Chapter 9: Hypothesis Testing

Sections

- 9.1 Problems of Testing Hypotheses
- 9.2 Testing Simple Hypotheses
- 9.3 Uniformly Most Powerful Tests
- Skip: 9.4 Two-Sided Alternatives
- 9.5 The $t$ Test
- 9.6 Comparing the Means of Two Normal Distributions
- 9.7 The $F$ Distributions
- 9.8 Bayes Test Procedures
- 9.9 Foundational Issues
### Uniformly Most Powerful (UMP) Tests

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

\[
\pi(\theta|\delta) \leq \pi(\theta|\delta^*) \quad \text{for all } \theta \in \Omega_1
\]

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

- First control the probability of type I error by setting the level (size) of the test low, then 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”
Example 1: Simple hypotheses.
    - $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$
    - LRT is UMP by Neyman-Pearson lemma (Theorem 9.2.2)

Example 2: One-sided hypotheses:
    - $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$
    - In a large class of problems (the distribution has a “monotone likelihood ratio”), we can show that “reject $H_0$ if $T \geq t$” is a UMP for some $T$ (Ch 9.3)

Example 3: Two-sided hypotheses:
    - $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$
    - UMP tests do not exist (Page 565)
Chapter 9: Hypothesis Testing

Sections

- 9.1 Problems of Testing Hypotheses
- Skip: 9.2 Testing Simple Hypotheses
- 9.3 Uniformly Most Powerful Tests
- Skip: 9.4 Two-Sided Alternatives
- 9.5 The $t$ Test
- 9.6 Comparing the Means of Two Normal Distributions
- 9.7 The $F$ Distributions
- 9.8 Bayes Test Procedures
- 9.9 Foundational Issues
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 $X_1, \ldots, X_n$ be i.i.d. $N(\mu, \sigma^2)$ and consider the hypotheses

$$H_0 : \mu \leq \mu_0 \quad \text{vs.} \quad H_1 : \mu > \mu_0$$

(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

- The $t$ test: a likelihood ratio test (see p. 583 - 585 in the book)
- Let
  \[ U = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\sigma'} \quad \text{where} \quad \sigma' = \left( \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \right)^{1/2} \]
- If $\mu = \mu_0$ then $U$ has the $t_{n-1}$ distribution
- Tests based on $U$ are called $t$ tests
The one-sided $t$-Test

- Let $T_{m}^{-1}$ be the quantile function of the $t_{m}$ distribution.
- The test $\delta$ that rejects $H_{0}$ in (1) if $U \geq T_{n-1}^{-1}(1 - \alpha_{0})$ has size $\alpha_{0}$ (Theorem 9.5.1).
- To calculate the p-value:

**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: $\bar{x}_n = 1.379$ 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 complete power function

- Need the power function to decide the sample size $n$
- The power function $\pi(\mu, \sigma^2|\delta)$ is a 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

$$\pi(\mu, \sigma^2|\delta) = 1 - T_{n-1}(c_1|\psi)$$
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 $\mu$
The other one-sided $t$-Test

- Now consider the hypothesis
  \[ H_0 : \mu \geq \mu_0 \quad \text{vs.} \quad H_1 : \mu < \mu_0 \]  
  (2)

- The test $\delta$ that rejects $H_0$ if $U \leq T_{n-1}^{-1}(\alpha_0)$ has size $\alpha_0$ (Corollary 9.5.1)

**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:

$$H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_1 : \mu \neq \mu_0$$  \hspace{1cm} (3)

- Size $\alpha_0$ test $\delta$: rejects $H_0$ iff $|U| \geq T_{n-1}^{-1}(1 - \alpha_0/2) = c$
- If $u$ is the observed value of $U$ then the p-value is $2(1 - T_{n-1}(|u|))$
- The power function is

$$\pi(\mu, \sigma^2 | \delta) = T_{n-1}(-c|\psi) + 1 - T_{n-1}(c|\psi)$$
Notes on one sample $t$ tests

- Paired $t$ tests are conducted in the same way.
- For large $n$, the distribution of the test statistic under $H_0$ is close to the standard normal, i.e., the corresponding test is close to a $Z$ test.
The two-sample $t$-test

Comparing the means of two populations
- $X_1, \ldots, X_m$ i.i.d. $N(\mu_1, \sigma^2)$ and $Y_1, \ldots, Y_n$ 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. $H_1 : \mu_1 \neq \mu_2$
### Two-sample $t$ statistic

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

$S^2_X = \sum_{i=1}^{m} (X_i - \overline{X}_m)^2$ and $S^2_Y = \sum_{i=1}^{n} (Y_i - \overline{Y}_n)^2$

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

- **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

$$\psi = \frac{\mu_1 - \mu_2}{\sigma \left( \frac{1}{m} + \frac{1}{n} \right)^{1/2}}$$
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 $U \geq T_{m+n-2}^{-1}(1 - \alpha_0)$
   - 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 $U \leq T_{m+n-2}^{-1}(\alpha_0)$
   - 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 $|U| \geq T_{m+n-2}^{-1}(1 - \alpha_0/2)$
   - p-value: $2(1 - T_{m+n-2}(|u|))$
Power function is now a function of 3 parameters: $\pi(\mu_1, \mu_2, \sigma^2 | \delta)$

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

Important difference: **Paired** $t$ test vs. two sample $t$ test

Two-sample $t$ test with unequal variances

- Proposed test-statistics do not have known distribution, but approximations have been obtained
- Approach 1: The Welch statistic

$$V = \frac{\bar{X}_m - \bar{Y}_n}{\left( \frac{S^2_X}{m(m-1)} + \frac{S^2_Y}{n(n-1)} \right)^{1/2}}$$

- can be approximated by a $t$ distribution
- Approach 2: The distribution of the likelihood ratio statistic can be approximated by the $\chi^2_1$ 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 → 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

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

The F-distributions are used in hypothesis testing, particularly in the context of analysis of variance (ANOVA), to compare variances from different populations. The graph shows different F distributions for various parameter values:

- Red line: $m = 10$, $n = 5$
- Green line: $m = 5$, $n = 10$
- Blue line: $m = 20$, $n = 20$

The x-axis represents the F statistic, and the y-axis represents the density of the F distribution.
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 \textit{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 $V \geq c = G_{m-1,n-1}^{-1}(1 - \alpha_0)$. 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 $= 1 - G_{m-1,n-1}(\nu)$, where $\nu$ is the observed value of $V$
The $F$ test – two sided alternative

Two sided alternative:

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad \text{vs.} \quad H_1 : \sigma_1^2 \neq \sigma_2^2$$

Equal-tailed two-sided $F$ test

Let $\delta$ be the $F$-test that rejects $H_0$ when

$$V \leq c_1 = G^{-1}_{m-1,n-1}(\alpha_0/2) \quad \text{or} \quad V \geq c_2 = G^{-1}_{m-1,n-1}(1 - \alpha_0/2).$$

Then $\delta$ is a level $\alpha_0$ test and the p-value is

$$2 \times \min\{1 - G_{m-1,n-1}(v), G_{m-1,n-1}(v)\}.$$