**91582:  Use statistical methods to make a formal inference (4 credits)**

***Key components of the statistical enquiry cycle for making a formal inference:***

* posing a comparison investigative question using a given multivariate data set
* selecting and using appropriate displays and summary statistics
* discussing sample distributions
* discussing sampling variability, including the variability of estimates
* making an appropriate formal statistical inference
* communicating findings in a conclusion.
1. **Example comparison investigative questions:**
* What is the difference between the median blood alcohol levels of the drivers in severe crashes and those in minor crashes in New Zealand, where alcohol was deemed to be the cause of the crash?
1. **Select and use appropriate displays and summary statistics:**





1. **Discuss sample distributions:**

**Central tendency**

* The median value for the blood alcohol level for drivers in serious crashes is 160mg/100ml, which is 25mg higher than for minor crashes which is 135 mg/100ml.
* There appear to be two modes evident in the severe crash sample, at around 160 and 190mg, however these are more likely to a peculiarity in the sample rather than a feature we would see in the population.

**Spread (dot plot, range, inter-quartile range)**

* From the dot plot it is evident that there is greater spread in the blood alcohol level of the serious crash data which has a range of 289mg/100ml, compared with a range of 182 mg/100ml.
* This greater spread is also evident in the middle 50% of the data as the interquartile range for the serious crash sample is 65.8mg/100ml compared with only 23mg/100ml for the minor crash sample.

**Shift / Overlap**

* As can be seen in the box plot there is a clear shift to the right of the middle 50% of the data for the serious crashes, when compared with the middle 50% of the data for the minor crashes.
* The median for the sample of severe crashes is higher than the upper quartile for the sample of minor crashes.
* There is, however, quite a bit of overlap in the interquartile ranges, with most of the middle 50% of the minor crash data sitting in the lower-middle 25% of the serious crash interquartile range.

**Unusual values**

* The severe crashes have some very high data points for blood alcohol level, with three values over 250mg/100ml.