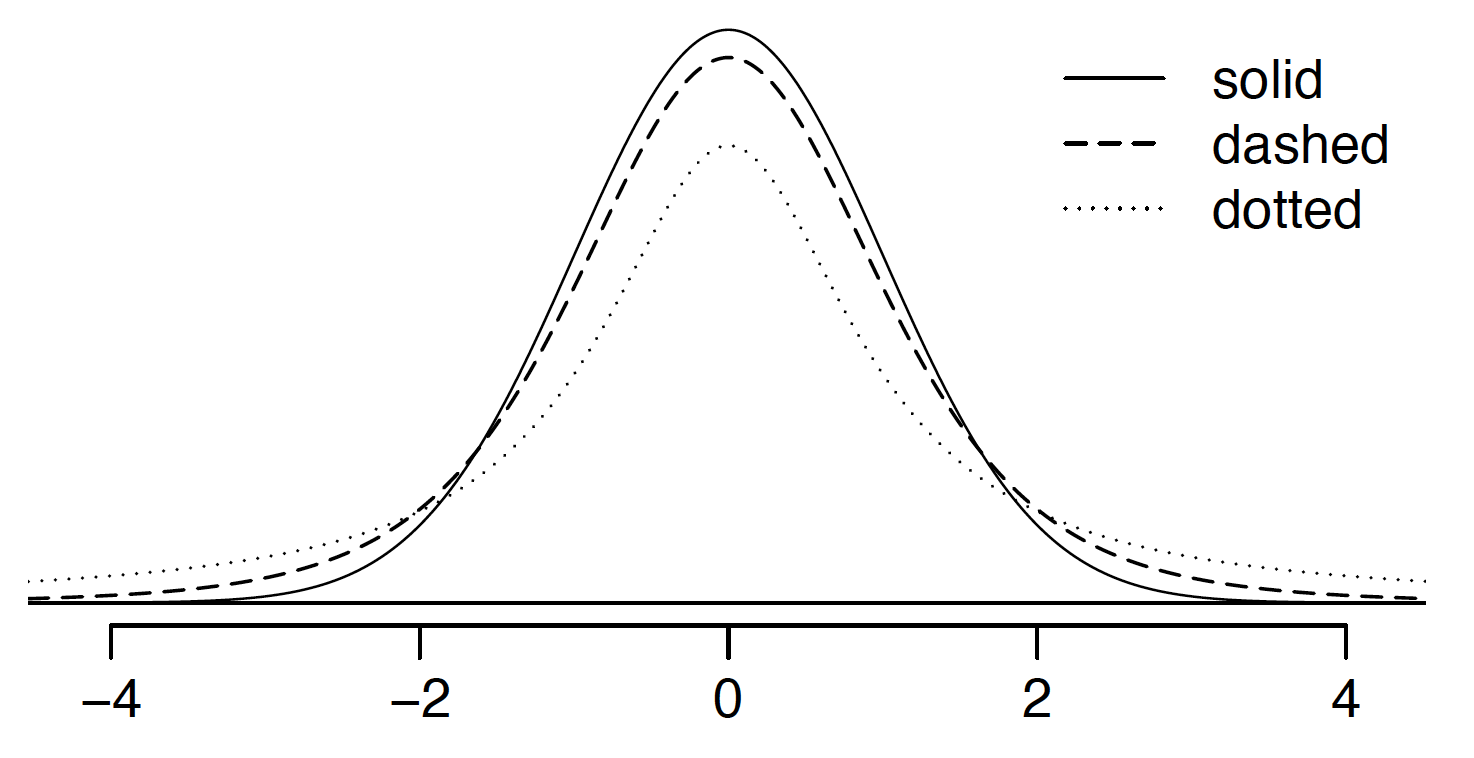
Problem set 5

## Section 7.1 problems

### Problem 1: -distribution

The figure below shows three unimodal and symmetric curves: the standard normal (z) distribution, the -distribution with 5 degrees of freedom, and the -distribution with 1 degree of freedom. Determine which is which, and explain your reasoning.

Note that another way of viewing this question is: which of these distributions puts the most probability in ‘extreme’ values (i.e. values far from 0), and which puts the least? (You can see this in the the figure – one distribution has the thinnest ‘tails’.)



### Problem 2: Find the p-value

An independent random sample is selected from an approximately normal population with an unknown standard deviation. Find the p-value for the given sample size and test statistic.. Also determine if the null hypothesis would be rejected at .

### Problem 3: Working backwards

A 99% confidence interval for a population mean is (13, 25). The population distribution is approximately normal and the population standard deviation is unknown. This confidence interval is based on a simple random sample of 31 observations. Calculate the sample mean, the margin of error, and the sample standard deviation.

### Problem 4: Environmental exposure to radon

Researchers interested in environmental exposure to radon in a small town in northern New Mexico sampled the blood of 47 park rangers subjected to possibly damaging levels of environmental radon levels while working in a primarily rural environment. The blood samples of these rangers had an average radon concentration of 744.3 Bq m and a SD of 82.7 m; a previous study of individuals from a different location in New Mexico, with no history of exposure, found an average blood level concentration of 417.2 m

1. Write down the hypotheses that would be appropriate for testing if the park rangers appear to have been exposed to a different concentration of radon
2. Explicitly state and check all conditions necessary for inference on these data.
3. Regardless of your answers in part (b), test the hypothesis that the park rangers have a higher radon exposure than the group in the previous study. Interpret your results in context.

## Section 7.2 problems

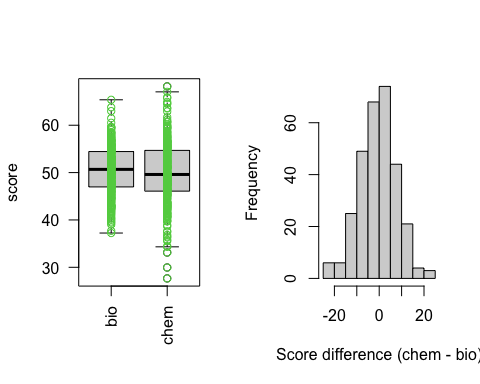
### Problem 5: True / False paired

Determine if the following statements are true or false. If false, explain.

1. Consider two datasets of different sizes. These can be analyzed as paired data if you average the extra observations first.
2. Consider two datasets that are paired with each other. Each observation in one data set has a natural correspondence with exactly one observation from the other data set.
3. Consider two sets of data that are paired with each other. Each observation in one data set is subtracted from the average of the other data set’s observations.
4. In a paired analysis we first take the difference of each pair of observations, and then we do inference on these differences.
5. If we randomly sample a set of individuals and then flip a coin to put each of them into one of two groups, it is natural to pair each individual in one group to someone in the other so long as the sample size is the same.

### Problem 6: Pre-med prowess

Duke University conducted a survey of pre-med students, collecting test data on chemistry, biology, physics, and several other subjects. Here we examine a simple random sample of 300 students from this survey. Side-by-side box plots of chemistry and biology scores as well as a histogram of the differences in scores are shown below.



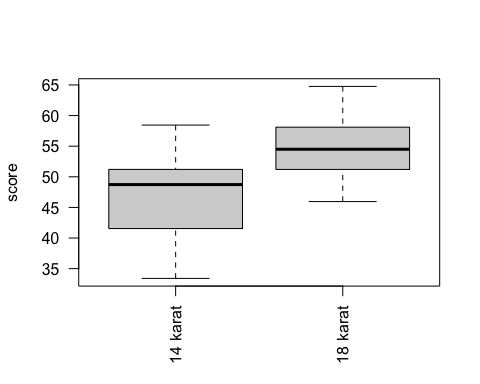
1. Is there a clear difference in the average chemistry and biology scores?
2. Are the chemistry and biology scores of each student independent of each other?
3. Create hypotheses appropriate for the following research question: is there an evident difference in the average scores of students in the chemistry and biology exam?
4. Check the conditions required to complete this test.
5. The average observed difference in scores is =-0.567, and the standard deviation of the differences is 8.259 points. Do these data provide convincing evidence of a difference between the average scores on the two exams?
6. What type of error might we have made? Explain what the error means in the context of the application.
7. Based on the results of this hypothesis test, would you expect a confidence interval for the average difference between the chemistry and biology scores to include 0? Explain your reasoning.

## Section 7.3 problems

### Problem 7: Gold, pt. 1

The unit of a karat, when used with gold, describes purity. 24-karat gold is pure gold, but pure gold is usually too soft to make jewelry and so is mixed with a metal like copper or silver. If, for example, a piece of jewelry is made of metal that is 18 parts gold and 6 parts copper, that is 18-karat gold.

We would expect the price of gold to differ by karat, but is this difference proportional to the purity increase? We explore this question using two random samples of gold, 14 karat and 18 karat, each sample of size 33, and compare the average prices of the gold. In order to be able to compare equivalent units, we first divide the price for each piece by its weight in karats. That is, for a 14 karat piece, we divide the price by 14. For a 18 karat piece, we divide the price by 18. The distributions and some sample statistics are shown below.



|  |  |  |
| --- | --- | --- |
|  | 14 karat | 18 karat |
| Mean | $47.38 | $55.41 |
| SD | $7.32 | $5.21 |
| n | 33 | 33 |

Conduct a hypothesis test to evaluate if there is a difference between the average standardized prices of 14 and 18 karat gold. Make sure to state your hypotheses clearly, check relevant conditions, and interpret your results in context of the data.

### Problem 8: Gold, pt. 2

Using the information provided in Problem 7, construct a 95% confidence interval for the average difference between the standardized prices of 14 and 18 karat gold. You may assume the conditions for inference are met.

### Problem 9: Netflix and eat

A group of researchers are interested in the possible effects of distracting stimuli during eating on the mindfulness experience of eating. They want to know whether people who do an activity while eating remember their food any differently than those who don’t. To test this hypothesis, they monitored food intake for a group of 52 students who were randomized into two equal groups. The treatment group ate dinner while watching Netflix, and the control group ate dinner without any added distractions. The 26 students in the treatment group who ate their dinner while watching Netflix were asked to do a serial-order recall of the dinner food items they ate. The average number of items recalled by the students in this group was 2.9, with a standard deviation of 1.8. The average number of items recalled by the students in the control group (no Netflix) was 5.1, with a standard deviation of 1.8. Do these data provide strong evidence that the average number of food items recalled by the students in the treatment and control groups are different?

### Problem 10: True / False comparing means

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

1. We use a pooled standard error for calculating the standard error of the difference between means when the variability in groups is constant.
2. When comparing means of two samples where and , we can use the normal model for the difference in means since .
3. As the degrees of freedom decreases, the t-distribution approaches normality.

## Section 7.4 problems

### Problem 11: Outpatient surveys

A Duke Outpatient clinic sends follow-up surveys regarding patient care. For example, one survey asks about how the patient felt about their provider’s listening skills, attention to detail, and care for their experience. Another survey asks about whether the patient felt that their issue had been thoroughly addressed. So far, patients respond to an average of just 3 surveys, and the standard deviation of the number of surveys is about 1.4. The public relations team at the clinic wants to try a new format that they think will encourage patients to complete more surveys. They plan to conduct a study where they randomize each patient to either get the new interface or the current interface. How many new patients do they need for each interface to detect an effect size of 0.5 surveys per patient, if the desired power level is 90%?

## Section 7.5 problems

### Problem 12: Communicating disease risk, part ii.

A study compared five different methods for informing the public about the disease risk associated with various activities. The five methods were a verbal delivery, text without pictures, text with pictures, a short video, and an interactive demonstration. 50 participants were randomly assigned, 10 to each method. After completing the risk learning module, participants took an exit exam.

1. What are the hypotheses for evaluating if the average exam scores are different for the different communication methods?
2. What are the degrees of freedom associated with the F-test for evaluating these hypotheses?
3. Suppose the p-value for this test is 0.023. What is the conclusion?

### Problem 13: Teaching team comparison

Suppose a professor at a large research institution has a lecture course with 185 students divided into 7 discussion sections. Each discussion section has its own teaching assistant who leads problem set reviews and covers labs. She is curious if the midterm exam scores in her class differ across discussion section. The summary table below shows the average midterm 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 |
|  | 92.48 | 91.34 | 90.1 | 91.38 | 88.99 | 91.03 | 88.84 |
|  | 5.26 | 5.38 | 7.07 | 4.2 | 4.17 | 5.84 | 3.12 |
|  | 30 | 32 | 23 | 18 | 29 | 25 | 28 |

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 | 6 | 306.68 | 51.11 | 1.95 | 0.0755 |
| Residuals | 178 | 4669.51 | 26.23 |  |  |

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.

### Problem 14: OIS Problem 7.42 (Work hours and education)

See OIS Problem 7.42 for problem statement and numeric values.

1. Write hypotheses for evaluating whether the average number of hours worked varies across the five groups.
2. Check conditions and describe any assumptions you must make to proceed with the test.
3. Below is part of the output associated with this test. Report the values for i. through vii.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
| section | i. | ii. | 501.54 | iii. | 0.0682 |
| Residuals | iv. | 267,382 | v. |  |  |
| Total | vi. | vii. |  |  |  |

1. What is the conclusion of the test?

### Problem 15: True / False ANOVA

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

Suppose we are comparing the means of four groups using ANOVA at a 5% significance level.

1. the appropriate to be used in pairwise comparisons is 0.05 / 4 = 0.0125 since there are four groups.

If the null hypothesis that the means of four groups are all the same is rejected, then …

1. we can then conclude that all the means are different from one another.
2. the pairwise analysis will identify at least one pair of means that are significantly different.

If we fail to reject the null hypothesis that the means of four groups are all the same, then …

1. we can then conclude that all the means are definitely the same.
2. the standardized variability between groups is lower than the standardized variability within groups.