# Chapter 9: Hypothesis Testing

## Sections

# Uniformly Most Powerful Tests

$$H_0 : \theta \in \Omega_0 \qquad \text{vs} \qquad H_1 : \theta \in \Omega_1$$

- A test $\delta^*$ is a *uniformly most powerful test* at level $\alpha_0$ if for any other level $\alpha_0$ test $\delta$

$$\pi(\theta|\delta) \leq \pi(\theta|\delta^*) \qquad \text{for all } \theta \in \Omega_1$$

  I.o.w: It has the lowest probability of type II error of any test, uniformly for all $\theta \in \Omega_1$.

# Uniformly Most Powerful Tests

$$H_0 : \theta \in \Omega_0 \qquad \text{vs} \qquad H_1 : \theta \in \Omega_1$$

- A test $\delta^*$ is a *uniformly most powerful test* at level $\alpha_0$ if for any other level $\alpha_0$ test $\delta$

$$\pi(\theta|\delta) \leq \pi(\theta|\delta^*) \qquad \text{for all } \theta \in \Omega_1$$

  I.o.w: It has the lowest probability of type II error of any test, uniformly for all $\theta \in \Omega_1$.
- We control the probability of type I error by setting the level (size) of the test low. We then want to control the probability of type II error.

# Uniformly Most Powerful Tests

$$H_0 : \theta \in \Omega_0 \qquad \text{vs} \qquad H_1 : \theta \in \Omega_1$$

- A test $\delta^*$ is a *uniformly most powerful test* at level $\alpha_0$ if for any other level $\alpha_0$ test $\delta$

  $$\pi(\theta|\delta) \leq \pi(\theta|\delta^*) \qquad \text{for all } \theta \in \Omega_1$$

  I.o.w: It has the lowest probability of type II error of any test, uniformly for all $\theta \in \Omega_1$.
- We control the probability of type I error by setting the level (size) of the test low. We then want to control the probability of type II error.
- If $\pi(\theta|\delta^*)$ is high for all $\theta \in \Omega_1$, the test is often called "powerful"
- In a large class of problems (the distribution has a "monotone likelihood ratio") we can find a uniformly most powerful test for one-sided hypotheses (Ch. 9.3).

# The *t*-Test

- The *t*-Test is a test for hypotheses concerning the mean parameter in the normal distribution when the variance is also unknown.
- The test is based on the *t* distribution

The setup for the next few slides:

- Let $X_1, \dots, X_n$ be i.i.d. $N(\mu, \sigma^2)$ and consider the hypotheses

$$H_0 : \mu \leq \mu_0 \qquad \text{vs.} \qquad H_1 : \mu > \mu_0 \tag{1}$$

The parameter space here is $-\infty < \mu < \infty$ and $\sigma^2 > 0$, i.e.

$$\Omega = (-\infty, \infty) \times (0, \infty)$$

And

$$\Omega_0 = (-\infty, \mu_0] \times (0, \infty) \quad \text{and} \quad \Omega_1 = (\mu_0, \infty) \times (0, \infty)$$

# The one-sided *t*-Test

- Let

$$U = \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{\sigma'} \quad \text{where} \quad \sigma' = \left( \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2 \right)^{1/2}$$

- If $\mu = \mu_0$ then $U$ has the $t$ distribution
- Tests based on $U$ are called *t tests*

# The one-sided $t$-Test

- Let $T_n^{-1}$ be the quantile function of the $t_n$ distribution

## Theorem 9.5.1

The test $\delta$ that rejects $H_0$ in (1) if $U \geq T_{n-1}^{-1}(1-\alpha_0)$ has size $\alpha_0$ and a power function with the following properties

(i) $\pi(\mu_0, \sigma^2|\delta) = \alpha_0$

(ii) $\pi(\mu, \sigma^2|\delta) < \alpha_0$ for $\mu < \mu_0$

(iii) $\pi(\mu, \sigma^2|\delta) > \alpha_0$ for $\mu > \mu_0$

(iv) $\pi(\mu, \sigma^2|\delta) \to 0$ as $\mu \to -\infty$

(v) $\pi(\mu, \sigma^2|\delta) \to 1$ as $\mu \to \infty$

# The complete power function

## For the one-sided $t$-test

To calculate the power function $\pi(\mu, \sigma^2|\delta)$ exactly we need the non-central $t_m$ distributions:

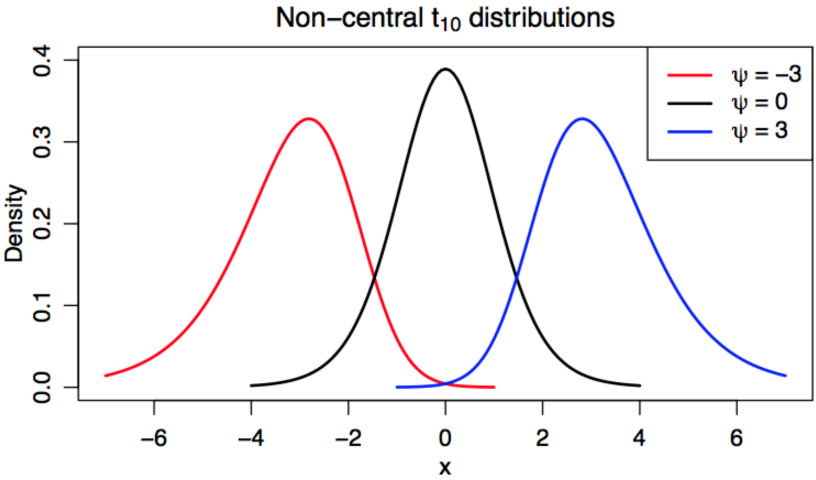
**Def: Non-central $t_m$ distributions**

Let $W \sim N(\psi, 1)$ and $Y \sim \chi^2_m$ be independent. The distribution of

$$X = \frac{W}{(Y/m)^{1/2}}$$

is called the *non-central t distribution with m degrees of freedom and non-centrality parameter* $\psi$

# Non-central $t_m$ distribution



# The complete power function

## For the one-sided *t*-test

**Theorem 9.5.3**

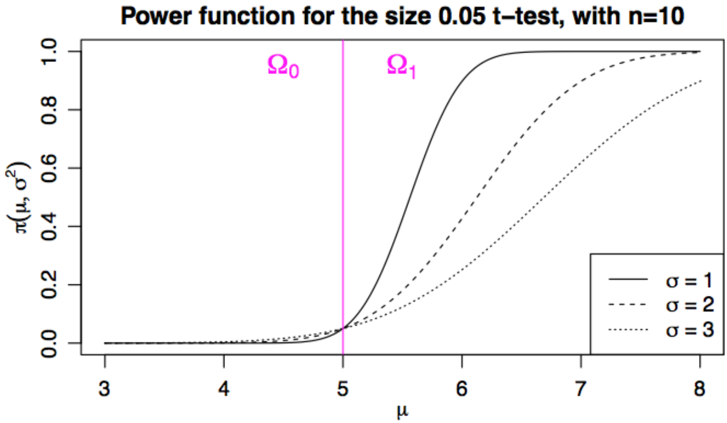
$U$ has the non-central $t_{n-1}$ distribution with non-centrality parameter $\psi = \sqrt{n}(\mu - \mu_0)/\sigma$.
The power function of the t-test that rejects $H_0$ in (1) if $U \geq T_{n-1}^{-1}(1 - \alpha_0) = c_1$ is

$$\pi(\mu, \sigma^2|\delta) = 1 - T_{n-1}(c_1|\psi)$$

- Can use the R function `1 - pt(q=c1, df=n-1, ncp=psi)`

# Power function for the one-sided *t*-test

## Example: $n = 10$, $\mu_0 = 5$, $\alpha_0 = 0.05$




Note that the power function is a function of both $\sigma^2$ and $\mu$

# p-value for the one-sided *t*-Test

**Theorem 9.5.2: p-values for *t* Tests**

Let $u$ be the observed value of $U$.
The p-value for the hypothesis in (1) is $1 - T_{n-1}(u)$.

Example: Acid Concentration in Cheese (Example 8.5.4)

- Have a random sample of $n = 10$ lactic acid measurements from cheese, assumed to be from a normal distribution with unknown mean and variance.
- Observed: $\overline{x}_n = 1.379$ and $\sigma' = 0.3277$
- Perform the level $\alpha_0 = 0.05$ *t*-test of the hypotheses

$$H_0 : \mu \leq 1.2 \quad \text{vs} \quad H_1 : \mu > 1.2$$

- Compute the p-value

# The other one-sided $t$-Test

- Now consider the hypotheses

$$H_0 : \mu \geq \mu_0 \qquad \text{vs.} \qquad H_1 : \mu < \mu_0 \tag{2}$$

## Corollary 9.5.1

The test $\delta$ that rejects $H_0$ if $U \leq T_{n-1}^{-1}(\alpha_0)$ has size $\alpha_0$ and a power function with the following properties

(i) $\pi(\mu_0, \sigma^2|\delta) = \alpha_0$

(ii) $\pi(\mu, \sigma^2|\delta) > \alpha_0$ for $\mu < \mu_0$

(iii) $\pi(\mu, \sigma^2|\delta) < \alpha_0$ for $\mu > \mu_0$

(iv) $\pi(\mu, \sigma^2|\delta) \to 1$ as $\mu \to -\infty$

(v) $\pi(\mu, \sigma^2|\delta) \to 0$ as $\mu \to \infty$

# Power function and p-value for the other one-sided $t$-Test

**Theorem 9.5.2: p-values for $t$ Tests**

Let $u$ be the observed value of $U$.
The p-value for the hypothesis in (2) is $T_{n-1}(u)$.

**Theorem 9.5.3**

$U$ has the non-central $t_{n-1}$ distribution with non-centrality parameter $\psi = \sqrt{n}(\mu - \mu_0)/\sigma$.
The power function of the t-test that rejects $H_0$ in (2) if $U \leq T_{n-1}^{-1}(\alpha_0) = c_2$ is

$$\pi(\mu, \sigma^2|\delta) = T_{n-1}(c_2|\psi)$$

# Two-sided $t$-test

Consider now the test with a two-sided alternative hypothesis:

$$H_0 : \mu = \mu_0 \qquad \text{vs.} \qquad H_1 : \mu \neq \mu_0 \tag{3}$$

- Let $\delta$ be the test that rejects $H_0$ iff $|U| \geq T_{n-1}^{-1}(1 - \alpha_0/2) = c$
- Then $\delta$ is a size $\alpha_0$ test
- The power function is

$$\pi(\mu, \sigma^2|\delta) = T_{n-1}(-c|\psi) + 1 - T_{n-1}(c|\psi)$$

- If $u$ is the observed value of $U$ then the p-value is $2(1 - T_{n-1}(|u|))$

The $t$ test is a likelihood ratio test (see p. 583 - 585 in the book)

# The two-sample $t$-test

Comparing the means of two populations

- $X_1, \dots, X_m$ i.i.d. $N(\mu_1, \sigma^2)$ and
  $Y_1, \dots, Y_n$ i.i.d. $N(\mu_2, \sigma^2)$
- The variance is the same for both samples, but unknown

We are interested in testing one of these hypotheses:

a) $H_0 : \mu_1 \leq \mu_2$ vs. $H_1 : \mu_1 > \mu_2$

b) $H_0 : \mu_1 \geq \mu_2$ vs. $H_1 : \mu_1 < \mu_2$

c) $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$

Power function is now a function of 3 parameters: $\pi(\mu_1, \mu_2, \sigma^2 | \delta)$

# Two-sample $t$ statistic

Let $$\overline{X}_m = \frac{1}{m}\sum_{i=1}^{m} X_i \quad \text{and} \quad \overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

$$S_X^2 = \sum_{i-1}^{m}(X_i - \overline{X}_m)^2 \quad \text{and} \quad S_Y^2 = \sum_{i-1}^{n}(Y_i - \overline{Y}_n)^2$$

$$U = \frac{\sqrt{m+n-2}\,(\overline{X}_m - \overline{Y}_n)}{\left(\frac{1}{m}+\frac{1}{n}\right)^{1/2}\left(S_X^2 + S_Y^2\right)^{1/2}}$$

- Theorem 9.6.1: If $\mu_1 = \mu_2$ then $U \sim t_{m+n-2}$
- Theorem 9.6.4: For any $\mu_1$ and $\mu_2$, $U$ has the non-central $t_{m+n-2}$ distribution with non-centrality parameter

$$\psi = \frac{\mu_1 - \mu_2}{\sigma\,(1/m + 1/n)^{1/2}}$$

# Two-sample $t$ test – summary

## Proofs similar to the regular $t$-test

a) $H_0 : \mu_1 \leq \mu_2$ vs. $H_1 : \mu_1 > \mu_2$

- Level $\alpha_0$ test: Reject $H_0$ iff $U \geq T^{-1}_{m+n-2}(1 - \alpha_0)$
- p-value: $1 - T_{m+n-2}(u)$

b) $H_0 : \mu_1 \geq \mu_2$ vs. $H_1 : \mu_1 < \mu_2$

- Level $\alpha_0$ test: Reject $H_0$ iff $U \leq T^{-1}_{m+n-2}(\alpha_0)$
- p-value: $T_{m+n-2}(u)$

c) $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$

- Level $\alpha_0$ test: Reject $H_0$ iff $|U| \geq T^{-1}_{m+n-2}(1 - \alpha_0/2)$
- p-value: $2(1 - T_{m+n-2}(|u|))$

The two-sample $t$-test is a likelihood ratio test (see p. 592)

# Two-sample *t* test – unequal variances

- We can extend the two sample $t$-test to a problem where the variances of the $X_i$'s and $Y_j$'s are not equal but the ratio of them is known, i.e. $\sigma_1^2 = k\sigma_2^2$ – Not very practical

In general, the problem where the variances are not equal is very hard.

# Two-sample *t* test – unequal variances

- We can extend the two sample *t*-test to a problem where the variances of the $X_i$'s and $Y_j$'s are not equal but the ratio of them is known, i.e. $\sigma_1^2 = k\sigma_2^2$ – Not very practical

In general, the problem where the variances are not equal is very hard.

- Proposed test-statistics do not have known distribution, but approximations have been obtained
- Example: The Welch statistic

$$V = \frac{\overline{X}_m - \overline{Y}_n}{\left(\frac{S_X^2}{m(m-1)} + \frac{S_Y^2}{n(n-1)}\right)^{1/2}}$$

can be approximated by a *t* distribution

# Two-sample *t* test – unequal variances

- We can extend the two sample *t*-test to a problem where the variances of the $X_i$'s and $Y_j$'s are not equal but the ratio of them is known, i.e. $\sigma_1^2 = k\sigma_2^2$ – Not very practical

In general, the problem where the variances are not equal is very hard.

- Proposed test-statistics do not have known distribution, but approximations have been obtained
- Example: The Welch statistic

$$V = \frac{\overline{X}_m - \overline{Y}_n}{\left(\frac{S_X^2}{m(m-1)} + \frac{S_Y^2}{n(n-1)}\right)^{1/2}}$$

  can be approximated by a *t* distribution
- Example: The distribution of the likelihood ratio statistic can be approximated by the $\chi_1^2$ distribution if the sample size is large enough

# F-distributions

- In light of the previous slide, it would be nice to have a test of whether the variances in the two normal populations are equal $\rightarrow$ need the $F_{m,n}$ distributions

**Def: $F_{m,n}$-distributions**

Let $Y \sim \chi^2_m$ and $W \sim \chi^2_n$ be independent. The distribution of

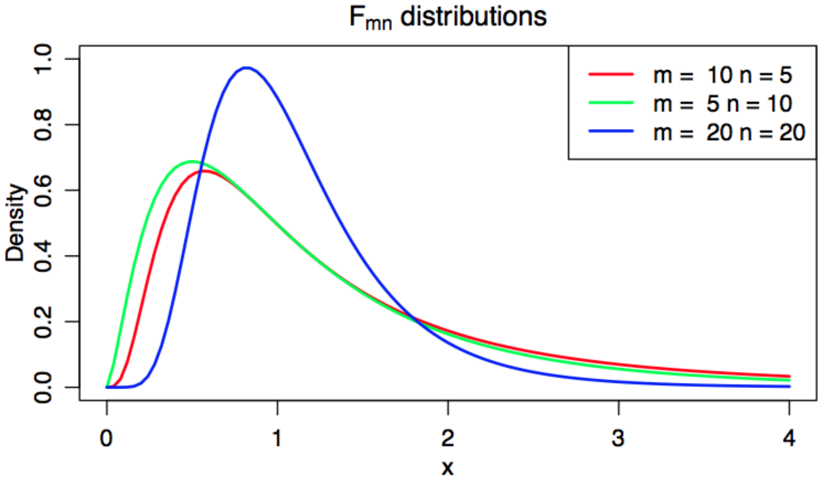
$$X = \frac{Y/m}{W/n} = \frac{nY}{mW}$$

is called the *F distribution with m and n degrees of freedom*

The pdf of the $F_{m,n}$ distribution is

$$f(x) = \frac{\Gamma\left((m+n)/2\right) m^{m/2} n^{n/2}}{\Gamma(m/2)\Gamma(n/2)} \frac{x^{m/2-1}}{(mx+n)^{(m+n)/2}} \quad x > 0$$

# F-distributions



# Properties of the F-distributions

- The 0.95 and 0.975 quantiles of the $F_{m,n}$ distribution is tabulated in the back of the book for a few combinations of $m$ and $n$

**Theorem 9.7.2: Two properties**

(i) If $X \sim F_{m,n}$ then $1/X \sim F_{n,m}$

(ii) If $Y \sim t_n$ then $Y^2 \sim F_{1,n}$

# Comparing the variances of two normals

Comparing the variances of two populations

- $X_1, \ldots, X_m$ i.i.d. $N(\mu_1, \sigma_1^2)$ and
  $Y_1, \ldots, Y_n$ i.i.d. $N(\mu_2, \sigma_2^2)$ All four parameters unknown

Consider the hypotheses:

(I) $H_0 : \sigma_1^2 \leq \sigma_2^2$ vs. $H_1 : \sigma_1^2 > \sigma_2^2$

and the test that rejects $H_0$ if $V \geq c$, where

$$V = \frac{S_X^2/(m-1)}{S_Y^2/(n-1)}$$

This test is called an *F-test*

- $\frac{\sigma_2^2}{\sigma_1^2} V \sim F_{m-1,n-1}$
- If $\sigma_1^2 = \sigma_2^2$ then $V \sim F_{m-1,n-1}$

# The $F$ test

Let $G_{m,n}(x)$ be the cdf of the $F_{m,n}$ distribution

**Theorem 9.7.4**

Let $\delta$ be the test that rejects $H_0$ in

$$H_0 : \sigma_1^2 \leq \sigma_2^2 \quad \text{vs.} \quad H_1 : \sigma_1^2 > \sigma_2^2$$

if $V \geq c = G^{-1}_{m-1,n-1}(1-\alpha_0)$. Then $\delta$ is of size $\alpha_0$ and

- $\pi(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2 | \delta) = 1 - G_{m-1,n-1}\left(\frac{\sigma_2^2}{\sigma_1^2} c\right)$ and
- p-value $= 1 - G_{m-1,n-1}(v)$, where $v$ is the observed value of $V$

# Course evaluation

**Course number: 4340**
**"Major": 53 (for graduate school)**

**I am Inst.1**

Reminders for students:

- Use only pencil or blue or black ink when completing the form
- Complete both sides of the form
- Utilize areas left for comments
- Do not fold, stable, rip or otherwise deface the evaluation form