Tests Associated with Normal Data

8.1 One Sample Z-test

Example 1. Curious George has a driver’s license that gives his weight as 187 pounds. He suspects he may have gained a little weight over the past few years. We examine the truth of the weight on George’s license using a test of significance. Recall 4 measurements in last month: 190.5, 189.0, 195.5, 187.0 with sample mean $\bar{x} = 190.5$. We assume that the data are normal with a population standard deviation $\sigma = 3$.

Are these observations compatible with a true mean weight of $\mu = 187$ pounds? To answer this question, we ask another: what is the probability of observing a sample mean of 190.5 or larger from a population with a true mean of 187? The answer is 0.01. Because this is small, we conclude that George has gained weight.

Suppose the sample mean of George’s 4 measurements is 188.5 pounds. What will be our conclusion? Still use test of significance, the probability we observe 188.5 or larger is 0.16. Since it is not particularly small, we do not reject that the true weight is $\mu = 187$.

Null Hypothesis

The statement being tested is called the null hypothesis, denoted as $H_0$. The test is designed to assess the strength of the evidence against the null hypothesis. Usually the null hypothesis is the default.

Alternative hypothesis

The statement we suspect is true instead of $H_0$ is called the alternative hypothesis and is abbreviated as $H_a$. In George’s weight example, we have $H_0 : \mu = 187$ vs. $H_a : \mu > 187$. Hypotheses always refer to some population or model, not to a particular outcome. We must state $H_0$ and $H_a$ in terms of population parameters. The alternative hypothesis can vary in
directions even under the same $H_0$. The example above is one-sided. If we only feel interested in whether George’s weight has changed since he renewed the driver’s licence. The hypotheses become

$$H_0 : \mu = 187 \text{ vs. } H_a : \mu \neq 187$$

which is two-sided.

**Test statistics**

The principle of hypotheses testing:

- the test is based on a statistic that estimates the parameter that appears in the hypothesis.
- values of the estimate far from the parameter value specified by $H_0$ give evidence against $H_0$. The alternative hypothesis determines which directions count as against $H_0$.

A test statistic measures compatibility between the null hypothesis and the data. We use it for the probability calculation. Usually it is a random variable with a distribution that we know. In George’s weight example, $\bar{x}$ has the $N(\mu, \sigma^2/n)$ distribution when the data are normal. If we assume $H_0 : \mu = 187$. The test statistic for this problem is the standardized version of $\bar{x}$:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

which has $N(0,1)$ distribution.

**p-value**

The probability, computed assuming that $H_0$ is true, that the test statistic would take a value as extreme or more extreme in favor of $H_a$ than actually observed is called the p-value of the test. The smaller the p-value, the stronger the evidence against $H_0$ provided by the data.

Example 2. George observed an SRS of size $n = 4$ from a normal distribution with $\sigma = 3$. He wants to test

$$H_0 : \mu = 187 \text{ vs. } H_a : \mu > 187$$

The test statistic

$$z = \frac{190.5 - 187}{3/\sqrt{4}} = 2.33$$

The p-value is the probability of observing a value of a standard normal at least as large as the one we observed, $z = 2.33$.

$$P(Z \geq 2.33) = 1 - 0.9901 = 0.01$$

**Statistical significance**

So far, we assume the model, calculate the p-values, but what is the conclusion? Do we accept or reject the $H_0$? *It depends!* Usually we set up a significance level $\alpha$, which most common is 1% or 5%. If the p-value is as small as or smaller than $\alpha$, we say the data are statistically significant at level $\alpha$. 

2
Recap the hypothesis testing

1. state the null and alternative hypotheses, and decide the significance level \( \alpha \).
2. calculate the test statistic.
3. find the p-value for the observed data.
4. state a conclusion.

Tests for a population mean

Assume the population has a distribution with mean \( \mu \) and standard deviation \( \sigma \). We have an simple random sample (SRS) with size \( n \). We could use the standardized sample mean as the test statistic. By central limit theorem, we could compare the \( z \) with a \( N(0,1) \) distribution. Therefore, testing \( H_0 : \mu = \mu_0 \) has a test statistic

\[
    z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}.
\]

The p-values for a test of \( H_0 \) against

\( H_a : \mu > \mu_0 \) is \( P(Z \geq z) \).

\( H_a : \mu < \mu_0 \) is \( P(Z \leq z) \).

\( H_a : \mu \neq \mu_0 \) is \( 2P(Z \geq |z|) \).

These p-values are exact if the population is normal and are approximately correct for large \( n \).

8.2 One Sample T-test

What happens if we do not know \( \sigma \)? First we introduce a new distribution called the \( t \) distribution. If we have an SRS of size \( n \) from a normal distribution \( N(\mu, \sigma^2) \). Then the sample mean

\[
    \bar{x} \sim N(\mu, \sigma^2 / n).
\]

when \( \sigma \) unknown, we estimate it with the sample standard deviation \( s \), then we estimate the standard deviation of \( \bar{x} \) by \( s / \sqrt{n} \). This quantity is called the standard error of the sample mean \( \bar{x} \).

\( t \) distributions

Suppose that an SRS of size \( n \) is drawn from an \( N(\mu, \sigma^2) \) population. Then the one-sample \( t \) statistic

\[
    t = \frac{\bar{x} - \mu}{s / \sqrt{n}}
\]

has the \( t \) distribution with \( n - 1 \) degrees of freedom.
Suppose that an SRS of size \( n \) is drawn from a population with mean \( \mu \). To test the
\( H_0 : \mu = \mu_0 \) based on this sample, compute the one-sample t statistic

\[
t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}.
\]

In terms of a random variable \( T \) having the \( t(n-1) \) distribution, the p-value for a test of \( H_0 \)
against

- \( H_a : \mu > \mu_0 \) is \( P(T \geq t) \)
- \( H_a : \mu < \mu_0 \) is \( P(T \leq t) \)
- \( H_a : \mu \neq \mu_0 \) is \( 2P(T \geq |t|) \)

Confidence interval for a population mean

The 100(1 - \( \alpha \))\% confidence interval for \( \mu \) is given by

\[
\bar{x} \pm t^* \frac{s}{\sqrt{n}}
\]

Here, \( t^* \) is the \( 1 - \alpha / 2 \) quantile of t-dist’n with \( df = n - 1 \).

SAS example: On pp. 104-105 in text. Age is approximately normal distributed. Test its
mean.

8.3 Matched pairs t procedure

Matched pairs are common when complete randomization is not possible. One situation is
before-and-after observations on the same subjects. Assume the difference has a normal distri-
bution (NOT the original records). We essentially apply a one-sample t-test on the differences.

SAS example: data on pp. 195; code on pp. 199 in text.

The National Endowment for the Humanities sponsors summer institutes to improve the skills
of high school teachers of foreign languages. One such institute hosted 20 French teachers
for 4 weeks. At the beginning of the period, the teacher were given the Modern Language
Association’s listening tests of understanding of spoken French. After 4 weeks of class, the
listening test was given again. frenchsummer.dat gives the pretest and posttest scores.

\[ Gain = Posttest - Pretest. \]

Suppose \( \mu \) is the mean improvement that would be achieved if the entire population of French
teachers attended this summer institute. We would like to test

\[ H_0 : \mu = 0 \text{ vs. } H_a : \mu > 0 \]
### R command:
\[ \begin{array}{l}
> \text{length(Gain)} \\
[1] 20 \\
> \text{mean(Gain)} \\
[1] 2.5 \\
> \text{sd(Gain)} \\
[1] 2.892822 \\
> t<- (\text{mean(Gain)}-0)/(\text{sd(Gain)}/\sqrt{20}) \\
> t \\
[1] 3.864855 \\
> 1-\text{pt(t, 19)} \\
[1] 0.000521637
\end{array} \]

##### or direct use \texttt{t.test}
\[ \begin{array}{l}
> \text{t.test(Gain)} \\
> \text{t.test(Posttest, Pretest, paired=T)}
\end{array} \]

/* SAS code to create the data set.*/
data frenchsummer;
infile "C:\Teaching\sta5206\chapter8\frenchsummer.dat" firstobs=2;
input teacher Pretest Posttest;
gain=Posttest-pretest;
run;

### 8.4 Comparing Two Means

A two-sample problem can arise from a randomized comparative experiment that randomly divides the subjects into two groups and exposes each group to a different treatment. Comparing random samples separately selected from two populations is also a two-sample problem. Introduce some notation:

<table>
<thead>
<tr>
<th>Population</th>
<th>Variable</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(x_1)</td>
<td>(\mu_1)</td>
<td>(\sigma_1)</td>
</tr>
<tr>
<td>2</td>
<td>(x_2)</td>
<td>(\mu_2)</td>
<td>(\sigma_2)</td>
</tr>
</tbody>
</table>

We want to compare the two populations means, either by giving a confidence interval for \(\mu_1 - \mu_2\) or by testing the hypothesis of difference, \(H_0 : \mu_1 - \mu_2 = d\).

Inference is based on two independent SRSs.

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample size</th>
<th>Sample Mean</th>
<th>Sample Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(n_1)</td>
<td>(\bar{x}_1)</td>
<td>(s_1)</td>
</tr>
<tr>
<td>2</td>
<td>(n_2)</td>
<td>(\bar{x}_2)</td>
<td>(s_2)</td>
</tr>
</tbody>
</table>
Two-sample z statistic

The natural estimator of the difference $\mu_1 - \mu_2$ is the difference between the sample means, $ar{x}_1 - \bar{x}_2$.

$$E(\bar{x}_1 - \bar{x}_2) = \mu_1 - \mu_2.$$ $$Var(\bar{x}_1 - \bar{x}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$  

If we assume both populations are normal, then $\bar{x}_1 - \bar{x}_2$ is also normal. Therefore, if we know the values of $\sigma_1$ and $\sigma_2$, we can construct two-sample z statistic:

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

which has the standard normal $N(0,1)$ distribution.

Two-sample t statistic

Now suppose we do not know $\sigma_1$ or $\sigma_2$. We only get a two-sample t statistic:

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

Unfortunately, this statistic does NOT have a t distribution. Nonetheless, we can approximate the distribution of the two-sample t statistic by using $t(k)$ distribution with an approximation for the degrees of freedom $k$. There are two approximation methods:

1. Use a value of $k$ that is calculated from the data. In general it will not be a whole number. Most softwares calculates the approximate df using Satterthwaite approximation as:

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

It is quite accurate when both sample sizes $n_1$ and $n_2$ are 5 or larger.

The number df given by the above formula is always at least as large as the smaller of $n_1 - 1$ and $n_2 - 1$. It is never larger than $n_1 + n_2 - 2$.

2. Use $k$ equal to the smaller of $n_1 - 1$ and $n_2 - 1$. (Only for statisticians’ amusement)

Two-Sample t confidence interval

The 100(1 − $\alpha$)% 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 \( 1 - \alpha/2 \) quantile of \( t \)-dist’n with \( df = k \).

**Keep it in mind:** \( t \) methods usually conservative. Small samples require special care.

SAS example: pp. 171 commitment and reward

**The Pooled Two-Sample \( t \) procedures**

Suppose both populations have the same common, still unknown, variance \( \sigma^2 \). Both sample s.d. \( s_1 \) and \( s_2 \) estimate \( \sigma \).

The resulting estimator of \( \sigma^2 \) is

\[
s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}
\]

This is called the **pooled estimator of** \( \sigma^2 \). The test statistic becomes

\[
t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{(\frac{1}{n_1} + \frac{1}{n_2})s_p^2}},
\]

with an exact \( t \)-distribution \( df = n_1 + n_2 - 2 \).