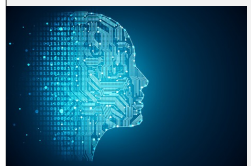


Data Science for Researchers and Scholars

Vasant G. Honavar
Dorothy Foehr Huck and J. Lloyd Huck Chair in Biomedical Data Sciences and Artificial Intelligence
Professor of Data Sciences, Informatics, Computer Science and Engineering, Bioinformatics & Genomics, Public Health Sciences and Neuroscience
Director, Center for Artificial Intelligence Foundations and Scientific Applications
Associate Director, Institute for Computational and Data Sciences
Pennsylvania State University

vhonavar@psu.edu
http://faculty.ist.psu.edu/vhonavar
http://ailab.ist.psu.edu

Estimation in the small sample setting

- When the sample size is small, the "large sample" estimation procedures are not appropriate.
- While point estimators remain the same, we have different confidence intervals for
  - ✓ μ, the mean of a normal population
  - ✓ $\mu_1 - \mu_2$, the difference between two normal population means

Sampling Distributionof the Sample Mean

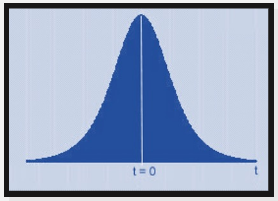
- When we take a sample from a normal population, the sample mean x has a normal distribution for any sample size n, and $Z = \frac{x-\mu}{\sigma/\sqrt{n}}$ has a standard normal distribution.

- But if σ is unknown, which is the case when we are estimating the population parameters from a sample, and we must use the sample standard deviation s to estimate it, the resulting statistic is not normal.
- The distribution of $\frac{x-\mu}{s/\sqrt{n}}$ is not normal although that of $\frac{x-\mu}{\sigma/\sqrt{n}}$ is.

Student’s t Distribution

- Fortunately, the sampling distribution of the statistic $t = \frac{x-\mu}{s/\sqrt{n}}$ is well known to statisticians, and is called the Student’s t distribution, with $n-1$ degrees of freedom.
- We can use this distribution to create estimation procedures for the population mean μ.

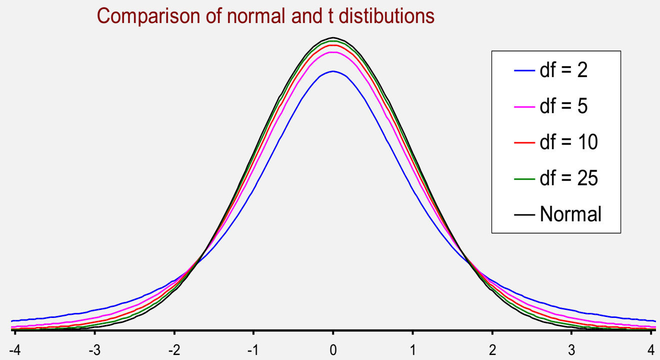
Properties of Student's t distribution



- Mound-shaped and symmetric about 0.
- More variable than z, with "heavier tails"

- Shape depends on the sample size n or the degrees of freedom, $n-1$.
- As n increases the shapes of the t and z distributions become almost identical.

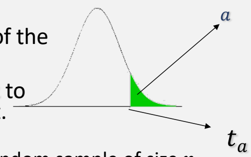
t distribution



df $=$ degrees of freedom $= n-1$

Using the t-Table

- Table gives the values of t that cut off the desired probability mass from the right tail of the t distribution.
- Use index *df* and the appropriate tail area a to find t_a, the value of t with area a to its right.



df	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$
1	3.078	6.314	12.706	31.821
2	1.886	2.920	4.303	6.965
3	1.638	2.353	3.182	4.541
4	1.533	2.132	2.776	3.747
5	1.476	2.015	2.571	3.365
6	1.440	1.943	2.447	3.143
7	1.415	1.895	2.365	2.998
8	1.397	1.860	2.306	2.896
9	1.383	1.833	2.262	2.821
10	1.372	1.812	2.228	2.764
11	1.363	1.796	2.201	2.718
12	1.356	1.782	2.179	2.681
13	1.350	1.771	2.160	2.650
14	1.345	1.761	2.145	2.624
15	1.341	1.753	2.131	2.602

For a random sample of size $n = 10$, find a value of t that cuts .025 off the right tail.

Row $= df = n - 1 = 9$

Column subscript $= a = .025$

$t_{.025} = 2.262$

Small Sample Confidence Interval for Population Mean

Small-Sample $(1-\alpha)100\%$ confidence interval of the population mean μ is

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

where $t_{\alpha/2}$ is the value of t that cuts off area $\alpha/2$ in the right tail of a t-distribution with $df = n-1$.

Assumption: population must be normal

Example

- Ten randomly selected students were each asked to list how many hours of television they watched per month. The results are

82	66	90	84	75
88	80	94	110	91

- Find a 90% confidence interval for the true mean number of hours of television watched per month by students.
  - Sample mean $x = 86$
  - Sample standard deviation $s = 11.842$
  - To find the 90% confidence interval around the mean
    - Find the critical value of t in row corresponding to $df = n - 1 = 9$ and column labeled $t_{0.05}$
    - Why? To find the 90% confidence interval, we need to exclude 5% of the area under the curve on either side

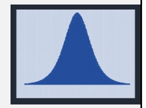
Example Continued

Hence the confidence interval around μ is given by

$$x \pm t\frac{s}{\sqrt{n}} = 86 \pm 1.833\frac{11.842}{\sqrt{10}}$$

$= 86 \pm 6.86$ or (79.14, 92.86)

Estimating the Difference between Two Means



- You can also create a $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$.

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

with $s^2 = \dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}$

and $t_{\alpha/2}$ is the critical value of t with degrees of freedom $n_1 + n_2 - 2$

- Remember the three assumptions:
- Original populations are normally distributed
- Samples are random and independent
- Equal population variances $\sigma_1^2 = \sigma_2^2$

Example

- A student recorded the mileage mileage given by his car while driving to school in his car.
- He kept track of the mileage for twelve different tanks of fuel, involving gasoline of two different octane ratings.
- Compute the 95% confidence interval for the difference of mean mileages.
- Assume that the variances of the two mileage distributions are identical, the mileages are normally distributed, and data represent IID samples

87 Octane	90 Octane
26.4, 27.6, 29.7	30.5, 30.9, 29.2
28.9, 29.3, 28.8	31.7, 32.8, 29.3

Example

Let 87 octane fuel be the first group and 90 octane fuel the second group, so we have

- $n_1 = n_2 = 6$
- $df = n_1 + n_2 - 2 = 6 + 6 - 2 = 10$
- $\overline{x_1} = 28.45, s_1 = 1.228$
- $\overline{x_2} = 30.73, s_2 = 1.392$
- $1 - \alpha = 0.95$, so $\alpha = 0.025$
- Critical value of $t = t_{.025} = 2.228$

df	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$
1	3.078	6.314	12.706	31.821
2	1.886	2.920	4.303	6.965
3	1.638	2.353	3.182	4.541
4	1.533	2.132	2.776	3.747
5	1.476	2.015	2.571	3.365
6	1.440	1.943	2.447	3.143
7	1.415	1.895	2.365	2.998
8	1.397	1.860	2.306	2.896
9	1.383	1.833	2.262	2.821
10	1.372	1.812	2.228	2.764
11	1.363	1.796	2.201	2.718
12	1.356	1.782	2.179	2.681
13	1.350	1.771	2.160	2.650
14	1.345	1.761	2.145	2.624
15	1.341	1.753	2.131	2.602

Example (contd.)

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{5 \times 1.508 + 5 \times 1.938}{10} = 1.723$$

$$\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2}\sqrt{s^2(1/n_1 + 1/n_2)}$$

$$= 28.45 - 30.73 \pm 2.228\sqrt{1.723 \times (1/5 + 1/5)}$$

$$= -2.28 \pm 1.849$$

The 95% confidence interval is (-4.129,-.431).

Interpretation: Average mileage of 87 octane fuel is worse than that of 90 octane fuel

Summary: Small sample Point Estimation

- For normal population, $(1-\alpha)100\%$ confidence interval of the population mean μ is given by

$$x \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$$

where $t_{\alpha/2}$ is the value of t that cuts off area ${}^{\alpha}/_{2}$ in the right tail of the t-distribution with degree of freedom $n-1$

Summary: Small Sample Difference of Means

- Assuming
  - Both populations are normal.
  - IID samples of the two populations
  - Two populations have a common variance
- The $(1-\alpha)100\%$ confidence interval of $\mu_1 - \mu_2$ is

$$\overline{x_1} - \overline{x_2} \pm t_{\alpha/2} s \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

- where $t_{\alpha/2}$ is the value of t that cuts off area $\alpha/2$ in the right tail of the t-distribution with degree of freedom $n-1$

Hypothesis testing

- Hypothesis testing is the process of determining whether the available data justifies rejecting the null hypothesis

Testing Hypotheses (against data)

- Hypothesis testing refers to the process of using statistical analysis to determine whether the differences between observed and hypothesized values of some statistics are due to random chance or to true differences in the samples.
  - Observed differences in mean A1c measures between groups of diabetes patients treated with Metformin as opposed to Placebo
  - All hypothesis tests have unavoidable, but quantifiable, risks of making the wrong conclusion

Example

- Suppose a vaccine maker claims that the mean efficacy μ of a COVID vaccine exceeds the minimum required for FDA approval.

FDA need to decide between two possibilities:

  - Null Hypothesis: The mean potency μ does not exceed the required minimum efficacy.
  - Alternative Hypothesis: The mean potency μ exceeds the required minimum efficacy.

- FDA needs to perform a hypothesis test based on the data presented to it

Five-Step Recipe for Hypothesis Testing

- Specify statistical hypotheses
  - null hypothesis H_0 and
  - alternative hypothesis H_a

  in terms of population parameters
- Identify and calculate the test statistic
- Identify distribution and find p-value
- Decide whether to reject or not to reject the null hypothesis
- State conclusion

Null and Alternative Hypothesis

- The null hypothesis, H_0
  - The hypothesis we wish to falsify
  - Assumed to be true until we can prove otherwise.

The alternative hypothesis, H_a
  - The hypothesis we wish to establish

Examples
  - Court Trial
    - H_0: innocent
    - H_a : guilty
  - Vaccine Trial
    - H_0: Mean vaccine efficacy μ does not exceed the required efficacy
    - H_a: μ exceeds required efficacy

Examples of Hypotheses

- You would like to determine if the Calc 1 grades have a mean of 3.0
  - H_0: $\mu = 3.0$
  - H_a: $\mu \neq 3.0$

  (Two-sided or two tailed alternative)

- Do the “12 ounce” cans of soda meet the claim on the label (on average)?
- The real concern has to do with reputational damage from cheating the customer
  - H_0: $\mu \geq 12$
  - H_a: $\mu < 12$

  (One-sided or one tailed alternative)

Remarks on Setting up the Hypotheses

- The distribution of test statistic requires the null hypothesis to be assumed to be true
- The alternate hypothesis should be what you are really attempting to show to be true.

  This is not always possible.

There are two possible decisions:

- reject the null hypothesis
- fail to reject the null hypothesis.
- Note that we say “fail to reject” rather than “accept” the null hypothesis.

Two Types of Errors

There are two types of errors which can occur in a statistical test:

- Type I error: reject the null hypothesis when it is true
- Type II error: fail to reject the null hypothesis when it is false

Actual Fact / Jury's Decision	Guilty	Innocent
Guilty	Correct	Error
Innocent	Error	Correct

Actual Fact / Your Decision	**H_0 true**	**H_0 false**
Fail to reject H_0	Correct	Type II Error
Reject H_0	Type I Error	Correct

Analogy

- Consider a COVID test where the hypotheses are equivalent to
  - H_0: the patient has COVID
  - H_a: the patient doesn't have COVID

In this setting,

- Type I error is equivalent to a false negative
  - Claiming that the patient does not have COVID when in fact, he does.)
- Type II error is equivalent to a false positive
  - Claiming that the patient has COVID when, in fact, he does not.

Two Types of Errors

Define:

- $\alpha = P(\text{Type I error}) = P(\text{Reject H}_0 \text{ when H}_0 \text{ is true})$
- $\beta = P(\text{Type II error}) = P(\text{Fail to reject H}_0 \text{ when H}_0 \text{ is false})$

- We want to keep the both α and β as small as possible.
- The value of α is controlled by the experimenter and is called the significance level.
- Generally, with all else held constant, decreasing one type of error causes the other to increase.

Balance Between α and β

- The only way to decrease both types of error simultaneously is to increase the sample size.
- No matter what decision is reached, there is always the risk of one of these errors.
- Balance:
  - Identify the largest significance level α as the maximum tolerable risk you want to have of making a type I error.
  - Employ a test procedure that makes type II error β as small as possible while maintaining type I error smaller than the given significance level α.

Test Statistic

- A test statistic is a quantity calculated from sample of data. Its value is used to decide whether or not the null hypothesis should be rejected.
- The choice of a test statistic will depend on the assumed probability model and the hypotheses under question.
- We then find sampling distribution of the test statistic and calculate the probability of rejecting the null hypothesis (**type I error**) if it is in fact true. This probability is called the p-value

P-value

- The p-value is a measure of inconsistency between the hypothesized value under the null hypothesis and the observed sample.
- The p-value is the probability, assuming that H_0 is true, of obtaining a test statistic value at least as inconsistent with H_0 as actually obtained.
- The p-value measures whether the test statistic is likely or unlikely, assuming H_0 is true.
  - Small p-values suggest that the null hypothesis is unlikely to be true.
  - The smaller it is, the more convincing is the rejection of the null hypothesis.
  - It indicates the strength of evidence for rejecting the null hypothesis H_0

Making Decisions

A decision as to whether H_0 should be rejected results from comparing the p-value to the chosen significance level α:

- H_0 should be rejected if p-value $\leq \alpha$.
- H_0 should not be rejected if p-value $> \alpha$.

- When p-value>α, say "fail to reject H_0" "there is insufficient evidence to reject H_0" rather than "accept H_a".
- Another way to make this decision is to use critical value and rejection region which we will not cover in this course

Five-Step Recipe for Hypothesis Testing

- Specify statistical hypotheses
  - null hypothesis H_0 and
  - alternative hypothesis H_a

  in terms of population parameters
- Identify and calculate the test statistic
- Identify distribution and find p-value
- Decide whether to reject or not to reject the null hypothesis
- State conclusion

Large Sample Test for Population Mean

Step 1: Specify the null and alternative hypothesis

- $H_0: \mu = \mu_0$ **versus** $H_a: \mu \neq \mu_0$ (two-sided test)
- $H_0: \mu = \mu_0$ **versus** $H_a: \mu > \mu_0$ (one-sided test)
- $H_0: \mu = \mu_0$ **versus** $H_a: \mu < \mu_0$ (one-sided test)

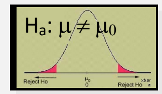
Step 2: Test statistic for large sample ($n \geq 30$)

$$z = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

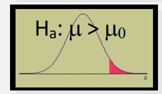
where n, $\bar{x}$ and s are sample size, mean and standard deviation

Intuition behind the Test Statistic

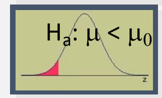
- If H_0 is true, the value of x should be close to μ_0, and z will be close to 0.
- If H_0 is false, x will be much larger or smaller than μ_0, and z will be much larger or smaller than 0, indicating that we should reject H_0.
  - z being much larger or smaller than 0 provides evidence against H_0



  - z being much larger than 0 provides evidence against H_0



  - z is much smaller than 0 provides evidence against H_0



Large Sample Test for Population Mean

Step 3: When n is large, the sampling distribution of z will be approximately standard normal under H_0.

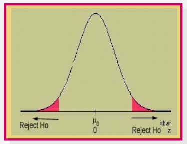
Compute sample statistic $$z^* = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

- z is defined for any possible sample.
  - Hence z is a random variable which can take many different values
  - The sampling distribution tells us the probability of each value.
- z^* is computed from the given data sample, and hence, a fixed number.

Large Sample Test for Population Mean

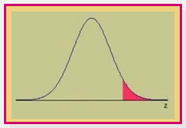
H_a: $\mu \neq \mu_0$ (two-sided test)

$$\text{p-value} = P(z < -|z^*|) + P(z > |z^*|) = 2P(z > |z^*|)$$



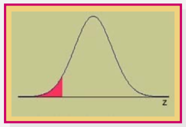
H_a: $\mu > \mu_0$ (one-sided test)

$$\text{p-value} = P(z > z^*)$$



H_a: $\mu < \mu_0$ (one-sided test)

$$\text{p-value} = P(z < z^*)$$



Example

- The daily yield for a chemical plant has averaged 880 tons for several years.
- The quality control manager wants to know if this average has changed.
- She randomly selects 50 days and records an average yield of 871 tons with a standard deviation of 21 tons.
- Conduct the test using α=.05.

$$H_0 : \mu = 880 \text{ vs } H_a : \mu \neq 880$$

Test statistic :

$$\mu_0 = 880, n = 50, \bar{x} = 871, s = 21$$

$$z^* = \frac{\bar{x} - \mu_0}{s / \sqrt{n}} = \frac{871 - 880}{21 / \sqrt{50}} = -3.03$$

Example - continued

p - value : this is a two - sided test

$$p - \text{value} = 2 \times P(z > 3.03) = 2(.0012) = .0024$$

- Decision: since p-value<α, we reject the hypothesis that μ=880.
- Conclusion: the average yield has changed and the change is statistically significant at level .05.
- What more does the the p-value tell us?:
  - The null hypothesis is very unlikely to be true.
  - If the significance level is set to be any value greater or equal to .0024, we would still reject the null hypothesis.
  - Thus, another interpretation of the p-value is the smallest level of significance at which H_0 would be rejected, and p-value is also called the observed significance level.

Example

- A homeowner randomly samples 64 homes similar to her own and found that the average selling price is $252,000 with a standard deviation of $15,000.
- Is this sufficient evidence to conclude that the average selling price is greater than $250,000? Use α = .01.

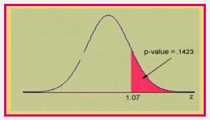
$$\text{H}_0 : \mu = 250{,}000 \quad \text{vs} \quad \text{H}_\text{a} : \mu > 250{,}000$$

Test statistic:

$$z^* = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{252{,}000 - 250{,}000}{15{,}000/\sqrt{64}} = 1.07$$

Example - continued

$$p\text{-value}:$$
$$\text{this is a one-sided test}$$
$$p-\text{value} = P(z > 1.07)$$
$$= .5 - .3577 = .1423$$



- Decision: Since the p-value is greater than $\alpha = .01$, H_0 is not rejected.
- Conclusion: There is insufficient evidence to indicate that the average selling price is greater than $250,000.

Small Sample Test for Population Mean

Step 1: Specify the null and alternative hypothesis

- **H_0: $\mu = \mu_0$** versus **H_a: $\mu \neq \mu_0$** (two-sided test)
- **H_0: $\mu = \mu_0$** versus **H_a: $\mu > \mu_0$** (one-sided test)
- **H_0: $\mu = \mu_0$** versus **H_a: $\mu < \mu_0$** (one-sided test)

Step 2: Test statistic for small sample

$$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

where n, $\bar{x}$ and s are sample size, mean and standard deviation

Step 3: When samples are from a normal population, under H_0, the sampling distribution of t has a Student's t distribution with $n - 1$ degrees of freedom

Small Sample Test for Population Mean

Step 3: Find p-value.

Compute sample statistic $t^* = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$

H_a: $\mu \neq \mu_0$ (two-sided test) $\quad \text{p - value} = 2P(t > |t^*|)$

H_a: $\mu > \mu_0$ (one-sided test) $\quad \text{p - value} = P(t > t^*)$

H_a: $\mu < \mu_0$ (one-sided test) $\quad \text{p - value} = P(t < t^*)$

Example

- A sprinkler system is designed so that the average time for the sprinklers to activate after being turned on is no more than 15 seconds.
- A test of 5 systems gave the following times: 17, 31, 12, 17, 13, 25
- Is the system working as specified? Test using $\alpha = .05$.

$$H_0 : \mu = 15 \text{ (working as specified)}$$
$$H_a : \mu > 15 \text{ (not working as specified)}$$

Example

Data: 17, 31, 12, 17, 13, 25

Calculate the sample mean and standard deviation.

$$\bar{x} = \frac{\sum x_i}{n} = \frac{115}{6} = 19.167$$

$$s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}} = \sqrt{\frac{2477 - \frac{115^2}{6}}{5}} = 7.387$$

Test statistic:

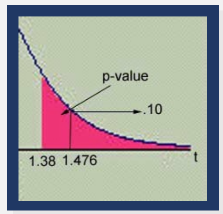
$$t^* = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{19.167 - 15}{7.387/\sqrt{6}} = 1.38$$

Degrees of freedom:

$$df = n - 1 = 6 - 1 = 5$$

Approximating the p-value

- Since the sample size is small, we need to assume a normal population and use *t* distribution.
- We can only approximate the *p*-value for the test using the Table.



df	$t_{.100}$	$t_{.050}$
1	3.078	6.314
2	1.886	2.920
3	1.638	2.353
4	1.533	2.132
5	1.476	2.015

Since the observed value of $t^* = 1.38$ is smaller than $t_{.10} = 1.476$, *p*-value > .10.

Example

- Decision: since the *p*-value is greater than .1, than it is greater than α = .05, H_0 is not rejected.
- Conclusion: there is insufficient evidence to indicate that the average activation time is greater than 15 seconds.

Exact p-values can be calculated by computers.

Large Sample Test for Difference Between Two Population Means

A random sample of size n_1 drawn from population 1 with mean μ_1 and variance σ_1^2.

A random sample of size n_2 drawn from population 2 with mean μ_2 and variance σ_2^2.

Large Sample Test for Difference Between Two Population Means

Step 1: Specify the null and alternative hypothesis

- $\mathbf{H_0: \mu_1 - \mu_2 = D_0}$ versus $\mathbf{H_a: \mu_1 - \mu_2 \neq D_0}$ (two-sided test)
- $\mathbf{H_0: \mu_1 - \mu_2 = D_0}$ **versus** $\mathbf{H_a: \mu_1 - \mu_2 > D_0}$ (one-sided test)
- $\mathbf{H_0: \mu_1 - \mu_2 = D_0}$ **versus** $\mathbf{H_a: \mu_1 - \mu_2 < D_0}$ (one-sided test)

where D_0 is some specified difference that you wish to test. $D_0 = 0$ when testing no difference.

Large Sample Test for Difference Between Two Population Means

- Step 2: Test statistic for large sample sizes when $n_1 \geq 30$ and $n_2 \geq 30$

$$z = \frac{\bar{x}_1 - \bar{x}_2 - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- Step 3: Under H_0, the sampling distribution of z is approximately standard normal

Large Sample Test for Difference Between Two Population Means

Step 3: Find p-value. Compute

$$z^* = \frac{\bar{x}_1 - \bar{x}_2 - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

H_a: $\mu_1 - \mu_2 \neq D_0$ (two-sided test) $\text{p-value} = 2P(z > |z^*|)$

H_a: $\mu_1 - \mu_2 > D_0$ (one-sided test) $\text{p-value} = P(z > z^*)$

H_a: $\mu_1 - \mu_2 < D_0$ (one-sided test) $\text{p-value} = P(z < z^*)$

Example

Avg Daily Intakes	Men	Women
Sample size	50	50
Sample mean	756	762
Sample Std Dev	35	30

- Is there a difference in the average daily intakes of dairy products for men versus women? Use a = .05.

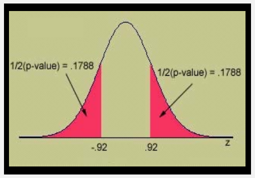
$$H_0 : \mu_1 - \mu_2 = 0 \text{ (same)} \qquad H_a : \mu_1 - \mu_2 \neq 0 \text{ (different)}$$

Test statistic :

$$z^* = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \qquad = \frac{756 - 762 - 0}{\sqrt{\frac{35^2}{50} + \frac{30^2}{50}}} = -.92$$

Example - continued

$$p\text{-value: two-sided test}$$
$$p\text{-value} = 2P(z > .92)$$
$$= 2(.1788) = .3576$$



Decision: since the *p*-value is greater than $\alpha = .05$, H_0 is not rejected.

Conclusion: there is insufficient evidence to indicate that men and women have different average daily intakes.

Small Sample Testing the Difference between Two Population Means

A random sample of size n_1 drawn from population 1 with normal distribution with mean μ_1 and variance σ^2.

A random sample of size n_2 drawn from population 2 with normal distribution with mean μ_2 and variance σ^2.

Note that both population are normally distributed with the same variances

Small Sample Testing the Difference between Two Population Means

Step 1: Specify the null and alternative hypothesis

- **H_0: $\mu_1 - \mu_2 = D_0$** versus **H_a: $\mu_1 - \mu_2 \neq D_0$** (two-sided test)
- **H_0: $\mu_1 - \mu_2 = D_0$** versus **H_a: $\mu_1 - \mu_2 > D_0$** (one-sided test)
- **H_0: $\mu_1 - \mu_2 = D_0$** versus **H_a: $\mu_1 - \mu_2 < D_0$** (one-sided test)

where D_0 is some specified difference that you wish to test. $D_0=0$ when testing no difference.

Small Sample Testing the Difference between Two Population Means

Step 2: Test statistic for small sample sizes

$$t = \frac{\bar{x}_1 - \bar{x}_2 - D_0}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where s^2 is calculated as $s^2 = \frac{(n_1 - 1)s_1^{\,2} + (n_2 - 1)s_2^{\,2}}{n_1 + n_2 - 2}$

Step 3: Under H_0, the sampling distribution of t has a Student's t distribution with $n_1 + n_2 - 2$ degrees of freedom

Small Sample Testing the Difference between Two Population Means

Step 3: Find p-value. Compute

$$t^* = \frac{\bar{x}_1 - \bar{x}_2 - D_0}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

H_a: $\mu_1 - \mu_2 \neq D_0$ (two-sided test) $\quad \text{p-value} = 2P(t > |t^*|)$

H_a: $\mu_1 - \mu_2 > D_0$ (one-sided test) $\quad \text{p-value} = P(t > t^*)$

H_a: $\mu_1 - \mu_2 < D_0$ (one-sided test) $\quad \text{p-value} = P(t < t^*)$

Example

Two training procedures are compared by measuring the time that it takes trainees to assemble a device. A different group of trainees are taught using each method. Is there a difference in the two methods? Use a = .01.

Time to Assemble	Method 1	Method 2
Sample size	10	12
Sample mean	35	31
Sample Std Dev	4.9	4.5

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_a: \mu_1 - \mu_2 \neq 0$$

Test statistic:

$$t = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Example

Time to Assemble	Method 1	Method 2
Sample size	10	12
Sample mean	35	31
Sample Std Dev	4.9	4.5

Calculate:

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

$$= \frac{9(4.9^2) + 11(4.5^2)}{20} = 21.942$$

Test statistic:

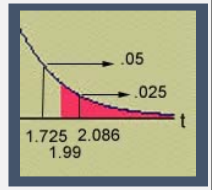
$$t^* = \frac{35 - 31}{\sqrt{21.942\left(\frac{1}{10} + \frac{1}{12}\right)}}$$

$$= 1.99$$

Example

$$p\text{-value: two-sided} \Rightarrow p\text{-value} = 2P(t > 1.99)$$
$$.025 < P(t > 1.99) < .05 \Rightarrow .05 < p - \text{value} < .1$$

$$df = n_1 + n_2 - 2$$
$$= 10 + 12 - 2 = 20$$



df	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	df
19	1.328	1.729	2.093	2.539	2.861	19
20	1.325	1.725	2.086	2.528	2.845	20

- Decision: since the *p*-value is greater than $\alpha = .01$, H_0 is not rejected.
- Conclusion: there is insufficient evidence to indicate a difference in the population means.

The Paired-Difference Test

- We have assumed that samples from two populations are independent.
- Sometimes the assumption of independent samples is intentionally violated, resulting in a matched-pairs or paired-difference test.
- By designing the experiment in this way, we can eliminate unwanted variability in the experiment
- Data:

Pair	1	2	...	n
Population 1	x_{11}	x_{12}	...	x_{1n}
Population 2	x_{21}	x_{22}	...	x_{2n}
Difference	$d_1 = x_{11} - x_{21}$	$d_2 = x_{12} - x_{22}$	...	$d_n = x_{1n} - x_{2n}$

The Paired-Difference Test

Step 1: Specify the null and alternative hypothesis

- **H_0: $\mu_1 - \mu_2 = D_0$ versus H_a: $\mu_1 - \mu_2 \neq D_0$** (two-sided test)
- **H_0: $\mu_1 - \mu_2 = D_0$ versus H_a: $\mu_1 - \mu_2 > D_0$** (one-sided test)
- **H_0: $\mu_1 - \mu_2 = D_0$ versus H_a: $\mu_1 - \mu_2 < D_0$** (one-sided test)

where D_0 is some specified difference that you wish to test. $D_0=0$ when testing no difference.

The Paired-Difference Test

Step 2: Test statistic for small sample sizes

$$t = \frac{\bar{d} - D_0}{s_d / \sqrt{n}}$$

where $\bar{d}$ and s_d are sample mean and standard deviation of the paired differences $d_1, \cdots, d_n$

Step 3: Under H_0, the sampling distribution of t has a Student's t distribution with n-1 degrees of freedom

The Paired-Difference Test

Step 3: Find p-value. Compute $t^* = \frac{\bar{d} - D_0}{s_d / \sqrt{n}}$

H_a: $\mu_1 - \mu_2 \neq D_0$ (two-sided test) $\text{p-value} = 2P(t > |t^*|)$

H_a: $\mu_1 - \mu_2 > D_0$ (one-sided test) $\text{p-value} = P(t > t^*)$

H_a: $\mu_1 - \mu_2 < D_0$ (one-sided test) $\text{p-value} = P(t < t^*)$

Paired t Test Example

- A weight reduction center advertises that participants in its program lose an average of at least 5 pounds during the first week of the participation.
- Because of numerous complaints, the state's consumer protection agency doubts this claim.
- To test the claim at the 0.05 level of significance, 12 participants were randomly selected.
- Their initial weights and their weights after 1 week in the program appear on the next slide.
- Set up and perform an appropriate hypothesis test.

Paired Sample Example continued

Member	Initial Weight	One Week Weight	Difference Initial -1week
1	195	195	0
2	153	151	2
3	174	170	4
4	125	123	2
5	149	144	5
6	152	149	3
7	135	131	4
8	143	147	-4
9	139	138	1
10	198	192	6
11	215	211	4
12	153	152	1

Paired Sample Example continued

- Each member serves as his/her own pair.
- weight changes=initial weight–weight after one week

$$H_0: \mu_1 - \mu_2 = 5 \text{ (working as claimed)}$$
$$H_a: \mu_1 - \mu_2 < 5 \text{ (not working as claimed)}$$

$$n = 12, \quad \bar{d} = 2.333, \quad s_d = 2.674$$
$$t^* = \frac{\bar{d} - D_0}{s_d / \sqrt{n}} = \frac{2.333 - 5}{2.674 / \sqrt{12}} = -3.45$$
$$df = n - 1 = 12 - 1 = 11$$

Paired Sample Example

continued

$$p\text{-value: one-sided} \Rightarrow p\text{-value} < .005$$

- Decision: since the *p*-value is smaller than $\alpha = .05$, H_0 is rejected.
- Conclusion: there is strong evidence that the mean weight loss is less than 5 pounds for those who took the program for one week.

Five-Step Recipe for Hypothesis Testing

- Specify statistical hypotheses
  - null hypothesis H_0 and
  - alternative hypothesis H_a

  in terms of population parameters
- Identify and calculate the test statistic
- Identify distribution and find p-value
- Decide whether to reject or not to reject the null hypothesis
- State conclusion

Summary: Statistical Hypothesis testing

Errors and Statistical Significance

- Type I error: reject the null hypothesis when it is true
- Type II error: fail to reject the null hypothesis when it is false
- The significance level α = P(type 1 error) and β =P(type 2 error)
- The *p*-value is the probability of observing a test statistic as extreme as or more than the one observed; also, the smallest value of α for which H_0 can be rejected
- When the *p*-value is less than the significance level α , the null hypothesis is rejected

Tests

- Tests for population mean
- Tests for difference between population means