1. The activation temperature of a sprinkler system, Y, is normally distributed with population mean $\mu = 130$ (deg F) and population variance $\sigma^2 = 2.25$ (deg F)^2. A random sample of $n = 20$ sprinklers is obtained (from a population described by this distribution) and the activation temperature is observed on each sprinkler.

(a) What is the sampling distribution of the sample mean $\overline{Y}$ of the $n = 20$ sprinkler temperatures? Be precise.
(b) Calculate the standard error of the sample mean $\overline{Y}$.
(c) Let S denote the sample standard deviation of the $n = 20$ sprinkler temperatures. Give the sampling distribution of

$$\frac{\overline{Y} - 130}{S/\sqrt{20}}.$$

(d) Let S^2 denote the sample variance of the $n = 20$ sprinkler temperatures. Fill in the blanks:

$$\begin{aligned} E(\overline{Y}) &= \underline{\qquad} \\ E(S^2) &= \underline{\qquad} \end{aligned}$$

Explain why your answers are correct.

2. Toxaphene is an insecticide associated with liver/kidney damage and with increased risk of cancer in humans. To investigate the effect of toxaphene exposure on weight gain in rats,

- $n_1 = 20$ rats were fed diets that contained a low dose of toxaphene (Group 1)
- $n_2 = 23$ rats were fed diets that contained no toxaphene (Group 2).

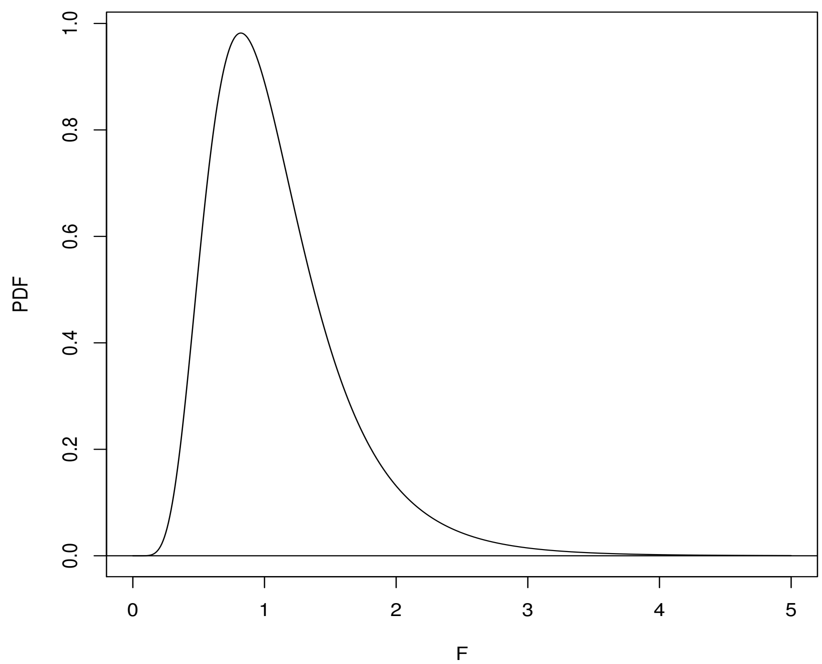
The weight gain (measured in grams) was recorded for each rat. Here are the sample variances for the two groups:

```
> var(low.dose) # Group 1
[1] 46.93
> var(no.toxaphene) # Group 2
[1] 27.81
```

(a) Based on these two independent samples of rats, could the population variances σ_1^2 and σ_2^2 possibly be equal? Recall that if $\sigma_1^2 = \sigma_2^2$, then the ratio of the sample variances

$$F = \frac{S_1^2}{S_2^2} \sim F(19, 22).$$

I have constructed the $F(19, 22)$ pdf below (see next page). Calculate the value of F in this example and place this value on the horizontal axis of the figure. What do you



think about the equal population variance assumption? Is there evidence for or against it? Strong evidence?

(b) The F sampling distribution in part (a) is not robust to departures from normality. Explain what this means and how you would diagnose the normality assumption for each sample.

(c) I used R to calculate a 95 percent confidence interval for

$$\mu_1 - \mu_2 = \text{difference of the population mean weight gains (in grams).}$$

I calculated one interval that assumes equal population variances and one interval that does not.

```
> t.test(low.dose,no.toxaphene,conf.level=0.95,var.equal=TRUE)$conf.int
[1] 20.48 27.96

> t.test(low.dose,no.toxaphene,conf.level=0.95,var.equal=FALSE)$conf.int
[1] 20.39 28.05
```

If researchers wanted to know how the population mean weight gains compare between the two groups of rats (low dose rats and no-toxaphene rats), what would you tell them? Assume that all necessary assumptions are justified.

3. In Connecticut, a random sample of $n = 200$ legally-registered automobiles was recently taken. Of the 200, only 124 passed the state's emission test for pollution.

I calculated a 90 percent confidence interval for the population proportion of automobiles that meet the state's emissions standards, to be $(0.56, 0.68)$. I used the formula

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

and recorded all calculations using 2 digits.

(a) What is the population here? What is the sample?
(b) Draw a detailed picture showing me where the value $z_{\alpha/2}$ comes from. I am looking for a very clear answer (i.e., if I have to guess from your picture, this is not good).
(c) An environmental engineer would like to design a larger study to estimate the population proportion. She would like to write a 95 percent confidence interval ($z_{0.05/2} \approx 1.96$) that will have margin of error equal to 0.02. How many cars will she need to sample?

4. The manager of a large taxi company in Los Angeles (with 1000s of cars) is trying to decide whether using radial tires (instead of using belted tires) improves his fleet's fuel economy on average.

- He randomly samples $n = 12$ cars equipped with radial tires and has them driven over a test course.
- Without changing drivers, the same cars are then equipped with belted tires and are driven through the same test course.

The gasoline consumption (recorded in km per liter) was recorded for each car and tire type:

Car	1	2	3	4	5	6	7	8	9	10	11	12
Radial	4.2	4.7	6.6	7.0	6.7	4.5	5.7	6.0	7.4	4.9	6.1	5.2
Belted	4.1	4.9	6.2	6.9	6.8	4.4	5.7	5.8	6.9	4.7	6.0	4.9

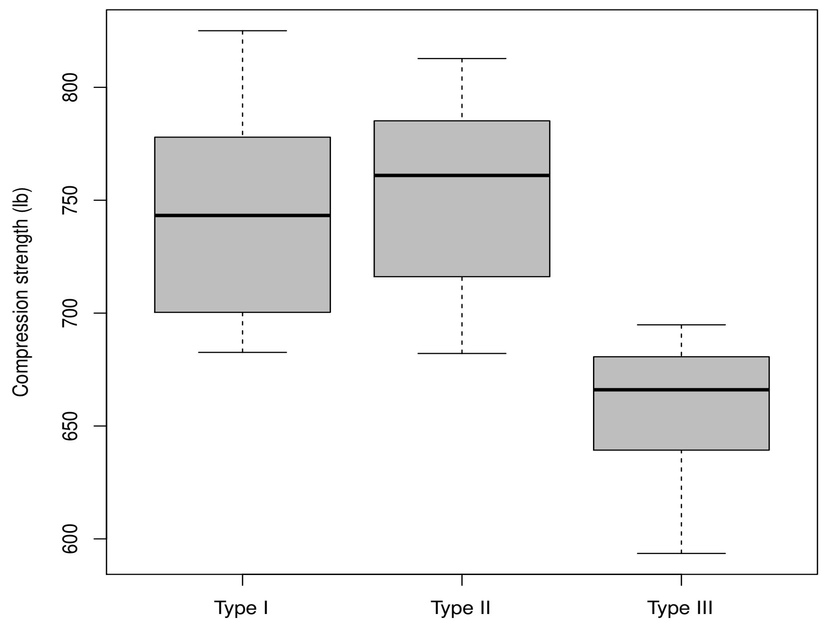
(a) Explain why this is a matched pairs study. Are the two samples independent or dependent? Why?
(b) I used R to write 95 and 99 percent confidence intervals for the population mean difference gasoline consumption between radial (Group 1) and belted (Group 2) tires.

```
> t.test(diff,conf.level=0.95)$conf.int
[1] 0.016 0.267

> t.test(diff,conf.level=0.99)$conf.int
[1] -0.035 0.318
```

- What does `diff` mean in the code above? Describe what this is.
- Pick one confidence interval above (tell me which one) and interpret it for the manager.
- Does it bother you that one interval includes "0" and the other doesn't? Explain why this might be happening.

5. A one-way classification analysis was used with three different types of boxes. Twelve boxes of each type were subjected to a compression test and the strength of each box was measured in lbs (36 boxes in all; 12 for each type).



Here is the analysis of variance (ANOVA) table for these data:

```
> anova(lm(strength ~ box.type))
Analysis of Variance Table

Response: strength
          Df Sum Sq Mean Sq F value    Pr(>F)
box.type   2  62732 31366.0   19.55 2.509e-06 ***
Residuals 33  52945  1604.4
```

(a) The F statistic (here, $F = 19.55$) is used to test two hypotheses: H_0 and H_1. Write out what these hypotheses are. You can do this using notation (that you clearly define)

or you can write this out in words. Also, tell me which hypothesis is more supported by the data (and why).
(b) Here is the R output to do a follow-up Tukey analysis:

```
> TukeyHSD(aov(lm(strength ~ box.type)),conf.level=0.95)
  Tukey multiple comparisons of means
    95% family-wise confidence level

                        diff        lwr       upr     p adj
Type.II-Type.I      7.675389  -32.44982  47.80060 0.8860678
Type.III-Type.I   -84.464660 -124.58987 -44.33945 0.0000331
Type.III-Type.II  -92.140049 -132.26526 -52.01484 0.0000083
```

- Give a brief summary of what these results tell us.
- If you were advising the investigators on which box type to use (to maximize population mean strength), which box type would you recommend?