1. (a) A reasonable answer is "all water wells in southeastern New Hampshire."
(b)

- The sample mean $\overline{x} = 48.2$ estimates the population mean μ, that is, the mean As concentration among all water wells in southeastern New Hampshire.
- The sample standard deviation $s = 43.6$ estimates the population standard deviation σ, that is, the standard deviation of the As concentrations for all water wells in southeastern New Hampshire.

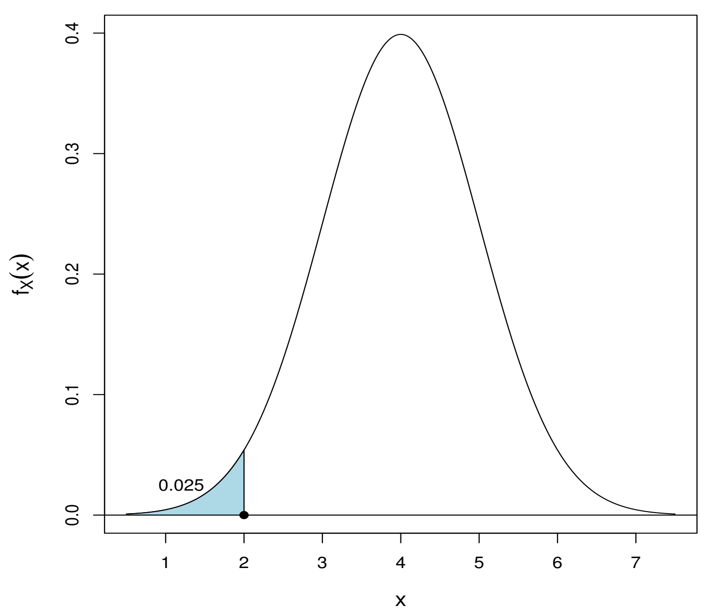
The units attached to $\overline{x} = 48.2$ is ppb. The units attached to $s = 43.6$ is ppb.
(c) The normal qq-plot shows a pronounced departure in the upper tail. This would not be consistent with a normal distribution.

In addition, the values of the sample mean $\overline{x} = 48.2$ and the sample standard deviation $s = 43.6$ are useful here. If a normal population distribution was suitable, then we know (from the 68-95-99.7% rule) approximately

- 68% of the data should be between $48.2 \pm 43.6 \longrightarrow (4.6, 91.8)$ ppb
- 95% of the data should be between $48.2 \pm 2(43.6) \longrightarrow (-39.0, 135.4)$ ppb
- 99.7% of the data should be between $48.2 \pm 3(43.6) \longrightarrow (-82.6, 179.0)$ ppb.

However, it is not possible for an As concentration to be negative. Therefore, these intervals suggest a normal population distribution assumption is not appropriate.
(d) We are 95% confident the population mean As concentration μ is between 32.5 and 63.9 ppb.

2. Here is the $\mathcal{N}(4, 1^2)$ pdf:



Note that 2 years is 2 standard deviations below the population mean of $\mu = 4$ years. Therefore, 95% of the car battery lifetimes should be between 2 years and 6 years (68-95-99.7% Rule). This leaves approximately 2.5% of the car battery lifetimes falling below 2 years. This is the percentage which will qualify for the warranty benefit.

(b) If σ is larger than 1 year, then the variability in the distribution of the car battery lifetimes will be larger than that in the picture on the last page. Assuming the population mean $\mu = 4$ years remains the same, larger variability will <u>increase</u> the percentage of lifetimes that are below 2 years (i.e., that will qualify for the warranty benefit).
(c) The confidence interval $(0.93, 1.54)$ says we are 95% confident the population variance σ^2 of the car battery lifetimes is between 0.93 and 1.54 $(\text{years})^2$. A 95% confidence interval for the population standard deviation σ is

$$(\sqrt{0.93}, \sqrt{1.54}) \implies (0.96, 1.24) \text{ years.}$$

We are 95% confident the population standard deviation σ is between 0.96 and 1.24 years. Note that "1" is included in this interval, so this interval is consistent with the manufacturer's claim the population standard deviation is 1 year.
(c) This means the interpretation of the intervals in part (c) depends critically on the normal population distribution assumption for the car battery lifetimes. If the population distribution is something else (e.g., gamma, exponential, Weibull, etc.), then the confidence intervals may be giving us misleading inferences.

3. (a) We are 90% confident the population proportion difference $\Delta = p_1 - p_2$ is between -0.006 and 0.126. A very large range of this interval is positive, but "0" is still included in the interval as a plausible value for the difference. If $\Delta = p_1 - p_2 = 0$, then we cannot say there is a difference between the population proportions when comparing the two groups.
(b) If the confidence level is lower, then the interval will be shorter. And, it won't have to be much below 90% because the 90% confidence interval just barely includes "0." So, I would say, "something slightly less than 90% should do it." In fact, I checked in R: an 86% confidence interval does barely exclude "0".

```
> prop.test(c(12,6),c(100,100),conf.level=0.86,correct=FALSE)$conf.int
[1] 0.001 0.112
```

(c) We set the margin of error

$$z_{\alpha/2}\sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n}} = 0.04$$

and solve for n. We have $z_{\alpha/2} \approx 1.96$ for 95% confidence and the estimates $\widehat{p}_1 = 0.12$ and $\widehat{p}_2 = 0.06$ from part (a). Thus,

$$\begin{aligned} 1.96\sqrt{\frac{0.12(1-0.12)}{n} + \frac{0.06(1-0.06)}{n}} = 0.04 \implies & \sqrt{\frac{0.162}{n}} = \frac{0.04}{1.96} \\ \implies & \frac{0.162}{n} = \left(\frac{0.04}{1.96}\right)^2 \\ \implies & \frac{n}{0.162} = \left(\frac{1.96}{0.04}\right)^2 \\ \implies & n = 0.162\left(\frac{1.96}{0.04}\right)^2 \approx 388.96. \end{aligned}$$

To write a 95% confidence interval for $\Delta = p_1 - p_2$ with margin of error equal to 0.04, we would need to observe about 389 individuals per group (that is, 389 nurses and 389 office workers).

4. (a) We are 95% confident the population mean difference $\Delta = \mu_1 - \mu_2$ is between -25.0 and -1.1. Note that this interval contains only negative values. That is, negative values of $\Delta = \mu_1 - \mu_2$ are consistent with the data. At the 95% confidence level, we would infer the population mean worm count for treated lambs (μ_1) is less than the population mean worm count for untreated lambs (μ_2).
(b) An acceptable answer to this is, "the variation in worms counts looks to be approximately the same for both groups (treated versus untreated)." In this situation, the two intervals will provide similar answers. The two intervals will provide different answers (potentially very different) when the variation levels between the groups are different.
(c) One acceptable way to answer this would be to say,

> *"No, it is possible to perform a matched-pairs experiment. We cannot treat a single lamb with the drug and not treat the lamb at the same time over the 6-month period. Therefore, it is not possible to observe the two experimental settings (treated/untreated) on a particular lamb at the same time."*

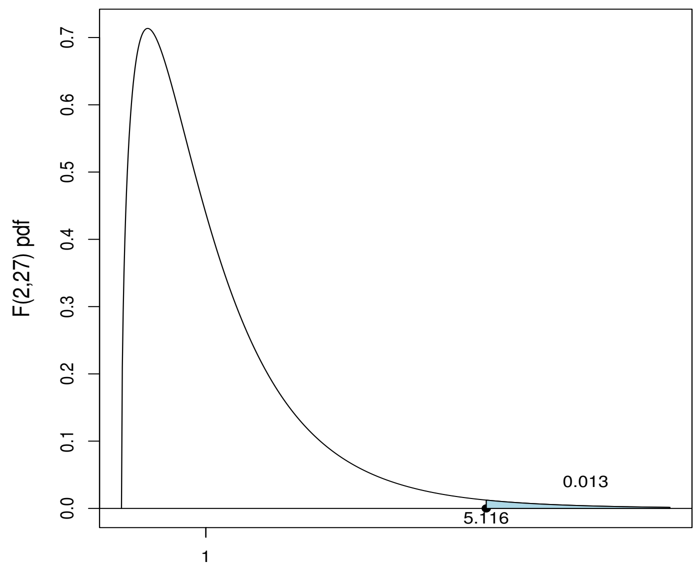
You could also argue that it is possible! Suppose that you split the 6 month period into 3 two-month periods.

- Stage 1: During the first two months, randomly assign the 24 lambs to the two experimental groups (12 treated/12 untreated). Record the worm counts at the end of the two months. Of course, this assumes one can measure these counts without killing the lamb.
- Stage 2: Wait two months (allowing the 12 treated lambs in the first stage to "return to normal"). This is called a "wash-out period."
- Stage 3: During the last two months, switch the 12 treated lambs in Stage 1 to now being untreated. Similarly, switch the 12 untreated lambs in Stage 1 to now be treated. Record the worm counts at the end of the two months.

This would give you treated and untreated worm counts for each of the 24 lambs, and you could analyze the data differences (treated minus untreated) as we did in a usual matched pairs analysis. This is an example of a *cross-over design*. It is a matched pairs design where each individual is treated under two different experimental conditions (with a wash-out period in between).

5. (a) When H_0 is true, the overall F statistic has an F distribution with 2 and 27 degrees of freedom, that is, $F \sim F(2, 27)$. The graph on the next page shows this distribution with the p-value identified. Recall the p-value is the area to the right of $F = 5.116$.
(b) We are 95% confident that

- the difference in population means for plastic and cork $\Delta_{31} = \mu_3 - \mu_1$ is between -634.9 and -60.9 gaskets per hour.
- the difference in population means for rubber and cork $\Delta_{21} = \mu_2 - \mu_1$ is between -570.8 and 3.4 gaskets per hour.
- the difference in population means for rubber and plastic $\Delta_{23} = \mu_2 - \mu_3$ is between -222.7 and 351.3 gaskets per hour.



Furthermore, because our familywise confidence level is 95%, this means we are 95% confident in all three statements combined as a group. Note that there is a significant difference between the mean number of gaskets produced for plastic and cork because the confidence interval for $\Delta_{31} = \mu_3 - \mu_1$,

$$(-634.9, -60.9),$$

contains only negative values (i.e., it excludes "0"). This suggests the population mean number of plastic gaskets (μ_3) is less than the population mean number of cork gaskets (μ_1). The other two confidence intervals contain "0," so we cannot conclude an ordering in the population means between

- rubber and cork
- rubber and plastic.

(c) The familywise 95% confidence interval for $\Delta_{21} = \mu_2 - \mu_1$ is $(-570.8, 3.4)$ gaskets per hour. Familywise confidence intervals are longer because the "95%" confidence level applies to all 3 pairwise intervals as a group. This means the confidence level associated with each individual interval must be larger than 95%.

A 95% confidence interval for $\Delta_{21} = \mu_2 - \mu_1$ by itself is $(-565.9, -1.3)$. The confidence level associated with $(-565.9, -1.3)$ is 95%, but it pertains only to this specific comparison (rubber and cork). It does not include the other 2 comparisons.

(d) I would be a little concerned about the equal-variance assumption for the three populations, that is,

$$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma^2.$$

The spread in the boxplot for the plastic sample is quite a bit less than the spread in the other two boxplots. We wouldn't expect to see this if the population variances were really equal. Of course, we have small sample sizes here (10), so we don't have a lot of information about the populations. We could be overreacting.

Either way, the equal-variance assumption in one-way classification is critical. If it is violated, this could distort our entire analysis.