Ministry of Transport
TE MANATŪ WAKA

New Zealand Crash Statistics

Mathematics and Statistics 91582 (3.10) version 1**:**
Use statistical methods to make a formal inference **Credits:** 4

Teacher guidelines

Context/setting

This activity requires students to carry out an investigation that uses statistical methods to make a formal inference about New Zealand Crash statistics.

This task has been motivated by recent successful media campaigns by the New Zealand Transport Association (for example http://www.nzta.govt.nz/about/advertising/drink-driving/legend.html). It would be appropriate to share one or two of these advertisements with the students to introduce them to the context. The multivariate dataset for the investigation is provided to the students. It is expected that students research background information into factors related to serious and minor injury crashes in New Zealand to give **purpose** and **relevancy** to their investigation. Major aspects of this research should be referenced appropriately.

This activity can be adapted to use another existing dataset. Any dataset provided needs to have appropriate motivation, contextual depth and relevance for the students. Details about the data collection methods need to be provided to enable students to inform themselves about the context and populations.

Resource requirements

The data set should be made available electronically for students to use for the assessment. The descriptions of the variables are given in the Student Instruction Sheet.

Ministry of Transport
TE MANATŪ WAKA

New Zealand Crash Statistics

Mathematics and Statistics 91582 (3.10) version 1**:**
Use statistical methods to make a formal inference **Credits:** 4

Student instructions

Achievement	Achievement with Merit	Achievement with Excellence
Use statistical methods to make a formal inference.	Use statistical methods to make a formal inference, with justification.	Use statistical methods to make a formal inference, with statistical insight.

Assessment Conditions

DUE DATE: Wednesday 27 June 2.30pm

You have one week to complete this assessment. It is expected that you will complete the majority of this report outside of class. You will need to complete background research and integrate this informed contextual knowledge throughout your report.

Checkpoint 1 – Monday 25 June

Please bring your assessment to class on Monday 25 June. Your teacher will check your progress at this stage. You will have access to the computer lab this period, or you are welcome to bring your own laptop to work on.

Submission Day – Wednesday 27 June

You will have access to the computer lab this period, or you are welcome to bring your own laptop to work on. By the end of the period, you must have submitted your work electronically, and provided a paper copy of your assessment.

Authenticity of student work

All submitted work must be your own work. You are permitted to discuss your work with others, use your notes and examples from class, and research background into the context as appropriate. (NOTE: All relevant research must be referenced.) You **MUST** complete your written report on **your own**.

Introduction

Following on from the successful "Legend" media campaign*, the Ministry of Transport commissioned a study of drivers from the age groups that have the highest risk of crashing and the lowest risk of crashing.

In 2011 there were 1409 serious or minor crashes where alcohol or drugs was recorded as a factor. A random sample was taken from these drivers and they were interviewed in person by researchers. Details of the information collected by these researchers are given below.

Use this data to conduct a statistical investigation and write a report on New Zealand crash statistics.

*http://www.nzta.govt.nz/about/advertising/drink-driving/legend.html

Variable definitions:

Variable	Description
Gender	• Male • Female
Age	Age in years at time of crash
Risk group	Highest risk age group 15 – 24 years (H) Lowest risk age group 50 – 59 years (L)
Licence type	Type of licence held at time of crash • Learners (L) • Restricted (R) • Full (F)
Crash severity	• Minor injury crash (M) • Serious injury crash (S)
Blood Alcohol Level	Recorded blood alcohol level recorded at time of crash *(In milligrams of alcohol per 100 millilitres of blood. Breath test results have been converted to a blood equivalent)*
Distance driven	Estimated distance driven in the last week *(in kilometres)*
Distance from home	Estimated distance away from home when crash occurred *(in kilometres)*
Vehicle age	Age of vehicle involved in crash
Insurance payout	Value of insurance claim
Medical expenses	Estimated medical costs as a result of the crash *(total costs to date)*
Time off work	Estimated number of days off work as a result of the crash (if applicable)

Your task

Problem 	• State the purpose of your investigation. **Remember to explain why you are investigating this situation; give a brief background and hypothesis to be investigated. You are expected to do background research to understand the situation.** • Identify the variables you wish to investigate. • Pose an appropriate comparison investigative question.
Analysis 	• Select and use appropriate displays and statistical measures. • Discuss and compare sample distributions
Conclusion 	• Make a formal statistical inference. • Conclude your investigation, reflecting on your hypothesis and justifying your formal inference **This may include:** - **Discussing sampling variability, including the variability of estimates.** - **Reflecting on the process you have used to make the formal inference**

Remember: The quality of thinking demonstrated in your report and your ability to link the context and populations to the different components of the statistical enquiry cycle will determine your overall grade. As you write your report, take care to link your discussion to the context and support your statements by referring to statistical evidence.

Ministry of Transport
TE MANATŪ WAKA

New Zealand Crash Statistics

Mathematics and Statistics 91582 (3.10) version 1**:**

Use statistical methods to make a formal inference **Credits:** 4

Final grades will be decided using professional judgement based on a holistic examination of the evidence provided against the criteria in the AS..

	Achieve Use statistical methods to make a formal inference	**Merit** Use statistical methods to make a formal inference, with justification	**Excellence** Use statistical methods to make a formal inference, with statistical insight
	The student shows evidence of using each component of the statistical enquiry cycle to make an inference Examples of using each component of the statistical enquiry cycle appear below	The student makes an inference, showing evidence of linking components of the statistical enquiry cycle to the context, and/or populations and referring to evidence in support of statements made. Examples of linking components of the statistical enquiry cycle to the context and/or populations and referring to evidence such as sample statistics, data values, or features of visual displays in support of statements made appear below	The student makes an inference with statistical insight, showing evidence of integrating statistical and contextual knowledge throughout the statistical enquiry cycle. They may reflect on the process or consider other explanations. Examples of integrating statistical and contextual knowledge throughout the statistical enquiry cycle appear below
Judgement	***Judgement**: The report covers all the points below to the depth indicated by examples. Minor errors or omissions should not withhold awarding of the grade.*	***Judgement**: The report covers all the points below to the depth indicated by examples. Understanding of key concepts (difference between sample statistics and population parameters; sampling variability; interpretation of the bootstrap confidence interval) should be evident. Minor errors or omissions should not withhold awarding of the grade.*	***Judgement**: The report covers all the points below to the depth indicated by examples. Understanding of key concepts (difference between sample statistics and population parameters; sampling variability, interpretation of the bootstrap confidence interval) should be evident. The student has reflected on the process or has given explanations by considering, in context, the effect of aspects such as sample representativeness on the estimate. They have discussed aspects of the investigation in context, such as re-examining the data to show a different aspect, or justify a point being made. Minor errors or omissions should not withhold awarding of the grade.*

Problem	Posing a comparison investigation question using a given multivariate data that clearly identifies the comparison and the population(s) *Question should be about the difference in medians or means*	Posing a comparison investigation question using a given multivariate data set including giving an explanation for the choice of variables for their investigation to pose an appropriate question and a hypothesis about what they may find *Question should be about the difference in medians or means*	Posing a comparison investigation question using a given multivariate data set including using research to develop the purpose for their investigation and using this contextual knowledge to pose a question. The purpose should include a hypothesis and justification of the hypothesis *Question should be about the difference in medians or means*

Analysis	Selecting and using appropriate displays and summary statistics. This includes producing dot plots, box plots, summary statistics including difference between sample medians or means and constructing a bootstrapped confidence interval Discussing and comparing sample distributions by identifying and comparing features of them. This could involve comparing the shift/centre, spread, shape and unusual features, using features of the displays and the summary statistics.	Selecting and using appropriate displays and summary statistics. This includes producing dot plots, box plots, summary statistics including difference between sample medians or means and constructing a bootstrapped confidence interval Discussing and comparing sample distributions by identifying and comparing features of them. This will involve comparing the shift/centre, spread, shape and unusual features with reference to features of displays and summary statistics **AND** links to context or investigative question.	Selecting and using appropriate displays and summary statistics. This includes producing dot plots, box plots, summary statistics including difference between sample medians or means and constructing a bootstrapped confidence interval Discussing and comparing sample distributions by seeking explanations for features of the data identified in displays or summary statistics and considering the impact of these on the context or investigative question.
Conclusion	Making an appropriate formal statistical inference: using the resampling (bootstrapping) method to construct a confidence interval. Communicating findings clearly in a conclusion using the formal inference to answer the investigative question.	Making an appropriate formal statistical inference. For example: using the resampling (bootstrapping) method to construct a confidence interval. Communicating findings clearly in a conclusion using the formal inference to answer the investigative question, justifying the call and making links to the context.	Making an appropriate formal statistical inference. For example: using the resampling (bootstrapping) method to construct a confidence interval. Communicating findings in a conclusion using the formal inference to answer the investigative question in context. This includes justifying the call **AND** linking back to the purpose of the investigation; considering other explanations for the findings; reflecting on the process (which may include discussion on sampling variability).

NOTES:

- Students are given a 'motivation' into the situation, but not too much of a direction in terms of hypothesis. It is expected that they then research themselves and develop their hypothesis and investigative question
- This research into the context will allow them to reflect on what they see with further depth
- There should be an obvious connection between a student's question, analysis and conclusion; they should have answered their question
- There should be population/sample understanding evident at all levels
- Understanding of sampling variability is implied through use of a confidence interval for population parameter.

HIGH NOT ACHIEVED – BECKY

Marker comments

PROBLEM – not about looking at the difference in medians between the two groups; some motivation to purpose of investigation given, but not directly relevant

ANALYSIS – Difference in sample medians not given; groups, variable and units used throughout. Comments are minimally descriptive only – no attempt at saying what these observations mean.

CONCLUSION – CI interpreted (in Analysis section) but population unclear and population parameter not mentioned, incorrect call, no justification

Low level of analysis comments along with incorrect means high NA rather than low A.

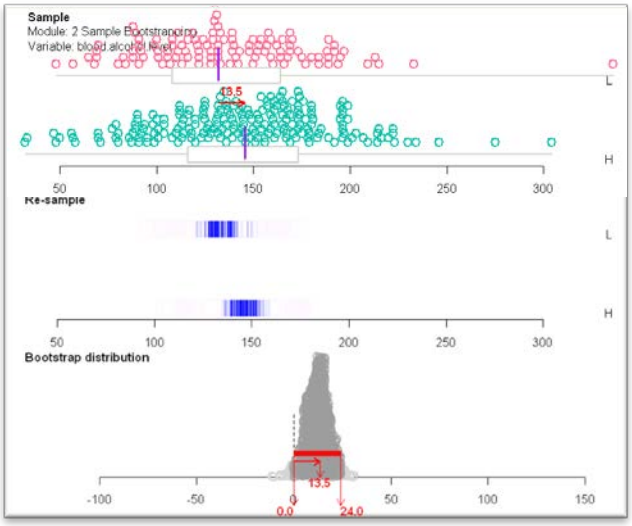
Problem

Do the drivers in the highest risk age group (15 – 24 years) tend to have a higher blood alcohol level at the time of crash than drivers in the lowest risk age group (50 – 59 years)?

I am asking this question there is a lot of argument about whether or not younger drivers are more dangerous on the roads than older drivers. I plan to find out, using real data, whether or not younger drivers have a higher blood alcohol level when driving (blood alcohol recorded at the time of crash). The driving age in 2011 up until the start of August was 15, but on the 1st of August, the driving age was put up to 16 years old, so there would have been drivers who were 15 years old even after the age was put up in 2011. The driving age in most other countries is 17 years.

As a learner driver in the high risk age group, this report relates to me, as it effects how other drivers see me on the road, as not every high risk driver is intoxicated when they drive, and those that are, may not be quite so different to the low risk age group. I think that not all drivers of the high risk age group are more dangerous than those of the low risk age group.

Analysis



From this bootstrap we can immediately see that the higher risk age group tends to have only a slightly higher blood alcohol level at the time of crash. The confidence interval shows that drivers of the higher risk age group tend to have a higher blood alcohol level at the time of crash, by somewhere between 0.0 and 24.0 milligrams of alcohol per 100 millilitres of blood.

This box and whisker/ dot plot shows that the blood alcohol level for both the low risk age group and the high risk age group have a similar spread, with one outlier in the low risk age group, with a blood alcohol level of well over 300 milligrams of alcohol per 100 millilitres of blood.

The initial sample median of the low risk age group is lower than that of the high risk age group. The dots of both the low and high risk groups are in the shape of a normal distribution bell curve, they are not skewed either way.

Looking at the box and whisker part of the graph, we can see that there is a large overlap between the high risk and low risk age groups, the upper quartile range of the high risk age group, only slightly higher than that of the low risk age group, and the lower quartile of the low risk age group, is only slightly lower than that of the high risk age group.

Conclusion

Each of the age groups have a range of 10 years from the lowest age to the highest age in that group. The high risk age group being 15 – 24 years, and the low risk group being 50 to 59 years. This factor leaves no room for bias. However, the number of data that we have for the high risk age (206) group is a lot higher than for the low risk age group (111). From this, one could assume that it is because the higher risk factor that there were more crashes that involved a driver of the high risk age group than that of the low risk age group in 2011. However, without having the entire population of data in front of us, this limitation means that we could not fairly make a call.

Another limitation of this data is that there is no data of the people who have died due to a car crash. We know this because the researchers interviewed the drivers in person to find out more data about their case. While this is rather morbid, it should be considered as a factor in this report.
I think that while there seems to be a slightly higher median of blood alcohol level of the drivers in the high risk age group, there is not enough of a difference between the two medians (low risk and high risk) to say that the younger drivers are more dangerous drivers (due to blood alcohol level at the time of crash).
From this report we can definitely say that not all drivers of the high risk age group are more dangerous on the roads than the drivers of the low risk age group.

HIGH ACHIEVE – JOSH

Marker comments

PROBLEM – Introduction well referenced and sourced. Question is about difference in medians but groups being compared not well defined; population not given.
ANALYSIS – Difference in sample medians given, descriptive comments in context and quantified, minimal reflection back to context and population. Unusual value acknowledged but no reasons given. Shape description minimal but thought about how this relates to population distribution and why (low level)
CONCLUSION – Good link to 25mg/100mL difference and relative risk from introduction but understanding of how that related directly to this sample unclear. CI interpreted correctly, but sample to population link unclear. *(Population link given in last paragraph.)* Call based on median-outside-box, then backed up with CI, justification with zero outside CI.
Sample → population link unclear in investigative question and CI interpretation; starting to make contextual links and thinking about the "so what" but comments not strong hence HIGH ACHIEVE *rather than* LOW MERIT.

Problem

Alcohol seems to be one of the largest causes of severe vehical accidents, this is because drinking impaires the brains ability to function. Particularly more in youth than adults. Drink driving can often be found in the media and how to tackle the situation is a very controversal matter.

As young people learning to drive and gaining access to alcohol, it is important for us to know how much more alcohol will affect us with our driving and to show others that they shouldn't drive under the influence of alcohol. We want to investigate the difference in median BAC and compare it to the severity of the accident.
I think that the more severe crashes will involve a higher driver median blood alcohol level for a just a few reasons:

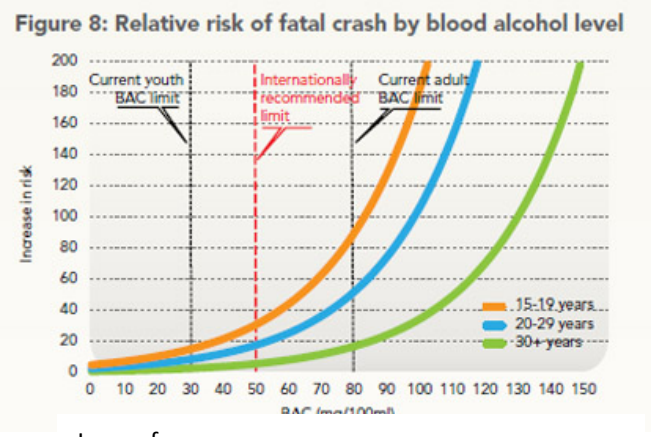


Image from:
www.transport.govt.nz/saferjourneys/Towardsasafesystem/Pages/Saferoaduse.aspx
(The graph needs updating because the youth BAC limit is now 0mg/100ml, BAC=Blood Alcohol Concentration)

- One New Zealand reseach paper stated "Miller has estimated from published studies that a blood alcohol concentration above 100 mg/100 mL increases the risk of involvement in any crash by 13 to 18 times and the risk of a fatal crash by as much as 50 to 90 times."
- The general fact that the more alcohol you consume, the greater the impact on your brain, such as increasing your reflex response time, delaying messages from your eyes to your brain and lack of balance.
    - Euphoria (BAC = 0.03% to 0.12%)
        - Overall improvement in mood and possible euphoria
        - Increased self-confidence
        - Shortened attention span
        - Impaired judgment
        - Impaired fine muscle coordination (http://en.wikipedia.org/wiki/Short-term_effects_of_alcohol)

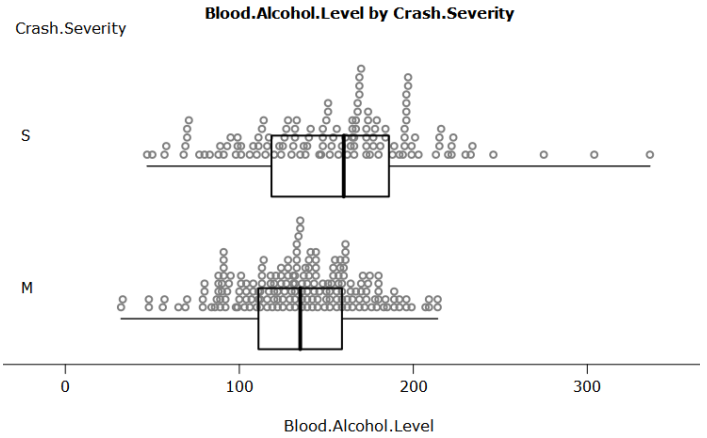
I would like to look at driver blood alcohol level and the severity of the crash. I wonder what the difference in median blood alcohol levels and severity of a car accident.

Analysis

Summary of Blood Alcohol Level by Crash Severity

	Min	1st Qu	Median	Mean	3rd Qu	Max	Std Dev	Sample Size
Minor	32	111.0	135	133.7	159	214	36.142	189
Severe	47	119.2	160	154.5	185	336	50.703	128

Overall comparison The blood alcohol levels for the severe crashes sample is more spread out than the minor crashes sample. Both samples have a lower range around 50mg/100mL to 100mg/100mL but only the severe sample goes past 225mg/100mL, and reaches over 300mg/100mL. The severe crashes sample has many values higher than the minor crashes sample.

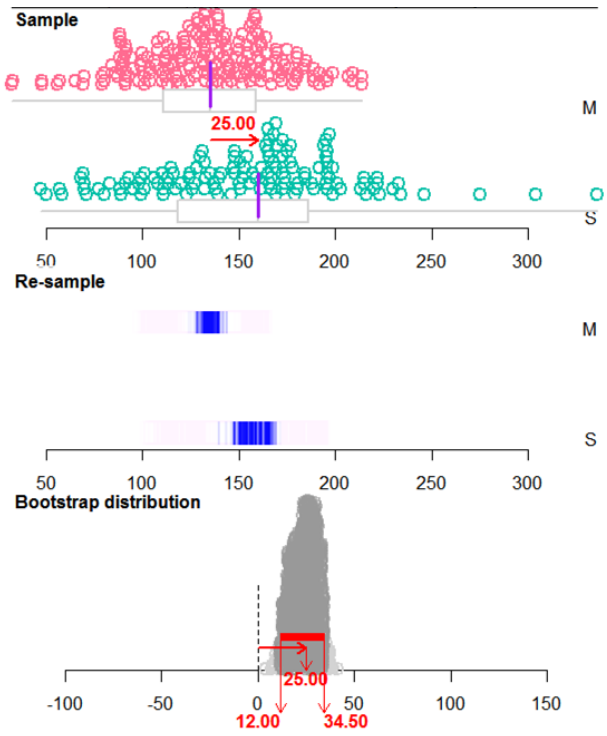


Centers The median for the minor crash blood alcohol level is 25mg/100mL lower than the median for severe crash blood alcohol, (135mg/L for minor compared to 160mg/l for severe).

Middle 50% - shift, overlap and spread. The middle 50% of the severe crash sample overlaps the middle 50% of the minor crash sample by a large amount, with the LQR of the severe crash sample well past the median of the minor crash sample. Both overlap each other's median, the median of the minor crashes in the LQR of the severe crashes and the median of the severe crashes in the UQR of the minor crashes. The middle 50% for the severe crash sample is more spread out than minor crash 50%, (IQR severe = 65.8mg/100mL, IQR minor = 48mg/100mL). This shows there is more variation in the severe crashes sample.

Unusual I think it is unusual to find people with blood alcohol levels as high as 300mg/100mL and greater. Otherwise it seems like normal data.

Shape The minor crash sample is left skewed while the severe crash data is more right skewed. Most of the data is between 100mg/100mL and 200mg/100mL. I think this would be similar back in the population because at different points of alcohol consumption you will become less and less able to drive.



Conclusion

From theses samples I can make the call that there is a visible difference in the blood alcohol levels for minor crashes and severe crashes, severe crashes having drivers with higher blood alcohol levels, but the two do overlap quite a lot so this difference isn't very large. Also, the graph showing the risk of fatal crashes by blood alcohol levels shows an exponential curve for all ages, meaning that the difference of 25mg/100mL between two points is much larger the further down the scale you go. From these samples I can see that the min blood alcohol levels are lower in the minor crashes than the min for the severe crashes, (minor=32mg/100mL, severe=47mg/100mL). I am pretty sure that the median blood alcohol level for severe crashes is between 12mg/100mL and 35mg/100mL more than the median blood alcohol level for minor crashes.

I can call from these samples that there is a difference in the severity of crashes related to the blood alcohol level of the driver because the median for the blood alcohol levels for severe crashes sample is about the same as the upper quartile of the minor crashes sample (severe median=154.9mg/100mL). Also reinforcing my call is the bootstrapping confidence interval which the difference in medians falls in and that it does not contain zero, which means that drivers with higher blood alcohol levels tend to cause more severe car accidents.
I am basing this conclusion on the bootstrap confidence interval I calculated, which involves re-sampling from my original sample of 317 alcohol influenced driver crashes. I am assuming that my original sample was eligible to represent the population of all accidents that were alcohol related. My sources state that there is a large difference caused by different alcohol blood levels, my calculations do show a similar trend although I would like to investigate other related aspects to this topic. My sample for severe crashes did have the top half past the upper quartile of the minor crashes sample as well.

Bibliography

- www.transport.govt.nz/saferjourneys/Towardsasafesystem/Pages/Saferoaduse.aspx
- http://202.68.89.83/NR/rdonlyres/EF63A419-E53D-41E3-9D28-6FCBED66507A/165595/49SCTIR_EVI_00DBHOH_BILL10329_1_A144412_ProfessorJ.pdf

---

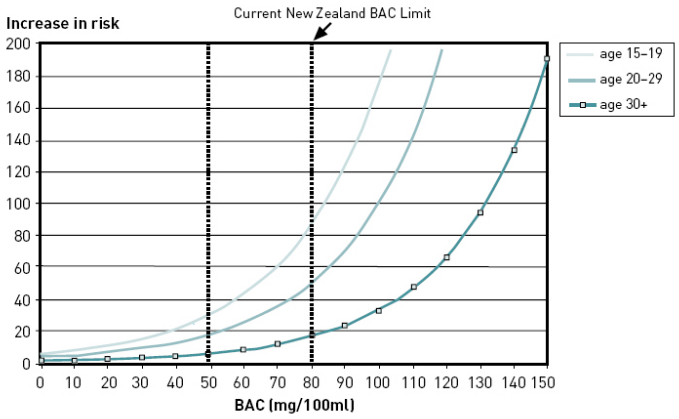
HIGH MERIT – LOGAN

Marker comments

PROBLEM – Risk graph given but unclear what risk is presented; hypothesis clear and justified; investigative question looking at difference in medians but population undefined (*given in hypothesis*).
ANALYSIS – Difference in sample medians given; descriptions in context and quantified; shape description also comparing mean and median; data re-categorised by legal BAL limit and described with link to population
CONCLUSION – CI interpreted but population parameter not given; correct call and justification; some thoughts about sample representativeness, but ideas not referenced.
Split of data a nice idea, but didn't really contribute much further depth to the analysis. Contextual links not strong, especially in analysis hence HIGH MERIT *rather than* LOW EXCELLENCE

Problem

Background Drink driving has become a largely publicized issue in New Zealand. Adverts showing the slogan 'Don't Drink and Drive' frequent our TV screens. Studies have shown that the risk of a person being involved in a crash is increased when they have higher blood alcohol content (BAC).



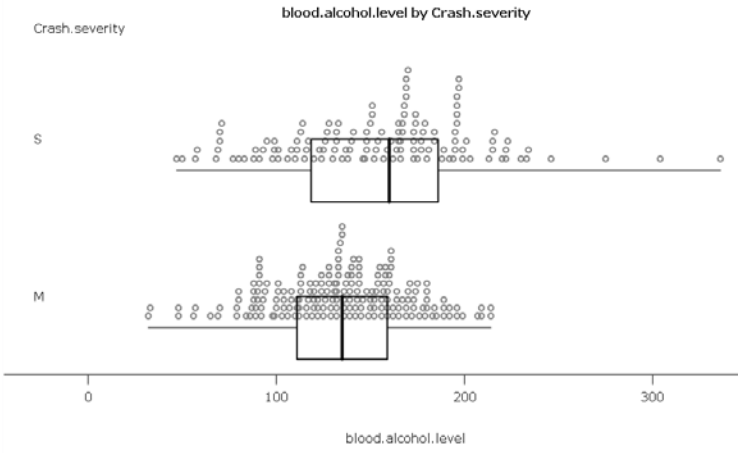
The graph is from the site, http://www.alac.org.nz/research-resources/nz-statistics/road-traffic-crashes-and-deaths. This is showing statistics from 2008 and clearly depicts a trend of increase in BAC meaning an increase in risk.
The Ministry of Transport site, http://www.transport.govt.nz/research/Documents/Alcohol-drug-crash-statistics-2011-(1).pdf, shows the same graph and also says that alcohol and/or drugs contributed to 121 fatal crashes, 398 crashes resulting in serious injury and 1011 crashes resulting in minor injuries in 2010.
Drink driving is an issue that affects us as it is a growing problem for teenagers over recent years. As many teenagers reach the age of being both allowed to drink and drive (separately), I want to investigate possible difference of severe and minor crashes due to blood alcohol levels.

Hypothesis I expect that severe crashes will have a higher median blood alcohol level than minor crashes in New Zealand because it has become well-known that alcohol impairs judgement and reflexes, also drink driving has been publicized again and again as something not to be done. Plus there have been numerous studies done on drink driving indicating that alcohol causes many crashes.

8

The investigation will look at the severity of crashes and blood alcohol level (milligrams per 100 milliliters of blood).

What is the difference in median of blood alcohol levels for severe and minor car crashes in 2011 where alcohol and/or drugs were a factor ?



Analysis

Overall Visual Comparison An initial look at the summary graph shows that the severe crashes appear to have a wider spread of blood alcohol levels than the minor crashes. The minor crashes seem to have a larger grouping of blood alcohol levels between 100 milligrams per 100 millilitres and 150 milligrams per 150 millilitres. The severe crashes have the highest blood alcohol levels in this sample.

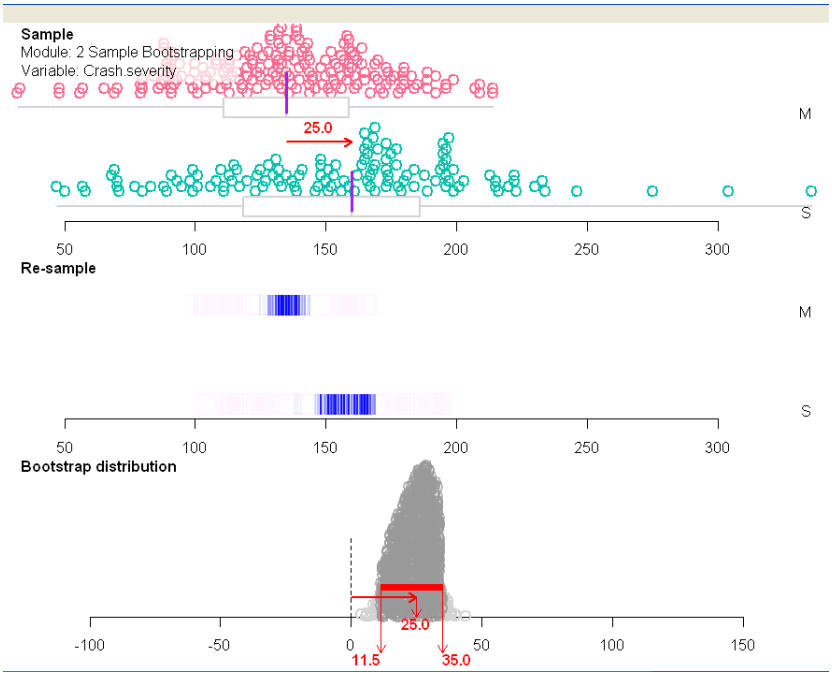
Summary of blood alcohol level by crash severity

	Minimum	1st Quarter	Median	Mean	3rd Quarter	Maximum	Standard Deviation	Sample Size
Minor	32	111	135	133.7	159	214	36.142	189
Severe	47	119.2	160	154.5	185	336	50.703	128

Centers The median alcohol level for minor crashes is 25 milligrams lower than for severe crashes (minor, 135mg compared to severe, 160mg). Yet, the median for severe crashes is only outside of the middle 50% of the minor crashes by 1 milligram (Severe median,160mg compared to the minor 3rd quarter, 159mg).

Middle 50% - Shift and Overlap, Spread

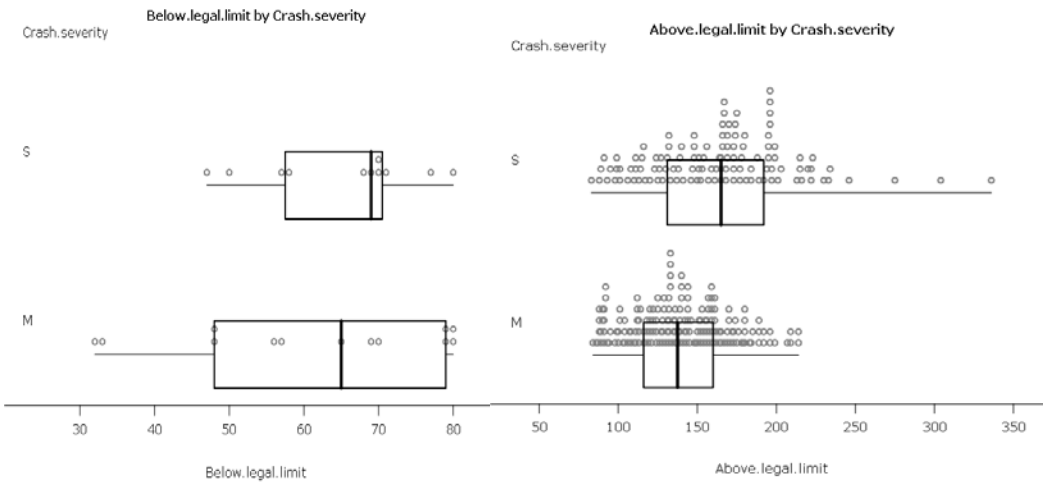
The middle 50% of the blood alcohol levels of these minor crashes overlaps the blood alcohol levels of these severe crashes by quite a bit, the lower quartile of these severe crashes being 119.2mg and the median of these minor crashes being 135mg. The middle 50% of the blood alcohol levels for these severe crashes is more spread out than the blood alcohol levels of these minor crashes, the interquartile range of these severe crashes being a difference of 65.8 milligrams per 100 millilitres and 48 milligrams per 100 millilitres for these minor crashes. This shows some more variation in the blood alcohol levels for these severe crashes than for these minor crashes.



Shape

The distribution of blood alcohol levels for these severe crashes and these minor crashes are both almost equal, with little skewness. Both have mean and medians that are very similar (severe crashes mean of 154.5 and median of 160 and minor crashes mean of 133.7 and median of 135).

The graphs show the data separated into blood alcohol levels that were of or below the legal limit (80mg per 100mL) and above the legal limit. When separated like this there is a right skew apparent for the severe crashes. This shows that in this sample, many of blood alcohol levels were not an extravagant amount higher than the limit. Although I think that in the population there would be more points of higher blood alcohol levels. The few points from this sample that we have of under the blood alcohol limit show a relatively even spread which I think is reasonable for the population because these crashes may or may not have been caused by the alcohol in the persons system.



Conclusion

From this sample it would be a reasonable assumption to say that there is a difference between the blood alcohol levels of severe crashes and minor crashes. More specifically, blood alcohol levels for severe crashes are higher than blood alcohol levels for minor crashes. From the sample we can be pretty sure that blood alcohol levels of severe crashes are 11.5 to 35mg per mL higher than for minor crashes.

From this sample it can be concluded that the blood alcohol levels of severe crashes are higher than of minor crashes as stated in the hypothesis, because the bootstrap confidence interval for the difference of blood alcohol levels for severe and minor crashes does not contain zero.

This conclusion is based on the bootstrap confidence interval that was calculated from the sample, which involved resampling from the original sample on 317 different points of data. The main assumption of the method used is that the sample represents the population of blood alcohol levels. The sample may not have been representative of the population as there were not as many higher blood alcohol levels for severe crashes as may have been expected. In the population there would likely be more points of higher blood alcohol levels, especially for the severe crashes. Yet the sample did show the higher blood alcohol levels for the severe crashes which was the reasoning for using the median for the analysis instead of the mean, as the mean would skew the results of the bootstrap confidence interval.

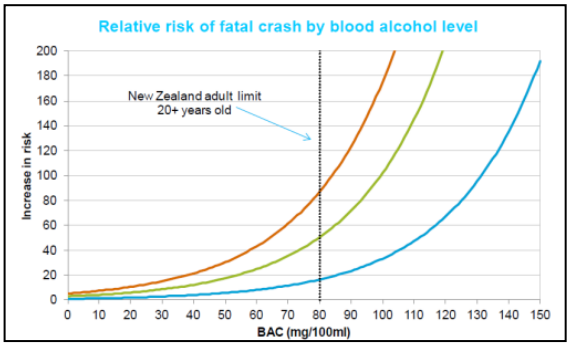
LOW EXCELLENCE – ALICE

Marker comments

PROBLEM – Risk graph given with clear interpretation related to investigation; hypothesis given, but direction of trend not clear; justification for hypothesis opinion rather than referenced; investigative question about difference in medians, population clearly defined.
ANALYSIS – Descriptions in context and quantified; data re-categorised by legal BAL limit and described with link to population; good contextual links and thinking evident about the population but based on personal opinion rather than factually referenced
CONCLUSION – CI interpreted, sample→ population inference clear; correct call; correct call and justification; comments relating findings back to context.
Split of data a nice idea, but didn't really contribute much further depth to the analysis. Contextual links show depth of thinking, but some personal interpretation of information rather than based on referenced material hence LOW EXCELLENCE *rather than* HIGH EXCELLENCE.

Problem

We often see in the media many advertisements trying to convince drivers that if they have consumed alcohol they should not be driving. This is because alcohol reportedly means you have a much higher risk of crashing because your decision making skills and awareness of what is around you is much worse after having drinking alcohol than when you have had no alcohol. The Ministry of Land Transport in a detailed report on http://www.transport.govt.nz/research/Documents/Alcohol-drug-crash-statistics-2011-(1).pdf looking at the effect alcohol and drugs have on driving based on crash statistics from the year ending December 31, 2011, has concluded that when your blood alcohol concentrate levels are 100mg per 100 ml your relative risk of having a fatal crash is much greater. For drivers aged between 15- 19 you are about 170 times more likely, for people aged between 20-29 you are about 105 times more and for those aged 30+ you are about 35 times more likely. A detailed summary is shown in the graph below-



As Year 13 students many of us are starting to turn 18 so will legally be allowed to drink alcohol soon. On top of this many of us already have our licenses. I want to investigate this possible connection between your Blood Alcohol Level and how severe a crash you are likely to have to give information to students who are transitioning into this new world and may face the social pressures that are around them to drink and drive.

I believe that there will be a strong connection between whether you have a serious injury crash or a minor injury crash and how high your Blood alcohol levels are. I think this because the amount of advertisements we see in the media to do with stopping people from drink driving must mean that there is a high level of danger involved with drinking and driving. Otherwise the government would not spend money on trying to advertise this fact.

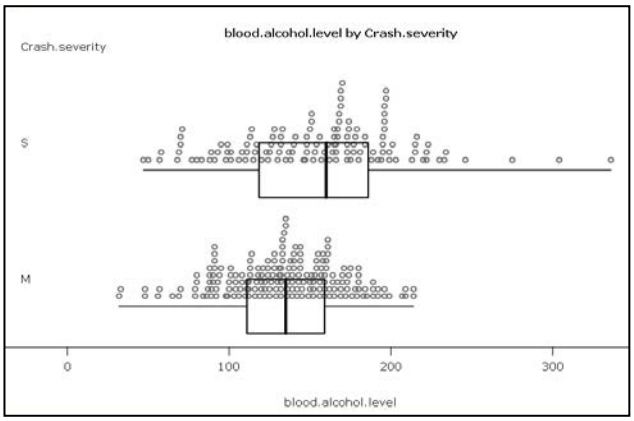
My variables I have chosen to examine are the crash severity in terms of if was a minor injury crash (M) or a serious injury crash (S) compared to how much alcohol was found in their blood at the time of the crash measured in milligrams of alcohol per 100 milliliters of blood. I have chosen these variables because I feel these are the best to express whether alcohol levels actually impact crashes severally.

What is the difference between the median Blood Alcohol Level for drivers who have had a Minor injury crash compared to the median Blood Alcohol Level for drivers who have had a Serious injury crash in New Zealand in 2011 in crashes where drugs and alcohol were deemed a factor?

Analysis

Summary of blood.alcohol.level by Crash.severity

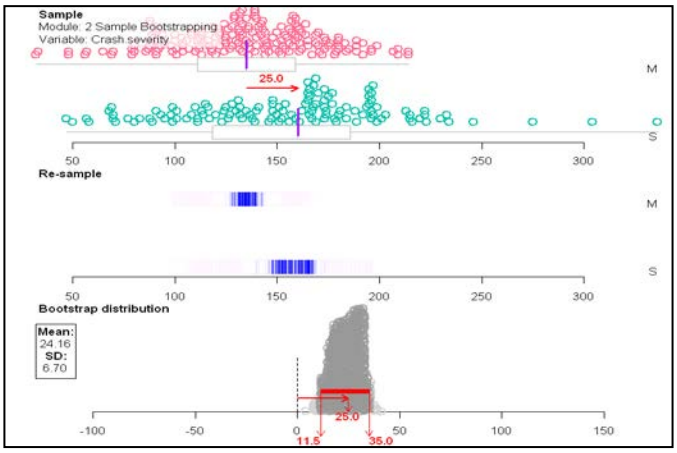
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Std.dev	Sample.Size
M	32	111.0	135	133.7	159	214	36.142	189
S	47	119.2	160	154.5	185	336	50.703	128



Overall Visual Comparisons: When I initially look at these samples my first impressions is that the serious injury crash severity BAL is more spread out than the minor injury crash BAL. For the minor injury crash there are less outliers at the top part of the spectrum. For both groups there are BAL's that are reasonably low though

Centers: The median Blood Alcohol Level for these minor injury crashes (M) are 25 milligrams of alcohol per 100 milliliters lower than the Median BAL for these serious injury crashes (S). This being that for M it is 135 mg/100ml and for S it is 160 mg/100 ml. Both these median figures are well over the legal limit in New Zealand for 20+ year olds, which is 80 mg/100ml. I do not find it surprising that these crash statistics show that the median BAL for serious and minor crashes are both over the legal limit because the government would enforce a limit that is safe and reasonable.

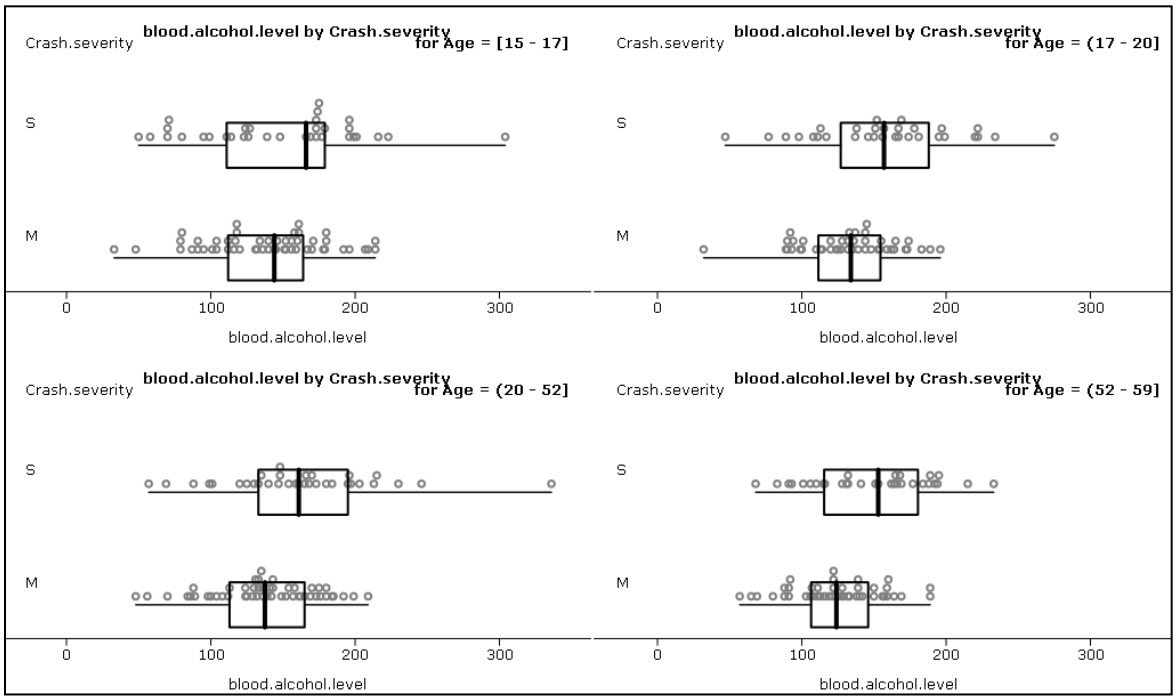
Middle 50%- Shift and overlap, spread: The middle 50% of these serious injury crashes (S) BAL's overlaps the middle 50% of these minor injury crashes (M) BAL's. The lower quartile for M is a BAL of 111mg/100ml, which is 8.2mg/100ml below the lower quartile for S, which is a BAL of 119.2mg/100ml. The upper quartile for these Minor injury crashes is 159 mg/100 ml which is a mere 1 mg/100ml less than the median for the serious injury crashes (160 mg/100ml). The middle 50% for these serious injury crashes is more spread out than the middle 50% of the minor injury crashes. The inter-quartile range for S being 65.8 mg/100ml, and the inter-quartile range for M being 48 mg/100 ml. This means that there is a bit more range in the variation of BAL for S than there is for M. This makes sense because it is possible to have a serious injury crash even without the influence of strong intoxication. Which is why the data for the serious injury crashes BAL's vary more.



Shape: The distribution for the serious injury crashes BAL's is right skewed, compared to distribution of the minor injury crashes, which is left skewed. Most of the data for S is between 32mg/100 ml and 220 mg/100 ml. With a few outliers ranging up to BAL's as big as 336mg/100 ml, these outliers are the majority of what is pulling the distribution out and causing it to be right skewed. This is similar in the case of the minor injury crashes just less severe. Most of the data in this case is between about 79mg/100ml and 200mg/100ml, with a few outliers in the bottom half. This is what causes the distribution to be left skewed in the minor injury crashes case. I believe that this makes sense in the context because in the minor injury crashes although there was alcohol in their blood that is not necessarily what caused the crash, and people with this much alcohol in their blood probably do get away with driving a lot of the time without a crash. It also makes sense in the case of the serious injury crashes being right skewed because not many people would consider driving with as much alcohol in their systems as the few outliers that did drive, resulting in a crash.

Unusual or interesting: Something I found quite interesting is the fact that there are a small but still significant number of crashes that are happening when the drivers BAL is under the legal adult limit in New Zealand which is 80mg/100 ml. This limit is different for those drivers aged under 20. Their legal limit is zero. Because of this I thought it would be interesting to look at how many of these crashes still crashed, when they were under the legal limit per what age group they are in. In other words, are the people who are crashing whose BAL's are under 80mg/100ml under 20 years old? In order to see this I constructed the graph on the following page which breaks up the data into the age groups (15-17), (17-20) (20-52) and 52-59)-

As you can see there is quite a mixed variation in the age groups of the crashes that were under the legal adult BAL. There were 21 people all together and out of these 21 people there were 10 that were over 20 so at the time of their crash were under the legal limit. Out of these 10 crashes, 4 of them were serious injury crashes. One limit of this data is though that we do not know if the drugs or alcohol in the drivers system was the cause of the crash. If we did know this we may be able to further analysis whether we think the BAL for adult drivers should be lower.



Conclusion

From these samples I can make the call that there is a difference in the median Blood Alcohol Level for drivers who have had a Minor injury crash compared to the median Blood Alcohol Level for drivers who have had a serious injury crash out of the crashes in New Zealand 2011 where drugs or alcohol were deemed a factor**.** That is that I can make the call that the median BAL for these serious injury crashes is higher than the median BAL for these Minor injury crashes. From these samples I am pretty sure that the median BAL for serious injury crashes out of crashes in New Zealand in 2011 where drugs or alcohol was deemed a factor is somewhere between 11.5 and 35.0 more than the median BAL for these minor injury crashes

I can make a call from these samples that there is a difference in the median Blood Alcohol Level for drivers who have had a Minor injury crash compared to the median Blood Alcohol Level for drivers who have had a serious injury crash out of the crashes in New Zealand 2011 where drugs or alcohol were deemed a factor because the median BAL for these serious injury crashes (160mg/100ml) is 1mg/100ml higher than the upper quartile for these minor injury crashes (159mg/100ml). This conclusion is also shown by the fact that my bootstrap confidence interval for the difference between the BAL of serious injury crashes and minor injury crashes does not contain zero. This indicates that the BAL for serious injury crashes tend to be higher than the BAL for minor injury crashes.

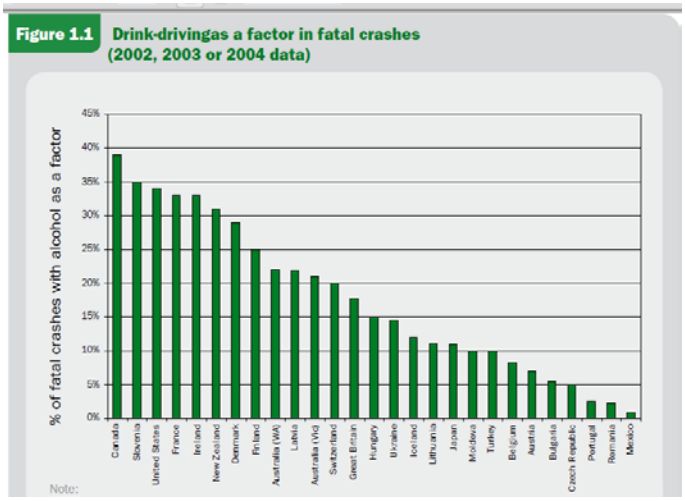
I am basing this conclusion on the bootstrap confidence interval I calculated, which involved re-sampling (1000 times) from my original random sample of 317 drivers who crashed in NZ in 2011 and drugs or alcohol were deemed a factor. This process is based on the assumption that my original sample was representative of all the 1409 crashes in NZ in 2011 where alcohol or drugs were deemed a factor. I believe that this was the case because it fits in with the 'Relative Risk of fatal crash by blood alcohol level' statistics I found at the start of this investigation. Both these statistics and my statistics showed that if you drink alcohol your risk of having a crash is much higher.

One thing I am interested and surprised about though is the fact that as seeming by these results you get to a certain point when you are so intoxicated you are almost definitely going to have a serious injury crash. Once your BAL gets past about 220 it looks like if you crash you will almost definitely come under the serious injury crash category. This is interesting and I hope will convince some of the students I have done this investigation for to not, in the future, get drunk before driving.

HIGH EXCELLENCE – TROY

Problem

After seeing the popular "Legend" drink-driving awareness ads, I discovered that over 40% of all drink-driving related crashes in New Zealand involve drink-drivers under the age of 24 years (NZTA Crash Statistics 2008-2010). The high proportion of accidents involving drink-driving teens has sparked controversy over licensing age, legal age to buy alcohol and alcohol limits for driving teens. Recently, there have been law changes as a result of these concerns, including raising the drivers licensing age by a year, a zero-alcohol limit on drivers under 20 and the proposed law to raise the legal age to purchase alcohol at liquor outside of a licensed establishment to 20 years. Recent statistics collected by the NZTA have found that around 200 young drivers are involved in a fatal or serious injury drink-driving crash in New Zealand. The percentage of drink-driving as a factor in fatal crashes varies between countries over the world, with New Zealand ranking 7th in the world for highest percentage of drink-driving related fatal crashes, with around 29% (2002-2004 WHO data).



As a year 13 student and teen driver, this issue impacts me greatly as I leave home to start tertiary education. I want to investigate this possible difference in crash statistics between drivers in the high risk group and low groups without the large amount of controversy surrounding the issue.

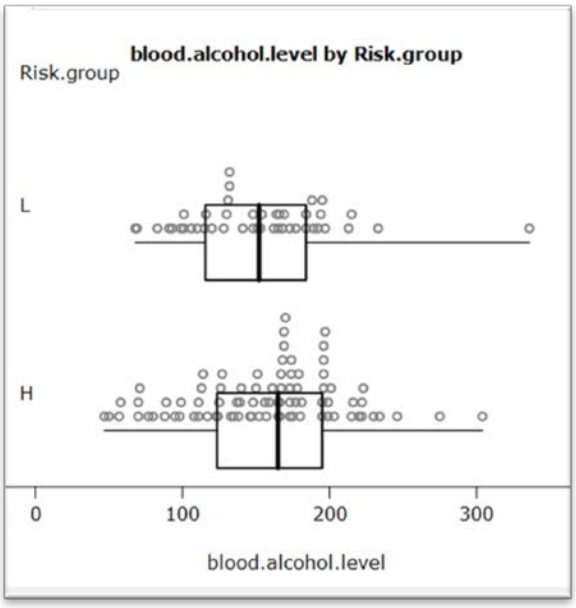
I think that drivers involved in serious drink-driving crashes in the higher risk group will have a higher median blood alcohol level than drivers in the lower risk group because:

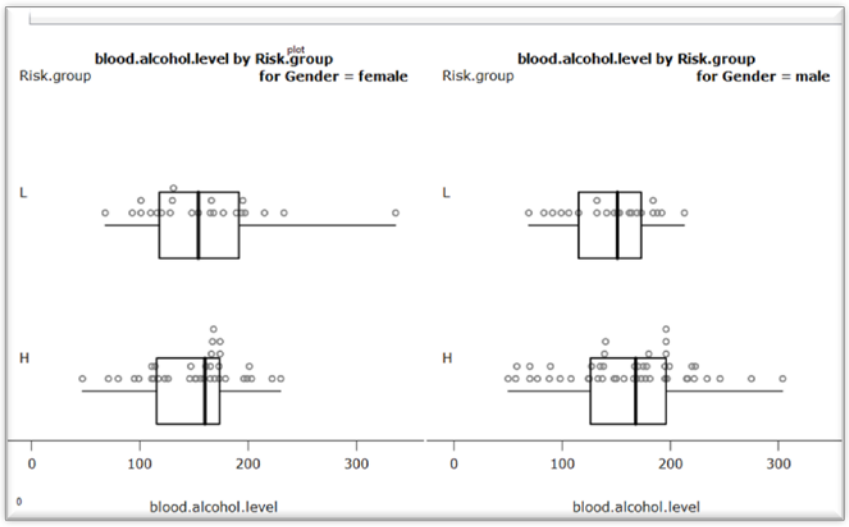
- Insurance risk groups are determined by age, older people are categorized in the low risk group and younger people in the high risk group.
- A recent survey has found that drivers in the first 3 years of driving are 20% more likely to be involved in a traffic collision than those who are more experienced (SADD). As most drivers get their license when they are 16, I assume that the higher risk in collision for less experienced drivers would have an influence on the drink-driving statistics.
- Young adults are more often involved in heavier drinking scenes due to immaturity of drinking and typical party scenarios encouraged by university life (personal experience).
- If young adults are involved in heavy drinking parties, more often than not, they are responsible for themselves and, therefore, are more likely to be tempted to drink drive, rather than stay somewhere other than their home.
- SADD (Students Against Driving Drunk) found that vehicle crashes remain the leading cause of death for young adults and, in 2007, 64% of young drink-drivers involved in fatal crashes were not wearing a safety belt. (NHTSA)

I will look at the risk group and blood alcohol level of drivers involved in serious traffic collisions. I have chosen to examine the difference in serious drink-driving crashes through the use of blood-alcohol levels, as the medical expenses data would be influenced by the age of the car, location of crash, number of occupants and medical insurance, amongst other variables which may hide any difference in the serious drink-driving crashes.

I wonder what the difference in median blood-alcohol level is between drink-drivers with in the High risk group and the drivers in the L risk group involved in serious-injury crashes where alcohol or drugs were a factor in New Zealand in 2011. (I will refer to the population as "NZ drunk drivers" for the remainder of this report).

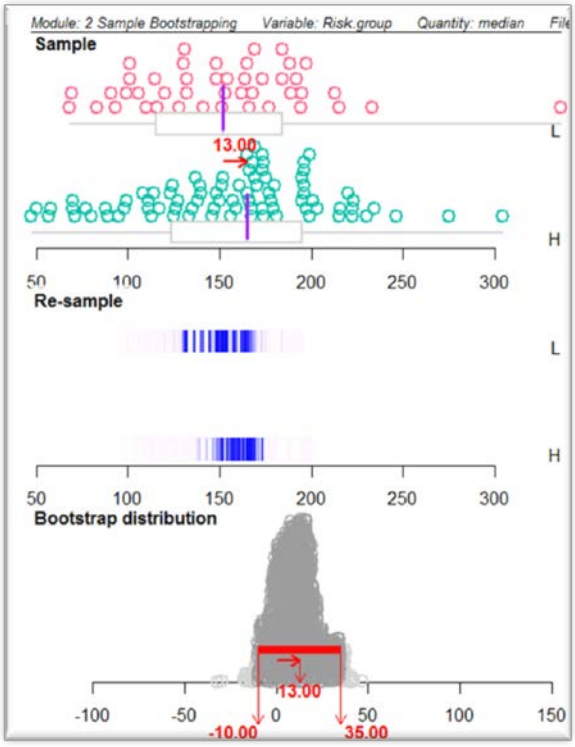
Analysis





Summary of Blood Alcohol Level by Risk Group

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Std.dev	Sample.Size
H	47	123.8	165	155.8	195	304	51.464	84
L	68	115.8	152	151.9	184	336	49.703	44



Overall visual comparisons: My initial impressions of these samples is that both groups have a similar spread; although the high risk group has a more rounded spread than the low risk groups more uniform spread. The bottom half of the high risk group's blood alcohol levels is more spread out than the low risk groups bottom half of blood alcohol levels. There doesn't appear to be a large difference in median blood alcohol levels from serious crashes where alcohol or drugs were a factor.

Centers: The median blood alcohol level for the low risk group is 13mg/100mL lower than the median blood alcohol for the high risk group (low risk group's 152mg/100mL compared to high risk group's 165mg/100mL). Both these figures are around double the legal alcohol limit for drivers over 20 (being 80mg/100mL) and this reflects the purpose of the law, as the alcohol limit is set in place to prevent intoxicated people from driving in an attempt to prevent serious accidents.

IQR – shift & overlap, spread: The middle 50% of these high risk group's blood alcohol levels overlaps the middle 50% of the low risk group's blood alcohol levels by a substantial amount, with the lower quartile for the high risk group's blood alcohol levels at 123.8mg/100mL; 28.2mg/100mL below the median blood alcohol level for the low risk group at 152mg/100mL. The median for the high risk group's blood alcohol level at 165mg/100mL is 19mg/100mL lower than the upper quartile of the low risk group's blood alcohol level of 184mg/100mL.

The middle 50% of the low risk group's blood alcohol level is similarly spread out than the high risk group (IQRlow = 68.2mg/100mL, IQRhigh = 71.2mg/100mL). This shows that there is little difference in variation in the middle 50% of blood alcohol levels for low risk groups and high risk groups. I think that this pattern may also be seen in the population

as high alcohol levels impair motor control and ability to control a motor vehicle, reasonably independent of age, and therefore, it would be expected that both groups would have similar values of median blood alcohol levels.

Unusual: I thought it was unusual that, in this sample, the lowest value for both high risk group and low risk group was under the legal limit, lowest value for high risk group of 47mg/100mL, 33mg/100mL lower than the legal limit for drivers over 20 and the lowest value for low risk group of 68mg/100mL, 12mg/100mL lower than the legal limit for drivers over 20. I think it is interesting that alcohol was classified as a factor in these serious crashes and leads me to believe that there must have been another factor in the crashes (possibly drugs or inexperience) as it is unlikely that the low alcohol levels would have meant a significant enough loss of motor control or impairment of judgment to result in the serious crash.

Shape: The distribution of blood alcohol levels for both the high risk group and low risk groups are relatively symmetrical, with the lower and upper quartile of the high risk group's blood alcohol levels (123.8mg/100mL and 195mg/100mL, respectively) are around 30mg/100mL from the median of 165mg/100mL and, similarly, the lower and upper quartile of the low risk group's blood alcohol levels (115.8mg/100mL and 184mg/100mL, respectively) are around 30mg/100mL from the median of 152mg/100mL, also.
I think that the distributions of blood alcohol levels in the population of NZ drunk drivers would also be symmetrical, with the risk of serious crash increasing as blood-alcohol level for drink-drivers increases, up until a point where the likelihood of a drunken person being able to even drive a car at all begins to lessen.

Conclusion

From these samples, I am unable to make a clear-cut call whether or not there is a substantial difference of blood-alcohol levels for NZ drunk drivers between drivers in the high risk group and low risk group.
From these samples, I am pretty sure that the median blood-alcohol levels for NZ drunk drivers in the high-risk group drivers is somewhere between 10mg/100mL less and 35mg/100mL more than the median blood-alcohol level for NZ drunk drivers in the low-risk group.

I am unable to make a clear-cut call whether or not there is a substantial difference of blood-alcohol levels for drivers involved in serious crashes where alcohol or drugs were a factor in New Zealand in 2011 between drivers in the high risk group and low risk group as there is a large amount of overlap of the middle 50% of both data sets, with the lower quartile for the high risk group's blood alcohol levels at 123.8mg/100mL; 28.2mg/100mL below the median blood alcohol level for the low risk group at 152mg/100mL and the median for the high risk group's blood alcohol level at 165mg/100mL is 19mg/100mL lower than the upper quartile of the low risk group's blood alcohol level of 184mg/100mL. My bootstrap confidence interval for the difference between population blood-alcohol levels for NZ drunk drivers contains a difference of zero, indicating that there is a possibility that there is little to no difference in median blood-alcohol level between the low-risk group drivers and high-risk group drivers.

I am basing this conclusion on the bootstrap confidence interval I calculated, which involves re-sampling from my original sample of 129 drivers involved in serious-injury crashes where drugs or alcohol were involved. The key assumption with this method is that my original sample was representative of the population of all NZ drunk drivers. As both of my sample medians were approximately twice the legal alcohol limit for drivers over 20 years, I think it reasonable to believe that my original sample is representative of the population. As the medians are approximately double the legal alcohol limit for drivers over 20 years, it is plausible that this would result in serious-crashes. To confirm this, I would need to investigate other categories in the data set further and compare them to known population values to make the decision finally. If it was wanted to get a more definite conclusion, the sample size would need to be increased to decrease the variation in the data.

My sample did highlight a few outliers in the low-risk group and high-risk groups, where some drivers were around double the median blood-alcohol level and others as low as a third of the median-blood alcohol level – this was the reason I chose to use the median as my summary statistic rather than mean. The mean would have been influenced by the large amount of small blood-alcohol level values in the high-risk group (sample mean for this high-risk group blood-alcohol levels was 155.8mg/100mL where as the sample median was 165mg/100mL).