

Chapter 7

Inference for numerical data

7.2 Inference for paired data

Are textbooks actually cheaper online? Here we compare the price of textbooks at UCLA's bookstore and prices at Amazon.com. Seventy-three UCLA courses were randomly sampled in Spring 2010, representing less than 10% of all UCLA courses.¹¹ A portion of this data set is shown in Table 7.9.

	dept	course	ucla	amazon	diff
1	Am Ind	C170	27.67	27.95	-0.28
2	Anthro	9	40.59	31.14	9.45
3	Anthro	135T	31.68	32.00	-0.32
4	Anthro	191HB	16.00	11.52	4.48
⋮	⋮	⋮	⋮	⋮	⋮
72	Wom Std	M144	23.76	18.72	5.04
73	Wom Std	285	27.70	18.22	9.48

Table 7.9: Six cases of the `textbooks` data set.

7.2.1 Paired observations and samples

Each textbook has two corresponding prices in the data set: one for the UCLA bookstore and one for Amazon. Therefore, each textbook price from the UCLA bookstore has a natural correspondence with a textbook price from Amazon. When two sets of observations have this special correspondence, they are said to be **paired**.

Paired data

Two sets of observations are *paired* if each observation in one set has a special correspondence or connection with exactly one observation in the other data set.

To analyze paired data, it is often useful to look at the difference in outcomes of each pair of observations. In the `textbook` data set, we look at the difference in prices, which is represented as the `diff` variable in the `textbooks` data. Here the differences are taken as

$$\text{UCLA price} - \text{Amazon price}$$

for each book. It is important that we always subtract using a consistent order; here Amazon prices are always subtracted from UCLA prices. If this difference is positive, the

⁹Choose Stats and let μ_0 be 100. Choose $>$ to correspond to H_A . $t = 2.39$ and $p\text{-value} = 0.012$.

¹⁰The interval is (105.21, 166.59).

¹¹When a class had multiple books, only the most expensive text was considered.

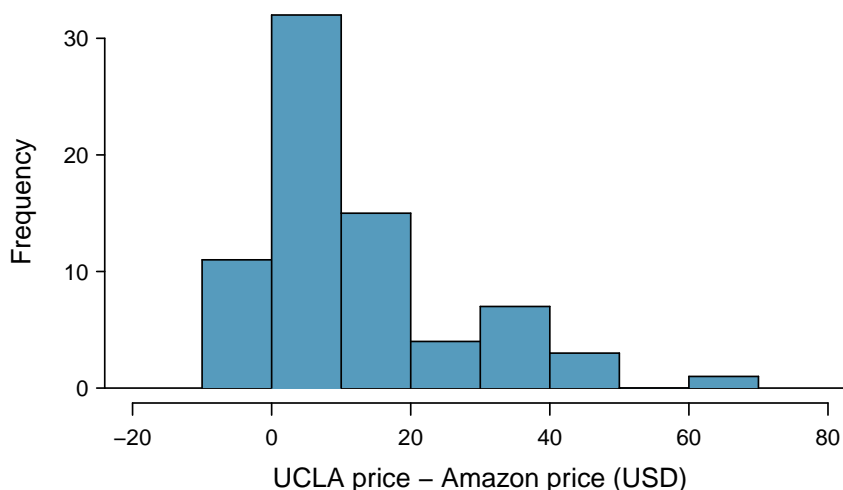


Figure 7.10: Histogram of the difference in price for each of the 73 books sampled. These data are strongly skewed.

UCLA price is higher. If this difference is negative, the Amazon price is higher. If this difference is zero, the two prices are equal. A histogram of these differences is shown in Figure 7.10. Using differences between paired observations is a common and useful way to analyze paired data.

- ⦿ **Guided Practice 7.14** The first difference shown in Table 7.9 is computed as $27.67 - 27.95 = -0.28$. Verify the differences are calculated correctly for observations 2 and 3.¹²

7.2.2 Hypothesis testing for paired data

To analyze a paired data set, we use the exact same tools that we developed in the previous section. Now we apply them to the differences in the paired observations.

n_{diff}	\bar{x}_{diff}	s_{diff}
73	12.76	14.26

Table 7.11: Summary statistics for the price differences. There were 73 books, so there are 73 differences.

- **Example 7.15** Set up and implement a hypothesis test to determine whether, on average, there is a difference between Amazon's price for a book and the UCLA bookstore's price.

There are two scenarios: there is no difference or there is some difference in average prices. The *no difference* scenario is always the null hypothesis:

$H_0: \mu_{diff} = 0$. There is no difference in the average textbook price.

¹²Observation 2: $40.59 - 31.14 = 9.45$. Observation 3: $31.68 - 32.00 = -0.32$.

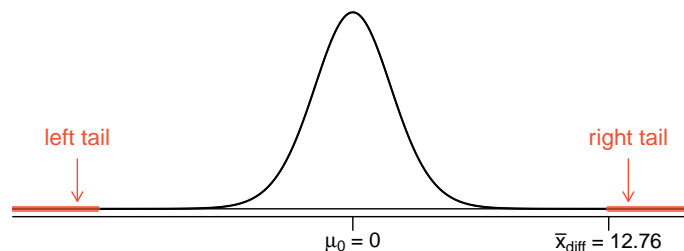


Figure 7.12: Sampling distribution for the mean difference in book prices, if the true average difference is zero.

$H_A: \mu_{diff} \neq 0$. There is a difference in average prices.

The standard deviation of all of the differences is unknown, so we will use the standard deviation of the sample differences. The observations are based on a simple random sample from less than 10% of all books sold at the bookstore, so independence is reasonable; the distribution of differences, shown in Figure 7.10, is strongly skewed, but this amount of skew is reasonable for this sized data set ($n = 73$). Because all three conditions are reasonably satisfied, we can conclude the t test is reasonable.

We compute the standard error associated with \bar{x}_{diff} using the standard deviation of the differences ($s_{diff} = 14.26$) and the number of differences ($n_{diff} = 73$):

$$SE_{\bar{x}_{diff}} = \frac{s_{diff}}{\sqrt{n_{diff}}} = \frac{14.26}{\sqrt{73}} = 1.67$$

To visualize the p-value, the sampling distribution of \bar{x}_{diff} is drawn as though H_0 is true, which is shown in Figure 7.12. The p-value is represented by the two (very) small tails.

To find the tail areas, we compute the test statistic, which is the t score of \bar{x}_{diff} under the null condition that the actual mean difference is 0:

$$t = \frac{\bar{x}_{diff} - 0}{SE_{\bar{x}_{diff}}} = \frac{12.76 - 0}{1.67} = 7.59 \quad df = 72$$

This t score is so large it isn't even in the table, which ensures the single tail area will be 0.0002 or smaller. A calculator gives a tail area as 4.5×10^{-11} . Since the p-value corresponds to both tails in this case and the t distribution is symmetric, the p-value can be estimated as twice the one-tail area:

$$\text{p-value} = 2 \times (\text{one tail area}) \approx 2 \times 4.5 \times 10^{-11} = 9 \times 10^{-11} \approx 0$$

Because the p-value is less than 0.05, we reject the null hypothesis. We have found convincing evidence that Amazon is, on average, cheaper than the UCLA bookstore for UCLA course textbooks.

Hypothesis test for paired data

1. State the name of the test being used: matched pairs t test.
2. Verify conditions.
 - Paired data from a random sample or experiment
 - Population of differences is known to be normal OR $n_{diff} \geq 30$ OR graph of sample differences is approximately symmetric with no outliers, making the assumption that population of differences is normal a reasonable one
3. Write the hypotheses in plain language, then set them up in mathematical notation.
 - $H_0 : \mu_{diff} = 0$
 - $H_0 : \mu_{diff} \neq$ or $<$ or > 0
4. Identify the significance level α .
5. Calculate the test statistic and df .

$$t = \frac{\text{point estimate} - \text{null value}}{\text{SE of estimate}}$$

Where the point estimate is \bar{x}_{diff} , $SE = \frac{s_{diff}}{\sqrt{n_{diff}}}$, and $df = n_{diff} - 1$.

6. Find the p-value and compare it to α to determine whether to reject or not reject H_0 .
7. Write the conclusion in the context of the question.

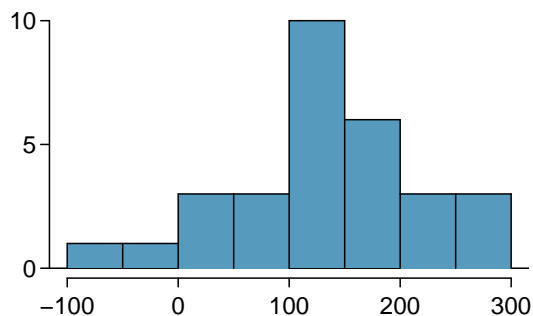


Figure 7.13: Sample distribution of: SAT score after course - SAT score before course. The distribution is approximately symmetric.

7.3 Difference of two means using the t distribution

It is also useful to be able to compare two means for small samples. For instance, a teacher might like to test the notion that two versions of an exam were equally difficult. She could do so by randomly assigning each version to students. If she found that the average scores on the exams were so different that we cannot write it off as chance, then she may want to award extra points to students who took the more difficult exam.

In a medical context, we might investigate whether embryonic stem cells can improve heart pumping capacity in individuals who have suffered a heart attack. We could look for evidence of greater heart health in the stem cell group against a control group.

In this section we use the t distribution for the difference in sample means. We will again drop the minimum sample size condition and instead impose a strong condition on the distribution of the data.

7.3.1 Sampling distribution for the difference of two means

In this section we consider a difference in two population means, $\mu_1 - \mu_2$, under the condition that the data are not paired. The methods are similar in theory but different in the details. Just as with a single sample, we identify conditions to ensure a point estimate of the difference $\bar{x}_1 - \bar{x}_2$ is nearly normal. Next we introduce a formula for the standard deviation of $\bar{x}_1 - \bar{x}_2$, which allows us to apply our general tools from Section 5.

We apply these methods to two examples: participants in the 2012 Cherry Blossom Run and newborn infants. This section is motivated by questions like “Is there convincing evidence that newborns from mothers who smoke have a different average birth weight than newborns from mothers who don’t smoke?”

We start by looking at the population mean and standard deviation for the run times of men and women participants in the 2009 Cherry Blossom Run. Table 7.15 summarizes these values.

	men	women
μ	87.65	102.13
σ	12.5	15.2

Table 7.15: Summary of the run time of participants in the 2009 Cherry Blossom Run.

¹⁸Enter the data into L1 and L2 on a calculator. Let $L3 = L1 - L2$. After selecting TTest, choose DATA, let μ_0 be 0, and let List be **L3**. Let Freq be 1 and select $>$. $t = 3.076$ and p-value = 0.0109.

¹⁹The data have already been entered into L1 and L2 and the differences should be in L3. After selecting TInterval, choose DATA, let List be **L3**. Let Freq be 1 and let C-Level be 0.95. The interval is (.80354, 7.0507).

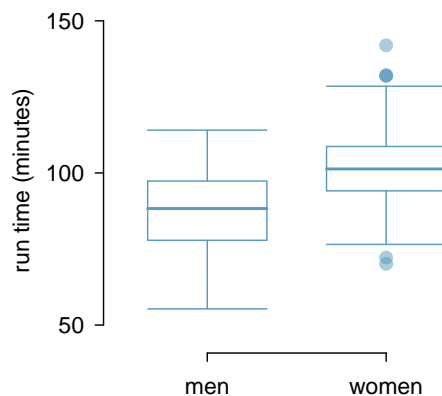


Figure 7.16: Side-by-side box plots for the sample of 2009 Cherry Blossom Run participants.

The two populations (men and women) are independent of one-another, so the data are not paired.²⁰ If we take two separate random samples of men and women from this race, what is the expected value for the difference in their average times? Not surprisingly, the expected value of $\bar{x}_w - \bar{x}_m$ is $\mu_1 - \mu_2$. We can quantify the variability in the point estimate, using the following formula for its standard deviation:

$$\begin{aligned} SD_{\bar{x}_w - \bar{x}_m} &= \sqrt{(SD_{\bar{x}_w})^2 + (SD_{\bar{x}_m})^2} \\ &= \sqrt{\left(\frac{\sigma_{\bar{x}_w}}{\sqrt{n_w}}\right)^2 + \left(\frac{\sigma_{\bar{x}_m}}{\sqrt{n_m}}\right)^2} \\ &= \sqrt{\frac{\sigma_w^2}{n_w} + \frac{\sigma_m^2}{n_m}} \end{aligned}$$

- ⊙ **Guided Practice 7.23** Let's say we take a random sample of 55 women and a random sample of 45 men. Use the SD formula for the difference of two means to compute the SD for the difference in the average run time for males and females.²¹

²⁰Probability theory guarantees that the difference of two independent normal random variables is also normal. Because each sample mean is nearly normal and observations in the samples are independent, we are assured the difference is also nearly normal.

²¹ $\sqrt{\frac{15.2^2}{55} + \frac{12.5^2}{45}} = 2.77$

Distribution of a difference of sample means

The sample difference of two means, $\bar{x}_1 - \bar{x}_2$, is nearly normal with mean $\mu_1 - \mu_2$ and standard deviation

$$SD_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad (7.24)$$

when each sample mean is nearly normal and all observations are independent. Recall that each sample mean will be nearly normal if the population is normal or if the sample size is at least 30.

7.3.2 Point estimates and standard errors for differences of means

In the example of two exam versions, the teacher would like to evaluate whether there is convincing evidence that the difference in average scores between the two exams is not due to chance.

It will be useful to extend the t distribution method from Section 7.1 to apply to a difference of means:

$$\bar{x}_1 - \bar{x}_2 \quad \text{as a point estimate for} \quad \mu_1 - \mu_2$$

First, we verify the small sample conditions (independence and nearly normal data) for each sample separately, then we verify that the samples are also independent. For instance, if the teacher believes students in her class are independent, the exam scores are nearly normal, and the students taking each version of the exam were independent, then we can use the t distribution for inference on the point estimate $\bar{x}_1 - \bar{x}_2$.

The formula for the standard error of $\bar{x}_1 - \bar{x}_2$, introduced in Section 7.3.1, also applies to small samples:

$$SE_{\bar{x}_1 - \bar{x}_2} = \sqrt{SE_{\bar{x}_1}^2 + SE_{\bar{x}_2}^2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad (7.25)$$

Because we will use the t distribution, we will need to identify the appropriate degrees of freedom. This can be done using a calculator or computer software. An alternative technique is to use the smaller of $n_1 - 1$ and $n_2 - 1$.²²

Using the t distribution for a difference in means

The t distribution can be used for inference when working with the standardized difference of two means if (1) each sample meets the conditions for using the t distribution and (2) the samples are independent. We estimate the standard error of the difference of two means using Equation (7.25).

7.3.3 Hypothesis testing for the difference of two means

Summary statistics for each exam version are shown in Table 7.17. The teacher would like to evaluate whether this difference is so large that it provides convincing evidence that Version B was more difficult (on average) than Version A.

²²This technique for degrees of freedom is conservative with respect to a Type 1 Error; it is more difficult to reject the null hypothesis using this df method.

Version	n	\bar{x}	s	min	max
A	30	79.4	14	45	100
B	27	74.1	20	32	100

Table 7.17: Summary statistics of scores for each exam version.

- ⊙ **Guided Practice 7.26** Construct a two-sided hypothesis test to evaluate whether the observed difference in sample means, $\bar{x}_A - \bar{x}_B = 5.3$, might be due to chance.²³
- ⊙ **Guided Practice 7.27** To evaluate the hypotheses in Guided Practice 7.26 using the t distribution, we must first verify assumptions. (a) Does it seem reasonable that the scores are independent within each group? (b) What about the normality condition for each group? (c) Do you think scores from the two groups would be independent of each other (i.e. the two samples are independent)?²⁴

After verifying the conditions for each sample and confirming the samples are independent of each other, we are ready to conduct the test using the t distribution. In this case, we are estimating the true difference in average test scores using the sample data, so the point estimate is $\bar{x}_A - \bar{x}_B = 5.3$. The standard error of the estimate can be calculated using Equation (7.25):

$$SE = \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}} = \sqrt{\frac{14^2}{30} + \frac{20^2}{27}} = 4.62$$

Finally, we construct the test statistic:

$$T = \frac{\text{point estimate} - \text{null value}}{SE} = \frac{(79.4 - 74.1) - 0}{4.62} = 1.15$$

If we have a calculator or computer handy, we can identify the degrees of freedom as 45.97. Otherwise we use the smaller of $n_1 - 1$ and $n_2 - 1$: $df = 26$.

- ⊙ **Guided Practice 7.28** Identify the p-value, shown in Figure 7.18. Use $df = 26$.²⁵

In Guided Practice 7.28, we could have used $df = 45.97$. However, this value is not listed in the table. In such cases, we use the next lower degrees of freedom (unless the computer also provides the p-value). For example, we could have used $df = 45$ but not $df = 46$. As before, we provide a summary of the steps to perform when carrying out such a test.

²³Because the teacher did not expect one exam to be more difficult prior to examining the test results, she should use a two-sided hypothesis test. H_0 : the exams are equally difficult, on average. $\mu_A - \mu_B = 0$. H_A : one exam was more difficult than the other, on average. $\mu_A - \mu_B \neq 0$.

²⁴(a) It is probably reasonable to conclude the scores are independent. (b) The summary statistics suggest the data are roughly symmetric about the mean, and it doesn't seem unreasonable to suggest the data might be normal. Note that since these samples are each nearing 30, moderate skew in the data would be acceptable. (c) It seems reasonable to suppose that the samples are independent since the exams were handed out randomly.

²⁵We examine row $df = 26$ in the t table. Because this value is smaller than the value in the left column, the p-value is larger than 0.200 (two tails!). Because the p-value is so large, we do not reject the null hypothesis. That is, the data do not convincingly show that one exam version is more difficult than the other, and the teacher should not be convinced that she should add points to the Version B exam scores.

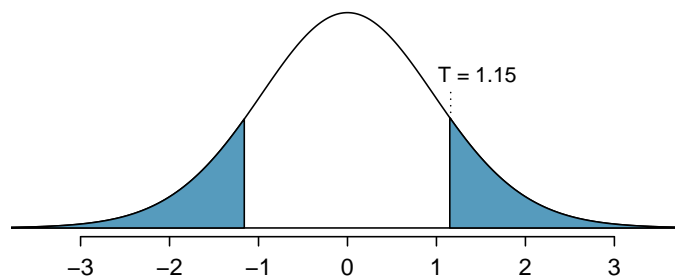


Figure 7.18: The t distribution with 26 degrees of freedom. The shaded right tail represents values with $T \geq 1.15$. Because it is a two-sided test, we also shade the corresponding lower tail.

Hypothesis test for the difference of two means

1. State the name of the test being used: 2-sample t test.
2. Verify conditions.
 - 2 independent random samples OR 2 randomly allocated treatments
 - Both populations known to be normal OR $n_1 \geq 30$ and $n_2 \geq 30$ OR graphs of both samples are approximately symmetric with no outliers, making the assumption that the populations are normal a reasonable one
3. Write the hypotheses in plain language, then set them up in mathematical notation.
 - $H_0 : \mu_1 = \mu_2$ or $\mu_1 - \mu_2 = 0$
 - $H_0 : \mu_1 \neq$ or $<$ or $> \mu_2$
4. Identify the significance level α .
5. Calculate the test statistic and df .

$$t = \frac{\text{point estimate} - \text{null value}}{\text{SE of estimate}}$$

Use a point estimate of $\bar{x}_1 - \bar{x}_2$, compute $SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$, and get the df from a calculator.

6. Find the p-value and compare it to α to determine whether to reject or not reject H_0 .
7. Write the conclusion in the context of the question.

	n	\bar{x}	s
ESCs	9	3.50	5.17
control	9	-4.33	2.76

Table 7.19: Summary statistics for the embryonic stem cell data set.

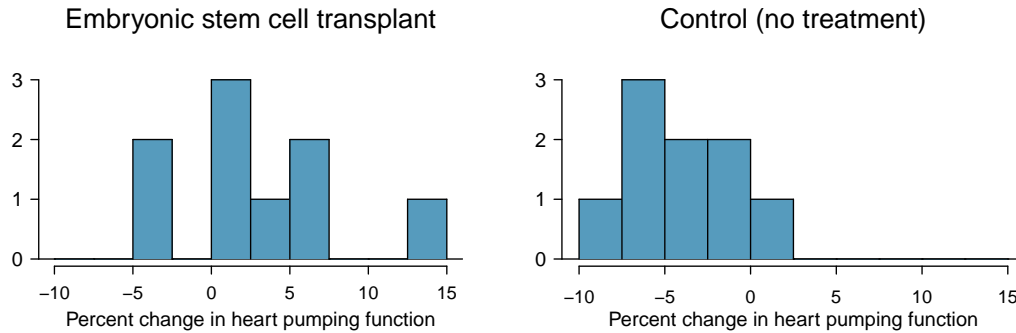


Figure 7.20: Histograms for both the embryonic stem cell group and the control group. Higher values are associated with greater improvement. We don't see any evidence of skew in these data; however, it is worth noting that skew would be difficult to detect with such a small sample.

- **Example 7.29** Do embryonic stem cells (ESCs) help improve heart function following a heart attack? Table 7.19 contains summary statistics for an experiment to test ESCs in sheep that had a heart attack. Each of these sheep was randomly assigned to the ESC or control group, and the change in their hearts' pumping capacity was measured. A positive value generally corresponds to increased pumping capacity, which suggests a stronger recovery. The sample data is graphed in Figure 7.20. Use the given information and an appropriate statistical test to answer the research question.

We will carry out a 2-sample t test. The first condition is met because it is stated that there were two randomly allocated treatments. For the second condition, we must look at a graphs of the data. The data are very limited, so we can only check for obvious outliers in the raw data in Figure 7.20. Since the distributions are (very) roughly symmetric, we will assume the populations are approximately normal.

$H_0: \mu_{esc} - \mu_{control} = 0$. The stem cells do not improve heart pumping function.

$H_A: \mu_{esc} - \mu_{control} > 0$. The stem cells do improve heart pumping function.

Let $\alpha = 0.05$. Now we compute the sample difference, the standard error for that point estimate, and the test statistic:

$$\bar{x}_{esc} - \bar{x}_{control} = 7.83 \quad SE = \sqrt{\frac{5.17^2}{9} + \frac{2.76^2}{9}} = 1.95 \quad T = \frac{7.83 - 0}{1.95} = 4.01$$

Using a calculator, $df = 12.2$ and $p\text{-value} = 8.4 \times 10^{-4}$. The p-value is much less than 0.05, so we reject the null hypothesis. The data provide convincing evidence that embryonic stem cells improve the heart's pumping function in sheep that have suffered a heart attack.

Appendix A

End of chapter exercise solutions

7 Inference for numerical data

7.1 (a) $df = 6 - 1 = 5$, $t_5^* = 2.02$ (column with two tails of 0.10, row with $df = 5$). (b) $df = 21 - 1 = 20$, $t_{20}^* = 2.53$ (column with two tails of 0.02, row with $df = 20$). (c) $df = 28$, $t_{28}^* = 2.05$. (d) $df = 11$, $t_{11}^* = 3.11$.

7.3 The mean is the midpoint: $\bar{x} = 20$. Identify the margin of error: $ME = 1.015$, then use $t_{35}^* = 2.03$ and $SE = s/\sqrt{n}$ in the formula for margin of error to identify $s = 3$.

7.5 (a) $H_0: \mu = 8$ (New Yorkers sleep 8 hrs per night on average.) $H_A: \mu < 8$ (New Yorkers sleep less than 8 hrs per night on average.) (b) Independence: The sample is random and from less than 10% of New Yorkers. The sample is small, so we will use a t distribution. For this size sample, slight skew is acceptable, and the min/max suggest there is not much skew in the data. $T = -1.75$. $df = 25 - 1 = 24$. (c) $0.025 < p\text{-value} < 0.05$. If in fact the true population mean of the amount New Yorkers sleep per night was 8 hours, the probability of getting a random sample of 25 New Yorkers where the average amount of sleep is 7.73 hrs per night or less is between 0.025 and 0.05. (d) Since $p\text{-value} < 0.05$, reject H_0 . The data provide strong evidence that New Yorkers sleep less than 8 hours per night on average. (e) No, as we rejected H_0 .

7.7 t_{19}^* is 1.73 for a one-tail. We want the lower tail, so set -1.73 equal to the T score, then solve for \bar{x} : 56.91.

7.9 (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

daily high temperature between January 1, 1968 and January 1, 2008 in the continental US.) $H_A: \mu_{diff} > 0$ (Average daily high temperature in January 1, 1968 was lower than average daily high temperature in January, 2008 in the continental US.) If you chose a two-sided test, that would also be acceptable. If this is the case, note that your p -value will be a little bigger than what is reported here in part (d). (c) Independence: locations are random and represent less than 10% of all possible locations in the US. The sample size is at least 30. We are not given the distribution to check the skew. In practice, we would ask to see the data to check this condition, but here we will move forward under the assumption that it is not strongly skewed. (d) $Z = 1.60 \rightarrow p\text{-value} = 0.0548$. (e) Since the $p\text{-value} > \alpha$ (since not given use 0.05), fail to reject H_0 . The data do not provide strong evidence of temperature warming in the continental US. However it should be noted that the p -value is very close to 0.05. (f) Type 2, since we may have incorrectly failed to reject H_0 . There may be an increase, but we were unable to detect it. (g) Yes, since we failed to reject H_0 , which had a null value of 0.

7.11 (a) (-0.03, 2.23). (b) We are 90% confident that the average daily high on January 1, 2008 in the continental US was 0.03 degrees lower to 2.23 degrees higher than the average daily high on January 1, 1968. (c) No, since 0 is included in the interval.

7.13 (a) Each of the 36 mothers is related to exactly one of the 36 fathers (and vice-versa), so there is a special correspondence between

the mothers and fathers. (b) $H_0 : \mu_{diff} = 0$. $H_A : \mu_{diff} \neq 0$. Independence: random sample from less than 10% of population. Sample size of at least 30. The skew of the differences is, at worst, slight. $Z = 2.72 \rightarrow$ p-value = 0.0066. Since p-value < 0.05, reject H_0 . The data provide strong evidence that the average IQ scores of mothers and fathers of gifted children are different, and the data indicate that mothers' scores are higher than fathers' scores for the parents of gifted children.

7.15 No, he should not move forward with the test since the distributions of total personal income are very strongly skewed. When sample sizes are large, we can be a bit lenient with skew. However, such strong skew observed in this exercise would require somewhat large sample sizes, somewhat higher than 30.

7.17 (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. With fewer than 10 observations, we would need to use the t distribution to model the sample mean. The normal probability plot of the differences shows an approximately straight line. There isn't a clear reason why this distribution would be skewed, and since the normal quantile plot looks reasonable, we can mark this condition as reasonably satisfied. (d) $T = 4.94$ for $df = 10 - 1 = 9 \rightarrow$ p-value < 0.01. (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 might believe this intersection is representative of all roads, i.e. there is higher traffic on Friday the 6th relative to Friday the 13th. However, we should be cautious of the required assumption for such a generalization.) (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.

7.19 (a) $H_0 : \mu_{diff} = 0$. $H_A : \mu_{diff} \neq 0$. $T = -2.71$. $df = 5$. $0.02 <$ p-value < 0.05. Since p-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.

7.21 (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. $T = 3.02$, $df = \min(11, 9) = 9 \rightarrow 0.01 <$ p-value < 0.02. Since p-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, since we rejected H_0 . (d) Yes, since p-value > 0.01, we would have failed to reject H_0 .

7.23 $H_0 : \mu_C = \mu_S$. $H_A : \mu_C \neq \mu_S$. $T = 3.48$, $df = 11 \rightarrow$ p-value < 0.01. Since p-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.

7.25 $H_0 : \mu_T = \mu_C$. $H_A : \mu_T \neq \mu_C$. $T = 2.24$, $df = 21 \rightarrow 0.02 <$ p-value < 0.05. 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.

7.27 Let $\mu_{diff} = \mu_{pre} - \mu_{post}$. $H_0 : \mu_{diff} = 0$:

Treatment has no effect. $H_A : \mu_{diff} > 0$: Treatment is effective in reducing Pd T scores, the average pre-treatment score is higher than the average post-treatment score. Note that the reported values are pre minus post, so we are looking for a positive difference, which would correspond to a reduction in the psychopathic deviant T score. Conditions are checked as follows. Independence: The subjects are randomly assigned to treatments, so the patients in each group are independent. All three sample sizes are smaller than 30, so we use t tests. Distributions of differences are somewhat skewed. The sample sizes are small, so we cannot reliably relax this assumption. (We will proceed, but we would not report the results of this specific analysis, at least for treatment group 1.) For all three groups: $df = 13$. $T_1 = 1.89$ ($0.025 < \text{p-value} < 0.05$), $T_2 = 1.35$ ($\text{p-value} = 0.10$), $T_3 = -1.40$ ($\text{p-value} > 0.10$). The only significant test reduction is found in Treatment 1, however, we had earlier noted that this result might not be reliable due to the skew in the distribution. Note that the calculation of the p-value for Treatment 3 was unnecessary: the sample mean indicated a increase in Pd T scores under this treatment (as opposed to a decrease, which was the result of interest). That is, we could tell without formally completing the hypothesis test that the p-value would be large for this treatment group.

7.29 $H_0: \mu_1 = \mu_2 = \dots = \mu_6$. 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-value is approximately 0. With such a small p-value, we reject H_0 . The data provide convincing evidence that the average weight of chicks varies across some (or all) feed supplement groups.