CONFIDENCE INTERVALS

Question 1.

The data below are taken from the paper entitled: *Mothers' dental attendance patterns and their children's dental attendance and dental health*, Gratrix *et al.*, 1990, BDJ 168: 441-443). It shows the mean DMFT (total of decayed, missing and filled teeth) of three groups of children along with their standard deviations.

Children's place of attendance	Number of children	Mean DMFT	Standard deviation of DMFT
General Dental Surgery	122	1.75	1.71
Community Dental Surgery	34	1.71	1.64
Non-attenders	21	1.52	1.72

a. Calculate the 95% confidence interval for each of the means:
  - Formula: $\left[\overline{X} - Z_{\alpha}\frac{s}{\sqrt{n}}, \overline{X} + Z_{\alpha}\frac{s}{\sqrt{n}}\right]$
  - 95% Confidence interval for General Dental Surgery: [1.75-1.96*(1.71/sqrt(122)); 1.75+1.96*(1.71/sqrt(122))] = [1.4466; 2.0534]
  - 95% Confidence interval for Community Dental Surgery: [1.71-1.96*(1.64/sqrt(34)); 1.71+1.96*(1.64/sqrt(34))] = [1.1587; 2.2613]
  - 95% Confidence interval for Non-attenders: [1.52-1.96*(1.72/sqrt(21)); 1.52+1.96*(1.72/sqrt(21))] = [0.7843; 2.2557]

b. Assuming that DMFT is a normally distributed variable, between what values would we expect 95% of the DMFT values for the 122 children attending the General Dental Surgery to lie?
  - Formula: m ± 1.96·s
  - The range = (1.75 – 1.96*1.71); (1.75 + 1.96*1.71)
  - The range = from -1.60 to 5.10

c. Consider your answer to the second part of this question, bearing in mind what DMFT is. What can you deduce about the distribution of DMFT? Is it legitimate or useful to carry out the calculations you did in the first part of the question?
  - The answer to question b) stated that 95% of the GDS population would have DMFT values lying between -1.60 and 5.10. As DMFT is the total number of decayed, missing and filled teeth it is clearly impossible for this to be a negative number. This means that one of the assumptions I made in my calculations must have been wrong. The most likely error is that DMFT is not a normally distributed variable so we would not expect 95% of a population's values to lie in the interval 'Mean - 1·96 standard deviations' to 'Mean + 1·96 standard deviations'.
  - It could be concluded that DMFT is not normally distributed. It could be guess that it has a positively skewed distribution as its mean (1.75) is close to its minimum possible value (0).
  - If DMFT is not normally distributed then the results from question a) are not so useful. It is mathematically legitimate to calculate the mean for a skewed distribution and the 95% confidence intervals will also be mathematically correct. In terms of clinical interpretation the median is a much more useful and accurate measure of location for skewed distribution.

Source of exercises: http://www.brettscaife.net/statistics/introstat/04infer/exercise.html

- For this problem it would be better to calculate the median with its 95% confidence interval.

Question 2.

True or false?

a. The 95% confidence interval is wider than the 90% confidence interval.

True

If we want to be more confident (95% confident rather than 90% confident) that our range of plausible values includes the true value then it makes sense that the interval has to be wider.

b. We would expect about 95% of a sample to lie within the 95% confidence interval of the sample's mean.

False

This statement is confusing standard deviations and standard errors. We would expect 95% of a normally distributed population to lie within the interval

'population mean - 1·96 standard deviations'

to 'population mean + 1·96 standard deviations'

A confidence interval is an expression of how confident we are about the value of a single quantity (such as the mean) not a description of a population or sample.

c. There is a 95% probability that the true value of the mean lies within the 95% confidence interval.

Just about true

In loose language there is nothing too wrong about this statement; but statisticians do not like it. The reason for this is it seems to imply that the mean that could have any of the possible values within the confidence interval. This is not the case: the mean only has one value, although we do not know it. A preferred statement would be "the 95% confidence interval has a 95% chance of including the mean".

Another way to think of it is to consider repeating the experiment. We would generally get a slightly different 95% confidence interval to the first experiment. The confidence interval changes but, obviously, the underlying (unknown) mean does not. Consequently, we prefer to make sure that any words that imply probability or variability are clearly attached to the words 'confidence interval' and not 'mean'.

d. A confidence interval is symmetric about the mean.

Frequently true

This is certainly true for the confidence intervals about the mean and the other confidence intervals you will calculate in these exercises. But there are also non-symmetric confidence intervals.

e. The 95% confidence interval is a plausible range of values for the population mean.

True

This is a good way to express the idea of a 95% confidence interval.

f. If we repeatedly sampled a population and constructed a 99% confidence interval for the mean of each sample then, in the long run, 99% of the confidence intervals would include the true value of the mean. If the statement is false explain why.

True

An absolutely accurate description of the way confidence intervals work.

Source of exercises: http://www.brettscaife.net/statistics/introstat/04infer/exercise.html

Question 3.
How wide would a 100% confidence interval be?

A 100% confidence interval would, in general, be infinitely wide.
The only way to be 100% sure of an estimate of anything is to include all of the possible values that it could take.
There will be some cases (for non-normal data) where a 100% confidence interval would not be infinitely wide. If we consider the estimate of the mean number of teeth that a population has. We know that the number of teeth can't be less than zero so the lower end of a 100% confidence interval will be zero. The upper end of the 100% confidence interval would be 32 if we were sure that there were no genetic oddities in our population. As we can't be certain of this we would have to find out what the world record for number of teeth is and set this as our upper limit.
The following table shows the numbers we multiply the standard error by when calculating various confidence intervals. Notice how the numbers increase much more rapidly than the increase in confidence level. The case of 100% confidence requiring an infinite interval is the extreme end of this series.

Question 4.
The data below are taken from the paper: *Fluoride release from glass-ionomer and compomer restorative materials: 6 month data* (Shaw *et al.*, 1998, Journal of Dentistry 26: 355-359). It shows the amount of fluoride released by three restorative materials (in vitro) after 6 months

Material	Number of specimens	Mean fluoride release (mg mm^{-2})	Standard deviation
Ketac-Fil	5	30.62	4.85
Chem-Fil Superior	5	12.69	2.55
Compoglass	5	10.35	1.02

Calculate the 95% confidence intervals for each of the means.

- 95% Confidence interval for Ketac-Fil: [30.62-1.96*(4.85/sqrt(5)); 30.62-1.96*(4.85/sqrt(5))] = [26.3688; 34.8712]
- 95% Confidence interval for Chem-Fil Superior: [12.69-1.96*(2.55/sqrt(5)); 12.69+1.96*(2.55/sqrt(5))] = [10.4548; 14.9252]
- 95% Confidence interval for Compoglass: [10.35-1.96*(1.02/sqrt(5)); 10.35+1.96*(1.02/sqrt(5))] = [9.4559; 11.2441]

What can you infer about how the different materials compare in the light of your results?

The 95% confidence intervals for *Chem-Fil Superior* and *Compoglass* overlap by a considerable amount. For example, a fluoride leakage of 11mg mm^{-2} is a plausible value for the means of both materials. We are unable to tell if there is a real difference between the two materials. By contrast none of the range values in the confidence interval for *Ketac-Fil* would be a plausible value for the mean leakage of either of the other two materials. We might conclude that there is a real difference between *Ketac-Fil* and the other two materials
This method of looking at the confidence intervals of the means of different samples is a good first step to seeing if a real difference is likely to exist between them. There is a more formal and correct approach: statistically testing for a difference and calculating a confidence interval for the difference.

Source of exercises: http://www.brettscaife.net/statistics/introstat/04infer/exercise.html

Question 5.

The data below are taken from the same paper referred to in *question 4*. They show the numbers of three different types of restoration included in the study and how many failed within 8 years. (Each of the materials is given a code letter to make it easier to refer to it in calculations and tables.)

Material	Code	Number of restorations	Number of failed restorations
Microfilled composite	M	55	9
Fine particle hybrid composite	C	52	8
Relatively coarse particle hybrid composite	P	54	5

Calculate the proportion of failures for each material, along with their confidence intervals.

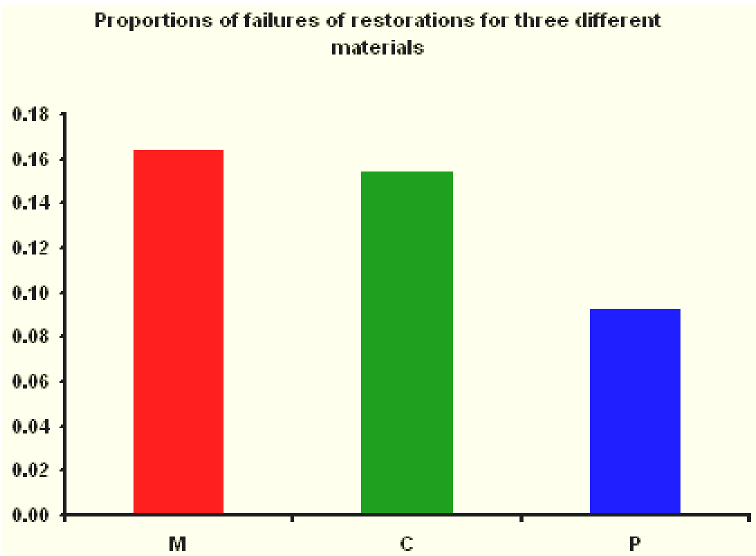
- Formula: $\left[f - Z_\alpha\sqrt{\frac{f(1-f)}{n}}; f + Z_\alpha\sqrt{\frac{f(1-f)}{n}}\right]$
- Computations

Code	f	95%CI - lower bound	95%CI - upper bound
M	=9/55 = 0.1636	= 0.1636-1.96*sqrt(0.1636*(1-0.1636)/55) = 0.0659	= 0.1636+1.96*sqrt(0.1636*(1-0.1636)/55) = 0.2614
C	=8/52 =0.1538	= 0.1538-1.96*sqrt(0.1538*(1-0.1538)/52) = 0.0558	= 0.1538+1.96*sqrt(0.1538*(1-0.1538)/52) = 0.2519
P	=5/54 =0.0926	= 0.0926-1.96*sqrt(0.0926*(1-0.0926)/54) = 0.0153	= 0.0926+1.96*sqrt(0.0926*(1-0.0926)/54) = 0.1699

Strictly speaking our confidence interval for material P is invalid. The formula we have used for the standard error is an approximation. It is a good approximation in almost all circumstances but it should not really be applied if the number of failures is 5 or less. (Or if 'sample size minus the number of failures' is 5 or less.) In this case we will carry on as if it does not matter. If you come across this problem in your own research you'll have to seek further advice.

Draw a graph or diagram which presents your results in a suitable manner.

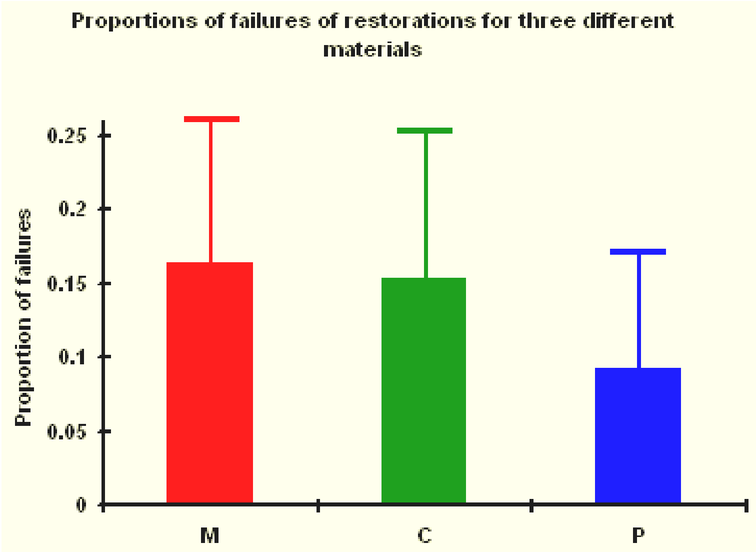
Now we need a diagram which describes our results in an honest and useful manner. You will sometimes see the results for a proportion (or a mean) presented in the style of a bar chart as so:



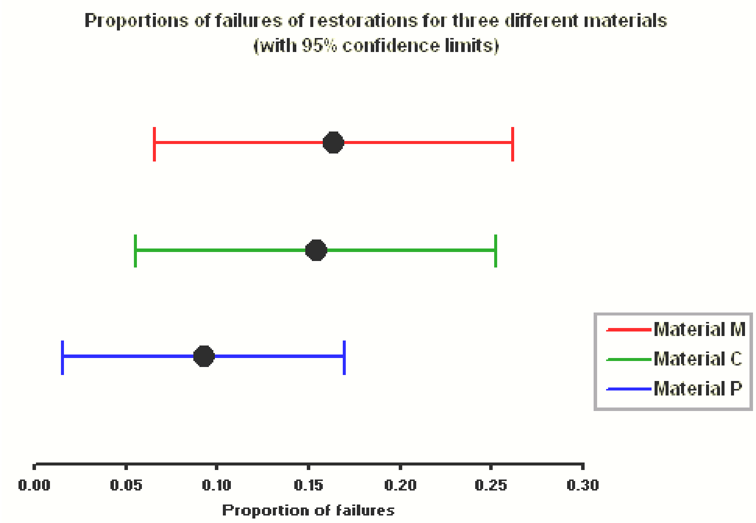
This is an incorrect way to present this sort of data. Firstly, it has no representation of how good the estimates of the means are: no confidence intervals. Secondly, bar chart type graphs are not suitable for presenting the values of single quantities such as proportions or

Source of exercises: http://www.brettscaife.net/statistics/introstat/04infer/exercise.html

means. They are best suited to representing *counts*. The brain's first interpretation of a bar chart tends to be along these lines.
It is also difficult to represent the precision of an estimate on a bar chart. Sometimes charts such as this are seen:



(These sort of plots are sometimes known as *detonator* plots after their resemblance to a classical plunger-type explosives detonator.) There is an attempt to represent the confidence interval here but it only goes one way! There is yet another version where a mirror image of the plunger goes down below the top of the column. At this point the graph starts to get cluttered and the effort of reading it starts to counteract the prime purpose of the graph: to simplify presentation.
A much better way of presenting the results is as so:



Here we see a clear point representing the proportion observed and lines representing the confidence intervals. It represents the data in a way that is easy to use. It aids our interpretation of the figures we calculated above.

Interpret your results.

It is clear that we are unable to see if there is any real difference between the three types of material. There is a wide range of values (from about 0·07 to about 0·17) which could be the proportion of failures for all three materials.
Note that there is a formal statistical test that can be carried out to confirm whether or not the differences seen between two proportions are significant.

Source of exercises: http://www.brettscaife.net/statistics/introstat/04infer/exercise.html

This result is not particularly surprising. If we want to detect the difference between two proportions then either the differences have to be very large or our sample sizes have to be very large.

In this case we might wonder if the difference we saw in failure rates for material P (about 9%) and the other two materials (about 15%) was a real difference, and the only reason we didn't see it as significant was because of the small sample size. We may want to recommend a further study to see if there really is a 6% difference in failure rates between material P and the other two materials. I calculated that we would need about 450 to 500 patients in each group to stand a reasonable chance of spotting this sort of difference. Obviously, we would have to be convinced of the potential clinical benefits of a 6% improvement in failure rates before we started such a large study. We will be looking at the problem of how big a sample needs to be in a future unit

Source of exercises: http://www.brettscaife.net/statistics/introstat/04infer/exercise.html