



Practice Exercises: Lesson 5.2 Solutions

Diez, D. M., Çetinkaya-Rundel, M., Barr, C. D. (2019). OpenIntro Statistics (4th ed.). OpenIntro.
<https://www.openintro.org/book/os/> CC BY-SA 3.0

STAT 1201
Introduction to Probability and Statistics

ONLINE AND DISTANCE EDUCATION

7.9 T is either -2.09 or 2.09. Then \bar{x} is one of the following:

$$\begin{aligned} -2.09 &= \frac{\bar{x} - 60}{\frac{8}{\sqrt{20}}} \rightarrow \bar{x} = 56.26 \\ 2.09 &= \frac{\bar{x} - 60}{\frac{8}{\sqrt{20}}} \rightarrow \bar{x} = 63.74 \end{aligned}$$

7.11 (a) We will conduct a 1-sample t -test. $H_0: \mu = 5$. $H_A: \mu \neq 5$. We'll use $\alpha = 0.05$. This is a random sample, so the observations are independent. To proceed, we assume the distribution of years of piano lessons is approximately normal. $SE = 2.2/\sqrt{20} = 0.4919$. The test statistic is $T = (4.6 - 5)/SE = -0.81$. $df = 20 - 1 = 19$. The one-tail area is about 0.21, so the p-value is about 0.42, which is bigger than $\alpha = 0.05$ and we do not reject H_0 . That is, we do not have sufficiently strong evidence to reject the notion that the average is 5 years.

(b) Using $SE = 0.4919$ and $t_{df=19}^* = 2.093$, the confidence interval is (3.57, 5.63). We are 95% confident that the average number of years a child takes piano lessons in this city is 3.57 to 5.63 years.

(c) They agree, since we did not reject the null hypothesis and the null value of 5 was in the t -interval.

7.13 If the sample is large, then the margin of error will be about $1.96 \times 100/\sqrt{n}$. We want this value to be less than 10, which leads to $n \geq 384.16$, meaning we need a sample size of at least 385 (round up for sample size calculations!).

7.15 Paired, data are recorded in the same cities at two different time points. The temperature in a city at one point is not independent of the temperature in the same city at another time point.

7.17 (a) Since it's the same students at the beginning and the end of the semester, there is a pairing between the datasets, for a given student their beginning and end of semester grades are dependent. (b) Since the subjects were sampled randomly, each observation in the men's group does not have a special correspondence with exactly one observation in the other (women's) group. (c) Since it's the same subjects at the beginning and the end of the study, there is a pairing between the datasets, for a subject student their beginning and end of semester artery thickness are dependent. (d) Since it's the same subjects at the beginning and the end of the study, there is a pairing between the datasets, for a subject student their beginning and end of semester weights are dependent.

7.19 (a) For each observation in one data set, there is exactly one specially corresponding observation in the other data set for the same geographic location. The data are paired. (b) $H_0: \mu_{diff} = 0$ (There is no difference in average number of days exceeding 90°F

in 1948 and 2018 for NOAA stations.) $H_A: \mu_{diff} \neq 0$ (There is a difference.) (c) Locations were randomly sampled, so independence is reasonable. The sample size is at least 30, so we're just looking for particularly extreme outliers: none are present (the observation off left in the histogram would be considered a clear outlier, but not a particularly extreme one). Therefore, the conditions are satisfied. (d) $SE = 17.2/\sqrt{197} = 1.23$. $T = \frac{2.9-0}{1.23} = 2.36$ with degrees of freedom $df = 197 - 1 = 196$. This leads to a one-tail area of 0.0096 and a p-value of about 0.019. (e) Since the p-value is less than 0.05, we reject H_0 . The data provide strong evidence that NOAA stations observed more 90°F days in 2018 than in 1948. (f) Type 1 Error, since we may have incorrectly rejected H_0 . This error would mean that NOAA stations did not actually observe a decrease, but the sample we took just so happened to make it appear that this was the case. (g) No, since we rejected H_0 , which had a null value of 0.

7.21 (a) $SE = 1.23$ and $t^* = 1.65$. $2.9 \pm 1.65 \times 1.23 \rightarrow (0.87, 4.93)$.

(b) We are 90% confident that there was an increase of 0.87 to 4.93 in the average number of days that hit 90°F in 2018 relative to 1948 for NOAA stations.

(c) Yes, since the interval lies entirely above 0.

7.23 (a) These data are paired. For example, the Friday the 13th in say, September 1991, would probably be more similar to the Friday the 6th in September 1991 than to Friday the 6th in another month or year.

(b) Let $\mu_{diff} = \mu_{sixth} - \mu_{thirteenth}$. $H_0: \mu_{diff} = 0$. $H_A: \mu_{diff} \neq 0$.

(c) Independence: The months selected are not random. However, if we think these dates are roughly equivalent to a simple random sample of all such Friday 6th/13th date pairs, then independence is reasonable. To proceed, we must make this strong assumption, though we should note this assumption in any reported results. Normality: With fewer than 10 observations, we would need to see clear outliers to be concerned. There is a borderline outlier on the right of the histogram of the differences, so we would want to report this in formal analysis results.

(d) $T = 4.93$ for $df = 10 - 1 = 9 \rightarrow$ p-value = 0.001.

(e) Since p-value < 0.05, reject H_0 . The data provide strong evidence that the average number of cars at the intersection is higher on Friday the 6th than on Friday the 13th. (We should exercise caution about generalizing the interpretation to all intersections or roads.)

(f) If the average number of cars passing the intersection actually was the same on Friday the 6th and 13th, then the probability that we would observe a test statistic so far from zero is less than 0.01.

(g) We might have made a Type 1 Error, i.e. incorrectly rejected the null hypothesis.

7.25 (a) $H_0 : \mu_{diff} = 0$. $H_A : \mu_{diff} \neq 0$. $T = -2.71$. $df = 5$. $p\text{-value} = 0.042$. Since $p\text{-value} < 0.05$, reject H_0 . The data provide strong evidence that the average number of traffic accident related emergency room admissions are different between Friday the 6th and Friday the 13th. Furthermore, the data indicate that the direction of that difference is that accidents are lower on Friday the 6th relative to Friday the 13th.

(b) (-6.49, -0.17).

(c) This is an observational study, not an experiment, so we cannot so easily infer a causal intervention implied by this statement. It is true that there is a difference. However, for example, this does not mean that a responsible adult going out on Friday the 13th has a higher chance of harm than on any other night.

7.27 (a) Chicken fed linseed weighed an average of 218.75 grams while those fed horsebean weighed an average of 160.20 grams. Both distributions are relatively symmetric with no apparent outliers. There is more variability in the weights of chicken fed linseed.

(b) $H_0 : \mu_{ls} = \mu_{hb}$. $H_A : \mu_{ls} \neq \mu_{hb}$.

We leave the conditions to you to consider.

$T = 3.02$, $df = \min(11, 9) = 9 \rightarrow p\text{-value} = 0.014$. Since $p\text{-value} < 0.05$, reject H_0 . The data provide strong evidence that there is a significant difference between the average weights of chickens that were fed linseed and horsebean.

(c) Type 1 Error, since we rejected H_0 .

(d) Yes, since $p\text{-value} > 0.01$, we would not have rejected H_0 .

7.29 $H_0 : \mu_C = \mu_S$. $H_A : \mu_C \neq \mu_S$. $T = 3.27$, $df = 11 \rightarrow p\text{-value} = 0.007$. Since $p\text{-value} < 0.05$, reject H_0 . The data provide strong evidence that the average weight of chickens that were fed casein is different than the average weight of chickens that were fed soybean (with weights from casein being higher). Since this is a randomized experiment, the observed difference can be attributed to the diet.

7.31 Let $\mu_{diff} = \mu_{pre} - \mu_{post}$. $H_0 : \mu_{diff} = 0$: Treatment has no effect. $H_A : \mu_{diff} \neq 0$: Treatment has an effect on P.D.T. scores, either positive or negative. Conditions: The subjects are randomly assigned to treatments, so independence within and between groups is satisfied. All three sample sizes are smaller than 30, so we look for clear outliers. There is a borderline outlier in the first treatment group. Since it is borderline, we will proceed, but we should report this caveat with any results. For all three groups: $df = 13$. $T_1 = 1.89 \rightarrow p\text{-value} = 0.081$, $T_2 = 1.35 \rightarrow p\text{-value} = 0.200$, $T_3 = -1.40 \rightarrow p\text{-value} = 0.185$. We do not reject the null hypothesis for any of these groups. As earlier noted, there is some uncertainty about if the method applied is reasonable for the first group.

7.33 Difference we care about: 40. Single tail of 90%: $1.28 \times SE$. Rejection region bounds: $\pm 1.96 \times SE$ (if 5% significance level). Setting $3.24 \times SE = 40$, solving in $SE = \sqrt{\frac{94^2}{n} + \frac{94^2}{n}}$, and solving for the sample size n gives 116 plots of land for each fertilizer.

7.35 Alternative.

7.37 $H_0 : \mu_1 = \mu_2 = \dots = \mu_6$. H_A : The average weight varies across some (or all) groups. Independence: Chicks are randomly assigned to feed types (presumably kept separate from one another), therefore independence of observations is reasonable. Approx. normal: the distributions of weights within each feed type appear to be fairly symmetric. Constant variance: Based on the side-by-side box plots, the constant variance assumption appears to be reasonable. There are differences in the actual computed standard deviations, but these might be due to chance as these are quite small samples. $F_{5,65} = 15.36$ and the $p\text{-value}$ is approximately 0. With such a small $p\text{-value}$, we reject H_0 . The data provide convincing evidence that the average weight of chicks varies across some (or all) feed supplement groups.

7.39 (a) H_0 : The population mean of MET for each group is equal to the others. H_A : At least one pair of means is different. (b) Independence: We don't have any information on how the data were collected, so we cannot assess independence. To proceed, we must assume the subjects in each group are independent. In practice, we would inquire for more details. Normality: The data are bound below by zero and the standard deviations are larger than the means, indicating very strong skew. However, since the sample sizes are extremely large, even extreme skew is acceptable. Constant variance: This condition is sufficiently met, as the standard deviations are reasonably consistent across groups. (c) See below, with the last column omitted:

	Df	Sum Sq	Mean Sq	F value
coffee	4	10508	2627	5.2
Residuals	50734	25564819	504	
Total	50738	25575327		

(d) Since $p\text{-value}$ is very small, reject H_0 . The data provide convincing evidence that the average MET differs between at least one pair of groups.

7.41 (a) H_0 : Average GPA is the same for all majors. H_A : At least one pair of means are different. (b) Since $p\text{-value} > 0.05$, fail to reject H_0 . The data do not provide convincing evidence of a difference between the average GPAs across three groups of majors. (c) The total degrees of freedom is $195 + 2 = 197$, so the sample size is $197 + 1 = 198$.

7.43 (a) False. As the number of groups increases, so does the number of comparisons and hence the modified significance level decreases. (b) True. (c) True. (d) False. We need observations to be independent regardless of sample size.

7.45 (a) H_0 : Average score difference is the same for all treatments. H_A : At least one pair of means are different. (b) We should check conditions. If we look back to the earlier exercise, we will see that the patients were randomized, so independence is satisfied. There are some minor concerns about skew, especially with the third group, though this may be acceptable. The standard deviations across the groups are reasonably similar. Since the p-value is less than 0.05, reject H_0 . The data provide convincing evidence of a difference between the average reduction in score among treatments. (c) We determined that at least two means are different in part (b), so we now conduct $K = 3 \times 2/2 = 3$ pairwise t -tests that each use $\alpha = 0.05/3 = 0.0167$ for a significance level. Use the following hypotheses for each pairwise test. H_0 : The two means are equal. H_A : The two means are different. The sample sizes are equal and we use the pooled SD, so we can compute $SE = 3.7$ with the pooled $df = 39$. The p-value for Trmt 1 vs. Trmt 3 is the only one under 0.05: p-value = 0.035 (or 0.024 if using s_{pooled} in place of s_1 and s_3 , though this won't affect the final conclusion). The p-value is larger than $0.05/3 = 1.67$, so we do not have strong evidence to conclude that it is this particular pair of groups that are different. That is, we cannot identify if which particular pair of groups are actually different, even though we've rejected the notion that they are all the same!

7.47 $H_0 : \mu_T = \mu_C$. $H_A : \mu_T \neq \mu_C$. $T = 2.24$, $df = 21 \rightarrow$ p-value = 0.036. Since p-value < 0.05, reject H_0 . The data provide strong evidence that the average food consumption by the patients in the treatment and control groups are different. Furthermore, the data indicate patients in the distracted eating (treatment) group consume more food than patients in the control group.

7.49 False. While it is true that paired analysis requires equal sample sizes, only having the equal sample sizes isn't, on its own, sufficient for doing a paired test. Paired tests require that there be a special correspondence between each pair of observations in the two groups.

7.51 (a) We are building a distribution of sample statistics, in this case the sample mean. Such a distribution is called a sampling distribution. (b) Because we are dealing with the distribution of sample means, we need to check to see if the Central Limit Theorem applies. Our sample size is greater than 30,

and we are told that random sampling is employed. With these conditions met, we expect that the distribution of the sample mean will be nearly normal and therefore symmetric. (c) Because we are dealing with a sampling distribution, we measure its variability with the standard error. $SE = 18.2/\sqrt{45} = 2.713$. (d) The sample means will be more variable with the smaller sample size.

7.53 (a) We should set 1.0% equal to 2.84 standard errors: $2.84 \times SE_{desired} = 1.0\%$ (see Example 7.37 on page 282 for details). This means the standard error should be about $SE = 0.35\%$ to achieve the desired statistical power.

(b) The margin of error was $0.5 \times (2.6\% - (-0.2\%)) = 1.4\%$, so the standard error in the experiment must have been $1.96 \times SE_{original} = 1.4\% \rightarrow SE_{original} = 0.71\%$.

(c) The standard error decreases with the square root of the sample size, so we should increase the sample size by a factor of $2.03^2 = 4.12$.

(d) The team should run an experiment 4.12 times larger, so they should have a random sample of 4.12% of their users in each of the experiment arms in the new experiment.

7.55 Independence: it is a random sample, so we can assume that the students in this sample are independent of each other with respect to number of exclusive relationships they have been in. Notice that there are no students who have had no exclusive relationships in the sample, which suggests some student responses are likely missing (perhaps only positive values were reported). The sample size is at least 30, and there are no particularly extreme outliers, so the normality condition is reasonable. 90% CI: (2.97, 3.43). We are 90% confident that undergraduate students have been in 2.97 to 3.43 exclusive relationships, on average.

7.57 The hypotheses should be about the population mean (μ), not the sample mean. The null hypothesis should have an equal sign and the alternative hypothesis should be about the null hypothesized value, not the observed sample mean. Correction:

$$H_0 : \mu = 10 \text{ hours}$$

$$H_A : \mu \neq 10 \text{ hours}$$

A two-sided test allows us to consider the possibility that the data show us something that we would find surprising.