Mathematical statistics

May 10th, 2019

Lecture 31: The paired t-test

Att.

- Final exam:

  Wednesday, 5/29/2019, 10:30am –12:30pm

  Ewing Hall Room 101

- Course evaluation
- Last homework due this Friday
- Department colloquium

  GORE 104

Overview

Week 1	Probability reviews
Week 2	Chapter 6: Statistics and Sampling Distributions
Week 4	Chapter 7: Point Estimation
Week 7	Chapter 8: Confidence Intervals
Week 10	Chapter 9, 10: Test of Hypothesis

Inferences based on two samples

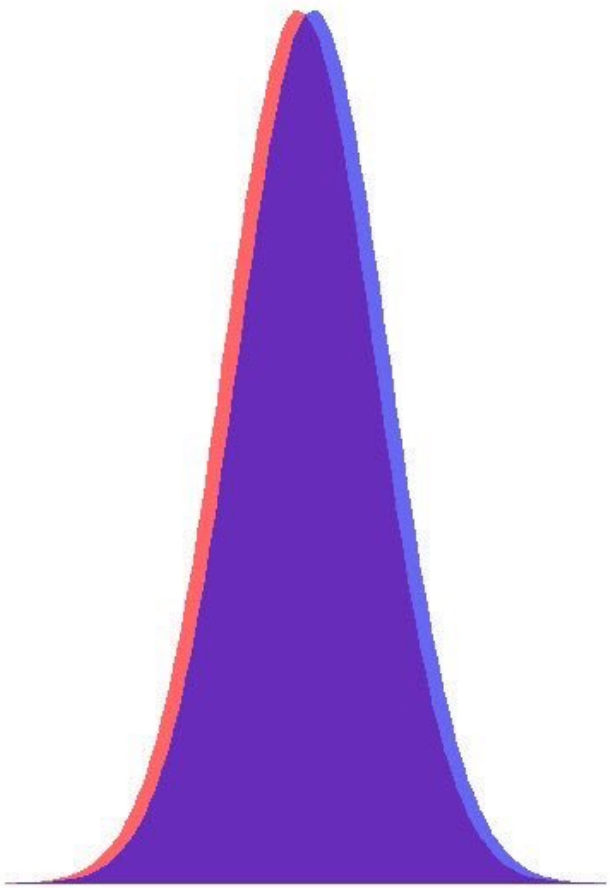
10.1 Difference between two population means

- z-test
- confidence intervals

10.2 The two-sample t test and confidence interval

10.3 Analysis of paired data

Difference between two population means



- Testing:

$$H_0 : \mu_1 = \mu_2$$
$$H_a : \mu_1 > \mu_2$$

- Works well, even if $|\mu_1 - \mu_2| << \sigma_1, \sigma_2$

Settings

- Independent samples
  1. $X_1, X_2, \ldots, X_m$ is a random sample from a population with mean μ_1 and variance σ_1^2.
  2. $Y_1, Y_2, \ldots, Y_n$ is a random sample from a population with mean μ_2 and variance σ_2^2.
  3. The X and Y samples are independent of each other.
- Paired samples
  1. There is only one set of n individuals or experimental objects
  2. Two observations are made on each individual or object

Remember Chapter 8?

- Section 8.1
  - Normal distribution
  - σ is known
- Section 8.2
  - ~~Normal distribution~~
    $\rightarrow$ Using Central Limit Theorem $\rightarrow$ needs $n > 30$
  - ~~σ is known~~
    $\rightarrow$ needs $n > 40$
- Section 8.3
  - Normal distribution
  - ~~σ is known~~
  - n is small

$\rightarrow$ Introducing t-distribution

Confidence intervals

When both population distributions are normal, standardizing $\overline{X} - \overline{Y}$ gives a random variable Z with a standard normal distribution. Since the area under the z curve between $-z_{\alpha/2}$ and $z_{\alpha/2}$ is $1 - \alpha$, it follows that

$$P\left(-z_{\alpha/2} < \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} < z_{\alpha/2}\right) = 1 - \alpha$$

Manipulation of the inequalities inside the parentheses to isolate $\mu_1 - \mu_2$ yields the equivalent probability statement

$$P\left(\overline{X} - \overline{Y} - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}} < \mu_1 - \mu_2 < \overline{X} - \overline{Y} + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}\right) = 1 - \alpha$$

Testing the difference between two population means

- Setting: independent normal random samples $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ with known values of σ_1 and σ_2. Constant Δ_0.
- Null hypothesis:

$$H_0 : \mu_1 - \mu_2 = \Delta_0$$

- Alternative hypothesis:
  (a) $H_a : \mu_1 - \mu_2 > \Delta_0$
  (b) $H_a : \mu_1 - \mu_2 < \Delta_0$
  (c) $H_a : \mu_1 - \mu_2 \neq \Delta_0$
- When $\Delta = 0$, the test (c) becomes

$$H_0 : \mu_1 = \mu_2$$
$$H_a : \mu_1 \neq \mu_2$$

Testing the difference between two population means

Proposition

Null hypothesis: H_0: $\mu_1 - \mu_2 = \Delta_0$

Test statistic value: $z = \dfrac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\dfrac{\sigma_1^2}{m} + \dfrac{\sigma_2^2}{n}}}$

Alternative Hypothesis	**Rejection Region for Level α Test**
H_a: $\mu_1 - \mu_2 > \Delta_0$	$z \geq z_\alpha$ (upper-tailed test)
H_a: $\mu_1 - \mu_2 < \Delta_0$	$z \leq -z_\alpha$ (lower-tailed test)
H_a: $\mu_1 - \mu_2 \neq \Delta_0$	either $z \geq z_{\alpha/2}$ or $z \leq -z_{\alpha/2}$ (two-tailed test)

Sample solution

1. The parameter of interest is $\mu_1 - \mu_2$, the difference between true mean GPA for the < 10 (conceptual) population and true mean GPA for the ≥ 10 population.

2. The null hypothesis is H_0: $\mu_1 - \mu_2 = 0$.

3. The alternative hypothesis is H_a: $\mu_1 - \mu_2 \neq 0$; if H_a is true then μ_1 and μ_2 are different. Although it would seem unlikely that $\mu_1 - \mu_2 > 0$ (those with low study hours have higher mean GPA) we will allow it as a possibility and do a two-tailed test.

4. With $\Delta_0 = 0$, the test statistic value is

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}$$

5. The inequality in H_a implies that the test is two-tailed. For $\alpha = .05$, $\alpha/2 = .025$ and $z_{\alpha/2} = z_{.025} = 1.96$. H_0 will be rejected if $z \geq 1.96$ or $z \leq -1.96$.

Solution

6. Substituting $m = 10$, $\bar{x} = 2.97$, $\sigma_1^2 = .36$, $n = 11$, $\bar{y} = 3.06$, and $\sigma_2^2 = .36$ into the formula for z yields

$$z = \frac{2.97 - 3.06}{\sqrt{\frac{.36}{10} + \frac{.36}{11}}} = \frac{-.09}{.262} = -.34$$

That is, the value of $\bar{x} - \bar{y}$ is only one-third of a standard deviation below what would be expected when H_0 is true.

7. Because the value of z is not even close to the rejection region, there is no reason to reject the null hypothesis. This test shows no evidence of any relationship between study hours and GPA. ■

Proposition

Provided that m and n are both large, a CI for $\mu_1 - \mu_2$ *with a confidence level of approximately* $100(1-\alpha)\%$ *is*

$$\bar{x} - \bar{y} \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}$$

where $-$ *gives the lower limit and* $+$ *the upper limit of the interval. An upper or lower confidence bound can also be calculated by retaining the appropriate sign and replacing* $z_{\alpha/2}$ *by* z_α.

Proposition

Use of the test statistic value

$$z = \frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}}$$

along with the previously stated upper-, lower-, and two-tailed rejection regions based on z critical values gives large-sample tests whose significance levels are approximately α. These tests are usually appropriate if both $m > 40$ and $n > 40$. A P-value is computed exactly as it was for our earlier z tests.

The two-sample t test and confidence interval

- For one-sample inferences:

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

- For two-sample inferences:

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}} \sim t_\nu$$

where ν is some appropriate degree of freedom (which depends on m and n).

2-sample t test: degree of freedom

THEOREM When the population distributions are both normal, the standardized variable

$$T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}} \tag{10.2}$$

has approximately a t distribution with df ν estimated from the data by

$$\nu = \frac{\left(\frac{s_1^2}{m} + \frac{s_2^2}{n}\right)^2}{\frac{(s_1^2/m)^2}{m-1} + \frac{(s_2^2/n)^2}{n-1}} = \frac{\left[(se_1)^2 + (se_2)^2\right]^2}{\frac{(se_1)^4}{m-1} + \frac{(se_2)^4}{n-1}}$$

where

$$se_1 = \frac{s_1}{\sqrt{m}} \qquad se_2 = \frac{s_2}{\sqrt{n}}$$

(round ν down to the nearest integer).

CIs for difference of the two population means

The **two-sample t confidence interval for $\mu_1 - \mu_2$** with confidence level $100(1-\alpha)\%$ is then

$$\bar{x} - \bar{y} \pm t_{\alpha/2,\nu}\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}$$

A one-sided confidence bound can be calculated as described earlier.

2-sample t procedures

The **two-sample *t* test** for testing H_0: $\mu_1 - \mu_2 = \Delta_0$ is as follows:

$$\text{Test statistic value: } t = \frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}}$$

Alternative Hypothesis	**Rejection Region for Approximate Level α Test**
H_a: $\mu_1 - \mu_2 > \Delta_0$	$t \geq t_{\alpha,\nu}$ (upper-tailed test)
H_a: $\mu_1 - \mu_2 < \Delta_0$	$t \leq -t_{\alpha,\nu}$ (lower-tailed test)
H_a: $\mu_1 - \mu_2 \neq \Delta_0$	either $t \geq t_{\alpha/2,\nu}$ or $t \leq -t_{\alpha/2,\nu}$ (two-tailed test)

A P-value can be computed as described in Section 9.4 for the one-sample t test.

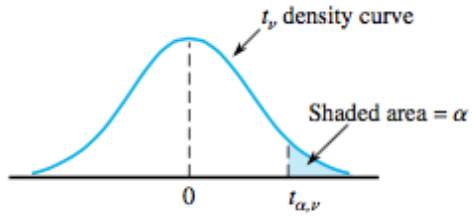
Example

A paper reported the following data on tensile strength (psi) of liner specimens both when a certain fusion process was used and when this process was not used:

No fusion	2748	2700	2655	2822	2511			
	3149	3257	3213	3220	2753			
	$m = 10$	$\bar{x} = 2902.8$	$s_1 = 277.3$					
Fused	3027	3356	3359	3297	3125	2910	2889	2902
	$n = 8$	$\bar{y} = 3108.1$	$s_2 = 205.9$					

The authors of the article stated that the fusion process increased the average tensile strength. With confidence level $\alpha = 0.05$, carry out a test of hypotheses to see whether the data supports this conclusion (and provide the P-value of the test)

Table A.5 Critical Values for t Distributions



	α						
ν	.10	.05	.025	.01	.005	.001	.0005
1	3.078	6.314	12.706	31.821	63.657	318.31	636.62
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.767
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745

Solution

1. Let μ_1 be the true average tensile strength of specimens when the no-fusion treatment is used and μ_2 denote the true average tensile strength when the fusion treatment is used.
2. H_0: $\mu_1 - \mu_2 = 0$ (no difference in the true average tensile strengths for the two treatments)
3. H_a: $\mu_1 - \mu_2 < 0$ (true average tensile strength for the no-fusion treatment is less than that for the fusion treatment, so that the investigators' conclusion is correct)

4. The null value is $\Delta_0 = 0$, so the test statistic is

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}}$$

5. We now compute both the test statistic value and the df for the test:

$$t = \frac{2902.8 - 3108.1}{\sqrt{\frac{277.3^2}{10} + \frac{205.9^2}{8}}} = \frac{-205.3}{113.97} = -1.8$$

Using $s_1^2/m = 7689.529$ and $s_2^2/n = 5299.351$,

$$\nu = \frac{(7689.529 + 5299.351)^2}{\frac{(7689.529)^2}{9} + \frac{(5299.351)^2}{7}} = \frac{168{,}711{,}004}{10{,}581{,}747} = 15.94$$

so the test will be based on 15 df.

Example

The following data summarizes the proportional stress limits for specimens constructed using two different types of wood:

Type of wood	Sample size	Sample mean	Sample sd
Red oak	14	8.48	0.79
Douglas fir	10	6.65	1.28

Assuming that both samples were selected from normal distributions, carry out a test of hypotheses with significance level $\alpha = 0.05$ to decide whether the true average proportional stress limit for red oak joints exceeds that for Douglas fir joints by more than 1 MPa. Provide the P-value of the test.

The paired samples setting

1. There is only one set of n individuals or experimental objects
2. Two observations are made on each individual or object

Example

Example

Consider two scenarios:

A. Insulin rate is measured on 30 patients before and after a medical treatment.

B. Insulin rate is measured on 30 patients receiving a placebo and 30 other patients receiving a medical treatment.

Notes

- In the independent case, we construct the statistics by looking at the distribution of

$$\bar{X} - \bar{Y}$$

which has

$$E[\bar{X}-\bar{Y}] = \mu_1-\mu_2, \qquad Var[\bar{X}-\bar{Y}] = Var(\bar{X})+Var(\bar{Y}) = \frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}$$

- With paired data, the X and Y observations within each pair are not independent, so $\bar{X}$ and $\bar{Y}$ are not independent of each other $\rightarrow$ the computation of the variance is in valid $\rightarrow$ could not use the old formulas

The paired t-test

- Because different pairs are independent, the D_i's are independent of each other
- We also have

$$E[D] = E[X - Y] = E[X] - E[Y] = \mu_1 - \mu_2 = \mu_D$$

- Testing about $\mu_1 - \mu_2$ is just the same as testing about μ_D
- Idea: to test hypotheses about $\mu_1 - \mu_2$ when data is paired:
  1. form the differences $D_1, D_2, \ldots, D_n$
  2. carry out a one-sample t-test (based on $n - 1$ df) on the differences.

Settings

Assumption

1. *The data consists of n independently selected pairs of independently normally distributed random variables* $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ *with* $E(X_i) = \mu_1$ *and* $E(Y_i) = \mu_2$.
2. *Let*

$$D_1 = X_1 - Y_1, \qquad D_2 = X_2 - Y_2, \ldots, \qquad D_n = X_n - Y_n,$$

so the D_i*'s are the differences within pairs.*

Confidence intervals

- A t confidence interval for for $\mu_D = \mu_1 - \mu_2$ can be constructed based on the fact that

$$T = \frac{\bar{D} - \mu_D}{S_D/\sqrt{n}}$$

  follows the t distribution with degree of freedom $n - 1$.
- The CI for μ_D is

$$\bar{d} \pm t_{\alpha/2,n-1}\frac{s_D}{\sqrt{n}}$$

- A one-sided confidence bound results from retaining the relevant sign and replacing $t_{\alpha/2,n-1}$ by $t_{\alpha,n-1}$.

The paired t-test

THE PAIRED t TEST

Null hypothesis: H_0: $\mu_D = \Delta_0$

Test statistic value: $t = \dfrac{\overline{d} - \Delta_0}{s_D/\sqrt{n}}$

(where $D = X - Y$ is the difference between the first and second observations within a pair, and $\mu_D = \mu_1 - \mu_2$)
(where $\overline{d}$ and s_D are the sample mean and standard deviation, respectively, of the d_i's)

Alternative Hypothesis	**Rejection Region for Level α Test**
H_a: $\mu_D > \Delta_0$	$t \geq t_{\alpha,n-1}$
H_a: $\mu_D < \Delta_0$	$t \leq -t_{\alpha,n-1}$
H_a: $\mu_D \neq \Delta_0$	either $t \geq t_{\alpha/2,n-1}$ or $t \leq -t_{\alpha/2,n-1}$

A P-value can be calculated as was done for earlier t tests.

Example

Example

Consider two scenarios:

A. Insulin rate is measured on 30 patients before and after a medical treatment.

B. Insulin rate is measured on 30 patients receiving a placebo and 30 other patients receiving a medical treatment.

What type of test should be used in each cased: paired or unpaired?

Example

Example

Suppose we have a new synthetic material for making soles for shoes. We'd like to compare the new material with leather – using some energetic kids who are willing to wear test shoes and return them after a time for our study. Consider two scenarios:

A. Giving 50 kids synthetic sole shoes and 50 kids leather shoes and then collect them back, comparing the average wear in each group

B. Give each of a random sample of 50 kids one shoe made by the new synthetic materials and one shoe made with leather

What type of test should be used in each cased: paired or unpaired?

Example

Consider an experiment in which each of 13 workers was provided with both a conventional shovel and a shovel whose blade was perforated with small holes. The following data on stable energy expenditure is provided:

Worker:	1	2	3	4	5	6	7
Conventional:	.0011	.0014	.0018	.0022	.0010	.0016	.0028
Perforated:	.0011	.0010	.0019	.0013	.0011	.0017	.0024

Worker:	8	9	10	11	12	13
Conventional:	.0020	.0015	.0014	.0023	.0017	.0020
Perforated:	.0020	.0013	.0013	.0017	.0015	.0013

Calculate a confidence interval at the 95% confidence level for the true average difference between energy expenditure for the conventional shovel and the perforated shovel (assuming that the differences follow normal distribution).

Example

Consider an experiment in which each of 13 workers was provided with both a conventional shovel and a shovel whose blade was perforated with small holes. The following data on stable energy expenditure is provided:

Worker:	1	2	3	4	5	6	7
Conventional:	.0011	.0014	.0018	.0022	.0010	.0016	.0028
Perforated:	.0011	.0010	.0019	.0013	.0011	.0017	.0024

Worker:	8	9	10	11	12	13
Conventional:	.0020	.0015	.0014	.0023	.0017	.0020
Perforated:	.0020	.0013	.0013	.0017	.0015	.0013

Carry out a test of hypotheses at significance level .05 to see if true average energy expenditure using the conventional shovel exceeds that using the perforated shovel; include a P-value in your analysis.

t-table

t \ ν	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1.6	.178	.125	.104	.092	.085	.080	.077	.074	.072	.070	.069	.068	.067	.065	.065	.065	.064	.064
1.7	.169	.116	.094	.082	.075	.070	.065	.064	.062	.060	.059	.057	.056	.055	.055	.054	.054	.053
1.8	.161	.107	.085	.073	.066	.061	.057	.055	.053	.051	.050	.049	.048	.046	.046	.045	.045	.044
1.9	.154	.099	.077	.065	.058	.053	.050	.047	.045	.043	.042	.041	.040	.038	.038	.038	.037	.037
2.0	.148	.092	.070	.058	.051	.046	.043	.040	.038	.037	.035	.034	.033	.032	.032	.031	.031	.030
2.1	.141	.085	.063	.052	.045	.040	.037	.034	.033	.031	.030	.029	.028	.027	.027	.026	.025	.025
2.2	.136	.079	.058	.046	.040	.035	.032	.029	.028	.026	.025	.024	.023	.022	.022	.021	.021	.021
2.3	.131	.074	.052	.041	.035	.031	.027	.025	.023	.022	.021	.020	.019	.018	.018	.018	.017	.017
2.4	.126	.069	.048	.037	.031	.027	.024	.022	.020	.019	.018	.017	.016	.015	.015	.014	.014	.014
2.5	.121	.065	.044	.033	.027	.023	.020	.018	.017	.016	.015	.014	.013	.012	.012	.012	.011	.011
2.6	.117	.061	.040	.030	.024	.020	.018	.016	.014	.013	.012	.012	.011	.010	.010	.010	.009	.009
2.7	.113	.057	.037	.027	.021	.018	.015	.014	.012	.011	.010	.010	.009	.008	.008	.008	.008	.007
2.8	.109	.054	.034	.024	.019	.016	.013	.012	.010	.009	.009	.008	.008	.007	.007	.006	.006	.006
2.9	.106	.051	.031	.022	.017	.014	.011	.010	.009	.008	.007	.007	.006	.005	.005	.005	.005	.005
3.0	.102	.048	.029	.020	.015	.012	.010	.009	.007	.007	.006	.006	.005	.004	.004	.004	.004	.004