

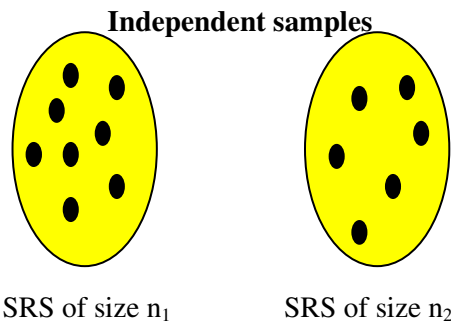
# Inference for Comparisons

## I. Using Inference for Comparisons: Two Population Means; Matched Pairs

Often researchers wish to make a comparison between two groups or assess the difference in two treatments or situations.

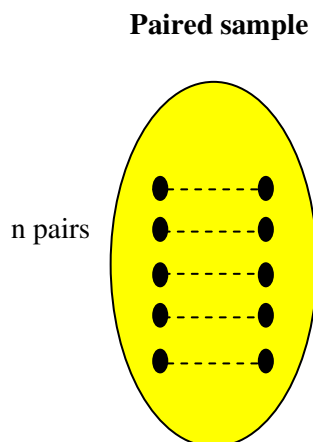
One experimental design is to either take **two independent samples**, one from each population, or to randomly assign subjects into each of two treatment groups. In these situations, the number of individuals in each sample or treatment group is potentially different.

In contrast, an alternative design is to use a **matched paired pair sample**. Perhaps each subject is measured twice, once under each set of treatment conditions. Or each experimental unit may be a pair of individuals, such as siblings or twins.



**Examples:**

- comparing mpg performance between two types of cars
- testing the effects of a drug by giving it to one group of subjects and placebo to the control group
- comparing the resting heart rates of men and women
- comparing the length of mating season for two different bird species



**Examples:**

- comparing patient weight before and after hospitalization
- comparing fish species diversity in lakes before and after heavy metal contamination
- testing effects of sunscreen applied to one arm of each subject compared with a placebo applied to the other arm
- testing effects of smoking in a sample of patients, each of which is compared with a nonsmoker closely matched by age, weight, and ethnic background
- testing effects of socioeconomic condition on dietary preferences by comparing identical twins raised in separate adoptive families that differ in their socioeconomic conditions

### 1. Hypothesis testing for the difference between two populations means— independent samples

We want to find the answer to the question:

**Can the difference between the sample means be due to chance?**

**Steps to follow:**

- **State the claims:**

$H_0: \mu_1 = \mu_2$  which is equivalent to  $\mu_1 - \mu_2 = 0$

$$H_a: \begin{cases} \mu_1 \neq \mu_2 & \text{which is equivalent to } \mu_1 - \mu_2 \neq 0 \\ \mu_1 < \mu_2 & \text{which is equivalent to } \mu_1 - \mu_2 < 0 \\ \mu_1 > \mu_2 & \text{which is equivalent to } \mu_1 - \mu_2 > 0 \end{cases}$$

- **Check conditions:**

- Observations are from two independent simple random samples of size  $n_1$  and  $n_2$ .
- The data comes from distributions that are approximately normally distributed, or the sample sizes are fairly large

- **Find the sample statistic,  $\bar{x}_1 - \bar{x}_2$ , in this case.**

- **Draw the sampling distribution of the difference in sample means based on the assumption that  $H_0$  is true and mark the value of  $\bar{x}_1 - \bar{x}_2$ . Then shade the area of interest on the graph.**

- **Find the test statistic:** 
$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- **Find the p-value** by using your calculator. You can use your calculator to find the test statistic and the p-value at the same time: STAT→TESTS→4: 2-SampTTest

- **Write your conclusion in context.**

## II. Confidence interval for the difference between two population means (CI for $\mu_1 - \mu_2$ )

Goal: estimate  $\mu_1 - \mu_2$  using the point estimate  $\bar{x}_1 - \bar{x}_2$ . That is, give a range of plausible values for  $\mu_1 - \mu_2$ .

Using your calculator: STAT→TESTS→0: 2-SampTInt

For two-sided tests, the CI can be used for testing: if the CI contains 0, we cannot reject the null hypothesis  $H_0: \mu_1 - \mu_2 = 0$ .

1. Determine whether the samples would be independent or consist of matched pairs.

- The effectiveness of Zantac for treating heartburn is tested by measuring gastric acid secretion in a group of patients treated with Zantac and another group of patients given a placebo.
- The effectiveness of Zantac for treating heartburn is tested by measuring gastric acid secretion in patients before and after the drug treatment.
- Comparing vitamin content of bread, immediately after baking versus 3 days later (the same loaves are used on day one and 3 days later).
- The effectiveness of a flu vaccine is tested by treating one group of subjects with the vaccine while another group of subjects is given placebos.

- e. The effectiveness of a tartar control toothpaste is tested in an experiment in which one twin used the regular toothpaste, and the other twin the tartar control toothpaste.
- f. Average fuel efficiency for 2005 vehicles is 21 miles per gallon. Is average fuel efficiency higher in the new generation “green vehicles?”
- g. A survey is conducted of teens from inner city schools to estimate the proportion who have tried drugs. A similar survey is conducted of teens from suburban schools.
- h. A psychologist measures the response times of subjects under two stimuli; each subject is observed under both of the stimuli, in a random order.
- i. An agronomist compares the yields of two varieties of soybean by planting each variety in 10 separate plots of land (a total of 20 plots).
- j. Lung cancer patients admitted in a hospital over a 12 month period are each matched with a non-cancer patient by age, sex, and race. To determine whether or not smoking is a risk factor for lung cancer, it is noted for each patient if he or she is a smoker.
- k. An advertising agency has come up with two different TV commercials for a household detergent. To determine which one is more effective, a test is conducted in which a sample of 100 adults is randomly divided into two groups. Each group is shown a different commercial, and the people in the group are asked to score the commercial.

2. A study published in the *Journal of American Academy of Business, Cambridge* (March, 2002) examined whether guests’ perception of the quality of service at five-star hotels in Jamaica differed by gender. Hotel guests were randomly selected from the lobby and restaurant areas and asked to rate 10 service-related items (e.g. “the personal attention you received from our employees”). Each item was rated on a five-point scale (from 1 = “much worse than I expected” to 5 = “much better than I expected”), and the sum of the item ratings for each guest was determined. A summary of the guest scores is provided in the following table:

| Gender         | Sample size | Mean score | Standard deviation |
|----------------|-------------|------------|--------------------|
| <b>Males</b>   | 127         | 39.08      | 6.73               |
| <b>Females</b> | 114         | 37.79      | 6.94               |

It was suspected that in general, female guests are not as satisfied with the quality of service as male guests.

Carry out the steps of the hypothesis testing; state the hypothesis, check conditions, find the test statistic, and the p-value. Write your conclusion in context.

3. Researchers randomly assigned participants either a tall, thin “highball” glass or a short, wide “tumbler,” each of which held 355 ml. Participants were asked to pour a shot (1.5 oz = 44.3 ml) of liquor into their glass. Did the shape of glass make a difference in how much liquor they poured? Here are the summaries:

| Highball  |         | Tumbler   |         |
|-----------|---------|-----------|---------|
| n         | 99      | n         | 99      |
| $\bar{y}$ | 42.2 oz | $\bar{y}$ | 47.9 oz |
| s         | 16.2 oz | s         | 17.9 oz |

a. Carry out the steps of hypothesis testing for the difference in how much liquid they poured.

$$H_0: \mu_{\text{highball}} = \mu_{\text{tumbler}} \quad \text{or} \quad \mu_{\text{highball}} - \mu_{\text{tumbler}} = 0$$

$$H_a: \mu_{\text{highball}} \neq \mu_{\text{tumbler}} \quad \text{or} \quad \mu_{\text{highball}} - \mu_{\text{tumbler}} \neq 0$$

Conditions: two independent simple random samples, and both samples sizes are over 30, so conditions are satisfied.

Using STAT → TESTS → 4: 2-SampTTest

$$t = -2.35 \quad \text{p-value: } 0.02$$

**Conclusion:** at the 5%, and 10% significance level we can reject the null hypothesis since the p-value is less than 5%. That is, we can reject the claim that there is no difference between the amount of liquid they poured, and we can conclude that the shape of glass make a difference in how much liquor they poured.

b. The 90% CI for the difference in means is (-9.71, -1.69).

Using this interval and your results from part a, state a conclusion in context.

Since the claim value, 0, is not in the 90% CI, it's not a plausible value. Therefore, based on this result, again, we can reject the null hypothesis at the 10% level. Same conclusion as in the previous part.

**4. A study was carried out to investigate the effectiveness of a treatment. 1000 subjects participated in the study, with 500 being randomly assigned to the “treatment group” and the other 500 to the “control (or placebo) group”. A statistically significant difference was reported between the responses of the two groups ( $P < .005$ ). Thus, we can conclude that**

- there is a large difference between the effects of the treatment and the placebo.
- there is strong evidence that the treatment is very effective.
- there is strong evidence that there is some difference in effect between the treatment and the placebo.
- there is little evidence that the treatment has any effect.
- there is evidence of a strong treatment effect.

6. The Excellent Drug Company claims its aspirin tablets will relieve headaches faster than any other aspirin on the market. To determine whether Excellent's claim is valid, random samples of size 15 are chosen from aspirins made by Excellent and the Simple Drug Company. An aspirin is given to each of the 30 randomly selected persons suffering from headaches and the number of minutes required for each to recover from the headache is recorded.

A 5% significance level test is performed to determine whether Excellent's (E) aspirin cures headaches significantly faster than Simple's (S) aspirin.

The appropriate set of hypotheses to be tested is:

- (a)  $H_0: \mu_E - \mu_S = 0$        $H_a: \mu_E - \mu_S > 0$
- (b)  $H_0: \mu_E - \mu_S = 0$        $H_a: \mu_E - \mu_S \neq 0$
- (c)  $H_0: \mu_E - \mu_S = 0$        $H_a: \mu_E - \mu_S < 0$
- (d)  $H_0: \mu_E - \mu_S < 0$        $H_a: \mu_E - \mu_S = 0$
- (e)  $H_0: \mu_E - \mu_S > 0$        $H_a: \mu_E - \mu_S = 0$

## 2. Inference for Matched Pairs

Recall that in a matched pairs design, subjects are matched in pairs and each treatment is given to one subject in each pair. Another situation involving matched pairs is before-after observations on the same objects.

### I. Hypothesis testing for the a mean population difference within pairs

We want to find the answer to the question:

**Can the observed mean of the pair differences be due to chance?**

**Steps to follow:**

- **State the claims:**
  - $H_0: \mu_d = 0$   
 $H_a: \mu_d > 0$       or       $\mu_d < 0$  or       $\mu_d \neq 0$
- **Check conditions:**
  - Observations consist of two measurements on each subject from a simple random sample of size  $n$ .
  - The data comes from a distribution that is approximately normally distributed, or the sample size.
- **Find the sample statistic,  $\bar{x}_d$ , in this case. ( $\bar{x}_d$  is the mean of the differences)**
- **Draw the sampling distribution of the sample proportions based on the assumption that  $H_0$  is true and mark the value of  $\bar{x}_d$ . Then shade the area of interest on the graph.**

- **Find the test statistic:**  $t = \frac{\bar{x}_d - 0}{s_d / \sqrt{n}}$  where  $s_d$  is the standard deviation of the differences.
- **Find the p-value** by using your calculator. You can use your calculator to find the test statistic and the p-value at the same time: STAT→TESTS→2: TTest  
Enter the differences into one of the lists and use Data in TTest, or use STAT and enter  $\bar{x}_d$ , and  $s_d$ .
- **Write your conclusion in context.**

## II. Confidence Interval for a mean population difference within pairs (CI for $\mu_d$ )

Goal: estimate  $\mu_d$  using the point estimate  $\bar{x}_d$ . That is, give a range of plausible values for  $\mu_d$ .

Using your calculator: STAT→TESTS→8: TInterval

For two-sided tests, the CI can also be used for testing: if the CI contains the value, 0, we cannot reject the null hypothesis  $H_0: \mu_d = 0$ , at the corresponding significance level.

**For the following examples, assume that the distribution of the differences is approximately normal.**

**1. A new treatment for depression is under investigation. Patients are given a psychological test before and after treatment in order to determine the effectiveness of the treatment. Researchers hope to find a difference in test scores before and after treatment. A 90% CI for the population average difference in test scores (before-after) is given by (0.25, 3.75), which was computed based on a random sample of 22 patients.**

- What is the population parameter of interest?
- Based on the CI, what is the value of the sample statistic (point estimate)?
- Determine if the following statement is true or false.

*We are 90% confident that the mean difference in sample test scores (before-after) of patients undergoing this treatment lies in the interval (0.25, .375).*

**True                      False**

- State the hypotheses in symbols:

$H_0:$

$H_a:$

- Based on the confidence interval, at a 10% significance level, would 0 be considered a reasonable value of the population parameter?
- What would be your decision for a hypothesis test done at the 10% significance level?
- Would a 95% confidence interval be wider or narrower than the 90% confidence interval?
- What would the decision be for a hypothesis test done at the 5% level?
- Assume the  $p$ -value for a 2-sided test was 0.09, with a t-test statistic of 1.78.

**2. Automobile insurance appraisers examine cars that have been in accidents in order to assess the cost of repairs. An insurance executive is concerned that some appraisers may tend to make**

**higher assessments than others. In one experiment, 10 cars that had recently been in accidents were shown to two appraisers. Each appraiser assessed the estimated repair cost for each car.**

a. State the hypotheses in symbols and in words:

$$H_0: \mu_d = 0$$

$$H_a: \mu_d \neq 0$$

The null hypothesis states that the mean of the differences in estimated repair costs between appraiser 1 and 2 is zero.

The alternative hypothesis states the mean of the differences in estimated repair costs between appraiser 1 and 2 is not zero.

Data, and summary:

```

A1      A2  Diff: A1 - A2
2150  1900  2150 - 1900 = 250
 860   880   860 - 880 = -20
1140  1100  1140 - 1100 = 40
1510  1420  1510 - 1420 = 90
1390  1430  1390 - 1430 = -40
1250  1150  1250 - 1150 = 100
 940   910   940 - 910 = 30
1710  1580  1710 - 1580 = 130
1020   980  1020 - 980 = 40
1190  1270  1190 - 1270 = -80
    
```

Sum of the differences: 540

The mean of the differences: 540/10 = 54

| Difference | Mean of Diff. | S.D. of Diff. | DF | T-Stat | P-value | 95% Lower Limit | 95% Upper Limit |
|------------|---------------|---------------|----|--------|---------|-----------------|-----------------|
| A1 - A2    | 54            | 94.77         | 9  | 1.80   | 0.1051  | -13.8           | 121.8           |

b. Show how the value 54 was obtained. See above next to the data.

c. Is there evidence that the two appraisers give significantly different assessments, on the average? Use a 0.05 significance level. In your conclusion use both the given p-value and confidence interval.

Since the p-value > 0.05, we don't have enough evidence to reject the null hypothesis. We don't have enough evidence to conclude that the two appraisers give significantly different assessments on average.

**3. An herbal medicine is tested on 16 randomly selected patients with sleep disorders. Each patient's amount of sleep (in hours) is measured for one night with the herbal medicine and for one night without the herbal medicine. Researchers claim that the patients' condition will improve using the herbal medicine. At the 1% significance level, does the data support the researchers' claim?**

Patient Without With

|    |     |     |
|----|-----|-----|
| 1  | 1.8 | 3.0 |
| 2  | 2.0 | 3.6 |
| 3  | 3.4 | 4.0 |
| 4  | 3.5 | 4.4 |
| 5  | 3.7 | 4.5 |
| 6  | 3.8 | 5.2 |
| 7  | 3.9 | 5.5 |
| 8  | 3.9 | 5.7 |
| 9  | 4.0 | 6.2 |
| 10 | 4.9 | 6.3 |
| 11 | 5.1 | 6.6 |
| 12 | 5.2 | 7.8 |
| 13 | 5.0 | 7.2 |
| 14 | 4.5 | 6.5 |
| 15 | 4.2 | 5.6 |
| 16 | 4.7 | 5.9 |

$$\text{Mean (Without-With)} = -1.525$$

$$\text{S.d. (Without-With)} = 0.5422$$

$$H_0: \mu_d = 0$$

$H_a: \mu_d < 0$  (because the difference is defined as “without – with”, and we want to show that the patients sleep more with the herbal medicine. So we hope to show that “with” is greater than “without”, so the difference is negative.)

Conditions: The observations are paired, and the sample was a random sample, and we assume that the data comes from a distribution that is approximately normally distributed (see note before question 1).

Test statistic: using STAT→TESTS→2: TTest : Highlight Stats, and enter  $\mu_0 = 0$ ,  $\bar{x} = -1.525$ ,  $S_x = 0.5422$ ,  $n = 16$ ,  $\mu: < \mu_0$

$$t = -11.25 \quad \text{p-value: } 5.19 \times 10^{-9}$$

**Conclusion:** since the p-value is extremely small, smaller than 1%, we have strong evidence against the null hypothesis. The data support the researcher's claim that the herbal medicine improves the condition of patients with sleeping disorder.

**4. In order to test the effect of Prozac on the well-being of depressed individuals, a questionnaire was administered to nine patients to gauge their level of well-being. The questionnaire was given to each subject both before and after treatment with Prozac.**

|   | moodpre | moodpost | diffnce |
|---|---------|----------|---------|
| 1 | 3.00    | 5.00     | 2.00    |
| 2 | .00     | 1.00     | 1.00    |
| 3 | 6.00    | 5.00     | -1.00   |
| 4 | 7.00    | 7.00     | .00     |
| 5 | 4.00    | 10.00    | 6.00    |
| 6 | 3.00    | 9.00     | 6.00    |
| 7 | 2.00    | 7.00     | 5.00    |
| 8 | 1.00    | 11.00    | 10.00   |
| 9 | 4.00    | 8.00     | 4.00    |

**The data is shown below. Higher scores indicate greater well-being (that is, Prozac is having a positive effect).**

Perform a test of significance to determine whether this study gives good evidence that Prozac has a positive effect. Follow the steps below:

- a. Define the parameter we wish to test. (Hint: It will involve the population of differences, of which the data above is a sample.)

The parameter of interest is the mean of the differences between the scores of well-being before and after the treatment with Prozac.

- b. Give the null and alternative hypotheses using symbols:

$$H_0: \mu_d = 0$$

$H_a: \mu_d > 0$  (Because the difference in the table is defined as “moodpost – moodpre”, and we want to show that “moodpost” is higher, meaning that Prozac has a positive effect on mood.)

- c. What two things do we need to assume in order for the test to be valid?

We need to assume that the sample was randomly chosen, and that the difference between the scores is approximately normally distributed.

- d. The sample mean of the differences is 3.67, and the sample standard deviation of the differences is 3.5. Compute the test statistic.

Test statistic: using STAT → TESTS → 2: TTest : Highlight Stats, and enter  $\mu_0 = 0$ ,  $\bar{x} = 3.67$ ,  $S_x = 3.5$ ,  $n = 9$ ,  $\mu: > \mu_0$

$$t = 3.15$$

By hand: 
$$t = \frac{\bar{x}_d - 0}{s_d / \sqrt{n}} = \frac{3.67 - 0}{3.5 / \sqrt{9}} = 3.15$$

- e. One of the values below is the correct  $p$ -value for the test. Which one must it be? Explain.  
1.62            -0.016            0.007

Even if you don't have a TI-83 calculator, you should be able to pick this  $p$ -value. Why? Because  $p$ -value is a probability so it cannot be greater than 1 (so the first answer cannot be correct), and it cannot be negative (so the second answer cannot be correct either.)

- f. Based on the  $p$ -value you selected in the previous part, how strong is the evidence that Prozac has a positive effect?

Since p-value is about 0.7%, we have very strong evidence against the null hypothesis.

- g. Make a decision based on a 1% significance level.

Since the p-value < 1%, we have enough evidence to reject the null hypothesis. We have good evidence to conclude that Prozac has a positive effect on mood.

## II. Using Inference to Compare Two Population Proportions

**Parameter of interest: a difference between two population proportions,  $p_1 - p_2$**

### **I. Hypothesis testing for the difference between two populations proportions—independent samples**

We want to find the answer to the question:

**Can the observed difference between the sample proportions be due to chance?**

**Steps to follow:**

- **State the claims:**

$H_0: p_1 = p_2$  which is equivalent to  $p_1 - p_2 = 0$

$$H_a: \begin{cases} p_1 \neq p_2 & \text{which is equivalent to } p_1 - p_2 \neq 0 \\ p_1 < p_2 & \text{which is equivalent to } p_1 - p_2 < 0 \\ p_1 > p_2 & \text{which is equivalent to } p_1 - p_2 > 0 \end{cases}$$

- **Check conditions:**

- Observations are from two independent simple random samples of size  $n_1$  and  $n_2$ .
- We must have  $n_1 \hat{p}_1 \geq 10$  and  $n_1(1 - \hat{p}_1) \geq 10$ , and  $n_2 \hat{p}_2 \geq 10$  and  $n_2(1 - \hat{p}_2) \geq 10$

- **Find the sample statistic,  $\hat{p}_1 - \hat{p}_2$ , in this case.**
- **Draw the sampling distribution of the difference between two proportions based on the assumption that  $H_0$  is true and mark the value of  $\hat{p}_1 - \hat{p}_2$ . Then shade the area of interest on the graph.**

- **Find the test statistic:**

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\tilde{p}(1 - \tilde{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \quad \text{where } \tilde{p} \text{ is the } \textit{pooled sample proportion}, \text{ computed as } \tilde{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

- **Find the p-value** either by using the z-table or your calculator. You can use your calculator to find the test statistic and the p-value at the same time: STAT→TESTS→6: 2-PropZTest

- Write your conclusion in context.

## **II. Confidence Interval for the Difference Between Two Population Proportions (CI for $p_1 - p_2$ )**

Goal: estimate  $p_1 - p_2$  using the point estimate  $\hat{p}_1 - \hat{p}_2$ . That is, give a range of plausible values for  $p_1 - p_2$ .

Using your calculator: STAT→TESTS→B: 2-PropZInt

For two-sided tests, the CI can also be used for testing: if the CI contains the value, 0, we cannot reject the null hypothesis  $H_0: p_1 = p_2$ , at the corresponding significance level.

---

1. A public health researcher wants to know how two high schools, one in the inner city and one in the suburbs, differ in the percentage of students who smoke. A random survey of students gives the following results:

| <i>Population</i> | <i>n</i> | <i>Smokers</i> |
|-------------------|----------|----------------|
| 1 (inner-city)    | 125      | 47             |
| 2 (suburban)      | 153      | 52             |

- a. State the parameters of interest in words, and in symbols as well.
  - b. State the null and alternative hypotheses.
  - c. Are the conditions for inference satisfied?
  - d. The 90% confidence interval for the difference between the two population proportions is: (-0.059, 0.131). State your conclusion based on this interval.
  - d. The test statistic is  $z = 0.63$ . Explain what this tells us.
  - e. The p-value is 0.53. Based on this, state your conclusion.
3. In 1997 a random sample of 200 low income families was taken and it was found that 43 of them have children who have no health insurance. In 2003 a similar survey of 270 families was taken and 42 were found not to have insurance. Follow the steps below to determine whether there is evidence from these samples that the proportion of low income children without insurance has changed. Use a 10% significance level.
    - a. State the null and alternative hypotheses.

$$H_0: p_1 = p_2 \quad \text{or} \quad p_1 - p_2 = 0$$

$$H_0: p_1 \neq p_2 \quad \text{or} \quad p_1 - p_2 \neq 0$$

where  $p_1$  represents the proportion of ALL low income families where the children had no health insurance in 1997, and  $p_2$  represent the proportion of ALL low income families where the children had no health insurance in 2003.

- b. Are the conditions satisfied for inference?

Observations are from two independent simple random samples.

$$\hat{p}_1 = \frac{43}{200} = 0.215 \qquad \hat{p}_2 = \frac{42}{270} = 0.156$$

We have  $n_1\hat{p}_1 \geq 10$  and  $n_1(1 - \hat{p}_1) \geq 10$ , and  $n_2\hat{p}_2 \geq 10$  and  $n_2(1 - \hat{p}_2) \geq 10$

- c. State your conclusion in context using the results from the confidence interval and the z-test shown in the computer output below:

```
Test and CI for Two Proportions
Sample      X      N   Sample P
1997       43    200   0.2150
2003       42    270   0.1555

Difference = p(1997) - p(2003)
Estimate for difference: 0.3706
90% CI: (-0.0006, 0.1194)
Test for difference = 0 (vs ≠ 0):  z = 1.655    p-value: 0.0978
```

This is borderline case. The p-value is barely less than 10%, but based on that we would reject the null hypothesis. On the other hand, since 0 is inside the 90% confidence interval, based on that we would not reject the null hypothesis. But again, 0 is just barely in there.

All we can say here is that we have very little evidence against the null hypothesis. We have very little evidence against the claim that there is no difference between the proportions of low income family where children don't have health insurance in 1997 and 2003.