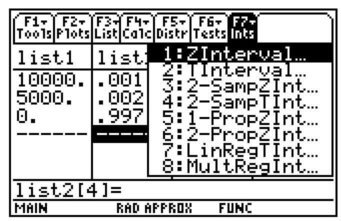
Chapter 10 – Inference for Proportions

We know the sample proportion, $\hat{p}$, is normally distributed if *both* $np \geq 10$ and $n(1-p) \geq 10$. This was seen in Chapter 18 of the text and the last chapter of this manual. With this fact we found probabilities based on the normal model of obtaining certain sample proportions (or sample means). Inference asks a different question. Based on a sample, what can we say about the true population proportion? Confidence intervals give ranges of believable values along with a statement giving our level of certainty that the interval contains the true value. Hypothesis tests are used to decide if a claimed value is or is not reasonable based on the sample. (We actually did this in the last chapter – now we'll formalize it).

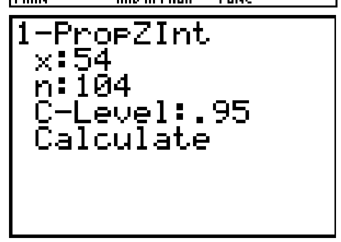
CONFIDENCE INTERVALS FOR A SINGLE PROPORTION

Sea fans in the Caribbean Sea have been under attack by a disease called *aspergillosis*. Sea fans that can take up to 40 years to grow can be killed quickly by this disease. In June 2000, members of a team from Dr. Drew Harvell's lab sampled sea fans at Las Redes Reef in Akumai, Mexico at a depth of 40 feet. They found that 54 of the 104 fans sampled were infected with the disease. What might this say about the prevalence of the disease in general? The observed proportion, $\hat{p} = 54/104 = 51.9\%$ is a point estimate of the true proportion, *p*. Other samples will surely give different results.

We can use the calculator to obtain a confidence interval for the true proportion of infected sea fans. Press STAT then arrow to TESTS. The first portion of the menu shows several hypothesis tests. We'll talk about them later. Arrowing down, we come to several possible intervals. The one we want here is choice A:1-PropZInt. Either arrow to it and press ENTER or press ALPHA MATH (A).

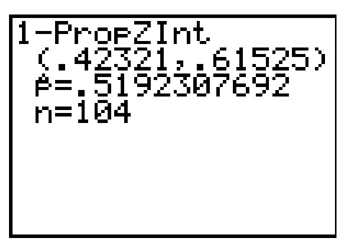


On a TI-89 all confidence intervals are located on the [F7] = 2nd F2 Ints menu. The 1-PropZInt is menu option 5. Its input boxes are labeled similarly to the prompts on the 83/84 series.



Here is the TI-83/84 input screen. Simply enter the number of observed "successes," which here is the number of infected sea fans, 54, press the down arrow, then enter the number of trials, the 104 fans that were observed, press the down arrow again to enter the desired level of confidence, finally press the down arrow again and press ENTER to calculate the results.

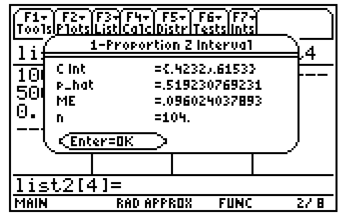
How much confidence? That is up to the individual researcher. The trade-off is that more confidence requires a wider interval (more possible values for the parameter). 95% is a typical value, but the level is generally specified in each problem.
Note: TI calculators use exact critical values when finding confidence intervals rather than using the approximate 68-95-99.7 Rule.



Here are the results. The interval is 0.423 to 0.615. Remember the calculator usually gives more decimal places than are really reasonable. Usually reporting proportions to tenths of a percent is more than enough. Your instructor may give other rules for where to round final answers. The output also gives the sample proportion, $\hat{p}$, and the sample size, *n*.

What can we say about the prevalence of disease in sea fans? Based on this sample, we are 95% confident that between 42.3% and 61.5% of Las Redes sea fans are infected by the disease, the sample proportion is 51.9%, and the margin of error is 9.6%.

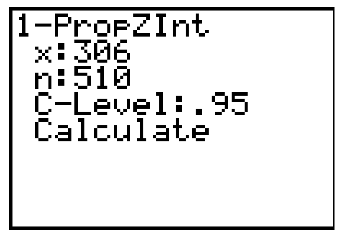
The TI-89 explicitly in its results gives the margin of error. If you are using a TI-83/84 and want to find the margin of error it is half the width of the interval. Compute $(high-low)/2$ to find it.



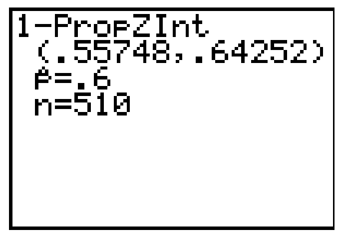
Another Example

In May 2006, the Gallup Poll asked 510 randomly selected adults the question, "Generally speaking, do you believe the death penalty is applied fairly or unfairly in the country today?" Of these, 60% answered, "Fairly." We'll build a confidence interval for the proportion of all U. S. adults who believe the death penalty is applied fairly. We had a random sample of U. S. adults which was less than 10% of the population, 60% of 510 (306) is more than 10, and 40% of 510 (204) is also more than 10, so the conditions for inference are met.

We want a 95% confidence interval. At right is the input screen. **Note:** If you are using a TI-83/84 you can actually type in the multiplication to find the number of "successes" (.6*510) and then round the result on this screen. If you are using a TI-89, you must do the multiplication on the home screen.



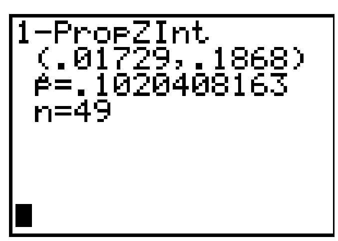
Here are the results. Based on our sample, we believe with 95% confidence that somewhere between 55.7% and 64.3% of all U. S. adults believe the death penalty is being applied fairly. Notice that since the interval only contains values above .5, we are convinced that a majority of the people in the country feels the death penalty is being applied fairly. The margin of error in these results is (.643-.557)/2 = 0.043 or 4.3%.



A SMALL SAMPLE CONFIDENCE INTERVAL FOR A PROPORTION

What can we do if we don't get the 10 successes and 10 failures in our sample? There is a simple adjustment to the above procedure due to Agresti/Couli (and Wilson). We merely add 2 additional "successes" and 4 "trials" to the observed data. This adjustment has been shown in numerous simulation studies to provide enough robustness for our usual procedure to actually give the desired confidence level (that is, 95% of 95% intervals *do* contain the true value.) We no longer call our estimated proportion $\hat{p}$, but $\tilde{p}$ to signify the difference.

Surgeons examined their results to compare two methods for a surgical procedure to alleviate pain on the outside of the wrist. A new method was being compared with the traditional method. Of the 45 operations with the traditional method, three were unsuccessful. What can we say about the failure rate for this type of operation? With only three failures, we can't calculate the interval as we have in the past. For these data, we find $\hat{p}=3/45=6.7\%,$ while using the adjustment, we find $\tilde{p}=5/49=10.2\%$. Using the adjustment, we find that we are 95% confident the traditional method will fail in somewhere between 1.7% and 18.8% of cases. Be careful – the calculator has only one symbol for the "observed proportion." You must know that you used the adjusted method and properly label your results!



HYPOTHESIS TESTS FOR A SINGLE PROPORTION

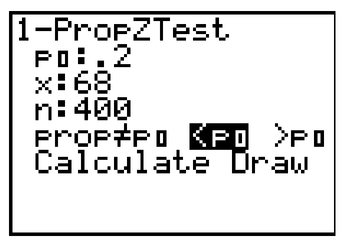
Confidence intervals in general give ranges of believable values for the parameter (in this case the proportion of whatever is being termed a "success.") Hypothesis tests assess the believability of a claim about the parameter. Certainly, if a claimed value is contained in a confidence interval it is plausible. If not, it is unreasonable. Formal tests of hypotheses assess the question somewhat differently. The results given include a test statistic (here a z-value based on the standard Normal model) and a p-value. The p-value is the probability of an observed sample result (or something more extreme), given the claimed value of the parameter. Large p-values argue in support of the claim (your result was likely to happen by chance), small ones argue against it; in essence, if the claimed value were true the likelihood of observing what was seen in the sample is very small.

A factory casts large ingots, which are then made into structural parts for cars and planes. If they crack while being formed, the crack may get into a critical part of the final product, which compromises its integrity. Airplane manufacturers insist that metal for their planes be defect-free, so the ingot must be made over at great expense if any cracking is detected. In one plant, only about 80% of the ingots have been free of cracks. In an attempt to reduce the cracking proportion, they institute changes to the process. Since then, 400 ingots have been cast, and only 17% of them were cracked. Has the cracking rate decreased, or was the 17% just due to luck?

If we assume the 400 ingots cast under the modified process represent a random sample from all possible ingots made with this process, the conditions are for inference are reasonable (.17*400 = 68 and .83*400 = 332).

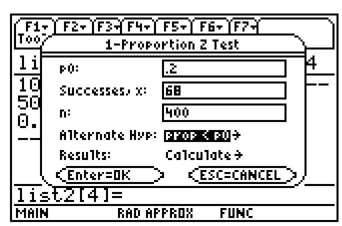
TI-83/84 Procedure

Press STAT and arrow to `TESTS`. Select choice `5:1-PropZTest` by either pressing 5 or arrowing to the selection and pressing ENTER. We are asked for p_0, the claimed value which is 20% in this case. Enter the proportion as the decimal 0.2. Then we need the number of "successes" (multiplying 0.17*400 = 68). On TI-83/84 calculators, the multiplication can be entered directly by the `x:` prompt. Press ENTER to do the computation, then press ▲ to round the result if needed. If you are using a TI-89, you must do the multiplication on the home screen. The number of trials, n, is 400. Then we need a direction for the alternate hypothesis (what we hope to show). This is usually obtained from the form of the question. In this case, we want to know if the proportion of cracked ingots has *decreased* which indicates we want the alternative $< p_0$. Move the blinking cursor to highlight this selection (if needed) and press ENTER to move the highlight. Finally, there are two choices for output. Selecting `Calculate` merely gives the results. Selecting `Draw` draws the normal curve and shades in the area corresponding to the p-value of the test. The TI-83/84 input screen for the test is at right.

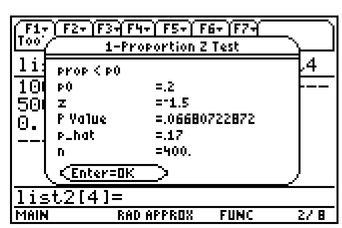


TI-89 Procedure

Hypothesis tests are located on the F6 = 2nd F1 `Tests` menu in the Statistics/List Editor Application. The input screen resembles that for the TI-83/84 series. Use the right arrow key to expand the alternate hypothesis choices and use the up or down arrows to select the one you want. Press ENTER to make the selection. You can similarly select `Calculate` or `Draw` options for the output.

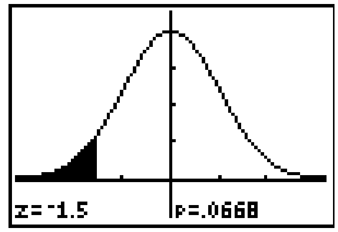


At right is the output from selecting `Calculate.` The first line of the output gives what the calculator understood the alternate hypothesis to be. Always check that this is what you intended, as it can make a difference in the p-value. The value of the test statistic is z = -1.5. This means that if there were no change, the observed 17% is 1.5 standard deviations above the mean. The p-value for the test is 0.067. This means that if the proportion of cracked ingots has not changed, we

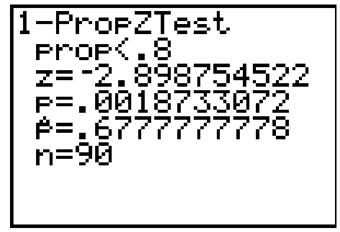


would observe a value of 17% or less only about 7 times in 100 samples of this size. Since this is not extremely rare, the factory management will have to make up their own minds about a demonstrated improvement in the process.

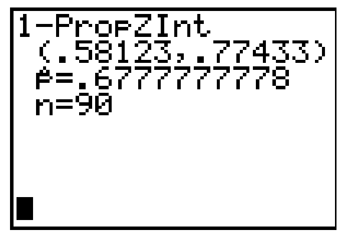
Here is the output when DRAW is selected. The test statistic and p-value are given. The area under the standard normal curve corresponding to the p-value is also shaded. Since the alternate hypothesis was $p < .20$, only area below the calculated test statistic corresponds to the p-value.



Another Example



The Department of Motor Vehicles claimed that over 80% of candidates pass the driving test, but a newspaper reporter's survey of 90 randomly selected local teens who had taken the test found only 61 had passed. The observed sample statistic is $\hat{p} = 61/90 = 0.678$. This is lower than the claimed passing rate. Is it enough lower to cast doubt on the claim made by the DMV? The teens were randomly selected, it's reasonable that they're independent of one another, they do represent less than 10% of teens who might take the test, and we have more than 10 each of failures and successes.

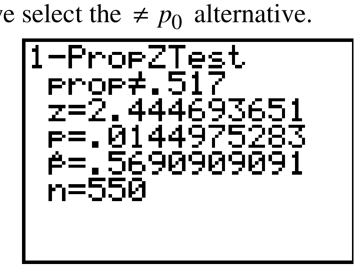


Here are the results. The z-statistic of -2.899 means that if the DMV is correct, the observed 67.8% is almost 3 standard deviations below the mean, which is extremely rare. The p-value is 0.0019. If the DMV were correct, we'd expect to see our sample results only about twice in 1000 samples. This extremely small p-value argues that these data show the DMV is wrong (at least for teen drivers.) What is the passing rate for teen drivers? Based on these data, we are 95% confident between 58.1% and 77.4% of teen drivers will pass the test. Notice that since the high end of the confidence interval is below 80%, this also supports our conclusion that the DMV's claim is wrong.

Yet another example

In some cultures, male children are valued more highly than females. In some countries with the advent of prenatal tests such as ultrasound, there is a fear that some parents will not carry pregnancies of girls to term. A study in Punjab, India[1] reports that in 1993 in one hospital 56.9% of the 550 live births were males. The authors report a baseline for this region of 51.7% male live births. Is the sample proportion of 56.9% evidence of a change in the percentage of male births?

The baseline proportion of males for the area is 51.7% in this case. This has been entered in decimal form as 0.517. Then we need the number of "successes" (Multiplying 0.569*550 gives 312.95 which rounds to 313). The multiplication can be entered directly by the x: prompt, then press ENTER, then up arrow to round the result if needed. The number of trials, n, is 550. Then we need a direction for the alternate hypothesis (what we hope to show). In this case, we want to know if the proportion has changed which means we select the $\neq p_0$ alternative.



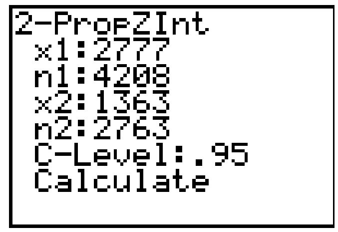
The value of the test statistic is z = 2.44. This means that if there were no change, the observed 56.9% is 2.44 standard deviations above the mean. The p-value for the test is 0.0145. This means that if the proportion of male births is still 51.7%, we would observe a value of 56.9% or greater only about 1 times in 100 samples. Since this is very rare, we will reject the null hypotheses and conclude that we believe that, based on these data, the true proportion of male births in Punjab is now greater than the baseline 51.7%.

---

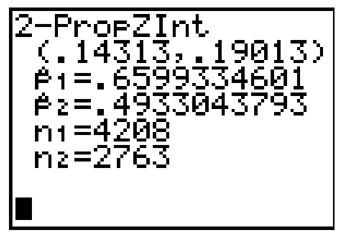
[1] "Fetal Sex determination in infants in Punjab, India: correlations and implications," E.E. Booth, M. Verna, R. S. Beri, *BMJ*, 1994; 309:1259-1261 (12 November).

CONFIDENCE INTERVALS FOR THE DIFFERENCE OF TWO PROPORTIONS

Do men take more risks than women? A recent seat belt study in Massachusetts found that, not surprisingly, male drivers wear seat belts less often than women, but that men's belt-wearing jumped more than 16 percentage points when a woman was a passenger[2] Seat belt use was recorded at 161 locations using random sampling methods developed by the National Highway Traffic Safety Administration (NHTSA). Of 4,208 male drivers with female passengers, 2,777 (66.0%) wore seat belts. But among the 2,763 male drivers with male passengers only 1,363 (49.3%) wore seat belts. What do we estimate the true gap in seat belt use to be? We want a confidence interval for the difference in the true proportions, $p_W - p_M$. If this interval contains 0, there is no statistical evidence of a difference. Note that the order makes no real difference, but since observed seat belt usage is higher with the woman passenger, most people tend to prefer positive differences.



From the `STAT`, `TESTS` menu select choice `B:2-PropZInt.` This is menu option 6 on the [F7] `Ints` menu on a TI-89. The calculator uses groups 1 and 2, not men and women, and calculates results based on $group1 - group2$. Decide (and keep note of) which number you assign to each group. Since it is usually easier to deal with positive numbers, we will call the women passengers group 1.



The calculated interval is 0.143 to 0.190. This means we are 95% confident, based on this poll, the proportion of men who buckle up with a woman passenger is between 14.3% and 19.0% more than the proportion of men who buckle up with a male passenger. Since 0 is not in the interval, there is a definite difference in male seat belt behavior due to the gender of the passenger (at least in Massachusetts.)

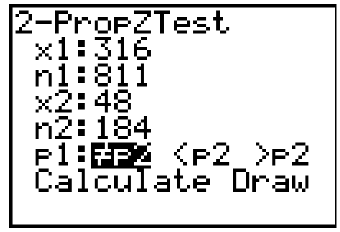
Note: A word of caution about these intervals: when looking at the results one must be clear which group was labeled Group 1 and which was Group 2. If I had called the male passengers Group 1 in the example above, the numerical results would be the same, but each end of the interval would be negative. In that case the interpretation would be that male drivers were between 14.3% and 19.0% *less* likely to buckle up with a male passenger than with a female.

HYPOTHESIS TESTS FOR A DIFFERENCE IN PROPORTIONS

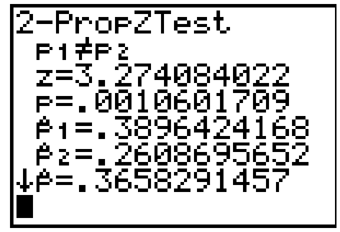
The National Sleep Foundation asked a random sample of 1010 U.S. adults questions about their sleep habits. The sample was selected in the fall of 2001 from random telephone numbers.[3] Of interest to us is the difference in the proportion of snorers by age group. The poll found that 26% of the 184 people age 30 or less reported snoring at least a few nights a week; 39% of the 811 people in the older group reported snoring. Is the observed difference of 13% real or merely due to sampling variation?

The null hypothesis is that there is no difference, which means $p_1 = p_2$ or $p_1 - p_2 = 0$. (The calculator and most computer statistics packages can only test assumed differences of 0; if there were an assumed difference, such as the belief that older people had 10% more snorers than young people, one would need to compute the test statistic "by hand.")

[2] Massachusetts Traffic Safety Research Program (June 2007)
[3] 2002 *Sleep in America Poll*, National Sleep Foundation. Washington D.C.

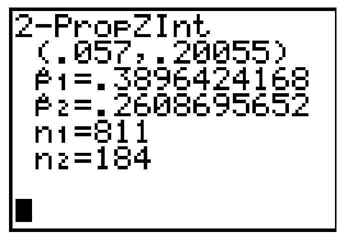


Decide which group will be Group 1. We will use the older people as Group 1. (There will be no difference in the results, but again, positive numbers are generally easier for most people to deal with.) From the STAT TESTS menu select choice 6:2-PropZTest. The number of snorers in the older group was 0.39*811 = 316.29 (rounded to 316); for the younger group the number of snorers was 0.26*184 = 47.84 which rounds to 48. The chosen alternate is $\neq p2$ since we just want to know if there is a difference.



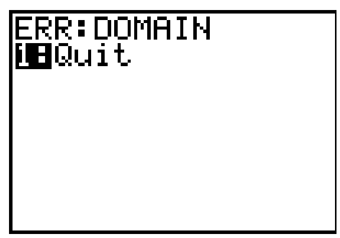
Here are the results. Be careful here: there are lots of p's floating around. We see the chosen alternative, $p1 \neq p2$, The test statistic is z =3.27, which means that if there were no difference in the proportion of snorers the observed difference (about 13%) is more than 3 standard deviations above the mean. The p-value for the test is given as $p = 0.0011$. The next values given are the observed proportions in each group, $\hat{p}_1$ and $\hat{p}_2$ then an overall $\hat{p}$ which represents the observed proportion, *if there were no difference in the groups*.

One can arrow down to see the sample sizes for the two groups as well. Since the p-value for the test is so small, we believe there is a difference in the rate of snorers based on this poll. We can further say the proportion of snorers is greater in older people than in those under 30 – their observed proportion (39.0%) is larger than for the observed proportion (26.1%) for the younger group.



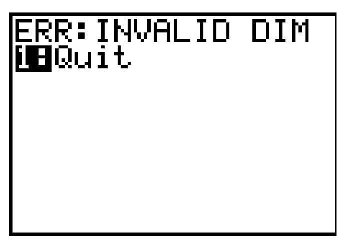
How big is the difference? We are 95% confident, based on this data the proportion of snorers in older adults is between 5.7% and 20.1% larger than for those under 30.

WHAT CAN GO WRONG?



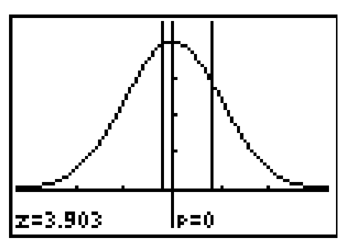
Err: Domain?

This error stems from one of two types of problems. Either a proportion was entered in a 1-PropZTest which was not in decimal form or the numbers of trials and/or successes was not an integer. Go back to the input screen and correct the problem.



Err:Invalid Dim?

This can be caused by selecting the DRAW option if another Statistics plot is turned on. Either go to the STAT PLOT menu ([2nd][Y=]) and turn off the plot or redo the test selecting CALCULATE.



What are the weird lines?

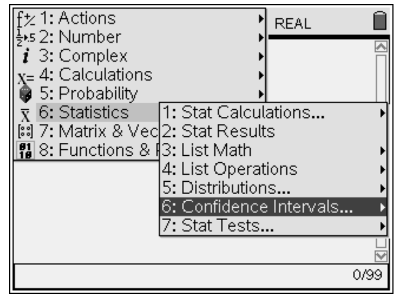
This is caused (as usual in graphing errors) by having an equation entered on the Y= screen or a STAT PLOT turned ON. Remember, the calculator always tries to plot everything it knows about. Clear the Y= screen, and turn STAT PLOTS off.

Bad Conclusions.

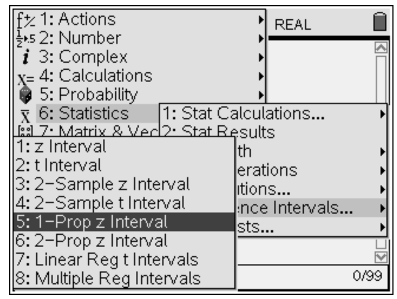
Small p-values for the test argue against the null hypothesis. If the p-value is small, one rejects the null hypothesis and believes the alternate is true. If the p-value is large, the null hypothesis is not rejected; this does *not* mean it is true – we simply haven't gotten enough evidence to show it's wrong. Be careful when writing conclusions to make them agree with the decision.

Commands for the TI-Nspire™ Handheld Calculator

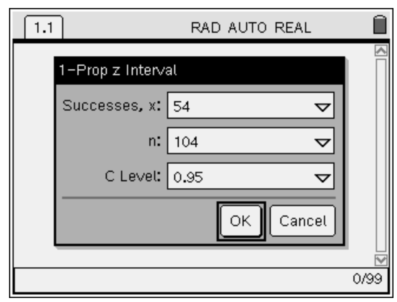
For statistical inference, start on a calculator page. For confidence intervals press (menu) , and then select Statistics and then confidence intervals.

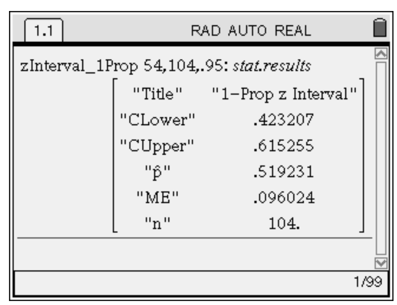


Select a one sample proportion.

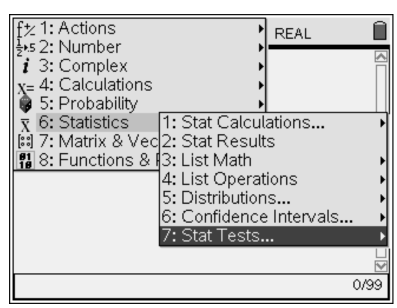


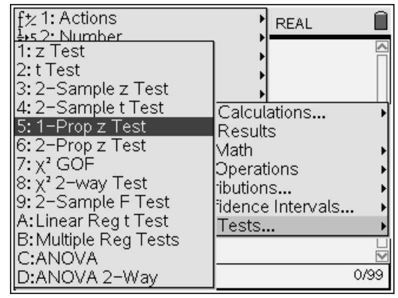
In the input box, type the number of successes, the number of trials, and the confidence level. For the Caribbean Sea example, use 54, 104, and .95.



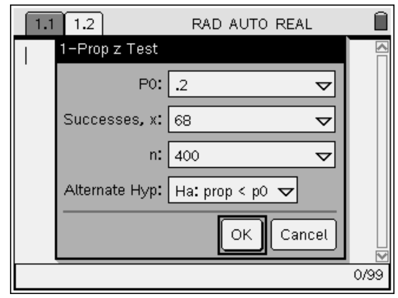


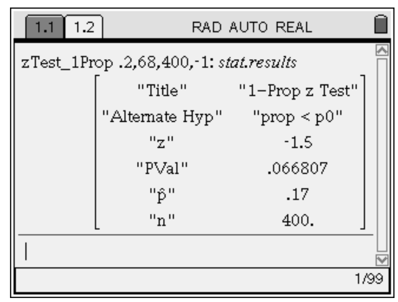
For a one sample proportion hypothesis test, press (menu), select Statistics, Stat Tests, and 1-Prop z Test.



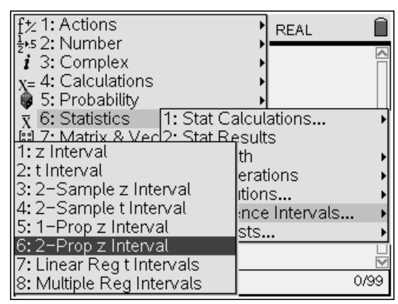


Type the information in the input box. The values for the Ingots example are shown.

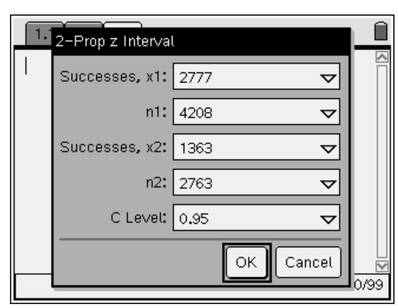


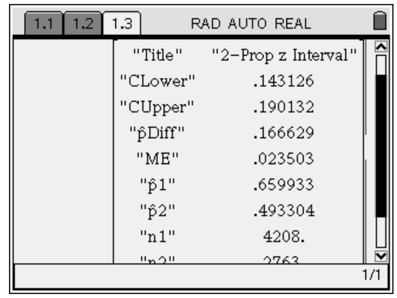


For a two sample proportion confidence interval, start on a calculator page. Press (menu) , select Statistics, Confidence intervals, and 2- Prop z Interval.

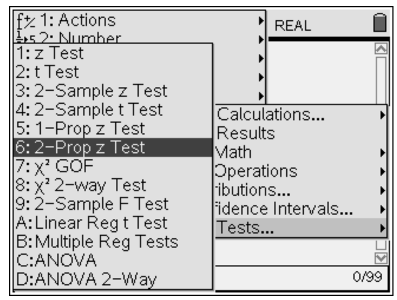


Type the information in the input box. The values for the seat belts example are shown.





For a two sample proportion hypothesis test, start on a calculator page. Press (menu) , and then select Statistics, Stat Tests, and 2- Prop z Test.



Type the information in the input box. The values for the sleep example are shown.