21BDS0340

Abhinav Dinesh Srivatsa

Exploratory Data Analysis Lab

Experiment 8.1

Code:

```
setwd("~/College Work/Year 4 Semester 1 (Sem 7)/Exploratory Data Analysis
Lab/Experiment 8-1")
data = read.csv("mtcars.csv")

library(psych)
library(moments)
library(ggplot2)
```

Output:

```
> setwd("~/College Work/Year 4 Semester 1 (Sem 7)/Exploratory Data Analysis
Lab/Experiment 8-1")
> data = read.csv("mtcars.csv")
>
> library(psych)
> library(moments)
> library(ggplot2)
```

Code:

```
# 1. measures of central tendency
# arithmetic mean
mean(data$hp)
# geometric mean
exp(mean(log(data$hp)))
# harmonic mean
1 / mean(1 / data$hp)
# median
median(data$hp)
# quantiles
quantile(data$hp)
# deciles
quantile(data$hp, probs = seq(.1, 1, by = .1))
# percentiles
quantile(data$hp, probs = seq(.01, 1, by = .01))
```

Output:

```
> # 1. measures of central tendency
> # arithmetic mean
> mean(data$hp)
[1] 146.6875
> # geometric mean
> exp(mean(log(data$hp)))
```

```
[1] 131.8837
> # harmonic mean
> 1 / mean(1 / data$hp)
[1] 118.2289
> # median
> median(data$hp)
[1] 123
> # quantiles
> quantile(data$hp)
   0%   25%   50%   75%  100%
 52.0  96.5 123.0 180.0 335.0
> # deciles
> quantile(data$hp, probs = seq(.1, 1, by = .1))
  10%   20%   30%   40%   50%   60%   70%   80%   90%  100%
 66.0  93.4 106.2 110.0 123.0 165.0 178.5 200.0 243.5 335.0
> # percentiles
> quantile(data$hp, probs = seq(.01, 1, by = .01))
    1%     2%     3%     4%     5%     6%     7%     8%     9%    10%    11%    12%
13%
 55.10  58.20  61.30  62.72  63.65  64.58  65.17  65.48  65.79  66.00  66.00  66.00
66.75
   14%    15%    16%    17%    18%    19%    20%    21%    22%    23%    24%    25%
26%
 74.50  82.25  90.00  91.54  92.16  92.78  93.40  94.02  94.64  95.26  95.88  96.50
97.48
   27%    28%    29%    30%    31%    32%    33%    34%    35%    36%    37%    38%
39%
 99.96 102.44 104.92 106.20 107.44 108.68 109.23 109.54 109.85 110.00 110.00 110.00
110.00
   40%    41%    42%    43%    44%    45%    46%    47%    48%    49%    50%    51%
52%
110.00 110.00 110.06 110.99 111.92 112.85 115.60 118.70 121.80 123.00 123.00 123.00
126.24
   53%    54%    55%    56%    57%    58%    59%    60%    61%    62%    63%    64%
65%
134.61 142.98 150.00 150.00 150.00 150.00 157.25 165.00 172.75 175.00 175.00 175.00
175.00
   66%    67%    68%    69%    70%    71%    72%    73%    74%    75%    76%    77%
78%
175.00 175.00 175.40 176.95 178.50 180.00 180.00 180.00 180.00 180.00 180.00 180.00
184.50
   79%    80%    81%    82%    83%    84%    85%    86%    87%    88%    89%    90%
91%
192.25 200.00 206.10 209.20 212.30 215.60 220.25 224.90 229.55 234.20 238.85 243.50
245.00
   92%    93%    94%    95%    96%    97%    98%    99%   100%
245.00 245.00 247.66 253.55 259.44 268.97 290.98 312.99 335.00
```

Code:

```
# 2. measures of dispersion
# range
```

```r
max(data$hp) - min(data$hp)
# inter quantile range
quantile(data$hp, 0.75) - quantile(data$hp, 0.25)
# inter decile range
quantile(data$hp, probs = seq(.1, 1, by = .1))["90%"] - quantile(data$hp, probs =
seq(.1, 1, by = .1))["10%"]
# mean absolute deviation
mad(data$hp)
# standard deviation
sd(data$hp)
# skewness
skewness(data$hp)
# kurtosis
kurtosis(data$hp)
```

Output:

```
> # 2. measures of dispersion
> # range
> max(data$hp) - min(data$hp)
[1] 283
> # inter quantile range
> quantile(data$hp, 0.75) - quantile(data$hp, 0.25)
 75%
83.5
> # inter decile range
> quantile(data$hp, probs = seq(.1, 1, by = .1))["90%"] - quantile(data$hp, probs =
seq(.1, 1, by = .1))["10%"]
  90%
177.5
> # mean absolute deviation
> mad(data$hp)
[1] 77.0952
> # standard deviation
> sd(data$hp)
[1] 68.56287
> # skewness
> skewness(data$hp)
[1] 0.7614356
> # kurtosis
> kurtosis(data$hp)
[1] 3.052233
```

Code:

```r
# 3. frequency distributions with plots
# frequency distribution
table(cut(mtcars$hp, breaks = 5))
# histogram
hist(data$hp)
# relative frequency distribution
prop.table(table(cut(mtcars$hp, breaks = 5)))
# cumulative frequency distribution
```

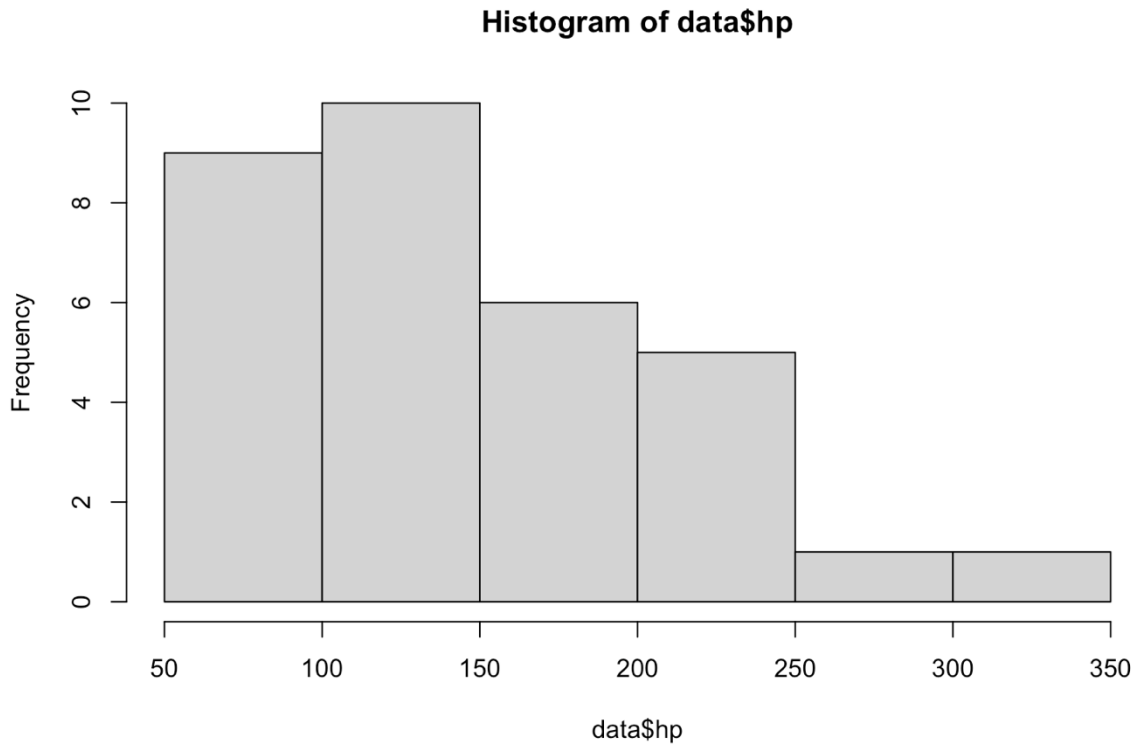
```
cumsum(table(cut(data$hp, breaks = 5)))
```

Output:

```
> # 3. frequency distributions with plots
> # frequency distribution
> table(cut(mtcars$hp, breaks = 5))

(51.7,109]  (109,165]  (165,222]  (222,278]  (278,335]
        10          9          8          4          1
> # histogram
> hist(data$hp)
> # relative frequency distribution
> prop.table(table(cut(mtcars$hp, breaks = 5)))

(51.7,109]  (109,165]  (165,222]  (222,278]  (278,335]
   0.31250    0.28125    0.25000    0.12500    0.03125
> # cumulative frequency distribution
> cumsum(table(cut(data$hp, breaks = 5)))
(51.7,109]  (109,165]  (165,222]  (222,278]  (278,335]
        10         19         27         31         32
```



Code:

```
# 4. plots from the categorical variable
# pie chart
pie(table(data$cyl))
# bar plot
barplot(table(mtcars$cyl))
```

Output: