

Lecture X - Inference with Two Sample Means

X. Inference with Two Sample Means

1. Logic of two-sample inference
2. Sampling distribution of mean differences
3. Independent-samples hypothesis testing
4. Non-independent-samples hypothesis testing

Logic of Two-Sample Inference

- In hypothesis testing with two sample means, we want to know how two groups differ with respect to some outcome of interest. Our research question with two-population hypothesis testing is this: Are these two samples drawn from a single population with mean μ , or are they subsets of two different populations with means μ_1 and μ_2 , respectively? Recall from our discussion of single-sample inference that there are two possible answers to this question.
 - Yes, our samples are drawn from a single population (with respect to the outcome we are interested in). Thus, the observed difference between the two sample means reflects a chance difference, a result of random sampling error.
 - No, our samples are drawn from two different populations. Thus, the observed difference between the two sample means reflects a systematic difference, a result of the fact that they actually comprise two different populations with respect to the outcome.

Sampling Distribution of Mean Differences

- Using probability theory, we can ask how likely it is that we would observe the difference between the two sample means if the population means are actually equal. We want to know if the difference is random or systematic. With this type of question, we resort to a new sampling or probability distribution, known as the sampling distribution of mean differences.
- In order to understand what a sampling distribution of mean differences is, let's conduct a thought experiment. Let's say that we draw a sample of males and a sample of females, ask them how many offenses they committed in the last year, calculate the mean number of offenses for each sample, and compute the difference between these two sample means ($\bar{x}_1 - \bar{x}_2$). Now, imagine that we continue to do this for an infinite number of samples, calculating a mean for each sample and taking the difference between the two means. This process results in a theoretical sampling distribution of differences between sample means. This distribution is defined by two parameters.
 - The mean of this sampling distribution is $\mu_{\bar{x}_1} - \mu_{\bar{x}_2}$, and is equal to the true difference between the population means, $\mu_1 - \mu_2$.
 - The standard deviation of this sampling distribution is $\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$, and is called the standard error of the difference between means, or the "standard error of the difference" for short.
- When we are considering the sampling distribution of mean differences, we may resort to an important theorem with which we are already familiar: the central limit theorem. That is, no matter the shape of the two population distributions, when the sample sizes become large (generally over 50 each) the sampling distribution of mean differences approximates a normal distribution. This theorem implies that we may conduct hypothesis tests for two population means.
 - Note that the general form for a standard score is the following:

$$z = \frac{\text{estimate-parameter}}{\text{standard deviation}}$$

We can use this general notation to construct the standard score for mean differences:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Since, under the null hypothesis, the population means are equivalent, we can remove the second term from the numerator. Thus, when both μ_1 and μ_2 are known, this leaves us with having to estimate:

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

Unfortunately, the population standard deviations will not always be known. This requires us to make an adjustment to the formula by substituting the sample standard deviations for the population standard deviations, and by adjusting for the “degrees of freedom” from each sample. In this case, the test statistic becomes:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}}}$$

Independent-Samples Hypothesis Testing

Let's pose a couple of research questions to practice hypothesis testing.

- Suppose we want to test the association between social disorganization and violent victimization. We gather data on 41 low-SES neighborhoods and find that the mean number of victimizations in the last year is 12.4 ($\sigma^2 = 18.49$). We also gather data on 21 high-SES neighborhoods and find that the mean number of victimizations is 5.2 ($\sigma^2 = 2.25$). Our research question is: Do people that live in low-SES neighborhoods have a different risk of victimization than people that live in high-SES neighborhoods? Use $\alpha = .01$.
 - Formally state hypotheses:* $H_1 : \mu_L \neq \mu_H \therefore \mu_L - \mu_H \neq 0$; $H_0 : \mu_L = \mu_H \therefore \mu_L - \mu_H = 0$
 - Obtain probability distribution:* z-distribution
 - Make decision rules:* $\alpha=.01$ (two-tailed), $z_{crit} =$, reject H_0 if $|TS|>$.
 - Calculate the test statistic:* $z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{12.4 - 5.2}{\sqrt{\frac{18.49}{41} + \frac{2.25}{21}}} = 9.638$
 - Make a decision about the null hypothesis:* Reject H_0 , low-SES and high-SES neighborhoods have significantly different rates of victimization.
- Prior research indicates that working youths are more heavily involved in delinquency, substance use, and other problem behaviors than non-working youths. We have a sample of 2243 working youths and find that the mean number of self-report problem behaviors is $\bar{x}_W=3.73$ ($s^2 = 12.95$). We also have a sample of 6679 non-working youths and find that the mean number of problem behaviors is $\bar{x}_N=2.22$ ($s^2 = 8.19$). Our research question is: Do working youths do engage in more problem behavior than non-working youths? Use $\alpha = .05$.
 - Formally state hypotheses:* $H_1 : \mu_W > \mu_N \therefore \mu_W - \mu_N > 0$; $H_0 : \mu_W = \mu_N \therefore \mu_W - \mu_N \leq 0$
 - Obtain probability distribution:* t-distribution, $df=8920$
 - Make decision rules:* $\alpha=.05$ (one-tailed), $t_{crit} = 1.645$, reject H_0 if $TS>1.645$.
 - Calculate the test statistic:* $t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1-1} + \frac{\sigma_2^2}{n_2-1}}} = \frac{3.73 - 2.22}{\sqrt{\frac{12.95}{2243-1} + \frac{8.19}{6679-1}}} = 18.045$
 - Make a decision about the null hypothesis:* Reject H_0 , working youths commit significantly more crimes than non-working youths.

- There is some suggestion that individuals who engage in more criminal behavior are, for a variety of reasons, also more likely to be victims of crime. You compare a sample of 10 self-report high-rate offenders with 10 self-report non-offenders on the number of times in the last five years they have been the victim of a personal crime. Use $\alpha = .10$ to test the hypothesis that a “criminal lifestyle” increases the likelihood of victimization.

x_C	x_C^2	x_N	x_N^2
5	25	3	9
8	64	5	25
10	100	2	4
8	64	8	64
6	36	2	4
5	25	4	16
2	4	3	9
7	49	1	1
3	9	0	0
5	25	0	0
$\sum x_C = 59$ $\bar{x}_C = 5.9$	$\sum x_C^2 = 401$ $s_C^2 = 5.29$	$\sum x_N = 28$ $\bar{x}_N = 2.8$	$\sum x_N^2 = 132$ $s_N^2 = 5.36$

1. *Formally state hypotheses:* $H_1 : \mu_C > \mu_N \therefore \mu_C - \mu_N > 0$; $H_0 : \mu_C = \mu_N \therefore \mu_C - \mu_N \leq 0$
2. *Obtain probability distribution:* t-distribution; $df = 10 + 10 - 2 = 18$
3. *Make decision rules:* $\alpha = .10$ (one-tailed), $t_{crit} = 1.330$, reject H_0 if $TS > 1.330$.
4. *Calculate the test statistic:* $t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1-1} + \frac{\sigma_2^2}{n_2-1}}} = \frac{5.9 - 2.8}{\sqrt{\frac{5.29}{10-1} + \frac{5.36}{10-1}}} = 2.850$
5. *Make a decision about the null hypothesis:* Reject H_0 , people who engage in crime at a high rate are significantly more likely to have been victimized.

* You collect data from a sample of 8 youths that report having gone out on a date in the last year, and a sample of 6 youths that report not having gone out on a date. Each youth provides information on their self-report engagement in a variety of problem behaviors. Using $\alpha = .01$, test the hypothesis that dating increases delinquency.

x_N	x_N^2	x_D	x_D^2
1	1	3	9
3	9	1	1
2	4	5	25
3	9	2	4
0	0	1	1
1	1	0	0
		3	9
		4	16
$\sum x_N = 10$ $\bar{x}_N = 1.67$	$\sum x_N^2 = 24$ $s_N^2 = 1.21$	$\sum x_D = 19$ $\bar{x}_D = 2.38$	$\sum x_D^2 = 65$ $s_D^2 = 2.46$

1. *Formally state hypotheses:* $H_1 : \mu_N < \mu_D \therefore \mu_N - \mu_D < 0$; $H_0 : \mu_N = \mu_D \therefore \mu_N - \mu_D \geq 0$
2. *Obtain probability distribution:* t-distribution; $df = 6 + 8 - 2 = 12$
3. *Make decision rules:* $\alpha = .01$ (one-tailed), $t_{crit} = 2.681$, reject H_0 if $TS < -2.681$.
4. *Calculate the test statistic:* $t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1-1} + \frac{\sigma_2^2}{n_2-1}}} = \frac{1.67 - 2.38}{\sqrt{\frac{1.21}{6-1} + \frac{2.46}{8-1}}} = -0.922$
5. *Make a decision about the null hypothesis:* Accept H_0 , youth who date do not engage in more problem behaviors than youth who do not date.

Non-Independent Samples Hypothesis Testing

- There is one very important assumption that must be met in order to carry out the hypothesis tests that we just completed. Our two samples must be independent of one another. In other words, both samples are randomly selected, and the fact that one case is selected into one sample does not affect the selection of a different case into the other sample. From our example, we know that selecting a worker into one sample had no effect on the probability of selecting a non-worker into the other sample. Thus, our samples of workers and non-workers were each independent of the other.
- There are times when this condition is deliberately violated. The most common is when we have a “before-and-after” design, in which we measure the dependent variable for each case at two points in time. During these two measurement periods, we typically introduce some sort of treatment, which we think of as the independent variable. What we are interested in finding out is whether the treatment had any effect on the outcome.
- Another instance in which we have non-independent samples is when we match individuals from two samples on important characteristics, such as when we match treatment and control groups on the basis of age, gender, and race.
- With this type of design, we must reformulate our statistical test to be based on the difference between the scores for each pair of observations. Thus, unlike the independent-samples hypothesis test, in which we test the difference between two sample means (i.e., $\bar{x}_1 - \bar{x}_2$), with a non-independent-samples hypothesis test, we test the difference between the pair of raw scores for each individual (i.e., $x_1 - x_2$). The sample size is not the total number of observations in each group, but rather the number of pairs of observations.
- The question we want to ask with a non-independent-samples hypothesis test is: Did the treatment have an effect on the outcome, or do the matched groups differ with respect to the outcome? We want to use a formula that takes into account that we are computing differences between raw scores. When we do this, we obtain:

$$TS = \frac{\bar{x}_D - \mu_D}{\sqrt{s_D^2/n - 1}} = \frac{\bar{x}_D - \mu_D}{s_D/\sqrt{n - 1}}$$

However, under the null hypothesis, we assume the treatment had no effect, which means the $\mu_D = 0$. This means that we can remove this term from the formula, which leaves us with:

$$TS = \frac{\bar{x}_D - \mu_D}{s_D/\sqrt{n - 1}} \text{ where } \bar{x}_D = \frac{\sum x_2 - x_1}{n} \text{ and } s_D = \sqrt{\frac{\sum (x_2 - x_1)^2}{n}} = \sqrt{\frac{\sum x_D^2}{n} - \bar{x}_D^2}$$

Let's work through a couple of examples.

- There is evidence that children with criminal parents are more likely to be criminal themselves than children with non-criminal parents. You collect data on a sample of 121 youths who have at least one parent with an official record of conviction, and match them on the basis of age, sex, race, and school performance with 121 youths whose parents do not have a record of conviction. You find that the mean difference ($x_D = x_2 - x_1 = x_C - x_N$) between these matched groups is 5.75 ($s_D = 4.25$). Your research question is as follows: Are youths with criminal parents more likely to be arrested than youths without criminal parents? Use $\alpha = .01$.

1. *Formally state hypotheses:* $H_1 : \mu_D < 0$; $H_0 : \mu_D \geq 0$
2. *Obtain probability distribution:* t-distribution; $df = 121 - 1 = 120$
3. *Make decision rules:* $\alpha = .01$ (one-tailed), $t_{crit} = 2.358$, reject H_0 if $TS > 2.358$.
4. *Calculate the test statistic:* $t = \frac{\bar{x}_D - \mu_D}{s_D / \sqrt{n - 1}} = \frac{5.75}{4.25 / \sqrt{121 - 1}} = 14.821$
5. *Make a decision about the null hypothesis:* Reject H_0 , having criminal parents significantly increases the likelihood of being arrested.

- We want to know if formal sanctions deter crime. We draw a sample of $n=10$ offenders and ask them the number of arrests they had prior to appearing in court. We then ask them the number of arrests they had after they appeared in court. Use $\alpha = .05$ to test the hypothesis that formal sanctions deter crime.

x_1	x_2	$x_D = x_2 - x_1$	x_D^2
3	1	-2	4
2	2	0	0
1	0	-1	1
1	1	0	0
2	1	-1	1
4	4	0	0
6	6	0	0
1	0	-1	1
1	1	0	0
2	3	-1	1
$\bar{x}_1 = 2.2$	$\bar{x}_2 = 1.8$	$\sum x_D = -4$	$\sum x_D^2 = 8$
		$\bar{x}_D = -0.4$	$s_D = \sqrt{\frac{8}{10} - .40^2} = 0.80$

1. *Formally state hypotheses:* $H_1 : \mu_D < 0$; $H_0 : \mu_D \geq 0$
2. *Obtain probability distribution:* t-distribution; $df = 10 - 1 = 9$
3. *Make decision rules:* $\alpha = .05$ (one-tailed), $t_{crit} = 1.833$, reject H_0 if $TS < -1.833$.
4. *Calculate the test statistic:* $t = \frac{\bar{x}_D - \mu_D}{s_D / \sqrt{n - 1}} = \frac{-0.40}{.80 / \sqrt{10 - 1}} = -1.500$
5. *Make a decision about the null hypothesis:* Accept H_0 , formal adjudication does not appear to deter crime.

- Suppose that we have a sample of five high-risk middle-school students with a history of poor academic performance. In order to help them improve, the school implements an after-school tutoring program. Their test scores before and after participation in the program are listed below. Use $\alpha = .05$ to test the hypothesis that the program is effective in raising test scores.

x_1	x_2	$x_D = x_2 - x_1$	x_D^2
60	62	2	4
66	66	0	0
65	67	2	4
60	64	4	16
59	65	6	36
$\bar{x}_1 = 62$	$\bar{x}_2 = 64.8$	$\sum x_D = 14$	$\sum x_D^2 = 60$
		$\bar{x}_D = 2.8$	$s_D = \sqrt{\frac{60}{5} - 2.8^2} = 2.04$

1. *Formally state hypotheses:* $H_1 : \mu_D > 0$; $H_0 : \mu_D \leq 0$
2. *Obtain probability distribution:* t-distribution; $df = 5 - 1 = 4$
3. *Make decision rules:* $\alpha = .05$ (one-tailed), $t_{crit} = 2.132$, reject H_0 if $TS > 2.132$.
4. *Calculate the test statistic:* $t = \frac{\bar{x}_D - \mu_D}{s_D / \sqrt{n - 1}} = \frac{2.8}{2.04 / \sqrt{5 - 1}} = 2.745$
5. *Make a decision about the null hypothesis:* Reject H_0 , the tutoring program significantly raised test scores.

- We saw earlier that youths who are employed are more likely to be delinquent than youths that are not employed. How about if we consider the same group of youths at two points in time, both before and after the transition to employment (the “treatment”)? If youth employment is criminogenic, we would expect that youths who begin should also increase their involvement in problem behavior. We collect data on a sample of 15 high-school students who at time one were non-workers, and who by time two were working part time. Use $\alpha = .05$ to test the hypothesis that the transition to work is criminogenic.

x_1	x_2	$x_D = x_2 - x_1$	x_D^2
4	5	1	1
0	0	0	0
1	2	1	1
1	1	0	0
3	1	-2	4
5	4	-1	1
2	3	1	1
0	2	2	4
1	3	2	4
3	2	-1	1
2	1	-1	1
3	5	2	4
5	5	0	0
3	4	1	1
2	5	3	9
$\bar{x}_1 = 2.33$	$\bar{x}_2 = 2.87$	$\sum x_D = 8$	$\sum x_D^2 = 32$
		$\bar{x}_D = 0.53$	$s_D = \sqrt{\frac{32}{15} - .53^2} = 1.36$

1. *Formally state hypotheses:* $H_1 : \mu_D > 0$; $H_0 : \mu_D \leq 0$
2. *Obtain probability distribution:* t-distribution; $df = 15 - 1 = 14$
3. *Make decision rules:* $\alpha = .05$ (one-tailed), $t_{crit} = 1.761$, reject H_0 if $TS > 1.761$.
4. *Calculate the test statistic:* $t = \frac{\bar{x}_D - \mu_D}{s_D / \sqrt{n - 1}} = \frac{.53}{1.36 / \sqrt{15 - 1}} = 1.458$
5. *Make a decision about the null hypothesis:* Accept H_0 , youth employment is not criminogenic.