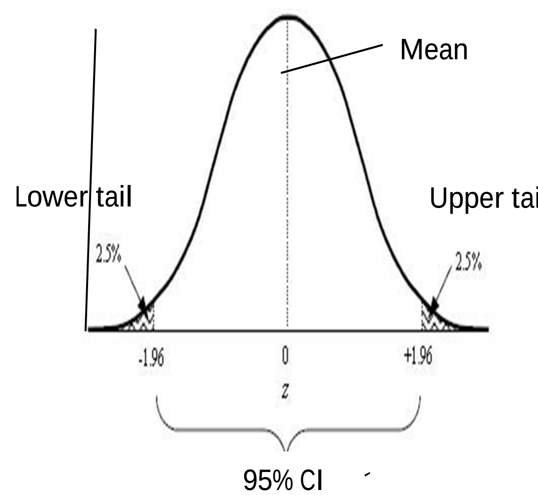
95% Confidence Interval

In epidemiology we measure quantities such as population characteristics e.g. age, sex, education etc. We also quantify disease using measures of occurrence e.g. prevalence, incidence, odds etc.; and to investigate the effect of risk factors, we calculate measures of effect such as risk ratio, odds ratio, etc. But how can we be sure that our results reflect the situation in the population? We usually do not collect information from all the population but from a sample of the population, therefore our results will not be exactly true. Imagine we wanted to estimate prevalence of smoking amongst 14000 students of FTE Sulaimani by collecting information from only 500 students. Imagine we found that the mean age of the students was 19 years and that 5% of students were smokers. Since we have not collected information from all students, these observed results of 19 years (mean age) and 5% (prevalence of smoking) will not be strictly correct i.e. if we collect information from all students rather than a sample, we will probably get a different result. But how close to the truth are our results? To what degree we can be certain (confident) that the result which we have observed from our study (the sample) is correct to all students (the population)? The 95% confidence interval answers this question.

A 95% confidence interval includes a range of values of the variable (e.g. mean age) within which the true population value of the variable will fall 95 times if we repeated the experiment 100 times. In other words we can be sure 95% that the true value (e.g. mean age) of the population will be within the range provided by the 95% confidence interval. Therefore 95% confidence interval is about generalization of the results on the population (statistical inference). Whenever we want to generalize our results on the larger population we have to calculate 95% interval.



As shown in the normal plot opposite, the 95% confidence interval falls within 2 standard deviations (or exactly 1.96 standard deviations) from the parameter estimate (e.g. the mean or the proportion). It covers 95% of the values i.e. the white area under the curve. 2.5% of the values fall outside the 9%% CI to the left and another 2.5% fall outside the 95% CI to the right. To calculate the 95% confidence interval, we have to know 2 things 1) the parameter estimate e.g. the mean or proportion, 2) the standard deviation (the standard error)

95% confidence interval for a mean

Whenever we estimate a sample mean and we want to generalize our results on the larger population i.e. we want to report the population mean, we have to calculate the 95% confidence interval of the mean. The sample mean is denoted as $\bar{x}$ and the population mean as μ. To calculate 95% confidence interval of μ we need to calculate the standard error (SE). Standard error is the standard deviation of the population.

Standard error of a mean is calculated from standard deviation of the sample mean (σ) divided by the square root of the sample size (n).

$$SE(\bar{x}) = \sigma / \sqrt{n}$$

Example: in a study of female genital mutilation (FGM) in Kurdistan, 348 females (sample size) were studied, the mean age of performing FGM was 4.6 years and the standard deviation was 2.4 years. Calculate the standard error of the mean age of FGM in Kurdistan.

SE= SD/√n = 2.4/√348=0.13

So the standard error of the population mean age of FGM is 0.13 years

Now, calculate the 95% confidence interval (95% CI) for the mean age

$$95\%\ CI\ (mean) = mean \pm 1.96 * standard\ error$$

95% CI (mean FGM age)=mean age ± 1.96*0.13

If we subtract (1.96*0.13) from the mean age we get the lower limit of the confidence interval
If we add the (1.96*0.13) to the mean age we get the upper limit of the confidence interval

Lower limit =4.6- (1.96*0.13)= 4.6 -0.26= 4.3 years

Upper limit =4.6- (1.96*0.13)= 4.6 + 0.26=4.9 years

Therefore the 95% confidence interval of the mean age of FGM will be from 4.3 years to 4.9 years

What does this mean? This means that we can be 95% of the time confident that the mean age of performing FGM in Kurdistan is somewhere between 4.2 years to 4.9 years.

Question
-What does happen to the 95% CI if the sample size is larger?
The 95% CI will be narrower because the SE will be smaller and therefore the amount we add to or subtract from the mean (i.e. two standard errors) will be smaller.

-What does happen to the 95% CI if the sample size is smaller?
The 95% CI will be wider because the SE will be bigger and therefore the amount we add to or subtract from the mean (i.e. two standard errors) will be bigger.

Question
The data below belong to the age of pre-school children who were burnt in 2008 in Sulaimani. Calculate 95% CI of the mean age of children who get burnt in Sulaimani:

Variable	Sample size	Mean	Std. Dev.
Age	944	2.3	1.3