

5.6.4 Power calculations for a difference of means

5.39 Increasing corn yield. A large farm wants to try out a new type of fertilizer to evaluate whether it will improve the farm's corn production. The land is broken into plots that produce an average of 1,215 pounds of corn with a standard deviation of 94 pounds per plot. The owner is interested in detecting any average difference of at least 40 pounds per plot. How many plots of land would be needed for the experiment if the desired power level is 90%? Assume each plot of land gets treated with either the current fertilizer or the new fertilizer.

5.40 Email outreach efforts. A medical research group is recruiting people to complete short surveys about their medical history. For example, one survey asks for information on a person's family history in regards to cancer. Another survey asks about what topics were discussed during the person's last visit to a hospital. So far, as people sign up, they complete an average of just 4 surveys, and the standard deviation of the number of surveys is about 2.2. The research group wants to try a new interface that they think will encourage new enrollees to complete more surveys, where they will randomize each enrollee to either get the new interface or the current interface. How many new enrollees do they need for each interface to detect an effect size of 0.5 surveys per enrollee, if the desired power level is 80%?

5.6.5 Comparing many means with ANOVA

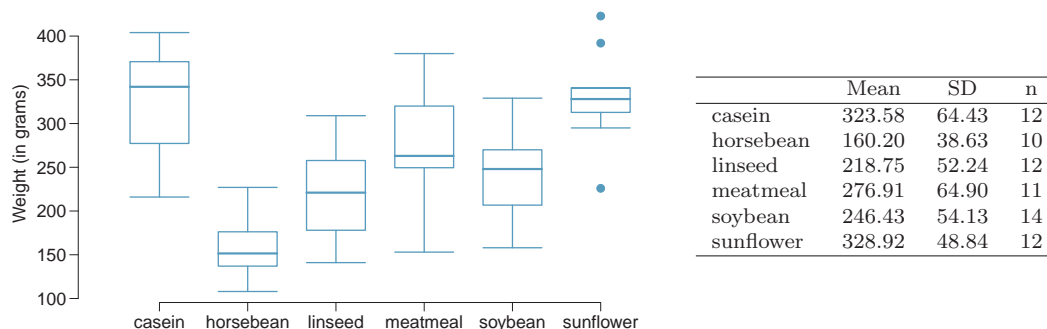
5.41 Fill in the blank. When doing an ANOVA, you observe large differences in means between groups. Within the ANOVA framework, this would most likely be interpreted as evidence strongly favoring the _____ hypothesis.

5.42 Which test? We would like to test if students who are in the social sciences, natural sciences, arts and humanities, and other fields spend the same amount of time studying for this course. What type of test should we use? Explain your reasoning.

5.43 Chicken diet and weight, Part III. In Exercises 5.31 and 5.33 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.



5.44 Teaching descriptive statistics. A study compared five different methods for teaching descriptive statistics. The five methods were traditional lecture and discussion, programmed textbook instruction, programmed text with lectures, computer instruction, and computer instruction with lectures. 45 students were randomly assigned, 9 to each method. After completing the course, students took a 1-hour exam.

- What are the hypotheses for evaluating if the average test scores are different for the different teaching methods?
- What are the degrees of freedom associated with the F -test for evaluating these hypotheses?
- Suppose the p -value for this test is 0.0168. What is the conclusion?

5.45 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.⁴⁶

<i>Caffeinated coffee consumption</i>					
	≤ 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?

⁴⁶M. Lucas et al. "Coffee, caffeine, and risk of depression among women". In: *Archives of internal medicine* 171.17 (2011), p. 1571.

5.46 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.

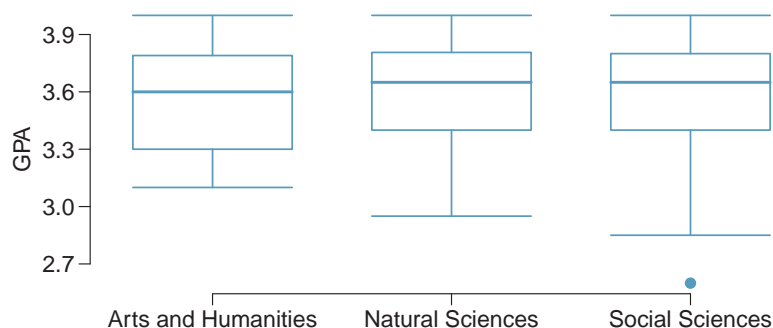
	Sec 1	Sec 2	Sec 3	Sec 4	Sec 5	Sec 6	Sec 7	Sec 8
n_i	33	19	10	29	33	10	32	31
\bar{x}_i	92.94	91.11	91.80	92.45	89.30	88.30	90.12	93.35
s_i	4.21	5.58	3.43	5.92	9.32	7.27	6.93	4.57

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.

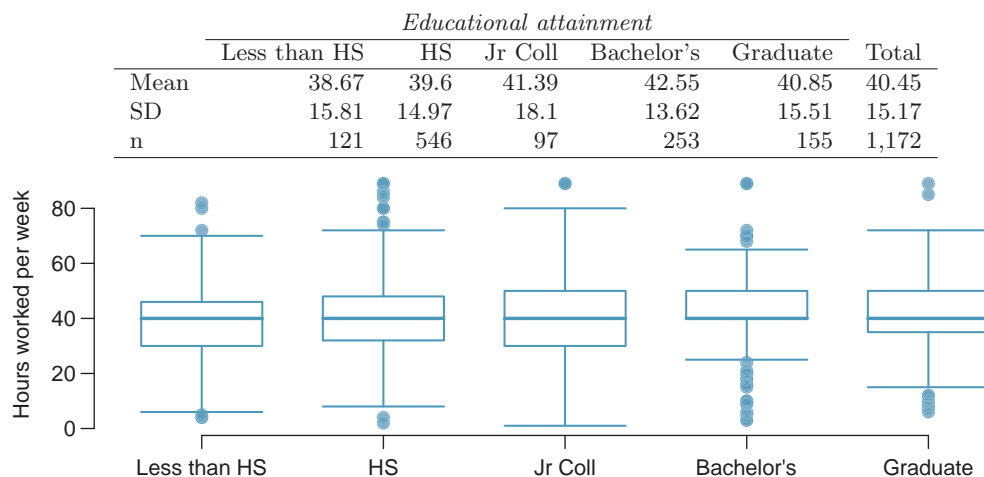
5.47 GPA and major. Undergraduate students taking an introductory statistics course at Duke University conducted a survey about GPA and major. The side-by-side box plots show the distribution of GPA among three groups of majors. Also provided is the ANOVA output.



	Df	Sum Sq	Mean Sq	F value	Pr(>F)
major	2	0.03	0.015	0.185	0.8313
Residuals	195	15.77	0.081		

- 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?

5.48 Work hours and education. The General Social Survey collects data on demographics, education, and work, among many other characteristics of US residents.⁴⁷ Using ANOVA, we can consider educational attainment levels for all 1,172 respondents at once. Below are the distributions of hours worked by educational attainment and relevant summary statistics that will be helpful in carrying out this analysis.



- 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"/>			

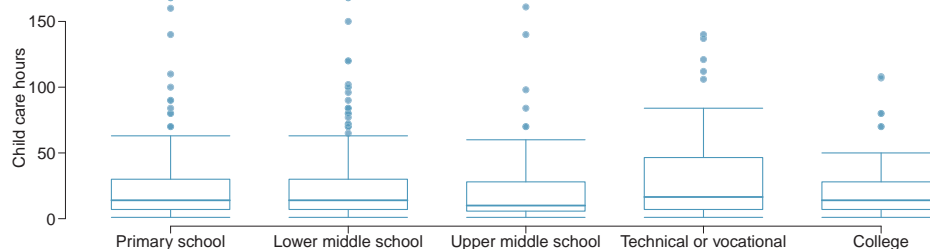
- What is the conclusion of the test?

5.49 True / False: ANOVA, Part I. Determine if the following statements are true or false in ANOVA, and explain your reasoning for statements you identify as false.

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

⁴⁷National Opinion Research Center, General Social Survey, 2010.

5.50 Child care hours. The China Health and Nutrition Survey aims to examine the effects of the health, nutrition, and family planning policies and programs implemented by national and local governments.⁴⁸ It, for example, 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?

5.51 Prison isolation experiment, Part II. Exercise 5.37 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 5.37 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		

$$s_{pooled} = 9.793 \text{ on } df = 39$$

- What are the hypotheses?
- What is the conclusion of the test? Use a 5% significance level.
- 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 answer.

5.52 True / False: ANOVA, Part II. 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 ...

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

⁴⁸UNC Carolina Population Center, China Health and Nutrition Survey, 2006.