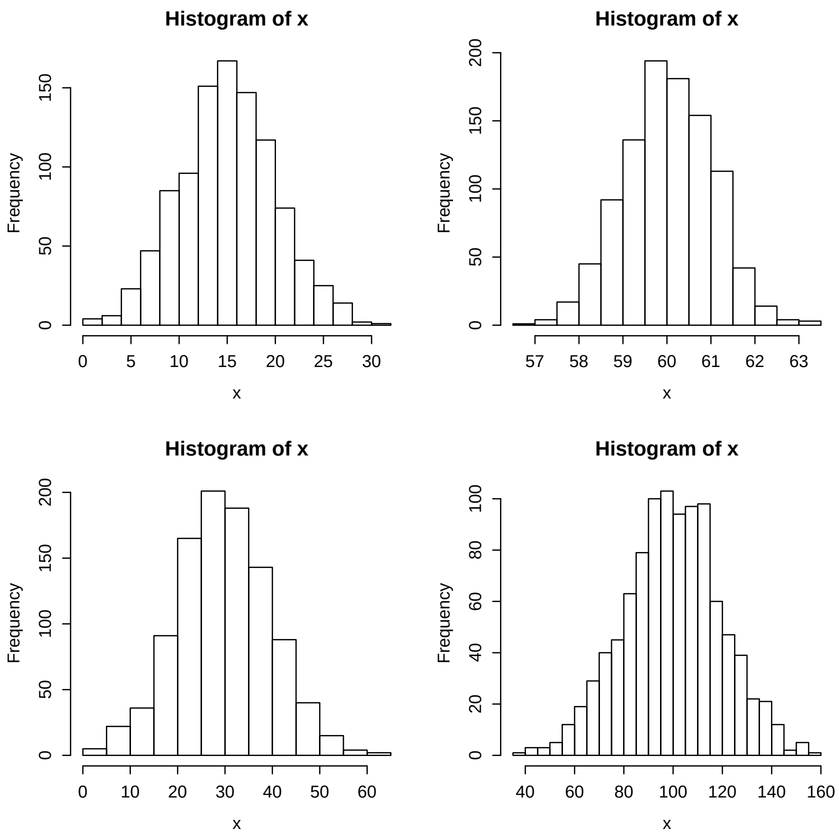
STM4PSD – Workshop 7

Histograms and sampling data

1. In the histograms below, $n = 1000$ observations have been randomly generated from a $N(\mu, \sigma^2)$ distribution for different choices of μ and σ. In no particular order, the μ parameters used for each are one of $15, 30, 60, 100$ and the σ parameter for each are one of $1, 5, 10, 20$.



Match the μ and σ parameters that must have been used to generate the sample data used for each histogram.

Top-left histogram: $\mu = \ldots\ldots$ and $\sigma = \ldots\ldots$
Top-right histogram: $\mu = \ldots\ldots$ and $\sigma = \ldots\ldots$
Bottom-left histogram: $\mu = \ldots\ldots$ and $\sigma = \ldots\ldots$
Bottom-right histogram: $\mu = \ldots\ldots$ and $\sigma = \ldots\ldots$

Confidence intervals

2. This question is based on a data set used in the *Mondolez International* Six Sigma Training workshop.

   The company Kraft wanted to see whether they should invest in the training of staff at certain locations to improve efficiency. Of interest is the time required (in hours) for designing advertisements for product promotions. In the initial stages, data from the New York office was collected with a mean design time of 20.19 hours and a sample standard deviation of 3.88. The sample size is $n = 43$.

   (a) Calculate the standard error for the mean.

   (b) Calculate an approximate 95% confidence interval for the mean design time in the New York office.

   (c) Do you think your confidence interval is narrow enough so that Kraft will find it useful? If not, what could Kraft do to obtain a narrower interval?

(d) Suppose that you need to report these results back to someone in the Kraft Head Office and need to do so in a simple-to-understand way. What would you say?

(e) Suppose that a Kraft employee who visited the New York office believes that the New York design team is slow, claiming that the average design could be 25 hours or more? Do you think your interval supports this employees claim? Justify your answer.

3. Suppose you were monitoring website traffic, and wanted to determine the proportion of visitors that click on a certain link. You randomly sampled 149 visitors, and found that 35 of them clicked on this link. Determine an approximate 95% confidence interval for the proportion p of visitors that click this link.

Although it is usually not wise to apply the theory of Chapter 8 to small sample sizes, in the following questions you will perform the calculations to some small samples just to familiarise yourself with the calculations, and observe the effect that different sample sizes can have on confidence intervals.

4. Suppose that you were investigating the shelf life of a certain product, and four items were randomly sampled. In days, the shelf life for each item was 14.3, 20.2, 13.5 and 17.4, respectively.

   (a) Calculate the sample mean $\overline{x}$ for this sample.

   (b) Calculate the sample variance s^2 for this sample.

   (c) Hence determine the standard error SE for the sample mean.

   (d) Using Result 8.7.1, give an approximate 95% confidence interval for the population mean μ.

5. Suppose instead that six products were sampled. Assume that the first four samples are the same as in the previous question, and that the fifth and sixth observed shelf lives were 20.0 days and 15.2 days, respectively.

   (a) Calculate the sample mean $\overline{x}$ for this sample. Do not try to re-use answers from Question 3 in your calculations; you will need to perform the calculations from the beginning.

   (b) Calculate the sample variance s^2 for this sample.

   (c) Hence determine the standard error SE for the sample mean.

   (d) Using Result 8.7.1, give an approximate 95% confidence interval for the population mean μ.

   (e) You should find that four of the six data points are not contained in the confidence interval. In other words, only 33% of the data is in the 95% confidence interval. Is this contradictory? Explain.

6. Suppose that, once again, the first four samples are the same as in Question 3, but this time the fifth and sixth samples are 12.2 and 23.3, respectively.

   (a) Using similar steps as above, determine an approximate 95% confidence interval for the population mean.

   (b) Compare this confidence interval with the ones you found in Questions 3 and 4. In particular, compare the sizes of each. Do you observe anything unusual?

7. A table of quantiles for $Z \sim N(0, 1)$ is given below:

p	0.8	0.85	0.9	0.95	0.975
$Q_Z(p)$	0.84	1.04	1.28	1.64	1.96

   Determine approximate 80% confidence intervals and 90% confidence intervals for the sample used in Question 5. Note that you do not need to recalculate the sample mean and standard error. You should refer to Result 8.7.3 if necessary.