CHAPTER 6:

6.1. Johnson Controls claims the activation temperature of its new sprinkler system for residential use (X, in deg F) is normally distributed with population mean $\mu = 130$ and population variance $\sigma^2 = 2.25$. A random sample of 20 sprinkler systems is observed and the activation temperatures $X_1, X_2, ..., X_{20}$ are recorded.
(a) What is the population in this example? Give a reasonable answer.
(b) Let $\overline{X}$ denote the sample mean of the 20 activation temperatures. What is the sampling distribution of the sample mean $\overline{X}$?
(c) Calculate the standard error of the sample mean $\overline{X}$. What does this measure?
(d) (Chapter 7) The sample mean and sample standard deviation of the 20 activation temperatures (`temp`) are shown below:

```
> mean(temp)
[1] 135.3
> sd(temp)
[1] 1.15
```

I calculated a 99% confidence interval for the population mean μ using R's `t.test` function:

```
> t.test(temp,conf.level=0.99)$conf.int
[1] 134.5 136.0
```

Is this confidence interval consistent with Johnson Controls' claim that the population mean activation temperature is 130 deg F? Explain.

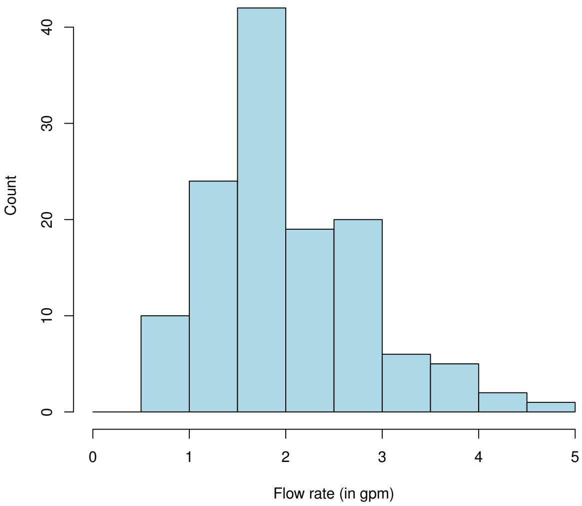
6.2. Shower flow rate is a measure of the water volume, in gallons per minute (gpm), that comes out of a shower head. An environmental engineer observed a random sample of $n = 129$ residential houses in Charlotte, NC, and measured the shower flow rate for each house's primary shower unit. A histogram of the data is shown at the top of the next page (left). A normal quantile-quantile plot is shown on the right.
(a) What is the population in this problem? Give a reasonable answer.
(b) I used R to calculate the sample mean and sample standard deviation:

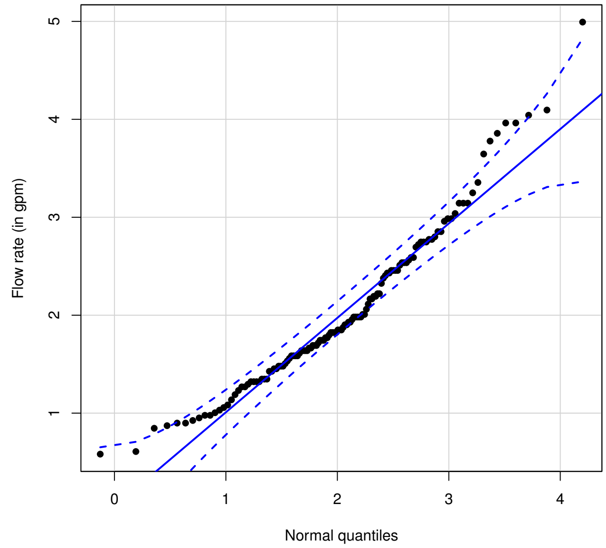
```
> mean(flow.rate)
[1] 2.04
> sd(flow.rate)
[1] 0.81
```

Estimate the standard error of the sample mean. What does this measure?
(c) A co-worker says,

> *"The histogram is skewed and there is some disagreement in the qq plot. We need to take a larger sample of houses so that the data are more normally distributed."*

Do you agree with this statement? Explain.





(d) (Chapter 7) I used R to calculate a 95% confidence interval for the population mean shower flow rate:

```
> t.test(flow.rate,conf.level=0.95)$conf.int
[1] 1.89 2.18
```

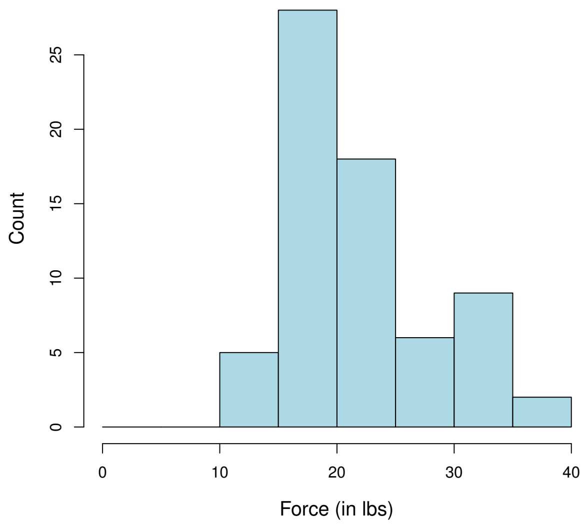
However, only 16 of the 129 houses in the sample above had a flow rate between these two end-points (which is a lot less than 95%). Is this a contradiction? If so, what must be going on here?

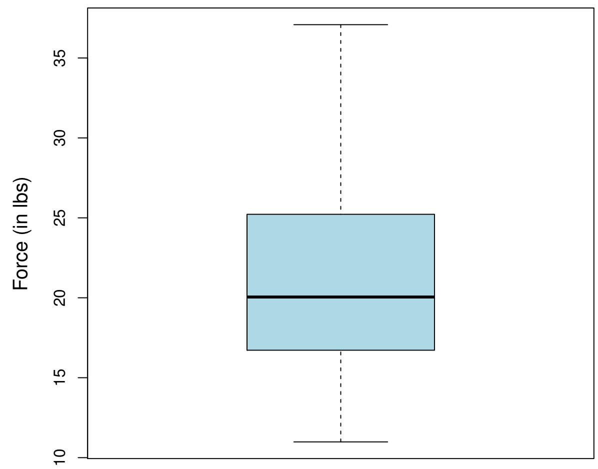
6.3. The force needed to remove a cap from a medicine bottle is an important characteristic. The force must be sufficiently high to prevent unauthorized entry (e.g., by children, etc.) but not so large that elderly patients cannot open the bottle. A random sample of $n = 68$ bottles was taken from a production process. The force (in lbs) was measured on each bottle by an automated device. The data are below:

14.7	18.9	27.6	24.3	24.1	28.7	22.2	21.9	16.5	17.6
22.9	16.0	16.1	18.4	30.4	16.2	14.5	15.3	25.7	15.3
16.1	15.2	15.2	19.8	19.1	10.5	22.1	17.1	15.6	17.5
20.2	17.5	20.3	15.9	17.7	20.7	24.2	27.6	17.3	33.0
31.9	27.2	21.1	21.3	26.0	31.2	34.3	32.5	24.9	16.9
37.1	36.1	34.4	20.2	20.0	21.8	14.0	15.0	19.6	15.7
30.4	24.3	15.6	17.2	17.8	21.1	34.9	24.9		

A histogram and a boxplot of the data are shown at the top of the next page. I used R to calculate the sample mean and the sample standard deviation of the data above:

```
> mean(force)
[1] 21.7
> sd(force)
[1] 6.4
```





(a) What do the sample mean and the sample standard deviation estimate? Explain in words−don't just write symbols. What are the units attached to these estimates?
(b) Recall the Central Limit Theorem, which says the sample mean

$$\overline{X} \sim \mathcal{AN}\left(\mu, \frac{\sigma^2}{n}\right),$$

where μ is the population mean and σ^2 is the population variance. Using this result, calculate an estimate of the standard error of the sample mean. Describe what this measures.
(c) (Chapter 7) I used R to calculate a 99% confidence interval for the population mean force required:

```
> t.test(force,conf.level=0.99)$conf.int
[1] 19.7 23.9
```

Interpret what this confidence interval means.

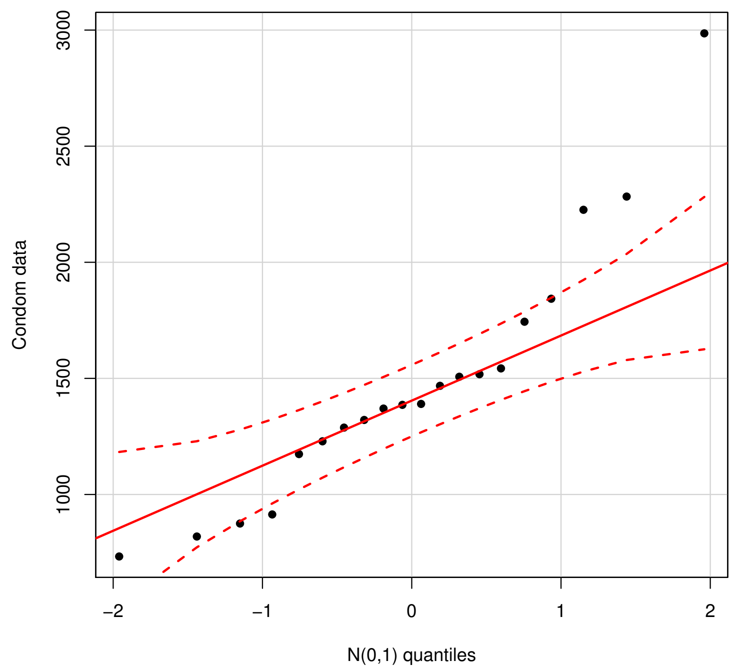
6.4. Condom failure during sexual intercourse can lead to unplanned pregnancy and the transmission of sexually transmitted diseases. The article "Fatigue testing in condoms" (*Polymer Testing*, **2009**; 567-571) examined the breaking strength of a random sample of $n = 20$ "Brand G" condoms. Individual condoms were placed on an apparatus that, when in motion, mimics sexual intercourse. Condom breaking strength was determined by recording the number of cycles until breakage. Here were the data:

2226	2283	875	733	1390	1744	1174	1468	1229	1386
1843	914	1518	1288	1507	1321	819	1370	1543	2986

Here are the sample mean and sample standard deviation for this sample of condoms:

```
> mean(cycles)
[1] 1481
> sd(cycles)
[1] 543
```

(a) Describe what you think the population is. Report point estimates for the population mean μ and the population variance σ^2. State the units attached to your estimates.
(b) Calculate an estimate of the standard error of the sample mean. Describe in words what the standard error measures.
(c) (Chapter 7) Here is a normal qq plot for the data:



An investigator looking at this plot concludes,

> *"These data are not normally distributed. Confidence intervals for the population mean breaking strength will not be useful."*

How would you respond to the investigator?

CHAPTER 7:

7.1. A geneticist wants to estimate the proportion of males in South Carolina (aged 18 or older) who have a certain minor blood disorder. From a study involving 200 male subjects (all South Carolinians), she determines that 36 have the disorder.
(a) What is the population here? What is the sample?
(b) Calculate a 90% confidence interval for the population proportion using

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

Interpret your interval. Note that for 90% confidence the quantile $z_{0.10/2} \approx 1.65$.
(c) The geneticist wants to use the results above to design a larger study. She would like to calculate a 99% confidence interval ($z_{0.01/2} \approx 2.58$) for the population proportion that has margin of error equal to 0.01. How many males will she have to sample? Comment on whether you think this is feasible and then discuss ways she could reduce the number of males needed.

7.2. A recent article in the *Magazine of Concrete Research* summarized an observational study involving the flexural strength of concrete beams. A random sample of $n = 27$ beams was tested and the strength of each beam (measured in MPa) was recorded. Here are the data:

5.9	7.2	7.3	6.3	8.1	6.8	7.0	7.6	6.8	6.5	7.0	6.3	7.9	9.0
8.2	8.7	7.8	9.7	7.4	7.7	9.7	7.8	7.7	11.6	11.3	11.8	10.7	

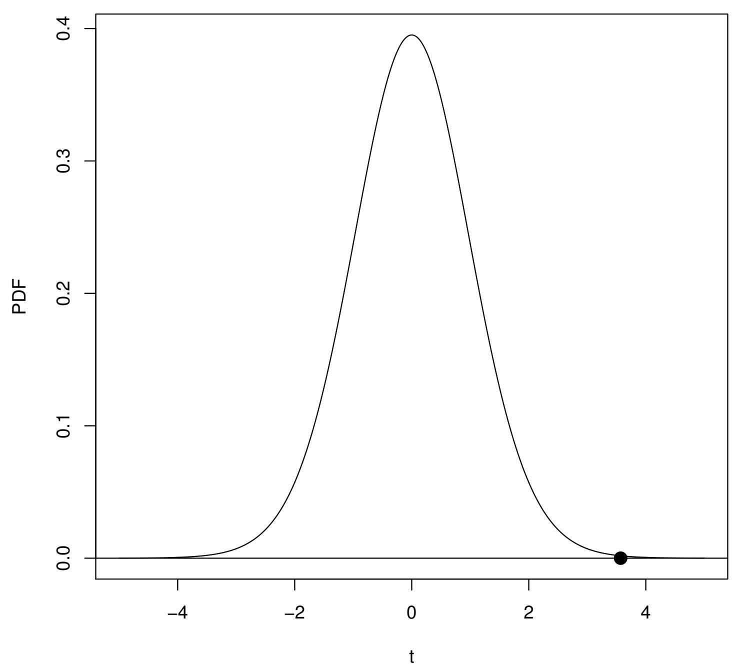
(a) Although I have not given you much information in the problem, describe what one might consider "the population" to be in this example.
(b) A 95% confidence interval for the population mean μ uses the formula

$$\overline{x} \pm t_{26,0.025} \frac{s}{\sqrt{27}}.$$

What part of this formula estimates the standard error of the sample mean? the margin of error?
(c) The manufacturer of this type of beam has to demonstrate to customers the population mean strength is 7.0 MPa. One way to investigate this would be to calculate the value of

$$t = \frac{\overline{x} - \mu}{s/\sqrt{27}}$$

and then compare t to its sampling distribution when the population mean strength is 7.0. What is the sampling distribution of t when the population mean is 7.0? What assumptions are you making here?
(d) When I calculated t and plotted it on its sampling distribution using the data above, here is what I observed:



Is this result ($t \approx 3.57$) most consistent with the population mean being equal to, smaller than, or larger than 7.0 MPa? Explain.

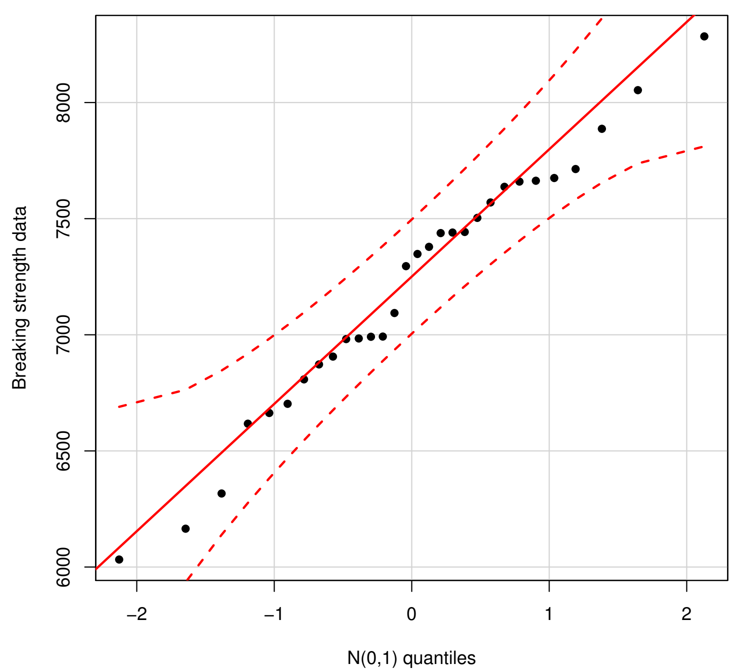
7.3. The 2008 article, "Development of novel industrial laminated planks in sweetgum lumber" (*Journal of Bridge Engineering*, **13**, 64-66), described the testing of composite beams used to add value to low-grade sweetgum lumber. A random sample of $n = 30$ beams was used. For each beam, investigators recorded its breaking strength (measured in lb/in^2). Here are the data:

6807.99	7637.06	6663.28	6165.03	6991.41	6992.23
6981.46	7569.75	7437.88	6872.39	7663.18	6032.28
6906.04	6617.17	6984.12	7093.71	7659.50	7378.61
7295.54	6702.76	7440.17	8053.26	8284.75	7347.95
7442.69	7886.87	6316.67	7713.65	7503.33	7674.99

(a) I used R to calculate a 90% confidence interval for the population mean beam breaking strength:

```
t.test(strength,conf.level=0.90)$conf.int
[1] 7035.15 7372.56
```

Interpret what this interval means.
(b) If you calculated a 95% confidence interval for the population mean breaking strength, would it be shorter than or longer than the interval above? Explain.
(c) Here is the qq plot for the breaking strength data under the normal population distribution assumption:



Suppose an investigator asks you, "Why does a linear trend in the qq plot support the normal distribution assumption? Could the population distribution be something else?" How would you respond to him? To answer his first question, explain precisely how a qq plot is formed and then explain the logic behind how it is interpreted.

7.4. In Connecticut, a random sample of 200 legally registered automobiles was selected from DOT records. Among the 200 automobiles selected, only 124 passed the state's emission test for pollution. I calculated a 90% confidence interval for the population proportion of automobiles that met 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 how $z_{\alpha/2}$ is determined. I am looking for a clear answer.
(c) An environmental engineer would like to design a larger study to estimate the population proportion. She would like to write a 99% confidence interval ($z_{0.01/2} \approx 2.58$) that will have margin of error equal to 0.01. How many cars will she need to sample? Comment on whether you think this is feasible and then list two ways she could reduce the number of automobiles needed.

CHAPTER 8:

8.1. Dental veneers are wafer-thin shells designed to cover the front surface of teeth. An orthodontist wants to compare the population mean lifetimes of veneers that differ only in what type of resin cement was used to attach them. Two types of cement were included: (1) light-cured resin cement and (2) self-adhesive resin cement.

The orthodontist sampled 14 patients who received a veneer applied to a top front tooth using light-cured resin cement. An independent sample was taken for 14 patients whose front-tooth veneer was applied using self-adhesive resin cement. The times to failure (in years) were recorded for each patient; these data are below. For example, the first patient's veneer in the light-cured resin cement group lasted for 7.6 years before it failed (e.g., it fell off, cracked, chipped, etc.).

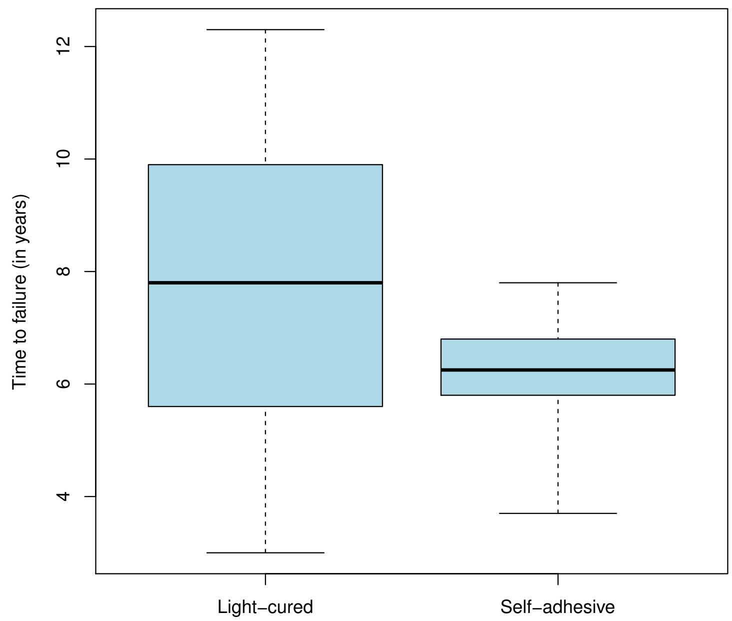
Light-cured		Self-adhesive	
7.6	3.0	5.4	3.7
12.3	9.9	6.2	7.8
5.6	9.6	6.3	6.0
6.7	8.6	6.5	5.0
5.6	10.7	6.0	6.8
4.4	10.0	7.0	6.6
8.0	7.3	6.8	5.8

Side-by-side boxplots of the two samples are shown at the top of the next page.

(a) If you wanted to write a confidence interval for $\Delta = \mu_1 - \mu_2$, the difference of the two population mean failure times between resin cement types (1 = light-cured; 2 = self-adhesive), select the confidence interval you would use:

- the one that assumed equal population variances $\sigma_1^2 = \sigma_2^2 = \sigma^2$.
- the one that did not assume the population variances were equal.

Explain why you chose the answer you did. In addition, what statistical procedure could be used to determine which assumption is more reasonable?



(b) I asked R to get both intervals in part (a), each at the 95% confidence level. Here are the intervals:

```
> t.test(light.cured,self.adhesive,conf.level=0.95,var.equal=TRUE)$conf.int
[1] 0.14 3.19
> t.test(light.cured,self.adhesive,conf.level=0.95,var.equal=FALSE)$conf.int
[1] 0.10 3.24
```

For the interval that you picked in part (a), interpret the corresponding interval here.
(c) The data above are from an observational study. The orthodontist simply selected the samples of patients from dental records. Could the orthodontist perform an experiment using a matched-pairs design to compare the population mean resin cement failure times μ_1 and μ_2? Explain how he could do this with 14 new patients and why this design might be better than using two independent samples.

8.2. A random sample of female workers from a large oil company was observed; among the $n = 541$ workers sampled, 120 were classified as "obese." The results from this study were published in a 2008 article in *Annals of Epidemiology*.
(a) (Chapter 7) Based on the information in the sample, calculate a 95% confidence interval for the population proportion of female workers who are obese. Use

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

and $z_{\alpha/2} = z_{0.05/2} \approx 1.96$. Interpret precisely what your interval means.
(b) The article also summarized the results of an independent random sample of male workers; among the 3612 workers sampled, 1084 were classified as obese. I calculated a 95% confidence

interval for the population proportion difference $\Delta = p_1 - p_2$ (female minus male) using

$$(\widehat{p}_1 - \widehat{p}_2) \pm 1.96\sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{541} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{3612}},$$

and I obtained $(-0.12, -0.04)$.

```
prop.test(c(120,1084),c(541,3612),conf.level=0.95,correct=FALSE)$conf.int
[1] -0.12 -0.04
```

Interpret precisely what this interval means. Is there evidence of a population-level difference between the obesity proprotions for females and males?
(c) Another variable recorded for each individual in the study was the number of missed work days annually due to health reasons. With data from the 541 females and 3612 males, I calculated the sample variances:

```
> var(female)
[1] 5.62
> var(male)
[1] 4.37
```

The sample variance for the 541 females is $s_1^2 = 5.62$, and the sample variance for the 3612 males is $s_2^2 = 4.37$. How might the population variances σ_1^2 and σ_2^2 compare? Describe an analysis you could do to answer this question. Provide as many details as possible, and cite any concerns you might have with the analysis.

8.3. Hexavalent chromium has been identified as an inhalation carcinogen and an air toxin linked to various cancers. An article published recently in the *Journal of Air and Waste Management Association* gave the data below on both indoor (I) and outdoor (O) concentrations (nanograms/m^3) for a random sample of $n = 33$ houses in southwestern Ontario province in Canada.

House	1	2	3	4	5	6	7	8	9	10	11
Indoor	0.07	0.08	0.09	0.12	0.12	0.12	0.13	0.14	0.15	0.15	0.17
Outdoor	0.29	0.68	0.47	0.54	0.97	0.35	0.49	0.84	0.86	0.28	0.32
House	12	13	14	15	16	17	18	19	20	21	22
Indoor	0.17	0.18	0.18	0.18	0.18	0.19	0.20	0.22	0.22	0.23	0.23
Outdoor	0.32	1.55	0.66	0.29	0.21	1.02	1.59	0.90	0.52	0.12	0.54
House	23	24	25	26	27	28	29	30	31	32	33
Indoor	0.25	0.26	0.28	0.28	0.29	0.34	0.39	0.40	0.45	0.54	0.62
Outdoor	0.88	0.49	1.24	0.48	0.27	0.37	1.26	0.70	0.76	0.99	0.36

(a) Explain why this is a matched pairs study. Are the indoor and outdoor samples independent or dependent? Explain.
(b) I asked R for the data differences (`diff`; indoor minus outdoor) and a 95% confidence interval for the population mean difference in concentrations $\Delta = \mu_I - \mu_O$ acknowledging the matched pairs structure:

```
> diff = indoor-outdoor
> t.test(diff,conf.level=0.95)$conf.int
[1] -0.56 -0.29
```

Interpret this interval, stating what it implies about the larger population of houses in this area.

(c) When I incorrectly analyzed the hexavalent chromium data as data from a two-independent sample design (Sample 1: Indoor; Sample 2: Outdoor), I got almost identical results when compared to the matched pairs analysis:

```
> t.test(indoor,outdoor,conf.level=0.95,var.equal=TRUE)$conf.int
[1] -0.56 -0.28
```

I thought this was surprising. What must be going on here?

8.4. The NFL's Scouting Combine provides an opportunity for participants to display their professional football potential. A special teams coach at this year's event was interested in comparing the population mean punting distance (in yards) between two types of footballs:

Type 1: Air-filled footballs

Type 2: Helium-filled footballs.

To learn how the population mean punting distances might compare, he recruited $n = 15$ punters and had each of them punt each type of ball. The order in which each punter punted an air-filled football and a helium-filled football was randomized. Here were the observed punt distances (measured in yards):

Punter	Air	Helium
1	56.4	55.2
2	44.8	47.9
3	43.7	41.8
4	37.1	38.3
5	33.8	33.5
6	37.1	40.2
7	39.9	43.2
8	33.2	35.5
9	42.8	46.1
10	48.7	50.6
11	32.7	37.3
12	44.6	48.1
13	40.4	40.3
14	44.9	46.1
15	46.1	47.5

(a) The special teams coach is good at coaching but he never went to his statistics class. Searching through an online resource, he found the words "two-sample t confidence interval" and "independent samples" and then calculated a 95% confidence interval for $\Delta = \mu_A - \mu_H$ to be $(-6.4, 3.0)$ yards. I reproduced his analysis in R:

```
> t.test(air,helium,conf.level=0.95,var.equal=TRUE)$conf.int
[1] -6.4  3.0
```

Explain what is wrong with this analysis.
(b) To compare the population means μ_A and μ_H using a confidence interval, the correct analysis involves analyzing the data differences on each punter; i.e.,

$$D_i = \text{Air}_i - \text{Helium}_i,$$

for $i = 1, 2, ..., 15$. Here is this analysis in R:

```
> diff = air-helium
> t.test(diff,conf.level=0.95)$conf.int
[1] -2.7 -0.6
```

Interpret this interval and describe what this suggests about how the population means μ_A and μ_H compare.
(c) Provide a statistical explanation why the confidence interval for $\Delta = \mu_A - \mu_H$ in part (b) is so much shorter than the incorrect interval in part (a).

8.5. I recently reviewed a grant proposal for the Hong Kong Research Grants Council. The proposal described an observational study performed last year involving two independent samples of Hong Kong area high school students:

- students who were "non-heavy" smartphone users ($n_1 = 101$)
- students who were "heavy" smartphone users ($n_2 = 103$).

(a) One variable measured on each student was whether s/he experienced sleep problems or insomnia. There were 8 students in Group 1 ("non-heavy") and 17 students in Group 2 ("heavy") who did. I calculated a 95% confidence interval for the population difference $p_1 - p_2$ ("non-heavy" minus "heavy") using

$$(\widehat{p}_1 - \widehat{p}_2) \pm 1.96\sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}},$$

and I obtained $(-0.175, 0.003)$.

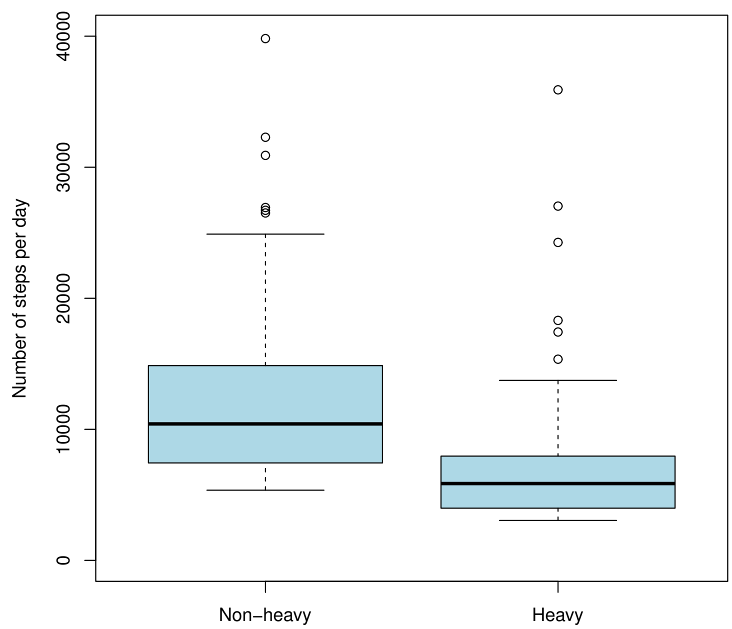
```
prop.test(c(8,17),c(101,103),conf.level=0.95,correct=FALSE)$conf.int
[1] -0.175 0.003
```

Give an interpretation of this interval.
(b) When I changed the confidence level in part (a) to 90%, I got this interval:

```
prop.test(c(8,17),c(101,103),conf.level=0.90,correct=FALSE)$conf.int
[1]  -0.160 -0.011
```

Does it bother you that one interval includes "0," and the other does not? Explain.

(c) Another variable recorded was the number of steps each student took per day (averaged over his/her time in the study). Students wore pedometers each day; these are devices that automatically record this information. Here are side-by-side boxplots of the data collected (again, independent samples; sample sizes given on the previous page):



If the investigators wanted to compare the population mean number of steps per day for these two groups, how would you advise them to do it? Describe what statistical procedure you would use and how you might check the underlying assumptions associated with this procedure. You don't have to perform any calculations here.

8.6. An herbal medicine is tested on $n = 16$ randomly selected patients with sleep disorders. Each patient's amount of sleep (in hours) is measured for one night without the herbal medicine and for one night with the herbal medicine. The data are on the next page.
(a) Explain why this is a matched pairs study.
(b) I calculated a 95% confidence interval for the difference of the population means ("without medicine" minus "with medicine") in R:

```
> diff = without - with
> t.test(diff,conf.level=0.95)$conf.int
[1] -1.8 -1.2
```

Interpret what this interval means precisely and then carefully describe what effect the herbal medicine has on the amount of sleep.
(c) Suppose you wanted to redesign this study using two independent samples of patients: 16 patients randomly assigned to take the herbal medicine and 16 different patients assigned to take a placebo (a pill designed to look like the medicine, but it contains nothing). Note that this design requires 32 patients total. Do you think the observed data from this study would produce a narrower or wider confidence interval when compared to the interval in part (b)? Carefully explain your answer.

Patient	Without	With
1	1.8	3.0
2	2.0	3.6
3	3.4	4.0
4	3.5	4.4
5	3.7	4.5
6	3.8	5.2
7	3.9	5.5
8	3.9	5.7
9	4.0	6.2
10	4.9	6.3
11	5.1	6.6
12	5.2	7.8
13	5.0	7.2
14	4.5	6.5
15	4.2	5.6
16	4.7	5.9

8.7. New York City's Yellow Taxi has about 15,000 taxis in its fleet. A manager is trying to decide whether using radial tires or 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.
- Using the same drivers, the same cars are then equipped with belted tires and are driven through the same test course.

The gasoline consumption (in kilometers 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 radial and belted samples independent or dependent?
(b) I used R to write 95% and 99% confidence intervals for $\Delta = \mu_1 - \mu_2$, the difference of the population mean gasoline consumption for the two types of tires (1 = radial; 2 = belted):

```
> t.test(diff,conf.level=0.95)$conf.int
[1] 0.02 0.27
> t.test(diff,conf.level=0.99)$conf.int
[1] -0.04 0.32
```

(i) What does `diff` mean in the code above? *Hint:* How are data from a matched pairs study analyzed?

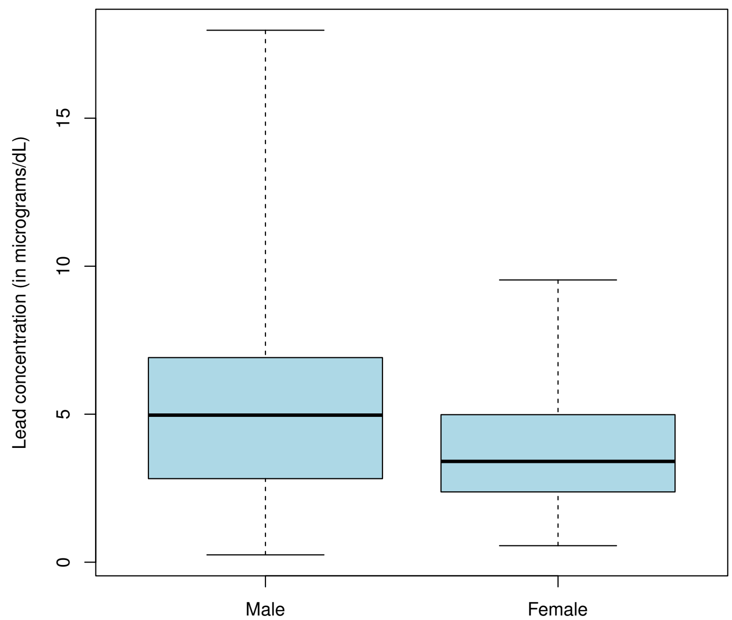
(ii) Pick one confidence interval above (tell me which one) and interpret it for the manager.

(c) The analysis above shows that different conclusions would be made about how the population means compare depending on which confidence level is used. Explain to the manager why this happens and why the two different conclusions would both be supported by the data above.

8.8. Public health officials in New Jersey wanted to compare lead levels in the blood of male and female hazardous waste workers employed in the state. Two independent random samples were observed:

- $n_1 = 152$ male workers
- $n_2 = 86$ female workers.

Lead concentrations (in micrograms per deciliter) were measured on each worker. Here are side-by-side boxplots of the data:



(a) If you wanted to write a confidence interval for $\Delta = \mu_1 - \mu_2$, the difference of the two population mean lead concentrations (1 = male; 2 = female), select the confidence interval you would use:

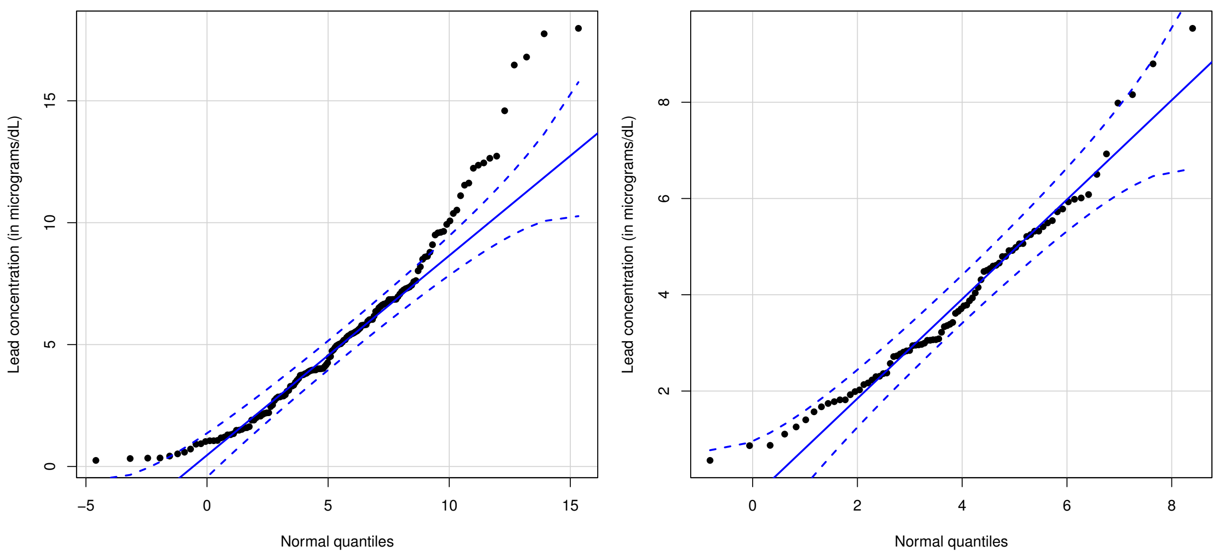
- the one that assumed equal population variances $\sigma_1^2 = \sigma_2^2 = \sigma^2$.
- the one that did not assume the population variances were equal.

Explain why you chose the answer you did. In addition, what statistical inference procedure could be used to determine which assumption is more reasonable?

(b) I asked R to get both confidence intervals in part (a), each at the 95% confidence level. Here are the intervals:

```
> t.test(male,female,conf.level=0.95,var.equal=TRUE)$conf.int
[1] 0.75 2.41
> t.test(male,female,conf.level=0.95,var.equal=FALSE)$conf.int
[1] 0.88 2.28
```

For the interval you picked in part (a), interpret what it means. What does your interval suggest about the mean lead concentrations for the two populations?
(c) Here are the normal quantile-quantile plots for the male (left) and female (right) lead concentrations:



Interpret the plots. Are you concerned about your conclusions in part (b)? Explain.

8.9. A recent article in *Consumer Reports* reported that 32 out of 80 randomly selected Perdue brand chicken broilers tested positively for either campylobacter or salmonella (or both), the leading bacterial causes of food-borne illness, whereas 42 out of 80 randomly selected Tyson brand chicken broilers tested positively.
(a) I used the formula

$$(\widehat{p}_1 - \widehat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}}$$

to write a 90% confidence interval for the population proportion difference $\Delta = p_1 - p_2$, where

$$p_1 = \text{population proportion of bacterial-positive Perdue broilers}$$
$$p_2 = \text{population proportion of bacterial-positive Tyson broilers,}$$

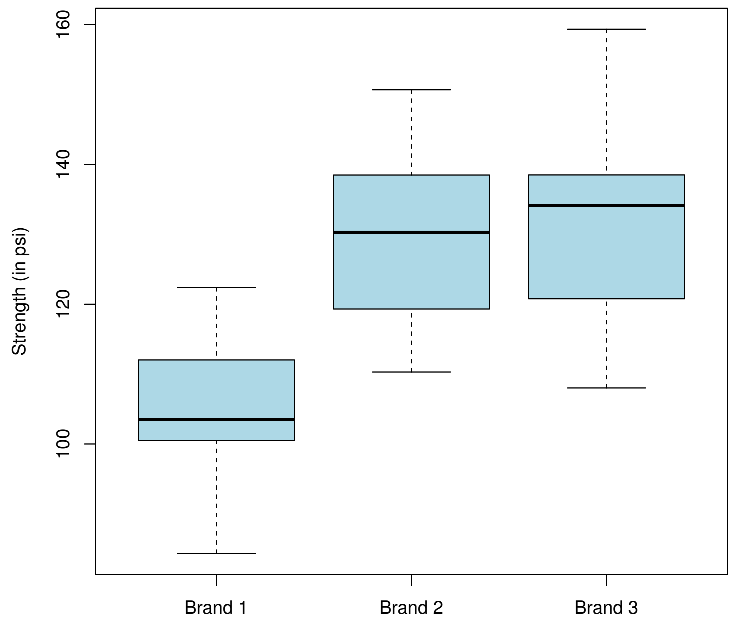
and I got $(-0.254, 0.004)$. Interpret this interval. What does it suggest about how the population proportions compare at the 90% confidence level?

```
> prop.test(c(32,42),c(80,80),conf.level=0.90,correct=FALSE)$conf.int
[1] -0.254  0.004
```

(b) At what confidence level do you think the interval above would exclude "0" as a plausible value for the population proportion difference Δ? There is no one right answer (but there are wrong ones).
(c) Suppose you want to plan a larger study that assumes a 99% confidence level ($z_{0.01/2} = 2.58$). Assuming $n_1 = n_2 = n$, find the smallest sample size n (per brand) that will produce a 99% confidence interval for $\Delta = p_1 - p_2$ with margin of error equal to 0.04. You can use the information in the problem to estimate any parameters needed for this calculation.

CHAPTER 9:

9.1. Investigators wanted to perform a one-way classification analysis with three brands of "double wall" boxes. Twelve boxes of each brand were subjected to a compression test and the strength of each box was measured in pounds per square inch (psi). There were 36 boxes in all; 12 of each brand. Here are side-by-side boxplots of the data:



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

```
> fit = lm(Strength ~ Brand)
> anova(fit)
Analysis of Variance Table

Response: Strength
          Df Sum Sq Mean Sq F value  Pr(>F)
Brand      2   5387  2693.6   16.29 1.2e-05
Residuals 33   5458   165.4
```

(a) The F statistic ($F = 16.29$) 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 must clearly define) or you can write this out in words. Which one of your hypotheses is more supported by the data? Why?
(b) Here is the R output to do a follow-up Tukey analysis:

```
> TukeyHSD(aov(fit),conf.level=0.95)

Tukey multiple comparisons of means
95% family-wise confidence level
                 diff    lwr   upr  p adj
Brand.2-Brand.1 25.25  12.37 38.13 0.0001
Brand.3-Brand.1 26.60  13.72 39.48 0.0001
Brand.3-Brand.2  1.35 -11.53 14.23 0.9643
```

(i) Explain what is meant by "95% family-wise confidence level."

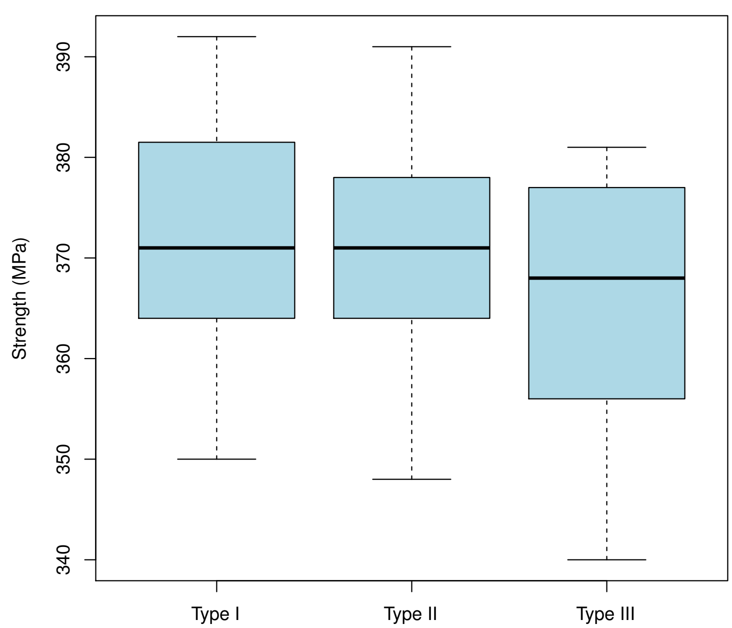
(ii) If you were advising the investigators on which box type to use (to maximize population mean strength), what would you tell them? Defend your conclusions with statistical evidence.

(c) One of the statistical assumptions in a one-way classification analysis is that the variances of the populations being compared are the same. Using the information on the preceding page, report an estimate of what this common population variance σ^2 is. What are the units attached to your estimate?

9.2. Wire rope is a device that helps to support and move an object or load. In the lifting and rigging industries, wire rope is attached to a crane or hoist and fitted with swivels, shackles, or hooks to attach to a load and move it in a controlled matter. It can also be used to lift and lower elevators or as a means of support for suspension bridges or towers. Engineers performed an observational study to compare the population mean strengths of three different types of wire rope. Fifteen specimens of each type were selected, and the strength (in MPa) was determined for each specimen. Here are the data:

Type I:	350	351	352	358	370	370	371	371	372	375	379	384	391	391	392
Type II:	348	354	359	363	365	368	369	371	373	374	376	380	383	388	391
Type III:	340	343	352	354	358	359	361	368	372	373	375	379	380	380	381

Here are side-by-side boxplots of the data:



(a) In performing an analysis of variance, what two hypotheses does the overall F statistic test? You can write your answer out in words, or you can use statistical symbols. If you use symbols, define what the symbols mean.

(b) Here is the R output from the analysis of variance:

```
Response: Strength
          Df Sum Sq Mean Sq F value Pr(>F)
Type       2  404.4  202.20    1.13  0.332
Residuals 42 7504.8  178.69
```

The overall F statistic is $F \approx 1.13$ and the probability value is ≈ 0.33. Using the hypotheses you described in part (a), make a decision as to which hypothesis is more supported by the data. Explain your decision.
(c) Although the overall F test is robust to normality departures, it is not robust to a violation in the equal population variance assumption among the three types of wire rope. Why do you think this is? *Hint:* Think about how the F statistic is created.
(d) The F statistic in the analysis above is close to 1. Sketch a new set of side-by-side boxplots (similar to the figure above) that would produce a really large F statistic, maybe like $F \approx 17$ that we saw with the mortar data in class.

9.3. A recent article in *American Journal of Clinical Nutrition* describes an investigation where 52 preterm infants were randomly assigned to different nutrition regimens:

1. B: Breast milk ($n_1 = 8$ infants)
2. CO: Corn-oil-based formula ($n_2 = 13$ infants)
3. SO: Soy-oil-based formula ($n_3 = 17$ infants)
4. SMO: Soy-and-marine-oil-based formula ($n_4 = 14$ infants).

The response measured on each infant was the total polyunsatruated fat (PUSF) percentage. The goal of the investigation was to determine if there were differences in the population mean PUSF percentage among the four groups. Side-by-side boxplots of the data are on the next page (top).

(a) Looking at the boxplots only, an investigator exclaims,

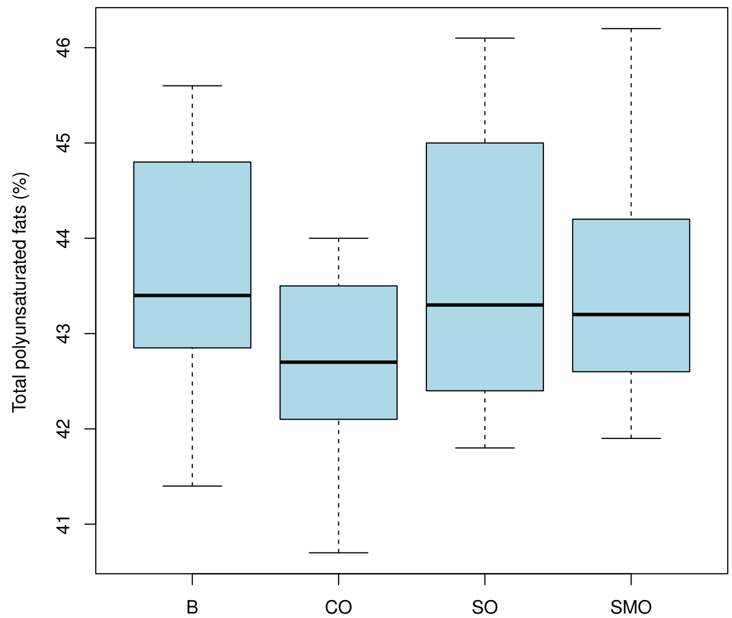
> *"I knew the corn-oil-based formula was best—see, the CO group produces the smallest population mean PUSF percentage."*

How would you respond to the investigator?
(b) Another investigator performs an analysis of variance (ANOVA) with the data; here is the R output from the analysis:

```
> anova(lm(PUSF.percentage ~ nutrition.regimen))

                  Df Sum Sq Mean Sq F value Pr(>F)
nutrition.regimen  3   8.12    2.71    1.60  0.201
Residuals         48  81.03    1.69
```

Interpret these results. Note that the overall F statistic is ≈ 1.60 (p-value ≈ 0.20).



(c) What statistical assumptions are needed for the analysis in part (b)?
(d) If you did a follow-up analysis using Tukey confidence intervals for the $\binom{4}{2} = 6$ pairwise population mean differences, what would you expect to see?

9.4. An observational study was performed to compare the population mean serum alkaline phosphatase (ALP) levels in children with seizures who were receiving anticonvulsant therapy. Forty-five children were found for the study and were categorized into one of four drug groups:

Group 1:	Control (no anticonvulsant drug and/or no history of having seizures)
Group 2:	Phenobarbital
Group 3:	Carbamazepine
Group 4:	Other anticonvulsants.

Using a blood sample from each child, the serum ALP level was recorded (in IU/L, international units per liter). Side-by-side boxplots of the data are on the next page (top).

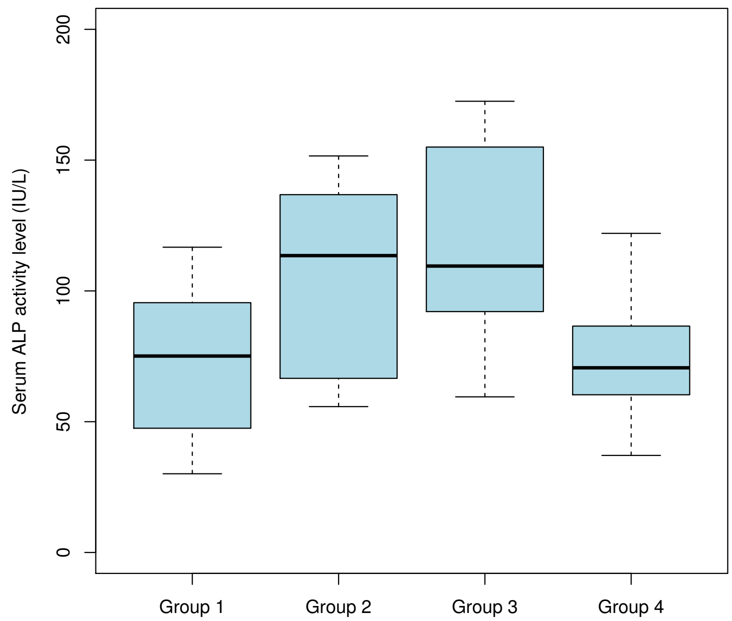
Let μ_i denote the population mean serum ALP level for the ith group ($i = 1, 2, 3, 4$). I used R to perform an analysis of variance (ANOVA) to test

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

Here is the output:

```
> anova(lm(alp.level~group))

Analysis of Variance Table
              Df Sum Sq  Mean Sq F value   Pr(>F)
drug.group     3  15509   5169.7    5.24    0.004 **
Residuals     41  40423    985.9
```



(a) Is the overall F statistic ($F \approx 5.24$) consistent with what we would expect when H_0 is true or when H_1 is true? Explain. If you wish, you can cite the p-value `Pr(>F)` in defending your decision.
(b) As a follow-up, one investigator wants to compare the population mean ALP level for children taking phenobarbital (Group 2) to the population mean ALP level for the control group (Group 1). Here is the 95% Tukey confidence interval for this population mean difference:

```
> TukeyHSD(aov(lm(alp.level~group)),conf.level=0.95)

Tukey multiple comparisons of means
95% family-wise confidence level
                  diff    lwr    upr  p adj
group.2-group.1  28.58  -5.15  62.33   0.12
```

The values `lwr` and `upr` are the lower and upper limits of this interval. Interpret the interval. How do the population mean ALP levels compare for these two groups?
(c) Interestingly, had the investigator not done the overall ANOVA and just focused on the Groups 2 and 1 to begin with, a 95% confidence interval for $\mu_2 - \mu_1$ based on the independent sample and equal population variance assumptions would be

```
> t.test(group.2,group.1,conf.level=0.95,var.equal=TRUE)$conf.int
[1]  3.75 53.43
```

Why is this interval (and its conclusion) so different than the interval in part (b)?
(d) The same investigator in part (c) asks you the following question:

> *"If the analysis of variance procedure is designed to compare population means, why isn't it called the analysis of means?"*

How would you respond?