Question 1.

(a) The sample mean $\overline{y} = 21.7$ lbs is an estimate of the population mean force required. The sample standard deviation $s = 6.4$ lbs is an estimate of the population standard deviation. By "population," we refer to the universe of all bottles produced in this production process. Our sample of 68 bottles is taken from this population, and $\overline{y}$ and s are estimates of the population mean μ and the population standard deviation σ, respectively.

(b) The standard error of a statistic (like $\overline{Y}$) is the standard deviation of the sampling distribution of the statistic. Here,

$$\text{se}(\overline{Y}) = \sqrt{\text{var}(\overline{Y})} = \sqrt{\frac{\sigma^2}{68}} = \frac{\sigma}{\sqrt{68}}.$$

From part (a), we know $s = 6.4$ estimates σ, so an estimate of the standard error is

$$\frac{6.4}{\sqrt{68}} \approx 0.78 \text{ lbs.}$$

This number quantifies how variable the sample mean $\overline{Y}$ is as an estimate of the population mean μ; i.e., it is a measure of variation or uncertainty in the estimate.

(c) We are 99 percent confident that the population mean force required μ is between 19.7 lbs and 23.9 lbs.

(d) Disagree. A confidence interval is an estimate of where the mean of the population falls. It does not refer to individual bottles in the population. Of course, the assumptions may still be violated (e.g., normality, etc.), but not for the 14/68 reason the engineer cites.

Question 2.

(a) The population is all males (18 and older) who live in South Carolina (that's probably around 2 million males). The sample is the 200 male subjects the geneticist observed.

(b) The sample proportion of South Carolina males who have the disorder is $\widehat{p} = \frac{36}{200} = 0.18$. A 90 percent confidence interval for the population proportion p is

$$0.18 \pm 1.65\sqrt{\frac{0.18(1-0.18)}{200}} \implies 0.18 \pm 0.045 \implies (0.135, 0.225).$$

We are 90 percent confident that the population proportion of males in South Carolina (aged 18 and older) who have this certain minor blood disorder p is between 0.135 and 0.225 (i.e., between 13.5% and 22.5%).

(c) In the confidence interval formula, we set the margin of error

$$z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = 0.01$$

and solve for n. With $z_{\alpha/2} = z_{0.01/2} \approx 2.58$ and $\widehat{p} = 0.18$, this gives

$$n = \left(\frac{2.58}{0.01}\right)^2 0.18(1-0.18) \approx 9825.$$

Therefore, she would have to sample 9,825 males to attain 99% confidence with a tight margin of error of 0.01.

Note: If you wanted to be conservative, you could substitute in $p_0 = 0.5$ as a guess for p, but this would inflate the number of males needed; i.e.,

$$n = \left(\frac{2.58}{0.01}\right)^2 0.5(1-0.5) \approx 16641.$$

How to reduce the number of males needed? She could

1. reduce the confidence level (e.g., to 95%, etc.)
2. increase the margin of error (e.g., to 0.03, etc.).

Doing either (or both) would reduce the required sample size. However, she would pay for this in the form of less confidence and less precision (i.e., a wider interval).

Question 3.
(a) Clearly, the boxplots suggest the samples have different amounts of variation (the light-cured sample is much more variable). Based on this alone, I would choose the interval that did not assume the population variances were equal. We could write a confidence interval for σ_1^2/σ_2^2 using the F distribution to determine which assumption was more reasonable. If this interval included "1," we could use the equal-variance interval; if not, use the unequal-variance interval.
(b) I picked the unequal-variance interval. We are 95 percent confident the difference in the population mean failure times $\mu_1 - \mu_2$ is between 0.10 and 3.24 years. Because this interval includes only positive values, this suggests the population mean failure time for the light-cured resin cement is larger than the population mean failure time for the self-adhesive resin cement.
(c) I can see how it would be possible to design a matched pairs study for future patients. First you would have to get them to agree to be in the experiment. You would also have to get each patient to agree to wear two veneers−one on each front top tooth.

- For the first patient, randomly assign the light-cured resin cement to one of the front top teeth; assign the self-adhesive resin cement to the other front top tooth.
- Do the same thing for the other 13 patients.
- Record the failure time for each veneer on each patient.

Obviously, this would take a long time to complete the study (maybe more than 10 years?), so the independent sample analysis using patients from past records is easier to do−all she has to do is search through old patient records to get the failure times.

The advantage of the matched pairs design, however, is that comparing the population mean failure times should be more precise. This is true because you are not comparing Patient A with the light-cured resin cement to Patient B with the self-adhesive resin cement. Instead you would be comparing the failure times for the two resin types within the same patient. This should lead to a more precise confidence interval for $\mu_1 - \mu_2$.

Question 4.
(a) The null hypothesis H_0 says the population mean strengths are the same for each type of wire rope. The alternative hypothesis H_1 says that the population mean strengths are different somehow (although it does not specify how the population means are different). In symbols,

$$H_0 : \mu_1 = \mu_2 = \mu_3$$
versus
$$H_1 : \text{the population means } \mu_i \text{ are not all equal.}$$

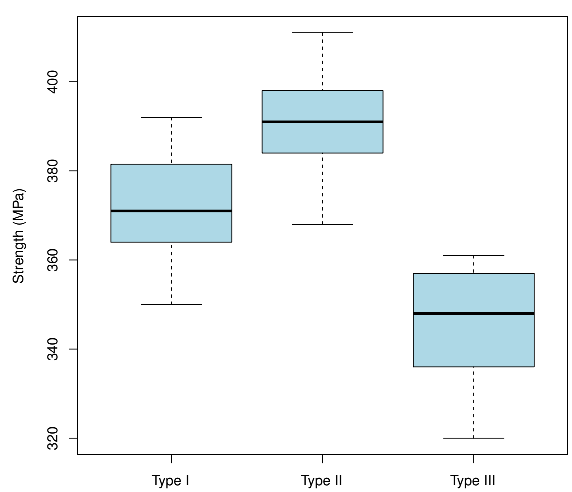
Here μ_1, μ_2, and μ_3 denote the population mean strengths of the three wire rope types, respectively.
(b) This value of F is very close to 1, so it is not surprising the p-value is large. If we performed the test at the $\alpha = 0.05$ significance level, we would not reject H_0 because the p-value > 0.05. In fact, H_0 would not be rejected at any reasonable level of significance. The data are consistent with what we would expect under H_0; i.e., when all three population mean strengths are equal.
(c) The statistical assumptions are

$$\begin{array}{ll} \text{Sample 1:} & Y_{11}, Y_{12}, ..., Y_{1n_1} \sim \mathcal{N}(\mu_1, \sigma^2) \\ \text{Sample 2:} & Y_{21}, Y_{22}, ..., Y_{2n_2} \sim \mathcal{N}(\mu_2, \sigma^2) \\ \text{Sample 3:} & Y_{31}, Y_{32}, ..., Y_{3n_3} \sim \mathcal{N}(\mu_3, \sigma^2). \end{array}$$

The F statistic is constructed by taking the ratio of two estimators of the common population variance σ^2. These estimators are MS_{trt} and MS_{res}. When H_0 is true,

$$\begin{aligned} E(\text{MS}_{trt}) &= \sigma^2 \\ E(\text{MS}_{res}) &= \sigma^2, \end{aligned}$$

so F should be around 1. However, if the population variances for the three wire types were different, then it would not be clear what MS_{trt} and MS_{res} are even estimating. MS_{res} would be estimating some unknown linear combination of the population variances, and I'm not sure what MS_{trt} would be estimating. For the overall F test to work, we need the population variances to be identical among the three wire rope types.
(d) A large F statistic is going to result when there is strong evidence against H_0. Therefore, draw side-by-side boxplots where the boxplots for each wire type are far away from each other with respect to location. For example, something like this:



The F statistic for these data was $F \approx 44$!