Sample-to-population inference progressions across senior curriculum

*__SAMPLE-TO-POPULATION INFERENCE:__ The process of drawing conclusions about *population parameters* based on a *sample* taken from the *population*.

Standard	91035 (1.10): Multivariate data	91264 (2.9): Use statistical methods to make an inference	91582 (3.10): Use statistical methods to make a formal inference
CL	Level 5 & Level 6	Level 7	Level 8
Inferential ideas	"Making the call" with L5 or L6 guidelines (based on position of medians and spread)	Using informal confidence interval for population medians (based on sample size & IQR)	Using a formal confidence interval to make an appropriate formal inference (based on the *resampling distribution* of the statistic of interest – *bootstrap confidence interval*)
Key ideas *Clear links between PROBLEM → PLAN → CONCLUSION & the situation should be evident throughout*	• Comparing overall distributions • Understanding of the measures (eg: IQR is a measure of how *spread out* the middle 50% is) • Some appreciation of sampling variability (sample size; if another sample was taken it might look like…)	• Comparing medians • Selecting a *random* sample in an appropriate way • Understanding sampling variability (impact of sample size and population spread)	• Researching context (relating the context-research-findings to what is seen in the data and to any conclusions made – the "so-what?" factor) • Working with the difference between means as well as medians
From Senior Secondary guides	• S6-1: …communicating features in context…	• S7-1: Uses *relevant contextual knowledge* when communicating findings	• S8-1: Uses *informed contextual knowledge* to support explanations and to communicate findings

Michelle Dalrymple November 2012

PPDAC Framework for **SAMPLE-TO-POPULATION INFERENCE** **Version 2**

Indicators (MERIT-ish)	Problem	• Comparison question clearly including – Variable that is being examined (*height in cm*) – Groups* that are being compared (*Year 11 boys and Year 11 girls*) – Population that inferences are being made about (*New Zealand Year 11 boys and New Zealand Year 11 girls*) – Direction of comparison (*boys tend to be taller than girls*) **Be careful that groups being selected to compare have a large enough sample size to be able to meaningfully complete the analysis and conclusion (ie groups should be approx. n > 20)*	• Comparison question clearly including – Variable that is being examined (*height in cm*) – Groups that are being compared (*Year 11 boys and Year 11 girls*) – Population that inferences are being made about (*New Zealand Year 11 boys and New Zealand Year 11 girls*) – Statistic (*median height)* – Direction of comparison (*median height of boys is greater than the median height of girls*) • Prediction of what students expect to see in their analysis	• Research into background of context to give purpose to the investigation • Comparison question clearly including – Variable that is being examined (*height in cm*) – Groups that are being compared (*Year 11 boys and Year 11 girls*) – Population that inferences are being made about (*New Zealand Year 11 boys and New Zealand Year 11 girls*) – Statistic (*DIFFERENCE in median heights between boys and girls (or means))* • Prediction of what students expect to see in their analysis and why
	Plan	Select variables to investigate from a given multivariate data set	Select random sample from a given population (containing multiple variables) • sampling method • sample size	Select variables to investigate from a given multivariate data set
	Data	Data given	Collect according to plan	Data given Students may choose to re-categorise data as appropriate to investigation
	Analysis	*Selecting and using appropriate displays and summary statistics & discussing sample distributions comparatively is expected across all levels. Statements should mention context (variable and groups – eg …heights of these boys compared to … heights of these girls), and evidence to support statements (generally this means numbers)* • Appropriate summary statistics (5 number summary) • Comparative descriptive statements of distributions – including overall picture, centres, shape, middle 50%, shift, overlap, spread (IQR), unusual or interesting features but note for distributional shape the discussion must be on the inferred population distributions. • Comparative statements should include the context, i.e. variable, units, values, population		
		• Dot plots – for each group • Box plots – for each group Comparative statements will include discussion of distributions (shape)	• Dot plots – for each group • Box plots – for each group • Informal confidence interval shown on box plot, and provided Comparative statements will include comparison of centres (medians) Comparative statements should be making some contextual links back to the population (the "so what?" factor)	• Dot plots – for each group • Box plots – for each group • Plot of the re-sampled distribution & formal confidence interval provided Comparative statements will include discussion about the difference in medians (or means) Comparative statements should be making some contextual links back to the population and initial research (the "so what?" factor)

Indicators (MERIT-ish)	Conclusion Justification	Conclusion should clearly cover • Informal inference – Sample → population link strong, with population clearly identified – Should reflect investigative question – Correct call (L5 or L6 call)	Conclusion should clearly cover • Interpretation of informal confidence intervals – Sample → population link strong – Some level of uncertainty evident ("pretty sure") – Population parameter identified ("population median height") • Correct call, with justification – Should reflect investigative question – Call based on whether intervals overlap or not – Direction of evidence (if intervals do not overlap) • Some understanding of sampling variability For example – Different random samples will give different statistics, and what impact this may have on making the call – Impact on confidence intervals of changing sample size • Linking call back to the context and starting to think about what this means (the "so what?" factor)	Conclusion should clearly cover • Interpretation of formal confidence interval – Sample → population link strong – Some level of uncertainty evident ("pretty sure") – Population parameter identified • Correct call, with justification – Should reflect investigative question – Call based on whether zero is contained within the interval or not – Direction of evidence (if zero outside of interval) • Linking back to the context and using initial research to help explain what this means (the "so what?" factor)
Step **DOWN** to ACHIEVE	Contextual links not as evident Statements not supported/justified			
Step **UP** to EXCELLENCE	Statistical insight, further depth of thinking demonstrated with clear contextual links, greater understanding of sampling variability, and understanding of coverage of the confidence interval(s) in 2.9 & 3.10, no errors demonstrated in understandings, interpretation or explanation of findings/data.			