Wellington, New Zealand High School Students

The file ***CASNZSec2013Wellington*** has data on *n = 827* high school students in Wellington, New Zealand. The data are from the Census @ School Project for the year 2013, and are simple random sample of all the students in Wellington who participated in the Census @ School Project.

1. The students were asked whether they had a Facebook page.
   a. What is your best thoughtful "guesstimate" of the proportion of Wellington NZ high school students who have a Facebook page?
   b. In the context of hypothesis testing, what symbol should be given to your guestimate? Should the symbol be p_0, *p*-hat, *p* or is this a *p*-value?
   c. Write in symbols the null and alternate hypotheses using your "guesstimate" making the alternate hypothesis two sided.

- Using the file ***CASNZSec2013Wellington*** get software to do a hypothesis test for your idea using the data collected in New Zealand. The variable *techfacebook* records whether or not a student had a Facebook page. The categories are "yes" and "no".
- Change the p_0 from the default in the program to whatever your "guesstimate" is, and use "yes" as success.

The questions following are based upon the output that you get.

2. a. Is the *p*-hat *reasonably likely* or *rare* given the null hypothesis based on your guesstimate p_0? Give a reason for your answer.
   b. Is the hypothesis test *statistically significant*? Give a reason for your answer.
   c. Interpretation of the *test*: Which of the following most closely matches the outcome that you see:
      (i) Our guess is definitely wrong, since the *p*-value shows that the *p*-hat is rare.
      (ii) The *p*-hat is consistent with our guess, since the *p*-value shows that the *p*-hat is reasonably likely, given our guess.
   d. Using the information in the output, show how the software calculated the test statistic.

- Next, use software and the file ***CASNZSec2013Wellington*** to get a 95% Confidence Interval for the population proportion of students having a Facebook page.

3. a. Show that the margin of error for the CI is 0.0282.
   b. What happens to the margin of error if the sample size is *increased*? Explain why.
   c. Determine whether your "guesstimate" p_0 is included in the Confidence Interval. If the p_0 is included in the range of the confidence interval, which should be true: (i) That the hypothesis test *is* statistically significant or (ii) That the hypothesis test *is not* statistically significant? Give a reason for the choice.
   d. Show how the software calculated the confidence interval.
4. What is the population? The Census@School data collection in Australia and New Zealand depends upon the voluntary participation of classroom teachers. When teachers or school administrators decide to participate, then all the students in the school are included in the data. Above it was stated that the data in the file ***CASNZSec2013Wellington*** are a SRS of students who participated.
   a. Here are some possible choices for a description of the population for these data. Choose one, and defend your choice against the other choices.
      - High school students in New Zealand
      - High school students in Welllington, New Zealand
      - High school students who participated in the C@S program in Wellington, New Zealand.
   b. Using your choice of population, give a standard interpretation of the confidence interval found.

5. Extensions?

Pros and Cons and Implications of Some Software Choices

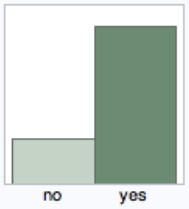
On the left are advantages; on the right disadvantages for each choice.

Graphing Calculators	
1 Students (and instructors) are already accustomed to using them from experience in other courses. 2. Instructions are included in many texts.	1. Poor graphics. 2. Only small data sets can conveniently be used. 3. Simulation possible, but not easily done. 4. Dynamics not good.
Excel	
1. Students (and instructors) are already accustomed to using them from experience in other courses. 2. Instructions are included in many texts. 3. Many students have access. 4. Good for "stepping-through" formulas. 5. Spreadsheet formula capability useful for students to learn. 6. Almost universal transporter for data.	1. Difficult to make some graphics (box plots and histograms, for example) without add-ins. 2. Add-ins work well only for Windows and not for Mac? 3. A business package, and not really a statistics package. 4. Simulations possible? 5. Dynamic capability?
Applets etc.	
1. Mostly free. (Examples below are free) 2. Simulation capabilities high, dynamic capabilities high. 3. Internet based, so mostly convenient. 4. Generally straightforward to use; the better ones make a link between "hands-on" simulation and the applet. ***GeoGebra (sp.) Distribution Calculator*** Combines graphics and calculations for many distributions, with dynamic capabilities. Can be downloaded, so can be used without an Internet connection. ***Rossman Chance Applets*** Works best if used with the teaching materials developed by the Tintle et al team. Can integrate hands-on work with software use. ***StatKey*** Works best if used with the teaching materials developed by the Lock et al team. Can be downloaded as an App to be used when there is no Internet connection	1. Data must usually be entered in some way. 2. Not a complete statistics package. 3. Internet based, so not available in some situations. (There are exceptions) 4. If not used carefully, can be a "black box" 5. Stand-alone, so not integrated within one statistical package for data analysis, and data manipulation. Only for quantitative data? Other Stat facilities comparable to calculators; data must be loaded. Data from other sources can be used, but must be entered into the Applets. Must have an Internet connection to run. Data from other sources can be used, but must be entered into the Applet.

Stat Packages Designed for Teaching	
1. Designed for statistics education in mind, so simple and straightforward to use, with dynamics built in. 2. Cost generally low.	1. Not a "real" statistics package. 2. Some techniques excluded.
Fathom/Tinkerplots	
Strong on dynamic features, simulation and data analysis in an integrated fashion. Data manipulation and data analysis easy. Cost low; free for 2014-2015, and either free or at very nominal cost beyond that? Can handle big data sets. Resides on one's own computer. Works on all platforms.	Development of package to keep up with OS changes? Community support less strong than previously. Further development?
StatCrunch	
Available through a browser, so available wherever there is a connection. Very intuitive: simple to learn and use. Cost low. Fairly extensive set of techniques included. A good set of Applets and simulation faculties that can be used with any data set. Strong community support in providing data sets, and correcting problems.	There may not be an Internet connection, or the connection may be"crowded." Data sets larger than two MB will not load.
Mainstream Stat Packages: (Minitab, SPSS, JMP, etc.)	
1. Complete statistical packages, which students may see in upper division work. 2. Cost may be high, but fro some packages now lower than in the past. 3. Generally beautiful graphics, and simulations can be done. 4. Can handle big data sets. 5. Resides on one's own computer. 6. Generally extensive tech support.	1. Usually more complicated than programs designed specifically for teaching (but may have support as well). 2. Not designed for teaching.
JMP	
Full range of analyses, with add-in designed for teaching, so dynamic facilities and simulation. Cost now reasonable for students ($20) and colleges ($500 for a five year license, is used). Active community support for teaching. Works on all platforms.	Number of choices overwhelming? Teaching add-ins still in the process of development. Necessity to write ones own guides?
Minitab and Minitab Express	
Cost reasonable for students Often shown in texts.	Minitab Express has limited options. Minitab 17 not for Mac

Distributions

techfacebook



Frequencies

Level	Count	Prob
no	181	0.21886
yes	646	0.78114
Total	827	1.00000

N Missing 0

2 Levels

Confidence Intervals

Level	Count	Prob	Lower CI	Upper CI	1-Alpha
no	181	0.21886	0.192018	0.248308	0.950
yes	646	0.78114	0.751692	0.807982	0.950
Total	827				

Note: Computed using score confidence intervals.

Estimate of CASNZSec2013Wellington.csv Estimate Proportion

Attribute (categorical): techfacebook

Interval estimate for population proportion of **yes** in **techfacebook**

```
Count:              646 out of 827, or 0.781137
Confidence level:   95.0 %
Estimate:           0.753 to 0.8093
```

95% confidence interval results:

Outcomes in : techfacebook

Success : yes

p : Proportion of successes

Method: Agresti-Coull

Variable	Count	Total	Sample Prop.	Std. Err.	L. Limit	U. Limit
techfacebook	646	827	0.78113664	0.01437796	0.75166183	0.80801174