

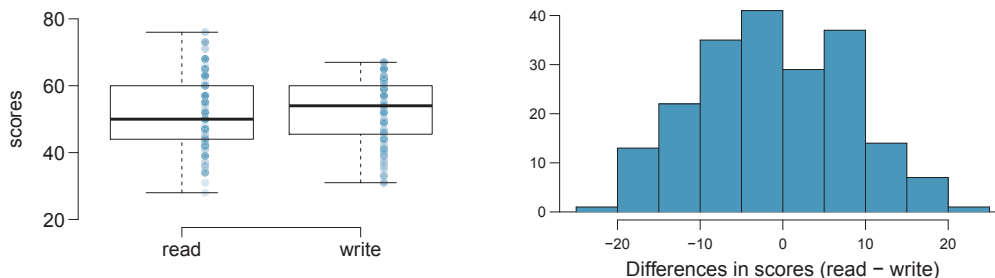
5.6 Exercises

5.6.1 Paired data

5.1 Global warming, Part I. Is there strong evidence of global warming? Let's consider a small scale example, comparing how temperatures have changed in the US from 1968 to 2008. The daily high temperature reading on January 1 was collected in 1968 and 2008 for 51 randomly selected locations in the continental US. Then the difference between the two readings (temperature in 2008 - temperature in 1968) was calculated for each of the 51 different locations. The average of these 51 values was 1.1 degrees with a standard deviation of 4.9 degrees. We are interested in determining whether these data provide strong evidence of temperature warming in the continental US.

- Is there a relationship between the observations collected in 1968 and 2008? Or are the observations in the two groups independent? Explain.
- Write hypotheses for this research in symbols and in words.
- Check the conditions required to complete this test.
- Calculate the test statistic and find the p-value.
- What do you conclude? Interpret your conclusion in context.
- What type of error might we have made? Explain in context what the error means.
- Based on the results of this hypothesis test, would you expect a confidence interval for the average difference between the temperature measurements from 1968 and 2008 to include 0? Explain your reasoning.

5.2 High School and Beyond, Part I. The National Center of Education Statistics conducted a survey of high school seniors, collecting test data on reading, writing, and several other subjects. Here we examine a simple random sample of 200 students from this survey. Side-by-side box plots of reading and writing scores as well as a histogram of the differences in scores are shown below.



- Is there a clear difference in the average reading and writing scores?
- Are the reading and writing scores of each student independent of each other?
- Create hypotheses appropriate for the following research question: is there an evident difference in the average scores of students in the reading and writing exam?
- Check the conditions required to complete this test.
- The average observed difference in scores is $\bar{x}_{read-write} = -0.545$, and the standard deviation of the differences is 8.887 points. Do these data provide convincing evidence of a difference between the average scores on the two exams?
- What type of error might we have made? Explain what the error means in the context of the application.
- Based on the results of this hypothesis test, would you expect a confidence interval for the average difference between the reading and writing scores to include 0? Explain your reasoning.

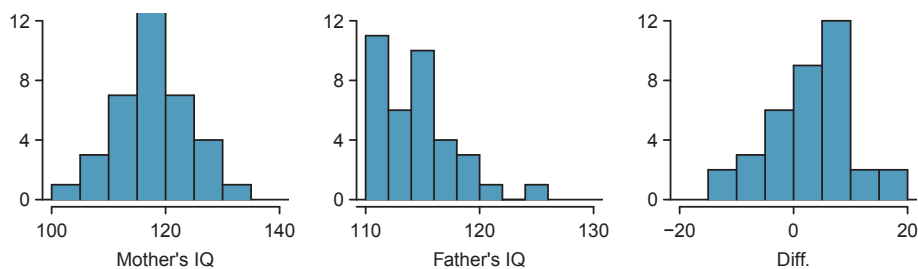
5.3 Global warming, Part II. We considered the differences between the temperature readings in January 1 of 1968 and 2008 at 51 locations in the continental US in Exercise 5.1. The mean and standard deviation of the reported differences are 1.1 degrees and 4.9 degrees.

- Calculate a 90% confidence interval for the average difference between the temperature measurements between 1968 and 2008.
- Interpret this interval in context.
- Does the confidence interval provide convincing evidence that the temperature was higher in 2008 than in 1968 in the continental US? Explain.

5.4 High school and beyond, Part II. We considered the differences between the reading and writing scores of a random sample of 200 students who took the High School and Beyond Survey in Exercise 5.3. The mean and standard deviation of the differences are $\bar{x}_{read-write} = -0.545$ and 8.887 points.

- Calculate a 95% confidence interval for the average difference between the reading and writing scores of all students.
- Interpret this interval in context.
- Does the confidence interval provide convincing evidence that there is a real difference in the average scores? Explain.

5.5 Gifted children. Researchers collected a simple random sample of 36 children who had been identified as gifted in a large city. The following histograms show the distributions of the IQ scores of mothers and fathers of these children. Also provided are some sample statistics.³⁵



	Mother	Father	Diff.
Mean	118.2	114.8	3.4
SD	6.5	3.5	7.5
n	36	36	36

- Are the IQs of mothers and the IQs of fathers in this data set related? Explain.
- Conduct a hypothesis test to evaluate if the scores are equal on average. Make sure to clearly state your hypotheses, check the relevant conditions, and state your conclusion in the context of the data.

5.6 Paired or not? In each of the following scenarios, determine if the data are paired.

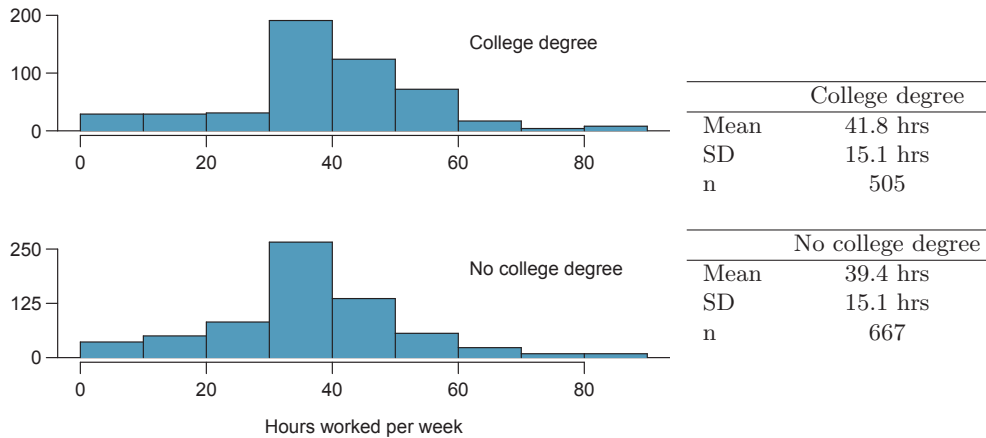
- We would like to know if Intel's stock and Southwest Airlines' stock have similar rates of return. To find out, we take a random sample of 50 days for Intel's stock and another random sample of 50 days for Southwest's stock.
- We randomly sample 50 items from Target stores and note the price for each. Then we visit Walmart and collect the price for each of those same 50 items.
- A school board would like to determine whether there is a difference in average SAT scores for students at one high school versus another high school in the district. To check, they take a simple random sample of 100 students from each high school.

³⁵F.A. Graybill and H.K. Iyer. *Regression Analysis: Concepts and Applications*. Duxbury Press, 1994, pp. 511–516.

5.6.2 Difference of two means

5.7 Math scores of 13 year olds, Part I. The National Assessment of Educational Progress tested a simple random sample of 1,000 thirteen year old students in both 2004 and 2008 (two separate simple random samples). The average and standard deviation in 2004 were 257 and 39, respectively. In 2008, the average and standard deviation were 260 and 38, respectively. Calculate a 90% confidence interval for the change in average scores from 2004 to 2008, and interpret this interval in the context of the application. (Reminder: check conditions.)³⁶

5.8 Work hours and education, Part I. The General Social Survey collects data on demographics, education, and work, among many other characteristics of US residents. The histograms below display the distributions of hours worked per week for two education groups: those with and without a college degree.³⁷ Suppose we want to estimate the average difference between the number of hours worked per week by all Americans with a college degree and those without a college degree. Summary information for each group is shown in the tables.



- What is the parameter of interest, and what is the point estimate?
- Are conditions satisfied for estimating this difference using a confidence interval?
- Create a 95% confidence interval for the difference in number of hours worked between the two groups, and interpret the interval in context.
- Can you think of any real world justification for your results? (*Note:* There isn't a single correct answer to this question.)

5.9 Math scores of 13 year olds, Part II. Exercise 5.7 provides data on the average math scores from tests conducted by the National Assessment of Educational Progress in 2004 and 2008. Two separate simple random samples, each of size 1,000, were taken in each of these years. The average and standard deviation in 2004 were 257 and 39, respectively. In 2008, the average and standard deviation were 260 and 38, respectively.

- Do these data provide strong evidence that the average math score for 13 year old students has changed from 2004 to 2008? Use a 10% significance level.
- It is possible that your conclusion in part (a) is incorrect. What type of error is possible for this conclusion? Explain.
- Based on your hypothesis test, would you expect a 90% confidence interval to contain the null value? Explain.

³⁶National Center for Education Statistics, NAEP Data Explorer.

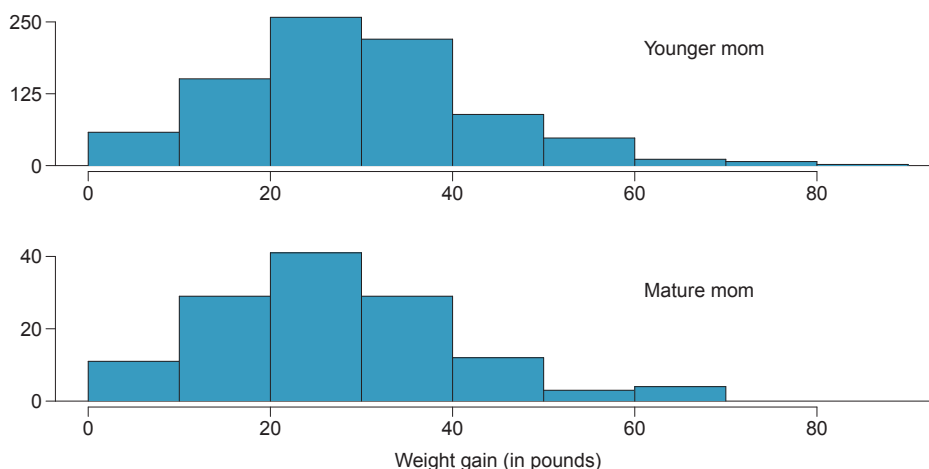
³⁷National Opinion Research Center, General Social Survey, 2010.

5.10 Work hours and education, Part II. The General Social Survey described in Exercise 5.8 included random samples from two groups: US residents with a college degree and US residents without a college degree. For the 505 sampled US residents with a college degree, the average number of hours worked each week was 41.8 hours with a standard deviation of 15.1 hours. For those 667 without a degree, the mean was 39.4 hours with a standard deviation of 15.1 hours. Conduct a hypothesis test to check for a difference in the average number of hours worked for the two groups.

5.11 Does the Paleo diet work? The Paleo diet allows only for foods that humans typically consumed over the last 2.5 million years, excluding those agriculture-type foods that arose during the last 10,000 years or so. Researchers randomly divided 500 volunteers into two equal-sized groups. One group spent 6 months on the Paleo diet. The other group received a pamphlet about controlling portion sizes. Randomized treatment assignment was performed, and at the beginning of the study, the average difference in weights between the two groups was about 0. After the study, the Paleo group had lost on average 7 pounds with a standard deviation of 20 pounds while the control group had lost on average 5 pounds with a standard deviation of 12 pounds.

- The 95% confidence interval for the difference between the two population parameters (Paleo - control) is given as $(-0.891, 4.891)$. Interpret this interval in the context of the data.
- Based on this confidence interval, do the data provide convincing evidence that the Paleo diet is more effective for weight loss than the pamphlet (control)? Explain your reasoning.
- Without explicitly performing the hypothesis test, do you think that if the Paleo group had lost 8 instead of 7 pounds on average, and everything else was the same, the results would then indicate a significant difference between the treatment and control groups? Explain your reasoning.

5.12 Weight gain during pregnancy. In 2004, the state of North Carolina released to the public a large data set containing information on births recorded in this state. This data set has been of interest to medical researchers who are studying the relationship between habits and practices of expectant mothers and the birth of their children. The following histograms show the distributions of weight gain during pregnancy by 867 younger moms (less than 35 years old) and 133 mature moms (35 years old and over) who have been randomly sampled from this large data set. The average weight gain of younger moms is 30.56 pounds, with a standard deviation of 14.35 pounds, and the average weight gain of mature moms is 28.79 pounds, with a standard deviation of 13.48 pounds. Calculate a 95% confidence interval for the difference between the average weight gain of younger and mature moms. Also comment on whether or not this interval provides strong evidence that there is a significant difference between the two population means.



5.19 Sleep habits of New Yorkers. New York is known as “the city that never sleeps”. A random sample of 25 New Yorkers were asked how much sleep they get per night. Statistical summaries of these data are shown below. Do these data provide strong evidence that New Yorkers sleep less than 8 hours a night on average?

n	\bar{x}	s	min	max
25	7.73	0.77	6.17	9.78

- Write the hypotheses in symbols and in words.
- Check conditions, then calculate the test statistic, T , and the associated degrees of freedom.
- Find and interpret the p-value in this context. Drawing a picture may be helpful.
- What is the conclusion of the hypothesis test?
- If you were to construct a 90% confidence interval that corresponded to this hypothesis test, would you expect 8 hours to be in the interval?

5.20 Fuel efficiency of Prius. Fueleconomy.gov, the official US government source for fuel economy information, allows users to share gas mileage information on their vehicles. The histogram below shows the distribution of gas mileage in miles per gallon (MPG) from 14 users who drive a 2012 Toyota Prius. The sample mean is 53.3 MPG and the standard deviation is 5.2 MPG. Note that these data are user estimates and since the source data cannot be verified, the accuracy of these estimates are not guaranteed.⁴⁰



- We would like to use these data to evaluate the average gas mileage of all 2012 Prius drivers. Do you think this is reasonable? Why or why not?
- The EPA claims that a 2012 Prius gets 50 MPG (city and highway mileage combined). Do these data provide strong evidence against this estimate for drivers who participate on fueleconomy.gov? Note any assumptions you must make as you proceed with the test.
- Calculate a 95% confidence interval for the average gas mileage of a 2012 Prius by drivers who participate on fueleconomy.gov.

5.21 Find the mean. You are given the following hypotheses:

$$H_0 : \mu = 60$$

$$H_A : \mu < 60$$

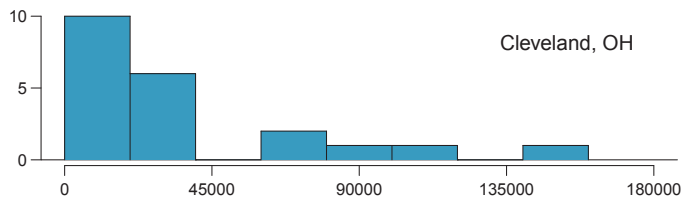
We know that the sample standard deviation is 8 and the sample size is 20. For what sample mean would the p-value be equal to 0.05? Assume that all conditions necessary for inference are satisfied.

5.22 t^* vs. z^* . For a given confidence level, t_{df}^* is larger than z^* . Explain how t_{df}^* being slightly larger than z^* affects the width of the confidence interval.

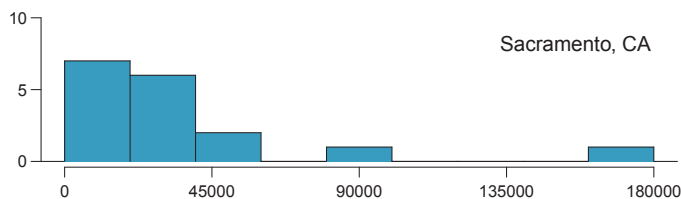
⁴⁰Fueleconomy.gov, Shared MPG Estimates: Toyota Prius 2012.

5.6.4 The t distribution for the difference of two means

5.23 Cleveland vs. Sacramento. Average income varies from one region of the country to another, and it often reflects both lifestyles and regional living expenses. Suppose a new graduate is considering a job in two locations, Cleveland, OH and Sacramento, CA, and he wants to see whether the average income in one of these cities is higher than the other. He would like to conduct a t test based on two small samples from the 2000 Census, but he first must consider whether the conditions are met to implement the test. Below are histograms for each city. Should he move forward with the t test? Explain your reasoning.



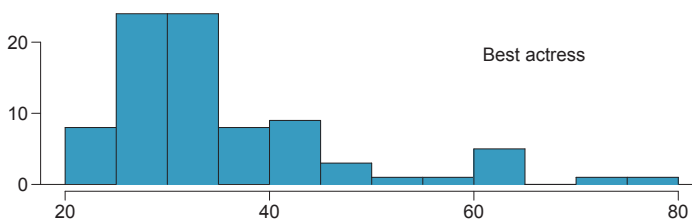
Cleveland, OH	
Mean	\$ 35,749
SD	\$ 39,421
n	21



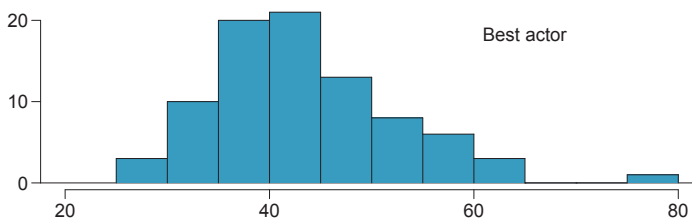
Sacramento, CA	
Mean	\$ 35,500
SD	\$ 41,512
n	17

Total personal income

5.24 Oscar winners. The first Oscar awards for best actor and best actress were given out in 1929. The histograms below show the age distribution for all of the best actor and best actress winners from 1929 to 2012. Summary statistics for these distributions are also provided. Is a t test appropriate for evaluating whether the difference in the average ages of best actors and actresses might be due to chance? Explain your reasoning.⁴¹



Best Actress	
Mean	35.6
SD	11.3
n	84

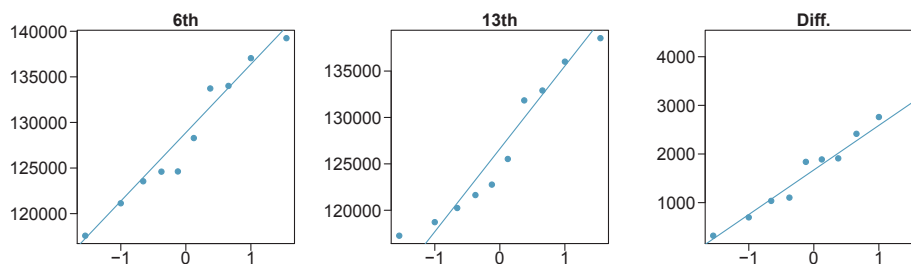


Best Actor	
Mean	44.7
SD	8.9
n	84

Ages (in years)

⁴¹Oscar winners from 1929 – 2012, data up to 2009 from the Journal of Statistics Education data archive and more current data from wikipedia.org.

5.25 Friday the 13th, Part I. In the early 1990's, researchers in the UK collected data on traffic flow, number of shoppers, and traffic accident related emergency room admissions on Friday the 13th and the previous Friday, Friday the 6th. The histograms below show the distribution of number of cars passing by a specific intersection on Friday the 6th and Friday the 13th for many such date pairs. Also given are some sample statistics, where the difference is the number of cars on the 6th minus the number of cars on the 13th.⁴²



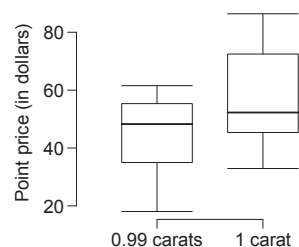
	6 th	13 th	Diff.
\bar{x}	128,385	126,550	1,835
s	7,259	7,664	1,176
n	10	10	10

- Are there any underlying structures in these data that should be considered in an analysis? Explain.
- What are the hypotheses for evaluating whether the number of people out on Friday the 6th is different than the number out on Friday the 13th?
- Check conditions to carry out the hypothesis test from part (b).
- Calculate the test statistic and the p-value.
- What is the conclusion of the hypothesis test?
- Interpret the p-value in this context.
- What type of error might have been made in the conclusion of your test? Explain.

5.26 Diamonds, Part I. Prices of diamonds are determined by what is known as the 4 Cs: cut, clarity, color, and carat weight. The prices of diamonds go up as the carat weight increases, but the increase is not smooth. For example, the difference between the size of a 0.99 carat diamond and a 1 carat diamond is undetectable to the naked human eye, but the price of a 1 carat diamond tends to be much higher than the price of a 0.99 diamond. In this question we use two random samples of diamonds, 0.99 carats and 1 carat, each sample of size 23, and compare the average prices of the diamonds. In order to be able to compare equivalent units, we first divide the price for each diamond by 100 times its weight in carats. That is, for a 0.99 carat diamond, we divide the price by 99. For a 1 carat diamond, we divide the price by 100. The distributions and some sample statistics are shown below.⁴³

Conduct a hypothesis test to evaluate if there is a difference between the average standardized prices of 0.99 and 1 carat diamonds. Make sure to state your hypotheses clearly, check relevant conditions, and interpret your results in context of the data.

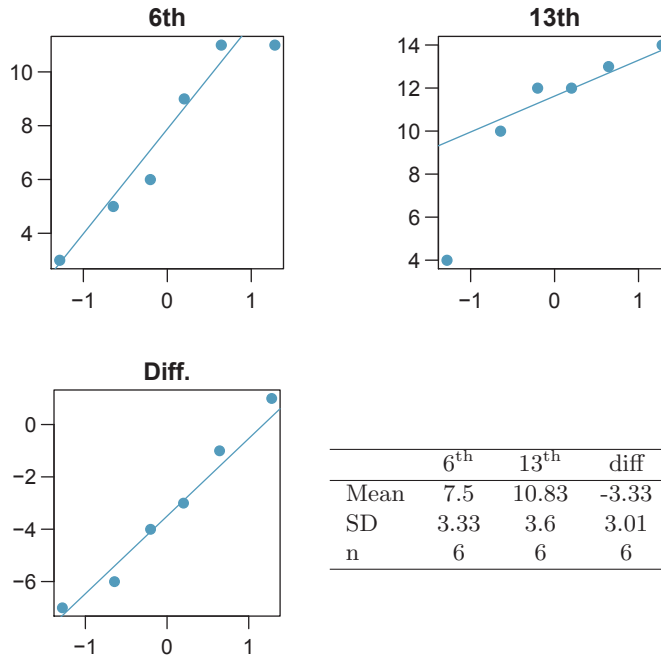
	0.99 carats	1 carat
Mean	\$ 44.51	\$ 56.81
SD	\$ 13.32	\$ 16.13
n	23	23



⁴²T.J. Scanlon et al. "Is Friday the 13th Bad For Your Health?" In: *BMJ* 307 (1993), pp. 1584–1586.

⁴³H. Wickham. *ggplot2: elegant graphics for data analysis*. Springer New York, 2009.

5.27 Friday the 13th, Part II. The Friday the 13th study reported in Exercise 5.25 also provides data on traffic accident related emergency room admissions. The distributions of these counts from Friday the 6th and Friday the 13th are shown below for six such paired dates along with summary statistics. You may assume that conditions for inference are met.

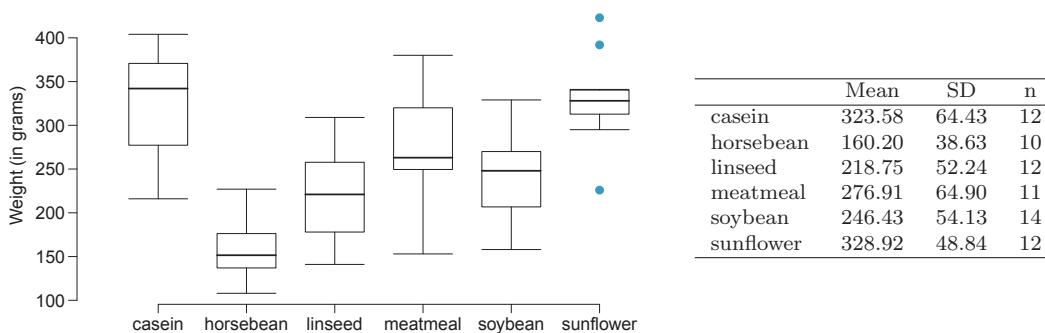


- Conduct a hypothesis test to evaluate if there is a difference between the average numbers of traffic accident related emergency room admissions between Friday the 6th and Friday the 13th.
- Calculate a 95% confidence interval for the difference between the average numbers of traffic accident related emergency room admissions between Friday the 6th and Friday the 13th.
- The conclusion of the original study states, “Friday 13th is unlucky for some. The risk of hospital admission as a result of a transport accident may be increased by as much as 52%. Staying at home is recommended.” Do you agree with this statement? Explain your reasoning.

5.28 Diamonds, Part II. In Exercise 5.26, we discussed diamond prices (standardized by weight) for diamonds with weights 0.99 carats and 1 carat. See the table for summary statistics, and then construct a 95% confidence interval for the average difference between the standardized prices of 0.99 and 1 carat diamonds. You may assume the conditions for inference are met.

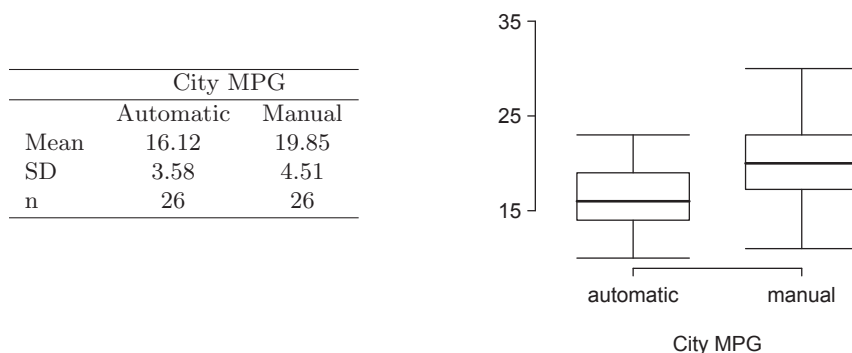
	0.99 carats	1 carat
Mean	\$ 44.51	\$ 56.81
SD	\$ 13.32	\$ 16.13
n	23	23

5.29 Chicken diet and weight, Part I. Chicken farming is a multi-billion dollar industry, and any methods that increase the growth rate of young chicks can reduce consumer costs while increasing company profits, possibly by millions of dollars. An experiment was conducted to measure and compare the effectiveness of various feed supplements on the growth rate of chickens. Newly hatched chicks were randomly allocated into six groups, and each group was given a different feed supplement. Below are some summary statistics from this data set along with box plots showing the distribution of weights by feed type.⁴⁴



- Describe the distributions of weights of chickens that were fed linseed and horsebean.
- Do these data provide strong evidence that the average weights of chickens that were fed linseed and horsebean are different? Use a 5% significance level.
- What type of error might we have committed? Explain.
- Would your conclusion change if we used $\alpha = 0.01$?

5.30 Fuel efficiency of manual and automatic cars, Part I. Each year the US Environmental Protection Agency (EPA) releases fuel economy data on cars manufactured in that year. Below are summary statistics on fuel efficiency (in miles/gallon) from random samples of cars with manual and automatic transmissions manufactured in 2012. Do these data provide strong evidence of a difference between the average fuel efficiency of cars with manual and automatic transmissions in terms of their average city mileage? Assume that conditions for inference are satisfied.⁴⁵



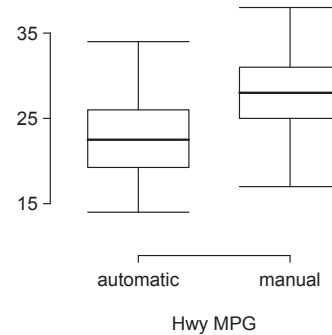
5.31 Chicken diet and weight, Part II. Casein is a common weight gain supplement for humans. Does it have an effect on chickens? Using data provided in Exercise 5.29, test the hypothesis that the average weight of chickens that were fed casein is different than the average weight of chickens that were fed soybean. If your hypothesis test yields a statistically significant result, discuss whether or not the higher average weight of chickens can be attributed to the casein diet. Assume that conditions for inference are satisfied.

⁴⁴Chicken Weights by Feed Type, from the `datasets` package in R..

⁴⁵U.S. Department of Energy, Fuel Economy Data, 2012 Datafile.

5.32 Fuel efficiency of manual and automatic cars, Part II. The table provides summary statistics on highway fuel economy of cars manufactured in 2012 (from Exercise 5.30). Use these statistics to calculate a 98% confidence interval for the difference between average highway mileage of manual and automatic cars, and interpret this interval in the context of the data. Assume the conditions for inference are satisfied.⁴⁶

	Hwy MPG	
	Automatic	Manual
Mean	22.92	27.88
SD	5.29	5.01
n	26	26



5.33 Gaming and distracted eating, Part I. A group of researchers are interested in the possible effects of distracting stimuli during eating, such as an increase or decrease in the amount of food consumption. To test this hypothesis, they monitored food intake for a group of 44 patients who were randomized into two equal groups. The treatment group ate lunch while playing solitaire, and the control group ate lunch without any added distractions. Patients in the treatment group ate 52.1 grams of biscuits, with a standard deviation of 45.1 grams, and patients in the control group ate 27.1 grams of biscuits, with a standard deviation of 26.4 grams. Do these data provide convincing evidence that the average food intake (measured in amount of biscuits consumed) is different for the patients in the treatment group? Assume that conditions for inference are satisfied.⁴⁷

5.34 Gaming and distracted eating, Part II. The researchers from Exercise 5.33 also investigated the effects of being distracted by a game on how much people eat. The 22 patients in the treatment group who ate their lunch while playing solitaire were asked to do a serial-order recall of the food lunch items they ate. The average number of items recalled by the patients in this group was 4.9, with a standard deviation of 1.8. The average number of items recalled by the patients in the control group (no distraction) was 6.1, with a standard deviation of 1.8. Do these data provide strong evidence that the average number of food items recalled by the patients in the treatment and control groups are different?

5.35 Prison isolation experiment, Part I. Subjects from Central Prison in Raleigh, NC, volunteered for an experiment involving an “isolation” experience. The goal of the experiment was to find a treatment that reduces subjects’ psychopathic deviant T scores. This score measures a person’s need for control or their rebellion against control, and it is part of a commonly used mental health test called the Minnesota Multiphasic Personality Inventory (MMPI) test. The experiment had three treatment groups:

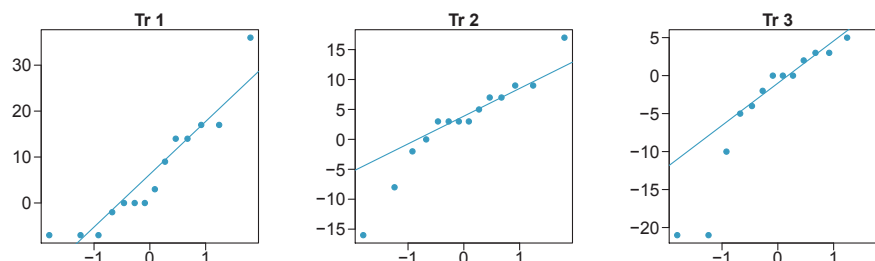
- (1) Four hours of sensory restriction plus a 15 minute “therapeutic” tape advising that professional help is available.
- (2) Four hours of sensory restriction plus a 15 minute “emotionally neutral” tape on training hunting dogs.
- (3) Four hours of sensory restriction but no taped message.

Forty-two subjects were randomly assigned to these treatment groups, and an MMPI test was administered before and after the treatment. Distributions of the differences between pre and

⁴⁶U.S. Department of Energy, Fuel Economy Data, 2012 Datafile.

⁴⁷R.E. Oldham-Cooper et al. “Playing a computer game during lunch affects fullness, memory for lunch, and later snack intake”. In: *The American Journal of Clinical Nutrition* 93.2 (2011), p. 308.

post treatment scores (pre - post) are shown below, along with some sample statistics. Use this information to independently test the effectiveness of each treatment. Make sure to clearly state your hypotheses, check conditions, and interpret results in the context of the data.⁴⁸



	Tr 1	Tr 2	Tr 3
Mean	6.21	2.86	-3.21
SD	12.3	7.94	8.57
n	14	14	14

5.36 True or false, Part I. Determine if the following statements are true or false, and explain your reasoning for statements you identify as false.

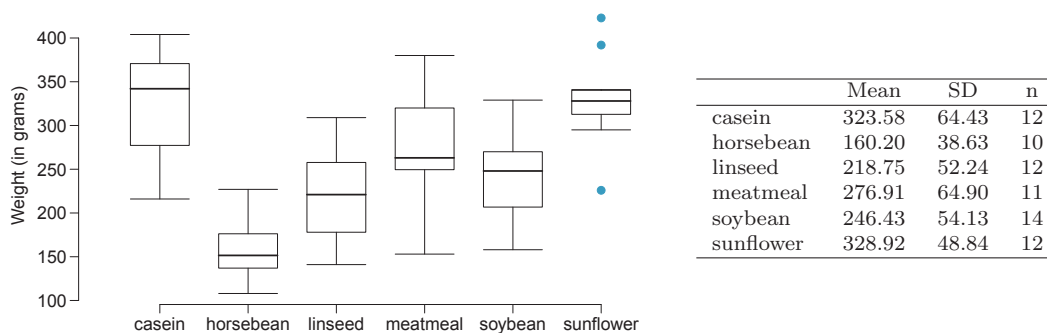
- When comparing means of two samples where $n_1 = 20$ and $n_2 = 40$, we can use the normal model for the difference in means since $n_2 \geq 30$.
- As the degrees of freedom increases, the T distribution approaches normality.
- We use a pooled standard error for calculating the standard error of the difference between means when sample sizes of groups are equal to each other.

5.6.5 Comparing many means with ANOVA

5.37 Chicken diet and weight, Part III. In Exercises 5.29 and 5.31 we compared the effects of two types of feed at a time. A better analysis would first consider all feed types at once: casein, horsebean, linseed, meat meal, soybean, and sunflower. The ANOVA output below can be used to test for differences between the average weights of chicks on different diets.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
feed	5	231,129.16	46,225.83	15.36	0.0000
Residuals	65	195,556.02	3,008.55		

Conduct a hypothesis test to determine if these data provide convincing evidence that the average weight of chicks varies across some (or all) groups. Make sure to check relevant conditions. Figures and summary statistics are shown below.



⁴⁸Prison isolation experiment.

5.38 Student performance across discussion sections. A professor who teaches a large introductory statistics class (197 students) with eight discussion sections would like to test if student performance differs by discussion section, where each discussion section has a different teaching assistant. The summary table below shows the average final exam score for each discussion section as well as the standard deviation of scores and the number of students in each section.

	Sec 1	Sec 2	Sec 3	Sec 4	Sec 5	Sec 6	Sec 7	Sec 8
n_i	33	19	10	29	33	10	32	31
\bar{x}_i	92.94	91.11	91.80	92.45	89.30	88.30	90.12	93.35
s_i	4.21	5.58	3.43	5.92	9.32	7.27	6.93	4.57

The ANOVA output below can be used to test for differences between the average scores from the different discussion sections.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
section	7	525.01	75.00	1.87	0.0767
Residuals	189	7584.11	40.13		

Conduct a hypothesis test to determine if these data provide convincing evidence that the average score varies across some (or all) groups. Check conditions and describe any assumptions you must make to proceed with the test.

5.39 Coffee, depression, and physical activity. Caffeine is the world's most widely used stimulant, with approximately 80% consumed in the form of coffee. Participants in a study investigating the relationship between coffee consumption and exercise were asked to report the number of hours they spent per week on moderate (e.g., brisk walking) and vigorous (e.g., strenuous sports and jogging) exercise. Based on these data the researchers estimated the total hours of metabolic equivalent tasks (MET) per week, a value always greater than 0. The table below gives summary statistics of MET for women in this study based on the amount of coffee consumed.⁴⁹

	<i>Caffeinated coffee consumption</i>					Total
	≤ 1 cup/week	2-6 cups/week	1 cup/day	2-3 cups/day	≥ 4 cups/day	
Mean	18.7	19.6	19.3	18.9	17.5	
SD	21.1	25.5	22.5	22.0	22.0	
n	12,215	6,617	17,234	12,290	2,383	50,739

- Write the hypotheses for evaluating if the average physical activity level varies among the different levels of coffee consumption.
- Check conditions and describe any assumptions you must make to proceed with the test.
- Below is part of the output associated with this test. Fill in the empty cells.

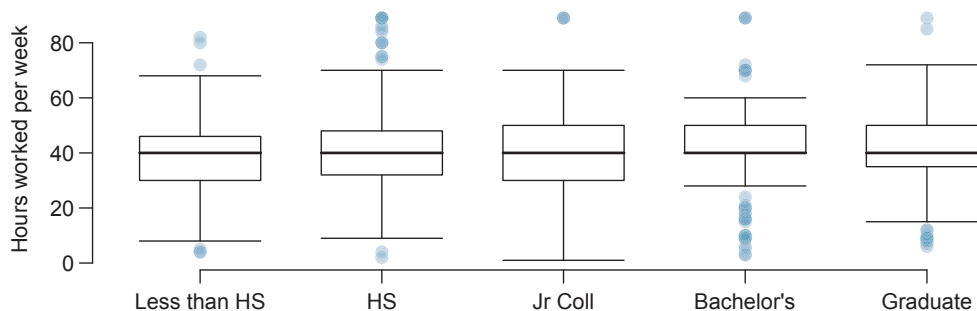
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
coffee	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	0.0003
Residuals	<input type="text"/>	25,564,819	<input type="text"/>		
Total	<input type="text"/>	25,575,327			

- What is the conclusion of the test?

⁴⁹M. Lucas et al. "Coffee, caffeine, and risk of depression among women". In: *Archives of internal medicine* 171.17 (2011), p. 1571.

5.40 Work hours and education, Part III. In Exercises 5.8 and 5.10 you worked with data from the General Social Survey in order to compare the average number of hours worked per week by US residents with and without a college degree. However, this analysis didn't take advantage of the original data which contained more accurate information on educational attainment (less than high school, high school, junior college, Bachelor's, and graduate school). Using ANOVA, we can consider educational attainment levels for all 1,172 respondents at once instead of re-categorizing them into two groups. Below are the distributions of hours worked by educational attainment and relevant summary statistics that will be helpful in carrying out this analysis.

	<i>Educational attainment</i>					
	Less than HS	HS	Jr Coll	Bachelor's	Graduate	Total
Mean	38.67	39.6	41.39	42.55	40.85	40.45
SD	15.81	14.97	18.1	13.62	15.51	15.17
n	121	546	97	253	155	1,172

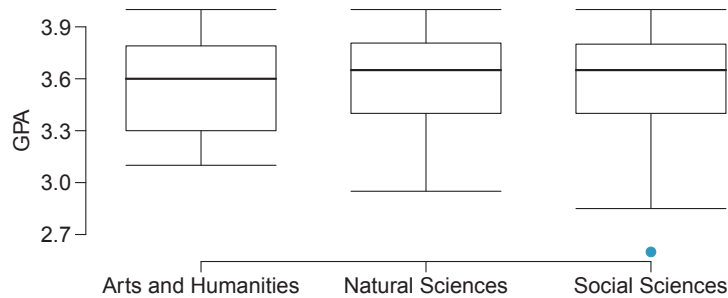


- Write hypotheses for evaluating whether the average number of hours worked varies across the five groups.
- Check conditions and describe any assumptions you must make to proceed with the test.
- Below is part of the output associated with this test. Fill in the empty cells.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
degree	<input type="text"/>	<input type="text"/>	501.54	<input type="text"/>	0.0682
Residuals	<input type="text"/>	267,382	<input type="text"/>		
Total	<input type="text"/>	<input type="text"/>			

- What is the conclusion of the test?

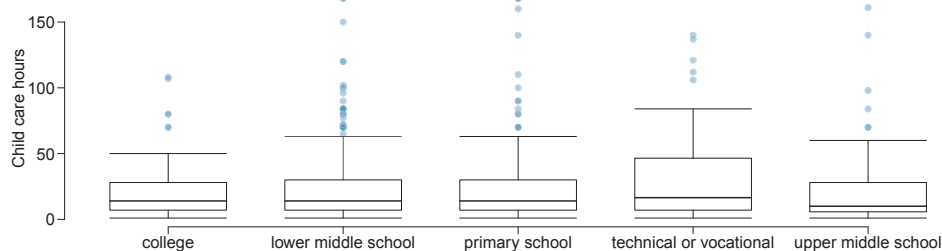
5.41 GPA and major. Undergraduate students taking an introductory statistics course at Duke University conducted a survey about GPA and major. The side-by-side box plots show the distribution of GPA among three groups of majors. Also provided is the ANOVA output.



	Df	Sum Sq	Mean Sq	F value	Pr(>F)
major	2	0.03	0.02	0.21	0.8068
Residuals	195	15.77	0.08		

- Write the hypotheses for testing for a difference between average GPA across majors.
- What is the conclusion of the hypothesis test?
- How many students answered these questions on the survey, i.e. what is the sample size?

5.42 Child care hours, Part II. Exercise 5.14 introduces the China Health and Nutrition Survey which, among other things, collects information on number of hours Chinese parents spend taking care of their children under age 6. The side by side box plots below show the distribution of this variable by educational attainment of the parent. Also provided below is the ANOVA output for comparing average hours across educational attainment categories.



	Df	Sum Sq	Mean Sq	F value	Pr(>F)
education	4	4142.09	1035.52	1.26	0.2846
Residuals	794	653047.83	822.48		

- Write the hypotheses for testing for a difference between the average number of hours spent on child care across educational attainment levels.
- What is the conclusion of the hypothesis test?

5.43 True or false, Part II. Determine if the following statements are true or false in ANOVA, and explain your reasoning for statements you identify as false.

- (a) As the number of groups increases, the modified significance level for pairwise tests increases as well.
- (b) As the total sample size increases, the degrees of freedom for the residuals increases as well.
- (c) The constant variance condition can be somewhat relaxed when the sample sizes are relatively consistent across groups.
- (d) The independence assumption can be relaxed when the total sample size is large.

5.44 True or false, Part III. Determine if the following statements are true or false, and explain your reasoning for statements you identify as false.

If the null hypothesis that the means of four groups are all the same is rejected using ANOVA at a 5% significance level, then ...

- (a) we can then conclude that all the means are different from one another.
- (b) the standardized variability between groups is higher than the standardized variability within groups.
- (c) the pairwise analysis will identify at least one pair of means that are significantly different.
- (d) the appropriate α to be used in pairwise comparisons is $0.05 / 4 = 0.0125$ since there are four groups.

5.45 Prison isolation experiment, Part II. Exercise 5.35 introduced an experiment that was conducted with the goal of identifying a treatment that reduces subjects' psychopathic deviant T scores, where this score measures a person's need for control or his rebellion against control. In Exercise 5.35 you evaluated the success of each treatment individually. An alternative analysis involves comparing the success of treatments. The relevant ANOVA output is given below.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
treatment	2	639.48	319.74	3.33	0.0461
Residuals	39	3740.43	95.91		

Spooled = 9.793 on *df* = 39

- (a) What are the hypotheses?
- (b) What is the conclusion of the test? Use a 5% significance level.
- (c) If in part (b) you determined that the test is significant, conduct pairwise tests to determine which groups are different from each other. If you did not reject the null hypothesis in part (b), recheck your solution.