

 Lagunita is retiring and will shut down at 12 noon Pacific Time on March 31, 2020. A few courses may be open for self-enrollment for a limited time. We will continue to offer courses on other online learning platforms; visit <http://online.stanford.edu>.

Course > Inference: Relationships $C \rightarrow Q$ > Matched Pairs > Matched Pairs: Example

 Bookmark this page

Matched Pairs: Example

Learning Objective: In a given context, carry out the inferential method for comparing groups and draw the appropriate conclusions.

Learning Objective: Specify the null and alternative hypotheses for comparing groups.

The "driving after having 2 beers" example is a case in which observations are paired by subject. In other words, both samples have the same subject, so that each subject is measured twice. Typically, as in our example, one of the measurements occurs before a treatment/intervention (2 beers in our case), and the other measurement after the treatment/intervention. Our next example is another typical type of study where the matched pairs design is used—it is a study involving twins.

Example

Researchers have long been interested in the extent to which intelligence, as measured by IQ score, is affected by "nurture" as opposed to "nature": that is, are people's IQ scores mainly a result of their upbringing and environment, or are they mainly an inherited trait? A study was designed to measure the effect of home environment on intelligence, or more specifically, the study was designed to address the question: "Are there significant differences in IQ scores between people who were raised by their birth parents, and those who were raised by someone else?"

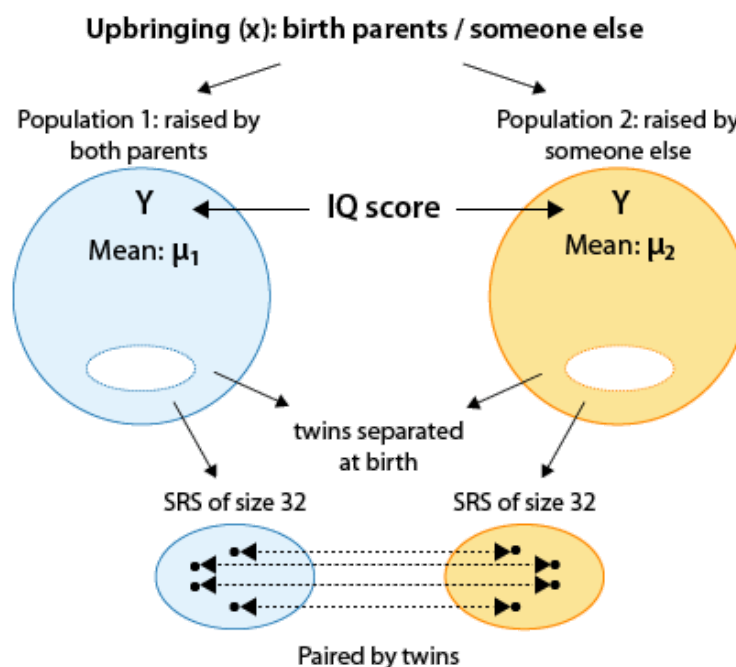
In order to be able to answer this question, the researchers needed to get two groups of subjects (one from the population of people who were raised by their birth parents, and one from the population of people who were raised by someone else) who are as similar as possible in all other respects. In

particular, since genetic differences may also affect intelligence, the researchers wanted to control for this confounding factor.

We know from our discussion on study design (in the Producing Data unit of the course) that one way to (at least theoretically) control for all confounding factors is randomization—randomizing subjects to the different treatment groups. In this case, however, this is not possible. This is an observational study; you cannot randomize children to either be raised by their birth parents or to be raised by someone else. How else can we eliminate the genetics factor? We can conduct a "twin study."

Because identical twins are genetically the same, a good design for obtaining information to answer this question would be to compare IQ scores for identical twins, one of whom is raised by birth parents and the other by someone else. Such a design (matched pairs) is an excellent way of making a comparison between individuals who only differ with respect to the explanatory variable of interest (upbringing) but are as alike as they can possibly be in all other important aspects (inborn intelligence). Identical twins raised apart were studied by Susan Farber, who published her studies in the book "Identical Twins Reared Apart" (1981, Basic Books). In this problem, we are going to use the data that appear in Farber's book in table E6, of the IQ scores of 32 pairs of identical twins who were reared apart.

Here is a figure that will help you understand this study:



Here are the important things to note in the figure:

1. We are essentially comparing the mean IQ scores in two populations that are defined by our (two-valued categorical) explanatory variable — upbringing (X), whose two values are: raised by birth parents, raised by someone else.

2. This is a matched pairs design (as opposed to a two independent samples design), since each observation in one sample is linked (matched) with an observation in the second sample. The observations are paired by twins.

Here is the data set:

| | Twin 1 | Twin 2 |
|----|--------|--------|
| 1 | 113 | 109 |
| 2 | 94 | 100 |
| 3 | 99 | 86 |
| 4 | 77 | 80 |
| 5 | 81 | 95 |
| 6 | 91 | 106 |
| 7 | 111 | 117 |
| 8 | 104 | 107 |
| 9 | 85 | 85 |
| 10 | 66 | 84 |
| 11 | 111 | 125 |
| 12 | 51 | 66 |
| 13 | 109 | 108 |
| 14 | 122 | 121 |
| 15 | 97 | 98 |
| 16 | 82 | 94 |
| 17 | 100 | 88 |
| 18 | 100 | 104 |
| 19 | 93 | 84 |
| 20 | 99 | 95 |
| 21 | 109 | 98 |
| 22 | 95 | 100 |

| | | |
|----|-----|-----|
| 23 | 75 | 86 |
| 24 | 104 | 103 |
| 25 | 73 | 78 |
| 26 | 88 | 99 |
| 27 | 92 | 111 |
| 28 | 108 | 110 |
| 29 | 88 | 83 |
| 30 | 90 | 82 |
| 31 | 79 | 76 |
| 32 | 97 | 98 |

Each of the 32 rows represents one pair of twins. Keeping the notation that we used above, twin 1 is the twin that was raised by his/her birth parents, and twin 2 is the twin that was raised by someone else. Let's carry out the analysis.

1. Stating the hypotheses.

Recall that in matched pairs, we reduce the data from two samples to one sample of differences:

| Pair | 1 | 2 | 3 | 4 | ... | 32 |
|---|-----|-----|----|----|-----|----|
| TWIN 1 (birth parents) | 113 | 94 | 99 | 77 | ... | 97 |
| TWIN 2 (someone else) | 109 | 100 | 86 | 80 | ... | 98 |
| Differences (twin 1 - twin 2) | 4 | -6 | 13 | -3 | ... | -1 |

and we state our hypotheses in terms of the mean of the differences, μ_d .

Since we would like to test whether there are differences in IQ scores between people who were raised by their birth parents and those who weren't, we are carrying out the two-sided test:

$$H_0 : \mu_d = 0$$

$$H_a : \mu_d \neq 0$$

Comment:

Again, some students find it easier to first think about the hypotheses in terms of μ_1 and μ_2 , and then write them in terms of μ_d . In this case, since we are testing for differences between the two populations, the hypotheses will be:

$$H_0 : \mu_1 - \mu_2 = 0$$

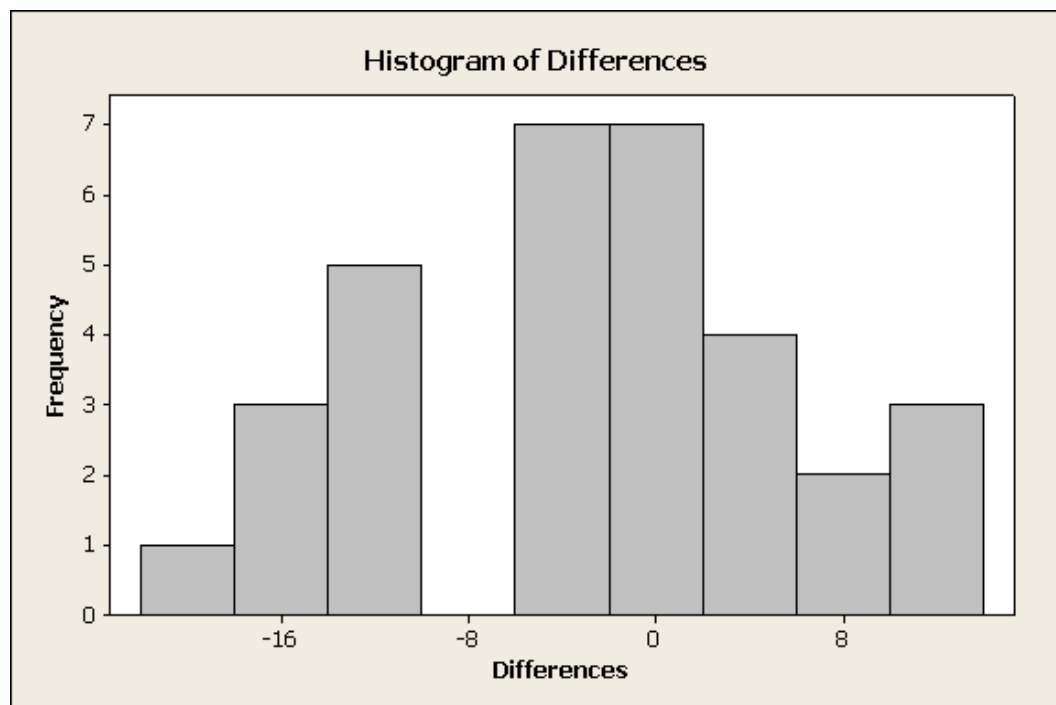
$$H_a : \mu_1 - \mu_2 \neq 0$$

and since $\mu_d = \mu_1 - \mu_2$ we get back to the hypotheses above.

2. Checking conditions and summarizing the data with a test statistic.

Is it safe to use the paired t-test in this case?

- a. Clearly, the samples of twins are not random samples from the two populations. However, in this context, they can be considered as random, assuming that there is nothing special about the IQ of a person just because he/she has an identical twin.
- b. The sample size here is $n = 32$. Even though it's the case that if we use the $n > 30$ rule of thumb our sample can be considered large, it is sort of a borderline case, so just to be on the safe side, we should look at the histogram of the differences just to make sure that we do not see anything extreme. (Comment: Looking at the histogram of differences in every case is useful even if the sample is very large, just in order to get a sense of the data. Recall: "Always look at the data.")



The data don't reveal anything that we should be worried about (like very extreme skewness or outliers), so we can safely proceed. Looking at the histogram, we note that most of the differences are negative, indicating that in most of the 32 pairs of twins, twin 2 (raised by someone else) has a higher IQ.

From this point we rely on statistical software, and find that:

- $t\text{-value} = -1.85$
- $p\text{-value} = 0.074$

Our test statistic is -1.85. Our data (represented by the average of the differences) are 1.85 standard errors below the null hypothesis (represented by the null value 0).

3. Finding the p-value.

The p-value is 0.074, indicating that there is a 7.4% chance of obtaining data like those observed (or even more extreme) assuming that H_0 is true (i.e., assuming that there are no significant differences in IQ scores between people who were raised by their natural parents and those who weren't).

4. Making conclusions.

Using the conventional significance level (cut-off probability) of 0.05, our p-value is not small enough, and we therefore cannot reject H_0 . In other words, our data do not provide enough evidence to conclude that whether a person was raised by his/her natural parents has an impact on the person's intelligence (as measured by IQ scores).

Learn By Doing

1/1 point (graded)

As we saw, the p-value of the test $H_0: \mu_d = 0$ vs. $H_a: \mu_d \neq 0$ is 0.074. This means that if we had carried out the one-sided test $H_0: \mu_d = 0$ vs. $H_a: \mu_d < 0$, the p-value would have been:

☐ 0.074 (the same)

☐ 0.148 (double)

☒ 0.037 (half) ✓

Answer

Correct: Indeed, the p-value of the one-sided test is half the p-value of the two-sided test.

[Submit](#)

Comment

This means that if, based on prior knowledge, prior research, or just a hunch, we had wanted to test the hypothesis that the IQ level of people raised by their birth parents is lower, on average, than the IQ level of people who were raised by someone else, we would have rejected H_0 and accepted that hypothesis (at the 0.05 significance level, since $0.037 < 0.05$).

It should be stressed, though, that one should set the hypotheses before looking at the data. It would be ethically wrong to look at the histogram of differences, note that most of the differences are negative, and then decide to carry out the one-sided test that the data seem to support. This is known as "data snooping," and is considered to be a very bad statistical practice.

Open Learning Initiative [↗](#)



[↗](#) Unless otherwise noted this work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License [↗](#).

© All Rights Reserved