Math 3070, Applied Statistics

Section 1

November 20, 2019

Lecture Outline, 11/20

Section 9.3

- T-Test for paired data

Paired t Test

To test a difference in means between the two normal populations, given a random sample of pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$, the correct procedure is to use the *paired t test*:

To perform the paired t test, take the differences $D_i = X_i - Y_i$ between corresponding observations in each pair and then perform a one-sample t test on the resulting differences D_i.

Test Statistic	Alternative hypothesis	Rejection region
$T = \dfrac{\overline{D} - \Delta_0}{S_D/\sqrt{n}}$	$H_a : \mu_1 - \mu_2 > \Delta_0$	$T > t_{\alpha,\nu}$
	$H_a : \mu_1 - \mu_2 < \Delta_0$	$T < -t_{\alpha,\nu}$
	$H_a : \mu_1 - \mu_2 \neq \Delta_0$	$\|T\| > t_{\alpha/2,\nu}$

Here S_D is the sample standard deviation of the differences $D_1, \ldots, D_n$, and $\nu = n - 1$. If n is large, say $n > 40$, then the assumption that the populations are normal may be dropped.

Example

Does the data provide significant evidence the mean zinc concentration in bottom water exceeds that of surface water?

Location	1	2	3	4	5	6
Bottom water (X_i)	.430	.266	.567	.531	.707	.716
Surface water (Y_i)	.415	.238	.390	.410	.605	.609
Difference (D_i)	.015	.028	.177	.121	.102	.107

We are testing $H_0 : \mu_1 - \mu_2 = 0$ against the alternative $H_a : \mu_1 - \mu_2 > 0$. We calculate $\overline{D} = .0917$, $\overline{S} = .0607$, so

$$T = \frac{\overline{D} - \Delta_0}{S/\sqrt{n}} = \frac{.0917 - 0}{.0607/\sqrt{6}} = 3.7$$

This gives a P-value of $P = P(T > 3.7) = .007$. Thus the data provides highly significant evidence that the mean zinc concentration in bottom water exceeds that of surface water.

Paired t Confidence Interval

Again suppose we have two normal populations with means μ_1 and μ_2 respectively, and we wish to construct a confidence interval for $\mu_1 - \mu_2$ based on a random sample of pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$.

We form the differences $D_i = X_i - Y_i$ and then simply construct the one-sample t confidence interval based on the D_i's:

Given paired data from two samples, a $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by

$$\overline{D} \pm \frac{t_{\alpha/2,n-1} S_D}{\sqrt{n}}$$

If n is large, say $n > 40$, then the assumption that the populations are normal may be dropped.

Example

Based on the given data, find a 95% confidence interval for the difference in mean zinc concentration in bottom water vs. surface water.

Location	1	2	3	4	5	6
Bottom water (X_i)	.430	.266	.567	.531	.707	.716
Surface water (Y_i)	.415	.238	.390	.410	.605	.609
Difference (D_i)	.015	.028	.177	.121	.102	.107

Here we have $\overline{D} = .0917$, $\overline{S} = .0607$, and $t_{\alpha/2,\nu} = t_{.025,5} = 2.571$, so the 95% confidence interval is given by

$$\begin{aligned} \overline{D} \pm \frac{t_{\alpha/2,\nu} S_D}{\sqrt{n}} &= .0917 \pm \frac{(2.571)(.0607)}{\sqrt{5}} \\ &= .0917 \pm .0698 \end{aligned}$$

Summary

Two-sample z C.I. for $\mu_1 - \mu_2$	$\overline{X} - \overline{Y} \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}$
Two-sample t C.I. for $\mu_1 - \mu_2$	$\overline{X} - \overline{Y} \pm t_{\alpha/2,\nu}\sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}$
Paired t C.I. for $\mu_1 - \mu_2$	$\overline{D} \pm \frac{t_{\alpha/2,n-1}S_D}{\sqrt{n}}$

Test	Null Hypothesis	Test Statistic
Two-sample z test	$H_0 : \mu_1 - \mu_2 = \Delta_0$	$Z = \frac{\overline{X}-\overline{Y}-\Delta_0}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}$
Two-sample t test	$H_0 : \mu_1 - \mu_2 = \Delta_0$	$T = \frac{\overline{X}-\overline{Y}-\Delta_0}{\sqrt{\frac{S_1^2}{m}+\frac{S_2^2}{n}}}$
Paired t test	$H_0 : \mu_1 - \mu_2 = \Delta_0$	$T = \frac{\overline{D}-\Delta_0}{S_D/\sqrt{n}}$