Chapter 8: Hypothesis Testing

ST2334 Probability and Statistics[1]
(Academic Year 2014/15, Semester 1)

Department of Statistics & Applied Probability (DSAP)
National University of Singapore (NUS)[1]

Outline

Introduction I

Learning Outcomes

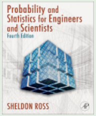
Questions to Address: ♦ Asymmetry between the null & the alternative hypothesis ♦ Difference between point estimation & hypothesis testing ♦ How to obtain the critical region for tests for various population settings ♦ Analog between c.i. & hypothesis test ♦ How to use c.i. to perform a 2-sided test ♦ Meaning & computation of the p-value
♦ How to test the mean of a normal population when the population variance is known or unknown ♦ How to test equality of means of two independent normal populations ♦ Assumption & application of the paired t-test

Introduction II

Learning Outcomes (continued)

Concept & Terminology: ♦ hypothesis ♦ status quo ♦ hypothesis test ♦ null/alternative hypothesis ♦ simple/composite hypothesis ♦ critical/rejection/acceptance region ♦ level of significance/significance level ♦ type I/II error ♦ 2-sided/1-sided test ♦ test statistic ♦ observed value of the test statistic ♦ *p*-value ♦ paired *t*-test for dependent samples

Mandatory Reading

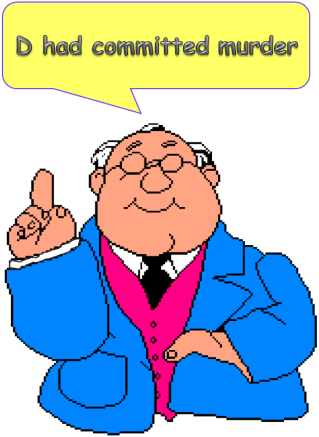


✏ Section 8.1 – Section 8.4 (except operation characteristic curve, power function, & any computation related to type II error)

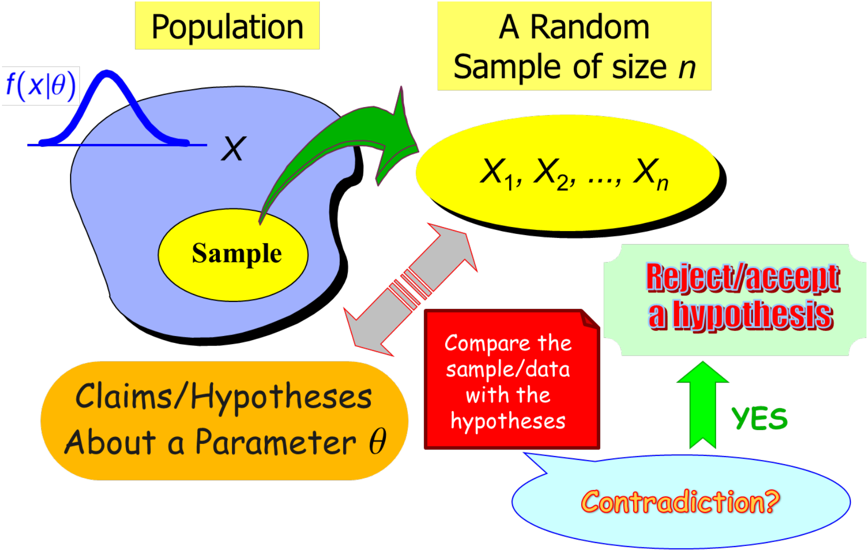
Example 1: A Criminal Trial

Someone was killed in a murder case. A person, called the defendant (D), is accused to have committed murder

- A hypothesis: a claim/an assertion
- A status quo: D is presumed to be innocent (*i.e.,* presumption of innocence)
- To verify this claim: Present evidence to show that D did indeed murder someone (*i.e.,* burden of proof)
- To conclude: The jury decides if they are convinced of guilty beyond a reasonable doubt. If they find there is not enough evidence, the case must be dismissed & D is deemed to be innocent



Hypothesis Testing: A Modification of Parameter Estimation I



Hypothesis Testing: A Modification of Parameter Estimation II

Under the *same set-up* as in (Ch.7) that a population of interest is modeled by the underlying distribution with unknown parameter $\theta \in \Theta$, rather than wishing to explicitly estimate the unknown parameter, we may be interested in *another inferential procedure* called

hypothesis testing

Questions to be addressed in a *hypothesis test*:

1. What are the *hypotheses about a parameter* θ?
2. How to *compare the sample/data with the hypotheses*?
3. According to *which criterion* do we say that there is any *contradiction* between the sample & the hypotheses?
4. What *conclusion* can be drawn – reject/accept a hypothesis?

The Null And The Alternative Hypotheses I

Definition

In a hypothesis test, there are 2 complementary hypotheses called the *null hypothesis* & the *alternative hypothesis*, denoted by H_0 & H_1, respectively

- The general formats of H_0 & H_1 are

$$H_0 : \theta \in \Theta_0 \qquad \& \qquad H_1 : \theta \in \Theta_1$$

where $\Theta_0, \Theta_1 \subset \Theta$ do *not overlap* (*i.e.,* $\Theta_0 \cap \Theta_1 = \varnothing$)

- *Equivalent* to investigating from which prob model

$$\{f(x|\theta)|\theta \in \Theta_0\} \qquad \text{or} \qquad \{f(x|\theta)|\theta \in \Theta_1\}$$

the sample/data are generated

The Null And The Alternative Hypotheses II

- *Asymmetry between H_0 & H_1*:
  1. H_0 is usually set as the *status quo*, *i.e.,* a claim/norm believed to be true by most people without any data/justification
  2. H_0 always contains the "$=$" sign, *e.g.,* $H_0 : \theta = \theta_0$
  3. We have a *non–trivial claim/assertion about θ which we wish it to be correct* & set it as H_1
  4. H_1 must be complementary to H_0 & is non–trivial (innovative/outstanding/totally different) compared with H_0 as it overturns the general belief in case it was true

$\Rrightarrow$ *Assume H_0 to be correct*, & try to find *evidence against H_0* from the data

Proof by Contradication of H_0!

The Null And The Alternative Hypotheses III

- *Difference between estimation (Ch.7) & hypothesis testing (Ch.8)*:
  - **Recall**: *In estimation* (Ch.7), we aim at *choosing the best value/values* in the whole parameter space Θ to serve as a guess of the true but unknown value θ_0
  - There are 2 meaningful & contrasting collections of possible values of θ in Θ set as H_0 & H_1 in the first place & our goal *in a hypothesis test* (Ch.8) is to choose 1 over the other, *i.e.*, *which collection, Θ_0 or Θ_1, contains θ_0?*
- 2 kinds of hypotheses according to the nature of Θ_i, $i = 0, 1$: If a hypothesis Θ_i consists of only 1 value in Θ, it is a *simple hypothesis* (*e.g.*, $\theta_0 = \theta_0$); otherwise it is a *composite hypothesis*

Hypothesis Test

Definition

A *hypothesis testing procedure* or *hypothesis test* is a rule that specifies:

1. For which sample values H_0 is rejected, &, in turn, H_1 is accepted to be true
2. For which sample values the decision is made not to reject H_0, &, in turn, accept H_0 as true

The set of the observations/data/sample values for which H_0 will rejected is called the *critical region (CR)*

- The CR is also called the *rejection region*. The complement of the CR is called the *acceptance region*
- Decision/conclusion of a test: *Reject or accept H_0* depending on the *level of significance (α) of the test*

Significance Levels I

No one knows the true value θ_0!

- *2 possible kinds of errors*:

$$\begin{cases} \underline{\textit{type I error}} & \Leftrightarrow & \text{erroneous rejection of a true } H_0 \\ \underline{\textit{type II error}} & \Leftrightarrow & \text{erroneous retention of a false } H_0 \end{cases}$$

Decision \ Truth	H_0 is correct	H_1 is correct
Reject H_0	Type I error	✔
Accept H_0	✔	Type II error

Significance Levels II

Definition

Significance level/level of significance $0 < \alpha < 1$ of a hypothesis test is a pre–specified small prob taken into account when the CR of a test is derived s.t. when the decision of a test is to reject H_0, the prob that H_0 is actually correct would never be $> \alpha$

- Such a test is called the *α level test*
- $P(\text{Commit a type I error}) = P(\text{Reject } H_0 | H_0 \text{ is true}) = \alpha$
- Rejecting a true H_0/committing a type I error is *more consequential/disastrous* than failing to reject a false H_0/committing a type II error
- **In practice**, the magnitude of α is controlled to be as small as possible, & is chosen to be as small as .05 or .01, *before analyzing the data*

Critical Region

Obtain The Critical Region Via ML Estimator of θ

The critical region (CR) is derived s.t. when the decision of a test concerning the value of a parameter θ is to reject H_0, the prob that H_0 is actually correct would never be $> \alpha$. A common approach to find CR relies on the sampling distribution of the ML estimator of θ

- **Recall**: The ML estimator of θ, $\hat{\theta} = d(\mathbf{X})$, yields an *ML estimate $\hat{\theta}_0$ based on the sample values $x_1, \ldots, x_n$*, & usually $\hat{\theta}_0$ is a good guess/proxy of the true value θ_0
- Under H_0 (*i.e.,* when H_0 is assumed to be true), *compare $\hat{\theta}_0$ with values in Θ_0* to see if

$\hat{\theta}_0$ is "far away" from Θ_0 $\Leftrightarrow$ Contradiction of assumption of H_0

2-Sided Test For a Normal Mean (Known σ^2) I

Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$ with known variance σ^2. Consider a 2-sided test concerning the unknown mean, which has

$$H_0 : \mu = \mu_0 \qquad \& \qquad H_1 : \mu \neq \mu_0$$

where $-\infty < \mu_0 < \infty$ is some pre-specified constant

1. $\overline{X}$ is the ML estimator of μ (Example 2 (Ch.7))
2. Sampling distribution of $\overline{X}$ under H_0: $\overline{X} \sim N\left(\mu_0, \dfrac{\sigma^2}{n}\right)$ (Equation (2) in Ch.6) as μ has a value μ_0 under the assumption that H_0 is true
3. Intuitively, the *critical region (CR)* should be of the form

$$\{X_1, \ldots, X_n : |\overline{X} - \mu_0| > c\}$$

for some suitably chosen value $c > 0$, *i.e.,* when $\overline{X}$ has the aforementioned sampling distribution, any observed value $\overline{x}$ should not be too far away from its mean μ_0

2-Sided Test For a Normal Mean (Known σ^2) II

4. To determine the value c, consider

$$P(\text{Commit a type I error}) = P(|\overline{X} - \mu_0| > c|H_0) = \alpha$$

where α is the pre-determined *level of significance*. Notice that the latter prob is equal to

$$P\left(|Z| > \frac{c}{\sigma/\sqrt{n}}\right) = 2 \times P\left(Z > \frac{c\sqrt{n}}{\sigma}\right)$$

as $\dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$ under H_0. Hence,

$$P\left(Z > \frac{c\sqrt{n}}{\sigma}\right) = \frac{\alpha}{2} \quad \Rrightarrow \quad \frac{c\sqrt{n}}{\sigma} = z_{\alpha/2} \quad \Rrightarrow \quad c = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

2-Sided Test For a Normal Mean (Known σ^2) III

CR of The 2-Sided Test For a Normal Mean (Known Variance)

The α level test which has $H_0 : \mu = \mu_0$ & $H_1 : \mu \neq \mu_0$, based on a sample from $N(\mu, \sigma^2)$ with known σ, has a *critical region*:

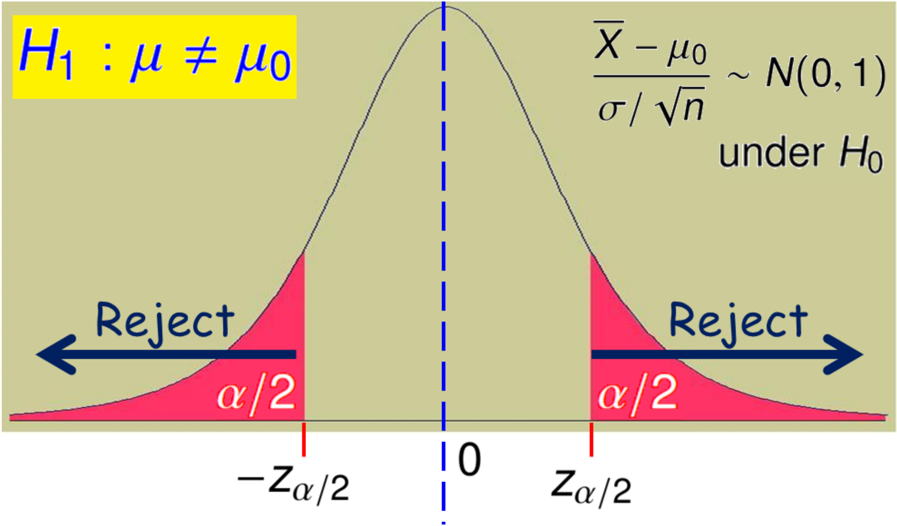
$$\left\{ \frac{|\overline{X} - \mu_0|}{\sigma / \sqrt{n}} > z_{\alpha/2} \right\} = \left\{ \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} < -z_{\alpha/2} \text{ or } \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} > z_{\alpha/2} \right\}$$

Definition

The r.v./statistic $\frac{|\overline{X} - \mu_0|}{\sigma / \sqrt{n}}$ in the above CR is called the *test statistic (TS)*

- Any TS should have a sampling distribution which does not depend on any unknown parameters
- **In practice**, one can compute an *observed value of the TS* based on the sample values $x_1, \ldots, x_n$ & see if such a value v falls in the CR

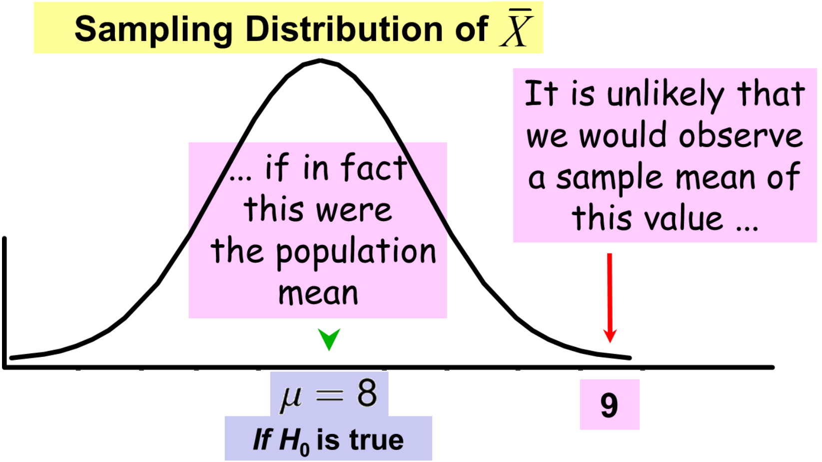
2-Sided Test For a Normal Mean (Known σ^2) IV



Example 2: 2-Sided Test For a Normal Mean (Known σ^2) I

1. A random sample of size 10 from a $N(\mu, 1)$ population with sample mean $\overline{x} = 9$ is given
2. To test $H_0 : \mu = 8$ VS $H_1 : \mu \neq 8$ at $\alpha = 5\%$ significance level
   - obtain the *CR* as
   $$\left\{ \frac{|\overline{X} - 8|}{1/\sqrt{10}} > 1.96 \right\}$$
   where TS $= \sqrt{10}|\overline{X} - 8|$
   - as $\overline{x} = 9$, compute the *observed value of TS*: $v = \sqrt{10}|9 - 8| = 3.16$
   $\Rightarrow$ v falls in the CR as $v > 1.96$
3. We *reject H_0 at 5% significance level* & have sufficient evidence to conclude that the true value of μ is different from 8
4. **Note** : In rejecting H_0 at 5% level, we should keep in mind that the prob of having mistakenly rejected a true H_0 (*i.e.,* the *prob of committing a type I error*) is equal to $\alpha = 5\%$

Example 2: 2-Sided Test For a Normal Mean (Known σ^2) II



... Hence, we reject $H_0 : \mu = 8$

Analog Between c.i. & Hypothesis Tests

- **Recall** (Equation (1), Ch.7): For the same set–up, a $100(1-\alpha)\%$ c.i. for μ is given by
$$\left(\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$
- Rearrange the CR of a 2–sided α level test as
$$\left\{\overline{X} < \mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \text{ or } \overline{X} > \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\}$$
which corresponds to an acceptance region
$$\left\{\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu_0 \leq \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\}$$
- Substituting $\overline{x}$ into the acceptance region yields the *same endpoints* as those of a $100(1-\alpha)\%$ c.i. for μ (*i.e.,* the 1st equation at this page)

Using a c.i. For Hypothesis Testing

When μ_0 is not inside a $100(1-\alpha)\%$ c.i. for μ, we reject $H_0 : \mu = \mu_0$ in a 2–sided α level test

The p-Value I

- Every test is based on a pre–specified significance level α, but no guidance is there about how to choose it
- Our test result depends on α: *One may cheat* by always choosing a larger α in order to reject H_0
- When we reject H_0

How strong is the evidence against H_0?

Definition

Based on any sample, the *p-value* is defined to be the smallest significance level at which H_0 would be rejected

- The prob under H_0 of *a result as or more extreme than that actually observed*
- *The smaller the p-value, the stronger the evidence against H_0*

The p-Value II

- Reporting the p-value in a test is more *flexible*: Any decision/conclusion is made by the user who makes his/her own choice of α

Computing The p-Value

The p-value is computed to be the prob

$$\begin{cases} P(TS \geq v), & \text{when } H_a : \theta \neq \theta_0 \text{ (i.e., for a 2-sided test)} \\ P(TS \geq v), & \text{when } H_a : \theta > \theta_0 \\ P(TS \leq v), & \text{when } H_a : \theta < \theta_0 \end{cases}$$

where v is the observed value of the TS based on the sample

- For Example 2, with $v = 3.16$, the p-value is given by

$$P(\sqrt{10}|\bar{X} - 8| \geq 3.16) = 2P(Z \geq 3.16) = 2(1 - .9992) = .0016$$

As long as $\alpha > .0016$, we would reject H_0. Indeed, a p-value of such a small magnitude provides very strong evidence against H_0

1-Sided Tests For a Normal Mean (Known σ^2) I

Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$ with known variance σ^2. Consider a *1–sided test* concerning the unknown mean, which has

$$H_0 : \mu = \mu_0 \text{ (or } \mu \leq \mu_0) \qquad \& \qquad H_1 : \mu > \mu_0$$

where $-\infty < \mu_0 < \infty$ is some pre-specified constant

1. $\overline{X}$ remains to be the ML estimator of μ as in the 2–sided test
2. Same sampling distribution of $\overline{X}$ under H_0: $\overline{X} \sim N\left(\mu_0, \dfrac{\sigma^2}{n}\right)$
3. Intuitively, the *critical region (CR)* should be of the form

$$\{X_1, \ldots, X_n : \overline{X} - \mu_0 > c\}$$

   for some suitably chosen value $c > 0$, *i.e.,* when $\overline{X}$ has the aforementioned sampling distribution, any observed value $\overline{x}$ should not be much greater than its mean μ_0

1-Sided Tests For a Normal Mean (Known σ^2) II

4. To determine the value c, consider

$$P(\text{Commit a type I error}) = P(\overline{X} - \mu_0 > c | H_0) = \alpha$$

where α is the pre-determined *level of significance*. Notice that the latter prob is equal to

$$P\left(\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{c}{\sigma/\sqrt{n}} \middle| H_0\right) = P\left(Z > \frac{c}{\sigma/\sqrt{n}}\right)$$

where $\text{TS} = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$ is the *test statistic (TS)*. Hence,

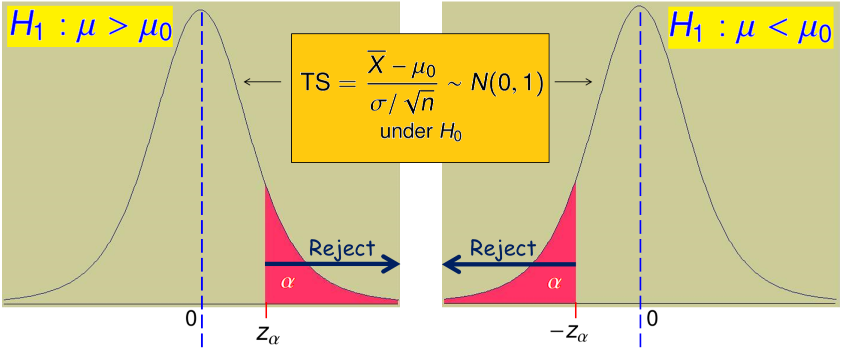
$$P\left(Z > \frac{c\sqrt{n}}{\sigma}\right) = \alpha \quad \Rrightarrow \quad \frac{c\sqrt{n}}{\sigma} = z_\alpha \quad \Rrightarrow \quad c = z_\alpha \frac{\sigma}{\sqrt{n}}$$

1-Sided Tests For a Normal Mean (Known σ^2) III

CR's of 1–Sided Tests For a Normal Mean (Known Variance)

The α level test which has $H_0 : \mu = \mu_0$ & $H_1 : \mu > \mu_0$ (*resp.*, $H_1 : \mu < \mu_0$), based on a sample from $N(\mu, \sigma^2)$ with known σ, has a *critical region*:

$$\left\{ \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} > z_\alpha \right\} \qquad \left(\textit{resp.}, \left\{ \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} < -z_\alpha \right\} \right)$$



Example 2: 1–Sided Test For a Normal Mean (Known σ^2)

1. A random sample of size 10 from a $N(\mu, 1)$ population with sample mean $\overline{x} = 9$ is given
2. To test $H_0 : \mu = 8$ VS $H_1 : \mu > 8$ at $\alpha = 5\%$ significance level
   - obtain the CR as
   $$\left\{ \frac{\overline{X} - 8}{1/\sqrt{10}} > 1.645 \right\}$$
   where TS $= \sqrt{10}(\overline{X} - 8)$
   - as $\overline{x} = 9$, compute the *observed value of TS*: $v = \sqrt{10}(9 - 8) = 3.16$
   $\Rrightarrow$ v falls in the CR as $v > 1.645$
3. We *reject H_0 at 5% significance level* & have sufficient evidence to conclude that the true value of μ is greater than 8
4. One can compute the *p-value* of this test as $P(Z \geq 3.16) = .0008$. As long as $\alpha > .0008$, we would reject H_0. This provides very strong evidence against H_0

2-Sided Test For a Normal Mean (Unknown σ^2) I

Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$ with both mean & variance unknown. Consider a *2-sided test*, which has

$$H_0 : \mu = \mu_0 \qquad \& \qquad H_1 : \mu \neq \mu_0$$

where $-\infty < \mu_0 < \infty$ is some pre-specified constant

1. $\overline{X}$ is the ML estimator of μ (`Example 3 (Ch.7)`)
2. Sampling distribution of $\overline{X}$ under H_0: $T = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$ (Equation (3), `Ch.6`) as μ has a value μ_0 under the assumption that H_0 is true
3. Assume that the *critical region (CR)* is of the form

$$\{X_1, \ldots, X_n : |T| > c\}$$

for some suitably chosen value $c > 0$

2-Sided Test For a Normal Mean (Unknown σ^2) II

4. To determine the value c, consider

$$P(\text{Commit a type I error}) = P\left(|T| > c|\, H_0\right) = \alpha$$

where α is the pre-determined *level of significance*. Notice that the latter prob is equal to

$$P\left(|t_{n-1}| > c\right) = 2 \times P\left(t_{n-1} > c\right)$$

Hence,

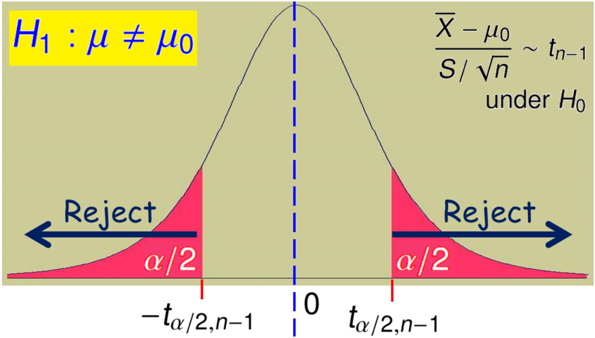
$$P\left(t_{n-1} > c\right) = \frac{\alpha}{2} \quad \Rrightarrow \quad c = t_{\alpha/2, n-1}$$

2-Sided Test For a Normal Mean (Unknown σ^2) III

CR of The 2-Sided Test For a Normal Mean (Unknown Variance)

The α level test which has $H_0 : \mu = \mu_0$ & $H_1 : \mu \neq \mu_0$, based on a sample from $N(\mu, \sigma^2)$ with unknown σ, has a *critical region*:

$$\left\{ TS = \frac{|\overline{X} - \mu_0|}{S/\sqrt{n}} > t_{\alpha/2, n-1} \right\}$$



Example 3: 2-Sided Test For a Normal Mean (Unknown σ^2) I

On a popular self-image test that results in normally distributed scores, the mean score for public-assistance recipients is expected to be 65. A random sample of 28 public-assistance recipients in a region is given the test. They achieve a mean score of 62.1, & their scores have a standard deviation of 5.83. Do the public-assistance recipients in this region test differently, on average, than what is expected, at the 2% level of significance?

Solution: Let μ be the mean self-image test score of all the public-assistance recipients in this region

1. Test $H_0 : \mu = 65$ VS $H_1 : \mu \neq 65$
2. Assume that the self-image test scores of all the public-assistance recipients in this region are normally distributed with unknown variance

Example 3: 2-Sided Test For a Normal Mean (Unknown σ^2) II

3. From a sample of size 28, $\overline{x} = 62.1$ & $s = 5.83$
4. The observed value of the TS is

$$v = \frac{|62.1 - 65|}{5.83/\sqrt{28}} = 2.632 > 2.473 = t_{.01,27}$$

5. Reject H_0 at 2% level of significance & have sufficient evidence to conclude that the public-assistance recipients in this region test differently, on average, than what is expected
6. One can compute the *p-value* of this test as

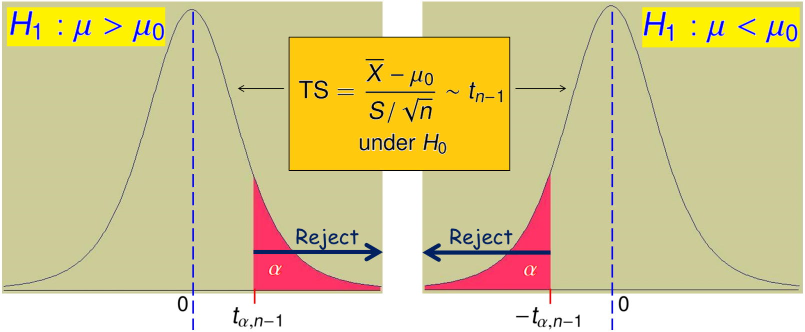
$$P(\text{TS} \geq 2.632) = 2P(t_{n-1} \geq 2.632) < 2P(t_{n-1} \geq 2.473) = .02$$

1-Sided Tests For a Normal Mean (Unknown σ^2)

CR's of 1-Sided Tests For a Normal Mean (Unknown Variance)

The α level test which has $H_0 : \mu = \mu_0$ & $H_1 : \mu > \mu_0$ (*resp.*, $H_1 : \mu < \mu_0$), based on a sample from $N(\mu, \sigma^2)$ with unknown σ, has a *critical region*:

$$\left\{ \frac{\overline{X} - \mu_0}{S/\sqrt{n}} > t_{\alpha,n-1} \right\} \qquad \left(resp., \left\{ \frac{\overline{X} - \mu_0}{S/\sqrt{n}} < -t_{\alpha,n-1} \right\} \right)$$



Example 4: 1-Sided Test For a Normal Mean (Unknown σ^2) I

A company requires that the mean down time of the computer server does not exceed 200 minutes. A sample of 10 most recent breakdown times of the server is as follows

$$210, \ 195, \ 197.5, \ 193, \ 205.2, \ 202, \ 186, \ 195.5, \ 202.5, \ 208.3$$

Would you conclude that the server is well-maintained at 10% level?

Solution: Let μ be the mean down time of the computer server

1. Test $H_0 : \mu = 200$ VS $H_1 : \mu < 200$
2. Assume that the down times of the computer server are normally distributed with unknown variance

Example 4: 1-Sided Test For a Normal Mean (Unknown σ^2) II

3. From the sample of size 10 from this normal population,
$\bar{x} = \frac{1}{10}(210 + 195 + \cdots + 208.3) = 199.5$ &
$s^2 = \frac{1}{10-1}(210^2 + 195^2 + \cdots + 208.3^2 - 10\bar{x}^2) = 55.576$
4. Observed value of the TS is

$$v = \frac{199.5 - 200}{\sqrt{55.576/10}} = -1.179 \not< -1.383 = -t_{.1,9}$$

5. Do not reject H_0 at 10% level of significance $\Rightarrow$ Insufficient evidence to conclude that the computer server is well-maintained
6. One can compute the *p-value* of this test as

$$P(t_{n-1} \leq -1.179) > P(t_{n-1} \leq -1.383) = .10$$

Test For Equality of 2 Normal Means (Known Variances) I

Let $X_1, \ldots, X_n$ be a random sample of size n from $N(\mu_1, \sigma_1^2)$ & let $Y_1, \ldots, Y_m$ be a random sample of size m from $N(\mu_2, \sigma_2^2)$ with *both variances known*, & suppose that the 2 samples are indept of each other. Consider a *2-sided test*, which has

$$H_0 : \mu_1 = \mu_2 \qquad \& \qquad H_1 : \mu_1 \neq \mu_2$$

or

$$H_0 : \mu_1 - \mu_2 = 0 \qquad \& \qquad H_1 : \mu_1 - \mu_2 \neq 0$$

1. $\overline{X} - \overline{Y}$ is the ML estimator of $\mu_1 - \mu_2$
2. Sampling distribution of $\overline{X} - \overline{Y}$ under H_0: $\overline{X} - \overline{Y} \sim N\left(0, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right)$
3. Intuitively, the *critical region (CR)* should be of the form

$$\{X_1, \ldots, X_n : |\overline{X} - \overline{Y}| > c\}$$

for some suitably chosen value $c > 0$

Test For Equality of 2 Normal Means (Known Variances) II

4. To determine the value c, consider

$$P(\text{Commit a type I error}) = P(|\overline{X} - \overline{Y}| > c|H_0) = \alpha$$

where α is the pre-determined *level of significance*. Notice that the latter prob is equal to

$$P\left(|Z| > k\right) = 2 \times P\left(Z > k\right)$$

as $\dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}} \sim N(0, 1)$ under H_0, where $k = \dfrac{c}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}}$. Hence,

$$P\left(Z > k\right) = \frac{\alpha}{2} \quad \Rightarrow \quad k = z_{\alpha/2} \quad \Rightarrow \quad c = z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}$$

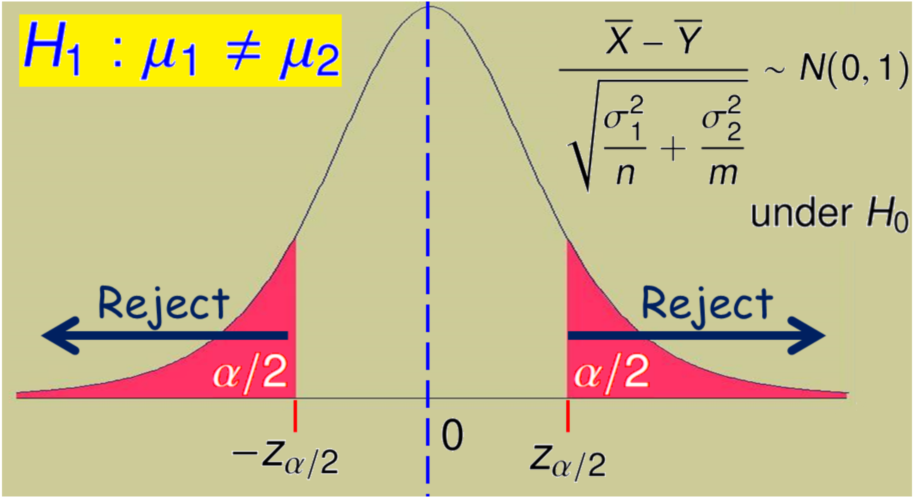
Test For Equality of 2 Normal Means (Known Variances) III

2-Sided Test For Equality of 2 Normal Means (Known Variances)

Consider the α level test which has $H_0 : \mu_1 = \mu_2$ & $H_1 : \mu_1 \neq \mu_2$ based on a sample of size n from $N(\mu_1, \sigma_1^2)$ & a sample of size m from $N(\mu_2, \sigma_2^2)$, where the samples are indept & both population variances are known

1. The *test statistic* is defined by $\text{TS} = \dfrac{|\overline{X} - \overline{Y}|}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}}$
2. The *critical region* is given by
$$\{\text{TS} > z_{\alpha/2}\}$$
3. The *p-value* is computed by
$$2P(Z \geq v)$$
when v is the observed value of the TS

Test For Equality of 2 Normal Means (Known Variances) IV



Example 5: Test For Equality of 2 Normal Means (Known Variances)

Refer to Example 7 (Ch.7), test if the average CAPs of students with religion & students without any religion are the same at the 1% level

Solution: Given $\sigma_1 = \sigma_2 = .5$

1. Test $H_0 : \mu_1 = \mu_2$ VS $H_1 : \mu_1 \neq \mu_2$
2. The 2 samples yield $n = m = 40$, & $\bar{x} - \bar{y} = -.18$
3. Observed value of the TS is $v = \dfrac{|-.18|}{\sqrt{\frac{.5^2}{40} + \frac{.5^2}{40}}} = 1.61 \not> 2.58 = z_{.005}$
4. Do not reject H_0 at 1% level $\Rightarrow$ Insufficient evidence to conclude that there is any difference in average CAPs of the 2 groups of students
5. p-value $= 2P(Z \geq 1.61) = 2(1 - .9463) = .1064$
6. Alternatively, based on the 99% c.i. for $\mu_1 - \mu_2$ (Example 7 (Ch.7)), 0 is inside $(-.11, .47)$ $\Rightarrow \mu_1 - \mu_2 = 0 \Leftrightarrow$ Do not reject H_0

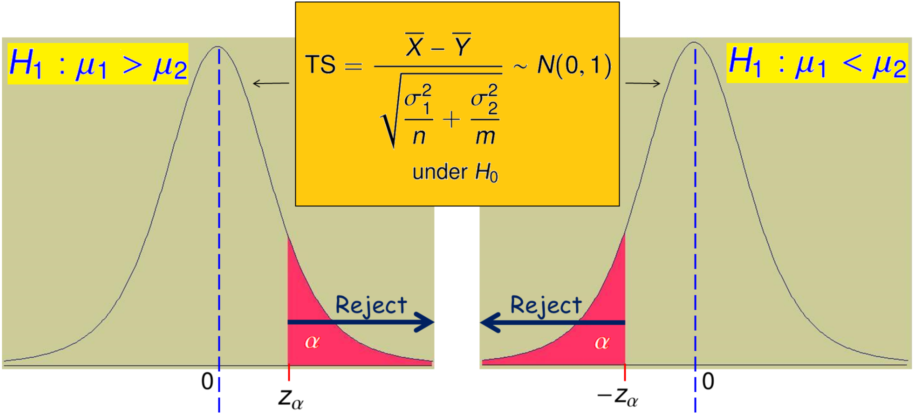
1-Sided Tests Comparing 2 Normal Means (Known Variances) I

1-Sided Tests For Comparison Between 2 Normal Means (Known Variances)

Consider the α level test which has $H_0 : \mu_1 = \mu_2$ & $H_1 : \mu_1 > \mu_2$ (*resp.*, $H_1 : \mu_1 < \mu_2$) based on a sample of size n from $N(\mu_1, \sigma_1^2)$ & a sample of size m from $N(\mu_2, \sigma_2^2)$, where the samples are indept & both population variances are known

1. The *test statistic* is defined by $\text{TS} = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}}$
2. The *critical region* is given by
$$\{\text{TS} > z_\alpha\} \qquad (\textit{resp.}, \ \{\text{TS} < -z_\alpha\})$$
3. The *p-value* is computed by $P(Z \geq v)$ (*resp.*, $P(Z \leq v)$) when v is the observed value of the TS

1-Sided Tests Comparing 2 Normal Means (Known Variances) II



Example 6: *1-Sided Test Comparing 2 Normal Means (Known Variances)*

Repeat the previous example by testing against $H_1 : \mu_1 < \mu_2$ at the 5% level

Solution:

1. Test $H_0 : \mu_1 = \mu_2$ VS $H_1 : \mu_1 < \mu_2$
2. Observed value of the TS is

$$v = \frac{-.18}{\sqrt{\frac{.5^2}{40} + \frac{.5^2}{40}}} = -1.61 \not< -1.645 = -z_{.05}$$

3. Do not reject H_0 at 5% level $\Rrightarrow$ Insufficient evidence to conclude that students with religion have a lower average CAP than students without any religion
4. p-value $= P(Z \leq -1.61) = 1 - .9463 = .0537$

Test For Equality of 2 Normal Means (Unknown But Equal Variances) I

Let $X_1, \ldots, X_n$ be a random sample of size n from $N(\mu_1, \sigma_1^2)$ & let $Y_1, \ldots, Y_m$ be a random sample of size m from $N(\mu_2, \sigma_2^2)$ with *both variances unknown but equal*, & suppose that the 2 samples are indept of each other. Consider a *2-sided test*, which has

$$H_0 : \mu_1 = \mu_2 \qquad \& \qquad H_1 : \mu_1 \neq \mu_2$$

1. $\overline{X} - \overline{Y}$ is the ML estimator of $\mu_1 - \mu_2$
2. Under H_0, sampling distribution of $T = \dfrac{\overline{X} - \overline{Y}}{S_p\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}} \sim t_{n+m-2}$, where

   S_p is defined at Page 36 (Ch.7)
3. Assume that the *critical region (CR)* is of the form

   $$\{X_1, \ldots, X_n : |T| > c\}$$

   for some suitably chosen value $c > 0$

Test For Equality of 2 Normal Means (Unknown But Equal Variances) II

4. To determine the value c, consider

$$P(\text{Commit a type I error}) = P(|T| > c|H_0) = \alpha$$

where α is the pre-determined *level of significance*. Notice that the latter prob is equal to

$$P\left(|t_{n+m-2}| > c\right) = 2 \times P\left(t_{n+m-2} > c\right)$$

Hence,

$$P\left(t_{n+m-2} > c\right) = \frac{\alpha}{2} \quad \Rrightarrow \quad c = t_{\alpha/2, n+m-2}$$

Test For Equality of 2 Normal Means (Unknown But Equal Variances) III

2-Sided Test For Equality of 2 Normal Means (Unknown But Equal Variances)

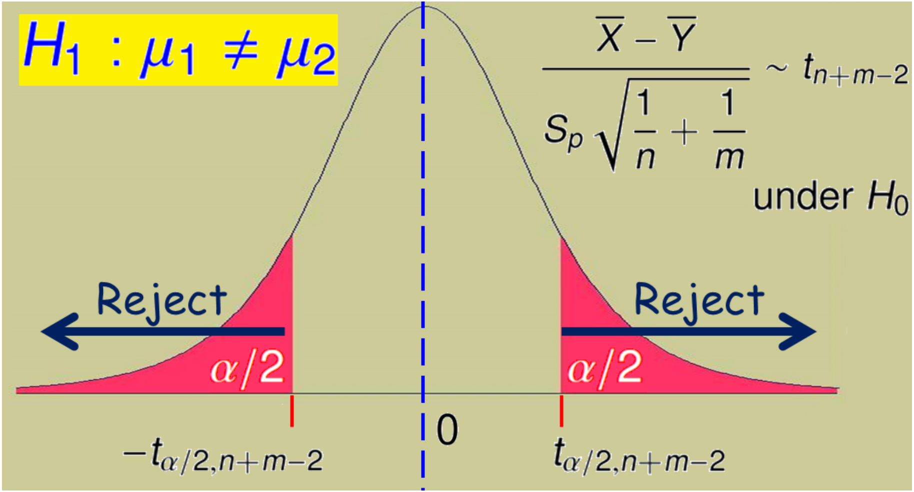
Consider the α level test which has $H_0 : \mu_1 = \mu_2$ & $H_1 : \mu_1 \neq \mu_2$ based on a sample of size n from $N(\mu_1, \sigma_1^2)$ & a sample of size m from $N(\mu_2, \sigma_2^2)$, where the samples are indept & both population variances are unknown but equal

1. The *test statistic* is defined by $\text{TS} = \dfrac{|\overline{X} - \overline{Y}|}{S_p\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}}$
2. The *critical region* is given by

$$\{\text{TS} > t_{\alpha/2, n+m-2}\}$$

3. The *p-value* is computed by $2P(t_{n+m-2} \geq v)$ when v is the observed value of the TS

Test For Equality of 2 Normal Means (Unknown But Equal Variances) IV



Example 7: Test For Equality of 2 Normal Means (Unknown But Equal Variances) I

Let us look at a packaging process for ground beef over a series of days. An interesting question is whether the true mean amount delivered by this process changes from day to day. Ten packages are accurately weighed (in grams) & randomly selected over the course of the day on two consecutive days. For the 1st day, the mean weight is 1391.50 & the sample variance of the weights is 28.409. For the 2nd day, the mean weight is 1398.65 & the sample variance of the weights is 51.696.

Solution: It is reasonable to assume that the weights of packages processed on each of the days are indept & normally distributed, & also there is a constant variance for all the weights of packages processed over the course of 1 day from day to day

1. Test $H_0 : \mu_1 = \mu_2$ VS $H_1 : \mu_1 \neq \mu_2$, where μ_1 & μ_2 are the mean weights of packages processed on the 2 consecutive days

Example 7: Test For Equality of 2 Normal Means (Unknown But Equal Variances) II

2. The 2 samples give $n = m = 10$, $\bar{x} - \bar{y} = 1391.50 - 1398.65 = -7.15$, $s_1^2 = 28.409$, & $s_2^2 = 51.696$, which yield

$$s_p^2 = \frac{(n-1)s_1^2 + (m-1)s_2^2}{n+m-2} = \frac{9(28.409) + 9(51.696)}{18} = 6.33^2$$

3. Observed value of the TS is

$$v = \frac{|-7.15|}{6.33\sqrt{\frac{1}{10} + \frac{1}{10}}} = 2.526 > 2.101 = t_{18,.025}$$

4. Reject H_0 at 5% level $\Rrightarrow$ Sufficient evidence to conclude that the true mean amounts delivered by the process on different days vary
5. p-value $= 2P(t_{18} \geq 2.526)$

1-Sided Tests Comparing 2 Normal Means (Unknown But Equal Variances) I

1-Sided Tests For Comparison Between 2 Normal Means (Unknown But Equal Variances)

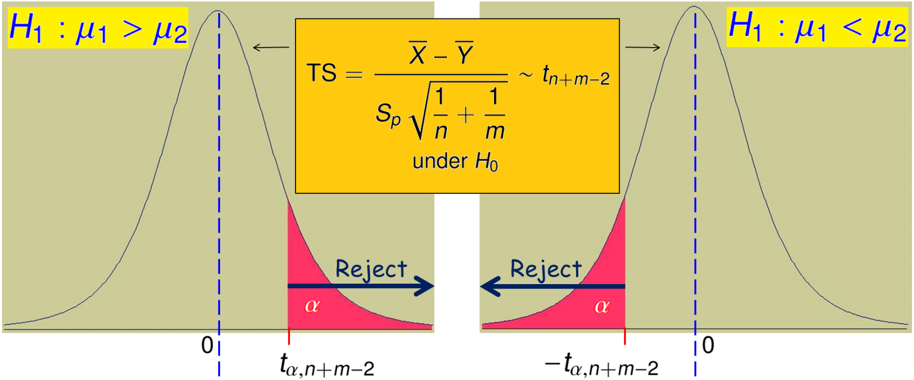
Consider the α level test which has $H_0 : \mu_1 = \mu_2$ & $H_1 : \mu_1 > \mu_2$ (*resp.*, $H_1 : \mu_1 < \mu_2$) based on a sample of size n from $N(\mu_1, \sigma_1^2)$ & a sample of size m from $N(\mu_2, \sigma_2^2)$, where the samples are indept & both population variances are unknown but equal

1. The *test statistic* is given by $\text{TS} = \dfrac{\overline{X} - \overline{Y}}{S_p \sqrt{\dfrac{1}{n} + \dfrac{1}{m}}}$
2. The *critical region* is given by

$$\{\text{TS} > t_{\alpha,n+m-2}\} \qquad (\textit{resp.},\ \{\text{TS} < -t_{\alpha,n+m-2}\})$$

3. The *p-value* is computed by $P(t_{n+m-2} \geq v)$ (*resp.*, $P(t_{n+m-2} \leq v)$) when v is the observed value of the TS

1-Sided Tests Comparing 2 Normal Means (Unknown But Equal Variances) II



$$\text{TS} = \frac{\overline{X} - \overline{Y}}{S_p\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}} \sim t_{n+m-2}$$

under H_0

Large-Sample Test For Equality of 2 Non-Normal Means (Unknown Variances) I

Let $X_1, \ldots, X_n$ be a random sample of size n from any non-normal population with mean μ_1 & variance σ_1^2 & let $Y_1, \ldots, Y_m$ be a random sample of size m from another non-normal population with mean μ_2 & variance σ_2^2, where *both variances are unknown*, & the 2 samples are indept of each other. Consider a *2-sided test*, which has

$$H_0 : \mu_1 = \mu_2 \qquad \& \qquad H_1 : \mu_1 \neq \mu_2$$

1. Sampling distribution of the ML estimator of $\mu_1 - \mu_2$, $\overline{X} - \overline{Y}$, under H_0:
   1. $E(\overline{X} - \overline{Y}) = 0$ & $\text{Var}(\overline{X} - \overline{Y}) = \dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}$
   2. $\overline{X} - \overline{Y} \dot{\sim} N\left(0, \dfrac{S_1^2}{n} + \dfrac{S_2^2}{m}\right) \Leftrightarrow \dfrac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}} \dot{\sim} N(0, 1)$ *when n, m are large*
2. The subsequent derivation follows as that at Page 36 of a test for equality of means from 2 normal populations with known variance

Large-Sample Test For Equality of 2 Non-Normal Means (Unknown Variances) II

Large-Sample 2-Sided Test For Equality of 2 Non-Normal Means (Unknown Variances)

Consider the α level test which has $H_0 : \mu_1 = \mu_2$ & $H_1 : \mu_1 \neq \mu_2$ based on a sample of size n from a non-normal population with mean μ_1 & variance σ_1^2 & a sample of size m from another non-normal population with mean μ_2 & variance σ_2^2, where the 2 samples are indept, both population variances are unknown, & n, m are large

1. The *test statistic* is defined by $\text{TS} = \dfrac{|\overline{X} - \overline{Y}|}{\sqrt{\dfrac{S_1^2}{n} + \dfrac{S_2^2}{m}}}$
2. The *critical region* is given by
$$\{\text{TS} > z_{\alpha/2}\}$$
3. The *p-value* is computed by $2P(Z \geq v)$ when v is the observed value of the TS

Large-Sample 1-Sided Tests Comparing 2 Non-Normal Means (Unknown Variances)

Large-Sample 1-Sided Tests For Comparison Between 2 Non-Normal Means (Unknown Variances)

Consider the α level test which has $H_0 : \mu_1 = \mu_2$ & $H_1 : \mu_1 > \mu_2$ (*resp.*, $H_1 : \mu_1 < \mu_2$) based on a sample of size n from a non-normal population with mean μ_1 & variance σ_1^2 & a sample of size m from another non-normal population with mean μ_2 & variance σ_2^2, where the 2 samples are indept, both population variances are unknown, & n, m are large

1. The *test statistic* is defined by $\text{TS} = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{S_1^2}{n} + \dfrac{S_2^2}{m}}}$
2. The *critical region* is given by

$$\{\text{TS} > z_\alpha\} \qquad (\textit{resp.}, \ \{\text{TS} < -z_\alpha\})$$

3. The *p-value* is computed by $P(Z \geq v)$ (*resp.*, $P(Z \leq v)$) when v is the observed value of the TS

The Paired t-Test I

So far in comparing difference in means of 2 populations, we always assume that we have indept samples from the 2 populations. On many occasions, however, the *2 samples are not indept* because they may involve the same experimental unit, one taken in scenario 1 & the other in scenario 2, *e.g.,*

- Salt-free diets are often prescribed to people with high blood pressure. People's diastolic blood pressures are measured before & after consuming a salt-free diet for 2 weeks. Study *change* in blood pressure after consumption of a salt-free diet
- An SoC student may wonder if reading ST2334 helps in studying computer science. Study *change* in CAPs of students prior to & after reading ST2334

Our interest is: **Is there any change?**

The Paired t-Test II

Data/Paired Sample

1. n pairs of observations (X_i, Y_i), $i = 1, \ldots, n$
2. X_i is the measurement from the ith experimental unit under scenario 1
3. Y_i is the measurement from the ith experimental unit under scenario 2
4. $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ forms a *paired sample of size n*

- $W_i = X_i - Y_i$ defines the *difference between measurements under scenario 1 & under scenario 2*, or the change in the measurement under different scenarios

Assumptions

All the differences W_i's are assumed to be a random sample of n from a normal population, $N(\mu_w, \sigma_w^2)$, with both mean & variance unknown

The Paired t-Test III

- By viewing $W_1, \dots, W_n$ as 1 random sample of size n from a normal population with unknown variance, tests concerning the mean μ_w is *equivalent* to "Tests of a Normal Mean (Unknown σ^2)" discussed at Pages 28-33

The Paired t-Test

The α level paired t-test which has $H_0 : \mu_w = 0$ & $H_1 : \mu_w \neq 0$, based on a paired sample of size n with differences i.i.d. from $N(\mu_w, \sigma_w^2)$ with unknown σ_w, has

1. *test statistic* defined by $\text{TS} = \dfrac{|\overline{W}|}{S_w/\sqrt{n}}$, where $\overline{W}$ & S_w are the sample mean & the sample sd of the differences
2. *critical region* given by
$$\{\text{TS} > t_{\alpha/2, n-1}\}$$
3. *p-value* computed by $2P(t_{n-1} \geq v)$ when v is the observed value of the TS

1-Sided Paired t-Tests

The α level test which has $H_0 : \mu_w = 0$ & $H_1 : \mu_w > 0$ (*resp.*, $H_1 : \mu_w < 0$), based on a paired sample of size n with differences i.i.d. from $N(\mu_w, \sigma_w^2)$ with unknown σ_w, has

1. *test statistic* defined by
$$\text{TS} = \frac{\overline{W}}{S_w/\sqrt{n}}$$
where $\overline{W}$ & S_w are the sample mean & the sample sd of the differences
2. *critical region* given by
$$\{\text{TS} > t_{\alpha,n-1}\} \qquad (\textit{resp.}, \{\text{TS} < -t_{\alpha,n-1}\})$$
3. *p-value* computed by
$$P(t_{n-1} \geq v) \qquad (\textit{resp.}, P(t_{n-1} \leq v))$$
when v is the observed value of the TS

Example 8: The Paired t-Test I

Ten people recently diagnosed with diabetes were tested to determine whether an educational program were effective in increasing their knowlegde of diabetes. They were given a test, before & after the educational program, concerning self-care aspects of diabetes. The scores on the test were as follows:

Patient	1	2	3	4	5	6	7	8	9	10
Before	75	62	67	70	55	59	60	64	72	59
After	77	65	68	72	62	61	60	67	75	68

Solution: The 2 samples are clearly not indept $\Rrightarrow$ Apply the paried t-test by imposing an *assumption* that the change in the scores after the educational program for every person diagnosed with diabetes is normally distributed with both mean μ_w & variance σ_w^2 unknown

Example 8: The Paired t-Test II

1. Test $H_0 : \mu_w = 0$ VS $H_1 : \mu_w > 0$ (a positive mean change: increase in knowledge of diabetes)
2. Compute change in scores for all 10 patients:

Patient	1	2	3	4	5	6	7	8	9	10
Change	2	3	1	2	7	2	0	3	3	9

$\Rrightarrow$ In this sample of size 10, $\overline{w} = 3.2$ & $s_w = 2.741$

3. Observed value of the TS is $v = \dfrac{3.2}{2.741/\sqrt{10}} = 3.69 > 2.821 = t_{.01,9}$
4. Reject H_0 at 1% level $\Rrightarrow$ Sufficient evidence to conclude that there is a change in the test score after the educational program
5. p-value $= P(t_9 \geq 3.69)$