7

TESTS OF HYPOTHESES

7.1 Introduction

In scientific activities, a great deal of attention is devoted to answering questions about the validity of theories or hypotheses concerning physical phenomena. Is a new drug effective? Does a lot of manufactured items contain an excessive number of defectives? Is the mean lifetime of a component at least some specified amount? Ordinarily, information about such phenomena can be obtained only by performing experiments the outcomes of which have some bearing on the hypotheses of interest.

The term hypotheses testing will refer to the process of trying to decide which hypotheses are true or false based on experimental evidence. To do it, we need formal model.

Definition 7.1.1. **(Simple/Composite Hypothesis)**. Let $X \sim f(x;\theta)$ (X is "evidence" and may be $X = (X_1, \ldots, X_n) \sim f(x_1, \ldots, x_n; \theta)$; θ can be vector). A statistical hypothesis is a statement about the distribution of X. If the hypothesis completely specifies $f(x;\theta)$, then it is referred to as a simple hypothesis; otherwise it is called composite.

Example 7.1.1.. $X \sim \mathcal{N}(\mu, 1)$, "$\mu = 0$" is a simple hypothesis; "$\mu \neq 0$" is composite since the distribution is not specified.

Example 7.1.2.. $X \sim \mathcal{N}(\mu, \sigma^2)$, "$\mu = 0, \sigma^2 = 1$" is a simple hypothesis; "$\mu = 0$" is composite since the σ^2 is not specified.

Definition 7.1.2. **(Null/Alternative Hypothesis)**. Null hypothesis, denoted "H_0", is usually an accepted theory; the context H_0: "not effect" usually is selected as the status quo or want to control the amount of evidence required to reject it.

Alternative hypothesis, denoted "H_a". is the new theory, intending to replace the old one; it would be costly to implement, but worthwhile if there is sufficient evidence.

In this chapter, we denote Ω (previously we used Ⓗ) as parameter space. Usually we denote values of θ under H_0 by Ω_0 and under H_a by Ω_a. We speak of testing $H_0 : \theta \in \Omega_0$ v.s. $H_a : \theta \in \Omega_a$, where $\Omega_0 \cap \Omega_a = \emptyset$ and $\Omega_0 \cup \Omega_a = \Omega$. In the case of simple hypotheses, these sets consisit of only one element each, $\Omega_0 = \{\theta_0\}$ and $\Omega_a = \{\theta_1\}$, where $\theta_0 \neq \theta_1$.

How to form a test? We take a sample (perform an experiment) where we observe $X \sim f(x;\theta)$. Based on observations of X, decide whether to reject H_0 in favor of H_a.

Definition 7.1.3. **(Critical Region)**. The critical region for testing H_0 v.s. H_a is the subset $\mathcal{S}$ of the sample space (range of the test statistic, X) that corresponds to rejecting H_0. I.e. $\mathcal{S}$ is the critical region if we reject H_0 if $X \in \mathcal{S}$ and do not reject H_0 if $X \in \mathcal{S}^c$[1].

7.2 Two Types of Errors

		True state of nature	
		H_0 true	H_0 false
Action taken	Reject H_0 in favor of H_a	Type I error	No error
	Do not reject H_0 in favor of H_a	No error	Type II error

Note that we speak of rejecting H_0 in favor of H_a or not rejecting H_0 ("status guo") in favor of H_a. Not rejecting H_0 is not the same as accepting H_a; it just means we have insufficient evidence to change from H_0 to H_a.

[1]Complement of $\mathcal{S}$.

$\boxed{\text{Definition 7.2.1.}}$ **(Power Function)**. Let θ be the true value of parameter, then

$$\Pi(\theta) = \mathbb{P}\{\text{Reject } H_0|\theta\} = \mathbb{P}\{X \in \mathcal{S}|\theta\}$$

is called the power function of a test. Here $\mathcal{S}$ is the critical region.

$\boxed{\text{Definition 7.2.2.}}$ **(Significant Level)**. If $\Omega_0 = \{\theta_0\}$, i.e. H_0 is simple, we call $\alpha = \mathbb{P}\{\text{Type I error}\}$ the significance level or size of the test.

If H_0 is composite, we call $\max_{\theta \in \Omega_0} \Pi(\theta) =$ size of the test. If $\Pi(\theta) \leq \alpha$ for all $\theta \in \Omega_0$, α is called a significance level. If $\max_{\theta \in \Omega_0} \Pi(\theta) < \alpha$, the test is a conservative significance level α test.

α controls the amount of evidence required; smaller the α, the more evidence required to reject H_0. Typical α is $0.1, 0.05, 0.01$.

Hence, for simple hypotheses $H_0 : \theta = \theta_0$ v.s. $H_a : \theta = \theta_1$, then

- $\Omega_0 = \{\theta_0\}$, $\alpha = \Pi(\theta_0) = \mathbb{P}\{\text{Type I error}\}$
- $\Omega_a = \{\theta_1\}$, $1 - \beta = \Pi(\theta_1) = 1 - \mathbb{P}\{\text{Type II error}\}$

Later we will try to determine $\mathcal{S}$ by controlling $\Pi(\theta) = \mathbb{P}\{\text{Reject } H_0|\theta\} \leq \alpha$ for all $\theta \in \Omega_0$ while trying to maximize $\mathbb{P}\{\text{Reject } H_0|\theta\}$ for all $\theta \in \Omega_a$. Before we get into examples, let's summarize the process of tests of hypotheses:

$\boxed{\text{Summary 7.2.1.}}$ **(Tests of Hypotheses)**.
Test: $H_0 : \theta \in \Omega_0$ v.s. $H_a : \theta \in \Omega_a$
Goal: to decide whether we reject H_0 or not in favor of H_a
Steps:

Step 1: do an experiment, i.e. observe values $x_1, \ldots, x_n$ of $X_1, \ldots, X_n$ i.i.d. from $f(x;\theta)$

Step 2: find test statistics (e.g. complete-sufficient statistics, pivotal quantity), $T = t(X_1, \ldots, X_n)$, for θ

Step 3: based on the distribution of T, find critical region, S, that satisfies

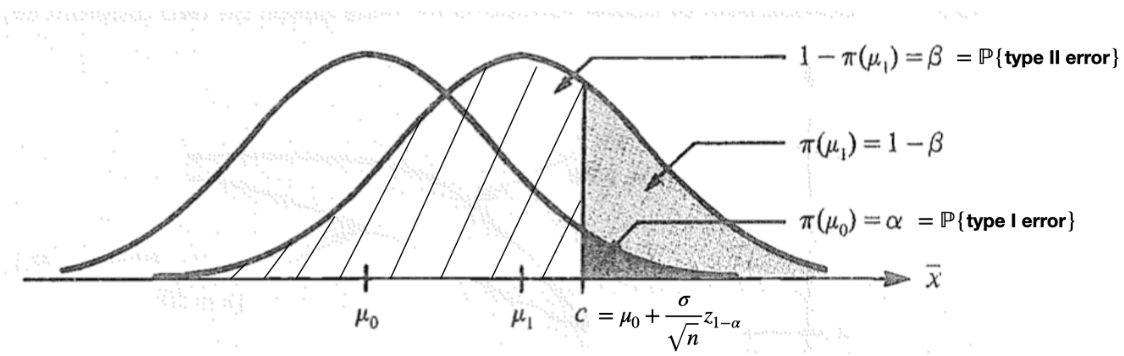
1. $\mathbb{P}\{\text{Reject } H_0|H_0 \text{ true}\} \leq \alpha$
2. $\max \mathbb{P}\{\text{Reject } H_0|H_a \text{ true}\}$

Step 4: the observation of T is $t = t(x_1, \ldots, x_n)$ based on the experiment, and

$$\begin{cases} \text{if } t \in S, \text{ we reject } H_0 \text{ in favor of } H_a \\ \text{if } t \notin S, \text{ we do not reject } H_0 \text{ in favor of } H_a \end{cases}$$

Example 7.2.1.. Suppose the number of years survival under a treatment is distributed as $\mathcal{N}(\mu_0, \sigma^2)$. Now new treatment introduced, purported to have survival distributed as $\mathcal{N}(\mu_1, \sigma^2)$, where $\mu_1 > \mu_0$. Let $X \sim \mathcal{N}(\mu, \sigma^2)$ and assume σ known, to test $H_0 : \mu = \mu_0$ v.s. $H_a : \mu = \mu_1$.

We do an experiment. Let $X_1, \ldots, X_n$ be a random sample from $\mathcal{N}(\mu, \sigma^2)$. $\overline{X}$ is complete sufficient for μ when σ is known. $\overline{X} \sim \mathcal{N}(\mu, \sigma^2/n)$. Because $\mu_1 > \mu_0$, we draw the distributions of $\overline{X}$ under the null and alternative hypotheses in the following plot and want to cut the reject region. Because to reject H_0 in favor of H_a, the value of $\bar{x}$ needs to be on the right starting from point c on the axis to infinity. In other words, the reject region is of form $\mathcal{S} = \{\bar{x} : \bar{x} > c\}$. That is, we will reject H_0 if $\bar{x} \geq c$ and we will not reject H_0 if $\bar{x} < c$.



To find the position of c, we need it to satisfy $\mathbb{P}\{\text{Reject } H_0 | H_0 \text{ true}\} = \mathbb{P}\{\overline{X} \geq c | \mu_0\} \leq \alpha$ and maximize $\mathbb{P}\{\text{Reject } H_0 | H_a \text{ true}\} = \mathbb{P}\{\overline{X} \geq c | \mu_1\}$. In other words, in the plot, we want the dark gray area smaller than α and the light gray area as big as possible, so the c needs to satisfy $\mathbb{P}\{\overline{X} \geq c | \mu_0\} = \alpha$.

Now because $\overline{X} \sim \mathcal{N}(\mu, \sigma^2/n)$,

$$\alpha = \mathbb{P}\{\overline{X} \geq c | \mu_0\} = \mathbb{P}\{\underbrace{\sqrt{n}(\overline{X} - \mu_0)/\sigma}_{\mathcal{N}(0,1)} \geq \sqrt{n}(c - \mu_0)/\sigma | \mu_0\}$$

$$\implies \sqrt{n}(c-\mu_0)/\sigma = z_{1-\alpha} \implies c = \mu_0 + z_{1-\alpha}\sigma/\sqrt{n}$$

So

$$\mathcal{S} = \{\overline{x} : \overline{x} > \mu_0 + z_{1-\alpha}\sigma/\sqrt{n}\}$$

I.e. we will reject H_0 if $\overline{x} \in \mathcal{S}$, i.e. if $\overline{x} > \mu_0 + z_{1-\alpha}\sigma/\sqrt{n}$.

Now let's discuss and draw the plot of the power function

$$\Pi(\mu) = \mathbb{P}\left\{\overline{X} \in \mathcal{S}|\mu\right\} = \mathbb{P}\{\overline{X} > \mu_0 + z_{1-\alpha}\sigma/\sqrt{n}|\mu\}$$

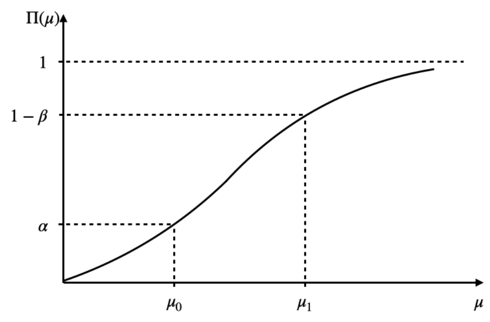
$$= \mathbb{P}\left\{\underbrace{\sqrt{n}\frac{\overline{X}-\mu}{\sigma}}_{\mathcal{N}(0,1)} > \sqrt{n}\frac{\mu_0-\mu}{\sigma} + z_{1-\alpha}\Big|\mu\right\} = 1 - \Phi\left(\sqrt{n}\frac{\mu_0-\mu}{\sigma} + z_{1-\alpha}\right)$$

So

$$\alpha = \mathbb{P}\{\text{type I error} = \Pi(\mu_0) = 1 - \Phi\left(z_{1-\alpha}\right)\}$$

$$1-\beta = 1 - \mathbb{P}\{\text{type II error}\} = \Pi(\mu_1) = 1 - \Phi\left(\sqrt{n}\underbrace{\frac{\mu_0-\mu_1}{\sigma}}_{<0} + z_{1-\alpha}\right) > \alpha$$

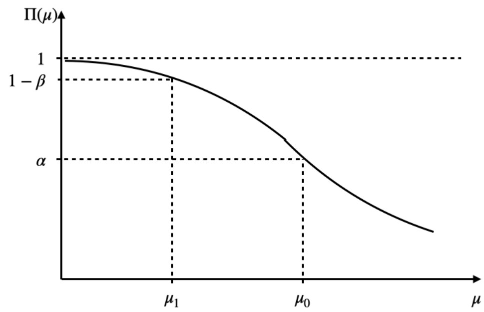
Now we can draw the plot of power function:



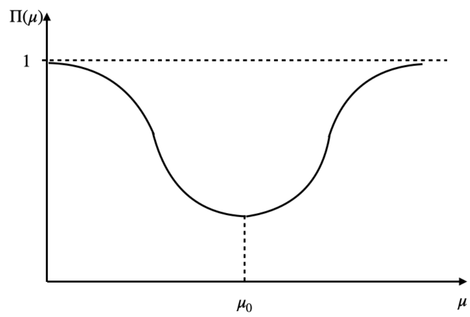
What if μ_1 is not given (i.e. $H_a : \mu > \mu_0$, which is composite)? The above test still works.

What if $H_0 : \mu \leq \mu_0$, $H_a : \mu > \mu_0$? The above test still works and $\max_{\mu \leq \mu_0} \Pi(\mu) = \alpha =$ size of test.

What if $H_0 : \mu \geq \mu_0$, $H_a : \mu < \mu_0$? Use the other tail, or $\mathcal{S} = \{\overline{x} : \overline{x} < \mu_0 + z_\alpha \sigma/\sqrt{n}\}$ and the power function becomes:



What if $H_0 : \mu = \mu_0$, $H_a : \mu \neq \mu_0$? Use two-sided test, or $\mathcal{S} = \{\overline{x} : \overline{x} < \mu_0 + z_{\alpha/2}\sigma/\sqrt{n}$ or $\overline{x} > \mu_0 + z_{1-\alpha/2}\sigma/\sqrt{n}\}$ and the power function becomes:



To summarize the above results:

$\boxed{\text{Theorem 7.2.1.}}$ **(Tests for the Mean of Normal (Variance Known)).** Suppose that $x_1 \ldots, x_n$ is an observed random sample from $\mathcal{N}\left(\mu, \sigma^2\right)$, where σ^2 is known, and let $z_0 = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$

1. A size α test of $H_0 : \mu \leq \mu_0$ versus $H_a : \mu > \mu_0$ is to reject H_0 if $z_0 \geq z_{1-\alpha}$. The power function for this test is

$$\Pi(\mu) = 1 - \Phi\left(z_{1-\alpha} + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right)$$

2. A size α test of $H_0 : \mu \geq \mu_0$ versus $H_a : \mu < \mu_0$ is to reject H_0 if $z_0 \leq -z_{1-\alpha}$. The power function for this test is

$$\Pi(\mu) = \Phi\left(-z_{1-\alpha} + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right)$$

3. A size α test of $H_0 : \mu = \mu_0$ versus $H_a : \mu \neq \mu_0$ is to reject H_0 if $z_0 \leq -z_{1-\alpha/2}$ or $z_0 \geq z_{1-\alpha/2}$.

4. The sample size required to achieve a size α test with power $1-\beta$ for an alternative value μ is given by

$$n = \frac{\left(z_{1-\alpha} + z_{1-\beta}\right)^2 \sigma^2}{\left(\mu_0 - \mu\right)^2}$$

for a one-sided test, and

$$n = \frac{\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2 \sigma^2}{\left(\mu_0 - \mu\right)^2}$$

for a two-sided test.

To obtain the results in part 4, e.g., go back to the power function analysis of the first case in Example 7.2.1, i.e. the following line

$$\Pi(\mu_1) = 1 - \Phi\left(\sqrt{n}\underbrace{\frac{\mu_0 - \mu_1}{\sigma}}_{<0} + z_{1-\alpha}\right) = 1 - \mathbb{P}\{\text{type II error}\} = 1 - \beta$$

which implies that

$$\Phi\left(\sqrt{n}\underbrace{\frac{\mu_0-\mu_1}{\sigma}}_{<0}+z_{1-\alpha}\right)=\beta \implies \sqrt{n}\underbrace{\frac{\mu_0-\mu_1}{\sigma}}_{<0}+z_{1-\alpha}=z_\beta$$

$$\implies n=\frac{\left(z_{1-\alpha}+z_{1-\beta}\right)^2\sigma^2}{(\mu_0-\mu)^2}$$

Theorem 7.2.2. **(Tests for the Mean of Normal (Variance Unknown))**. The last theorem handles μ with σ^2 known. Now let $X_1,\dots,X_n$ be a random sample (with observed values $x_1,\dots,x_n$) from $\mathcal{N}(\mu,\sigma^2)$, with both μ, σ^2 unknown. Want to test:

1. $H_0:\mu\le\mu_0$ v.s. $H_a:\mu>\mu_0$
   We reject H_0 if $\sqrt{n}\frac{\overline{x}-\mu_0}{s}\ge t_{1-\alpha}(n-1)$ or if $\overline{x}\ge\mu_0+\frac{s}{\sqrt{n}}t_{1-\alpha}(n-1)$
2. $H_0:\mu\ge\mu_0$ v.s. $H_a:\mu<\mu_0$
   We reject H_0 if $\sqrt{n}\frac{\overline{x}-\mu_0}{s}\le t_{\alpha}(n-1)$ or if $\overline{x}\le\mu_0+\frac{s}{\sqrt{n}}t_{\alpha}(n-1)$.
3. $H_0:\mu=\mu_0$ v.s. $H_a:\mu\ne\mu_0$
   We reject H_0 if $\sqrt{n}\frac{\overline{x}-\mu_0}{s}\ge t_{1-\alpha/2}(n-1)$ or $\sqrt{n}\frac{\overline{x}-\mu_0}{s}\le t_{\alpha/2}(n-1)$, i.e. if $\overline{x}\ge\mu_0+\frac{s}{\sqrt{n}}t_{1-\alpha/2}(n-1)$ or $\overline{x}\le\mu_0+\frac{s}{\sqrt{n}}t_{\alpha/2}(n-1)$.

Proof. (Ad-Hoc Approaches) For part 1, how to proceed? Note that these are composite hypotheses. To reduce by sufficiency leads us to consider test procedure depends on $\overline{X}=\frac{1}{n}\sum_{i=1}^n X_i$, $S^2=\frac{1}{n-1}\sum_{i=1}^n(X_i-\overline{X})^2$.

Follow the same discussion in Example 7.2.1, we would like to test that rejects H_0 for large values of $\overline{X}$. How about using $\overline{X}$ as the test statistic so $\mathcal{S}=\{\overline{x}:\overline{x}\ge c\}$? To find c, we solve

$$\alpha=\mathbb{P}\{\overline{X}\in S|\mu_0,\sigma^2\}=\mathbb{P}\{\overline{X}\ge c|\mu_0,\sigma^2\}$$

$$=\mathbb{P}\{\underbrace{\sqrt{n}\frac{\overline{X}-\mu_0}{\sigma}}_{\sim\mathcal{N}(0,1)}\ge\sqrt{n}\frac{c-\mu_0}{\sigma}|\mu_0,\sigma^2\}=1-\Phi\left(\sqrt{n}\frac{c-\mu_0}{\sigma}\right)$$

which depends on σ!

So we try the pivotal quantity used to setting confidence interval on μ for σ^2 unknown. Recall $\sqrt{n}\frac{\overline{X}-\mu}{S} \sim t(n-1)$, so we can use $\overline{X}$ and S^2 as the test statistics, so $\mathcal{S} = \{(\overline{x}, s^2) : \sqrt{n}\frac{\overline{x}-\mu_0}{s} \geq c\}$. Now

$$\alpha = \mathbb{P}\{(\overline{X}, S^2) \in \mathcal{S} | \mu_0, \sigma^2\} = \mathbb{P}\left\{ \underbrace{\sqrt{n}\frac{\overline{X}-\mu_0}{S}}_{t(n-1)} \geq c | \mu_0, \sigma^2 \right\}$$

So we can get size α-test with $c = t_{1-\alpha}(n-1)$. I.e. we reject H_0 if $\sqrt{n}\frac{\overline{x}-\mu_0}{s} \geq t_{1-\alpha}(n-1)$ or if $\overline{x} \geq \mu_0 + \frac{s}{\sqrt{n}} t_{1-\alpha}(n-1)$.

Student can verify that part 2,3 give size α tests. $\square$

In general, the approach is to reduce by sufficiency. We use pivotal function that would be used for setting confidence intervals and then set up reject region on intuitive grounds.

The following theorem gives tests for variance σ^2 with μ unknown. It uses $(n-1)S^2/\sigma_0^2 \sim \chi^2(n-1)$ to test hypotheses about σ^2.

Theorem 7.2.3. **(Tests for Variance of Normal (Mean Unknown)).** Let $x_1, \ldots, x_n$ be an observed random sample from $\mathcal{N}(\mu, \sigma^2)$, and let $H(c; v)$ be the CDF of $\chi^2(v)$ and let

$$v_0 = (n-1)s^2/\sigma_0^2$$

1. A size α test of $H_0 : \sigma^2 \leq \sigma_0^2$ versus $H_a : \sigma^2 > \sigma_0^2$ is to reject H_0 if $v_0 \geq \chi^2_{1-a}(n-1)$. The power function for this test is

$$\Pi\left(\sigma^2\right) = 1 - H\left(\left(\sigma_0^2/\sigma^2\right)\chi^2_{1-\alpha}(n-1); n-1\right)$$

2. A size α test of $H_0 : \sigma^2 \geq \sigma_0^2$ versus $H_a : \sigma^2 < \sigma_0^2$ is to reject H_0 if $v_0 \leq \chi^2_a(n-1)$. The power function for this test is

$$\Pi\left(\sigma^2\right) = H\left(\left(\sigma_0^2/\sigma^2\right)\chi^2_a(n-1); n-1\right)$$

3. A size α test of $H_0 : \sigma^2 = \sigma_0^2$ versus $H_a : \sigma^2 \neq \sigma_0^2$ is to reject H_0 if $v_0 \leq \chi^2_{\alpha/2}(n-1)$ or $v_0 \geq \chi^2_{1-\alpha/2}(n-1)$

The following theorem tests of σ_2^2/σ_1^2 using $\frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} \sim F(n_1 - 1, n_2 - 1)$.

Theorem 7.2.4. **(Two-Sample Tests for Variables)**. Suppose that $x_1, \ldots, x_{n_1}$ and $y_1, \ldots, y_{n_2}$ are observed values of independent random samples from $\mathrm{N}\left(\mu_1, \sigma_1^2\right)$ and $\mathrm{N}\left(\mu_2, \sigma_2^2\right)$, respectively, and let

$$f_0 = \frac{s_1^2}{s_2^2} d_0$$

1. A size α test of $H_0 : \sigma_2^2/\sigma_1^2 \leq d_0$ versus $H_a : \sigma_2^2/\sigma_1^2 > d_0$ is to reject H_0 if $f_0 \leq 1/F_{1-\alpha}\left(n_2 - 1, n_1 - 1\right)$.
2. A size α test of $H_0 : \sigma_2^2/\sigma_1^2 \geq d_0$ versus $H_a : \sigma_2^2/\sigma_1^2 < d_0$ is to reject H_0 if $f_0 \geq F_{1-\alpha}\left(n_1 - 1, n_2 - 1\right)$.
3. A size α test of $H_0 : \sigma_2^2/\sigma_1^2 = d_0$ versus $H_a : \sigma_2^2/\sigma_1^2 \neq d_0$ is to reject H_0 if $f_0 \leq 1/F_{1-\alpha/2}\left(n_2 - 1, n_1 - 1\right)$ or $f_0 \geq F_{1-\alpha/2}\left(n_2 - 1, n_2 - 1\right)$.

The following theorem tests of $\mu_2 - \mu_1$, with σ^2 unknown but same, using

$$\frac{\overline{Y} - \overline{X} - (\mu_2 - \mu_1)}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \sim t(n_1 + n_2 - 2)$$

Theorem 7.2.5. **(Two-Sample Tests for Means)**. Suppose that $x_1, \ldots, x_{n_1}$ and $y_1, \ldots, y_{n_2}$ are observed values of independent random samples from $\mathrm{N}\left(\mu_1, \sigma_1^2\right)$ and $\mathrm{N}\left(\mu_2, \sigma_2^2\right)$, respectively, where $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Let

$$t_0 = \frac{\bar{y} - \bar{x} - d_0}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

1. A size α test of $H_0 : \mu_2 - \mu_1 \leq d_0$ versus $H_a : \mu_2 - \mu_1 > d_0$ is to reject H_0 if $t_0 \geq t_{1-\alpha}\left(n_1 + n_2 - 2\right)$.
2. A size α test of $H_0 : \mu_2 - \mu_1 \geq d_0$ versus $H_a : \mu_2 - \mu_1 < d_0$ is to reject H_0 if $t_0 \leq -t_{1-\alpha}\left(n_1 + n_2 - 2\right)$.
3. A size α test of $H_0 : \mu_2 - \mu_1 = d_0$ versus $H_a : \mu_2 - \mu_1 \neq d_0$ is to reject H_0 if $t_0 \leq -t_{1-\alpha/2}\left(n_1 + n_2 - 2\right)$ or $t_0 \geq t_{1-\alpha/2}\left(n_1 + n_2 - 2\right)$.