STA218
Inference about comparing two populations

Al Nosedal.
University of Toronto.
Fall 2018

November 15, 2018

Two-sample problems

- The goal of inference is to compare the responses to two treatments or to compare the characteristics of two populations.
- We have a separate sample from each treatment or each population.

Conditions for inference comparing two means

- We have two SRSs, from two distinct populations. The samples are independent. That is, one sample has no influence on the other. Matching violates independence, for example. We measure the same response variable for both samples.
- Both populations are Normally distributed. The means and standard deviations of the populations are unknown. In practice, it is enough that the distributions have similar shapes and that the data have no strong outliers.

The Two-Sample Procedures

Draw an SRS of size n_1 from a Normal population with unknown mean μ_1, and draw and independent SRS of size n_2 from another Normal population with unknown mean μ_2. A confidence interval for $\mu_1 - \mu_2$ is given by

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Here t^* is the critical value for the $t(k)$ density curve with area C between $-t^*$ and t^*. The degrees of freedom k are equal to the smaller of $n_1 - 1$ and $n_2 - 1$.

The Two-Sample Procedures

To test the hypothesis $H_0 : \mu_1 = \mu_2$, calculate the two-sample t statistic

$$t^* = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

and use P-values or critical values for the $t(k)$ distribution.

Degrees of freedom (Option 1)

Option 1. With software, use the statistic t with accurate critical values from the approximating t distribution.
The distribution of the two-sample t statistic is very close to the t distribution with degrees of freedom df given by

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{1}{n_1-1}\right)\left(\frac{s_1^2}{n_1}\right)^2 + \left(\frac{1}{n_2-1}\right)\left(\frac{s_2^2}{n_2}\right)^2}$$

This approximation is accurate when both sample sizes n_1 and n_2 are 5 or larger.

Degrees of freedom (Option2)

Option 2. Without software, use the statistic t with critical values from the t distribution with degrees of freedom equal to the smaller of $n_1 - 1$ and $n_2 - 1$. These procedures are always conservative for any two Normal populations.

Example

A company selects 22 sales trainees who are randomly divided into two experimental groups - one receives type A and the other type B training. The salespeople are then assigned and managed without regard to the training they have received. At the year's end, the manager reviews the performances of salespeople in these groups and finds the following results:

	A Group	B Group
Average Weekly Sales	$\bar{x}_1 = \$1,500$	$\bar{x}_2 = \$1,300$
Standard Deviation	$s_1 = \$225$	$s_2 = \$251$

Example

a) Set up the null and alternative hypotheses needed to attempt to establish that type A training results in higher mean weekly sales than does type B training.
b) Because different sales trainees are assigned to the two experimental groups, it is reasonable to believe that the two samples are independent. Assuming that the Normality assumption holds, test the hypotheses you set up in part a) at level of significance 0.05.

Solution

1. State hypotheses. $H_0 : \mu_1 = \mu_2$ vs $H_a : \mu_1 > \mu_2$, where μ_1 is the mean weekly sales for all individuals assigned to type A training and μ_2 is the mean weekly sales for all individuals assigned to type B training.
2. Test statistic. $t^* = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = 1.9678$ ($\bar{x}_1 = 1500$, $\bar{x}_2 = 1300$,
$s_1 = 225$, $s_2 = 251$, $n_1 = 11$ and $n_2 = 11$)
3. P-value. Using Table C, we have $df = 10$, and
$0.025 <$ P-value < 0.05.
4. Conclusion. Since P-value < 0.05, we reject H_0. There is strong evidence that type A training results in higher mean weekly sales than does type B training.

Calculate a 95 percent confidence interval for the difference between the mean weekly sales obtained when type A training is used and the mean weekly sales obtained when type B training is used.

Solution

1. Find $\bar{x}_1 - \bar{x}_2$. From what we did earlier:
$\bar{x}_1 - \bar{x}_2 = 1500 - 1300 = 200$
2. Find SE = Standard Error. We already know that: $s_1 = 225$, $s_2 = 251$, $n_1 = 11$ and $n_2 = 11$

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{225^2}{11} + \frac{251^2}{11}} = 101.6348$$

3. Find $m = t^* SE$. From Table C, we have $df = 10$ and 95% confidence level, then $t^* = 2.228$. Hence,
$m = (2.228)(101.6348) = 226.4423$.
4. Find Confidence Interval.
$\bar{x}_1 - \bar{x}_2 \pm t^* SE = 200 \pm 226.4423$ from -26.4423 to 426.4423.

Logging in the rain forest

"Conservationists have despaired over destruction of tropical rain forest by logging, clearing, and burning". These words begin a report on a statistical study of the effects of logging in Borneo. Here are data on the number of tree species in 12 unlogged forest plots and 9 similar plots logged 8 years earlier:
Unlogged: 22 18 22 20 15 21 13 13 19 13 19 15
Logged : 17 4 18 14 18 15 15 10 12
Does logging significantly reduce the mean number of species in a plot after 8 years? State the hypotheses and do a t test. Is the result significant at the 5% level?

Solution

Does logging significantly reduce the mean number of species in a plot after 8 years?

1. State hypotheses. $H_0 : \mu_1 = \mu_2$ vs $H_a : \mu_1 > \mu_2$, where μ_1 is the mean number of species in unlogged plots and μ_2 is the mean number of species in plots logged 8 years earlier.

2. Test statistic. $t^* = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = 2.1140$ ($\bar{x}_1 = 17.5$, $\bar{x}_2 = 13.6666$, $s_1 = 3.5290$, $s_2 = 4.5$, $n_1 = 12$ and $n_2 = 9$)

3. P-value. Using Table C, we have $df = 8$, and $0.025 <$ P-value < 0.05.

4. Conclusion. Since P-value < 0.05, we reject H_0. There is strong evidence that the mean number of species in unlogged plots is greater than that for logged plots 8 year after logging.

Logging in the rainforest, continued

Use the data from the previous exercise to give a 99% confidence interval for the difference in mean number of species between unlogged and logged plots.

Solution

1. Find $\bar{x}_1 - \bar{x}_2$. From what we did earlier:
$\bar{x}_1 - \bar{x}_2 = 17.5 - 13.6666 = 3.8334$
2. Find SE = Standard Error. We already know that: $s_1 = 3.5290$, $s_2 = 4.5$, $n_1 = 12$ and $n_2 = 9$

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{3.5290^2}{12} + \frac{4.5^2}{9}} = 1.8132$$

3. Find $m = t^* SE$. From Table C, we have $df = 8$ and 99% confidence level, then $t^* = 3.355$. Hence,
$m = (3.355)(1.8132) = 6.0832$.
4. Find Confidence Interval.
$\bar{x}_1 - \bar{x}_2 \pm t^* SE = 3.8334 \pm 6.0832$ from -2.2498 to 9.9166.

Exercise

In random samples of 25 from each of two Normal populations, we found the following statistics:
$\bar{x}_1 = 524$ and $s_1 = 129$
$\bar{x}_2 = 469$ and $s_2 = 141$
Estimate the difference between the two population means with 95% confidence.

Solution

1. Find $\bar{x}_1 - \bar{x}_2$. In this case: $\bar{x}_1 - \bar{x}_2 = 524 - 469 = 55$
2. Find SE = Standard Error. We already know that: $s_1 = 129$, $s_2 = 141$, $n_1 = 25$ and $n_2 = 25$

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{129^2}{25} + \frac{141^2}{25}} = 38.2215$$

3. Find $m = t^* SE$. From Table C, we have $df = 24$ and 95% confidence level, then $t^* = 2.064$. Hence, $m = (2.064)(38.2215) = 78.8892$.
4. Find Confidence Interval. $\bar{x}_1 - \bar{x}_2 \pm t^* SE = 55 \pm 78.8892$ from -23.8892 to 133.8892.

Exercise

In random samples of 12 from each of two Normal populations, we found the following statistics:
$\bar{x}_1 = 74$ and $s_1 = 18$
$\bar{x}_2 = 71$ and $s_2 = 16$
Test with $\alpha = 0.05$ to determine whether we can infer that the population means differ.

Solution

1. State hypotheses. $H_0 : \mu_1 = \mu_2$ vs $H_a : \mu_1 \neq \mu_2$.
2. Test statistic. $t^* = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = 0.4315$ ($\bar{x}_1 = 74$, $\bar{x}_2 = 71$,

$s_1 = 18$, $s_2 = 16$, $n_1 = 12$ and $n_2 = 12$)
3. P-value. Using Table C, we have $df = 11$, and P-value > 0.50.
4. Conclusion. Since P-value $> \alpha = 0.05$, we **can't** reject H_0.
There is not enough evidence to infer that the population means differ.

Matched pairs t procedures

To compare the responses to the two treatments in a matched pairs design, find the difference between the responses within each pair. Then apply the one-sample t procedures to these differences. A matched pairs design compares just two treatments. Choose pairs of subjects that are as closely matched as possible. Assign one of the treatments to one of the subjects in a pair by tossing a coin or reading odd and even digits from a table of random digits (or by generating them with a computer). The other subject gets the remaining treatment. Sometimes each "pair" in a matched pairs design consists of just one subject, who gets both treatments one after the other.

Example

A manufacturer wanted to compare the wearing qualities of two different types of automobile tires, A and B. In the comparison, a tire of type A and one of type B were randomly assigned and mounted on the rear wheels of each of five automobiles. The automobiles were then operated for a specified number of miles, and the amount of wear was recorded for each tire. These measurements appear in a table below. Do the data provide sufficient evidence to indicate a difference in mean wear for tire types A and B? Test using $\alpha = 0.05$.

Auto	1	2	3	4	5
Tire A	10.6	9.8	12.3	9.7	8.8
Tire B	10.2	9.4	11.8	9.1	8.3

Solution

You can verify that the mean and standard deviation of the five **difference** measurements are $\bar{d} = 0.48$ and $s_d = 0.0837$.

Step 1. State Hypotheses. $H_0 : \mu_d = 0$ vs $H_a : \mu_d \neq 0$.
Step 2. Find test statistic. $t^* = \frac{\bar{d} - 0}{s_d/\sqrt{n}} = \frac{0.48}{0.0837/\sqrt{5}} = 12.8$
Step 3. Compute P-value. Using Table C, $P - value < 0.001$
Step 4. Conclusion. Since $P - value < \alpha = 0.05$, we reject H_0.
There is ample evidence of a difference in the mean amount of wear for tire types A and B.

Example

Find a 95% confidence interval for $(\mu_A - \mu_B) = \mu_d$ using the data from our previous example.

Solution

A 95% confidence interval for the difference between the mean wear is

$$\bar{d} \pm t^* \frac{s_d}{\sqrt{n}}$$

$$0.48 \pm (2.776)\frac{0.0837}{\sqrt{5}}$$

$$0.48 \pm 0.1039$$

Exercise

In an effort to determine whether a new type of fertilizer is more effective than the type currently in use, researchers took 12 two-acre plots of land scattered throughout the county. Each plot was divided into two equal-size subplots, one of which was treated with the new fertilizer. Wheat was planted, and the crop yields were measured.

Exercise

Plot	1	2	3	4	5	6	7	8	9	10	11	12
Current	56	45	68	72	61	69	57	55	60	72	75	66
New	60	49	66	73	59	67	61	60	58	75	72	68

Exercise

a. Can we conclude at the 5% significance level that the new fertilizer is more effective than the current one?

b. Estimate with 95% confidence the difference in mean crop yields between the two fertilizers.

Solution a)

You can verify that the mean and standard deviation of the twelve **difference** measurements are $\bar{d}$ = new - current = 1 and $s_d = 3.0151$.

Step 1. State Hypotheses. $H_0 : \mu_d = 0$ vs $H_a : \mu_d > 0$.
Step 2. Find test statistic. $t^* = \frac{\bar{d} - 0}{s_d/\sqrt{n}} = \frac{1}{3.0151/\sqrt{12}} = 1.1489$
Step 3. Compute P-value. Using Table C (df=11), $0.10 < P - value < 0.15$. Exact P-value = 0.1375, using R.
Step 4. Conclusion. Since $P - value > \alpha = 0.05$, we **can't** reject H_0. There is not enough evidence to infer that the new fertilizer is better.

R Code

```
# Step 1. Entering data;

current=c(56, 45, 68, 72, 61, 69, 57, 55, 60, 72, 75, 66);

new=c(60, 49, 66, 73, 59, 67, 61, 60, 58, 75, 72, 68);

diff=new-current;

# Step 2. T test;

t.test(diff,alternative="greater");
```

R Code

```
##
##  One Sample t-test
##
## data:  diff
## t = 1.1489, df = 11, p-value = 0.1375
## alternative hypothesis: true mean is greater than 0
## 95 percent confidence interval:
##  -0.5631171        Inf
## sample estimates:
## mean of x
##         1
```

Solution b)

A 95% confidence interval for the difference between the mean crop yields between the two fertilizers is

$$\bar{d} \pm t^* \frac{s_d}{\sqrt{n}}$$

$$1 \pm (2.201) \frac{3.0151}{\sqrt{12}}$$

$$1 \pm 1.9157$$

R Code

```
# Finding CI;

 t.test(diff,conf.level=0.95);
```

R Code

```
##
##  One Sample t-test
##
## data:  diff
## t = 1.1489, df = 11, p-value = 0.275
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.9157117  2.9157117
## sample estimates:
## mean of x
##         1
```

Large-sample confidence interval for comparing two proportions

Draw an SRS of size n_1 from a population having proportion p_1 of successes and draw an independent SRS of size n_2 from another population having proportion p_2 of successes. When n_1 and n_2 are large, an approximate level C confidence interval for $p_1 - p_2$ is

$$(\hat{p}_1 - \hat{p}_2) \pm z^* SE$$

In this formula the standard error SE of $\hat{p}_1 - \hat{p}_2$ is

$$SE = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

and z^* is the critical value for the standard Normal density curve with area C between $-z^*$ and z^*.

Hypotheses Tests for a Proportion

To test the hypothesis $H_0 : p_1 = p_2$ first find the pooled proportion $\hat{p}$ of successes in both samples combined. Then compute the z_* statistic, $z_* = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

In terms of a variable Z having the standard Normal distribution, the approximate P-value for a test of H_0 against

$H_a : p_1 > p_2$: is : $P(Z > z_*)$

$H_a : p_1 < p_2$: is : $P(Z < z_*)$

$H_a : p_1 \neq p_2$: is : $2P(Z > |z_*|)$

Example

A hospital administrator suspects that the delinquency rate in the payment of hospital bills has increased over the past year. Hospital records show that the bills of 48 of 1284 persons admitted in the month of April have been delinquent for more than 90 days. This number compares with 34 of 1002 persons admitted during the same month one year ago. Do these data provide sufficient evidence to indicate an increase in the rate of delinquency in payments exceeding 90 days? Test using $\alpha = 0.10$.

Solution

Let p_1 and p_2 represent the proportions of all potential hospital admissions in April of this year and last year, respectively, that would have allowed their accounts to be delinquent for a period exceeding 90 days, and let $n_1 = 1284$ admissions this year and the $n_2 = 1002$ admissions last year represent independent random samples from these populations.

Solution

Step 1. State Hypotheses. $H_0 : p_1 = p_2$ vs $H_a : p_1 > p_2$
Step 2. Find test statistic. $\hat{p}_1 = \frac{48}{1284} = 0.0374$ and
$\hat{p}_2 = \frac{34}{1002} = 0.0339$

$\hat{p} = \frac{x_1+x_2}{n_1+n_2} = \frac{48+34}{1284+1002} = 0.0359$

$z_* = \frac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = 0.45$

Solution

Step 3. Compute P-value.
$P - value = P(Z > z^*) = P(Z > 0.45) = 1 - P(Z < 0.45) = 0.3264$

Step 4. Conclusion. Since $P - value > \alpha = 0.10$, we **cannot** reject the null hypothesis that $p_1 = p_2$. The data present insufficient evidence to indicate that the proportion of delinquent accounts in April of this year exceeds the corresponding proportion last year.

R Code

```
successes=c(48,34);

totals=c(1284,1002);

prop.test(successes,totals,alternative="greater",
correct=FALSE);
```

R Code

```
##
##  2-sample test for equality of proportions without conti
##  correction
##
## data:  successes out of totals
## X-squared = 0.19381, df = 1, p-value = 0.3299
## alternative hypothesis: greater
## 95 percent confidence interval:
##  -0.009368431  1.000000000
## sample estimates:
##     prop 1     prop 2
## 0.03738318 0.03393214
```

Exercise

These statistics were calculated from two random samples:
$\hat{p}_1 = 0.60 \quad n_1 = 225 \quad \hat{p}_2 = 0.56 \quad n_2 = 225.$
Calculate the P-value of a test to determine whether there is evidence to infer that the population proportions differ.

Exercise

After sampling from two binomial populations we found the following.
$\hat{p}_1 = 0.18 \quad n_1 = 100 \quad \hat{p}_2 = 0.22 \quad n_2 = 100.$
Estimate with 90% confidence the difference in population proportions.

R Code

```
successes=c(18, 22);

totals=c(100, 100);

prop.test(successes,totals, conf.level=0.90,
correct=FALSE);
```

R Code

```
##
##  2-sample test for equality of proportions without conti
##  correction
##
## data:  successes out of totals
## X-squared = 0.5, df = 1, p-value = 0.4795
## alternative hypothesis: two.sided
## 90 percent confidence interval:
##  -0.13293059  0.05293059
## sample estimates:
## prop 1 prop 2
##   0.18   0.22
```

Exercise

One hundred normal-weight people and 100 obese people were observed at several Chinese-food buffets. For each researchers recorded whether the diner used chopsticks or knife and fork. The table shown here was created.

	Normal Weight	Obese
Used chop sticks	26	7
Used knife and fork	74	93

Is there sufficient evidence at the 10% significance level to conclude that obese Chinese food eaters are less likely to use chop sticks?

Solution

Let p_1 represent the proportion of all Normal Weight Chinese food eaters that use chop sticks and p_2 represent the proportion of all obese Chinese food eaters that use chop sticks.

Solution

Step 1. State Hypotheses. $H_0 : p_1 = p_2$ vs $H_a : p_1 > p_2$
Step 2. Find test statistic. $\hat{p}_1 = \frac{26}{100} = 0.26$ and $\hat{p}_2 = \frac{7}{100} = 0.07$

$$\hat{p} = \frac{x_1+x_2}{n_1+n_2} = \frac{26+7}{100+100} = 0.165$$

$$z_* = \frac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = 3.6195 \approx 3.62$$

Solution

Step 3. Compute P-value.
$P - value = P(Z > z^*) = P(Z > 3.62) = 1 - P(Z < 3.62) < 1 - 0.9998 = 0.0002$ (you can find the exact P-value using R).

Step 4. Conclusion. Since $P - value < 0.0002 < \alpha = 0.10$, we **reject the null hypothesis** that $p_1 = p_2$. There is enough evidence to conclude that obese Chinese food eaters are less likely to use chop sticks.