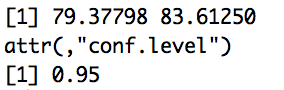
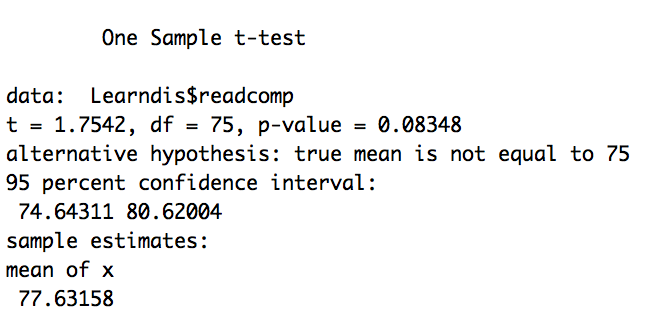
**CHAPTER 11 SOLUTIONS**

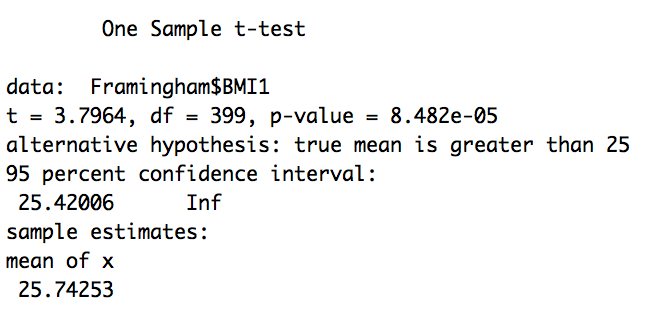
1. In this case, we perform a *t*-test instead of a *z*-test because ** is not assumed to be known and needs to be estimated from the sample.
2. The one sample *t*-test is robust to violations of the normality assumption in this case because the sample size, *N* = 105, is large.
3. (79.38, 83.61) obtained using the R command **t.test(Learndis$iq)$conf.int**

****

1. H0: **100 and H1: ** ≠ 100
2. The null hypothesis can be rejected because, according to the 95 percent CI, 100 is not included in the CI, and as such, is not a plausible value for the mean IQ of all children diagnosed with learning disabilities in this city.
3. Because all of the values in the confidence interval are below 100, we may further conclude that children diagnosed with learning disabilities in this city have average IQ scores that are statistically significantly lower than 100.
4. The one sample *t*-test is robust to violations of the normality assumption in this case because the sample size, *N* = 76, is large.
5. H0: **75 and H1: ** ≠ 75
6. *p* = .083. Obtained using the R command **t.test(Learndis$readcomp, mu = 75)**



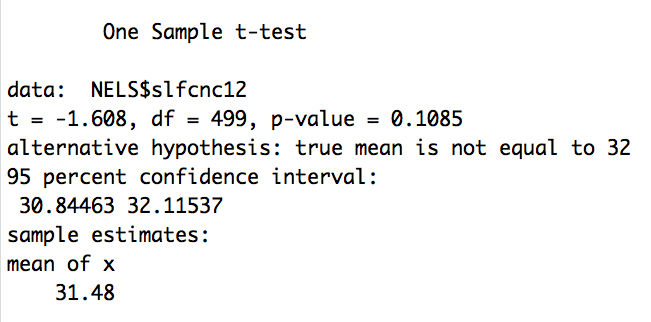
1. We cannot reject the null hypothesis in favor of the alternative.
2. The mean reading comprehension score for children attending public elementary school in the city who are diagnosed with learning disabilities (*M* = 77.63, *SD* = 13.08) is not statistically significantly different from 75, *t*(75) = 1.754, *p* = .08.
3. (74.64, 80.62)
4. The null hypothesis cannot be rejected because, according to the 95 percent CI, 75 is included in the CI, and as such, is a plausible value for the mean reading comprehension score of all children diagnosed with learning disabilities in this city.
5. More powerful since the standard error of the mean would have been smaller, producing a larger obtained *t-*value and a narrower CI.
   1. The output is obtained using the R command **t.test(Framingham$BMI1, mu = 25)**



1. The one sample *t*-test is robust to violations of the normality assumption in this case, because the sample size, *N* = 400, is large.
2. H0: **25 and H1: ** > 25
3. Because the test is one-tailed, we take the *p*-value associated with Ha: ** > 25. The *p*-value, *p* < .0001, suggests that the mean body mass index of population of non-institutionalized adults as measured during the first examination is, in fact, greater than 25.
4. 

The mean body mass index is .19 standard deviations higher than 25, a small effect, according to Cohen’s rule-of-thumb guidelines.

1. Confidence intervals are not appropriate for conducting tests of directional hypotheses.
2. H0: **120 and H1: ** < 120
3. Because the sample mean (*M* = 130.36) is larger than 120, we cannot reject the null in favor of the alternative hypothesis. The direction of difference alone is sufficient to determine that the mean systolic blood pressure of the population of non-institutionalized adults as measured during the first examination is not less than 120 mmHg.
4. The one sample *t*-test is robust to violations of the normality assumption in this case, because the sample size, *N* = 500, is large.
5. (4.58, 4.81)
6. The average family size for this population is statistically significantly higher than 4.5.
7. *d* = (4.69 – 4.5)/1.319 = .15. The average family size is .15 standard deviations higher than 4.5, a negligible effect, according to Cohen’s rule-of-thumb guidelines. This is an example of a situation in which we have statistical significance but not practical significance.
   1. According to the 95 percent confidence interval, the average twelfth grade self-concept score of the population of college-bound students who have always been at grade level is between 30.84 and 32.12. Because 32 is contained in the confidence interval, 32 is a plausible self-concept average for this population. The population mean is therefore not statistically significantly different from 32.
   2. According to the *p*-value derived from the results of a one-sample *t*-test, (*t*(499) = -1.61, *p* = .11), the average twelfth grade self-concept score of the population of college-bound students who have always been at grade level is not statistically significantly different from 32.



* 1. According to the *p*-value derived from the results of a one-sample *t*-test, (*t*(499) = 1.03, *p* = .15), the population of college-bound students who have always been at grade level on average do not take a statistically significantly greater number of units of mathematics in high school than 3.6.

a) To restrict the sample to students from the Northeast, we specify NELS$unitengl[NELS$region == "Northeast"] in R.

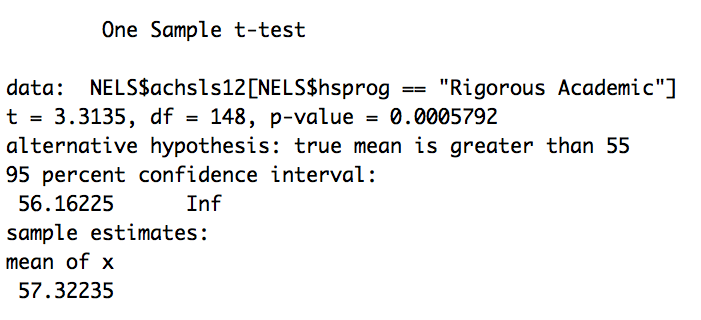
b) The one sample *t*-test is robust to violations of the normality assumption in this case because the sample size, *N* = 106, is large.

c) (3.957, 4.153)

d) No, the estimate is greater than 3 years.

e) *d* = 2.06, indicating that the number of years of high school English taken by this population of college-bound students from the Northeast is approximately 2.06 standard deviations greater than 3 years. According to Cohen’s rule-of-thumb guidelines, this is a very large effect.

* 1. The R command to generate the output is **t.test(NELS$achsls12[NELS$hsprog == "Rigorous Academic"], mu = 55, alternative = "greater")**



a) The one sample *t*-test is robust to violations of the normality assumption in this case because the sample size, *N* = 149, is large.

b) H0: ** 55 and H1: ** 55.

c) Because the sample mean is 57.32, the direction of the alternative hypothesis is supported.

d) *p* = .0006. This value is derived from a one sample *t*-test, *t*(148) = 3.31, *p* = .0006.

e) Yes, because the *p*-value is less than .05.

f) The mean social studies achievement score for this population (*M* = 57.32, *SD* = 8.56) is statistically significantly larger than 55, *t*(148) = 3.31, *p* = .0005.

g) The mean social studies achievement score for this population is .27 standard deviations larger than 55, a small to moderate effect, according to Cohen’s rule-of-thumb guidelines.

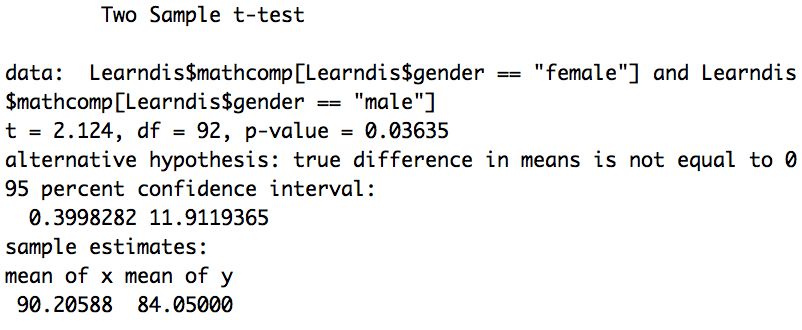
h) The standard deviation value of 8.56 is an estimate of the population standard deviation, ** which measures the spread in the population of the social studies achievement scores themselves. The standard error of the mean value of .70 is an estimate of**, which measures the spread of the social studies achievement means based on samples of size 149 randomly selected from this population.

i) Larger. The denominator of the *t*-statistic would have been larger, the *t*-value would have been smaller, and the area to its right would have been larger.

a) The one sample *t*-test is not robust to violations of the normality assumption in this case because the sample size, *N* = 26, is small. However, the skewness ratio is -1.53, which is less than 2 in magnitude, suggesting that the data are not skewed and the normality assumption is tenable.

b) (27.56, 34.13)

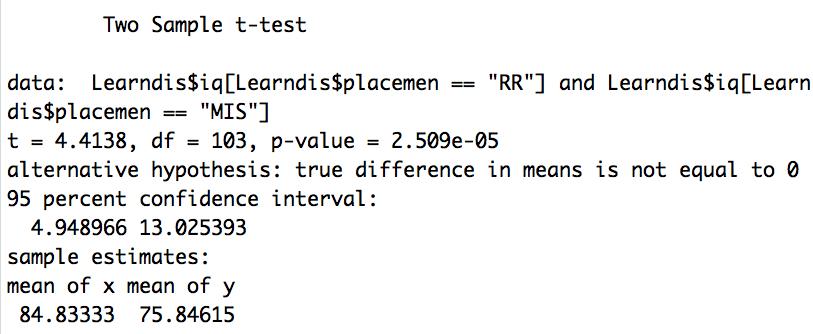
* 1. The output is generated using the R command **t.test(Learndis$mathcomp[Learndis$gender == "female"], Learndis$mathcomp[Learndis$gender == "male"], var.equal = T)**

****

a) The independent samples *t*-test is robust to violations of the normality assumption in this case because the sample size is greater than 30 for both males and females.

1. Equal variances because Levene’s test indicates the tenability of the equality of homogeneity of variances assumption in this case, *F*(1, 92) = .57 *p* = .45.
2. H0: males = females and H1: males ≠ females
3. The results of the independent samples *t*-test indicate that for public elementary students with learning disabilities in this city, females, on average, (*M* = 90.21, *SD* = 14.42) have statistically significantly higher math comprehension scores than males (*M* = 84.05, *SD* = 12.96), *t*(92) = -2.12, *p* = .04.
4. For this population of learning disabled students, on average, females score approximately .46 standard deviations higher than males, which represents a moderate effect according to Cohen’s rule-of-thumb guidelines.
   1. The output is generated using the R command

**t.test(Learndis$iq[Learndis$placemen == "RR"], Learndis$iq[Learndis$placemen == "MIS"], var.equal = T)**

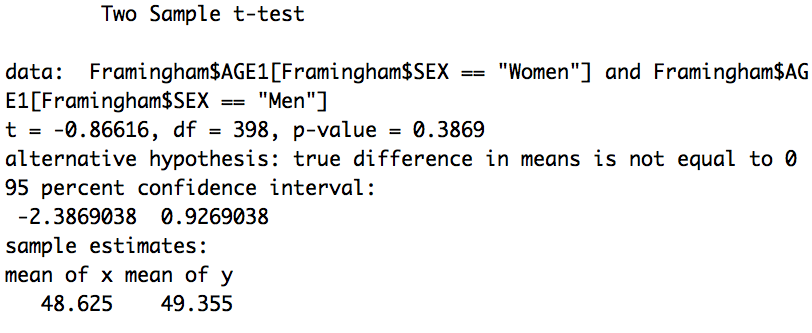


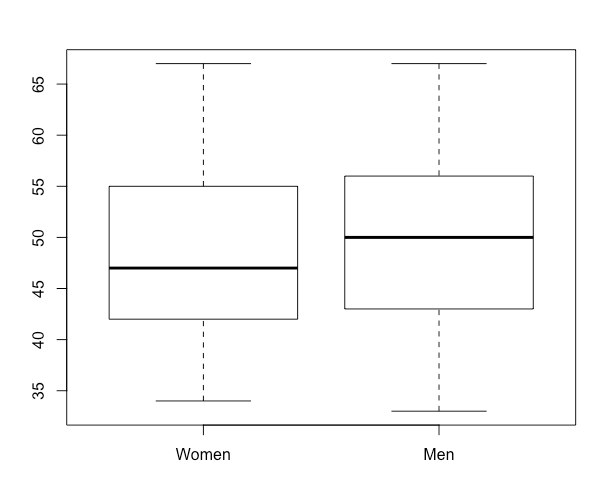
a) The independent samples *t*-test is robust to violations of the normality assumption in this case because the sample size is greater than 30 for both those in a self-contained classroom and those in the resource room.

1. Equal variances because Levene’s test indicates the tenability of the equality of homogeneity of variances assumption in this case, *F*(1, 103) = 1.30, *p* = .26.
2. The 95 percent confidence interval of the difference is (4.95, 13.03), which does not contain zero and indicates that the difference is statistically significant. The average intellectual ability score of all public elementary school students diagnosed with learning disabilities in the city in the resource room (*M* = 84.83, *SD* = 10.37) was between 4.95 and 13.03 points higher than that of the students in the self-contained classroom (*M* = 75.85, *SD* = 9.57).
3. Public elementary students diagnosed with learning disabilities in the city with a resource room placement score approximately .89 standard deviations higher, on average, than students in a self-contained classroom from the same population. This represents a large effect according to Cohen’s rule of thumb guidelines.
   1. The output is generated using the R commands

**t.test(Framingham$AGE1[Framingham$SEX == "Women"], Framingham$AGE1[Framingham$SEX == "Men"], var.equal = T)**

**boxplot(Framingham$AGE1[Framingham$SEX == "Women"], Framingham$AGE1[Framingham$SEX == "Men"], names = c("Women", "Men"))**

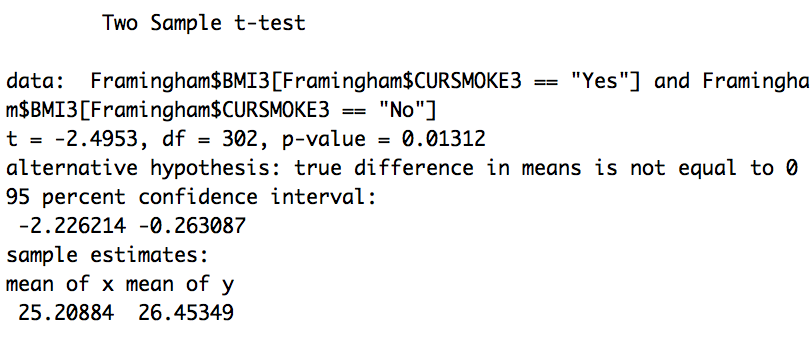
****

****

From the boxplot, it appears that the normality assumption is tenable because the distributions are fairly symmetric. The equality or homogeneity of variance assumption appears tenable as well because the interquartile ranges are similar. Finally, the median for men is larger than it is for women, so given the reasonably symmetric nature of the distributions, a likely result of the independent samples *t*-test is that the mean age of men is statistically significantly higher than it is of women.

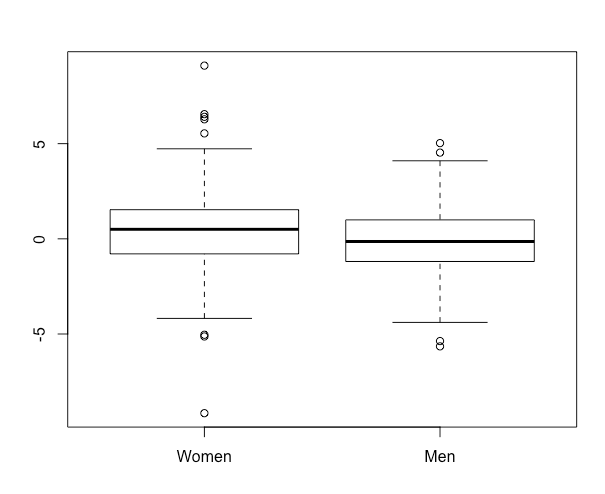
1. The independent samples *t*-test is robust to violations of the normality assumption in this case because the sample size is greater than 30 for both males and females.
2. Equal variances because Levene’s test indicates the tenability of the equality of homogeneity of variances assumption in this case, *F*(1, 398) = 1.21 *p* = .27.
3. According to the *p*-value, the mean age of females at initial examination (*M* = 48.63, *SD* = 8.18) is not statistically significantly different from that of males (*M* = 49.36, *SD* = 8.67), *t*(398) = -0.87, *p* = .39.
4. According to the 95 percent CI, on average, the difference in age between males and females is (-2.39, 0.93), which includes zero, as expected from the answer to part (d).
5. Given a non-directional alternative hypothesis, the independent groups *t*-test conducted at the ** level gives results with respect to statistical significance that are consistent with the (1- ** ) percent CI.
   1. The output is generated using the R command

**t.test(Framingham$BMI3[Framingham$CURSMOKE3 == "Yes"], Framingham$BMI3[Framingham$CURSMOKE3 == "No"], var.equal = T)**



1. According to the results of the independent samples *t*-test (*t*(302) = -2.50), the *p*-value of .007 indicates that on average, the body mass index of the smokers at the third examination (*M* = 25.21, *SD* = 4.53) is statistically significantly lower than that of non-smokers (*M* = 26.45, *SD* = 3.99) for all non-institutionalized adults.
2. Less powerful. In general, given that the direction for the sample is the same as the one that has been hypothesized, one-tailed tests are more powerful than corresponding two-tailed tests because the area of the tail for rejecting the null hypothesis is larger in a one-tailed test than in a two-tailed test.

1. Use **Framingham$BMIDIFF = Framingham$BMI3 - Framingham$BMI1** to create the new variable, and **Framingham$BMIDIFF[Framingham$ID == 1]** to obtain the score for the first person in the dataset. The score on BMIDIFF of the first person in the dataset is -0.64, indicating that the body mass index score for that person decreased by 0.64 BMI units from examination 1 (1956) to examination 3 (1968).
2. According to the boxplot, both distributions appear to be reasonably symmetric, although they each have several outliers. The female distribution appears to be more heterogeneous than males when one considers the outliers. According to the medians, the men appear to show a decrease in BMI while the women appear to show an increase.



1. The independent samples *t*-test is robust to violations of the normality assumption in this case because the sample size is greater than 30 for both males and females.
2. Equal variances because Levene’s test indicates the tenability of the equality of homogeneity of variances assumption in this case, *F*(1, 302) = 1.96 *p* = .16.
3. According to the *p*-value derived from an independent groups *t*-test on means, on average, the change in body mass index for women (*M* = .50, *SD* = 2.34) is statistically significantly different from that for men (*M* = -.13, *SD* = 1.91), *t*(302) = -2.57, *p* = .01. While the body mass index for women increased from 1956 to 1968, it decreased for men.
4. According to the value of Cohen’s *d*, the mean change in body mass index for women is .29 standard deviations higher than that for men. This represents a small to moderate effect.
5. Because there are only 27 participants in the Framingham dataset who used antihypertensive medication, the tenability of the normality assumption needs to be evaluated. In this case, it is not tenable because the distribution of systolic blood pressure scores at time 3 is severely positively skewed for those who were not taking blood pressure medication (skewness ratio = 4.64).
6. We use the following R command to create the log-transformed variable  
   **Framingham$SYSBPlg = log(Framingham$SYSBP3)**

We then use the following R commands to calculate the skewness and skewness ratios. In this series, the values are calculated for SYSBP3lg.

**skew(Framingham$SYSBPlg[Framingham$BPMEDS3 == "currently used"])**

**skew.ratio(Framingham$SYSBPlg[Framingham$BPMEDS3 == "currently used"])**

**skew(Framingham$SYSBPlg[Framingham$BPMEDS3 == "Not currently used"])**

**skew.ratio(Framingham$SYSBPlg[Framingham$BPMEDS3 == "Not currently used"])**

The log transformation is effective as it brings the skewness ratio into acceptable limits.

1. The test is performed using the R command  
   **t.test(Framingham$SYSBPlg[Framingham$BPMEDS3 == "currently used"], Framingham$SYSBPlg[Framingham$BPMEDS3 == "Not currently used"], var.equal = T)**

According to the results of the independent samples *t*-test, the mean of the logarithm of systolic blood pressure at time 3 of those currently taking anti-hypertensive medication (*M* = 2.20, *SD* = .05) is statistically significantly different from that for those not currently taking the medication (*M* = 2.13, *SD* = .07), *t*(256) = 5.20, *p* < .0005.

1. No, because we do not know what their blood pressure readings were at time 1. Anti-hypertensive medication is taken to reduce blood pressure. It may be that it successfully reduced the blood pressure of the people taking it, on average, even though it was still higher for that group at time 3 than it was for the group not taking it.
   1. An independent samples *t*-test was conducted to determine whether, among the population of non-institutionalized adults, there are differences in total serum cholesterol by sex. Because the sample size for both men and women is greater than 30, the test results are robust to violations of the normality assumption. Furthermore, the result of Levene’s test indicates the tenability of the equality of homogeneity of variances assumption in this case, *F*(1, 284) = .50 *p* = .48.

From the *p*-value derived from the independent groups *t*-test, we may note that the mean serum cholesterol for women (*M* = 245.57, *SD* = 48.31) is statistically significantly higher than that for men (*M* = 226.70, *SD* = 42.31), *t*(284) = 3.51, *p* = .001. Alternatively, using the confidence interval, (8.30, 29.45), which does not contain 0, we see that the mean serum cholesterol of women (*M* = 245.57, *SD* = 48.31) is statistically significantly higher than that for men (*M* = 226.70, *SD* = 42.31). According to the value of Cohen’s *d*, the mean serum cholesterol of women is approximately .42 standard deviations higher than that for men, suggesting, according to Cohen’s guidelines, a weak to moderate effect.

a) An independent samples *t*-test is more appropriate than a one-sample *t*-test because the question relates to two population means: the twelfth grade math achievement of those who owned a computer and the twelfth grade math achievement of those who did not own a computer. An independent samples *t*-test is more appropriate than a paired samples *t*-test because each twelfth grade math achievement score of a student who owned a computer is not paired with one from someone who did not own a computer. Another way to see that an independent samples *t*-test is more appropriate than a paired samples *t*-test is that the variable computer is dichotomous and the variable achmat12 is interval. A paired samples *t*-test involves two variables that are at least interval.

1. Because there are 263 students whose parents did not own a computer when they were in eighth grade and 237 students whose parents did own a computer when they were in eighth grade and both of these sample sizes are greater than 30, a violation of the normality assumption will not compromise the validity of the hypothesis test results in this case.
2. Levene’s test indicates the tenability of the equality of homogeneity of variances assumption in this case, *F*(1, 498) = 2.21 *p* = .14.
3. H0: computer owned = no computer and H1: computer owned ≠ no computer
4. *p*  .0005.
5. Among the population of college-bound students who have always been at grade level, students whose families owned a computer in eighth grade (*M* = 58.36, *SD* = 7.93) scored statistically significantly higher in twelfth grade math achievement, on average, than those whose families did not own a computer (*M* = 55.60, *SD* = 7.62), *t*(498) = 3.97, *p* < .0005.
6. (-4.13, -1.40) or (1.40, 4.13)
7. Because 0 is not contained in the confidence interval and because of the values of the sample means, among the population of college-bound students who have always been at grade level students whose families owned a computer in eighth grade (*M* = 58.36, *SD* = 7.93) scored statistically significantly higher in twelfth grade math achievement, on average, than those whose families did not own a computer (*M* = 55.60, *SD* = 7.62).
8. Among the population of college-bound students who have always been at grade level students whose families owned a computer in eighth grade scored approximately .36 standard deviations higher in twelfth grade math achievement, on average, than those whose families did not own a computer, a small to moderate effect.
9. The 99 percent CI is longer. While the 95 percent CI is (-4.13, -1.40) or (1.40, 4.13), the 99 percent CI is (-4.56, -.97) or (.97, 4.56).

a) A two-tailed, independent samples *t*-test was performed to determine whether the average twelfth grade self-concept score is different for smokers and non-smokers. Prior to the analysis itself, an analysis of the underlying assumptions was conducted. Because there are 429 students who never smoked and 71 students who did and both of these sample sizes are greater than 30, a violation of the normality assumption would likely not compromise the validity of the results of the hypothesis test. Because *p* = .27 is greater than .05 for Levene’s homogeneity of variances test, no statistically significant difference was detected in the population variances, so the homogeneity of variances assumption may be considered to be met for these data. According to the results of the independent samples *t*-test, among the population of college-bound students who have always been at grade level, on average, non-smokers (*M* = 31.96, *SD* = 7.11) score statistically significantly differently from smokers (*M* = 28.58, *SD* = 7.31) in twelfth grade self-concept *t*(498) = 3.70, *p* < .0005. Non-smokers score approximately .47 standard deviations higher in twelfth grade self-concept, on average, than smokers score, a moderate effect according to Cohen’s rule-of-thumb guidelines.

b) A two-tailed, independent samples *t*-test was performed to determine whether students who take advanced math in eighth grade have different expectations for future income than students who do not. Prior to the analysis itself, an analysis of the underlying assumptions was conducted. Because there are 241 students who did not take advanced math in eighth grade and 209 who did and both of these sample sizes are greater than 30, a violation of the normality assumption would likely not compromise the validity of the results of the hypothesis test. Because *p* = .44 is greater than .05 for Levene’s homogeneity of variances test, no statistically significant difference was detected in the population variances, so the homogeneity of variances assumption may be considered to be met for these data. According to the results of the independent samples *t*-test, among the population of college-bound students who have always been at grade level, students who take advanced math in eighth grade (*M* = $52,191.39, *SD* = $43,307.17) do not have statistically significant differences in future income expectations than students who do not (*M* = $51,696.27, *SD* = $69,560.29), *t*(448) = 0.09, *p* = .93.

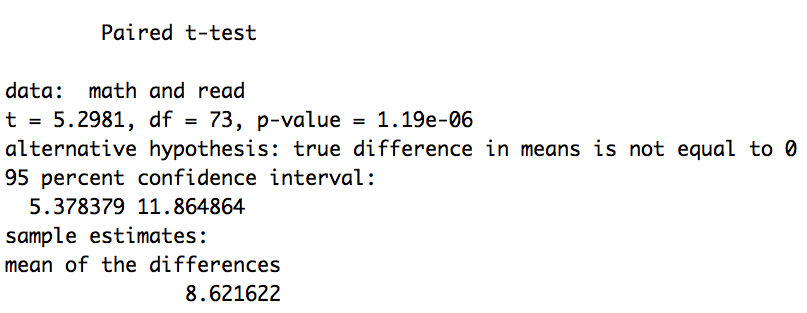
c) A one-tailed, independent samples *t*-test was performed to determine whether students who take advanced math in eighth grade have better eighth grade self-concepts than students who do not. Prior to the analysis itself, an analysis of the underlying assumptions was conducted. Because there are 265 students who did not take advanced math in eighth grade and 226 who did, and both of these sample sizes are greater than 30, a violation of the normality assumption would likely not compromise the validity of the results of the hypothesis test. Because *p* = .78 is greater than .05 for Levene’s homogeneity of variances test, no statistically significant difference was detected in the population variances, so the homogeneity of variances assumption may be considered to be met for these data. According to the results of the independent samples *t*-test, among the population of college-bound students who have always been at grade level, students who take advanced math in eighth grade (*M* = 21.61, *SD* = 5.91) have statistically significantly greater self-concept in eighth grade than students who do not (*M* = 20.54, *SD* = 5.99), *t*(489) = 1.97, *p* = .02. According to an effect size analysis, however, the magnitude of the effect is small according to Cohen’s guidelines. Students who take advanced math in eighth grade score approximately only .18 standard deviations higher in eighth grade self-concept than students who do not take advanced math in eighth grade.

d) An independent samples *t*-test was performed to determine whether females have better twelfth grade self-concept than males. Prior to the analysis itself, an analysis of the underlying assumptions was conducted. Because there are 227 males and 273 females, and both of these sample sizes are greater than 30, a violation of the normality assumption would probably not compromise the the validity of the results of the hypothesis test. An inspection of the sample means indicates that among the population of college-bound students who have always been at grade level, females (*M* = 30.42, *SD* = 7.16) do not have statistically significantly better twelfth grade self-concept than males (*M* = 32.75, *SD* = 7.13).

e) A one-tailed, independent samples *t*-test was performed to determine whether females do better on twelfth grade reading achievement tests than males. Prior to the analysis itself, an analysis of the underlying assumptions was conducted. Because there are 227 males and 273 females and both of these sample sizes are greater than 30, a violation of the normality assumption would probably not compromise the validity of the results of the hypothesis test. Because *p* = .04 is less than .05 for Levene’s homogeneity of variances test, a statistically significant difference was detected in the population variances, so the homogeneity of variances assumption is not met for these data. For this reason, the independent samples *t*-test with equal variances not assumed was used for the analysis. According to the results of the independent samples *t*-test for unequal variances, among the population of college-bound students who have always been at grade level, females (*M* = 55.84, *SD* = 7.48) do not have statistically significantly better twelfth grade reading achievement scores than males (*M* = 55.31, *SD* = 8.56), *t*(452.53) = -.74, *p* = .23.

f) A one-tailed, independent samples *t*-test was performed to determine whether those who attended nursery school tend to have smaller families than those who did not. Prior to the analysis itself, an analysis of the underlying assumptions was conducted. Because there are 139 students who did not attend nursery school and 281 who did and both of these sample sizes are greater than 30, a violation of the normality assumption would likely not compromise the validity of the results of the hypothesis test. Because *p* = .03 is less than .05 for Levene’s homogeneity of variances test, a statistically significant difference was detected in the population variances, so the homogeneity of variances assumption is not met for these data. For this reason, the independent samples *t*-test with equal variances not assumed was used for the analysis. According to the results of the independent samples *t*-test for unequal variances, among the population of college-bound students who have always been at grade level, students who attended nursery school (*M* = 4.54, *SD* = 1.16) have statistically significantly smaller families than those who did not (*M* = 4.80, *SD* = 1.47), *t*(226.79) = 1.81, *p* = .04. Among the population of college-bound students who have always been at grade level, students who attended nursery school have families that are only approximately .20 standard deviations smaller than those who did not, a small effect according to Cohen’s guidelines.

1. A one-tailed, independent samples *t*-test was performed to determine whether students who live in the Northeast have mean ses that is higher than that of those who live in the West. Prior to the analysis itself, an analysis of the underlying assumptions was conducted. Because there are 106 students from the Northeast and 93 students from the West and both of these sample sizes are greater than 30, a violation of the normality assumption would probably not compromise the validity of the results of the hypothesis test. Because *p* = .16 is greater than .05 for Levene’s homogeneity of variances test, no statistically significant difference was detected in the population variances, so the homogeneity of variances assumption may be considered met for these data. According to the results of the independent samples *t*-test, among the population of college-bound students who have always been at grade level, the average ses of students from the Northeast (*M* = 20.25, *SD* = 6.99) is statistically significantly higher than the average ses of students from the West (*M* = 18.38, *SD* = 6.68), *t*(197) = 1.93, *p* = .03.
   1. A two-tailed, independent samples *t*-test is performed to determine whether there is a difference in the average blood pressure reduction between the calcium and the placebo groups for African American males. Prior to the *t*-test itself, an analysis of the underlying assumptions is performed. The normality assumption is found to be tenable by looking at the skewness ratios. For both the placebo group (skew = .53, skewness ratio = .80) and the calcium group (skew = .31, skewness ratio = .45) the skew and skewness ratio values are within acceptable limits, the normality assumption is tenable. The results of Levene’s test indicate that the homogeneity of variances assumption is tenable (*p* = .051), although barely. The results of the independent samples *t*-test indicate that no statistically significant difference is detected in blood pressure reduction between the placebo (*M* = -.64, *SD* = 5.87) and calcium (*M* = 4.90, *SD* = 8.67) groups among African American males, *t*(19) = -1.73, *p* = .10.
   2. The test is performed using the R command **t.test(math, read, paired = T)**.

****

a) The sample sizes are large enough so that failure to meet the normality assumption will not compromise the validity of the results of the *t*-test.

b) Because the sample sizes are equal in the paired samples *t*-test, failure to meet the homogeneity of variances assumption will not compromise the validity of test results in this case.

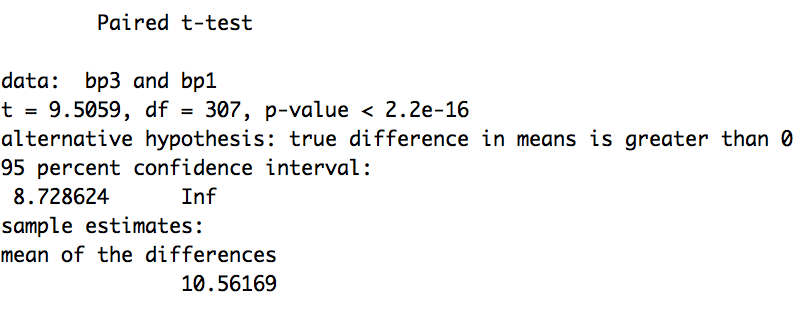
c) H0: mathcomp = readcomp and H1: mathcomp ≠ readcomp

d) The results of the paired samples *t*-test indicate that students diagnosed with learning disabilities in the city scored statistically significantly higher in math comprehension, on average, (*M* = 86.41, *SD* = 14.59) than they did in reading comprehension (*M* = 77.78, *SD* = 13.14), *t*(73) = 5.298, *p* < .0005.

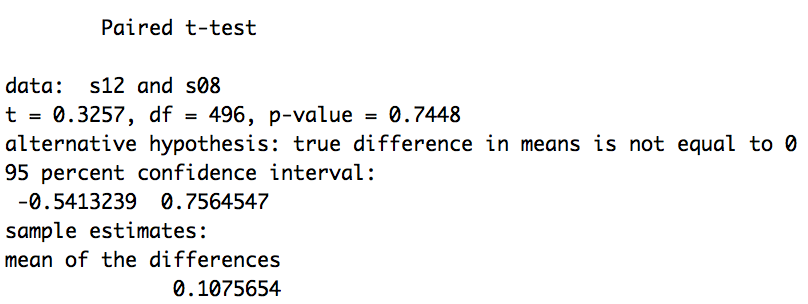
e) Because 0 is not contained in the confidence interval (5.38, 11.86), we know that performance on the two tests is statistically significantly different. Looking at the sample means, we conclude that public elementary school students in the city diagnosed with learning disabilities score between 5.38 and 11.87 points higher in math comprehension, on average, (*M* = 86.41, *SD* = 14.59) than they do in reading comprehension (*M* = 77.78, *SD* = 13.14).

f) Public elementary school children in the city score approximately .62 standard deviations (of the paired differences) higher in math comprehension, on average, than they do in reading comprehension: a moderate effect, supporting the result of the hypothesis test.

* 1. The test is performed using the R command **t.test(bp3, bp1, paired = T, alternative = "greater")**.

****

1. *P* < .00005. The means are in the hypothesized direction.
2. According to the *p*-value, the mean systolic blood pressure at time 3 (in 1968) (*M* = 137.99, *SD* = 22.10) is statistically significantly higher than at time 1(in 1956) (*M* = 127.43, *SD* = 19.15), *t*(307) = 9.51, *p* < .0005.
3. According to the value of Cohen’s *d*, the mean systolic blood pressure at time 3 (1968) is .51 standard deviations (of the paired differences) higher than at time 1 (1956), a moderate effect according to Cohen’s guidelines.
   1. The test is needed would be a paired *t* test, which requires using only individuals with a value for both DIABP1 and DIABP3. Limiting to these individuals can be done in R using the following commands:  
        
      **d1 = Framingham$DIABP1[complete.cases(Framingham$DIABP1) & complete.cases(Framingham$DIABP3)]  
        
      d3 = Framingham$DIABP3[complete.cases(Framingham$DIABP1) & complete.cases(Framingham$DIABP3)]**   
        
      However, when we check the means for d1 and d3 we see that actually running the test is unnecessary here: the direction of the mean change (from 80.77 in 1956 to 81.67 in 1968) suggests an increase rather than a decrease in diastolic blood pressure.
   2. The test is performed using the R command **t.test(s12, s08, paired = T)**.

****

a) A paired samples *t*-test is more appropriate than a one-sample *t*-test because the question relates to two population means: eighth grade science achievement and twelfth grade science achievement. A paired samples *t*-test is more appropriate than an independent samples *t*-test because each eighth grade science achievement score is paired with the twelfth grade science achievement score of the same person. The paired samples *t*-test may be viewed as equivalent to a one-sample *t*-test on the difference scores calculated as the difference between achsci08 and achsci12, two interval-leveled variables.

b) Because there are 497 scores for both eighth and twelfth grade science achievement and both of these numbers are larger than 30, a violation of the normality assumption would not compromise the validity of the results of the hypothesis test.

c) H0: achsci08 = achsci12 and H1: achsci08 ≠ achsci12

d) *p* = 0.74.

e) The *p*-value suggests that the change in science achievement from *M* = 55.67 (*SD* = 9.39) in eighth grade to *M* = 55.78 (*SD* = 8.57) in twelfth grade for this population of students is not statistically significantly different from zero, *t*(496) = 0.33, *p* = 0.75.

f) The 95 percent confidence interval for the mean difference in science achievement between eighth and twelfth grade is (-0.76, 0.54) or (-0.54, 0.76).

g) Because 0 is contained in the 95 percent confidence interval, zero is a plausible value for the mean difference in science achievement from eighth to twelfth grade for this population. Said differently, the level of science achievement in eighth grade relative to all students in the grade is not statistically significantly different from that in twelfth grade.

A one-tailed, paired samples *t*-test was performed to determine whether among the population of college-bound students who have always been at grade level, the level of self concept, relative to all students in the grade, increases from eighth (slfcnc08) to twelfth grade (slfcnc12). Prior to the analysis itself, an analysis of the underlying assumptions was conducted. Because all 500 students in the dataset filled out the self concept questionnaires in both eighth and twelfth grade, the paired sample size is greater than 30 and a violation of the normality assumption would not compromise the validity of the results of the hypothesis test. Because the sample sizes are equal, a violation of the homogeneity of variances assumption also would not compromise the validity of the results of the hypothesis test. According to the results of the paired samples *t*-test, among the population of college-bound students who have always been at grade level, the level of self concept, relative to all students in the grade, increases statistically significantly from eighth (*M* = 21.06, *SD* = 5.97) to twelfth (*M* = 31.48, *SD* = 7.23) grade, *t*(499) = 33.06, *p* < .0005. Among the population of college-bound students who have always been at grade level, the level of self concept, relative to all students in the grade, is 1.57 standard deviations (of the paired differences) higher in twelfth grade than in eighth grade, a large effect size according to Cohen’s guidelines.

a) (6)

b) (5)

c) (2)

d) (4)

e) (3)

a) A one-tailed, independent samples *t*-test was performed to determine whether among college-bound students who are always at grade level, those who attended nursery school tend to have higher socioeconomic statuses (ses) than those who did not. Prior to the *t*-test itself, an analysis of the underlying assumptions was performed. The normality assumption was not an issue because the sample sizes were far greater than 30 for both nursery groups. The results of Levene’s test indicated that the homogeneity of variances assumption was met (*p* = .53). The results of the independent samples *t*-test indicated that students who attended nursery school (*M* = 20.74, *SD* = 6.51) had a statistically significantly higher SES, on average, than those who did not (*M* = 15.07, *SD* = 6.24), *t*(418) = 8.51, *p* < 0.0005. Those who attended nursery school have ses scores that are approximately 0.88 standard deviations higher than those who did not, a large effect according to Cohen’s guidelines.

b) A two-tailed, paired samples *t*-test was performed to determine whether among college-bound students who are always at grade level, self-concept differed in eighth and tenth grades. Prior to the *t*-test itself, an analysis of the underlying assumptions was performed. The normality assumption was not an issue because the sample sizes were far greater than 30 for the eighth and tenth grade scores. The homogeneity of variances assumption was also not a concern because there were the same number of eighth and tenth grade scores. The results of the paired samples *t*-test indicated that students had statistically significantly higher self-concept scores, on average, in tenth grade (*M* = 22.62, *SD* = 6.87) than they did in eighth grade (*M* = 21.06, *SD* = 5.97), *t*(499) = 5.69, *p* < .0005. The tenth grade self-concept scores were approximately 0.24 standard deviations higher than their eighth grade self-concept scores, a relatively small effect according to Cohen’s guidelines.

c) A two-tailed, independent samples *t*-test was performed to determine whether among college-bound students who are always at grade level, those who attended public school tend to perform differently in twelfth grade math achievement from those who attended private school. Prior to the *t*-test itself, an analysis of the underlying assumptions was performed. The normality assumption was not an isssue because the sample sizes were far greater than 30 for both school groups. The results of Levene’s test indicated that the homogeneity of variances assumption was met (*p* = .32). The results of the independent samples *t*-test indicated that students who attended private school (*M* = 58.85, *SD* = 7.45) had a statistically significantly higher twelfth grade math achievement, on average, than those who did not (*M* = 56.22, *SD* = 7.93), *t*(498) = -3.30, *p* = .001. Students in private schools scored approximately .34 standard deviations higher in twelfth grade math achievement than those in public schools, a small to moderate effect according to Cohen’s guidelines.

d) A two-tailed, one sample *t*-test was performed to determine whether among college-bound students who are always at grade level, families typically have four members. Prior to the *t*-test itself, an analysis of the normality assumption was performed. Because the sample size was far greater than 30, the normality assumption was not a concern. The results of the one sample *t*-test indicated that families (*M* = 4.69, *SD* = 1.32) have statistically significantly more than four members, on average, *t*(499) = 11.73, *p* < .0005. The family size was approximately .52 standard deviations higher than 4, a moderate effect according to Cohen’s guidelines.

e) A one-tailed, paired samples *t*-test was performed to determine whether college-bound students who are always at grade level take more years of English than math. Prior to the *t*-test itself, an analysis of the underlying assumptions was performed. The normality assumption was not a concern because the sample sizes were far greater than 30 for both subjects. The homogeneity of variances assumption was also not a concern because the two groups were paired and therefore, the number in each group was the same. The results of the paired samples *t*-test indicates that students took statistically significantly more years of English (*M* = 4.13, *SD* = .66) than math (*M* = 3.64, *SD* = .81), on average, *t*(499) = 11.39, *p* < .0005. Students took approximately .67 standard deviations more years of English than math, a moderate to large effect according to Cohen’s guidelines.

* 1. The R commands to generate the output for this exercise are

**boxplot(NELS$achmat08, NELS$achmat12, names = c("Eighth", "Twelfth"))**

**t.test(NELS$achmat12, NELS$achmat08, paired = T)**

**math\_diff = NELS$achmat12 - NELS$achmat08**

**t.test(math\_diff)**

1. Based on the boxplots, it appears as though the performance in math is slightly higher for the twelfth graders. A hypothesis test is necessary to determine whether the difference is statistically significant.

b) The results of the paired samples *t*-test indicate that among college-bound high school students who are always at grade level, there is not a statistically significant difference in math achievement between twelfth grade (*M* = 56.91, *SD* = 7.88) and eighth grade (*M* = 56.59, *SD* = 9.34), *t*(499) = 1.25, *p* = .21.

c) The results of the one sample *t*-test indicate that among college-bound high school students who are always at grade level, the difference between math achievement in twelfth grade and eighth grade (*M* = .32, *SD* = 5.67) is not statistically significantly different from 0, *t*(499) = 1.25, *p* = .21. Thus, level of math achievement, on average, does not change from eighth to twelfth grades.

d) The *t*-ratios are identical. The two approaches are equivalent.

e) In general, the independent and paired *t*-tests will yield different results, especially when the pairing of subjects across groups controls for extraneous factors related to the dependent variable. In this case, the paired sample *t*-test will be more powerful than the independent samples *t*-test, thereby increasing the chance of obtaining statistical significance.

* 1. A one-sample *t*-test on means should not be used in this case because it is not appropriate to calculate the mean of an ordinal leveled variable, such as late12.

a) Because the focus here is on drawing a conclusion about the students in the NELS dataset, this question requires the use of only descriptive statistics. Based on the higher mean value, students in the NELS dataset who never smoked cigarettes (*M* = 57.51, *SD* = 7.71), on average, outperformed those who did (*M* = 53.24, *SD* = 8.00) in terms of twelfth grade math achievement.

b) Because the focus here is in drawing a conclusion about the population of college-bound students who are always on grade level, this question requires the use of inferential statistics. The question to be addressed is whether the observed difference in means is likely to be due to chance. Prior to performing the independent samples *t*-test, an analysis of the underlying assumptions was performed. Based on the results of the independent samples *t*-test, among college-bound students who are always at grade level, those who never smoked cigarettes (*M* = 57.51, *SD* = 7.71), on average, outperformed those who did (*M* = 53.24, *SD* = 8.00) in terms of twelfth grade math achievement, *t*(498) = 4.31, *p* < .0005.

*The p-values for exercises 11.32 – 11.42 were estimated using Table 2 in Appendix C.*

a) .02

b) Because the sample mean does not support the direction of the alternative hypothesis, the result is not statistically significant.

c) .01

a) .20

b) Because the sample means do not support the direction of the alternative hypothesis, the result is not statistically significant.

c) .10

a) .20 < *p* < .30

b) .10 < *p* < .15

c) Because the sample means do not support the direction of the alternative hypothesis, the result is not statistically significant.

a) .05 < *p* < .10

b) Because the sample mean does not support the direction of the alternative hypothesis, the result is not statistically significant.

c) .025 < *p* < .05

Mean difference = 81.50 – 85 = -3.5



*df* = 105 – 1 = 104

Using Table 2, using the approximation *df* = 100, .001 < *p* < .005

Using R with the command **2\*pt(-3.28, 104)** we obtain *p* = .0014

The mean intellectual ability score for all children attending public elementary school in the city (*M* = 81.5, *SD* = 10.94) is statistically significantly lower than 85, *t*(104) = -3.28, *p* = .001.

*Df* = 105 – 1 = 104

The standard error of the mean = 

Using Table 2, using the approximation *df* = 100, *tc* = 1.984

Using R with the command **qt(.025, 104, lower.tail=F)** we obtain *tc* = 1.983

The 95 percent confidence interval is 81.5 ± 1.983\*1.068

= 81.5 ± 2.118

= (79.38, 83.62)

The mean intellectual ability score for all children attending public elementary school in the city (*M* = 81.5, *SD* = 10.94) is statistically significantly lower than 85 because the 95 percent confidence interval is (79.38, 83.62), which does not contain 85.

Mean difference = 84.05 – 90.21 = -6.16 or 90.21 – 84.05 = 6.16



*Df* = 60 + 34 - 2 = 92

Using Table 2, using the approximation *df* = 90, .02 < *p* < .04

Using R with the command **2\*pt(2.13, 92, lower.tail = F)** we obtain *p* = .036

Using Table 2, using the approximation *df* = 90, *tc* = 1.987

Using R with the command **qt(.025, 92, lower.tail=F)** we obtain *tc* = 1.986

The 95 percent confidence interval is -6.16 ± 1.986\*2.898

= -6.16 ± 5.76

= (-11.92, -.40)

With 6.16 as the mean difference, the confidence interval is (.40, 11.92)

The results of the independent samples *t*-test indicate that female public elementary students with learning disabilities in the city (*M* = 90.21, *SD* = 14.42) are statistically significantly different from males from the same population with respect to math comprehension (*M* = 84.05, *SD* = 12.96), *t*(92) = -2.13, *p* = .04.

Mean difference = 86.41– 77.78 = 8.63 or 77.78 – 86.41 = -8.63



*Df* = 74 - 1 = 73

Using Table 2, using the approximation *df* = 70, *p* < .001

Using R with the command **2\*pt(5.30, 73, lower.tail = F)** we obtain *p* = .0000012

Using Table 2, using the approximation *df* = 70, *tc* = 1.994

Using R with the command **qt(.025, 73, lower.tail = F)** we obtain *tc* = 1.993

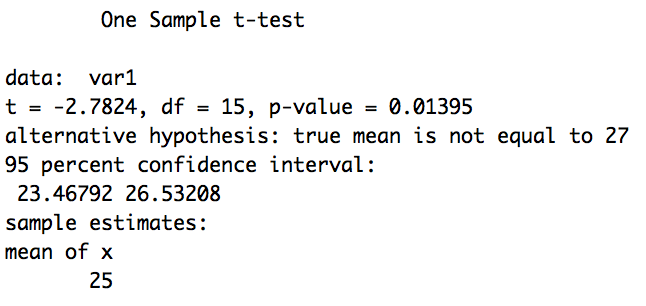
The 95 percent confidence interval is 8.63 ± 1.992\*1.627

= 8.63 ± 3.24 = (5.39, 11.87)

With -8.63 as the mean difference, the confidence interval is (-11.87, -5.39)

The results of the paired samples *t*-test indicate that students diagnosed with learning disabilities in the city, on average, score statistically significantly different in math (*M* = 86.41, *SD* = 14.59) and reading comprehension (*M* = 77.78, *SD* = 13.14), *t*(73) = 5.30, *p* = .0000012.

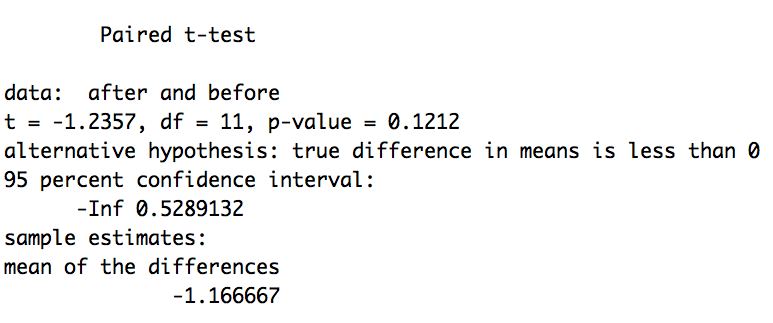
a) The computation can be performed by hand. Alternatively it can be performed using R. Using R, we enter the data and use the command **t.test(var1, mu = 27)** toobtain the following output:

****

The 95 percent CI for ** is (23.47, 26.53). Because 27 is not among these values, which are all below 27, we may conclude that 27 is not a plausible population mean value, and that in fact, the mean of the population is likely to be lower than 27.

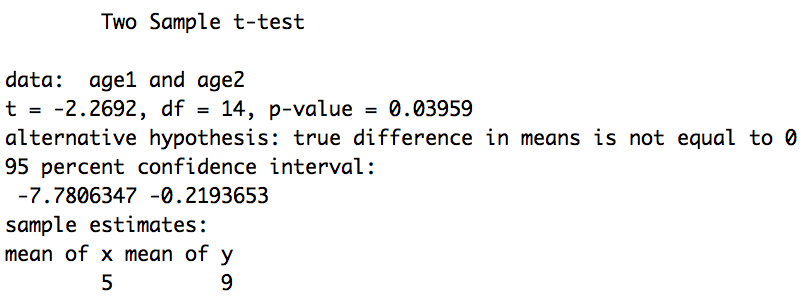
b) Because *p* < .05, we conclude that the mean in the population is statistically significantly different from 27, *t*(15) = -2.78, *p* = .014, and based on the value of the sample mean, likely to be lower than 27.

* 1. The computation can be performed by hand. Alternatively it can be performed using R. Using R, we enter the data and use the command **t.test(after, before, paired = T, alternative = "less")** toobtain the following output:



The results of the paired samples *t*-test indicate that the mental alertness scores are not statistically different before (*M* = 9.08, *SD* = 5.45) and after (*M* = 7.92, *SD* = 4.50) taking the drug, *t*(11) = -1.24, *p* = .12.

* 1. The computation can be performed by hand. Alternatively it can be performed using R. Using R, we enter the data and use the command **t.test(age1, age2, var.equal = T)** toobtain the following output:



The results of the independent samples *t*-test indicate that the scores are statistically significantly different between 2-year-olds (*M* = 9.00, *SD* = 3.70) and 1-year-olds (*M* = 5.00, *SD* = 3.34), *t*(14) = -2.27, *p* = .04.

* 1. The *t* distribution extends infinitely in both directions without crossing the horizontal axis; hence, there is always area under the curve beyond the observed *t*-value.
  2. It decreases.
  3. It decreases.
  4. No. It would not be ethical to do so. If she conducts a more powerful one-tailed test, but hypothesizes incorrectly about the direction of the mean difference, the penalty is that the results are not statistically significant. A one-tailed test is more powerful than a two-tailed test, as long as the mean differences are in the hypothesized direction. If she conducts a more powerful one-tailed test and hypothesizes correctly about the direction of the mean difference, but her results are not statistically significant, the results of the two-tailed test will also not be statistically significant.
  5. No. Once statistical significance is established, the values of the sample means are used to determine the nature of the mean difference.

1. Using the R command **pwr.t.test(d = 0.5, sig.level = 0.05, power = 0.8, type = "one.sample", alternative = "two.sided")**,we see that *N* = 34.
2. Using the R command **pwr.t.test(d = 0.5, sig.level = 0.05, power = 0.8, type = "one.sample", alternative = "greater")**,we see that *N* = 27.
3. Using the R command **pwr.t.test(d = 0.2, sig.level = 0.05, power = 0.8, type = "one.sample", alternative = "two.sided")**,we see that the appropriate sample size is *N* = 199.
4. Using the R command **pwr.t.test(d = 0.2, sig.level = 0.05, power = 0.8, type = "one.sample", alternative = "greater")**, we see that the appropriate sample size is *N* = 156.
5. The one-tailed test is inherently more powerful, as it requires a smaller sample size than that of a two-tailed test to detect the same size effect with the same power.
6. Using the R command **pwr.t.test(d = 0.2, sig.level = 0.05, power = 0.8, type = "two.sample", alternative = "two.sided"),** we see that the appropriate sample size is *N* = 788 (about 394 in each group).
7. Using the R command **pwr.t.test(d = 0.2, sig.level = 0.05, power = 0.8, type = "two.sample", alternative = "greater")**,we see that the appropriate sample size is *N* = 620 (about 310 in each group).
8. Using the R command **pwr.t.test(d = 0.2, sig.level = 0.05, power = 0.8, type = "paired", alternative = "two.sided")**, we see that the appropriate sample size is *N* = 199.
9. Using the R command **pwr.t.test(d = 0.2, sig.level = 0.05, power = 0.8, type = "paired", alternative = "greater")**,we see that the appropriate sample size is *N* = 156.