


7.3: Claims on Dependent Paired Variables

Learning Objectives

- Distinguish between dependent and independent samples
- Develop and apply hypothesis testing for dependent paired variables

 [Section 7.3 Excel File](#): (contains all of the data sets for this section)

Review and Preview

Recall the two studies about weekly recreational screen time (one real and one fabricated) from [Text Exercise 7.2.5](#). It might be tempting to conclude that the average weekly recreational screen time for emerging adults in 2024 is different from the average weekly recreational screen time for emerging adults in 2020 since we have evidence to say that the average in 2024 is different from 28.5 hours, but we must exercise a little caution. The original study (the real one) reported an estimate for the population mean using a confidence interval with a margin of error of 11.6 hours. This is a large margin of error. The average weekly recreational screen time for emerging adults in 2020 could be anywhere from 16.9 hours to 40.1 hours. We do not know precisely where it falls. So, using 28.5 hours as the conclusive average for 2020 is questionable. We will always be using estimates of parameters unless we conduct a census on the population. One might ask how can we ever proceed with these sorts of comparisons? Did not the standard population means from other problems come from interval estimates as well? The short answer is yes, they did, but there is more at play.

We first note that we can control the size of the margin of error by balancing confidence level and sample size. A more precise estimate can be obtained using a larger sample. If the margin of error were only 0.1 hours, we might feel more confident in treating the population mean as 28.5. We can also approach the problem using a different frame of reference. The general idea is that we are comparing two populations so we should make comparisons using the data from both populations. We compare them by collecting a random sample from each population and then analyzing the differences in the samples.

One methodology compared recreational screen time in 2018 to 2020. The original study used data from 2018 and 2020 from the same set of people. The researchers could study the difference in recreational screen time by each member of the sample. They had one sample from 2018 and another sample from 2020, but they were dependent upon each other because they consisted of the same set of people. We describe such a situation as one with dependent samples. Tests using dependent samples, often referred to as tests on dependent paired variables, provide strong results because they reduce the influence of confounding variables; there is less variation across one subject as a single treatment is applied than the variation present across the members of the population, but this is not the only way two populations can be compared.

Imagine the difficulty of keeping track of hundreds of participants over the course of months or years. Is it possible to make comparisons between 2018 and 2020 without conducting such a longitudinal study? The answer is yes. Two samples can be taken independently of each other. A random sample may be taken from one population and then another random sample may be taken from the other population. In the context of recreational screen time, a random sample may be taken in 2018 and then another random sample may be taken in 2020. Here we are not guaranteed that the same people will be in the two samples. It is possible that there is overlap, but the fact that a person was in the first sample does not affect the probability that they are in the second sample. We describe such a situation as one with independent samples. We will not address the methodologies involved in such claims in this text, but the interested reader is encouraged to study it independently. We now begin our development of testing claims on dependent paired variables.

Claims on Dependent Paired Variables

Researchers in medicine, education, and business are often interested in studying the effect of some treatment, educational practice, or product. It is quite natural to assess the patient, student, or consumer prior to some treatment and then assess them once the treatment has been in effect. Consider medical research: doctors can conduct pre-assessments and post-assessments to gauge the impact of a particular medical intervention on a random sample of patients. The doctors could simply compare the pre-assessment and post-assessment averages as if the samples were independently gathered, but there is a connection between the samples that is not being acknowledged, namely, that the same patients have two assessment values. We can measure effect of the medical intervention on each patient by considering the difference in the pre-assessment and post-assessment. In studying these paired differences, it is like we are studying a single sample and can utilize techniques already developed in this chapter to test claims.

A common concern of many people, especially in the medical community, is the consumption of chicken eggs. Previous [research](#) seems to indicate the possibility of a tie to heart disease and diabetes, but studies require independent attempts at reproducing the same results to verify that they weren't produced by chance. Suppose a medical researcher designs and conducts the following study to test the impact eating 2 chicken eggs a day has on LDL cholesterol (low-density lipoprotein cholesterol (the bad cholesterol)) levels in the body.

Given the varying conclusions of the previous medical research and the number of confounding variables that cloud their results, this particular researcher decides to test whether or not there is any effect in adopting the consumption of 2 chicken eggs a day and will test at a significance level of 0.05.

Participants are randomly sampled from the population at large. Each participant is asked to abstain from eating chicken eggs for the span of 3 months to normalize the sample to a diet without chicken eggs. Each participant's LDL cholesterol is measured the morning after completing the 3 month normalization period and is expected to have been fasting from midnight the night before. The participants then eat 2 chicken eggs scrambled using a teaspoon of olive oil each day for breakfast for an entire month. Participants are expected to maintain their regular diet otherwise. At the end of a month, participants again have their LDL cholesterol measured in the same fashion as before.

Table 7.3.1: Initial and Final LDL Cholesterol Readings

Participant #	Initial LDL (mg/dL)	Final LDL (mg/dL)
1	189	189
2	110	101
3	155	158
4	97	94
5	83	73
6	75	73
7	182	189
8	177	180
9	160	151
10	185	184
11	72	72
12	169	171
13	87	86
14	112	118
15	112	118
16	107	104
17	168	174
18	190	194
19	120	126
20	122	125
21	175	167
22	168	178
23	106	104

Participant #	Initial LDL (mg/dL)	Final LDL (mg/dL)
24	108	110
25	93	99
26	129	139
27	95	94
28	63	68
29	176	170
30	186	191
31	171	175
32	154	154
33	78	76
34	156	164
35	170	160

To analyze the results of this hypothetical medical study (the results were fabricated for the purposes of the book), we treat the two samples as dependent samples given that the variables of interest (LDL cholesterol levels before and after) can be matched by participant. We are interested in the change in cholesterol level after having the medical intervention of eating 2 scrambled chicken eggs a day for a month. To compute the change, we will need to compute the difference between the final measurement and the initial measurement, Final LDL – Initial LDL. A positive difference indicates that the LDL level increased; while, a negative difference indicates that the LDL level decreased. We will conduct our analyses on the values of these differences. To emphasize the fact that we are studying the differences of dependent paired variables, we will utilize the following notation for means and standard deviations: μ_d , \bar{x}_d , σ_d , and s_d .

With this notation in hand, let us formulate our hypotheses regarding the average value of these differences. The researcher wants to determine whether eating 2 chicken eggs a day has any effect on LDL levels. This would be an increase or decrease. If there is no effect, the average of the differences will be 0. If there is an effect, the average of the differences will not be 0. We adopt the former as our null hypothesis because chicken eggs are a relatively cheap source of protein and other nutrients that have been consumed consistently in larger quantities for a long time.

$$H_0 : \mu_d = 0 \text{ mg/dL}$$

$$H_1 : \mu_d \neq 0 \text{ mg/dL}$$

Having our hypotheses in hand, we compute the differences to analyze and ensure that we met the requirements necessary to conduct the hypothesis test. We have a random sample with a sample of 35 participants. Just like in our previous tests, we need either that the underlying distribution, the distribution of all these differences, is normal or that the sample is large enough for the Central Limit Theorem to assert that the sampling distribution of sample means is approximately normal. Since $n = 35$, we will proceed using the latter as our justification.

Table 7.3.2: Initial and Final LDL Cholesterol Readings with Differences

Participant #	Initial LDL (mg/dL)	Final LDL (mg/dL)	Difference (Final - Initial) (mg/dL)
1	189	189	0
2	110	101	–9
3	155	158	3
4	97	94	–3

Participant #	Initial LDL (mg/dL)	Final LDL (mg/dL)	Difference (Final - Initial) (mg/dL)
5	83	73	-10
6	75	73	-2
7	182	189	7
8	177	180	3
9	160	151	-9
10	185	184	-1
11	72	72	0
12	169	171	2
13	87	86	-1
14	112	118	6
15	112	118	6
16	107	104	-3
17	168	174	6
18	190	194	4
19	120	126	6
20	122	125	3
21	175	167	-8
22	168	178	10
23	106	104	-2
24	108	110	2
25	93	99	6
26	129	139	10
27	95	94	-1
28	63	68	5
29	176	170	-6
30	186	191	5
31	171	175	4
32	154	154	0
33	78	76	-2
34	156	164	8
35	170	160	-10

We do not know anything about the population parameters, so we will have to conduct our test using the t -transformation test statistic; hypotheses tests on dependent paired variables in this context are often referred to as a paired t -tests. We compute the sample mean and standard deviation using the difference values in the fourth column, compute the test statistic under the

assumption that the null hypothesis is true, and produce a visualization for computing the p -value. Notice that since the fourth column has 35 data points, we will use $n = 35$ in our computations, despite the fact that we recorded 70 values in total. $\bar{x}_d \approx 0.8286 \text{ mg/dL}$. $s_d \approx 5.6386 \text{ mg/dL}$.

$$t \approx \frac{0.8286 - 0}{\frac{5.6386}{\sqrt{35}}} \approx 0.8694.$$

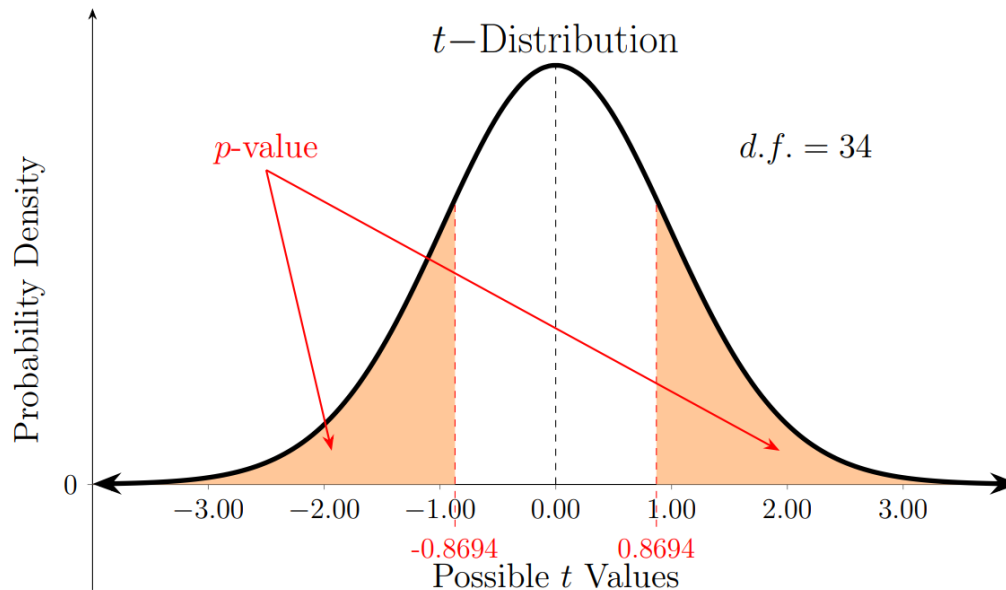


Figure 7.3.1: t -distribution for LDL Cholesterol Readings

$$p\text{-value} \approx 2 \cdot \text{T.DIST}(-0.8694, 34, 1) \approx 2 \cdot 0.1954 \approx 0.3908$$

Given that the p -value is greater than the α value, we fail to reject the null hypothesis. There is not sufficient evidence to say that eating 2 chicken eggs per day in the manner specified in the study alters the amount of LDL cholesterol in one's system over the course of a month.

? Text Exercise 7.3.1

An athletic training company executive officer recently discovered the [knees-over-toes guy](#), a trainer with a seemingly effective approach to living well through exercise focused on whole body movement, flexibility, and overall strength. The trainer claims that his approach helps people dunk basketballs. As this is an area of strategic growth for his company, the executive officer was enticed and decided to test the strategy on his basketball clients for a year to assess the growth in the height of the clients vertical jump. A random sample of 31 male clients was selected to participate in the study. Initial and final vertical jumps were measured in inches (see table below). Conduct the hypothesis test at a 0.02 significance level. Note that program is real, but this study is fabricated for the purposes of the book.

Table 7.3.3: Initial and Final Jump Height in Inches

Client #	Initial Jump Height (in)	Final Jump Height (in)
1	18	25
2	22	24
3	24	25
4	16	17
5	18	24

Client #	Initial Jump Height (in)	Final Jump Height (in)
6	22	24
7	24	28
8	26	29
9	15	20
10	18	19
11	14	20
12	16	16
13	24	31
14	16	21
15	24	32
16	24	26
17	24	25
18	14	22
19	18	20
20	23	29
21	15	22
22	18	19
23	25	32
24	18	23
25	17	20
26	16	19
27	22	23
28	20	22
29	25	29
30	24	24
31	14	14

Answer

We treat the two samples as dependent samples given that the variables of interest (vertical jump height) came from the same participant pool and we can match the values by participant. We are again interested in the change in the variable of interest after having some intervention; in this case, the intervention is a particular form of athletic training. To compute the change, we will need to compute the difference between the final measurement and the initial measurement. Again, a positive difference indicates that the intervention increased the jump height; while, a negative difference indicates that the jump height decreased. We will again conduct our analyses on the values of these differences.

The company officer will only be interested in the new program if it increases clients' jump heights. Increasing the jump height would result in a positive difference on average. The company officer does not want to assume that the program is effective without evidence; we, therefore, have the following hypotheses for our test.

$$H_0 : \mu_d \leq 0 \text{ in}$$

$$H_1 : \mu_d > 0 \text{ in}$$

Since the study used a random sample of 31 clients, the hypothesis test can be conducted. We compute the differences in the following table.

Table 7.3.4 Initial and Final Jump Height with Differences in Inches

Client #	Initial Jump Height (in)	Final Jump Height (in)	Difference (Final - Initial) (in)
1	18	25	7
2	22	24	2
3	24	25	1
4	16	17	1
5	18	24	6
6	22	24	2
7	24	28	4
8	26	29	3
9	15	20	5
10	18	19	1
11	14	20	6
12	16	16	0
13	24	31	7
14	16	21	5
15	24	32	8
16	24	26	2
17	24	25	1
18	14	22	8
19	18	20	2
20	23	29	6
21	15	22	7
22	18	19	1
23	25	32	7
24	18	23	5
25	17	20	3
26	16	19	3
27	22	23	1
28	20	22	2
29	25	29	4

Client #	Initial Jump Height (in)	Final Jump Height (in)	Difference (Final - Initial) (in)
30	24	24	0
31	14	14	0

We will again conduct a paired t -test. $\bar{x}_d \approx 3.5484$ inches. $s_d \approx 2.5928$ inches.

$$t \approx \frac{3.5484 - 0}{\frac{2.5928}{\sqrt{31}}} \approx 7.6198$$

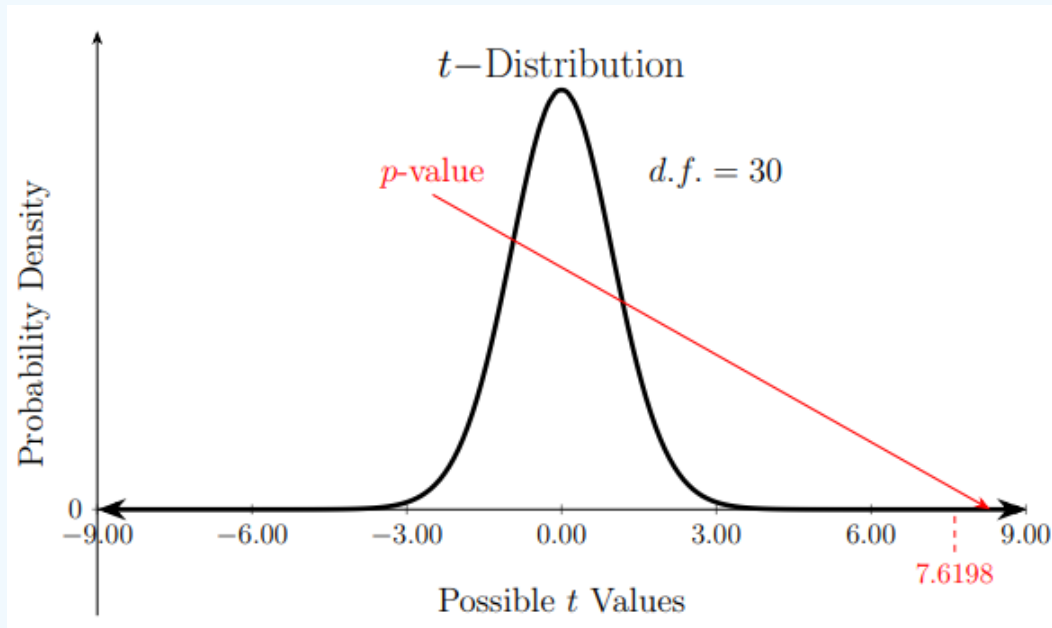


Figure 7.3.2 Right-tailed test with $t = 7.6198$

$$p\text{-value} \approx 1 - \text{T.DIST}(7.6198, 30, 1) \approx 8.4615 \cdot 10^{-9}$$

Given that the p -value is less than the α value, we reject the null hypothesis. There is sufficient evidence to say that over the course of a year using the knees-over-toes guy's training regiment the average height of clients' vertical jumps increased.

? Text Exercise 7.3.2

Many people have been concerned with carbon emissions from automobiles. Various governments have enacted policies that set emission standards and goals for new cars. A government is giving automobile manufacturers 10 years to reach the emission standards, but each year the manufacturers have to show that progress has been made by reducing carbon dioxide emissions across updated models within each class of vehicles in the amount of at least 10 grams of carbon dioxide per mile driven.

An automobile manufacturer's analyses indicate that they will not meet the emission progression threshold for their four-door sedans. They are aware of certain studies that state that the fuels with higher ethanol concentrations produce less emissions. It happens that the motors in this class of cars work well with pure gasoline and gasoline blended with ethanol. Without the time to redesign enough models to meet the progression requirements, the company considers selling their four-door sedans as requiring gasoline blended with a high ethanol concentration. They are hoping the difference from the fuel will be enough to satisfy the requirements. With all the varieties in models, the company makes over 600 different four-door sedans. They

randomly select 31 models to test the carbon dioxide emissions and then compare the results to the results of the previous year. The results are presented in the table below. Test the hypothesis at the 0.05 significance level.

Table 7.3.5: Four-Door Sedan Emissions

Model of Four-Door Sedan #	Emissions from Last Year (g/mi)	Emissions from This Year with Blend (g/mi)
1	483	471
2	468	456
3	409	401
4	457	452
5	461	447
6	403	396
7	408	396
8	414	398
9	429	422
10	443	428
11	467	460
12	386	369
13	350	343
14	396	381
15	476	461
16	363	347
17	465	453
18	398	392
19	426	417
20	489	472
21	454	444
22	449	442
23	400	387
24	380	365
25	383	378
26	371	357
27	423	406
28	437	423
29	379	374
30	351	338
31	397	385

Answer

We treat the two samples as dependent samples given that the variables of interest (carbon dioxide emissions per mile driven) are paired by particular models of four-door sedans. We are again interested in the change in the variable of interest after having some intervention; in this case, blended fuel. To compute the change, we will need to compute the difference between the final measurement and the initial measurement. Again, a positive difference indicates that the intervention increased emission rates; while, a negative difference indicates a decrease in emission rates. We will again conduct our analyses on the values of these differences.

The company will only be interested if switching fuel specifications decreases carbon dioxide emission by at least 10 grams per mile driven on average. The company does not want to assume that this is the case without evidence. We form the following hypotheses.

$$H_0 : \mu_d \geq -10 \text{ g/mi}$$

$$H_1 : \mu_d < -10 \text{ g/mi}$$

Since the study used a random sample of 31 models of four-door sedans, the hypothesis test can be conducted. We compute the differences in the following table.

Table 7.3.6 Four-Door Sedan Emissions with Differences

Model of Four-Door Sedan #	Emissions from Last Year (g/mi)	Emissions from This year with Blend (g/mi)	Difference (g/mi)
1	483	471	-12
2	468	456	-12
3	409	401	-8
4	457	452	-5
5	461	447	-14
6	403	396	-7
7	408	396	-12
8	414	398	-16
9	429	422	-7
10	443	428	-15
11	467	460	-7
12	386	369	-17
13	350	343	-7
14	396	381	-15
15	476	461	-15
16	363	347	-16
17	465	453	-12
18	398	392	-6
19	426	417	-9
20	489	472	-17
21	454	444	-10
22	449	442	-7

Model of Four-Door Sedan #	Emissions from Last Year (g/mi)	Emissions from This year with Blend (g/mi)	Difference (g/mi)
23	400	387	-13
24	380	365	-15
25	383	378	-5
26	371	357	-14
27	423	406	-17
28	437	423	-14
29	379	374	-5
30	351	338	-13
31	397	385	-12

We will again conduct a paired t -test. $\bar{x}_d \approx -11.4194\text{g/mi}$. $s_d \approx 4.0148\text{g/mi}$.

$$t \approx \frac{-11.4194 - (-10)}{\frac{4.0148}{\sqrt{31}}} \approx -1.9684$$

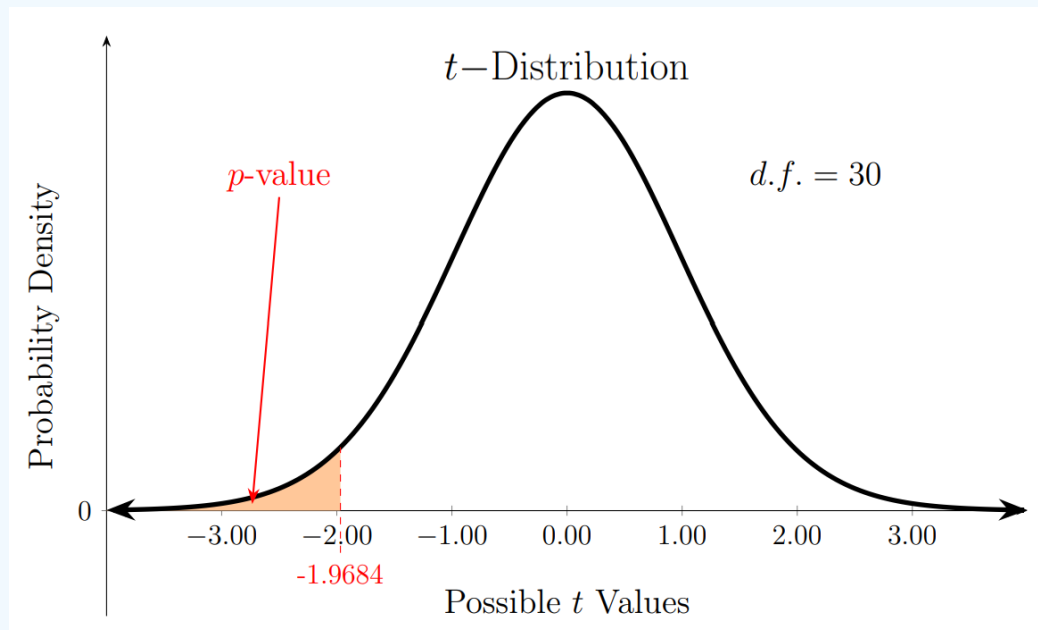


Figure 7.3.3 Left-tailed test with $t = -1.9684$

$$p\text{-value} \approx \text{T.DIST}(-1.9684, 30, 1) \approx 0.0292$$

Given that the p -value is less than the α value, we reject the null hypothesis. There is sufficient evidence to say that switching the fuel classification of the company's four-door sedans to requiring a ethanol-gasoline blended fuel with high concentrations of ethanol will allow the company to meet the emission progress standards set by the government. The progress may not reflect the intent of the law but seems to pass the letter of the law.

- **12.1: Testing a Single Mean** by [David Lane](#) is licensed [Public Domain](#). Original source: <https://onlinestatbook.com>.