Testing Means

Learning Objectives

After completing this module, the student will be able to

• carry out a statistical test of significance to test for means when the sample size is large
• carry out a statistical test of significance to test for means when the sample size is small

Knowledge and Skills

• Concepts: Z-score, t-test

Prerequisites

• Normal distribution
• Sample average
• Sample standard deviation
• P-value
Large Sample Size

Many experiments are designed to determine whether the sample comes from a population with a specified mean. For instance, a study by Mackowiak et al. (1992) investigated the average oral temperature in 130 healthy adults aged 18-40. The average in that group was 98.2°F instead of 98.6°F, which has been the standard since the 19th century. The data is provided in the Excel spreadsheet “t_test_workbook.xlsx” in the first tab “BodyTemp.” The body temperature is listed in the first column. The second column lists the gender (1=male, 2=female), and the third column lists the heart rate.

In-class Activity 1: Calculate the average body temperature (use the function AVERAGE) and the sample standard deviation (use the function STDEV).

What can we say about the average body temperature in this group of healthy individuals?

In the following, we will develop a statistical hypothesis test to determine whether this group’s body temperature is significantly different from what we expect based on the standard 98.6°F. We follow the steps that were developed in the Module Hypothesis Testing.

1. Formulation of a null hypothesis and alternative

The null hypothesis is that the sample comes from a population that has the expected mean of 98.6°F, that is,

\[ H_0 : \mu = 98.6^\circ F \]

The alternative is that the average body temperature is different from this expected value, that is,

\[ H_1 : \mu \neq 98.6^\circ F \]

This alternative specifies a two-sided test.

2. Construction of a test statistic that can discriminate between the null hypothesis and the alternative. Calculate the probability distribution under the null hypothesis.

The measurements that were taken are the body temperatures of 130 individuals. Since we are interested in testing whether the mean is a specific value, this suggests that we should calculate the sample average \( \bar{X} \). You already did this in In-class Activity 1 and found

\[ \bar{X} = 98.2^\circ F \]
We make the assumption that the body temperature is normally distributed with mean \( \mu \) and variance \( \sigma^2 \) (or standard deviation \( \sigma \)). The normality assumption is quite often a good approximation of the true distribution when we take a large number of measurements that are distributed around a specific value.

Recall that if we standardize the sample average of \( n \) observations, denoted by \( \bar{X}_n \), we obtain the Z-score

\[
Z = \frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}}
\]

which is normally distributed with mean 0 and variance 1.

In our example, we know \( \mu \), namely, \( \mu = 98.6^\circ F \). We don’t know the standard deviation \( \sigma \). If the sample size is large, however, we can replace the standard deviation \( \sigma \) by the sample standard deviation

\[
S_n = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2}
\]

You calculated this quantity in In-class Activity 1 using the Excel function STDEV and obtained \( S = 1.34 \).

Under the null hypothesis, with \( n = 130 \), our test statistic is therefore

\[
Z = \frac{98.2 - 98.6}{1.34 / \sqrt{130}} = -2.98
\]

3. Calculation of the corresponding \( p \)-value followed by decision of whether or not to reject the null hypothesis.

To find the \( p \)-value, recall that this is a two-sided test. We thus need to find out the probability that a standard normally distributed random variable is either less than -2.98 or greater than 2.98. This can be written as

\[
P(|Z| > 2.98) = 2(0.001441) = 0.002882 \approx 0.003
\]

In Excel, this number is calculated as “=2*NORMDIST(-2.98,0,1,TRUE).”
The p-value is therefore 0.003, which is highly significant and we reject the null hypothesis that the body temperature of healthy adults aged 18-40 is 98.6.

The authors of the research study write in their conclusion that “[T]hirty-seven degrees centigrade (98.6°F) should be abandoned as a concept relevant to clinical thermometry; 37.2°C (98.9°F) in the early morning and 37.7°C (99.9°F) overall should be regarded as the upper limit of the normal oral temperature range in healthy adults aged 40 years or younger, and several of Wunderlich’s other cherished dictums should be revised.”

In-class Activity 2: In the third column of the “BodyTemp” tab in the spreadsheet, the heart rate of each of the 130 individuals is listed. Develop a statistical test to determine whether the average heart rate is 72 beats per minute. Assume that heart rate is normally distributed.

Small Sample Size
In the large sample case, we assumed that the test statistic

\[
\frac{\bar{X}_n - \mu}{S_n / \sqrt{n}}
\]

is approximately normally distributed. This is a generally the case when the population is nearly normal and the sample size is large. When the sample size is not large, then, even if the population is nearly normal, the test statistic is no longer approximately normally distributed. The reason for this is that the sample standard deviation \(S_n\) is then no longer a good approximation for the population standard deviation \(\sigma\).

The distribution of the standardized test statistic, which we denote by \(T\), that is,

\[
T = \frac{\bar{X}_n - \mu}{S_n / \sqrt{n}}
\]

was calculated by the chemist W.S. Gossett. He published the result in 1908 in the journal *Bioometrika* under the name “Student.” Gossett worked for the Guinness Brewery and was not allowed to publish under his real name. This distribution is therefore known as the Student’s t-distribution and the test as the **Student’s t-test**. The t-distribution depends on the sample size and so there is a different distribution for each sample size. If the sample size is \(n\), then the appropriate distribution is the t-
distribution with \( n-1 \) degrees of freedom. The number of “degrees of freedom” of this distribution is one less than the sample size. Excel has a function “TDIST” that calculates the tail probability of the \( t \)-distribution. Namely, depending on whether the test is one-sided or two-sided,

\[
P(T > x) = \text{TDIST}(x, \text{degrees freedom}, 1) \\
P(|T| > x) = \text{TDIST}(x, \text{degrees freedom}, 2)
\]

To appreciate the difference between the \( t \)-distribution for small \( n \) and the standard normal distribution, we plot the density functions for both distributions in a single figure:

We see that the \( t \)-distribution has fatter tails than the normal distribution.

**Example:** The recommendation is that healthy adults should get 7 hours of sleep. 14 college students were asked to report the number of hours they slept before an exam. The following table lists the number of hours each student slept:

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>6.5</td>
<td>7.2</td>
<td>6.2</td>
<td>5.5</td>
<td>7.4</td>
<td>7.6</td>
</tr>
<tr>
<td>5.7</td>
<td>7.5</td>
<td>6.9</td>
<td>6.5</td>
<td>6.7</td>
<td>7.1</td>
</tr>
<tr>
<td>5.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Test whether the students slept less than recommended.
The null hypothesis is that the average is 7 hours and the alternative is that the students slept less than 7 hours:

\[ H_0 : \mu = 7 \]
\[ H_1 : \mu < 7 \]

The test statistic is the t-statistic and to calculate it, we need to compute the average and sample standard deviation of the data above. We find that the average number of hours of sleep is 6.6 hours and the sample standard deviation is 0.73. The test statistic is then

\[ t = \frac{6.6 - 7}{0.73 / \sqrt{14}} = -1.99 \]

The test is a one-sided test and we need to calculate the probability that the test statistic \( T \) is less than the value of the test statistic, \(-1.99\). Since the sample size is 14, the “degrees of freedom” is 13. Now, using Excel, we find

\[ P(T < -1.99) = P(T > 1.99) = \text{TDIST}(1.99,13,1) = 0.0340 \]

We therefore conclude that the result is highly significant and reject the null hypothesis.

**In-class Activity 3**

In a different study, 12 college students were asked on a Monday after Spring Break how many hours they had slept the night before. They reported the following data

<table>
<thead>
<tr>
<th>8.2</th>
<th>7.2</th>
<th>7.3</th>
<th>8.2</th>
<th>6.9</th>
<th>7.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1</td>
<td>6.8</td>
<td>6.9</td>
<td>7.3</td>
<td>6.7</td>
<td>7.1</td>
</tr>
</tbody>
</table>

Develop a statistical test to decide whether this group of students slept the recommended 7 hours per night.

**Large Sample versus Small Sample**

A sample is typically considered small if it is smaller than 30. That is, if the sample size is at least 30, the Z-score can be used instead of the t-statistic.
References


Homework

In Problems 1-4, calculate the probabilities

1. Assume that $Z$ is normally distributed with mean 0 and variance 1. Find
   a. $P(Z > 1.3)$
   b. $P(Z \leq 0.9)$
   c. $P(Z \geq -2.2)$
   d. $P(-1.4 < Z < 1.6)$
   e. $P(|Z| < 0.9)$
   f. $P(|Z| \geq 1.8)$

2. Assume that $X$ is normally distributed with mean 3 and variance 4. Find
   a. $P(X < 6)$
   b. $P(X > 1.5)$
   c. $P(|X - 3| > 1.9)$

3. Assume that $T$ is t-distributed with 9 degrees of freedom. Find
   a. $P(T > 0.3)$
   b. $P(T \leq 0.7)$
   c. $P(T \geq -0.2)$
   d. $P(-0.4 < T < 0.6)$
   e. $P(|T| < 1.4)$
   f. $P(|T| \geq 1.2)$

4. Assume that $T$ is t-distributed. Find
   a. $P(T < 2.4)$ if the degrees of freedom is 3
   b. $P(|T| < 1.9)$ if the degrees of freedom is 24
5. Assume that the population distribution is normal with mean 21 and standard deviation 3. Assume that a sample of size 26 is taken from this population. What is the probability that the sample average will exceed 22?
6. Assume that the population is normally distributed with mean 15. The standard deviation is not known. A sample of size 12 is taken. What is the likelihood that the sample average is less than 13 if the sample standard deviation is 2.8.
7. Set up a spreadsheet that would allow you to plot the cumulative distribution functions of a normal distribution with mean 0 and variance 1 and the t-distribution with k degrees of freedom in a single graph. Produce such a graph when (a) k=3, (b) k=10, and (c) k=30.
8. In a clinical trial, glucose levels of 45 patients were measured. The average glucose level in this sample was 93 mg/dl and the sample standard deviation was 14 mg/dl. Test whether this group has a higher glucose level than a comparison group where the glucose level is 90 mg/dl. Assume that glucose levels are approximately normally distributed.
9. In a clinical trial, glucose levels of 15 patients were measured. The average glucose level in this sample was 95 mg/dl and the sample standard deviation was 12 mg/dl. Test whether this group has a higher glucose level than a comparison group where the glucose level is 90 mg/dl. Assume that glucose levels are approximately normally distributed.
10. You live on a street where the speed limit was reduced from 30 mph to 25 mph. You suspect that drivers ignore the new speed limit and decide to measure the speed of passing cars. You collect data on 12 cars and find that they were passing you at the following speed (in mph): 27, 31, 24, 29, 22, 23, 29, 31, 33, 35, 27, 25, and 28. Does your data indicate a reduction in speed? Determine the power of your test. How big a sample would you need if you wanted the power of the test to be 0.9?
11. Age-related memory impairment is studied in Drosophila. In a study by Tamura et al. (2003), Drosophila were exposed to the odor of MCH (4-methylcyclohexanol) and simultaneously given electric shocks to trigger avoidance when exposed to the odor. To test whether Drosophila remembered the experience, they are put in a T-maze, where one arm of the maze has the odor and the other arm does not. Drosophila who remember the odor will avoid the arm with the odor, those who do not, will randomly choose one arm or the other of the maze. A performance index (PI) can be calculated, which is equal to 0 if the Drosophila choose randomly and equal to 100 if they always avoid the arm with the odor. Ten flies participate in the experiment at ages 1 day, 10 days, and 20 days. The table on the right summarizes the outcome of the experiment (SEM is the standard error of the mean and is defined as the sample standard deviation divided by the square root of the sample size). Test whether 10-day old and 20-day old Drosophila are significantly different from the performance index 57 observed in 1-day old Drosophila. Assume normality of the PI score.

<table>
<thead>
<tr>
<th>Age</th>
<th>PI (mean PI ± SEM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 day</td>
<td>57 ± 4</td>
</tr>
<tr>
<td>10 days</td>
<td>51 ± 7</td>
</tr>
<tr>
<td>20 days</td>
<td>36 ± 8</td>
</tr>
</tbody>
</table>

Citation: Neuhauser, C. Testing Means.
Created: December 5, 2009  Revisions:
Copyright: © 2009 Neuhauser. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial Share Alike License, which permits unrestricted use, distribution, and reproduction in any medium, and allows others to translate, make remixes, and produce new stories based on this work, provided the original author and source are credited and the new work will carry the same license.
Funding: This work was partially supported by a HHMI Professors grant from the Howard Hughes Medical Institute.