**Overview of Formal Inference Report Writing** AS91582

***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.

***Aiming for Excellence***

|  |  |  |
| --- | --- | --- |
| **Achievement** | **Merit** | **Excellence** |
| http://www.pamhook.com/plug-ins/soloSymbolgen/soloImage.php?color=a90a34 Use statistical methods to make a formal inference. | http://www.pamhook.com/plug-ins/soloSymbolgen/soloImage.php?color=a90a34 Use statistical methods to make a formal inference, with justification. This involves linking components of the statistical enquiry cycle to the context, and/or to the populations, and referring to evidence such as sample statistics, data values, or features of visual displays in support of statements made. | Use statistical methods to make a formal inference, with statistical insight. http://www.pamhook.com/plug-ins/soloSymbolgen/soloImage.php?color=a90a34 This involves integrating statistical and contextual knowledge throughout the statistical enquiry cycle, and may include reflecting about the process; considering other relevant explanations. |

|  |  |  |  |
| --- | --- | --- | --- |
| **Multivariate Data**  | ☹ | 😐 | ☺ |
| * Multivariate data includes multiple discrete and / or continuous variables about a particular topic.
 |  |  |  |
| * In this assessment you will be compare a continuous variable for two groups using dot plots and box plots (from iNZight) and writing a report.
 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Overview of Report (Use headings 1-7 to organise your report)** | ☹ | 😐 | ☺ |
| I notice / I wonder (**do not** include these notes in your final report)  |  |  |  |
| 1. Introduction / Background  |  |  |  |
| 2. Discuss sample distributions |  |  |  |
| 3. Discuss sampling variability |  |  |  |
| 4. Make an appropriate formal statistical inference  |  |  |  |
| 5. Conclusion |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. **Writing Introductions when comparing multivariate Data**
 | **Exemplar statements** ***Italicised statements give alternative statements for describing the data*** | ☹ | 😐 | ☺ |
| **Description and** **Investigative Question** | * Description of topic / variables from topic (one sentence).
 | * Measurements such as height, weight and lean body mass are useful ways of comparing athletes’ health and performance.
* Such measurements are likely to be different in males and females due to the different physical characteristics of athletes.
* The lean body mass of a person is calculated as body weight – (body weight x body fat%) and is a measure of the mass of a person’s organs, blood, bones, muscle and skin[[1]](#footnote-1).
 |  |  |  |
| * Comparative Question
* Difference in Parameter (median) Variable for Groups in Population
 | * This report will investigate the difference between the median lean body mass (LBM) of female athletes in Australia and the median lean body mass of male athletes in AIS?
 |
| * Aim / Interest (Why worth investigating? Questions?)
 | * An understanding of the differences between the lean body mass of male and female athletes might be useful because it could help tailor specific programmes for male and for female athletes linked to their actual physical attributes.
* The lean body mass is also a useful way of tracking changes in body fat (as compared with body mass) as losing muscle is not the desired outcome of weight loss for athletes.[[2]](#footnote-2)
 |
| **Data / Survey** | * Sample / source
 | * The source of this data is a sample of 120 athletes from the Australian Institute of Sport.
 |  |  |  |
| * Definition and description of variable, groups and parameter.
 | * The variable being investigated is the lean body mass of both female and male athletes. The parameter being compared is the median lean body mass.
 |
| * Important aspects of data collection details / validity.
 | * This data is likely to be valid as it is collected by doctors at the AIS.
* *This data was collected by xxx and therefore may not be a valid measure of …*
 |
| **Research** | * Link findings to what you know / have researched about the variables.
* Why might one group have a different ‘variable’ than the other?
 | * In this context it seems likely that the lean body mass for male athletes would be higher than female athletes as they tend to be more muscular.
* One important aspect of this data is that all of the data points are for athletes and so any findings may not be applicable to non-athletes.
 |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. **Describe sample distribution**
 | **Exemplar statements** | ☹ | 😐 | ☺ |
| **Graph** | *Inzight Graph: Dot and Box plots*Inzight – summary table |   |  |  |  |
| **Central Tendency** | * Visual and numeric
* Compare sample / population
* Context / research
 | * The median lean body mass for male athletes is 74.5kg which is 19.58kg higher than the median lean body mass for female athletes of 54.92kg.
* This difference is quite large and appears to fit with what research suggests about the physical differences between male and females, i.e. that male athletes would have more muscle mass and thus a larger LBM.
* I did expect the male athletes to be heavier, although I did not expect the difference to be this large.
* I expected… I was surprised… This makes sense because…
 |  |  |  |
| **Shape**  | * Shape: n-shaped, bimodal, rectangular, u-shaped
* Skew: right / left
* Symmetrical
* Compare sample / population
* Context / research
 | * Both sets of data are n-shaped with the majority of LBMs in the middle of the data and fewer athletes with very small or very large LBMs.
* There is a possibility of bimodality in the male data with a group of athletes centered around 70kg and a second group centred around 75-80 kg.
* This bimodal shape may suggest two groups of male athletes, perhaps those that tend to compete in sports where additional muscle is an advantage (such as weight lifting) and those where being more lean is an advantage (such as long distance running).
* There is evidence of right skew in the male data caused by three athletes with quite large LBMs of around 100kg.
* I expected… I was surprised… This makes sense because…
 |  |  |  |
| **Spread** | * Visual and numeric
* IQR or Standard deviation
* Compare sample / population
* Context / research
 | * In the sample distribution of male athlete LBM tends to be more spread out than the female athletes LBM. This is confirmed by the difference in the standard deviations, with the male athlete Std.dev of 9.9kg compared with the female athlete Std.dev of 6.9kg. This shows that the male athlete LBM has greater variation than the female athletes LBM.
* This isn't what I expected because... This is what I expected because…
* This greater spread is also evident in the middle 50% of the data as the interquartile range for the male athletes LBM is 12.75kg, whereas the IQR for the female athletes is 7.47kg. This also shows that there is more variation in the male athlete LBM than the female LBM.
 |  |  |  |
| **Shift / Overlap** | * Visual and numeric
* Compare sample / population
* Context / research
 | * There is a clear shift evident in the box plots, with the middle 50% of the male athletes having a much greater LBM than the middle 50% of the female athletes. There is no overlap between the middle 50% of data for these two groups meaning more than 75% of the male LBMs are greater than 75% of the LBMs for the female athletes.
 |  |  |  |
| **Unusual values** | * Visual and numeric
* Context / research
 | * There are three male athletes that have rather larger LBMs (of approx. 100kg) which do skew the male data to the right. These three data values contribute to the larger spread seen in the male sample.
* These athletes could possibly be weightlifters, or some other sport that would support a very high LBM (link to research….)
 |  |  |  |
| **Context / Research** | * The differences observed in the two samples are supported by research / do not appear to be supported by research such as…
* These differences are to be expected because…
 | * The differences observed between the two samples are also supported by research which found that males tend to have an average skeletal muscle of 42% of body weight, whereas for females it is about 35%.[[3]](#footnote-3)
* These differences are to be expected because…
 |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. **Discuss sampling variability**
 | **Exemplar statements** | ☹ | 😐 | ☺ |
|  | * Recognises sampling variability, including variability of estimates.
* A biased or poor sample will give poor population estimates
 | * I decided to use medians as my sample statistic as the means would be influenced by the few large values in the male data.
* I know that these sample medians would be highly likely to change if I was to select a different sample from the population due to sampling variability.
* This sample does appear to be representative of Lean Body Mass in athletes and therefore is likely to be useful for generating an estimate of the population medians.
 |  |  |  |
|  | * Discussion of Sample variability
* Link to bootstrap distribution
 | * In order to make a formal inference a bootstrap distribution of re-sample medians has been found as I know this is similar to the distribution of medians from repeated random sampling.
* I have used this bootstrap distribution to model the sampling variability in the data, and come up with a confidence interval based on this bootstrap distribution.
 |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. **Make an appropriate formal statistical inference**
 | **Exemplar statements** | ☹ | 😐 | ☺ |
|  |  |  |  |  |  |
|  | * Inference about **population**
* Population parameter identified (mean or median)
* In context – what does it mean?
* Level of uncertainty. “I can be reasonably sure that…” ‘it’s a safe bet that”
* Evidence
 | * I am fairly sure that, for all athletes in AIS, the median lean body mass of male athletes is more than the median lean body mass of female athletes and that the difference in the medians is between xx and yy kg.
* Therefore I can make the call, given the sample I have, that the median LBM of male athletes is more than the median LBM of female athletes.
* This data shows me that it is quite likely that back in the AIS population there is a difference between the medians for the LBM of male and female athletes.
* These findings agree with my initial hypothesis – (or research) that male athletes would have a higher LBM as they tend to have greater muscle mass than female athletes.
 |  |  |  |
|  | * Evidence from bootstrap confidence interval
 | * My bootstrap confidence interval is from xx to yy. This does not include 0 and therefore I am reasonably confident that there is a difference in the LBM of male and female athletes in the AIS population.
 |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. **Conclusion**
 | **Exemplar statements** | ☹ | 😐 | ☺ |
| **Summary** | * Make the call
* Formal inference used to answer question with justification and links to context.
* Includes interpretation of the confidence interval.
* Bootstrapping evidence
* An understanding of sampling variability is also evident.
* Findings clearly communicated and linked to population and context.
 | * This statistical investigation means I can make the call, with the sample that I have, that the lean body mass for male athletes is higher than for female athletes in Australia as there is a significant gap in the median values of 55 and 75kg.
* This means there is a very good chance that the median LBM for males is higher than the median LBM for females back in the AIS population.
* This is backed up by the bootstrapped confidence interval, as a difference of 0 is well outside the confidence interval of 16.59 to 22.89kg. The bootstrapped confidence interval was achieved by resampling the data set again, with replacement, for the same number of samples. This gives us many different samples that mimic the original population. This method gave me a range of lean body masses that will cover the median of the population we sampled most of the time.
 |  |  |  |
| **Link to sample** | * Representative sample?
* Choice of parameter
 | * All of this is based on the assumption that my original sample was representative, and I have no reason to suspect it was not. The sample was of a fairly large number of athletes, and had a good spread of data, so it is likely that is is a good representation of the population of athletes at the AIS.
* I chose to examine the median as my parameter in this investigation due to the presence of outliers in the male LBM data that would influence the mean.
 |  |  |  |
| **Link to rcontext / eserach** | * Further reflection on process or further explanations for findings.
* Summary of what this means in context / research / future investigations
* Usefulness / Limitations / Improvements / Possible uses / Future Investigations
 | * One thing I didn’t consider in my study was the dfferent sports or disciplines practiced by the athletes. This might influence LBM because…
* I could have improved my investigation by first analysing sport data and then comparing the genders for these different sports.
* Possible limitations of this investigation include…
* These findings may be useful because…
* Further factors that may influence LBM are …
* In this study I did not identify the cause of the differences in the two genders LBMs …
 |  |  |  |

1. http://www.builtlean.com/2011/08/24/lean-body-mass-definition-formula/ [↑](#footnote-ref-1)
2. http://www.builtlean.com/2011/08/24/lean-body-mass-definition-formula/ [↑](#footnote-ref-2)
3. Cite website / reference [↑](#footnote-ref-3)