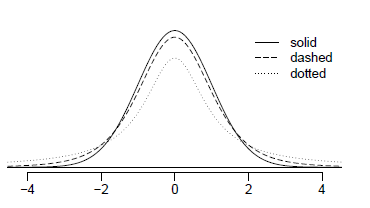
**Chapter 5**(these exercises are presented between pages 257 and 273 of this text): **5.2, 5.12, 5.20, 5.30, 5.36, 5.38, 5.40, 5.44, 5.48, 5.52.** For each exercise, answer all parts with well composed sentences including numbers or graphical images depending on the nature of the exercise. Begin each exercise with the appropriate number, and label the subparts of each exercise with letters following the conventions used in the text. When you have finished the exercises, submit a Word document with your answers.

# 5.2 t-distribution.

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



* Dotted 🡺 1 degrees of freedom
* Dashed 🡺 5 degrees of freedom
* Solid 🡺 Standard normal(z) distribution
* Dotted and dash lines have thicker tails compared to solid.
* Larger the degree of freedom more closely the distribution approximates the standard normal (z) model. Dotted line is far away from normal distribution so associated with degree of freedom = 1. Dashed line is lesser far away from standard distribution so associated with degree of freedom =5. df(dotted) < df(dashed).
* Solid line has very less fat tail so associated with standard normal (z) distribution.

# 5.12 Auto exhaust and lead exposure.

Researchers interested in lead exposure due to car exhaust sampled the blood of 52 police officers subjected to constant inhalation of automobile exhaust fumes while working traffic enforcement in a primarily urban environment. The blood samples of these officers had an average lead concentration of 124.32 µg/l and a SD of 37.74 µg/l; a previous study of individuals from a nearby suburb, with no history of exposure, found an average blood level concentration of 35 µg/l.

1. Write down the hypotheses that would be appropriate for testing if the police officers appear to have been exposed to a higher concentration of lead.

* Ho 🡺 µ = 35
* HA 🡺 µ > 35
* **Null Hypothesis:** Average lead concentration of the police officers is same as that of the individuals from nearby suburbs with no exposure. Ho 🡺 µ = 35
* **Alternative Hypothesis:** Average lead concentration of the police officers is greater than that of the individuals from nearby suburbs with no exposure. HA 🡺 µ > 35

1. Explicitly state and check all conditions necessary for inference on these data.

* It does not mention about random selection of the police officers.
* Researchers are interested in studying auto exhaust and lead exposure in general across the population, choosing police officers is convenience sampling.
* Sample size is 52, but the population size has not been mentioned. For t-distribution n (sample size) < 10% of the population. So there should be at-least 520 police offices having the same exposure to lead as that of the sample ones.
* Normality: If most observations lie around 2.5 standard deviation from the mean then normality condition is satisfied.

1. Test the hypothesis that the downtown police officers have a higher lead exposure than the group in the previous study. Interpret your results in context.

SD = 37.74

n = 52

SE = SD /√ n

SE = 37.74/√52 = 5.234

Tdf = (124.32 – 35)/5.234 = 17.067

p-value 🡺 1- pt(17.067, df=51) = 0 (Close to 0 not exactly 0)

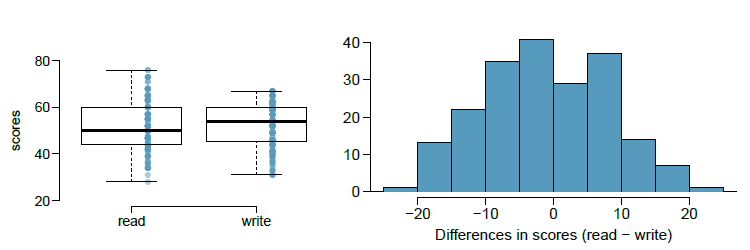
* **So in summary p-value is close to 0**
* Smaller p-value stronger the data to favor HA over HO**,** p-value < 0.05 corresponds to sufficient evidence to reject HO in favor of Ha.
* These enough evidence to prove that average lead concentration of the police officers is greater than that of the individuals from nearby suburbs with no exposure

1. Based on your preceding result, without performing a calculation, would a 99% confidence interval for the average blood concentration level of police officers contain 35 µg/l?

* Since p-value is close to 0, so 99% confidence interval would not include average 35 µg/l

# 5.20. 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.



1. Is there a clear difference in the average reading and writing scores?

* Histogram
  + It is unimodal with peak at near difference = 0, which is clear indicator that there is **no clear difference in average reading and writing score**. Histogram is slightly right skewed. The distribution is fairly normal around 0.
* Box Plot
  + Medians for reading and writing are slightly different, but overall box plot for reading and writing is similar with exception of upper whisker for reading been higher than reading.

1. Are the reading and writing scores of each student independent of each other?

* Reading and writing score are paired data. Scores are in general independent across students, but **reading and writing score of individual student are not independent of each other**.

1. 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?

* Null Hypothesis: **Difference** in the average scores of students in the **reading** and **writing** exam is **equal to zero**.
  + HO: µread - µwrite = 0
* Alternative Hypothesis: **Difference** in the average scores of students in the **reading** and **writing** exam is **not equal to zero**
  + HA: µread - µwrite != 0

1. Check the conditions required to complete this test.

* Random sample data of 200 students was collected.
* Reading and writing score for each student are not independent as it’s a paired data, but selection of student scores is independent.
* Data is close to normal distribution (Bell curve) evident from histogram and box plot. Box plot does indicate some skews

1. The average observed difference in scores is xread-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?

SEdiff = 8.887/√200 = 0.628

T = (xdiff – 0)/SE = (-0.545 – 0)/0.628 = -0.868

pt(-0.868, df=199) = 0.1932199

Since HA : µread - µwrite != 0 we consider both sides of the curve so

2 \* 0.1932199 = 0.386

Since the p-value is 0.386 which is larger than 0.05 **we fail to reject Ho,** we **cannot reject null hypothesis**

There is no convincing evidence of difference of average scores in reading and writing exams for students

1. What type of error might we have made? Explain what the error means in the context of the application.



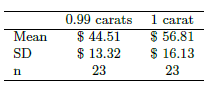
* We may possibly make a Type 2 error 🡺 Rejecting Alternative hypothesis even though alternative is true. We may have incorrectly concluded that there is no difference in the average score difference between reading and writing exams for students.
* In other words Type 2 error 🡺 Failing to reject NULL hypothesis (Ho) even though alternative hypothesis (Ha) is true.

1. 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.

* Based on the hypothesis test we have failed to reject NULL hypothesis as there was not much evidence of difference in average scores of reading and writing exam scores for students. From this it is evident that difference in score = 0. So confidence interval should include 0.

# 5.30 Diamonds, Part II.

In Exercise 5.28, 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.



The sample difference of two means, x1 - x2, can be modeled using the t-distribution and the standard error



SE = √((13.32\*13.32)/23) + (16.13\*16.13)/23)

SE = 4.362

Sample difference of 2 means = 56.81 - 44.51 = 12.3

point estimate +- tdf \* SE

df (degree of freedom) = n -1 = 23-1 = 22

tdf = t22 = 2.07 (For 95% confidence interval based on the t distribution probability table)

12.3 + (2.07 \* 4.362) = 21.33

12.3 - (2.07 \* 4.362) = 3.27

Therefore the confidence interval is **(3.27, 21.33)**.

For 95% confidence interval the average difference between the standardized prices of 0.99 and 1 carat diamonds is between **3.27** and **21.33**

# 5.36 Gaming and distracted eating, Part II.

The researchers from Exercise 5.35 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?

Let’s define Null and Alternative hypothesis

* Null Hypothesis: H0
  + There is no difference in the average number of items recalled by the patients in treatment and control group
  + In statistical notation µc - µt = 0, where µc represents control group and µt represents treatment group
* Alternative Hypothesis: HA
  + There is some difference in the average number of items recalled by the patients in treatment and control group
  + In statistical notation µc - µt != 0

|  |  |  |
| --- | --- | --- |
|  | Treatment Group | Control Group |
| **Mean** | 4.9 | 6.1 |
| **SD** | 1.8 | 1.8 |
| **n** | 22 | 22 |
|  |  |  |

xc – xt = 6.1 -4.9 = 1.2



In this case SD and sample size are same, so

SE = √((4.9\*4.9)/22) + (6.1\*6.1)/22) = 0.54272042

tdf = (Point Estimate – Null Value) / SE

t21 = (1.2 – 0)/ 0.54272042= 2.211083194

Based on the t21 distribution probability table the p-value for 2 sided tail value is between 0.050 0.020, which is lower than the significance value of 0.05

As p-value is less than the significance level for 0.05 we reject the Null hypothesis. There is concrete evidence of difference in the average number of items recalled by the patients in treatment and control group.

# **5.38.** True / False: comparing means.

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

1. When comparing means of two samples where n1 = 20 and n2 = 40, we can use the normal model for the difference in means since n2 >= 30.

* **False**. The sample difference of two means, x1 - x2, can be modeled using the t-distribution and the standard error.



* + When each sample mean can itself be modeled using a t-distribution and the samples are independent. To calculate the degrees of freedom, use the smaller of n1 - 1 and n2 - 1.
  + In this case n1 = 20 and n2 = 40, so have to choose n1 = 20 as it smaller than n2. As n value considered is 20 which does not satisfy the condition n>=30 for normal model. Hence we cannot use the normal model.

1. As the degrees of freedom increases, the t-distribution approaches normality.

* **True.** The degrees of freedom describe the shape of the t-distribution. The larger the degrees of freedom, the more closely the distribution approximates the normal model. When the degrees of freedom is about 30 or more, the t-distribution is nearly indistinguishable from the normal distribution.

1. 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.

* **False.** 
  + We use pooled standard error when two population standard deviation are equal. In this case sample size of groups are equal to each other. If sample pool size is equal you cannot pool SE.
  + A pooled standard deviation is only appropriate when background research indicates the population standard deviations are nearly equal.

# **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%?

SD = 2.2

Z-score that would give us a lower tail of 80%: 0.84

Additionally, the rejection region always extends 1.96 \* SE from the center of the null distribution for α = 0:05. This allows us to calculate the target distance between the center of the null and alternative distributions in terms of the standard error:

0.84 \* SE + 1.96 \* SE = 2.8 \* SE

The distance between the null and alternative distributions' centers to equal the minimum effect size of interest, 0.5 surveys per enrollee, which allows us to set up an equation between this difference and the standard error:

2.2 = 2.8 \* √ ((2.2)\*(2.2)/n) + ((2.2)\*(2.2)/n)

n = ((2.8\*2.8)/ (0.5\*0.5))\*((2.2\*2.2) + (2.2\*2.2)) = 303.5648

**304** new enrollees they need for each interface to detect an effect size of 0.5 surveys per enrollee, if the desired power level is 80%

# **5.44** Teaching descriptive statistics.

A study compared five different methods for teaching descriptive statistics. The five methods were traditional lecture and discussion, programmed text-book 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.

(a) What are the hypotheses for evaluating if the average test scores are different for the different teaching methods?

* H0: Null Hypothesis: The average/mean test scores for exam across all 5 different teaching methods is same. Notationally, we write µ1 = µ2 = µ3 = µ4 = µ5
* HA: Alternative Hypothesis: The average/mean score for exam varies by the teaching method.

(b) What are the degrees of freedom associated with the F-test for evaluating these hypotheses?

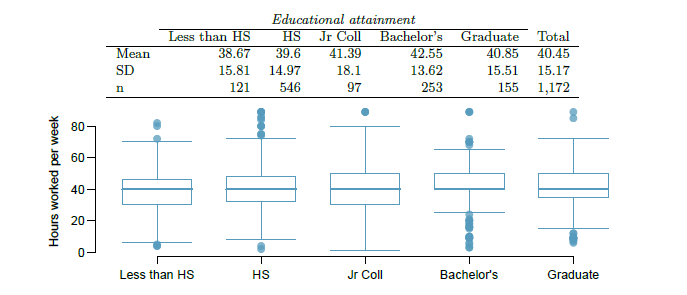
* F = MSG/MSE
* For MSG degree of freedom dfg = k -1 where k = number of groups, so
  + dfg = 5 -1 = 4
* For MSE degree of freedom dfE = n - k where k = number of groups, so
  + dfg = 45 - 5 = 40

(c) Suppose the p-value for this test is 0.0168. What is the conclusion?

* A small p-value 0.0168 is less than 0.05, indicating there is evidence to reject the null hypothesis at a significance level of 0.05.
* Data does provide strong evidence that the average/mean score for exam varies by the teaching method. At least two group means are significantly different from each other.

# 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.



(a) Write hypotheses for evaluating whether the average number of hours worked varies across the five groups.

* H0: Null Hypothesis
  + Difference in mean of worked hours across 5 groups = 0. In other words, the average worked hours is equal among all 5 groups.
* HA: Alternative Hypothesis:
  + Difference in mean of worked hours across 5 groups != 0. In other words, the average worked hours is not equal among all 5 groups.

(b) Check conditions and describe any assumptions you must make to proceed with the test.

* Assumptions
  + Sample size is 1172, so assumption is all 1172 respondents are independent of each other. The observations are independent within and across groups.
  + Sample is less than 10% of the population.
  + The data within each group are nearly normal.
  + The variability across the groups is about equal. The variance within each group must be approximately equal. This assumption can be verified by side by side box plots, variability is similar but not identical, and standard deviation varies from one group to another.

(c) Below is part of the output associated with this test. Fill in the empty cells.

dfg = k -1

dfd = n – k

k = 5

n = 121 + 546 + 97 +253 + 155 = 1172

dfg = 5 -1 = 4

dfd = 1172 – 5 = 1167

MSG = 501.54

SSE = 267382



SSG = dfg \* MSG = 4 \* 501.54 = 2006.16

SSE = SST – SSG

267382 = SST - 2006.16

SST = 267382 - 2006.16 =



MSE = 267382/1167 = 229.1191088

F = MSG/MSE = 501.54/229.1191088

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Df | Sum Sq | Mean Sq | F Value | Pr(>F) |
| degree | **4** | **2006.16** | 501.54 | **2.18899** | 0.0682 |
| residual | **1167** | 267382 | **229.119** |  |  |
| Total | **1171** | **269388.16** |  |  |  |

(d) What is the conclusion of the test?

As p-value = 0.0682 is greater than 0.05 we cannot reject the NULL hypothesis. The average worked hours is equal among all 5 groups.

# 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...

(a) We can then conclude that all the means are different from one another.

* **False** 🡺 Using ANOVA, alternative hypothesis dictates that at least one mean is different. It does not guarantee that all means are different from one another.

(b) The standardized variability between groups is higher than the standardized variability within groups.

* **True** 🡺The larger the observed variability in the sample means (MSG) relative to the within- group observations (MSE), the larger F will be and the stronger the evidence against the null hypothesis.

(c) The pairwise analysis will identify at least one pair of means that are significantly different.

* **True**.

(d) The appropriate α to be used in pairwise comparisons is 0.05 / 4 = 0.0125 since there are four groups.

* False 🡺 There are 4 groups, so there are 6 possible pairwise comparisons

The Bonferroni correction suggests that a more stringent significance level is more appropriate for these tests:



where K is the number of comparisons being considered (formally or informally).

If there are k groups, then usually all possible pairs are compared and K = k(k-1)/2

k = 4

K = (4\*(4 – 3))/2 = 6

α\* = 0.05/6 = 0.008333333