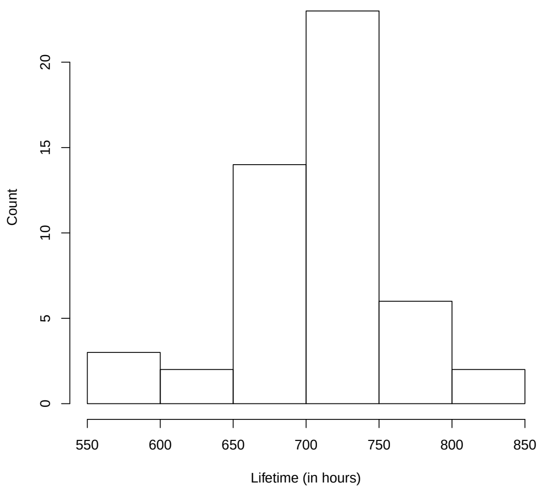
Note: This homework assignment covers Chapter 7.

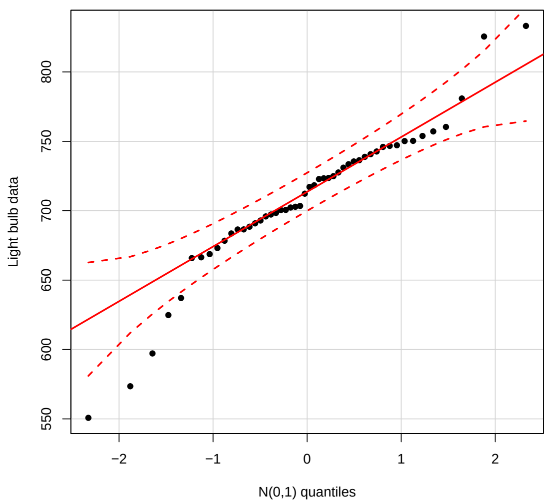
Disclaimer: If you use R, include all R code and output as attachments. Do not just "write in" the R code you used. Also, don't just write the answer and say this is what R gave you. If my grader can't see how you got an answer, it is wrong. I want to see your code and your answers accompanying your code (like in the notes).

1. A new brand of light bulb is advertised as having a population mean lifetime of 750 hours. The price of these bulbs is favorable, so a potential customer has decided to go ahead with a purchase arrangement unless it can be conclusively demonstrated that the population mean lifetime μ is smaller than what is advertised. A random sample of $n = 50$ bulbs was selected and the lifetime (Y, measured in hours) of each bulb was determined. The following R output gives the summary statistics:

```
> mean(bulbs)
[1] 708.91
> sd(bulbs)
[1] 52.44
```

A histogram and normal quantile-quantile (qq) plot are shown below.





(a) Using the sample information, write a 95 percent confidence interval for the population mean lifetime μ. Interpret the interval using plain English. On the basis of the confidence interval, how would you advise the customer to proceed (i.e., whether s/he should go ahead with the purchase arrangement)?
(b) The customer is concerned that the underlying assumptions required for part (a) are not met. Write a short discussion that addresses this point.
(c) In addition to the population mean lifetime, the customer is also concerned with excessive variation in the light bulb lifetimes. Write a 95 percent confidence interval for the population standard deviation σ and interpret the interval. What about the underlying assumptions in this part? Are you concerned?

2. The United States Army recently commissioned a study to assess how deeply a bullet penetrates ceramic body armor. In the study, a cylindrical clay model was layered under an armor vest. A projective was then fired, causing an indentation in the clay. The deepest impression in the clay was measured as an indication of survivability of someone wearing the armor. Here are the data that were observed; measurements (Y, measured in mm) were made using a manually controlled digital caliper. The data are already ordered from low to high.

22.4	23.6	24.0	24.9	25.5	25.6	25.8	26.1	26.4	26.7	27.4	27.6	28.3
29.0	29.1	29.6	29.7	29.8	29.9	30.0	30.4	30.5	30.7	30.7	31.0	31.0
31.4	31.6	31.7	31.9	31.9	32.0	32.1	32.4	32.5	32.5	32.6	32.9	33.1
33.3	33.5	33.5	33.5	33.5	33.6	33.6	33.8	33.9	34.1	34.2	34.6	34.6
35.0	35.2	35.2	35.4	35.4	35.4	35.5	35.7	35.8	36.0	36.0	36.0	36.1
36.1	36.2	36.4	36.6	37.0	37.4	37.5	37.5	38.0	38.7	38.8	39.8	41.0
42.0	42.1	44.6	48.3	55.0								

(a) Using the sample information, calculate a 95 percent confidence interval for the population mean μ and interpret your interval. Explain clearly (in words) what μ represents in the context of this problem. State precisely the assumptions needed for your interval to make sense. Check assumptions as warranted (discuss robustness).
(b) A co-worker of yours looks at your interval in part (a) and says, *"Ninety-five percent of the data values above should be in this interval."* How would you respond to him?
(c) Calculate a 95 percent confidence interval for the population standard deviation σ and interpret your interval. Explain clearly (in words) what σ represents in the context of this problem. State precisely the assumptions needed for your interval to make sense. Check assumptions as warranted (discuss robustness).

3. A random sample of 80 blood donations at a large blood bank reveals that 42 were type A blood.
(a) Do these data suggest that the percentage of type A donations differs from 40%, the percentage of the population having type A blood? Answer this question by using the sample information above and writing a 95 percent confidence interval for p, the population proportion. Interpret your interval and discuss your findings; also check the rules of thumb needed to ensure your confidence interval is valid.
(b) The 40 percent figure cited in part (a) is the population percentage for Caucasians only. For the African-American population, determine a sample size needed to estimate the population proportion p of type A blood individuals. Your requirements are that a confidence interval for p must be no wider than 0.05 and you want to have 99 percent confidence.

4. In some instances, it is not desired to estimate a population mean μ with a confidence interval; instead, it is desired to *predict* the value of a future observation, say Y^*. As an example, suppose that a meat inspector is inspecting packages of "low-fat" beef.

- An estimation problem might involve estimating the mean percentage of fat content in the population of packages prepared in a given week.
- A prediction problem would involve predicting the fat percentage for the next package prepared.

On the surface, these might sound like the same problem. However, they are different. In the first problem, we are trying to estimate a parameter μ that describes the entire population. In the second problem, we are trying to predict the value of one individual Y^* from the population.

It is easy to obtain a **prediction interval** for the one-sample problem we have considered so far. Suppose that $Y_1, Y_2, ..., Y_n$ is a random sample from a $\mathcal{N}(\mu, \sigma^2)$ distribution. A natural point estimator for the population mean μ is $\overline{Y}$, and, in fact, we know that

$$\overline{Y} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

However, to predict a new observation Y^*, not only do we need to account for the variation (σ^2/n) in estimating the population mean μ, but we additionally need to account for the variation (σ^2) attached to the new value Y^*. We combine these two sources of variability and this leads us to the following distributional result:

$$Y^* - \overline{Y} \sim \mathcal{N}\left(0,\ \sigma^2 + \frac{\sigma^2}{n}\right) \implies Z = \frac{Y^* - \overline{Y}}{\sqrt{\sigma^2 + \sigma^2/n}} = \frac{Y^* - \overline{Y}}{\sigma\sqrt{1 + 1/n}} \sim \mathcal{N}(0,1).$$

If the population variance σ^2 is unknown, we replace σ^2 with its natural point estimator S^2 (the sample variance) and we obtain

$$t = \frac{Y^* - \overline{Y}}{S\sqrt{1 + 1/n}} \sim t(n-1). \quad (1)$$

From this result, it follows that

$$\overline{Y} \pm t_{n-1,\alpha/2} S\sqrt{1 + 1/n} \quad (2)$$

is a $100(1-\alpha)$ **percent prediction interval** for the new observation Y^*.
(a) Starting with the sampling distribution result in Equation (1), provide the algebraic argument showing that the prediction interval in Equation (2) is correct. Start by writing

$$1 - \alpha \;=\; P\left(-t_{n-1,\alpha/2} < \frac{Y^* - \overline{Y}}{S\sqrt{1+1/n}} < t_{n-1,\alpha/2}\right)$$

and now isolate Y^* in the center of the probability using algebra. For additional guidance, look at the derivation of the corresponding confidence interval for μ in the notes.
(b) Return to Problem 1 on this assignment. Calculate a 95 percent prediction interval for Y^*, the lifetime of the next bulb sampled. Interpret the interval.
(c) Compare the length of the prediction interval you just calculated in part (b) to the length of the confidence interval for μ you calculated in Problem 1(a). Explain conceptually why the prediction interval is wider.