.5** com uma **frequência absoluta de 21 indivíduos**, **5.68 % ( frequência relativa) de pessoas com um nível de BMI dos 18.5 aos 22** com uma **frequência absoluta de 76 indivíduos**, **13.68% ( frequência relativa) de pessoas com um nível de BMI dos 22 aos 25.5** com uma **frequência absoluta de 183 indivíduos**, **20.92 % ( frequência relativa) de pessoas com um nível de BMI dos 25.5 aos 29** com uma **frequência absoluta de 280 indivíduos**, **21.23% ( frequência relativa) de pessoas com um nível de BMI dos 29 aos 32.5** com uma **frequência absoluta de 284 indivíduos**, **17.71% ( frequência relativa) de pessoas com um nível de BMI dos 32.5 aos 36** com uma **frequência absoluta de 237 indivíduos**, **11.14% ( frequência relativa) de pessoas com um nível de BMI dos 36 aos 39.5** com uma **frequência absoluta de 149 indivíduos**, **5.53% ( frequência relativa) de pessoas com um nível de BMI dos 39.5 aos 43** com uma **frequência absoluta de 74 indivíduos**, **1.49% ( frequência relativa) de pessoas com um nível de BMI dos 43 aos 46.5** com uma **frequência absoluta de 20 indivíduos**, **0.82% ( frequência relativa) de pessoas com um nível de BMI dos 46.5 aos 50** com uma **frequência absoluta de 11 indivíduos** e **0.22% ( frequência relativa) de pessoas com um nível de BMI dos 50 aos 53.5** com uma **frequência absoluta de 3 indivíduos**.

Enquadramento BMI

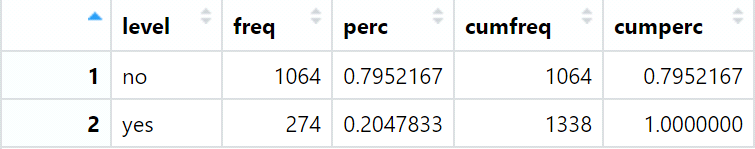


*Figura 11 - Tabela de valores BMI*

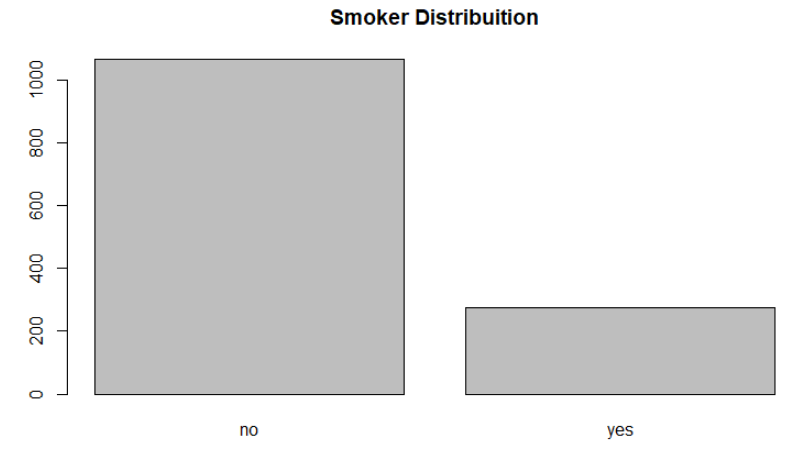
A amostra possui uma **media de BMI igual a 30.6634**, **mediana igual a 30.4** e com um **desvio padrão de 6.098187** que prova um certo agrupação sobretudo nos 25 a 30 e nos 30 aos 35.

**Nota: É discutível se podemos considerar os obesos extremos e os abaixo do peso como outliers.**

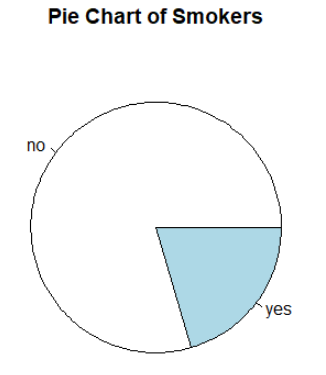
Análise Fumadores



*Figura 12- Tabela de frequências Fumadores*



*Figura 13- Grafico de barras da quantidade de fumadores*

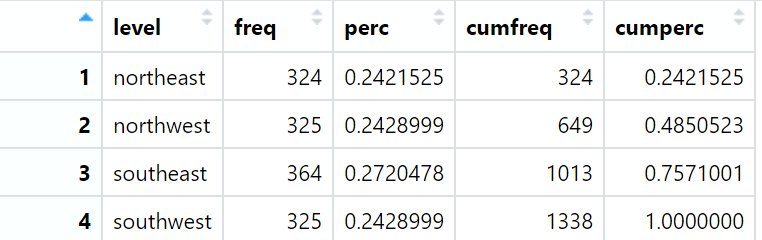


*Figura 14- Pie Chart da quantidade de fumadores*

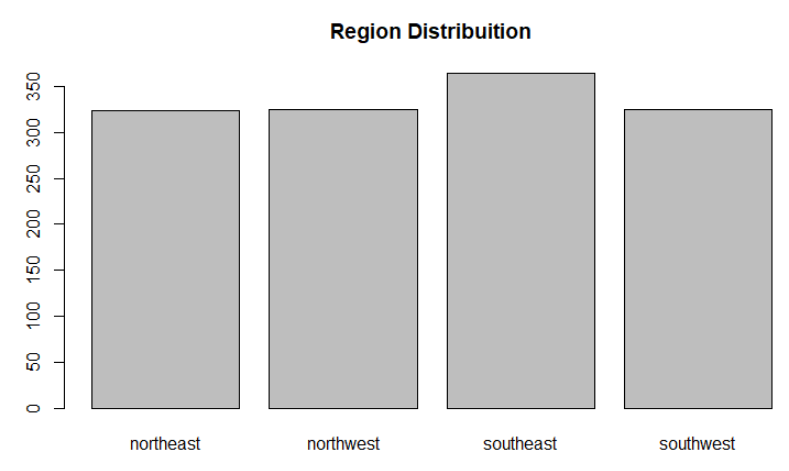
Como podemos ver temos **79.52% (frequência relativa) de pessoas que não são fumadores** com uma **frequência absoluta de 1064 indivíduos** e **20.48% (frequência relativa) de pessoas que são fumadores** com uma **frequência absoluta de 274 indivíduos**.

Existe uma grande dispersão entre fumadores e não fumadores sendo os não fumadores a maioria.

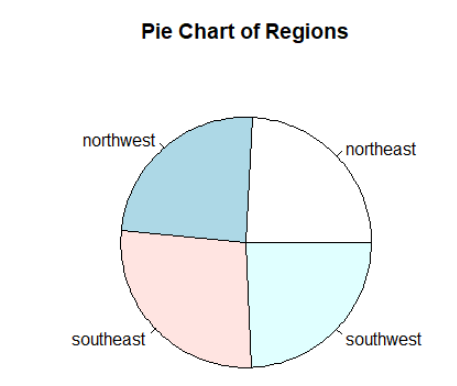
Análise das Regiões



*Figura 15 - Tabela de frequências Regiões*



*Figura 16 - Grafico de barras da distribuição por Regiões*

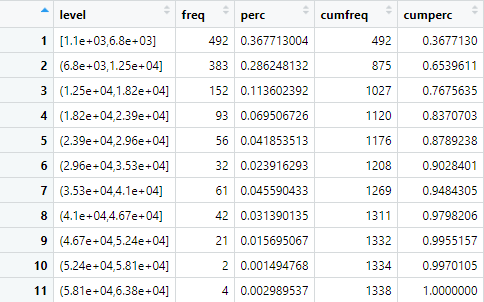


*Figura 17 - Pie Chart da distribuição por Regiões*

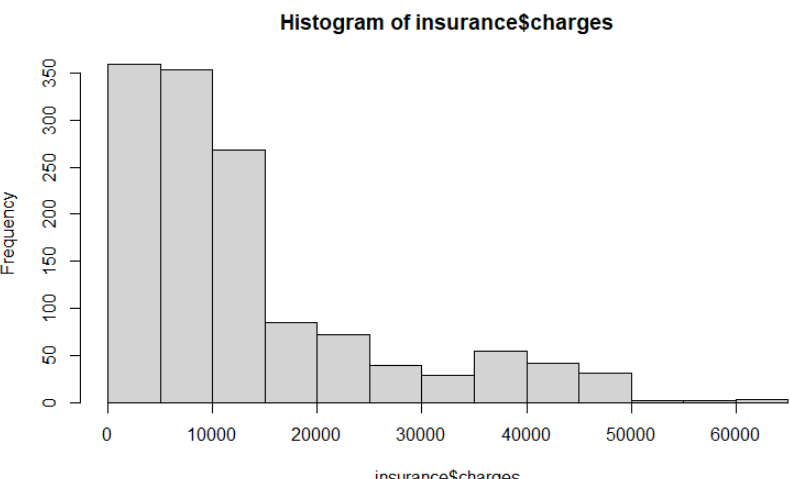
Como podemos ver temos **24.22% (frequência relativa) de pessoas pertencem a zona “northeast”** com uma **frequência absoluta de 324 indivíduos**, **24.30% (frequência relativa) de pessoas pertencem a zona “northwest”** com uma **frequência absoluta de 325 indivíduos**, **27.20% (frequência relativa) de pessoas pertencem a zona “southeast”** com uma **frequência absoluta de 364 indivíduos** e **24.29% (frequência relativa) de pessoas pertencem a zona “southwest”** com uma **frequência absoluta de 325 indivíduos**.

A amostra esta distribuída bastante uniformemente, mas podemos verificar cerca de 364 em southeast tornando a maioria.

Análise das Charges



*Figura 18 - Tabela de frequências Charges*



*Figura 19 - Histograma de Charges*

Como podemos ver temos **36.8% (frequência relativa) de pessoas que pagam entre “1100 e 6800”** com uma **frequência absoluta de 492 indivíduos**, **28.6% (frequência relativa) de pessoas que pagam entre 6800 e 12500** com uma **frequência absoluta de 383 indivíduos**, **11.36% (frequência relativa) de pessoas pagam entre 12500 e 18200** com uma **frequência absoluta de 152 indivíduos**, **06.95% (frequência relativa) de pessoas que pagam entre 18200 e 23900** com uma **frequência absoluta de 93 indivíduos**, **04.18% (frequência relativa) de pessoas que pagam entre 23900 e 29600** com uma **frequência absoluta de 56 indivíduos**, **02.39% (frequência relativa) de pessoas que pagam entre 29600 e 35300** com uma **frequência absoluta de 32 indivíduos**, **04.55% (frequência relativa) de pessoas que pagam entre 35300 e 41000** com uma **frequência absoluta de 61 indivíduos**, **03.13% (frequência relativa) de pessoas pagam entre 41000 e 46700** com uma **frequência absoluta de 42 indivíduos**, **01.57% (frequência relativa) de pessoas pagam entre 46700 e 52400** com uma **frequência absoluta de 21 indivíduos**, **0.15% (frequência relativa) de pessoas que pagam entre 52400e 58100** com uma **frequência absoluta de 2 indivíduos** e **0.30% (frequência relativa) de pessoas que pagam entre 58100 e 63800** com uma **frequência absoluta de 4 indivíduos**.

A amostra está distribuída bastante uniformemente, permite verificar que a maioria das pessoas esta concentrada entre os 1100 e os 182000 decrescendo à medida que as despesas aumentam.

A média das despesas é de 13270.42, mediana de 9382.033.

Relações Encontradas

**Fumadores x Seguro**

Após a análise verificamos que os fumadores pagam em média 32050.23 euros enquanto os não fumadores pagam em média 8434.268, **concluindo que pessoas não fumadoras pagam menos em seguro em comparação a pessoas fumadoras.**

**Obesidade x Seguro**

Após a análise verificamos que as pessoas consideradas obesas (que apresentam um BMI superior a 30 de acordo com a figura 11) pagam em média 15560.93 euros enquanto as pessoas abaixo de 30 pagam em média 10713.67 euros, **concluindo que pessoas com níveis altos de BMI pagam mais seguro que as pessoas com níveis mais baixos.**

**Filhos x Seguro**

Após a análise verificamos que as pessoas que não tem filhos pagam em média 12365.98 euros, pessoas com 1 filho pagam em média 12731.17 euros e pessoas com mais de 1 filho pagam em média 14847.4 euros**, concluindo que pessoas sem filhos ou 1 filho pagam aproximadamente o mesmo mas no entanto pessoas com mais do que 1 filho pagam mais em seguros que os outros.**