Ch. 9 – Inferences Based on Two Samples

Two-sample z Test

Suppose we are given two independent normal random samples:

- $X_1, \ldots, X_m$ from a $N(\mu_1, \sigma_1^2)$ distribution
- $Y_1, \ldots, Y_n$ from a $N(\mu_2, \sigma_2^2)$ distribution

If we know the variances σ_1^2 and σ_2^2, we may use a **two-sample z test** to test the null hypothesis $H_0 : \mu_1 - \mu_2 = \Delta_0$:

Test Statistic	Alternative hypothesis	Rejection region
$Z = \dfrac{\overline{X} - \overline{Y} - \Delta_0}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}$	$H_a : \mu_1 - \mu_2 > \Delta_0$	$Z > z_\alpha$
	$H_a : \mu_1 - \mu_2 < \Delta_0$	$Z < -z_\alpha$
	$H_a : \mu_1 - \mu_2 \neq \Delta_0$	$\lvert Z \rvert > z_{\alpha/2}$

If m and n are large (say, $m > 40$ and $n > 40$), then we may use sample variances S_1^2 and S_2^2 in place of σ_1^2 and σ_2^2 and may drop the assumption that the distributions are normal.

Example

A random sample of 20 specimens of cold-rolled steel had an average yield strength of 29.8 ksi. For a random sample of 25 two-sided galvanized steel specimens the average was 34.7 ksi. Assuming that the two yield-strength distributions are normal with $\sigma_1 = 4.0$ and $\sigma_2 = 5.0$, does the data provide significance evidence (at the $\alpha = .01$ level) for a difference between the mean yield strength of the two types of specimens?

We want to test the null hypothesis $H_0 : \mu_1 - \mu_2 = 0$ against the alternative $H_a : \mu_1 - \mu_2 \neq 0$. We calculate the test statistic:

$$Z = \frac{\overline{X} - \overline{Y} - \Delta_0}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} = \frac{29.8 - 34.7 - 0}{\sqrt{\frac{(4.0)^2}{20} + \frac{(5.0)^2}{25}}} = -3.65$$

The P-value for the test is $P(|Z| > 3.65) = 2\Phi(-3.65) = .00026$. This provides strong evidence for a difference in the mean yield strengths of the two types of specimens.

z Confidence Interval for Difference of Two Means

Suppose we are given two independent normal random samples:

- $X_1, \ldots, X_m$ from a $N(\mu_1, \sigma_1^2)$ distribution
- $Y_1, \ldots, Y_n$ from a $N(\mu_2, \sigma_2^2)$ distribution

Assume we know the variances σ_1^2 and σ_2^2.

A $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by

$$\overline{X} - \overline{Y} \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}$$

If m and n are large (say, $m > 40$ and $n > 40$), then we may use sample variances S_1^2 and S_2^2 in place of σ_1^2 and σ_2^2 and may drop the assumption that the distributions are normal.

Example

A random sample of 20 specimens of cold-rolled steel had an average yield strength of 29.8 ksi. For a random sample of 25 two-sided galvanized steel specimens the average was 34.7 ksi. Assuming that the two yield-strength distributions are normal with $\sigma_1 = 4.0$ and $\sigma_2 = 5.0$, find a 95% confidence interval for the difference in mean yield strength between the two types of specimens?

$$\begin{aligned}
&\overline{X} - \overline{Y} \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}} \\
&= 29.8 - 34.7 \pm 1.96\sqrt{\frac{(4.0)^2}{20} + \frac{(5.0)^2}{25}} \\
&= -4.9 \pm 2.63
\end{aligned}$$

Two-Sample t Test (Welch's t Test)

Suppose we are given two independent normal random samples:

- $X_1, \ldots, X_m$ from a $N(\mu_1, \sigma_1^2)$ distribution
- $Y_1, \ldots, Y_n$ from a $N(\mu_2, \sigma_2^2)$ distribution

If we don't know the variances σ_1^2 and σ_2^2, we may use a **two-sample t test** to test the null hypothesis $H_0 : \mu_1 - \mu_2 = \Delta_0$:

Test Statistic	Alternative hypothesis	Rejection region
$T = \dfrac{\overline{X} - \overline{Y} - \Delta_0}{\sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}}$	$H_a : \mu_1 - \mu_2 > \Delta_0$	$T > t_{\alpha,\nu}$
	$H_a : \mu_1 - \mu_2 < \Delta_0$	$T < -t_{\alpha,\nu}$
	$H_a : \mu_1 - \mu_2 \neq \Delta_0$	$\lvert T\rvert > t_{\alpha/2,\nu}$

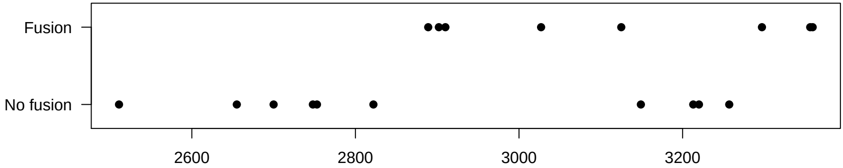
Here the degrees of freedom ν is estimated by

$$\nu = \frac{\left(\dfrac{S_1^2}{m} + \dfrac{S_2^2}{n}\right)^2}{\dfrac{(S_1^2/m)^2}{m-1} + \dfrac{(S_2^2/n)^2}{n-1}}$$

Example

The deterioration of many municipal pipeline networks across the country is a growing concern. One technology proposed for pipeline rehabilitation uses a flexible liner threaded through existing pipe. An article 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	3149	2700	2655	2822	2511	3257	3213	3220	2753
Fusion	3027	3356	3359	3297	3125	2910	2889	2902		



Example

Does the data provide significant evidence for a difference in the mean tensile strength of the two types of specimens?

We will test the null hypothesis $H_0 : \mu_1 - \mu_2 = 0$ against the alternative $H_a : \mu_1 - \mu_2 \neq 0$.
The specimens with no fusion have $\overline{X} = 2902.8$ and $S_1 = 277.3$, while those with fusion have $\overline{Y} = 3108.1$ and $S_2 = 205.9$.

$$T = \frac{\overline{X} - \overline{Y} - \Delta_0}{\sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}} = \frac{2902.8 - 3108.1 - 0}{\sqrt{\frac{(277.3)^2}{10} + \frac{(205.9)^2}{8}}} \approx -1.8$$

$$\nu = \frac{\left(\frac{S_1^2}{m} + \frac{S_2^2}{n}\right)^2}{\frac{(S_1^2/m)^2}{m-1} + \frac{(S_2^2/n)^2}{n-1}} = \frac{(7689.5 + 5299.3)^2}{\frac{(7689.5)^2}{9} + \frac{(5299.3)^2}{7}} = 15.9 \approx 16$$

The P-value for the test is

$$P(|T| > 1.8) = 2P(T > 1.8) = 2 \cdot .045 = .090$$

t Confidence Interval for Difference of Two Means

Suppose we are given two independent normal random samples:

- $X_1, \ldots, X_m$ from a $N(\mu_1, \sigma_1^2)$ distribution
- $Y_1, \ldots, Y_n$ from a $N(\mu_2, \sigma_2^2)$ distribution

Assume we *do not* know the variances σ_1^2 and σ_2^2.

A $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by

$$\overline{X} - \overline{Y} \pm t_{\alpha/2,\nu}\sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}$$

Here, as before the degrees of freedom ν is estimated by

$$\nu = \frac{\left(\frac{S_1^2}{m} + \frac{S_2^2}{n}\right)^2}{\frac{(S_1^2/m)^2}{m-1} + \frac{(S_2^2/n)^2}{n-1}}$$

Example

Based on the pipeline liner data, find a 95% confidence interval for the difference in mean tensile strength between the two types of specimens (no fusion vs. fusion).

The specimens with no fusion had $\overline{X} = 2902.8$ and $S_1 = 277.3$, while those with fusion had $\overline{Y} = 3108.1$ and $S_2 = 205.9$. We calculated that the appropriate degrees of freedom was $\nu \approx 16$. This leads to a critical value of $t_{.025,16} = 2.120$. A 95% confidence interval for the difference $\mu_1 - \mu_2$ is then given by

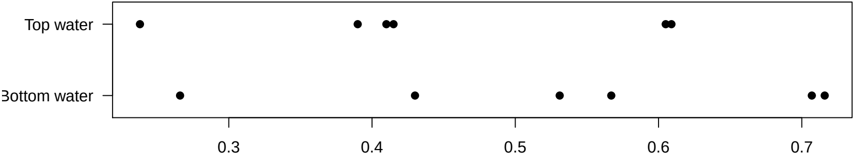
$$\begin{aligned}
&\overline{X} - \overline{Y} \pm t_{\alpha/2,\nu}\sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}} \\
&= 2902.8 - 3108.1 \pm 2.120\sqrt{\frac{(277.3)^2}{10} + \frac{(205.9)^2}{8}} \\
&= -205.3 \pm 241.6
\end{aligned}$$

Problem

An article [1] reports on a study in which six river locations were selected and the zinc concentration (mg/L) determined for both surface water and bottom water at each location:

Location	1	2	3	4	5	6
Bottom water	.430	.266	.567	.531	.707	.716
Surface water	.415	.238	.390	.410	.605	.609

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



[1] "Trace Metals of South Indian River" (Envir. Studies, 1982: 62–66)

Problem

We want to test the null hypothesis $H_0 : \mu_1 - \mu_2 = 0$ against an alternative $H_0 : \mu_1 - \mu_2 > 0$ based on the data:

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

It may seem natural to treat this as a two-sample problem: We could calculate $\overline{X} = .536$, $S_1 = .171$, $\overline{Y} = .444$, and $S_2 = .142$:

$$T = \frac{\overline{X} - \overline{Y} - \Delta_0}{\sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}} = \frac{.536 - .444 - 0}{\sqrt{\frac{(.171)^2}{6} + \frac{(.142)^2}{6}}} \approx 1.0$$

$$\nu = \left(\frac{S_1^2}{m} + \frac{S_2^2}{n}\right)^2 \div \left(\frac{(S_1^2/m)^2}{m-1} + \frac{(S_2^2/n)^2}{n-1}\right) = 9.7 \approx 10$$

$$P = P(T > 1.0) = .170$$

However, this method would be *incorrect* because the two samples are not independent of each other!

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
$T = \dfrac{\overline{D} - \Delta_0}{S_D/\sqrt{n}}$	$H_a : \mu_1 - \mu_2 > \Delta_0$	$T > t_{\alpha,\nu}$
	$H_a : \mu_1 - \mu_2 < \Delta_0$	$T < -t_{\alpha,\nu}$
	$H_a : \mu_1 - \mu_2 \neq \Delta_0$	$\lvert T \rvert > t_{\alpha/2,\nu}$

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

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

Large-Sample z Test for Equality of Two Proportions

Suppose proportion p_1 of one population has a certain characteristic, while proportion p_2 of a second population does.

We want to test the hypothesis $H_0 : p_1 = p_2$ based on a random sample of m individuals from the first population and n individuals from the second.

Given sample proportions $\hat{p}_1 = \frac{X}{m}$ and $\hat{p}_2 = \frac{Y}{n}$, the **z test** is

Test Statistic	Alternative hyp.	Rejection region
$Z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{m}+\frac{1}{n}\right)}}$	$H_a : p_1 > p_2$	$Z > z_\alpha$
	$H_a : p_1 < p_2$	$Z < -z_\alpha$
	$H_a : p_1 \neq p_2$	$\lvert Z \rvert > z_{\alpha/2}$

Here $\hat{p} = \frac{X+Y}{m+n}$ is the **pooled sample proportion**.

This is a large-sample test, appropriate only if $X, m-X, Y, n-Y$ are all at least 10.

Example

An article reported that of 549 study participants who regularly used aspirin after being diagnosed with colorectal cancer, there were 81 colorectal cancer-specific deaths, whereas among 730 similarly diagnosed individuals who did not subsequently use aspirin, there were 141 colorectal cancer-specific deaths[2]. Does this data suggest that the regular use of aspirin after diagnosis will decrease the incidence rate of colorectal cancer-specific deaths?

Here we want to test the null hypothesis $H_0 : p_1 = p_2$ against the alternative $H_a : p_1 < p_2$. We have $m = 549$, $n = 730$, $\hat{p}_1 = \frac{81}{549}$, $\hat{p}_2 = \frac{141}{730}$, $\hat{p} = \frac{81+141}{549+730}$.

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{m}+\frac{1}{n}\right)}} = -2.13$$

$$P = P(Z < -2.13) = .0166$$

[2] "Aspirin Use and Survival After Diagnosis of Colorectal Cancer" (J. of the Amer. Med. Assoc., 2009: 649–658)

z Confidence Interval for Difference of Two Proportions

Suppose proportion p_1 of one population has a certain characteristic, while proportion p_2 of a second population does.

Assume we have a random sample of m individuals from the first population and n individuals from the second, and let $\hat{p}_1 = \frac{X}{m}$ and $\hat{p}_2 = \frac{Y}{n}$ be the sample proportions.

An approximate $100(1-\alpha)\%$ confidence interval for $p_1 - p_2$ is

$$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{m} + \frac{\hat{p}_2(1-\hat{p}_2)}{n}}$$

Note that the pooled sample proportion $\hat{p}$ does *not* appear in the confidence interval for a difference of two proportions.

Again this is a large-sample procedure, appropriate only if $X, m-X, Y, n-Y$ are all at least 10.

Example

Recall the data from the colorectal cancer study: 81 deaths out of 549 participants who took aspirin, and 141 out of 730 who did not take aspirin. Find an approximate 95% confidence interval for the difference between the two proportions.

We have $m = 549$, $n = 730$, $\hat{p}_1 = \frac{81}{549}$, $\hat{p}_2 = \frac{141}{730}$; an approximate 95% confidence interval is given by

$$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{m} + \frac{\hat{p}_2(1-\hat{p}_2)}{n}}$$
$$= -.046 \pm .041$$

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 $\mu_1 - \mu_2$	$\overline{D} \pm \frac{t_{\alpha/2,n-1}S_D}{\sqrt{n}}$
Approximate C.I. for $p_1 - p_2$	$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{m} + \frac{\hat{p}_2(1-\hat{p}_2)}{n}}$

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	$H_0 : \mu_1 - \mu_2 = \Delta_0$	$T = \frac{\overline{D}-\Delta_0}{S_D/\sqrt{n}}$
Approximate z test for proportions	$H_0 : p_1 = p_2$	$Z = \frac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{m}+\frac{1}{n}\right)}}$