DMI: Statistics of One Variable

MDM4U

Jin Hyung Park

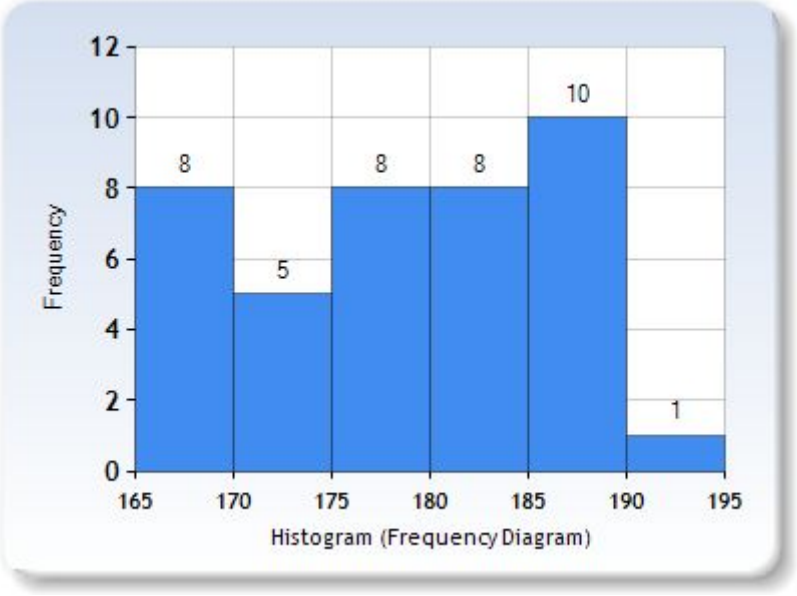
Height

Arranged Data

165 165 166 166 167 167 168 169 170 171 171 172 173 175 176 177 177 177 177 179 179 180 181 182 182 183 184 184 184 185 185 185 187 187 188 188 188 189 189 190

Histogram

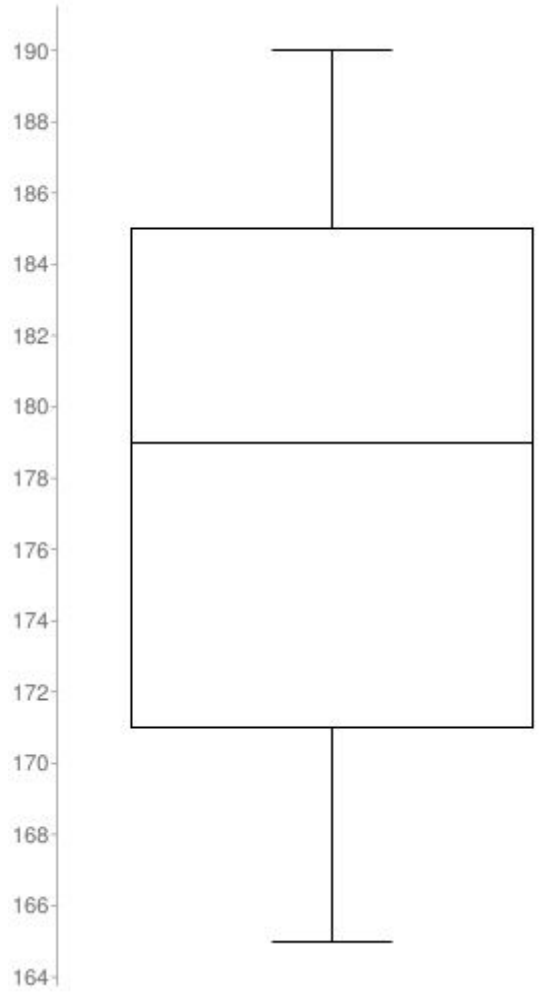
Frequency Table	
Class	**Count**
165-169	8
170-174	5
175-179	8
180-184	8
185-189	10
190-195	1



Box and Whisker Plot

- Minimum: 165
- Lower Quartile: 171
- Median: 179

- Upper Quartile: 184.5
- Maximum: 190
- Mean: 178.2



Central Tendencies:

- Mean:

$$\frac{\Sigma(data)}{40}$$
$$= \frac{7128}{40}$$
$$= 178.2$$
$$\therefore 178.2\text{cm}$$

- Mode:

Height	Frequency
165	2
166	2

167	2
168	1
169	1
170	1
171	2
172	1
173	1
174	0
175	1
176	1
177	4
178	0
179	2
180	1
181	1
182	2
183	1
184	3
185	3
186	0
187	2
188	3
189	2
190	1

Mode: 177

- Median:
  179cm

Dispersion & Distribution

- Range:
  $190 - 165$
  $= 25$
  $\therefore 25$
- Lower Quartile:
  $\frac{1}{4} \times (40 + 1)th\ value$
  $= 10.25^{th}\ value\ (which\ is\ on\ the\ 11^{th}\ value)$
  $= 171$
  $\therefore 171cm$
- Upper Quartile:
  $\frac{3}{4} \times (40 + 1)th\ value$
  $= 30.75^{th}\ value\ (about\ 31^{th}\ value)$
  $= 185$
  $\therefore 185cm$
- Interquartile Range:
  $185 - 171$
  $= 14$
  $\therefore 14$
- Standard Deviation:

x	x-mean		(x-mean)^2
189	189-178.2	10.8	116.64
167	167-178.2	-11.2	125.44
181	181-178.2	2.8	7.84
185	185-178.2	6.8	46.24
179	179-178.2	0.8	0.64
185	185-178.2	6.8	46.24
169	169-178.2	-9.2	84.64
166	166-178.2	-12.2	148.84
182	182-178.2	3.8	14.44
172	172-178.2	-6.2	38.44
171	171-178.2	-7.2	51.84
179	179-178.2	0.8	0.64
184	184-178.2	5.8	33.64
177	177-178.2	-1.2	1.44
187	187-178.2	8.8	77.44
180	180-178.2	1.8	3.24
177	177-178.2	-1.2	1.44

183	183-178.2	4.8	23.04
166	166-178.2	-12.2	148.84
188	188-178.2	9.8	96.04
177	177-178.2	-1.2	1.44
187	187-178.2	8.8	77.44
165	165-178.2	-13.2	174.24
184	184-178.2	5.8	33.64
170	170-178.2	-8.2	67.24
171	171-178.2	-7.2	51.84
168	168-178.2	-10.2	104.04
165	165-178.2	-13.2	174.24
177	177-178.2	-1.2	1.44
184	184-178.2	5.8	33.64
185	185-178.2	6.8	46.24
176	176-178.2	-2.2	4.84
175	175-178.2	-3.2	10.24
167	167-178.2	-11.2	125.44
190	190-178.2	11.8	139.24
188	188-178.2	9.8	96.04
173	173-178.2	-5.2	27.04
182	182-178.2	3.8	14.44
189	189-178.2	10.8	116.64
188	188-178.2	9.8	96.04

$$\sigma = \sqrt{\frac{\sum(x-mean)^2}{40}}$$

$$\sigma = \sqrt{\frac{2462.4}{40}}$$

$\therefore \sigma \fallingdotseq 7.84$

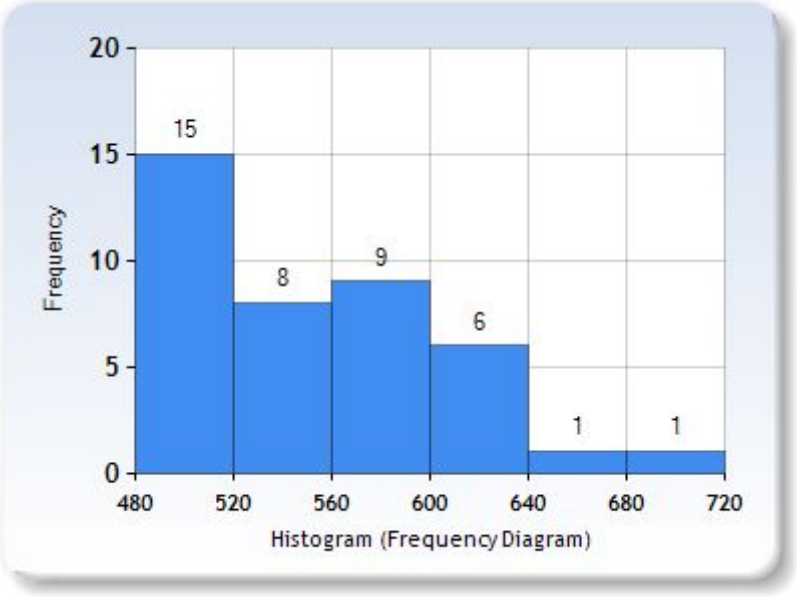
Running Time(secs)

Arranged Data

480,484,484,484,489,493,499,499,502,503,507,508,511,517,519,520,524,534,536,544,546,547,556,565,568,570,572,574,580,583,583,586,608,612,621,625,632,635,647,704

Histogram

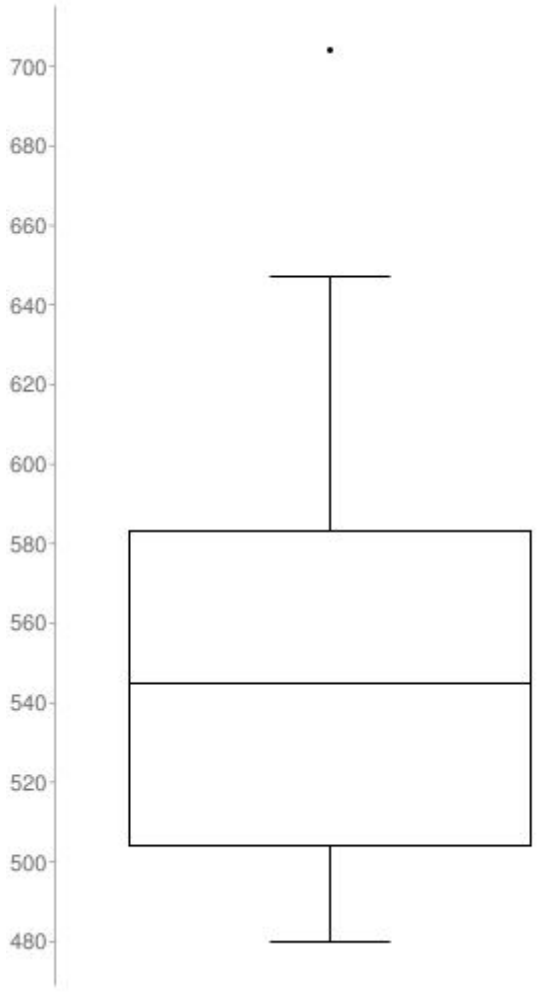
Frequency Table	
Class	**Count**
480-519	15
520-559	8
560-599	9
600-639	6
640-679	1
680-720	1



Box and Whisker Plot

- Minimum: 480
- Lower Quartile: 505
- Median: 545
- Upper Quartile: 583

- Maximum: 704



Central Tendencies

- Mean:

  $\frac{\Sigma(data)}{40}$

  $= \frac{22051}{40}$

  $= 551.275$

  $\therefore$ about 551
- Mode:

Running Time(secs)	Frequency
480	1
484	3
489	1
493	1
499	2
502	1

503	1
507	1
508	1
511	1
517	1
519	1
520	1
524	1
534	1
536	1
544	1
546	1
547	1
556	1
565	1
568	1
570	1
572	1
574	1
580	1
583	2
586	1
608	1
612	1
621	1
625	1
632	1
635	1
647	1
704	1

Mode: 484

- Median:
  $\frac{544+546}{2} = \frac{1090}{2} = 545$
  545

Dispersion & Distribution

- Range:

  $704 - 480$

  $= 224$

  $\therefore 224$

- Lower Quartile:

  $\frac{1}{4} \times (40 + 1)th\ value$

  $= 10.25^{th}\ value\ (between\ 10^{th}\ value\ and\ 11^{th}\ value)$

  $= \frac{503+507}{2}$

  $\therefore 505\ secs$

- Upper Quartile:

  $\frac{3}{4} \times (40 + 1)th\ value$

  $= 30.75^{th}\ value\ (about\ 31^{th}\ value)$

  $= 583$

  $\therefore 583\ secs$

- Interquartile Range:

  $583 - 505$

  $= 78$

  $\therefore 78$

- Standard Deviation:

x	x-mean		(x-mean)^2
493	493-551.275	-58.275	3395.975625
704	704-551.275	152.725	23324.92563
517	517-551.275	-34.275	1174.775625
507	507-551.275	-44.275	1960.275625
574	574-551.275	22.725	516.425625
502	502-551.275	-49.275	2428.025625
647	647-551.275	95.725	9163.275625
621	621-551.275	69.725	4861.575625
499	499-551.275	-52.275	2732.675625
544	544-551.275	-7.275	52.925625
572	572-551.275	20.725	429.525625
565	565-551.275	13.725	188.375625
508	508-551.275	-43.275	1872.725625
583	583-551.275	31.725	1006.475625
484	484-551.275	-67.275	4525.925625
489	489-551.275	-62.275	3878.175625
556	556-551.275	4.725	22.325625

534	534-551.275	-17.275	298.425625
612	612-551.275	60.725	3687.525625
503	503-551.275	-48.275	2330.475625
547	547-551.275	-4.275	18.275625
524	524-551.275	-27.275	743.925625
608	608-551.275	56.725	3217.725625
484	484-551.275	-67.275	4525.925625
583	583-551.275	31.725	1006.475625
586	586-551.275	34.725	1205.825625
632	632-551.275	80.725	6516.525625
635	635-551.275	83.725	7009.875625
570	570-551.275	18.725	350.625625
511	511-551.275	-40.275	1622.075625
484	484-551.275	-67.275	4525.925625
546	546-551.275	-5.275	27.825625
580	580-551.275	28.725	825.125625
625	625-551.275	73.725	5435.375625
520	520-551.275	-31.275	978.125625
536	536-551.275	-15.275	233.325625
568	568-551.275	16.725	279.725625
480	480-551.275	-71.275	5080.125625
519	519-551.275	-32.275	1041.675625
499	499-551.275	-52.275	2732.675625

$$\sigma = \sqrt{\frac{\sum(x-mean)^2}{40}}$$

$$\sigma = \sqrt{\frac{115227.975}{40}}$$

$$\therefore \sigma \fallingdotseq 53.67$$

Conclusion

Both histograms show a distribution that is skewed to the left. Even though most of the height data has most of the data points near the mean, there is a large number of individuals in the lowest bracket. Interestingly, there is a similar leftward skew in the running time data, although the expected relationship should be an inverse relationship. The standard deviations of the datasets display relatively well distributed data sets without skewing toward the middle or too flat of a distribution.

It isn't possible to conclude on a relationship between the two variables from a one variable analysis. However, in the last part of the assignment, I was able to perform a regression

analysis to determine a linear inverse relationship between the two variables, as discussed in the Two Variable Analysis assignment.

Limitations and Improvements

Improvements could have been made to the collection of data. Firstly, the dataset could have included more individuals to make sure that a better distribution with skewing could be formed. It is undoubtedly true that the height of males in the population of humans show a very normally distributed pattern. Therefore, more data points would have made the height data more accurate. In addition, the age distribution of the individuals are too narrow because of the data collection being performed on military servicemen. As such, there will be some limitations on how much the data can be used to infer information about the relationship between height and running time across all age groups of men.