Reading Confidence Intervals

Chris Wild, University of Auckland

The recent video on analysing numeric outcomes used the NHANES-1000 dataset and the variable **AgeFirstMarij** (age of first use of marijuana - among those who *had* used marijuana). Output from iNZight for "Get Inference" included the following (values have been rounded):

```
AgeFirstMarij
Group means with 95% Confidence Intervals
ci.lower    estimate    ci.upper
   16.5         16.9        17.4
```

Here the confidence interval extends from 16.5 (ci.lower) to 17.4 (ci.upper). How can we read, interpret and communicate these values?

We would typically say something like this:
"With 95% confidence, the population mean for age-of-first-use is somewhere between 16.5 and 17.4."

The pattern used is:
"With 95% confidence, the population *quantity* **for** *variable* **is somewhere between** *ci.lower* **and** *ci.upper* **."**

- *quantity* is the type of quantity being estimated (e.g. mean, median, percentage, slope, ..)
- *variable* is the variable we are using (e.g. height, income, blood pressure, ..)
- *ci.lower* and *ci.upper* are the lower and upper confidence limits (numbers) for the true or population value of the quantity
  (there are many variations that would work just as well.)

AgeFirstMarij by Education.reord

Group Means with 95% Confidence Intervals

	ci.lower	estimate	ci.upper
8thGrade	10.984	13.83	16.6813
9_11thGrade	14.916	16.20	17.4984
HighSchool	15.135	16.02	16.9084
SomeCollege	16.090	17.12	18.1497
CollegeGrad	17.097	17.72	18.3366

The table above has come from looking at **AgeFirstMarij** broken out by (re-ordered) **Education** and clicking "Get Inference" (Normal) in iNZight. We have been given separate confidence intervals for mean-age-of-first-use for each Education group. For the **CollegeGrad** group the confidence interval goes from 17.097 (round to 17.1) and 18.3365 (round to 18.3). We have a little more information to add to our sentence. We'd say something like this:

With 95% confidence, the population mean-age-of-first-use for the CollegeGrad group is somewhere between 17.1 years an 18.3 years.

The pattern used is:
"With 95% confidence, the population *quantity* **for** *variable* **for the** *subgroup* **is somewhere between** *ci.lower* **and** *ci.upper* **."**
(there are many variations that would work just as well.)

Confidence Intervals for Differences (A - B)

Confidence intervals for differences in a set of proportions

Differences between Proportions (col - row)

.....(estimates panel removed) .....

95% Confidence Intervals:

	CollegeGrad	SomeCollege	HighSchool	9_11thGrade
SomeCollege	-0.084018			
	0.084018			
HighSchool	0.060912	0.060912		
	0.20762	0.20762		
9_11thGrade	0.14451	0.14451	0.018915	
	0.27508	0.27508	0.13213	
8thGrade	0.19824	0.19824	0.072317	0.0041212
	0.31645	0.31645	0.17384	0.090984

C.I. limits for difference between **College Grad** and **SomeCollege**

C.I. limits for difference between **HighSchool** and **8thGrade**

Above are confidence intervals for a set of differences in proportions produced using a re-ordering of the Education variable in NHANES-1000 using "Get Inference" (Normal) in iNZight. The picture shows how the interval-information is laid out. Each pair of numbers gives the lower and upper confidence limits for the difference between the proportion in the **column-name** category **minus** the proportion in the **row-name** category.

We have highlighted two intervals:

- Marked up in blue is the difference **HighSchool - 8thGrade**. The interval for this difference goes from 0.072 to 0.174.

*With 95% confidence, the population proportion in the **HighSchool** category is larger than the proportion in the **8thGrade** category by somewhere between 0.072 and 0.174.*

- Marked up in red is the difference **CollegeGrad - SomeCollege**. The interval for this difference goes from -0.084 to 0.084.

*With 95% confidence, the population proportion in the **CollegeGrad** category is somewhere between being smaller than for the **SomeCollege** category by up to 0.084 and larger by up to 0.084.*

We have approximately followed these patterns (which were written for means):

(*ci.lower and ci.upper both positive*)

With 95% confidence, the population *quantity* **for** *variable* **is larger for** *group A* **than it is for** *group B***by somewhere between** *ci.lower***and larger by***ci.upper*.

(*ci.lower negative and ci.upper positive*)

With 95% confidence, the population *quantity* **for** *variable* **is somewhere between being smaller for** *group A* **than it is for** *group B* **by** *ci.lower* **and being larger by** *ci.upper*.

(*ci.lower and ci.upper both negative*)
Reverse the order of the comparison and make the numbers positive.