EXERCISE 1.26 (FTC cigarette rankings)

(a) The relative frequency histogram, based on nicotine contents of 500 cigarette brands, is fairly mound shaped with the highest frequency at 0.975. As the sample size is quite large, the rule of thumb suggests that approximately 95% of the measurements will lie within 2-standard deviations of their mean.

(b) \( \bar{y} = 0.8425, s = 0.3455250 \)

So, the interval \( \bar{y} \pm 2s \) comes out to be \((0.15145, 1.53355)\)

(c) Based on the answer in part (a), it can be said that approximately 95% of cigarettes have nicotine content that falls within the interval formed in part (b).

(d) Looking carefully at the given relative frequency histogram, it can be seen that 94% of the observed nicotine levels in cigarette brands fall within the interval formed in part (b). Thus it can be concluded that the answer agrees quite well with our estimation in part (c).

---

EXERCISE 1.46 (Impact of cooking on air particles)

Firstly it should be noted that the confidence interval is NOT \( \bar{y} \pm 2s \).

The 100(1 - \( \alpha \))% confidence interval for the population mean is given by

\[ \bar{y} \pm t_{\alpha/2 : n-1} \left( \frac{s}{\sqrt{n}} \right) \]

If you are not sure why the cut-off point of a \( t \)-distribution is used instead of that of a standard normal distribution, you can go through the little derivation below (or you can skip it, if you don’t like it 😊)

We know if \( \{y_i : i = 1(1)n\} \) are independently and identically distributed as \( N(\mu, \sigma^2) \) random variables, then

\[ Z = \frac{\bar{y} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1) \]
Since \( \sigma^2 \) is unknown, we replace it by its unbiased estimator \( s^2 \) (sample variance, with denominator \( n - 1 \)).

We also know that

\[
\chi^2 = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}
\]

Thus,

\[
\frac{\bar{y} - \mu}{s/\sqrt{n}} = \frac{\bar{y} - \mu}{\sigma/\sqrt{n}} = \frac{Z}{\sqrt{\frac{s^2}{\sigma^2}}} \sim t_{n-1}
\]

Thus \( 100(1 - \alpha)\% \) confidence interval for \( \mu \) is obtained as

\[
1 - \alpha = P \left( -t_{\alpha/2 : n-1} < \frac{\bar{y} - \mu}{s/\sqrt{n}} < t_{\alpha/2 : n-1} \right)
\]

This implies,

\[
1 - \alpha = P \left\{ \bar{y} - t_{\alpha/2 : n-1} \left( \frac{s}{\sqrt{n}} \right) < \mu < \bar{y} + t_{\alpha/2 : n-1} \left( \frac{s}{\sqrt{n}} \right) \right\}
\]

**Interpretation of the above probability statement:** the probability that the random interval \( \bar{y} \pm t_{\alpha/2 : n-1} \left( \frac{s}{\sqrt{n}} \right) \) contains the true population mean is equal to \( 1 - \alpha \).

**NOTE:** It should be noted that \( \mu \) here is not a random variable; it is just a parameter. So it would be wrong if one interprets the above probability statement as the probability that \( \mu \) lies in the interval is \( 1 - \alpha \). \( \mu \) is just a parameter and cannot have any probability statements attached to it.

(a) Based on the given data, \( \bar{y} = 1.0733, s = 0.23157, t_{0.025 : 5} = 2.571 \)

So the 95\% confidence interval for the mean decay rate is obtained as \( (0.83027, 1.316395) \).
**Interpretation:** One can be 95% confident that the interval (0.83027, 1.316395) will cover the true value of the mean decay rate of fine particles produced from oven cooking or toasting.

(b) Here the ‘95% confidence level’ implies that if we were to employ our interval estimator on repeated occasions, 95% of the intervals constructed on each of those repetitions would contain the true population mean. (One must realize that the interval obtained in part (a) has no significance as such; with each repetition we will get new data points and hence new confidence intervals. So when it comes to an interval containing the true population mean based on another trial, there will be a new confidence interval which will cover the true population mean with 95% confidence; the interval in (a) will be of no use in that case.)

**NOTE:** I have put a CROSS beside this part, if you have written that the interval in (a) would cover the true value of the population mean with 95% confidence on repeated occasions, or if you have been vague about the interval, just to point out how grave the mistake actually is.

(c) For this inference to be valid, the distribution of the population of decay rates should be *normally distributed.*