MA5701: Statistical Methods

Chapter 1 : Data and Statistics

Kui Zhang, Mathematical Sciences

Exercise – Tree Data

- The discipline of forest science is a frequent user of statistics. An important activity is to gather data on the physical characteristics of a random sample of trees in a forest. The resulting data may be used to estimate the potential yield of the forest, to obtain information on the genetic composition of a particular species, or to investigate the effect of environmental conditions. The following data set consists of measurements of three characteristics of 64 sample trees of a particular species.

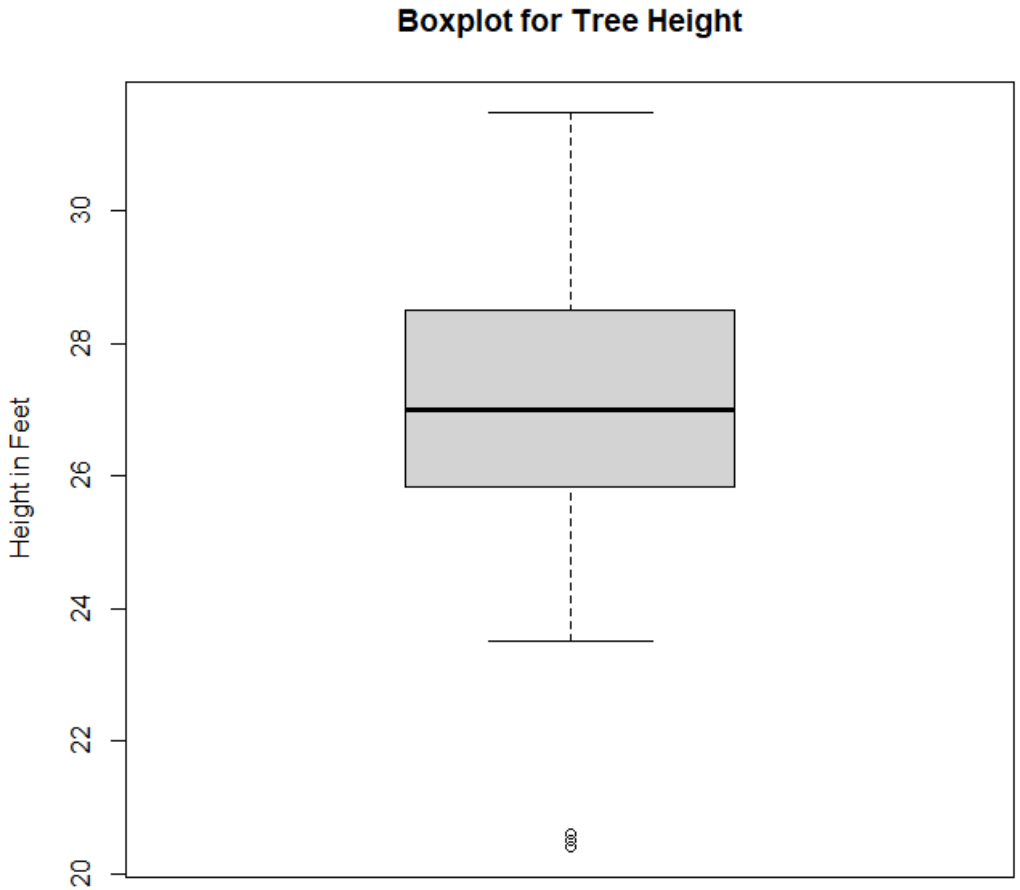
Exercise – Tree Data

- The data look like this:

obs	dfoot	hcrn	ht	obs	dfoot	hcrn	Ht
1	4.1	1.5	24.5	23	4.3	2.0	25.6
2	3.4	4.7	25.0	24	2.7	3.0	20.4
3	4.4	2.8	29.0	25	4.3	2.0	25.0

- **dfoot:** the diameter of the tree at one foot above ground level, measured in inches
- **hcrn:** the height to the base of the crown measured in feet
- **ht:** the total height of the tree measured in feet

Boxplot for Height – Tree Data



Exercise - Boxplot for Height from Tree Data

- Mean: 26.96
- Median: 27.00
- Maximum (largest) observation: 31.50
- Minimum (smallest) Observation: 20.40
- First Quartile (Q_1): 25.875
- Third Quartile (Q_3): 28.50

Exercise - Boxplot for Height from Tree Data

- Interquartile Range?
- Step?
- Upper Inner Fence (UIF)?
- Lower Inner Fence (LIF)?
- Upper Outer Fence (UOF)?
- Lower Outer Fence (LOF)?
- Are smallest or largest observations here outliers?

Exercise - Boxplot for Height from Tree Data

- **Interquartile Range:** $28.50 - 25.875 = 2.625$
- Step: $2.625 * 1.5 = 3.9375$
- **Upper Inner Fence (UIF):** $28.50 + 3.9375 = 32.4375$
- **Lower Inner Fence (LIF):** $25.875 - 3.9375 = 21.9375$
- **Upper Outer Fence (UOF):** $28.50 + 2 * 3.9375 = 36.375$
- **Lower Outer Fence (LOF):** $25.875 - 2 * 3.9375 = 18.0$
- **Smallest observation:** 20.40 < LIF but > LOF, mild outlier
- **Largest observation:** 31.50 < UIF, not outlier