

1º TRABALHO DE GRUPO

Métodos Estatísticos

Síntese

Neste relatório, é possível verificar toda a análise relativamente apenas à Estatística Descritiva de um conjunto de dados previamente fornecido pelo docente responsável.

GRUPO 1:

Alexandre Coelho - 190221093

Sérgio Veríssimo - 190221128

Tim Rodrigues - 190221131

Índice

Introdução.....	4
Descrição dos dados.....	4
Variáveis Qualitativas Nominais.....	5
Holiday – Feriados	5
Seasons – Estações do ano.....	7
Functioning.Day – Dia Funcional	9
Variável Qualitativa Ordinal	11
Date – Data.....	11
Variáveis Quantitativas Discretas.....	12
Hour – Hora do dia	12
Rented Bike Count – Contagem de alugueis de bicicletas	13
Variáveis Quantitativas Contínuas	16
Temperature – Temperatura	16
Humidity – Humidade	22
Wind Speed (ms) – Velocidade do vento	28
Visibility (10m) – Visibilidade	34
Dew point Temperature (Celsius) – Temperatura do orvalho	40
Solar Radiation – Radiação Solar.....	46
Rainfall (mm) – Precipitação (mm).....	52
Snowfall (mm) – Queda de neve (mm)	58
Conclusão	64
Notas complementares.....	65
Referências bibliográficas	67
Bibliografia	67
Equações utilizadas	67

Índice de figuras

Figura 1 - Gráfico de Barras das frequências absolutas da variável 'Holiday'	6
Figura 2 - Diagrama Circular das frequências absolutas da variável 'Holiday'	7
Figura 3 - Gráfico de Barras das frequências absolutas da variável 'Seasons'	8
Figura 4 - Diagrama Circular das frequências absolutas da variável 'Seasons'	9
Figura 5 - Gráfico de Barras das frequências absolutas da variável 'Functioning.Day'	10
Figura 6 - Diagrama Circular das frequências absolutas da variável 'Functioning.Day'	11
Figura 7 - Histograma das classes da variável 'Temperature'	17
Figura 8 - Diagrama de extremos e quartis da variável 'Temperature'	19
Figura 9 - Histograma das classes da variável 'Humidity'	23
Figura 10 - Diagrama de extremos e quartis da variável 'Humidity'	25
Figura 11 - Histograma das classes da variável 'Wind Speed'	29
Figura 12 - Diagrama de extremos e quartis da variável 'Wind Speed'	31
Figura 13 - Histograma das classes da variável 'Visibility'	35
Figura 14 - Diagrama de extremos e quartis da variável 'Visibility'	37
Figura 15 - Histograma das classes da variável 'Dew point Temperature'	41
Figura 16 - Diagrama de extremos e quartis da variável 'Dew Point Temperature'	43
Figura 17 - Histograma das classes da variável 'Solar Radiation'	47
Figura 18 - Diagrama de extremos e quartis da variável 'Solar Radiation'	49
Figura 19 - Histograma das classes da variável 'Rainfall'	53
Figura 20 - Diagrama de extremos e quartis da variável 'Rainfall'	55
Figura 21 - Histograma das classes da variável 'Snowfall'	59
Figura 22 - Diagrama de extremos e quartis da variável 'Snowfall'	61

Índice de tabelas

Tabela 1 - Tabela de Frequências da variável 'Holiday'	6
Tabela 2 - Tabela de Frequências da variável 'Seasons'	8
Tabela 3 - Tabela de Frequências da variável 'Functioning.Day'	10
Tabela 4 - Tabela de Frequências da variável 'Hour'	13
Tabela 5 - Tabela de Frequências da variável 'Rented Bike Count'	14
Tabela 6 - Tabela de Frequências da variável 'Temperature'	16
Tabela 7- Tabela de Número de Bicicletas por classe de Temperatura	17
Tabela 8 - Tabela de Frequências da variável 'Humidity'	22
Tabela 9 - Tabela de Frequências da variável 'Wind Speed'	28
Tabela 10 - Tabela de Frequências da variável 'Visibility'	34
Tabela 11- Tabela de Frequências da variável 'Dew point Temperature'	40
Tabela 12 - Tabela de Frequências da variável 'Solar Radiation'	46
Tabela 13 - Tabela de Frequências da variável 'Rainfall'	52
Tabela 14 - Tabela de Frequências da variável 'Snowfall'	58

Introdução

Conforme requisitado pelo enunciado do 1º Trabalho de Grupo, foi implementada uma análise detalhada, apenas dentro da área de Estatística Descritiva de um conjunto de dados fornecido pelo docente responsável. O conjunto de dados em questão retrata a procura de aluguer de bicicletas na cidade de Seoul na Coreia do Sul, sendo que é composto por uma amostra constituída por 8760 alugueis. Este conjunto de dados encontra-se descrito no ponto que se encontra a seguir neste relatório. Ou seja, mais resumidamente temos como população as bicicletas alugáveis em Seoul na Coreia do Sul e como amostra 8760 bicicletas que foram efetivamente alugadas. Durante a análise aos dados com recurso ao *RStudio*, surgiram alguns constrangimentos, e também abordamos certos métodos para a resolução de alguns problemas. Todos estes métodos, e constrangimentos podem ser consultados na secção de [Notas Complementares](#).

Descrição dos dados

Os dados encontram-se categorizados e divididos em 14 variáveis diferentes. Estas variáveis são:

- **Date** – Que indica a data do aluguer da bicicleta;
- **Rented Bike Count** – A frequência absoluta de bicicletas numa determinada data;
- **Hour** – A hora do dia em que foram efetuados os alugueis de bicicletas;
- **Temperature (Celsius)** – A temperatura numa certa hora e data;
- **Humidity** – A percentagem de humidade numa certa hora e data;
- **Wind Speed (ms)** – A velocidade do vento numa certa hora e data;
- **Visibility (10m)** - Visibilidade num espaço de 10 metros, numa certa hora e data;
- **Dewpoint temperature (Celsius)** – Temperatura de orvalho, numa certa hora e data;
- **Solar Radiation** – Radiação solar, numa certa hora e data;
- **Rainfall (mm)** – Precipitação, numa certa hora e data;
- **Snowfall (mm)** – Queda de neve, numa certa hora e data;
- **Seasons** – Estações do ano;
- **Holiday** – Se o aluguer foi efetuado num feriado ou não;
- **Functioning.Day** – Se o aluguer foi feito a horas funcionais ou não.

Estas variáveis podem ser agrupadas consoante a sua classificação:

- **Variável Qualitativa Nominal** – Holiday, Seasons e Functioning.Day;
- **Variável Qualitativa Ordinal** – Date;
- **Variável Quantitativa Discreta** – Hour, Rented Bike Count;
- **Variável Quantitativa Contínua** – Temperature, Humidity, Wind Speed (ms), Visibility (10m), Dewpoint temperature (Celsius), Solar Radiation, Rainfall (mm), Snowfall (mm).

Variáveis Qualitativas Nominais

Começando por fazer uma descrição dos dados relativamente às variáveis qualitativas nominais, iremos retratar toda a análise das mesmas, referindo previamente, que não foi possível aplicar todas as medidas de localização, devido a não serem aplicáveis em variáveis qualitativas. A única medida de localização que poderíamos aplicar e que eventualmente aplicamos, foi a moda.

Optamos também pela utilização de gráficos de barras e pelo diagrama circular para representação gráfica destas variáveis, por ser a que é frequentemente utilizada e recomendada para visualização de dados que se encaixem neste tipo de variáveis.

Holiday – Feriados

Esta variável, como mencionada acima, apenas é referente à indicação de que o aluguer de bicicletas foi efetuado durante um feriado ou não. Ou seja, é uma variável que é constituída apenas por duas respostas possíveis – “Holiday” ou “No Holiday”.

Como referido acima, não existiu aplicação de medidas de localização como a média e quantis, devido a não serem aplicados a variáveis qualitativas nominais. Entretanto foi aplicada a moda nesta variável e foi determinado que no meio de 8760 alugueis, a moda foi “No Holiday” ou sim, com uma frequência absoluta de 8328, ou uma frequência relativa de 0.951 (95.1%). Noutra perspetiva, “Holiday” teve apenas uma frequência absoluta de 432, ou uma frequência relativa de 0.049 (4.9%). Tudo isto pode ser visualizado na tabela de frequências da variável que se encontra presente na [Tabela 1 – Tabela de Frequências da variável ‘Holiday’](#).

Feriado (Holiday)	Frequência Absoluta	Frequência Relativa (em %)	N_i ^[1]	F_i ^[2] (em %)
Holiday	8328	95.1	8328	95.1
No Holiday	432	4.9	8760	100
	8760	100		

Tabela 1 - Tabela de Frequências da variável 'Holiday'

Também anteriormente referido, estamos limitados à representação gráfica dos dados, devido a este ser uma variável qualitativa nominal, pelo que obtemos os gráficos presentes na [Figura 1 – Gráfico de Barras das frequências absolutas da variável 'Holiday'](#) e na [Figura 2 – Diagrama Circular das frequências absolutas da variável 'Holiday'](#).

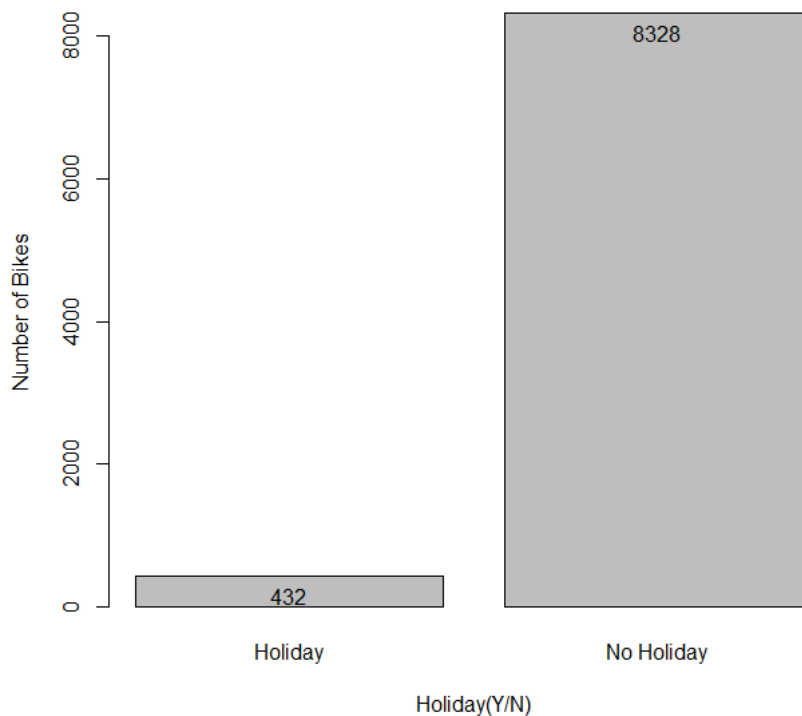


Figura 1 - Gráfico de Barras das frequências absolutas da variável 'Holiday'

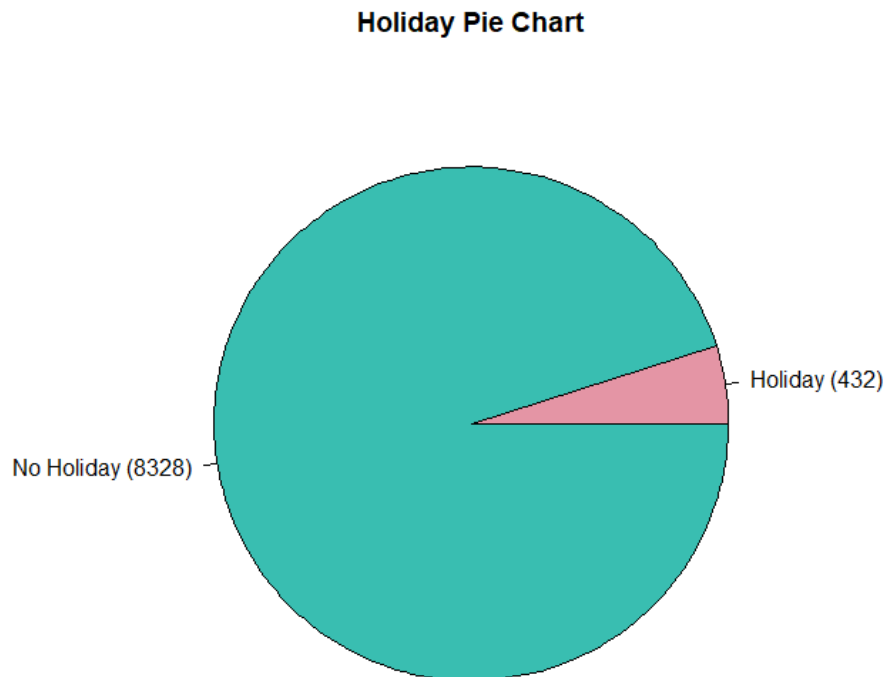


Figura 2 - Diagrama Circular das frequências absolutas da variável 'Holiday'

Seasons – Estações do ano

Esta variável, como mencionada acima, apenas é referente à estação do ano em que o aluguer de bicicletas foi efetuado. Sendo estações do ano, apenas teríamos quatro opções, “*Spring*” (Primavera), “*Summer*” (Verão), “*Autumn*” (Outono) e “*Winter*” (Inverno).

Como referido acima, não existiu aplicação de medidas de localização como a média e quantis, devido a não serem aplicados a variáveis qualitativas nominais. Entretanto foi aplicada a moda nesta variável e foi determinada a existência de duas entradas de dados como moda, ou seja, relativamente a esta variável, podemos afirmar que é bimodal, sendo as respetivas modas “*Spring*” (Primavera) e “*Summer*” (Verão), cada uma delas com uma frequência absoluta de 2208 bicicletas num conjunto de 8760 ou uma frequência relativa de 25.2% (cada uma). É possível consultar as frequências absolutas e as frequências relativas de cada uma das estações na respetiva tabela de frequências da variável que se encontra presente na [Tabela 2 – Tabela de Frequências da variável ‘Seasons’](#).

Estações do Ano (Seasons)	Frequência absoluta	Frequência Relativa (em %)	N_i ^[1]	F_i ^[2] (em %)
Spring (Primavera)	2208	25.2	2208	25.2
Summer (Verão)	2208	25.2	4416	50.4
Winter (Inverno)	2184	24.9	6600	75.3
Autumn (Outono)	2160	24.7	8760	100
	8760	100		

Tabela 2 - Tabela de Frequências da variável 'Seasons'

Também anteriormente referido, estamos limitados à representação gráfica dos dados, devido a este ser uma variável qualitativa nominal, pelo que obtemos os gráficos presentes na [Figura 3 – Gráfico de Barras das frequências absolutas da variável 'Seasons'](#) e na [Figura 4 – Diagrama Circular das frequências absolutas da variável 'Seasons'](#).

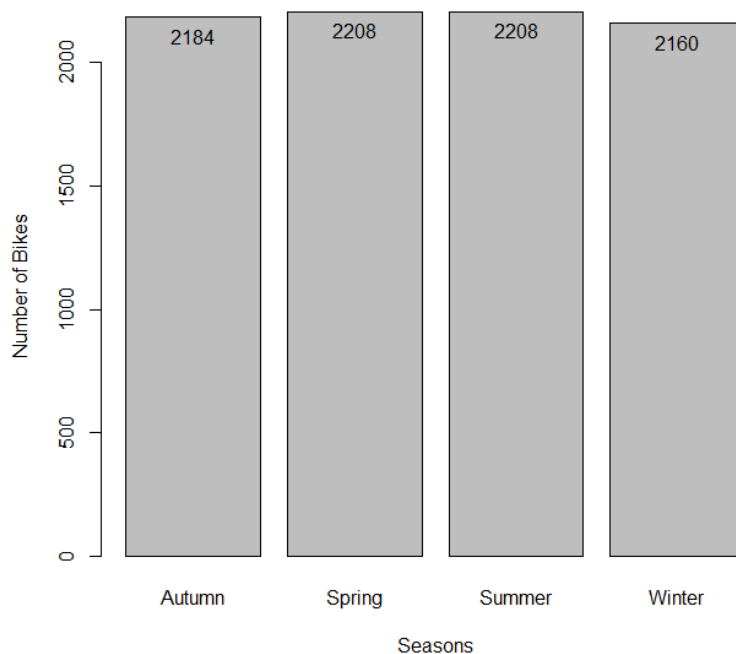


Figura 3 - Gráfico de Barras das frequências absolutas da variável 'Seasons'

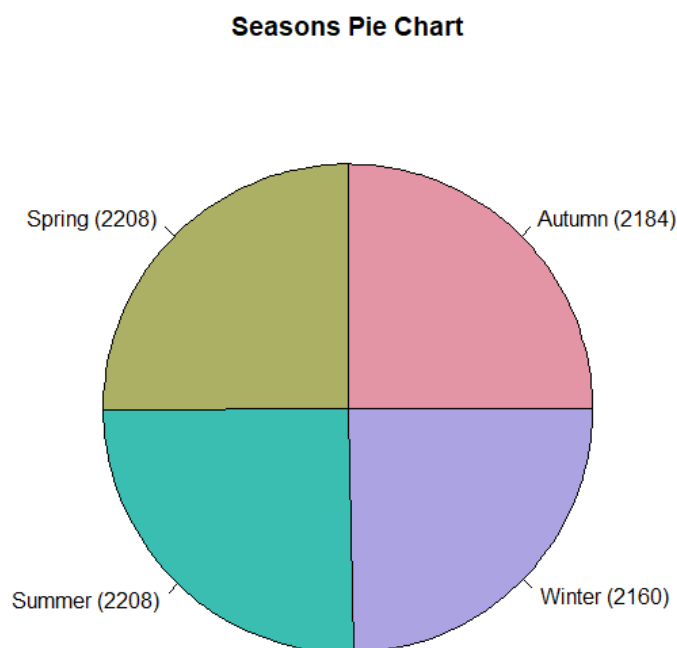


Figura 4 - Diagrama Circular das frequências absolutas da variável 'Seasons'

Functioning.Day – Dia Funcional

Esta variável, como mencionada acima, apenas é referente à informação se os alugueis das bicicletas foram efetuados durante horas funcionais ou não.

Como referido acima, não existiu aplicação de medidas de localização como a média e quantis, devido a não serem aplicados a variáveis qualitativas nominais. Entretanto foi aplicada a moda nesta variável e foi determinado que na amostra total de 8760 alugueis, a moda foi “Yes”, ou seja, foram alugadas mais bicicletas em horas funcionais, do que em horas não funcionais. Esta moda é constituída por uma frequência absoluta 8465 alugueis de 8760 ou por uma frequência relativa 0.966 (96.6%). Todas as frequências podem ser consultadas na tabela de frequências da variável, que se encontra inteiramente estruturada na [Tabela 3 – Tabela de Frequências da variável 'Functioning.Day'](#).

Dia funcional (Functioning Day)	Frequência Absoluta	Frequência Relativa (em %)	N_i ^[1]	F_i ^[2] (em %)
Yes	8465	96.2	8465	96.2
No	295	3.4	8760	100
	8760	100		

Tabela 3 - Tabela de Frequências da variável 'Functioning.Day'

Em consonância com as variáveis qualitativas nominais anteriormente retratadas, também indicamos que apenas utilizamos a representação gráfica dos dados através de diagrama circular ou de um gráfico de barras. Estes gráficos encontram-se na [Figura 5 – Gráfico de Barras das frequências absolutas da variável 'Functioning.Day'](#) e na [Figura 6 – Diagrama Circular das frequências absolutas da variável 'Functioning.Day'](#), anexados em baixo.

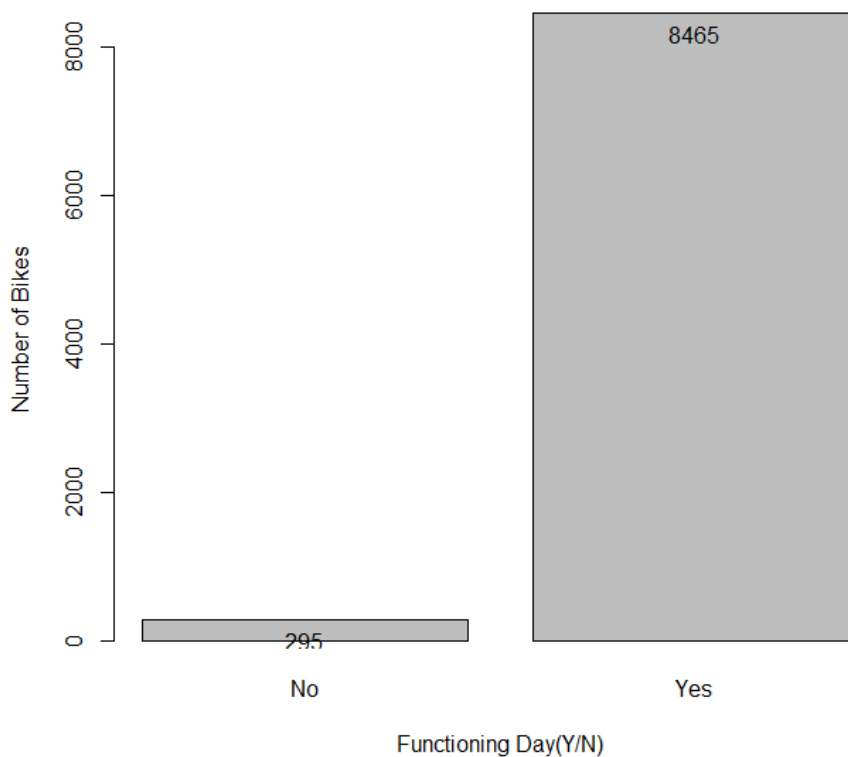


Figura 5 - Gráfico de Barras das frequências absolutas da variável 'Functioning.Day'

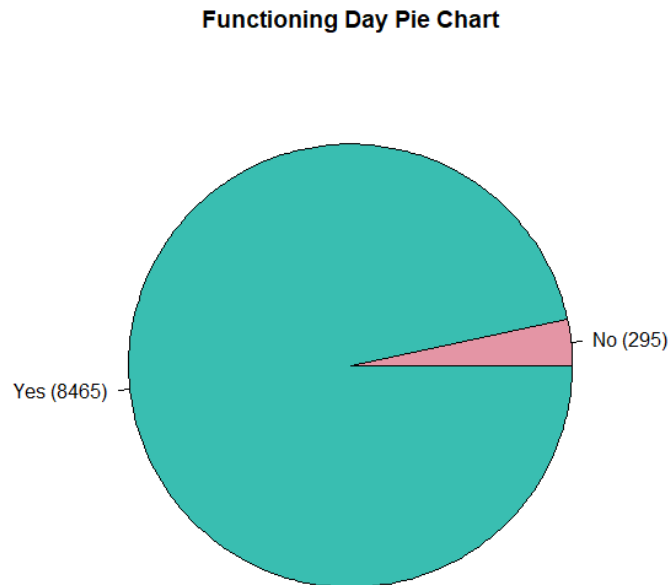


Figura 6 - Diagrama Circular das frequências absolutas da variável 'Functioning.Day'

Variável Qualitativa Ordinal

Date – Data

Relativamente a esta variável não conseguimos apontar qualquer dado estatístico, devido a se encontrar demonstrado como “Ano/Dia/Mês”. Não é possível arquitetar gráficos sobre esta variável, nem determinar a moda, que é a única medida de localização aplicável. Então decidimos optar pela transformação da data em dias julianos, que mudaria a classificação da variável, mas talvez permitisse uma análise mais facilitada desta variável, mas tal não se verificou, devido à composição desta variável no ficheiro de dados fornecido. Encontra-se demasiado repetida e dependente da variável de hora. Não serve como indicativo individual de algo. Para existir uma praticidade, esta variável tem que ser estudada em conjunto com a variável de número de alugueis de bicicleta (Rented Bike Count), para nos retornar indicativos de qual a data em que existiu um número menor de alugueis de bicicletas, ou qual foi a data em que existiu um número maior de alugueis de bicicletas.

Variáveis Quantitativas Discretas

Hour – Hora do dia

Esta variável, diferencia-se das outras variáveis previamente referidas neste relatório, devido a permitir a utilização de todas as medidas de localização e também de ser representada graficamente por um histograma, contrariamente à representação anterior que era o Diagrama Circular e o Gráfico de Barras.

Após separarmos os valores de hora em classes através da Regra de Sturges que se compõe consequentemente das seguintes equações:

$$k = \lceil 1 + \log_2 8760 \rceil = 14$$

Sendo que a nossa amostra ou n , é equivalente a 8760. Posteriormente determinado o k , podemos verificar a amplitude que é calculada da seguinte forma:

$$h = \frac{23 - 0}{14} = 1.64$$

Não é possível a construção gráfica destes dados, devido a serem dados quantitativos discretos, mas agrupados em classes, mas conseguimos representar a tabela de frequências com as suas respetivas classes. Essa tabela encontra-se aqui em baixo, na [*Tabela 4 – Tabela de frequências da variável 'Hour'*](#).

c_i ^[3]	x'_i ^[4]	n_i ^[5]	f_i ^[6] (em %)	N_i ^[1]	F_i ^[2] (em %)
[0,1.643[0.8215	730	8.33	730	8.33
[1.643,3.286[2.4645	730	8.33	1460	16.67
[3.286,4.929[4.1075	365	4.17	1825	20.83
[4.929,6.572[5.7505	730	8.33	2555	29.17
[6.572,8.215[7.3935	730	8.33	3285	37.50
[8.215,9.858[9.0365	365	4.17	3650	41.67
[9.858,11.5[10.6795	730	8.33	4380	50.00
[11.5,13.14[12.3225	730	8.33	5110	58.33
[13.14,14.79[13.9655	365	4.17	5475	62.50
[14.79,16.43[15.6085	730	8.33	6205	70.83
[16.43,18.07[17.2515	730	8.33	6935	79.17
[18.07,19.72[18.8945	365	4.17	7300	83.33
[19.72,21.36[20.5375	730	8.33	8030	91.67
[21.36,23]	22.1805	730	8.33	8760	100
		8760	100		

Tabela 4 - Tabela de Frequências da variável 'Hour'

Após consulta destes dados podemos verificar pela forma bruta que é multimodal, mas existe um pequeno detalhe nesta observação. Todos estes cálculos, podem ser efetuados em variáveis com esta qualificação, desde que se apresentem bem representadas no conjunto de dados que nos é fornecido. Neste caso a variável 'Hour' determina as horas do dia, variável essa, que poderia ser o indicador de muitos casos concretos analiticamente. Poderia ser determinante para saber a hora do dia em que os utilizadores mais efetuam os alugues, ou a hora do dia em que é efetuado menos alugueis, numa certa data. Esta variável, embora tenhamos determinado as classes e o histograma, acaba por se tratar por uma perda de tempo, por estar representada como está, individualmente no conjunto de dados. Será útil posteriormente na análise da variável seguinte, 'Rented Bike Count'.

Rented Bike Count – Contagem de alugueis de bicicletas

Esta contagem de alugueis de bicicleta determina a quantidade de bicicletas que foram alugadas numa determinada hora de um determinado dia. Sendo uma contagem, foi previamente classificada como variável quantitativa discreta. Tal classificação permitia-nos a representação gráfica através de um histograma, caso a variável não estivesse agrupada em classes, mas tal não se verificou, por isso apenas determinamos todas as medidas de localização. Inicialmente, procedemos ao agrupamento dos dados em classes, e para tal determinamos a amplitude através das seguintes equações:

$$k = \lceil 1 + \log_2 8760 \rceil = 14$$

Sendo que a amostra desta variável ou n é igual a 8760. Posteriormente prosseguimos para a equação da amplitude:

$$h = \frac{3556 - 0}{14} = 254$$

E como tal conseguimos chegar à nossa tabela de frequências, que se encontra presente na tabela abaixo, denominada de [*Tabela 5 – Tabela de Frequências da variável 'Rented Bike Count'*](#).

c_i ^[3]	x'_i ^[4]	n_i ^[5]	f_i ^[6] (em %)	N_i ^[1]	F_i ^[2] (em %)
[0,254[127	2817	32.16	2817	32.16
[254, 508[381	1575	17.98	4392	50.14
[508, 762[635	1070	12.21	5462	62.35
[762, 1016[889	933	10.65	6395	73.00
[1016, 1270[1143	792	9.04	7187	82.04
[1270, 1524[1397	476	5.43	7663	87.48
[1524, 1778[1651	338	3.86	8001	91.34
[1778, 2032[1905	310	3.54	8311	94.87
[2032, 2286[2159	220	2.51	8531	97.39
[2286, 2540[2413	143	1.63	8674	99.02
[2540, 2794[2667	35	0.40	8709	99.42
[2794, 3048[2921	21	0.24	8730	99.66
[3048, 3302[3175	23	0.26	8753	99.92
[3302, 3556]	3429	7	0.08	8760	100
		8760	100		

Tabela 5 - Tabela de Frequências da variável 'Rented Bike Count'

Falemos agora dos cálculos das medidas de localização que se podem aplicar a esta variável. Começando pela moda, através de uma função que definimos no R, determinamos que a moda é 0 bicicletas. Pela forma “bruta”, temos que a classe modal é a classe [0,254[, cujo valor modal é o seu representante da classe, ou seja 127. Para nos conformamos com uma resposta concreta, optaríamos pelo método de King, mas como não obtemos classe anterior para obter a frequência relativa anterior, não podemos efetuar os devidos cálculos. Pelo que decidimos optar pelo valor bruto de 127, como a nossa moda. A média foi determinada através do representante de classe e da frequência absoluta, pelo que obtivemos o seguinte:

$$\bar{X} = \frac{6250202,5}{8760} \approx 713.4934$$

Relativamente aos quartis, determinamos da seguinte forma:

$$1^{\text{º}} \text{ Quartil} - Q_{0.25} \approx 197$$

$$2^{\text{º}} \text{ Quartil (Mediana)} - Q_{0.5} \approx 254 + 254 \times \frac{(0.5 - 0.3216)}{0.1798} \approx 506$$

$$3^{\text{º}} \text{ Quartil} - Q_{0.75} \approx 762 + 254 \times \frac{0.75 - 0.73}{0.0904} \approx 818.1947$$

Apesar de não ser das variáveis mais relevantes para um estudo concreto, podemos determinar em conjunto com outras variáveis, algumas estatísticas mais relevantes, como por exemplo, o dia em que existiu mais alugueis de bicicletas foi 2018/06/19 pelas 18 horas, com uma frequência absoluta de 3556 alugueis de bicicletas (podemos indicar também que é o valor máximo). Quanto ao menor valor, existem diversas datas e horas em que existiram 0 alugueis de bicicletas (podemos indicar também que é o nosso valor mínimo).

Variáveis Quantitativas Contínuas

Temperature – Temperatura

Como a sua nomenclatura indica, esta variável mantém o registo da temperatura numa determinada hora de uma determinada data. É uma medição, que é influenciável no aluguer de bicicletas, pois considera-se que quando está uma temperatura agradável, existe uma maior probabilidade de alguém alugar uma bicicleta. Tal pode ser consultado através dos cálculos da moda. Também podemos verificar a partir da [Tabela 6 - Tabela de Frequências da variável 'Temperature'](#), que onde existiram as menores frequências absolutas foram em temperaturas extremamente baixas, ou temperaturas extremamente altas, para o padrão de temperatura da Coreia do Sul. Esta quantificação não é 100% correta, pois não tem em conta o número de bicicletas que foram alugadas concretamente nessa hora e data, apenas tem a quantidade de alugueis que foram efetuados. Para tal, é necessário existir uma separação do número de bicicletas alugadas em cada classe de temperatura. Tal separação pode ser consultada na [Tabela 7 – Tabela de Número de Bicicletas](#) por classe de Temperatura. Nessa tabela, podemos verificar que apesar da frequência absoluta referida na tabela de frequência, como valor modal, ou seja, o maior número de alugueis de bicicletas foi efetuado numa temperatura entre 18.97°C e 23.06°C, o maior número de bicicletas alugada, foi durante temperaturas entre 23.06°C e 27.14°C com um número total de 1168973 bicicletas.

c_i ^[3]	x'_i ^[4]	n_i ^[5]	f_i ^[6] (em %)	N_i ^[1]	F_i ^[2] (em %)
[-17.8,-13.71)	-15.755	67	0.76	67	0.76
[-13.71,-9.629)	-11.6693	180	2.05	247	2.82
[-9.629,-5.543)	-7.5836	377	4.30	624	7.12
[-5.543,-1.457)	-3.4978	590	6.74	1214	13.86
[-1.457,2.629)	0.5878	797	9.10	2011	22.96
[2.629,6.714)	4.6735	883	10.08	2894	33.04
[6.714,10.8)	8.7593	882	10.07	3776	43.11
[10.8,14.89)	12.8450	829	9.46	4605	52.57
[14.89,18.97)	16.9307	943	10.76	5548	63.33
[18.97,23.06)	21.0164	1174	13.40	6722	76.74
[23.06,27.14)	25.1021	964	11.00	7686	87.74
[27.14,31.23)	29.1878	694	7.92	8380	95.66
[31.23,35.31)	33.2736	306	3.49	8686	99.16
[35.31,39.4)	37.3593	74	0.84	8760	100
		8760	100		

Tabela 6 - Tabela de Frequências da variável 'Temperature'

c_i ^[3]	Número de Bicicletas
[-17.8,-13.71)	7052
[-13.71,-9.629)	28076
[-9.629,-5.543)	68263
[-5.543,-1.457)	124907
[-1.457,2.629)	219635
[2.629,6.714)	360278
[6.714,10.8)	486854
[10.8,14.89)	581042
[14.89,18.97)	754007
[18.97,23.06)	1133761
[23.06,27.14)	1168973
[27.14,31.23)	822375
[31.23,35.31)	341939
[35.31,39.4)	64571

Tabela 7- Tabela de Número de Bicicletas por classe de Temperatura

A demonstração gráfica destes dados, é efetuada através de um histograma, que fora anteriormente gerado com recurso a comandos para o efeito. Este histograma, encontra-se presente na [Figura 7 – Histograma das classes da variável 'Temperature'](#), anexado em baixo.

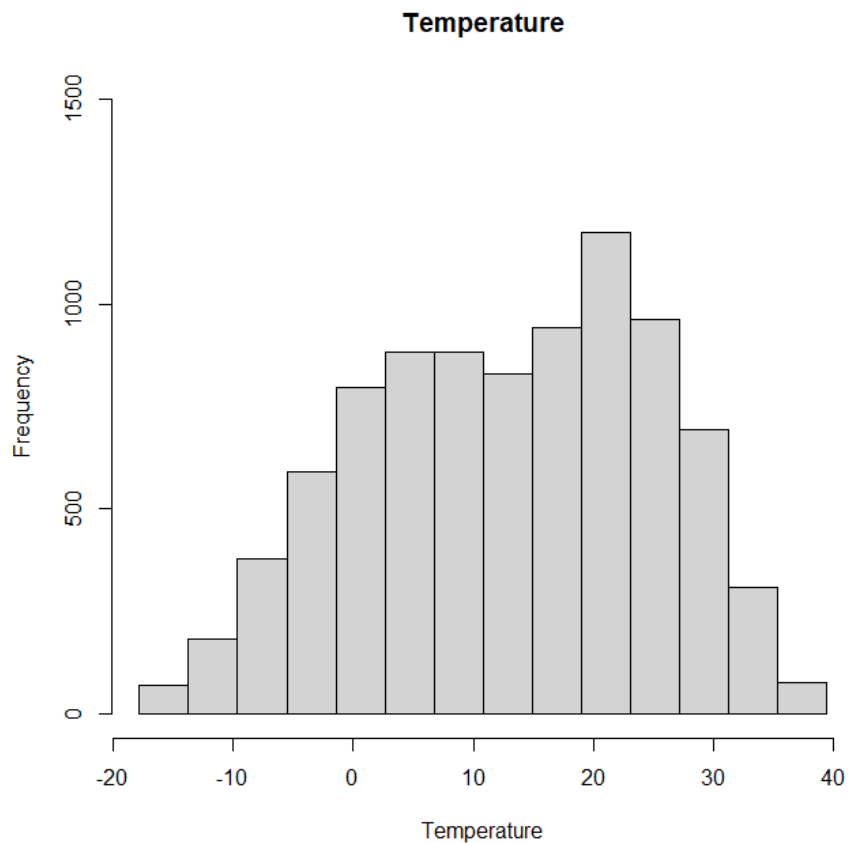


Figura 7 - Histograma das classes da variável 'Temperature'

Relativamente aos cálculos estatísticos efetuados, conseguimos obter todas as medidas de localização e todas as medidas de dispersão, bem como o achatamento e a curtose. Os resultados obtidos encontram-se demonstrados sucintamente em baixo.

Medidas de Localização

Moda – Relativamente à moda, temos diversas maneiras de a obter. Inicialmente podemos indicar que a Classe Modal é [18.97,23.06[, pois tem a maior frequência absoluta, 1174 alugueis. Calculando de forma bruta e com recurso às classes, temos que a moda é igual ao representante da classe da Classe Modal, ou seja, valor modal é 21.0164. Também com recurso ao *RStudio* e com uma função previamente criada para a moda, obtivemos a moda, sem recurso ao agrupamento de classes, e obtivemos uma variável bimodal, com os valores 20.5 e 19.1. Ambos estes valores se encontram dentro da Classe Modal e podem ser indicativos de que a moda se encontra mesmo nessa classe. Mas sendo os dados agrupados em classes, esses valores apenas são meros indicativos, e não tem peso na moda concreta desta variável. Para calcular a moda, também podemos utilizar o método de King, que se compõe da seguinte forma:

$$moda \approx 18.97 + (23.06 - 18.97) \times \frac{0.134}{0.0946 + 0.134} \approx 21.3675$$

Média – Relativamente à média, optamos por utilizar um comando no *RStudio* que nos fizesse imediatamente a média. O resultado obtido foi 12.8292. Tal valor poderia ser obtido através da equação $\frac{\sum_{i=1}^k n_i x'_i}{n}$, através das frequências absolutas. Mas devido a existirem um número considerável de fatores para chegar a esse resultado, tornar-se-ia demasiado extenso para colocar neste relatório sucinto.

Quantis – Quanto aos quantis, obtivemos os resultados através da execução de comandos que têm o propósito de calcular os quantis. Como, queremos chegar a um diagrama de quartis, decidimos calcular apenas os quartis da variável, como tal, o primeiro quartil é igual a 3.5, o segundo quartil (ou mediana) é igual a 13.7 e o terceiro quartil é igual a 22.5. Também poderíamos analisar a tabela de frequências e determinar de forma bruta o valor dos quantis,

sendo que desta forma, o primeiro quartil é 4.6735, o segundo quartil (ou mediana) é igual a 12.8450, e o terceiro quartil é igual a 21.0164. [\(verificar nota que indica como chegamos a estes valores\)](#)

Mas para ser mais concreto, calculamos os valores aproximados através das seguintes equações:

$$1^{\text{o}} \text{ Quartil} - Q_{0.25} \approx 2.629 + 4.0858 \times \frac{(0.25 - 0.2296)}{0.1008} \approx 3.456$$

$$2^{\text{o}} \text{ Quartil (Mediana)} - Q_{0.5} \approx 10.8 + 4.0858 \times \frac{(0.5 - 0.4311)}{0.0946} \approx 13.776$$

$$3^{\text{o}} \text{ Quartil} - Q_{0.75} \approx 18.97 + 4.0858 \times \frac{0.75 - 0.6333}{0.1340} \approx 22.528$$

Posteriormente, fora gerado o diagrama de extremos e quartis, que se encontra disposto na [Figura 8 – Diagrama de extremos e quartis da variável ‘Temperature’](#).

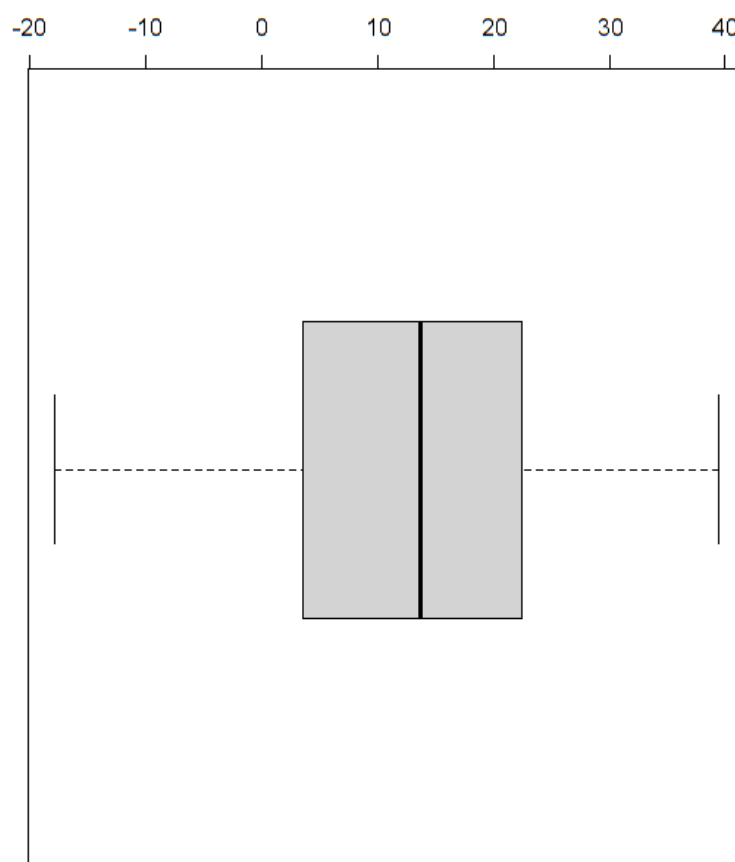


Figura 8 - Diagrama de extremos e quartis da variável ‘Temperature’

Medidas de Dispersão

Amplitude Total – É determinado através do valor mínimo e máximo da variável, sendo que neste caso ***amplitude*** = $39.4 - (-17.8) = 57.2$. Este resultado não é um bom indicativo da amplitude, devido ao $\min(x_i)$ ser um valor negativo.

Amplitude Interquartil – Determinamos o valor com recurso à execução do comando específico que efetua esta medida. O resultado obtido foi 19, com base nos valores $Q_{0.25} = 3.5$ e $Q_{0.75} = 22.5$. Para termos um resultado mais concreto, podemos efetuar a amplitude interquartil com os valores aproximados que foram determinados com as equações anteriormente referidas:

$$AIQ = Q_{0.75} - Q_{0.25} = 22.528 - 3.456 = 19.072$$

Variância – Novamente, com recurso a um comando específico do R, que me calcula imediatamente o resultado, foi obtido que $s^2 \approx 142.6788$. Tal comando, é baseado na seguinte equação:

$$s^2 \approx \frac{\sum_{i=1}^k n_i (x'_i - \bar{x})^2}{n - 1}$$

Desvio Padrão – Facilmente obtido, sendo que já obtemos a variância no ponto anterior, mas decidimos optar pelo comando do R, que me efetua imediatamente o cálculo, sendo que

$$s \approx 11.94483$$

Este resultado está baseado na equação:

$$s \approx \sqrt{s^2}$$

Coefficiente de variação – Determinado com recurso ao desvio padrão e à média, podemos determinar facilmente esta medida de dispersão. Com recurso à equação:

$$CV = \frac{s}{\bar{x}} * 100$$

E com a passagem desta equação para o R, utilizando as variáveis anteriormente calculadas com os respectivos comandos de R (média e desvio padrão), obtivemos que:

$$CV = \frac{s}{\bar{x}} * 100 = 92.71829\%$$

Caracterização da Distribuição de Frequências

Finalizando a análise desta variável apenas necessitamos de demonstrar a caracterização da distribuição de frequências, sendo que apenas existem as medidas de curtose e as medidas de assimetria.

Medidas de Assimetria – Embora pudéssemos verificar através do histograma e do diagrama de extremos e quartis que existe uma assimetria negativa, ou até através da comparação da moda, com a média e com a mediana (média < mediana < moda), optamos por complementar com o cálculo do b_1 , com recurso a um comando R para o resultado. Sendo que o resultado da assimetria deu:

$$b_1 \approx \frac{1}{n} \sum_{i=1}^n \left(\frac{(x_i - \bar{x})}{s} \right)^3 \approx -0.1982916$$

Medidas de Curtose – O achatamento ou curtose, fora determinado através de um comando R para o efeito. Sendo que o resultado deu:

$$b_2 \approx \frac{1}{n} \sum_{i=1}^n \left(\frac{(x_i - \bar{x})}{s} \right)^4 - 3 \approx 2.162007$$

Valor este que indica que existe uma forte concentração de valores em torno da média, existindo uma variação pouco elevada, ou seja, que o grau de curtose é do tipo curva leptocúrtica, alongada, pois $b_2 > 0$.

Humidity – Humidade

Esta variável possui valores relacionados com a humidade que estaria no aluguer de bicicletas de uma determinada hora numa determinada data. Poderá existir uma relação com o aluguer de bicicletas com o nível de humidade, mas não é tão influenciável como a temperatura. A humidade poderá estar relacionada com a temperatura, pois quando existe uma temperatura alta com uma humidade baixa, poderá ser uma situação desconfortável para se andar de bicicleta, ou alugar uma bicicleta. A análise desta variável teve apenas o intuito de conseguirmos demonstrar diversas estatísticas sobre a humidade em Seoul, com recurso ao aluguer de bicicletas. A tabela de frequências desta variável encontra-se presente na [Tabela 8 – Tabela de Frequências da variável 'Humidity'](#).

c_i ^[3]	x'_i ^[4]	n_i ^[5]	f_i ^[6] (em %)	N_i ^[1]	F_i ^[2] (em %)
[0,7[3.5	17	0.19	17	0.19
[7,14[10.5	6	0.07	23	0.26
[14,21[17.5	119	1.36	142	1.62
[21,28[24.5	346	3.95	488	5.57
[28,35[31.5	728	8.31	1216	13.88
[35,42[38.5	854	9.75	2070	23.63
[42,49[45.5	987	11.27	3057	34.90
[49,56[52.5	1053	12.02	4110	46.92
[56,63[59.5	987	11.27	5097	58.18
[63,70[66.5	925	10.56	6022	68.74
[70,77[73.5	811	9.26	6833	78.00
[77,84[80.5	774	8.84	7607	86.84
[84,91[87.5	594	6.78	8201	93.62
[91,98]	94.5	559	6.38	8710	100
		8760	100		

Tabela 8 - Tabela de Frequências da variável 'Humidity'

A demonstração gráfica destes dados, é efetuada através de um histograma, que fora anteriormente gerado com recurso a comandos para o efeito. Este histograma, encontra-se presente na [Figura 9 – Histograma das classes da variável 'Humidity'](#), anexado em baixo.

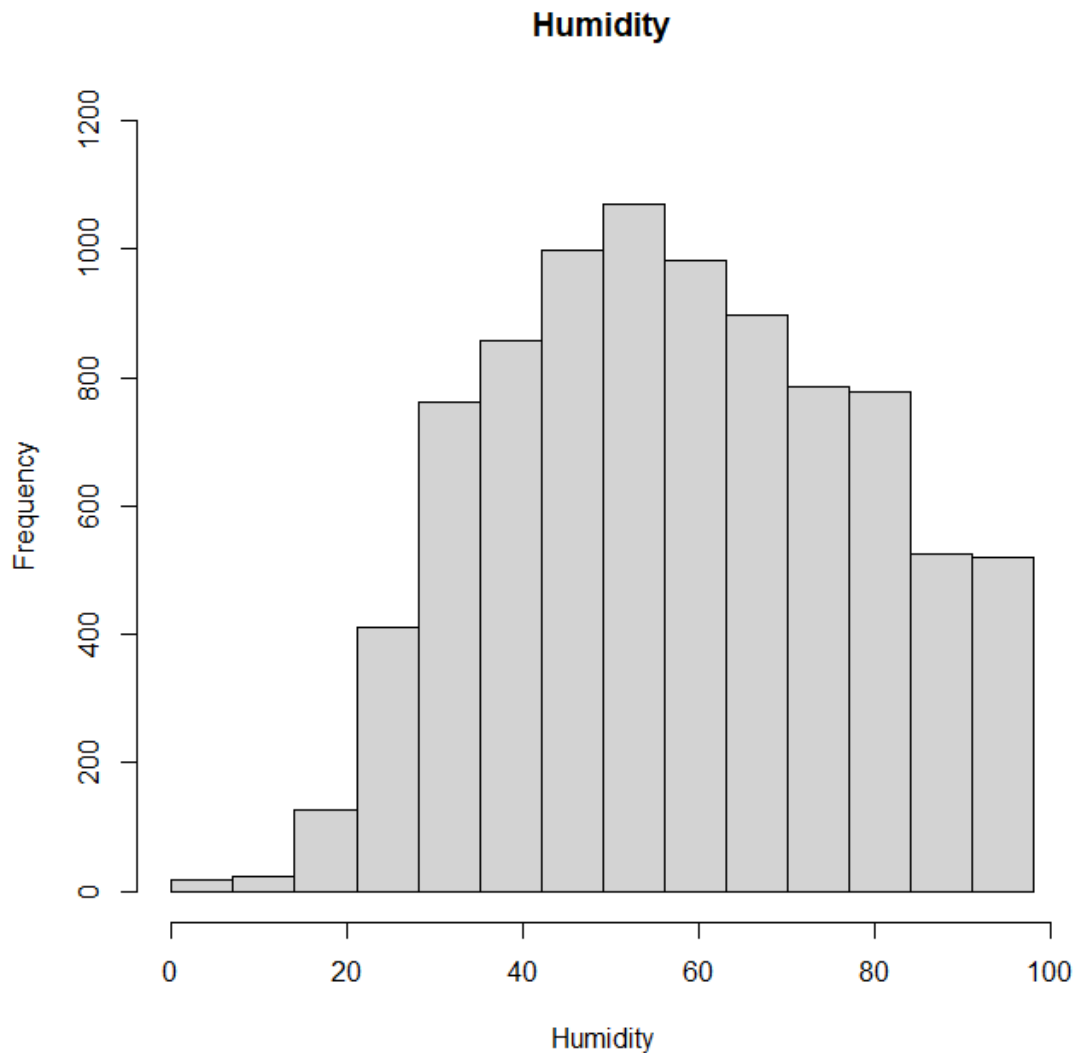


Figura 9 - Histograma das classes da variável 'Humidity'

Relativamente aos cálculos estatísticos efetuados, conseguimos obter todas as medidas de localização e todas as medidas de dispersão, bem como o achatamento e a curtose. Os resultados obtidos encontram-se demonstrados sucintamente em baixo.

Medidas de Localização

Moda – Relativamente à moda, temos diversas maneiras de a obter. Inicialmente podemos indicar que a Classe Modal é [49,56[, pois tem a maior frequência absoluta, 1053 alugueis. Calculando de forma bruta e com recurso às classes, temos que a moda é igual ao representante da classe da Classe Modal, ou seja, valor modal é 52.5. Também com recurso ao *RStudio* e com uma função previamente criada para a moda, obtivemos a moda, sem recurso ao agrupamento de classes, e obtivemos uma variável bimodal, com os valores 53 e 97. Mas sendo os dados agrupados em classes, esses valores apenas são meros indicativos, e não tem peso na moda concreta desta variável. Para calcular a moda, também podemos utilizar o método de King, que se compôs da seguinte forma:

$$moda \approx 49 + (56 - 49) \times \frac{0.1202}{0.1127 + 0.1202} \approx 52.6127$$

Média – Relativamente à média, optamos por utilizar um comando no *RStudio* que nos fizesse imediatamente a média. O resultado obtido foi 58.22626. Tal valor poderia ser obtido através da equação $\frac{\sum_{i=1}^k n_i x'_i}{n}$, através das frequências absolutas. Mas devido a existirem um número considerável de fatores para chegar a esse resultado, tornar-se-ia demasiado extenso para colocar neste relatório sucinto.

Quantis – Quantos aos quantis, obtivemos os resultados através da execução de comandos que têm o propósito de calcular os quantis. Como, queremos chegar a um diagrama de quantis, decidimos calcular apenas os quartis da variável, como tal, o primeiro quartil é igual a 42, o segundo quartil (ou mediana) é igual a 57 e o terceiro quartil é igual a 74. Também poderíamos analisar a tabela de frequências e determinar de forma bruta o valor dos quantis, sendo que desta forma, o primeiro quartil é 45.5, o segundo quartil (ou mediana) é igual a 59.5, e o terceiro quartil é igual a 73.5. [\(verificar nota que indica como chegamos a estes valores\)](#)

Mas para ser mais concreto, calculamos os valores aproximados através das seguintes equações:

$$1^{\text{o}} \text{ Quartil} - Q_{0.25} \approx 42 + 7 \times \frac{(0.25 - 0.2363)}{0.1127} \approx 42.8509$$

$$2^{\text{o}} \text{ Quartil (Mediana)} - Q_{0.5} \approx 56 + 7 \times \frac{(0.5 - 0.4692)}{0.1127} \approx 57.9130$$

$$3^{\text{o}} \text{ Quartil} - Q_{0.75} \approx 70 + 7 \times \frac{(0.75 - 0.6874)}{0.0926} \approx 74.7322$$

Posteriormente, fora gerado o diagrama de extremos e quartis, que se encontra disposto na [Figura 10 – Diagrama de extremos e quartis da variável 'Humidity'](#).

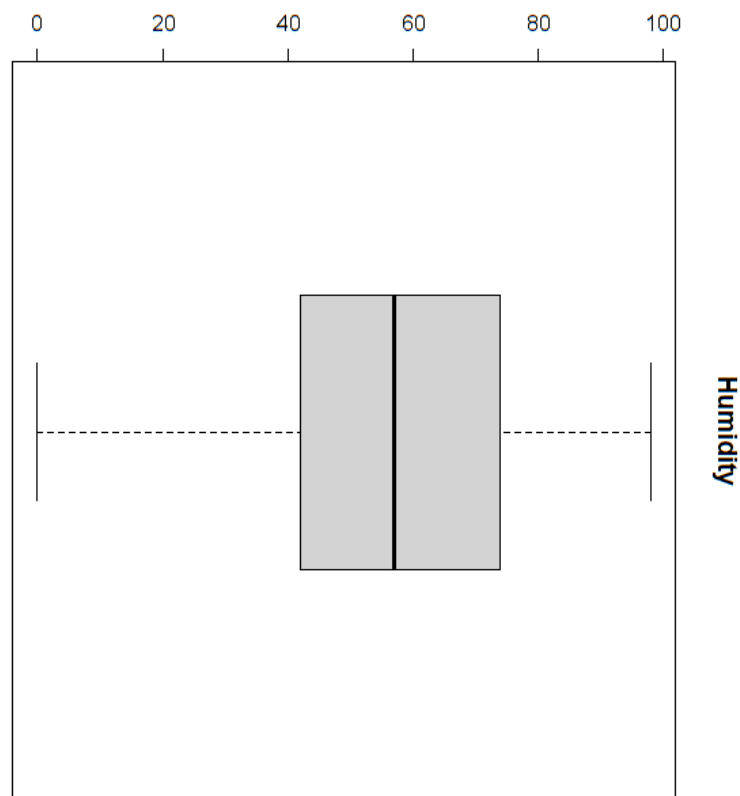


Figura 10 - Diagrama de extremos e quartis da variável 'Humidity'

Medidas de Dispersão

Amplitude Total – É determinado através do valor mínimo e máximo da variável, sendo que neste caso ***amplitude*** = $98 - 0 = 98$.

Amplitude Interquartil – Determinamos o valor com recurso à execução do comando específico que efetua esta medida. O resultado obtido foi 32, com base nos valores $Q_{0.25} = 42$ e $Q_{0.75} = 74$. Para termos um resultado mais concreto, podemos efetuar a amplitude interquartil com os valores aproximados que foram determinados com as equações anteriormente referidas:

$$AIQ = Q_{0.75} - Q_{0.25} = 74.7322 - 42.8509 = 31.8813$$

Variância – Novamente, com recurso a um comando específico do R, que me calcula imediatamente o resultado, foi obtido que $s^2 \approx 414.6279$. Tal comando, é baseado na seguinte equação:

$$s^2 \approx \frac{\sum_{i=1}^k n_i (x'_i - \bar{x})^2}{n - 1}$$

Desvio Padrão – Facilmente obtido, sendo que já obtemos a variância no ponto anterior, mas decidimos optar pelo comando do R, que me efetua imediatamente o cálculo, sendo que

$$s \approx 20.36241$$

Este resultado está baseado na equação:

$$s \approx \sqrt{s^2}$$

Coefficiente de variação – Determinado com recurso ao desvio padrão e à média, podemos determinar facilmente esta medida de dispersão. Com recurso à equação:

$$CV = \frac{s}{\bar{x}} * 100$$

E com a passagem desta equação para o R, utilizando as variáveis anteriormente calculadas com os respectivos comandos de R (média e desvio padrão), obtivemos que:

$$CV = \frac{s}{\bar{x}} * 100 = 34.97119\%$$

Caracterização da Distribuição de Frequências

Finalizando a análise desta variável apenas necessitamos de demonstrar a caracterização da distribuição de frequências, sendo que apenas existem as medidas de curtose e as medidas de assimetria.

Medidas de Assimetria – Embora pudéssemos verificar através do histograma e do diagrama de extremos e quartis que existe uma assimetria positiva, ou até através da comparação da moda, com a média e com a mediana (média > mediana > moda), optamos por complementar com o cálculo do b_1 , com recurso a um comando R para o resultado. Sendo que o resultado da assimetria deu:

$$b_1 \approx \frac{1}{n} \sum_{i=1}^n \left(\frac{(x_i - \bar{x})}{s} \right)^3 \approx 0.05956877$$

Medidas de Curtose – O achatamento ou curtose, fora determinado através de um comando R para o efeito. Sendo que o resultado deu:

$$b_2 \approx \frac{1}{n} \sum_{i=1}^n \left(\frac{(x_i - \bar{x})}{s} \right)^4 - 3 \approx 2.196215$$

Valor este que indica que existe uma forte concentração de valores em torno da média, existindo uma variação pouco elevada, ou seja, que o grau de curtose é do tipo curva leptocúrtica, alongada, pois $b_2 > 0$.

Wind Speed (ms) – Velocidade do vento

Esta variável possui valores relacionados com a velocidade do vento que estaria no aluguer de bicicletas de uma determinada hora numa determinada data. Poderá existir uma relação com o aluguer de bicicletas com a temperatura, mas não é tão influenciável. A análise desta variável teve apenas o intuito de conseguirmos demonstrar diversas estatísticas sobre a velocidade do vento em Seoul, com recurso ao aluguer de bicicletas. A tabela de frequências desta variável encontra-se presente na [Tabela 9 – Tabela de Frequências da variável 'Wind Speed'](#).

c_i ^[3]	x'_i ^[4]	n_i ^[5]	f_i ^[6] (em %)	N_i ^[1]	F_i ^[2] (em %)
[0,0.529[0.2645	811	9.26	811	9.26
[0.529,1.06[0.7931	1795	20.49	2606	29.75
[1.06,1.59[1.3216	1865	21.29	4471	51.04
[1.59,2.11[1.8502	1675	19.12	6146	70.16
[2.11,2.64[2.3788	997	11.38	7143	81.54
[2.64,3.17[2.9073	702	8.01	7845	89.55
[3.17,3.7[3.4359	436	4.98	8281	94.53
[3.7,4.23[3.9645	285	3.25	8566	97.79
[4.23,4.76[4.4931	119	1.36	8685	99.14
[4.76,5.29[5.0216	49	0.56	8734	99.70
[5.29,5.81[5.550	18	0.21	8752	99.91
[5.81,6.34[6.0788	3	0.03	8755	99.94
[6.34,6.87[6.6073	1	0.01	8756	99.95
[6.87,7.4[7.1359	4	0.04	8760	100
		8760	100		

Tabela 9 - Tabela de Frequências da variável 'Wind Speed'

A demonstração gráfica destes dados, é efetuada através de um histograma, que fora anteriormente gerado com recurso a comandos para o efeito. Este histograma, encontra-se presente na [Figura 11 – Histograma das classes da variável 'Humidity'](#), anexado em baixo.

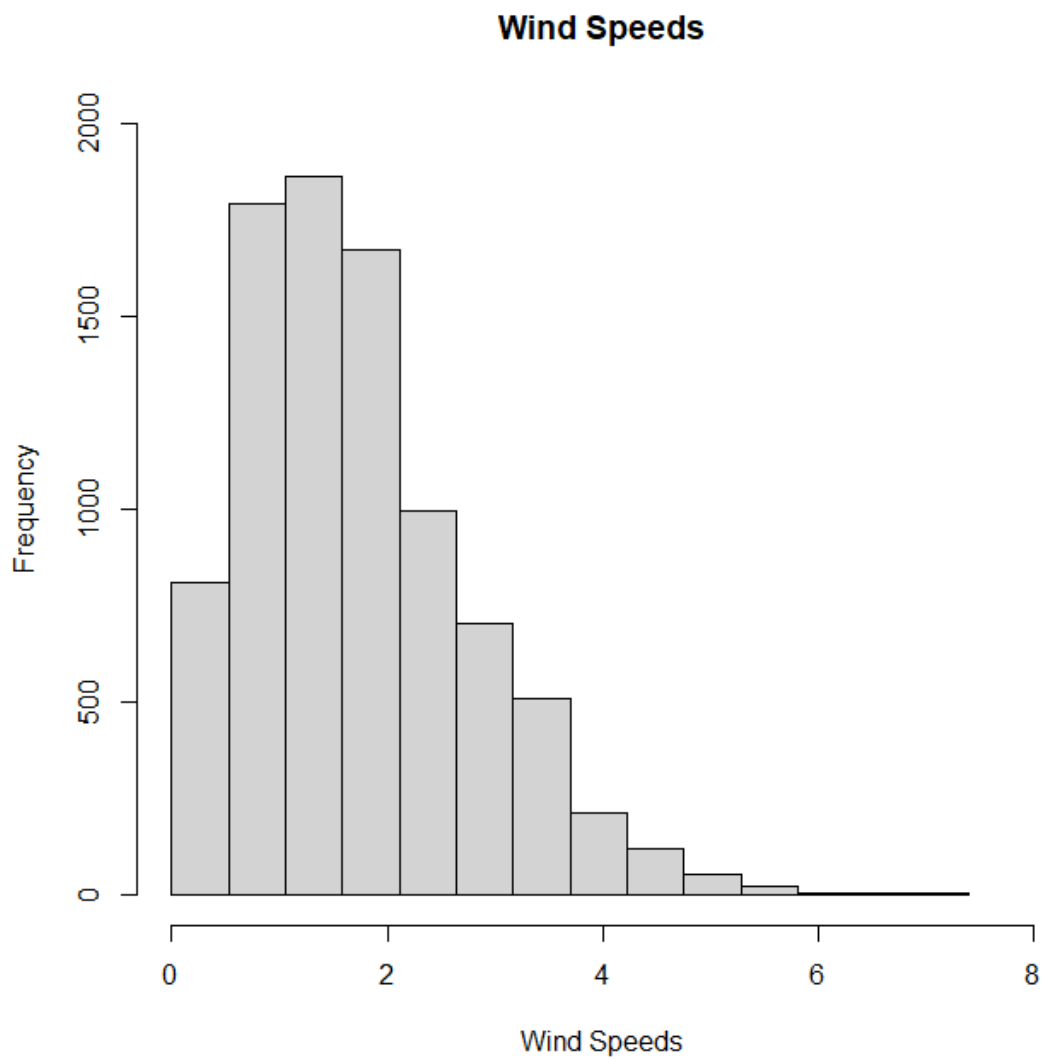


Figura 11 - Histograma das classes da variável 'Wind Speed'

Relativamente aos cálculos estatísticos efetuados, conseguimos obter todas as medidas de localização e todas as medidas de dispersão, bem como o achatamento e a curtose. Os resultados obtidos encontram-se demonstrados sucintamente em baixo.

Medidas de Localização

Moda – Relativamente à moda, temos diversas maneiras de a obter. Inicialmente podemos indicar que a Classe Modal é [1.06,1.59[, pois tem a maior frequência absoluta, 1865 alugueis. Calculando de forma bruta e com recurso às classes, temos que a moda é igual ao representante da classe da Classe Modal, ou seja, valor modal é 1.3216. Também com recurso ao *RStudio* e com uma função previamente criada para a moda, obtivemos a moda, sem recurso ao agrupamento de classes, e obtivemos o valor 1.1. Mas sendo os dados agrupados em classes, esses valores apenas são meros indicativos, e não tem peso na moda concreta desta variável. Para calcular a moda, também podemos utilizar o método de King, que se compõe da seguinte forma:

$$moda \approx 1.06 + (1.59 - 1.06) \times \frac{0.1912}{0.2049 + 0.1912} \approx 1.315834$$

Média – Relativamente à média, optamos por utilizar um comando no *RStudio* que nos fizesse imediatamente a média. O resultado obtido foi 1.724909. Tal valor poderia ser obtido através da equação $\frac{\sum_{i=1}^k n_i x'_i}{n}$, através das frequências absolutas. Mas devido a existirem um número considerável de fatores para chegar a esse resultado, tornar-se-ia demasiado extenso para colocar neste relatório sucinto.

Quantis – Quantos aos quantis, obtivemos os resultados através da execução de comandos que têm o propósito de calcular os quantis. Como, queremos chegar a um diagrama de quantis, decidimos calcular apenas os quartis da variável, como tal, o primeiro quartil é igual a 0.9, o segundo quartil (ou mediana) é igual a 1.5 e o terceiro quartil é igual a 2.3. Também poderíamos analisar a tabela de frequências e determinar de forma bruta o valor dos quantis, sendo que desta forma, o primeiro quartil é 0.7931, o segundo quartil (ou mediana) é igual a 1.3216, e o terceiro quartil é igual a 2.3788. [\(verificar nota que indica como chegamos a estes valores\)](#)

Mas para ser mais concreto, calculamos os valores aproximados através das seguintes equações:

$$1^{\text{o}} \text{ Quartil} - Q_{0.25} \approx 0.529 + (1.06 - 0.529) \times \frac{(0.25 - 0.0926)}{0.2049} \approx 0.9369034$$

$$2^{\text{o}} \text{ Quartil (Mediana)} - Q_{0.5} \approx 1.06 + (1.59 - 1.06) \times \frac{(0.5 - 0.2975)}{0.2129} \approx 1.56411$$

$$3^{\text{o}} \text{ Quartil} - Q_{0.75} \approx 2.11 + (2.64 - 2.11) \times \frac{(0.75 - 0.716)}{0.1138} \approx 2.268348$$

Posteriormente, fora gerado o diagrama de extremos e quartis, que se encontra disposto na [Figura 12 – Diagrama de extremos e quartis da variável 'Wind Speed'](#).

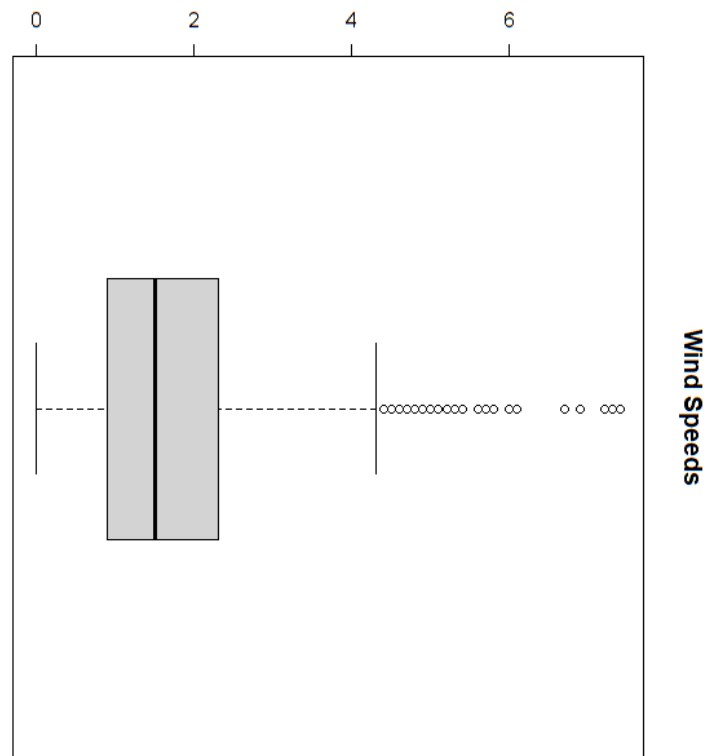


Figura 12 - Diagrama de extremos e quartis da variável 'Wind Speed'

Medidas de Dispersão

Amplitude Total – É determinado através do valor mínimo e máximo da variável, sendo que neste caso ***amplitude*** = $7.4 - 0 = 7.4$.

Amplitude Interquartil – Determinamos o valor com recurso à execução do comando específico que efetua esta medida. O resultado obtido foi 1.4, com base nos valores $Q_{0.25} = 0.9$ e $Q_{0.75} = 2.3$. Para termos um resultado mais concreto, podemos efetuar a amplitude interquartil com os valores aproximados que foram determinados com as equações anteriormente referidas:

$$AIQ = Q_{0.75} - Q_{0.25} = 2.268348 - 0.9369034 = 1.331445$$

Variância – Novamente, com recurso a um comando específico do R, que me calcula imediatamente o resultado, foi obtido que $s^2 \approx 1.073918$. Tal comando, é baseado na seguinte equação:

$$s^2 \approx \frac{\sum_{i=1}^k n_i (x'_i - \bar{x})^2}{n - 1}$$

Desvio Padrão – Facilmente obtido, sendo que já obtemos a variância no ponto anterior, mas decidimos optar pelo comando do R, que me efetua imediatamente o cálculo, sendo que

$$s \approx 1.0363$$

Este resultado está baseado na equação:

$$s \approx \sqrt{s^2}$$

Coefficiente de variação – Determinado com recurso ao desvio padrão e à média, podemos determinar facilmente esta medida de dispersão. Com recurso à equação:

$$CV = \frac{s}{\bar{x}} * 100$$

E com a passagem desta equação para o R, utilizando as variáveis anteriormente calculadas com os respectivos comandos de R (média e desvio padrão), obtivemos que:

$$CV = \frac{s}{\bar{x}} * 100 = 60.07854\%$$

Caracterização da Distribuição de Frequências

Finalizando a análise desta variável apenas necessitamos de demonstrar a caracterização da distribuição de frequências, sendo que apenas existem as medidas de curtose e as medidas de assimetria.

Medidas de Assimetria – Embora pudéssemos verificar através do histograma e do diagrama de extremos e quartis que existe uma assimetria positiva, ou até através da comparação da moda, com a média e com a mediana (média > mediana > moda), optamos por complementar com o cálculo do b_1 , com recurso a um comando R para o resultado. Sendo que o resultado da assimetria deu:

$$b_1 \approx \frac{1}{n} \sum_{i=1}^n \left(\frac{(x_i - \bar{x})}{s} \right)^3 \approx 0.8908022$$

Medidas de Curtose – O achatamento ou curtose, fora determinado através de um comando R para o efeito. Sendo que o resultado deu:

$$b_2 \approx \frac{1}{n} \sum_{i=1}^n \left(\frac{(x_i - \bar{x})}{s} \right)^4 - 3 \approx 3.72608$$

Valor este que indica que existe uma forte concentração de valores em torno da média, existindo uma variação pouco elevada, ou seja, que o grau de curtose é do tipo curva leptocúrtica, alongada, pois $b_2 > 0$.

Visibility (10m) – Visibilidade

Esta variável possui valores relacionados com a visibilidade no espaço de 10m, que estaria no aluguer de bicicletas de uma determinada hora numa determinada data. Conseguimos associar um número inferior de alugueis de bicicletas, no caso de a visibilidade ser baixa, logicamente, e também analisando as frequências, nas tabelas de frequências presente na [Tabela 10 – Tabela de Frequências da variável ‘Visibility’](#).

c_i ^[3]	x'_i ^[4]	n_i ^[5]	f_i ^[6] (em %)	N_i ^[1]	F_i ^[2] (em %)
[27,167.9[97.45	81	0.92	81	0.92
[167.9,308.9[238.4	382	4.36	463	5.29
[308.9,449.8[379.35	446	5.09	909	10.38
[449.8,590.7[520.25	398	4.54	1307	14.92
[590.7,731.6[661.15	363	4.14	1670	19.06
[731.6,872.6[802.05	347	3.96	2017	23.03
[872.6,1013[942.97	402	4.59	2419	27.61
[1013,1154[1083.9	385	4.39	2804	32.01
[1154,1295[1224.83	372	4.25	3176	36.26
[1295,1436[1365.76	418	4.77	3594	41.03
[1436,1577[1506.69	414	4.73	4008	45.75
[1577,1718[1647.62	443	5.06	4451	50.81
[1718,1859[1788.55	632	7.21	5083	58.03
[1859,2000]	1929.47	3677	41.97	8760	100
		8760	100		

Tabela 10 - Tabela de Frequências da variável ‘Visibility’

A demonstração gráfica destes dados, é efetuada através de um histograma, que fora anteriormente gerado com recurso a comandos para o efeito. Este histograma, encontra-se presente na [Figura 13 – Histograma das classes da variável ‘Visibility’](#), anexado em baixo.

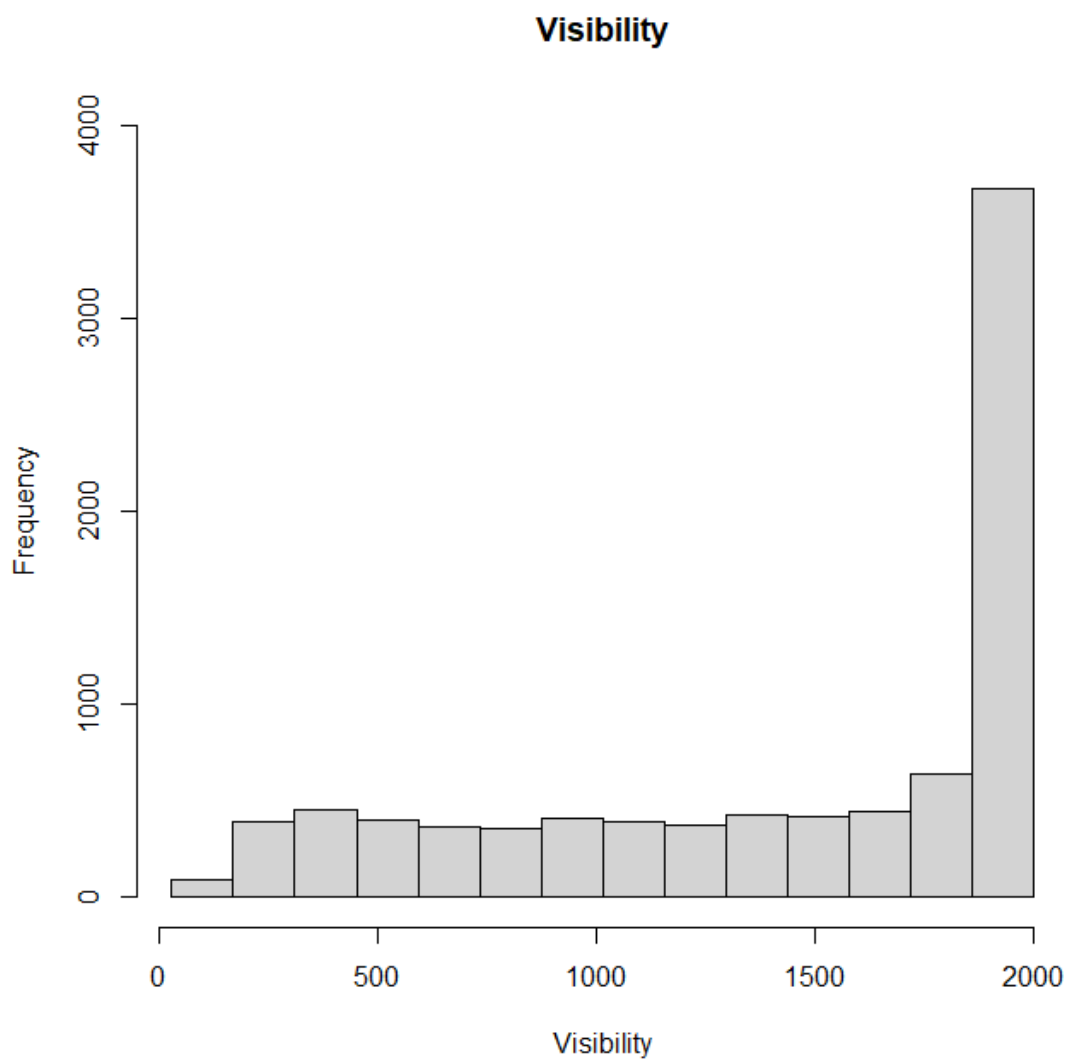


Figura 13 - Histograma das classes da variável 'Visibility'

Relativamente aos cálculos estatísticos efetuados, conseguimos obter todas as medidas de localização e todas as medidas de dispersão, bem como o achatamento e a curtose. Os resultados obtidos encontram-se demonstrados sucintamente em baixo.

Medidas de Localização

Moda – Relativamente à moda, temos diversas maneiras de a obter. Inicialmente podemos indicar que a Classe Modal é [1859,2000[, pois tem a maior frequência absoluta, 3677 alugueis. Calculando de forma bruta e com recurso às classes, temos que a moda é igual ao representante da classe da Classe Modal, ou seja, valor modal é 1929.47. Também com recurso ao *RStudio* e com uma função previamente criada para a moda, obtivemos a moda, sem recurso ao agrupamento de classes, e obtivemos o valor 2000. Mas sendo os dados agrupados em classes, esses valores apenas são meros indicativos, e não tem peso na moda concreta desta variável. Para calcular a moda, também podemos utilizar o método de King, que se compõe da seguinte forma:

$$moda \approx 1859 + (2000 - 1859) \times \frac{0}{0.0721 + 0} = 1859$$

Média – Relativamente à média, optamos por utilizar um comando no *RStudio* que nos fizesse imediatamente a média. O resultado obtido foi 1436.826. Tal valor poderia ser obtido através da equação $\frac{\sum_{i=1}^k n_i x'_i}{n}$, através das frequências absolutas. Mas devido a existirem um número considerável de fatores para chegar a esse resultado, tornar-se-ia demasiado extenso para colocar neste relatório sucinto.

Quantis – Quanto aos quantis, obtivemos os resultados através da execução de comandos que têm o propósito de calcular os quantis. Como, queremos chegar a um diagrama de quantis, decidimos calcular apenas os quartis da variável, como tal, o primeiro quartil é igual a 940, o segundo quartil (ou mediana) é igual a 1698 e o terceiro quartil é igual a 2000. Também poderíamos analisar a tabela de frequências e determinar de forma bruta o valor dos quantis, sendo que desta forma, o primeiro quartil é 942.97, o segundo quartil (ou mediana) é igual a 1647.62, e o terceiro quartil é igual a 1929.47. [\(verificar nota que indica como chegamos a estes valores\)](#)

Mas para ser mais concreto, calculamos os valores aproximados através das seguintes equações:

$$1^{\circ} \text{ Quartil} - Q_{0.25} \approx 872.6 + (1013 - 872.6) \times \frac{(0.25 - 0.2303)}{0.0414} \approx 939.4087$$

$$2^{\circ} \text{ Quartil (Mediana)} - Q_{0.5} \approx 1577 + (1718 - 1577) \times \frac{(0.5 - 0.4575)}{0.0506} \approx 1695.429$$

$$3^{\circ} \text{ Quartil} - Q_{0.75} \approx 1859 + (2000 - 1859) \times \frac{(0.75 - 0.5803)}{0.4197} \approx 1916.011$$

Posteriormente, fora gerado o diagrama de extremos e quartis, que se encontra disposto na [Figura 14 – Diagrama de extremos e quartis da variável 'Visibility'](#).

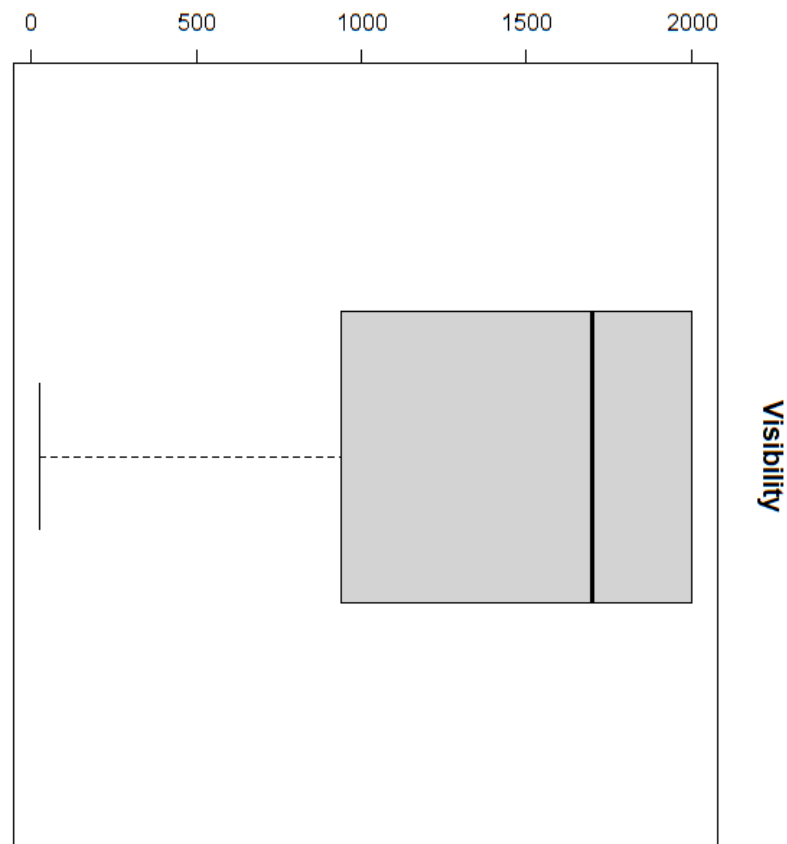


Figura 14 - Diagrama de extremos e quartis da variável 'Visibility'

Medidas de Dispersão

Amplitude Total – É determinado através do valor mínimo e máximo da variável, sendo que neste caso ***amplitude*** = $2000 - 27 = 1973$.

Amplitude Interquartil – Determinamos o valor com recurso à execução do comando específico que efetua esta medida. O resultado obtido foi 940, com base nos valores $Q_{0.25} = 1698$ e $Q_{0.75} = 2000$. Para termos um resultado mais concreto, podemos efetuar a amplitude interquartil com os valores aproximados que foram determinados com as equações anteriormente referidas:

$$AIQ = Q_{0.75} - Q_{0.25} = 1916.011 - 939.4087 = 976.6023$$

Variância – Novamente, com recurso a um comando específico do R, que me calcula imediatamente o resultado, foi obtido que $s^2 \approx 370027.3$. Tal comando, é baseado na seguinte equação:

$$s^2 \approx \frac{\sum_{i=1}^k n_i (x'_i - \bar{x})^2}{n - 1}$$

Desvio Padrão – Facilmente obtido, sendo que já obtemos a variância no ponto anterior, mas decidimos optar pelo comando do R, que me efetua imediatamente o cálculo, sendo que

$$s \approx 608.2987$$

Este resultado está baseado na equação:

$$s \approx \sqrt{s^2}$$

Coefficiente de variação – Determinado com recurso ao desvio padrão e à média, podemos determinar facilmente esta medida de dispersão. Com recurso à equação:

$$CV = \frac{s}{\bar{x}} * 100$$

E com a passagem desta equação para o R, utilizando as variáveis anteriormente calculadas com os respectivos comandos de R (média e desvio padrão), obtivemos que:

$$CV = \frac{s}{\bar{x}} * 100 = 42.33629\%$$

Caracterização da Distribuição de Frequências

Finalizando a análise desta variável apenas necessitamos de demonstrar a caracterização da distribuição de frequências, sendo que apenas existem as medidas de curtose e as medidas de assimetria.

Medidas de Assimetria – Embora pudéssemos verificar através do histograma e do diagrama de extremos e quartis que existe uma assimetria negativa, ou até através da comparação da moda, com a média e com a mediana (média < mediana < moda), optamos por complementar com o cálculo do b_1 , com recurso a um comando R para o resultado. Sendo que o resultado da assimetria deu:

$$b_1 \approx \frac{1}{n} \sum_{i=1}^n \left(\frac{(x_i - \bar{x})}{s} \right)^3 \approx -0.7016663$$

Medidas de Curtose – O achatamento ou curtose, fora determinado através de um comando R para o efeito. Sendo que o resultado deu:

$$b_2 \approx \frac{1}{n} \sum_{i=1}^n \left(\frac{(x_i - \bar{x})}{s} \right)^4 - 3 \approx 2.037884$$

Valor este que indica que existe uma forte concentração de valores em torno da média, existindo uma variação pouco elevada, ou seja, que o grau de curtose é do tipo curva leptocúrtica, alongada, pois $b_2 > 0$.

Dew point Temperature (Celsius) – Temperatura do orvalho

Esta variável possui valores relacionados com a temperatura do orvalho, que estaria no aluguel de bicicletas de uma determinada hora numa determinada data. É uma variável que não é influenciável no aluguel das bicicletas, e que apenas foi estudada por interesse na distribuição de dados relacionados com a temperatura do orvalho na cidade de Seoul, na Coreia do Sul. As frequências desta variável, encontram-se na [Tabela 11 – Tabela de Frequências da variável 'Dew point Temperature'](#).

c_i ^[3]	x'_i ^[4]	n_i ^[5]	f_i ^[6] (em %)	N_i ^[1]	F_i ^[2] (em %)
[-30.6,-26.47[- 28.535	51	0.58	51	0.58
[-26.47,-22.34[- 24.405	143	1.63	194	2.21
[-22.34,-18.21[- 20.276	381	4.35	575	6.56
[-18.21,-14.09[- 16.147	456	5.21	1031	11.77
[-14.09,-9.957[- 12.018	488	5.57	1519	17.34
[-9.957,-5.829[- 7.889	509	5.81	2028	23.15
[-5.829,-1.7[-3.76	722	8.24	2750	31.39
[-1.7,2.429[0.369	1048	11.96	3798	43.36
[2.429,6.557[3.76	871	9.94	4669	53.30
[6.557,10.69[7.889	991	11.31	5660	64.61
[10.69,14.81[12.018	916	10.46	6576	75.07
[14.81,18.94[16.147	774	8.84	7350	83.90
[18.94,23.07[20.276	1106	12.63	8456	96.53
[23.07,27.2[24.405	304	3.47	8760	99.99
		8760	100		

Tabela 11- Tabela de Frequências da variável 'Dew point Temperature'

A demonstração gráfica destes dados, é efetuada através de um histograma, que fora anteriormente gerado com recurso a comandos para o efeito. Este histograma, encontra-se presente na [Figura 15 – Histograma das classes da variável 'Dew point Temperature'](#), anexo em baixo.

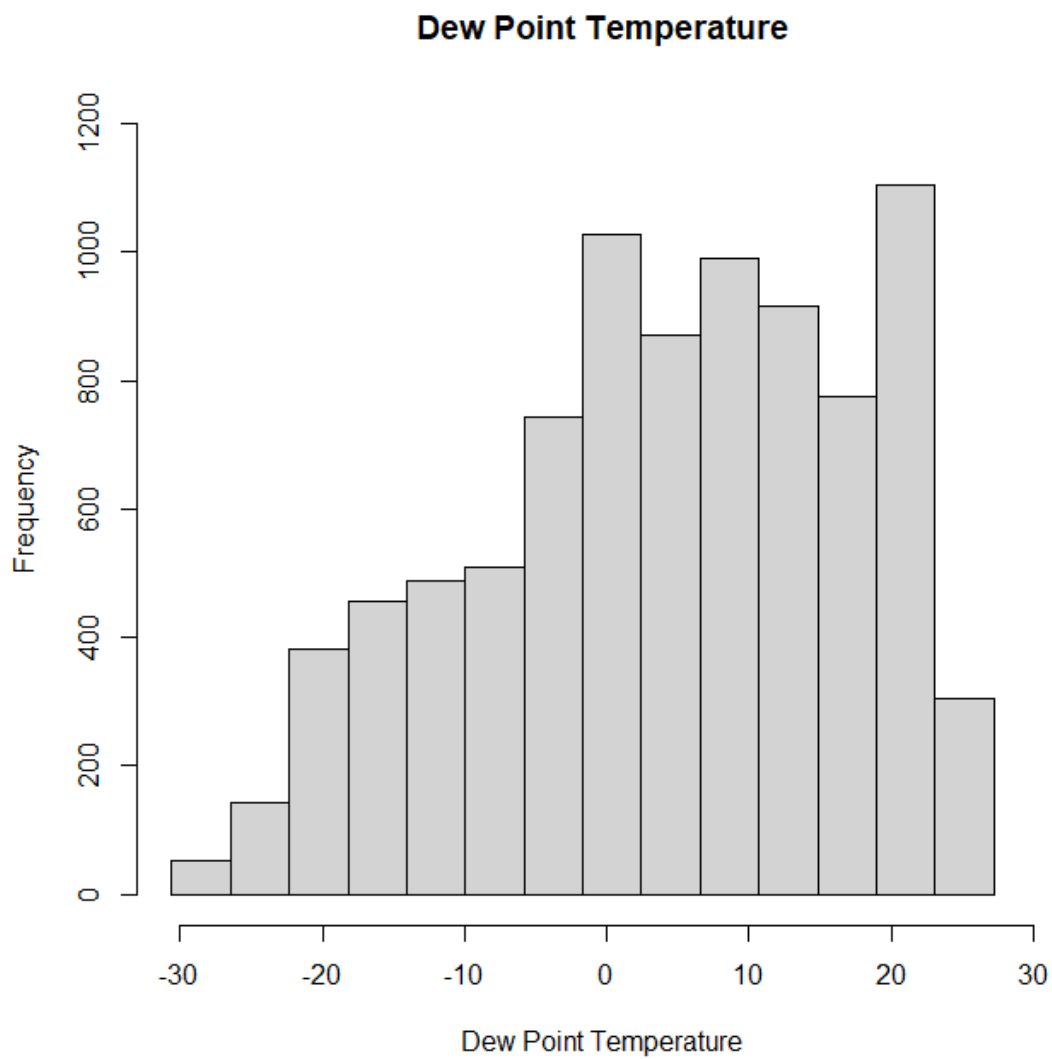


Figura 15 - Histograma das classes da variável 'Dew point Temperature'

Relativamente aos cálculos estatísticos efetuados, conseguimos obter todas as medidas de localização e todas as medidas de dispersão, bem como o achatamento e a curtose. Os resultados obtidos encontram-se demonstrados sucintamente em baixo.

Medidas de Localização

Moda – Relativamente à moda, temos diversas maneiras de a obter. Inicialmente podemos indicar que a Classe Modal é [18.94,23.07[, pois tem a maior frequência absoluta, 1106 alugueis. Calculando de forma bruta e com recurso às classes, temos que a moda é igual ao representante da classe da Classe Modal, ou seja, valor modal é 20.276. Também com recurso ao *RStudio* e com uma função previamente criada para a moda, obtivemos a moda, sem recurso ao agrupamento de classes, e obtivemos o valor 0. Mas sendo os dados agrupados em classes, esses valores apenas são meros indicativos, e não tem peso na moda concreta desta variável. Para calcular a moda, também podemos utilizar o método de King, que se compõe da seguinte forma:

$$moda \approx 18.94 + (23.07 - 18.94) \times \frac{0.0347}{0.0884 + 0.0347} = 20.10418$$

Média – Relativamente à média, optamos por utilizar um comando no *RStudio* que nos fizesse imediatamente a média. O resultado obtido foi 4.073813. Tal valor poderia ser obtido através da equação $\frac{\sum_{i=1}^k n_i x'_i}{n}$, através das frequências absolutas. Mas devido a existirem um número considerável de fatores para chegar a esse resultado, tornar-se-ia demasiado extenso para colocar neste relatório sucinto.

Quantis – Quantos aos quantis, obtivemos os resultados através da execução de comandos que têm o propósito de calcular os quantis. Como, queremos chegar a um diagrama de quartis, decidimos calcular apenas os quartis da variável, como tal, o primeiro quartil é igual a -4.7, o segundo quartil (ou mediana) é igual a 5.1 e o terceiro quartil é igual a 14.8. Também poderíamos analisar a tabela de frequências e determinar de forma bruta o valor dos quantis, sendo que desta forma, o primeiro quartil é -3.76, o segundo quartil (ou mediana) é igual a 3.76, e o terceiro quartil é igual a 12.018. [\(verificar nota que indica como chegamos a estes valores\)](#)

Mas para ser mais concreto, calculamos os valores aproximados através das seguintes equações:

$$1^{\text{o}} \text{ Quartil} - Q_{0.25} \approx -5.829 + (-1.7 + 5.829) \times \frac{(0.25 - 0.2315)}{0.0824} \approx 6.756021$$

$$2^{\text{o}} \text{ Quartil (Mediana)} - Q_{0.5} \approx 2.429 + (6.557 - 2.429) \times \frac{(0.5 - 0.4336)}{0.094} \\ \approx 5.344949$$

$$3^{\text{o}} \text{ Quartil} - Q_{0.75} \approx 10.69 + (14.81 - 10.69) \times \frac{0.75 - 0.6461}{0.1046} \approx 14.78243$$

Embora tenhamos calculado de forma concreta os valores, o melhor indicativo dos quartis é a forma bruta anteriormente feita, devido ao máximo e mínimo da classe que se colocou na equação do 1º Quartil serem valores negativos.

Posteriormente, fora gerado o diagrama de extremos e quartis, que se encontra disposto na [Figura 14 – Diagrama de extremos e quartis da variável 'Visibility'](#).

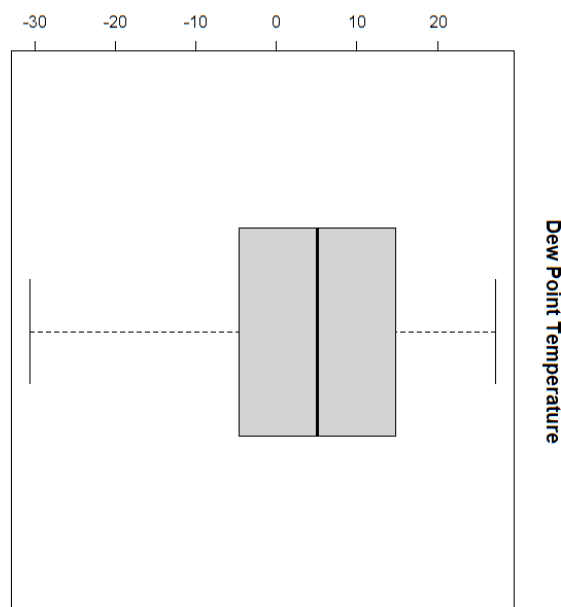


Figura 16 - Diagrama de extremos e quartis da variável 'Dew Point Temperature'

Medidas de Dispersão

Amplitude Total – É determinado através do valor mínimo e máximo da variável, sendo que neste caso **amplitude** = $27.2 - (-30.6) = 57.8$. Este resultado não é um bom indicativo da amplitude, devido ao $\min(x_i)$ ser um valor negativo

Amplitude Interquartil – Determinamos o valor com recurso à execução do comando específico que efetua esta medida. O resultado obtido foi 940, com base nos valores $Q_{0.25} = -4.7$ e $Q_{0.75} = 14.8$. Para termos um resultado mais concreto, podemos efetuar a amplitude interquartil com os valores aproximados que foram determinados com as equações anteriormente referidas (valor não se encontra correto, devido a existir um quartil com valor negativo):

$$AIQ = Q_{0.75} - Q_{0.25} = 14.8 - (-4.7) = 19.5$$

Variância – Novamente, com recurso a um comando específico do R, que me calcula imediatamente o resultado, foi obtido que $s^2 \approx 170.5732$. Tal comando, é baseado na seguinte equação:

$$s^2 \approx \frac{\sum_{i=1}^k n_i (x'_i - \bar{x})^2}{n - 1}$$

Desvio Padrão – Facilmente obtido, sendo que já obtemos a variância no ponto anterior, mas decidimos optar pelo comando do R, que me efetua imediatamente o cálculo, sendo que

$$s \approx 13.06037$$

Este resultado está baseado na equação:

$$s \approx \sqrt{s^2}$$

Coefficiente de variação – Determinado com recurso ao desvio padrão e à média, podemos determinar facilmente esta medida de dispersão. Com recurso à equação:

$$CV = \frac{s}{\bar{x}} * 100$$

E com a passagem desta equação para o R, utilizando as variáveis anteriormente calculadas com os respectivos comandos de R (média e desvio padrão), obtivemos que:

$$CV = \frac{s}{\bar{x}} * 100 = 320.5933\%$$

Caracterização da Distribuição de Frequências

Finalizando a análise desta variável apenas necessitamos de demonstrar a caracterização da distribuição de frequências, sendo que apenas existem as medidas de curtose e as medidas de assimetria.

Medidas de Assimetria – Embora pudéssemos verificar através do histograma e do diagrama de extremos e quartis que existe uma assimetria negativa, ou até através da comparação da moda, com a média e com a mediana (média < mediana < moda), optamos por complementar com o cálculo do b_1 , com recurso a um comando R para o resultado. Sendo que o resultado da assimetria deu:

$$b_1 \approx \frac{1}{n} \sum_{i=1}^n \left(\frac{(x_i - \bar{x})}{s} \right)^3 \approx -0.3672355$$

Medidas de Curtose – O achatamento ou curtose, fora determinado através de um comando R para o efeito. Sendo que o resultado deu:

$$b_2 \approx \frac{1}{n} \sum_{i=1}^n \left(\frac{(x_i - \bar{x})}{s} \right)^4 - 3 \approx 2.244317$$

Valor este que indica que existe uma forte concentração de valores em torno da média, existindo uma variação pouco elevada, ou seja, que o grau de curtose é do tipo curva leptocúrtica, alongada, pois $b_2 > 0$.

Solar Radiation – Radiação Solar

Esta variável possui valores relacionados com a radiação solar, que estaria no aluguer de bicicletas de uma determinada hora numa determinada data. É uma variável que pode ter uma influência no aluguer de bicicletas, pois quando existe uma radiação solar menor, existe mais pessoas a andarem de bicicleta. Pode ser verificado através das frequências desta variável, que se encontram na [Tabela 12 – Tabela de Frequências da variável 'Solar Radiation'](#).

c_i ^[3]	x'_i ^[4]	n_i ^[5]	f_i ^[6] (em %)	N_i ^[1]	F_i ^[2] (em %)
[0,0.2514[0.1257	5329	60.83	5329	60.83
[0.2514,0.5029[0.3771	543	6.20	5872	67.03
[0.5029,0.7543[0.6285	416	4.75	6288	71.78
[0.7543,1.006[0.88	397	4.53	6685	76.31
[1.006,1.257[1.1314	402	4.59	7087	80.90
[1.257,1.509[1.3771	235	2.68	7322	83.58
[1.509,1.76[1.6285	263	3.00	7585	86.59
[1.76,2.011[1.88	264	3.01	7849	89.60
[2.011,2.263[2.1314	208	2.37	8057	91.97
[2.263,2.514[2.3828	224	2.56	8281	94.53
[2.514,2.766[2.6343	173	1.97	8454	96.51
[2.766,3.017[2.8857	158	1.80	8612	98.31
[3.017,3.269[3.1371	107	1.22	8719	99.53
[3.269,3.52[3.3885	41	0.47	8760	100
		8760	100		

Tabela 12 - Tabela de Frequências da variável 'Solar Radiation'

A demonstração gráfica destes dados, é efetuada através de um histograma, que fora anteriormente gerado com recurso a comandos para o efeito. Este histograma, encontra-se presente na [Figura 17 – Histograma das classes da variável 'Solar Radiation'](#), anexado em baixo.

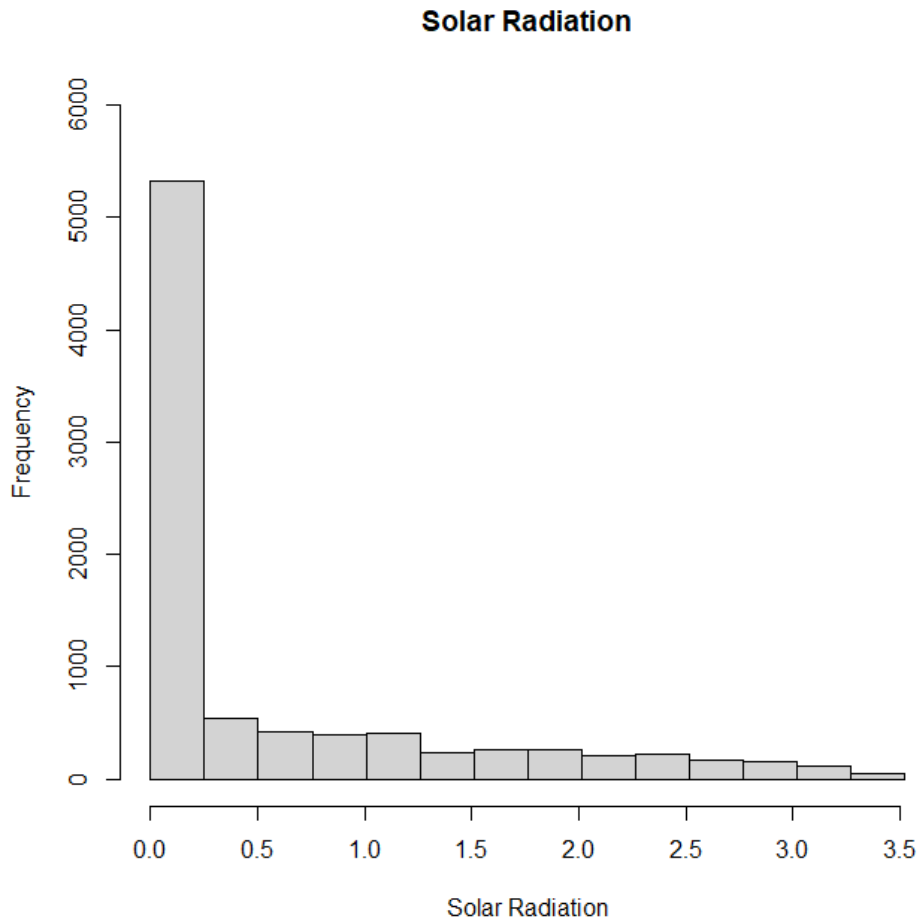


Figura 17 - Histograma das classes da variável 'Solar Radiation'

Relativamente aos cálculos estatísticos efetuados, conseguimos obter todas as medidas de localização e todas as medidas de dispersão, bem como o achatamento e a curtose. Os resultados obtidos encontram-se demonstrados sucintamente em baixo.

Medidas de Localização

Moda – Relativamente à moda, temos diversas maneiras de a obter. Inicialmente podemos indicar que a Classe Modal é $[0,0.2514[$, pois tem a maior frequência absoluta, 5329 alugueis. Calculando de forma bruta e com recurso às classes, temos que a moda é igual ao representante da classe da Classe Modal, ou seja, valor modal é 0.1257. Também com recurso ao *RStudio* e com uma função previamente criada para a moda, obtivemos a moda, sem recurso ao agrupamento de classes, e obtivemos o valor 0. Mas sendo os dados agrupados em classes, esses

valores apenas são meros indicativos, e não tem peso na moda concreta desta variável. Para calcular a moda, também podemos utilizar o método de King, que se compõe da seguinte forma:

$$moda \approx 0 + (0.2514 - 0) \times \frac{0.0620}{0 + 0.0620} = 0.2514$$

Média – Relativamente à média, optamos por utilizar um comando no *RStudio* que nos fizesse imediatamente a média. O resultado obtido foi 0.5691107. Tal valor poderia ser obtido através da equação $\frac{\sum_{i=1}^k n_i x'_i}{n}$, através das frequências absolutas. Mas devido a existirem um número considerável de fatores para chegar a esse resultado, tornar-se-ia demasiado extenso para colocar neste relatório sucinto.

Quantis – Quanto aos quantis, obtivemos os resultados através da execução de comandos que têm o propósito de calcular os quantis. Como, queremos chegar a um diagrama de quartis, decidimos calcular apenas os quartis da variável, como tal, o primeiro quartil é igual a 0, o segundo quartil (ou mediana) é igual a 0.01 e o terceiro quartil é igual a 0.93. Também poderíamos analisar a tabela de frequências e determinar de forma bruta o valor dos quantis, sendo que desta forma, o primeiro quartil é 0.1257, o segundo quartil (ou mediana) é igual a 0.1257, e o terceiro quartil é igual a 0.88. [\(verificar nota que indica como chegamos a estes valores\)](#)

Mas para ser mais concreto, calculamos os valores aproximados através das seguintes equações:

$$1^{\text{o}} \text{ Quartil} - Q_{0.25} \approx 0 + (0.2514 - 0) \times \frac{(0.25 - 0)}{0.6093} \approx 0.1031512$$

$$2^{\text{o}} \text{ Quartil (Mediana)} - Q_{0.5} \approx 0 + (0.2514 - 0) \times \frac{(0.5 - 0)}{0.6093} \approx 0.2063023$$

$$3^{\text{o}} \text{ Quartil} - Q_{0.75} \approx 0.7543 + (1.006 - 0.7543) \times \frac{0.75 - 0.7178}{0.0453} \approx 0.9332126$$

Posteriormente, fora gerado o diagrama de extremos e quartis, que se encontra disposto na [Figura 18 - Diagrama de extremos e quartis da variável 'Solar Radiation'](#).

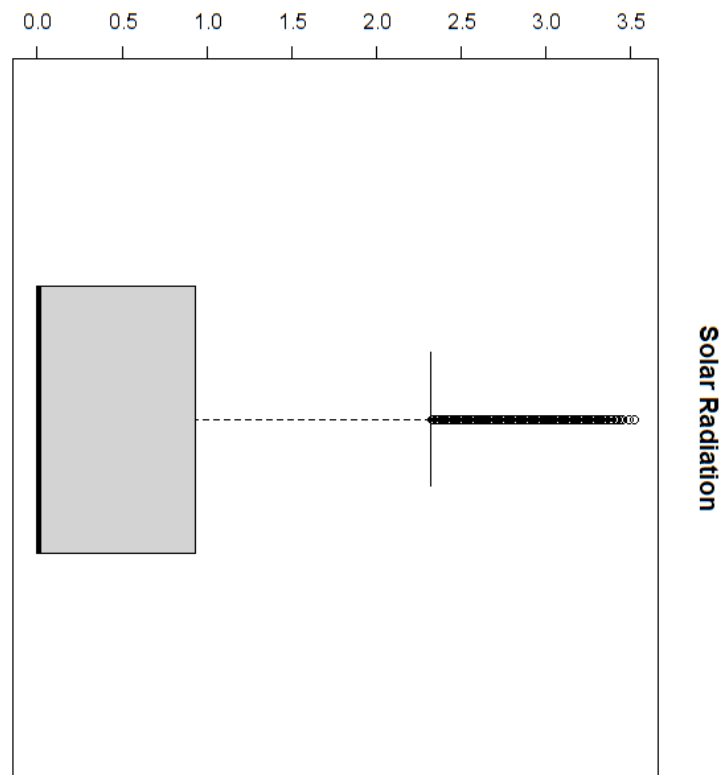


Figura 18 - Diagrama de extremos e quartis da variável 'Solar Radiation'

Medidas de Dispersão

Amplitude Total – É determinado através do valor mínimo e máximo da variável, sendo que neste caso ***amplitude*** = $3.52 - 0 = 3.52$.

Amplitude Interquartil – Determinamos o valor com recurso à execução do comando específico que efetua esta medida. O resultado obtido foi 0.93, com base nos valores $Q_{0.25} = 0$ e $Q_{0.75} = 0.93$. Para termos um resultado mais concreto, podemos efetuar a amplitude interquartil com os valores aproximados que foram determinados com as equações anteriormente referidas (valor não se encontra correto, devido a existir um quartil com valor negativo):

$$AIQ = Q_{0.75} - Q_{0.25} = 0.93 - 0 = 0.93$$

Variância – Novamente, com recurso a um comando específico do R, que me calcula imediatamente o resultado, foi obtido que $s^2 \approx 0.75472$. Tal comando, é baseado na seguinte equação:

$$s^2 \approx \frac{\sum_{i=1}^k n_i (x'_i - \bar{x})^2}{n - 1}$$

Desvio Padrão – Facilmente obtido, sendo que já obtemos a variância no ponto anterior, mas decidimos optar pelo comando do R, que me efetua imediatamente o cálculo, sendo que

$$s \approx 0.8687462$$

Este resultado está baseado na equação:

$$s \approx \sqrt{s^2}$$

Coefficiente de variação – Determinado com recurso ao desvio padrão e à média, podemos determinar facilmente esta medida de dispersão. Com recurso à equação:

$$CV = \frac{s}{\bar{x}} * 100$$

E com a passagem desta equação para o R, utilizando as variáveis anteriormente calculadas com os respetivos comandos de R (média e desvio padrão), obtivemos que:

$$CV = \frac{s}{\bar{x}} * 100 = 152.6498\%$$

Caracterização da Distribuição de Frequências

Finalizando a análise desta variável apenas necessitamos de demonstrar a caracterização da distribuição de frequências, sendo que apenas existem as medidas de curtose e as medidas de assimetria.

Medidas de Assimetria – Embora pudéssemos verificar através do histograma e do diagrama de extremos e quartis que existe uma assimetria positiva, ou até através da comparação da moda, com a média e com a mediana (média > mediana > moda), optamos por complementar com o

cálculo do b_1 , com recurso a um comando R para o resultado. Sendo que o resultado da assimetria deu:

$$b_1 \approx \frac{1}{n} \sum_{i=1}^n \left(\frac{(x_i - \bar{x})}{s} \right)^3 \approx 1.503782$$

Medidas de Curtose – O achatamento ou curtose, fora determinado através de um comando R para o efeito. Sendo que o resultado deu:

$$b_2 \approx \frac{1}{n} \sum_{i=1}^n \left(\frac{(x_i - \bar{x})}{s} \right)^4 - 3 \approx 4.125105$$

Valor este que indica que existe uma forte concentração de valores em torno da média, existindo uma variação pouco elevada, ou seja, que o grau de curtose é do tipo curva leptocúrtica, alongada, pois $b_2 > 0$.

Rainfall (mm) – Precipitação (mm)

Como a sua nomenclatura indica, esta variável mantém o registo da precipitação numa determinada hora de uma determinada data. É uma medição, que é influenciável no aluguer de bicicletas, pois considera-se que quando não existe precipitação, existe uma maior probabilidade de alguém alugar uma bicicleta. Tal pode ser consultado através dos cálculos da moda.

c_i [3]	x'_i [4]	n_i [5]	f_i [6] (em %)	N_i [1]	F_i [2] (em %)
[0,2.5[1.25	8610	98.29	8610	98.29
[2.5,5[3.75	79	0.90	8689	99.19
[5,7.5[6.25	31	0.35	8720	99.54
[7.5,10[8.75	17	0.19	8737	99.74
[10,12[11.25	4	0.05	8741	99.78
[12,15[13.75	6	0.07	8747	99.85
[15,18[16.25	3	0.03	8750	99.89
[18,20[18.75	5	0.06	8755	99.94
[20,22[21.25	2	0.02	8757	99.97
[22,25[23.75	1	0.01	8758	99.98
[25,28[26.25	0	0.00	8758	99.98
[28,30[28.75	1	0.01	8759	99.99
[30,32[31.25	0	0.00	8759	99.99
[32,35]	33.75	1	0.01	8760	100
		8760	100		

Tabela 13 - Tabela de Frequências da variável 'Rainfall'

A demonstração gráfica destes dados, é efetuada através de um histograma, que fora anteriormente gerado com recurso a comandos para o efeito. Este histograma, encontra-se presente na [Figura 19 - Histograma das classes da variável 'Rainfall'](#), anexado em baixo.

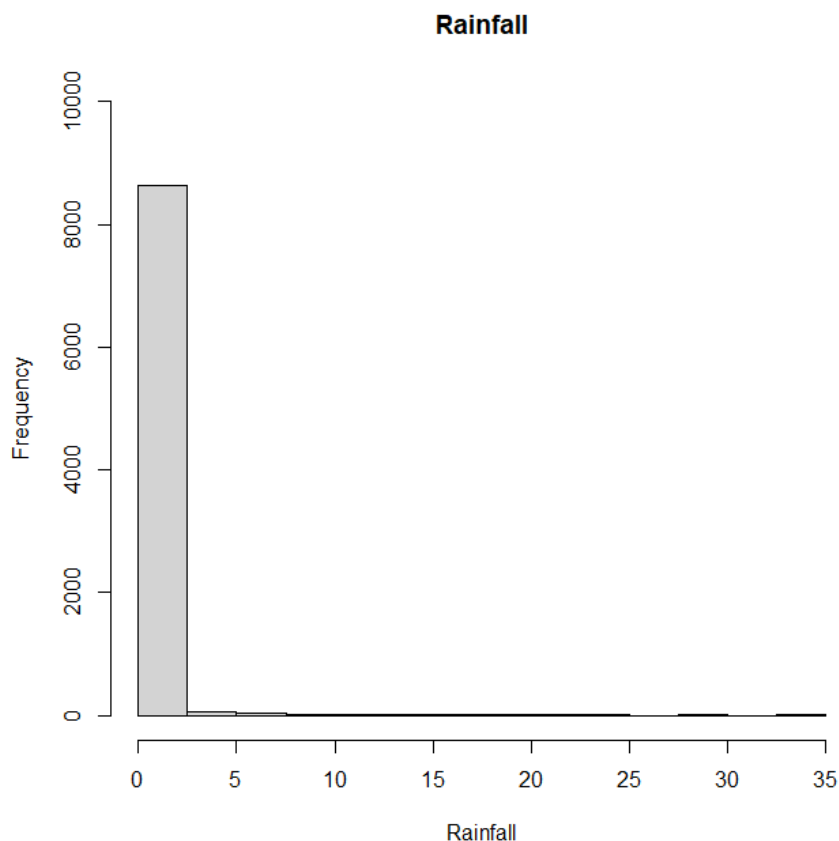


Figura 19 - Histograma das classes da variável 'Rainfall'

Relativamente aos cálculos estatísticos efetuados, conseguimos obter todas as medidas de localização e todas as medidas de dispersão, bem como o achatamento e a curtose. Os resultados obtidos encontram-se demonstrados sucintamente em baixo.

Medidas de Localização

Moda – Relativamente à moda, temos diversas maneiras de a obter. Inicialmente podemos indicar que a Classe Modal é $[0, 2.5]$, pois tem a maior frequência absoluta, 8610. Calculando de forma bruta e com recurso às classes, temos que a moda é igual ao representante da classe da Classe Modal, ou seja, valor modal é 1.25. Também com recurso ao RStudio e com uma função previamente criada para a moda, obtivemos a moda, sem recurso ao agrupamento de classes, e obtivemos uma variável modal, com o valor 0. Este valor encontra-se dentro da Classe Modal e pode ser indicativo de que a moda se encontra mesmo nessa classe. Mas sendo os dados agrupados em classe, este valor é apenas um mero indicativo, e não tem peso na moda concreta

desta variável. Para calcular a moda, também podemos utilizar o método de King, que se compõe da seguinte forma:

$$moda \approx 0 + (2.5 - 0) \times \frac{0.009}{0 + 0.090} \approx 0.25$$

Média – Relativamente à média, optamos por utilizar um comando no RStudio que nos fizesse imediatamente a média. O resultado obtido foi 0.1486872. Tal valor poderia ser obtido através da equação $\frac{\sum_{i=1}^k n_i x'_i}{n}$, através das frequências absolutas. Mas devido a existirem um número considerável de fatores para chegar a esse resultado, tornar-se-ia demasiado extenso para colocar neste relatório sucinto.

Quantis – Quanto aos quantis, obtivemos os resultados através da execução de comandos que têm o propósito de calcular os quantis. Como, queremos chegar a um diagrama de quantis, decidimos calcular apenas os quartis da variável, como tal, o primeiro quartil é igual a 0, o segundo quartil (ou mediana) é igual a 0 e o terceiro quartil é igual a 0. Também poderíamos analisar a tabela de frequências e determinar de forma bruta o valor dos quantis, sendo que desta forma, o primeiro quartil é 1.25, o segundo quartil (ou mediana) é igual a 1.25, e o terceiro quartil é igual a 1.25. [\(verificar nota que indica como chegamos a estes valores\)](#)

Mas para ser mais concreto, calculamos os valores aproximados através das seguintes equações:

$$1^{\circ} \text{ Quartil} - Q_{0.25} \approx 0 + 2.5 \times \frac{(0.25 - 0)}{0.9829} \approx 0.6358$$

$$2^{\circ} \text{ Quartil (Mediana)} - Q_{0.5} \approx 0 + 2.5 \times \frac{(0.5 - 0)}{0.9829} \approx 1.2717$$

$$3^{\circ} \text{ Quartil} - Q_{0.75} \approx 0 + 2.5 \times \frac{(0.75 - 0)}{0.9829} \approx 1.9076$$

Embora tenhamos calculado de forma concreta os valores, o melhor indicativo dos quartis é a forma bruta anteriormente feita, devido a existir uma grande concentração dos dados perto do zero.

Posteriormente, fora gerado o diagrama de extremos e quartis, que se encontra disposto na [Figura 20 - Diagrama de extremos e quartis da variável 'Rainfall'](#).

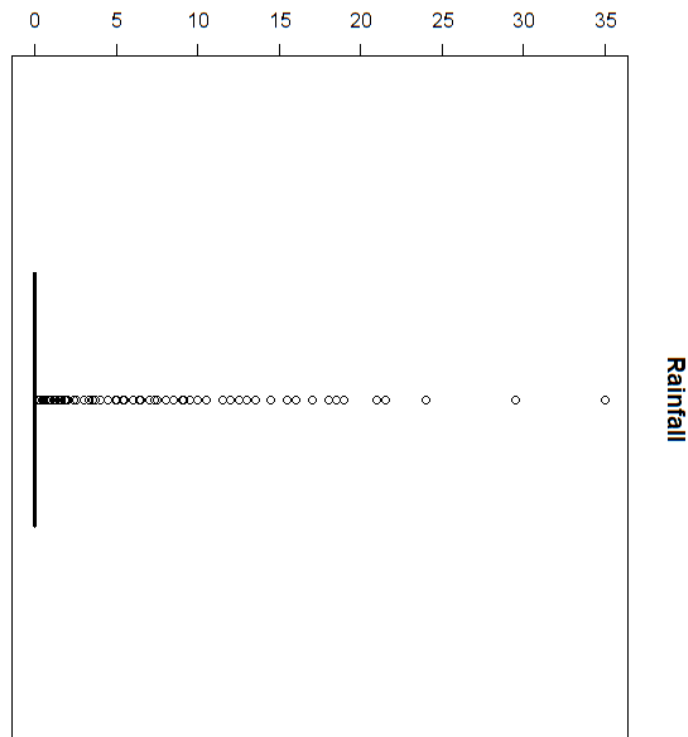


Figura 20 - Diagrama de extremos e quartis da variável 'Rainfall'

Medidas de Dispersão

Amplitude Total – É determinado através do valor mínimo e máximo da variável, sendo que neste caso ***amplitude*** = $35 - 0 = 35$.

Amplitude Interquartil – Determinamos o valor com recurso à execução do comando específico que efetua esta medida. O resultado obtido foi 0, com base nos valores $Q_{0.25} = 0$ e $Q_{0.75} = 0$. Para termos um resultado mais concreto, podemos efetuar a amplitude interquartil com os valores aproximados que foram determinados com as equações anteriormente referidas:

$$AIQ = Q_{0.75} - Q_{0.25} = 1.9076 - 0.6358 = 1.2718$$

Variância – Novamente, com recurso a um comando específico do R, que me calcula imediatamente o resultado, foi obtido que $s^2 \approx 1.272819$. Tal comando, é baseado na seguinte equação:

$$s^2 \approx \frac{\sum_{i=1}^k n_i (x'_i - \bar{x})^2}{n - 1}$$

Desvio Padrão – Facilmente obtido, sendo que já obtemos a variância no ponto anterior, mas decidimos optar pelo comando do R, que me efetua imediatamente o cálculo, sendo que

$$s \approx 1.128193$$

Este resultado está baseado na equação:

$$s \approx \sqrt{s^2}$$

Coefficiente de variação – Determinado com recurso ao desvio padrão e à média, podemos determinar facilmente esta medida de dispersão. Com recurso à equação:

$$CV = \frac{s}{\bar{x}} * 100$$

E com a passagem desta equação para o R, utilizando as variáveis anteriormente calculadas com os respetivos comandos de R (média e desvio padrão), obtivemos que:

$$CV = \frac{s}{\bar{x}} * 100 = 758.7693\%$$

Caracterização da Distribuição de Frequências

Finalizando a análise desta variável apenas necessitamos de demonstrar a caracterização da distribuição de frequências, sendo que apenas existem as medidas de curtose e as medidas de assimetria.

Medidas de Assimetria – Embora pudéssemos verificar através do histograma e do diagrama de extremos e quartis que existe uma assimetria negativa, ou até através da comparação da moda, com a média e com a mediana (média < mediana < moda), optamos por complementar com o cálculo do b_1 , com recurso a um comando R para o resultado. Sendo que o resultado da assimetria deu:

$$b_1 \approx \frac{1}{n} \sum_{i=1}^n \left(\frac{(x_i - \bar{x})}{s} \right)^3 \approx 14.53074$$

Medidas de Curtose – O achatamento ou curtose, fora determinado através de um comando R para o efeito. Sendo que o resultado deu:

$$b_2 \approx \frac{1}{n} \sum_{i=1}^n \left(\frac{(x_i - \bar{x})}{s} \right)^4 - 3 \approx 287.8278$$

Valor este que indica que existe uma forte concentração de valores em torno da média, existindo uma variação pouco elevada, ou seja, que o grau de curtose é do tipo curva leptocúrtica, alongada, pois $b_2 > 0$.

Snowfall (mm) – Queda de neve (mm)

Como a sua nomenclatura indica, esta variável mantém o registo da queda de neve numa determinada hora de uma determinada data. É uma medição, que é influenciável no aluguer de bicicletas, pois considera-se que quando não existe queda de neve, existe uma maior probabilidade de alguém alugar uma bicicleta. Tal pode ser consultado através dos cálculos da moda.

c_i [3]	x'_i [4]	n_i [5]	f_i [6] (em %)	N_i [1]	F_i [2] (em %)
[0,0.629[0.3145	8446	96.42	8446	96.42
[0.629,1.26[0.9431	137	1.56	8583	97.98
[1.26,1.89[1.5716	34	0.39	8617	98.37
[1.89,2.51[2.2002	62	0.71	8679	99.08
[2.51,3.14[2.8288	28	0.32	8707	99.39
[3.14,3.77[3.4573	27	0.31	8734	99.70
[3.77,4.4[4.0859	16	0.18	8750	99.89
[4.4,5.03[4.7145	4	0.05	8754	99.93
[5.03,5.66[5.3431	1	0.01	8755	99.94
[5.66,6.29[5.9716	1	0.01	8756	99.95
[6.29,6.91[6.6002	0	0.00	8756	99.95
[6.91,7.54[7.2288	2	0.02	8758	99.98
[7.54,8.17[7.8573	0	0.00	8758	99.98
[8.17,8.8[8.4859	2	0.02	8760	100
		8760	100		

Tabela 14 - Tabela de Frequências da variável 'Snowfall'

A demonstração gráfica destes dados, é efetuada através de um histograma, que fora anteriormente gerado com recurso a comandos para o efeito. Este histograma, encontra-se presente na [Figura 21 - Histograma das classes da variável 'Snowfall'](#), anexado em baixo.

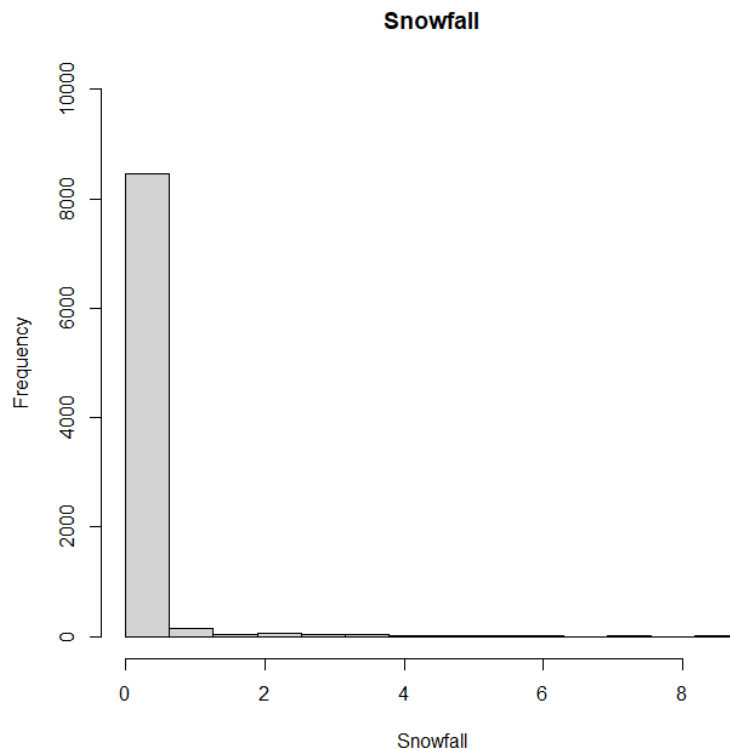


Figura 21 - Histograma das classes da variável 'Snowfall'

Relativamente aos cálculos estatísticos efetuados, conseguimos obter todas as medidas de localização e todas as medidas de dispersão, bem como o achatamento e a curtose. Os resultados obtidos encontram-se demonstrados sucintamente em baixo.

Medidas de Localização

Moda – Relativamente à moda, temos diversas maneiras de a obter. Inicialmente podemos indicar que a Classe Modal é $[0, 0.629[$, pois tem a maior frequência absoluta, 8446. Calculando de forma bruta e com recurso às classes, temos que a moda é igual ao representante da classe da Classe Modal, ou seja, valor modal é 0.3145. Também com recurso ao *RStudio* e com uma função previamente criada para a moda, obtivemos a moda, sem recurso ao agrupamento de classes, e obtivemos uma variável modal, com o valor 0. Este valor encontra-se dentro da Classe Modal e pode ser indicativo de que a moda se encontra mesmo nessa classe. Mas sendo os dados agrupados em classe, este valor é apenas um mero indicativo, e não tem peso na moda concreta desta variável. Para calcular a moda, também podemos utilizar o método de King, que se compõe da seguinte forma:

$$moda \approx 0 + (0.629 - 0) \times \frac{0.0156}{0 + 0.156} \approx 0.629$$

Média – Relativamente à média, optamos por utilizar um comando no *RStudio* que nos fizesse imediatamente a média. O resultado obtido foi 0.07506849. Tal valor poderia ser obtido através da equação $\frac{\sum_{i=1}^k n_i x'_i}{n}$, através das frequências absolutas. Mas devido a existirem um número considerável de fatores para chegar a esse resultado, tornar-se-ia demasiado extenso para colocar neste relatório sucinto.

Quantis – Quanto aos quantis, obtivemos os resultados através da execução de comandos que têm o propósito de calcular os quantis. Como, queremos chegar a um diagrama de quartis, decidimos calcular apenas os quartis da variável, como tal, o primeiro quartil é igual a 0, o segundo quartil (ou mediana) é igual a 0 e o terceiro quartil é igual a 0. Também poderíamos analisar a tabela de frequências e determina de forma bruta o valor dos quantis, sendo que desta forma, o primeiro quartil é 0.3145, o segundo quartil (ou mediana) é igual a 0.3145, e o terceiro quartil é igual a 0.3145. [\(verificar nota que indica como chegamos a estes valores\)](#)

Mas para ser mais concreto, calculamos os valores aproximados através das seguintes equações:

$$1^{\text{o}} \text{ Quartil} - Q_{0.25} \approx 0 + 0.629 \times \frac{(0.25 - 0)}{0.9642} \approx 0.1630$$

$$2^{\text{o}} \text{ Quartil (Mediana)} - Q_{0.5} \approx 0 + 0.629 \times \frac{(0.5 - 0)}{0.9642} \approx 0.3261$$

$$3^{\text{o}} \text{ Quartil} - Q_{0.75} \approx 0 + 0.629 \times \frac{(0.75 - 0)}{0.9642} \approx 0.4892$$

Posteriormente, fora gerado o diagrama de extremos e quartis, que se encontra disposto na [Figura 22 - Diagrama de extremos e quartis da variável 'Snowfall'](#).

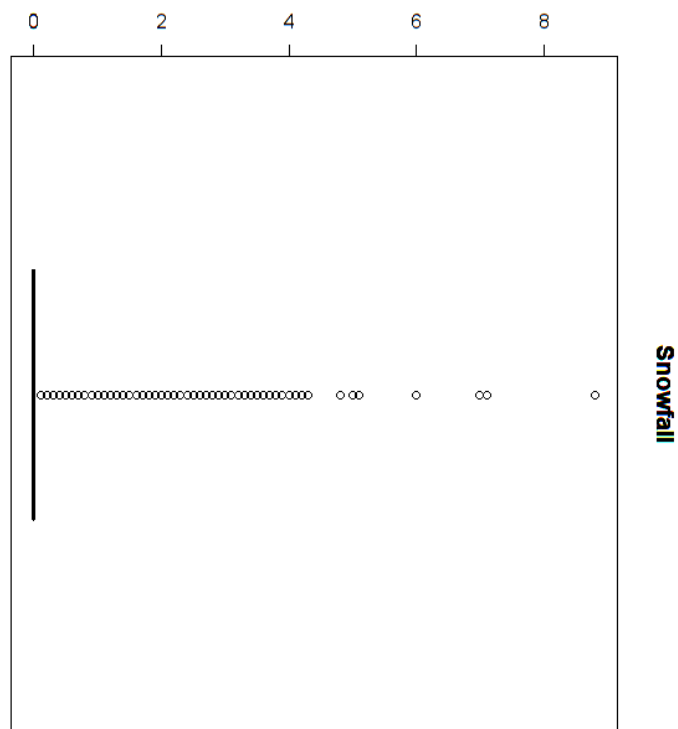


Figura 22 - Diagrama de extremos e quartis da variável 'Snowfall'

Medidas de Dispersão

Amplitude Total – É determinado através do valor mínimo e máximo da variável, sendo que neste caso **amplitude** = $8.8 - 0 = 8.8$.

Amplitude Interquartil – Determinamos o valor com recurso à execução do comando específico que efetua esta medida. O resultado obtido foi 0, com base nos valores $Q_{0.25} = 0$ e $Q_{0.75} = 0$. Para termos um resultado mais concreto, podemos efetuar a amplitude interquartil com os valores aproximados que foram determinados com as equações anteriormente referidas:

$$AIQ = Q_{0.75} - Q_{0.25} = 0.4892 - 0.1630 = 0.3262$$

Variância – Novamente, com recurso a um comando específico do R, que me calcula imediatamente o resultado, foi obtido que $s^2 \approx 0.1907472$. Tal comando, é baseado na seguinte equação:

$$s^2 \approx \frac{\sum_{i=1}^k n_i (x'_i - \bar{x})^2}{n - 1}$$

Desvio Padrão – Facilmente obtido, sendo que já obtemos a variância no ponto anterior, mas decidimos optar pelo comando do R, que me efetua imediatamente o cálculo, sendo que

$$s \approx 0.4367462$$

Este resultado está baseado na equação:

$$s \approx \sqrt{s^2}$$

Coeficiente de variação – Determinado com recurso ao desvio padrão e à média, podemos determinar facilmente esta medida de dispersão. Com recurso à equação:

$$CV = \frac{s}{\bar{x}} * 100$$

E com a passagem desta equação para o R, utilizando as variáveis anteriormente calculadas com os respetivos comandos de R (média e desvio padrão), obtivemos que:

$$CV = \frac{s}{\bar{x}} * 100 = 581.7969\%$$

Caracterização da Distribuição de Frequências

Finalizando a análise desta variável apenas necessitamos de demonstrar a caracterização da distribuição de frequências, sendo que apenas existem as medidas de curtose e as medidas de assimetria.

Medidas de Assimetria – Embora pudéssemos verificar através do histograma e do diagrama de extremos e quartis que existe uma assimetria negativa, ou até através da comparação da moda, com a média e com a mediana (média < mediana < moda), optamos por complementar com o cálculo do b_1 , com recurso a um comando R para o resultado. Sendo que o resultado da assimetria deu:

$$b_1 \approx \frac{1}{n} \sum_{i=1}^n \left(\frac{(x_i - \bar{x})}{s} \right)^3 \approx 8.439355$$

Medidas de Curtose – O achatamento ou curtose, fora determinado através de um comando R para o efeito. Sendo que o resultado deu:

$$b_2 \approx \frac{1}{n} \sum_{i=1}^n \left(\frac{(x_i - \bar{x})}{s} \right)^4 - 3 \approx 96.74911$$

Valor este que indica que existe uma forte concentração de valores em torno da média, existindo uma variação pouco elevada, ou seja, que o grau de curtose é do tipo curva leptocúrtica, alongada, pois $b_2 > 0$.

Conclusão

Esta análise relativamente complexa, mas sucinta, requereu que aprofundássemos o nosso conhecimento na área da estatística descritiva, decrementando assim as dificuldades na componente teórica da unidade curricular de Métodos Estatísticos.

Também surgiram alguns constrangimentos relativamente ao script de R, alguns que foram resumidos e apresentados nas notas complementares, e outros apenas foram resolvidos no imediato, com recurso à consulta de sites para o efeito, ou até mesmo, apoio dos professores. Essa aprendizagem que retiramos em ambos os aspetos, na análise e na programação, dar-nos-á uma maior valência, caso surja interesse dos membros no grupo na área de análise de dados e dar-nos-á um incremento no leque de capacidades que adquirimos na trajetória do curso.

Notas complementares

[Nota 1] – **Biblioteca *fdth*** – Relativamente a esta biblioteca, existiu um constrangimento devido ao facto de esta, não permitir a colocação de um parêntesis reto na última classe, ou seja, ter a última classe com o seguinte formato $[b_{k-1}; b_k]$, na utilização do comando *fdt*, que retorna a tabela de frequências de uma variável, separada por classes. Tal constrangimento, levou a que na execução do comando, nos demonstrasse uma tabela de frequências incompleta, com necessidade de acrescentar às frequências absolutas e frequências relativas da última classe, a ocorrência de valores que efetivamente seriam iguais ao máximo. Como tal, bastou-nos utilizar o valor da amostra (neste caso 8760), e efetuar a diferença com o último valor obtido na frequência absoluta acumulada através do comando *fdt*, para obtermos o valor de frequência absoluta correto da última classe. Para o valor de frequência relativa, bastou-nos pegar nesse valor de frequência absoluta já corrigido, e dividir pelo valor da amostra. Este constrangimento foi comunicado ao professor numa aula, e também fora devidamente corrigido com a resolução anterior. A correção deste constrangimento, permitiu-nos uma fluida e correta passagem de valores da consola do *RStudio*, após execução do comando *fdt*, para as tabelas de frequências que se encontram anexadas neste documento [ir para índice de tabelas](#).

[Nota 2] – **Arredondamentos nas Tabelas** – Nas tabelas de frequências presentes neste documento, queríamos deixar claro, que nos representantes de classe, que possuíam um número considerável de casas decimais, devido à não aplicação de arredondamento do valor de amplitude, efetuamos um arredondamento na quarta casa decimal, deixando apenas quatro casas decimais presentes nos valores. Todos os outros valores, como frequências relativas ou até valores nas classes, fora tudo determinado pelo comando *fdt* da biblioteca *fdth*.

[Nota 3] – **Estudo das variáveis** – Para complementar o que fora mencionado nos tópicos das variáveis hora e data, existiu um estudo com alguns avanços, mas são apenas demonstrativos de como o faríamos caso aplicássemos a uma variável com classificação igual. Os valores que as variáveis possuem, ou a sua forma de estarem descritos, não permitem um estudo assertivo sobre os seus valores, pois individualmente, não possuem qualquer contribuição relevante para a estatística do aluguer de bicicletas. Relativamente ao agrupamento em classes da variável hora, que é quantitativa discreta, só possui o intuito de demonstrar, que quando estamos perante uma variável quantitativa discreta com muitos valores distintos, devemos proceder dessa forma.

[Nota 4] – **Determinação de quartis** – Embora não especificado na secção em que foi feita o cálculo de quartis das variáveis quantitativas contínuas, a determinação dos quartis, foi efetuada tendo em conta os valores que se encontram na coluna de frequências relativas acumuladas, ou seja, o 1º Quartil é determinado como sendo o representante de classe (x'_i) da primeira classe cuja frequência relativa acumulada é superior a 0.25 (25%). O 2º Quartil cuja frequência relativa acumulada é superior a 0.5 (50%) e o 3º Quartil cuja frequência relativa acumulada é superior a 0.75 (75%). Embora não seja indicativo de que os quartis se encontram devidamente demonstrados/calculados, pode e deve servir para efetuarmos os cálculos do valor aproximado.

Referências bibliográficas

Bibliografia

- Slides disponibilizados pelos professores, no moodle (Capítulo 1 – Estatística Descritiva).
- Fichas das aulas práticas (1-3).

Equações utilizadas

[1] – Frequência Absoluta acumulada determinada através de $N_k = \sum_{i=1}^k n_i = n$;

[2] – Frequência Relativa acumulada determina através de $F_k = \sum_{i=1}^k f_i = 1$;

[3] – Classes, com representação $c_k =]b_{k-1}; b_k]$ com $b_k = b_{k-1} + h$;

[4] – Representante de Classe, determinado através da equação $x'_k = \frac{(b_{k-1} + b_k)}{2}$;

[5] – Frequência Absoluta determinada através da contagem de casos por classe ou n_k ;

[6] – Frequência Relativa determinada através de $\frac{n_i}{N}$.