Question 1.
(a) The population might be the collection of all Brand G condoms produced by its manufacturer. There are other possible answers here too. A point estimate for the population mean is $\overline{y} = 1481$ cycles. A point estimate for the population variance is $s^2 = (543)^2 = 294849$ (cycles)^2.
(b) The standard error of the sample mean $\overline{Y}$ is

$$\text{se}(\overline{Y}) = \frac{\sigma}{\sqrt{n}} = \frac{\sigma}{\sqrt{20}},$$

where σ is the population standard deviation. To estimate this quantity, estimate σ with the sample standard deviation $s = 543$. The estimated standard error is

$$\frac{s}{\sqrt{20}} = \frac{543}{\sqrt{20}} \approx 121.4.$$

This standard error measures how variable the sample mean $\overline{Y}$ is as an estimator for the population mean μ.
(c) First, it is hard to get a definite answer on the true population distribution, especially with only 20 observations from it. Second, with such a small sample, we know that qq plots can produce suspicious results even if the population model is correct. Third, the investigator has forgotten that confidence intervals for population means are robust to normality departures−this means that these intervals can still be used even when the normality assumption is violated.

Question 2.
(a) The sample proportion of obese female workers is $\widehat{p} = \frac{120}{541} \approx 0.222$. A 95 percent confidence interval for the population proportion is

$$\frac{120}{541} \pm 1.96\sqrt{\frac{\frac{120}{541}\left(1 - \frac{120}{541}\right)}{541}} \implies (0.187, 0.257).$$

We are 95 percent confident that the population proportion of obese female workers is between 0.187 and 0.257.
(b) We are 95 percent confident that the difference in the population proportions of obese workers $p_1 - p_2$ is between -0.116 and -0.040. Because this interval does not include "0," we can conclude (at the 95 percent confidence level) that the population proportion of obese female workers is less than the population proportion of obese male workers.
(c) We could calculate a confidence interval for the ratio of the population variances σ_2^2/σ_1^2. If this interval excluded the value "1," this would suggest that the population variances are different. The confidence interval (we discussed in class) is based on the F distribution. **Note:** This confidence interval would depend heavily on the assumption that the two population distributions are normal. Because Y, the number of missed work days, is integer-valued, the population distributions here are discrete. The confidence interval for σ_2^2/σ_1^2 described above could give misleading results.

Question 3.
(a) Both indoor and outdoor observations are recorded at each house. The samples therefore are not independent (i.e., indoor and outdoor concentrations at the same house are potentially dependent).

(b) We are 95 percent confident that the difference in the population mean hexavalent chromium concentrations is between -0.561 and -0.286 nanograms/m^3. Because this interval does not include "0," we can conclude (at the 95 percent confidence level) that the population mean indoor hexavalent chromium concentration is less than the population mean outdoor hexavalent chromium concentration in the population.
(c) It must mean that recording two observations on the same house, on average, did not produce largely different results than had these concentrations been recorded on independent samples; i.e., 33 houses with indoor concentrations and 33 houses with outdoor concentrations. In other words, pairing did not reduce the variability involved with making this comparison.

Question 4.
(a) The investigator may be correct for the sample information (i.e., for the 52 infants in the study). However, no one (except the Oracle) can know whether or not this assertion applies to the population of infants, especially when looking at the sample information (boxplots) only.
(b) The F statistic is used to test

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$
$$\text{versus}$$
$$H_1 : \text{the population means } \mu_i \text{ are not all equal.}$$

The large p-value here (p-value = 0.2009) provides very little evidence to reject H_0. In other words, we can not conclude that the population mean PUSF percentages are different among the four groups of nutrition.
(c) The one-way classification analysis assumes that

1. The samples are independent. This would be reasonable if the nutrition regimens were randomly assigned to the infants.
2. The populations (one for each nutrition) are normally distributed.
3. The populations (one for each nutrition) have equal variances.

(d) In part (b), we did not reject H_0, the hypothesis that says all population means are equal. Therefore, I would expect each of the 6 pairwise confidence intervals to include "0."