Discussing Distributions

A **distribution** is a list of all of the possible outcomes of a random variable along with either their corresponding frequency or probability values (we would speak of the **frequency distribution** or **probability distribution** respectively).

The distribution gives insight into **how likely** or **how common** the various outcomes are.

Measures of Centre

The **mean** is the sum of all values divided by the number of values.

The **median** is the ‘middle’ value of the sorted data set.

The **mode** is the most likely value in the data set.

The **modality** of a distribution is the number of modes (peaks).

The **skewness** of a distribution is a measure of how asymmetric it is (the tendency to be distorted to the left or right.)

Measures of Spread

The **range** is the difference between the smallest and largest values. (max - min)

The **interquartile range** is the range between the first and third quartiles (Q3 - Q1).

A **quartile** is a point in a distribution where a multiple of a quarter of the distribution lies above and below that point.

The **variance** and **standard deviation** of a distribution are measures of how much each value differs from the mean.

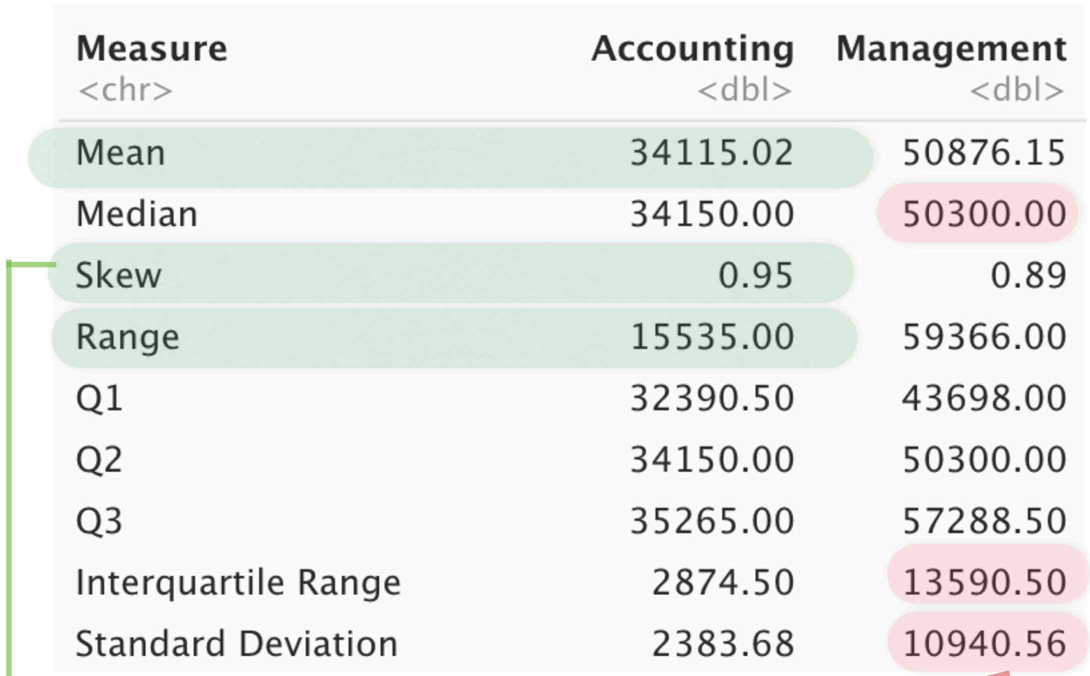
Case Study: Tyrell Company Salaries

Figure 1 (bottom left): salary histogram (unfaceted)

Figure 2 (bottom right): box-plots. Split by job position.

Figure 3 (top right): summary table.

Figure 3: summary table

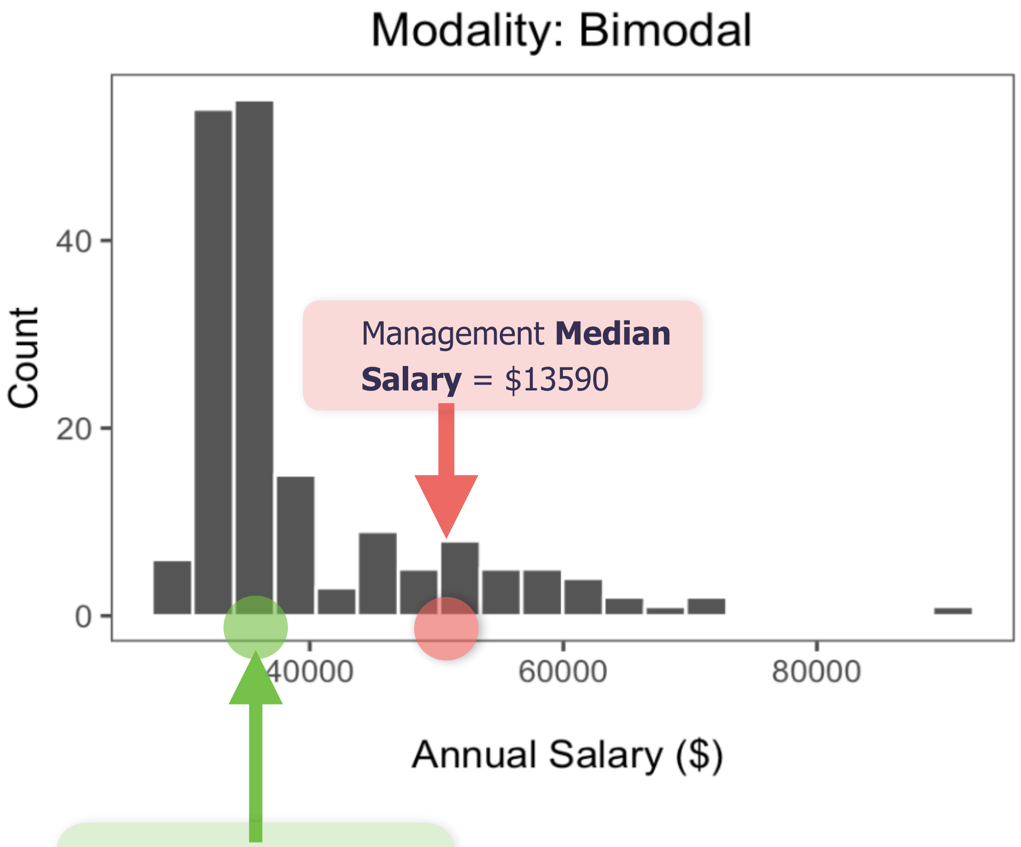


Measure <chr>	Accounting <dbl>	Management <dbl>
Mean	34115.02	50876.15
Median	34150.00	50300.00
Skew	0.95	0.89
Range	15535.00	59366.00
Q1	32390.50	43698.00
Q2	34150.00	50300.00
Q3	35265.00	57288.50
Interquartile Range	2874.50	13590.50
Standard Deviation	2383.68	10940.56

Accounting **Skew** = 0.95 ~ Right-Skewed

Management **Standard Deviation** = $10940 ~ each point deviates from the mean ($50300) by an average of $10940

Figure 1: salary histogram

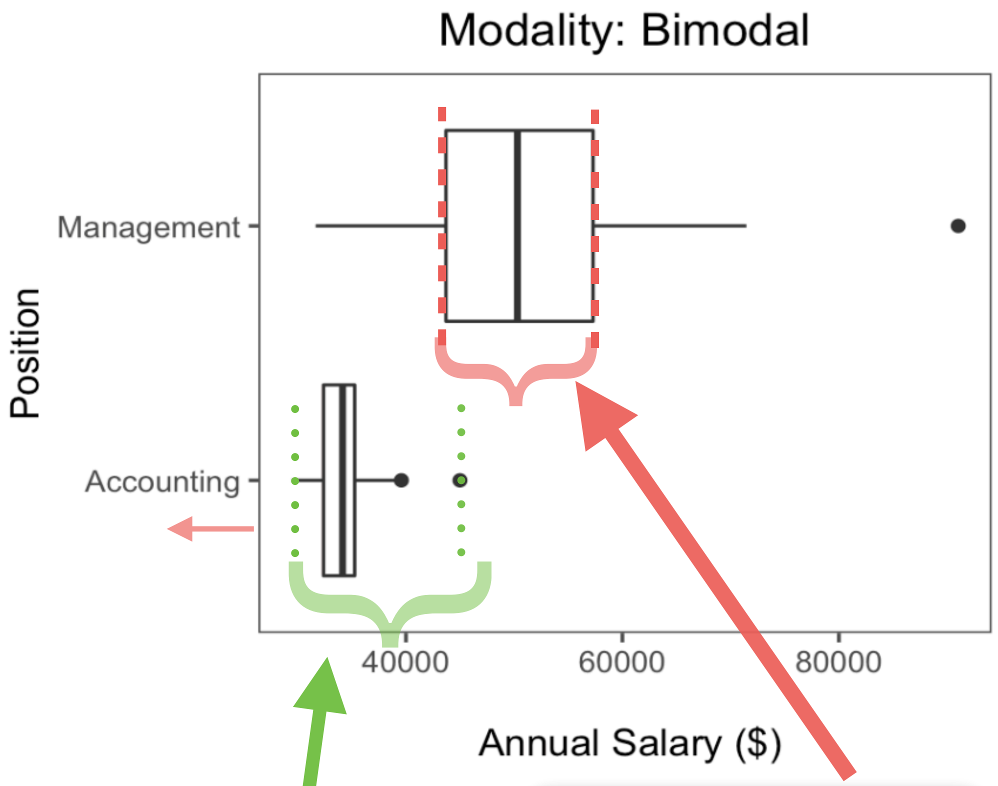


Modality: Bimodal

Management **Median Salary** = $13590

Accounting **Mean Salary** = $34115.02

Figure 2: box plots



Modality: Bimodal

Management **IQR** = $13590

Accounting **Range** = $15535