Lecture 4: Hypotheses Tests and Confidence Regions

Måns Thulin

Department of Mathematics, Uppsala University
thulin@math.uu.se

Multivariate Methods • 6/4 2011

Outline

- Testing hypotheses in p dimensions
  - Hypotheses about $\boldsymbol{\mu}$ for the MVN
- Hotelling's T^2
- Confidence regions
  - T^2
  - Bonferroni's inequalities
- Large sample approximations

Reminder: testing hypotheses

- Hypothesis: null and alternative
- Test statistic
- Quantiles
- Significance level α
- p-value
- Power of test
- t-test, z-test, $\dots$

Testing hypotheses in p dimensions

- Several problems occur when we wish to test hypotheses for p dimensional random variables.
  - Dependencies between variables makes testing complicated
  - Usually we want to test a hypothesis for the p variables' joint distribution as well as hypotheses for each of the p marginal distributions
  - The number of parameters in the hypotheses can be large
  - A huge number of possible statistics exist
  - Choice between many univariate tests and one multivariate test
- Example: the multivariate normal distribution has $\frac{1}{2}p(p+3)$ parameters.
  - For $p = 5$, the MVN has 20 parameters
  - For $p = 10$, the MVN has 65 parameters
  - For $p = 100$, the MVN has 5150 parameters
  - It is difficult to reliably estimate or test hypotheses about all of these!

Testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu_0}$ – an example

Consider $\mathbf{X} = (X_1, X_2)$ belonging to a bivariate normal distribution $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. We wish to test $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu_0} = (182, 182)$.
For simplicity, we assume that X_1 and X_2 are independent and that both have a known variance $\sigma^2 = 100$.

Given 25 observations $\mathbf{x_1}, \ldots, \mathbf{x_{25}}$, with $\bar{x}_1 = 185.72$ and $\bar{x}_2 = 183.84$ we could test the hypotheses

$$H_0^{(1)} : \mu_1 = 182$$
$$H_0^{(2)} : \mu_2 = 182$$

with two z-tests with significance level a.
The test statistics would be

$$z_1 = \frac{\bar{X}_1 - 182}{10/\sqrt{25}} \sim N(0,1) \text{ under } H_0$$
$$z_2 = \frac{\bar{X}_2 - 182}{10/\sqrt{25}} \sim N(0,1) \text{ under } H_0$$

Testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu_0}$ – an example

H_0 is rejected if either $H_0^{(1)}$ or $H_0^{(2)}$ is rejected.
Then, since X_1 and X_2 are independent,

$$\alpha_{(simultaneous)} = P_{H_0}(H_0 \text{ is rejected}) =$$
$$P_{H_0}(H_0^{(1)} \text{ is rejected}) + P_{H_0}(H_0^{(2)} \text{ is rejected}) -$$
$$P_{H_0}(H_0^{(1)} \text{ is rejected}) \cdot P_{H_0}(H_0^{(2)} \text{ is rejected}) =$$
$$P(|z_1| > \lambda_{a/2}) + P(|z_2| > \lambda_{a/2}) - P(|z_1| > \lambda_{a/2}) \cdot P(|z_2| > \lambda_{a/2}) =$$
$$2a - a^2.$$

Thus $a = 1 - \sqrt{1-\alpha}$ would yield a simultaneous test with significance level α.
We note that if X_1 and X_2 were not independent, calculating the simultaneous significance level could be difficult.

Let $\alpha = 0.05$. Then $a = 1 - \sqrt{1-\alpha} \approx 0.0253$ and $\lambda_{a/2} \approx 2.24$.
With $\bar{x}_1 = 185.72$ and $\bar{x}_2 = 183.84$ we have that $|z_1| = 1.86 < 2.24$ and $|z_2| = 0.92 < 2.24$. Thus *the null hypothesis is not rejected*.

Testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu_0}$ – an example

On the other hand, if H_0 is true, then

$$z_3 = \frac{1}{\sqrt{2}}(z_1 + z_2) \sim N(0, 1)$$

and we could use this to perform a single z-test of the hypothesis. We have that $z_3 = 1.966$. With $\alpha = 0.05$, we get $\lambda_{\alpha/2} = 1.960$, so $|z_3| > \lambda_{\alpha/2}$, which means that we would *reject the null hypothesis!*

Furthermore, if H_0 is true, we also have that

$$z_4 = z_1^2 + z_2^2 \sim \chi_2^2.$$

We could use z_4 to test the hypothesis; we get that $z_4 = 4.306 < \chi_2^2(0.05)$, so *the null hypothesis would not be rejected*.

Which test should we trust? Are there other, better, tests?

Testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu_0}$ – univariate case

Consider a univariate sample $x_1, \ldots, x_n$ from $N(\mu, \sigma^2)$ where μ and σ^2 are unknown. Assume that we wish to test the hypothesis $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$.

In this setting, we would use the t-test. The use of the t-test is motivated by the likelihood ratio concept.

The test statistic is

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim t_{n-1} \quad \text{under } H_0.$$

H_0 is rejected at significance level α if $|t| > t_{n-1}(\alpha/2)$.

This is equivalent to studying

$$t^2 = \frac{(\bar{X} - \mu_0)^2}{s^2/n} = n(\bar{X} - \mu_0)(s^2)^{-1}(\bar{X} - \mu_0) \sim F_{1,n-1}.$$

Testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu_0}$ – T^2

It would therefore seem natural to study a multivariate generalization of

$$t^2 = \frac{(\bar{X} - \mu_0)^2}{s^2/n} = n(\bar{X} - \mu_0)(s^2)^{-1}(\bar{X} - \mu_0),$$

namely,

$$T^2 = n(\bar{\mathbf{X}} - \boldsymbol{\mu_0})'\mathbf{S}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu_0}).$$

This statistic is called *Hotelling's* T^2 after Harold Hotelling, who showed that, under H_0,

$$\frac{n-p}{(n-1)p} \cdot T^2 \sim F_{p,n-p}.$$

Testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu_0}$ – T^2

The T^2 test therefore rejects $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu_0}$ at level α if

$$T^2 > \frac{(n-1)p}{n-p} F_{p,n-p}(\alpha).$$

Similarly, the p-value of the test is obtained as

$$p = P(T^2 > x)$$

where x is the observed value of the statistic.

Testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu_0}$ – T^2

Some further remarks regarding Hotelling's T^2:

- T^2 is invariant under transformations of the kind $\mathbf{CX} + \mathbf{d}$ where $\mathbf{X}$ is the data matrix, $\mathbf{C}$ is a non-singular matrix and $\mathbf{d}$ is a vector.
- Hotelling's T^2 is the likelihood ratio test (see J&W Sec 5.3) and has some optimality properties.
- Under the alternative $H_1 : \mu = \mu_1$,

$$\frac{n-p}{(n-1)p} \cdot T^2 \sim F\Big(p, n-p, (\boldsymbol{\mu_1} - \boldsymbol{\mu_0})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu_1} - \boldsymbol{\mu_0})\Big),$$

  where $F(n_1, n_2, A)$ is a noncentral F-distribution with degrees of freedom n_1 and n_2 and noncentrality parameter A. The power of the T^2-test against H_1 can thus be easily obtained.

Confidence regions

A (univariate) confidence interval for the parameter θ is an interval that covers the true parameter value with probability $1 - \alpha$ (before sampling).

A confidence region for the p-dimensional parameter $\boldsymbol{\theta}$ is a region in p-dimensional space that covers the true parameter value with probability $1 - \alpha$ (before sampling).

Confidence regions: μ in the univariate case

We return to the univariate setting where we have a sample $x_1, \ldots, x_n$ from $N(\mu, \sigma^2)$ with μ and σ^2 unknown.
The null hypothesis $H_0 : \mu = \mu_0$ is *not rejected* at level α if

$$\left|\frac{\bar{x} - \mu_0}{s/\sqrt{n}}\right| \leq t_{n-1}(\alpha/2)$$

or, equivalently, if

$$\bar{x} - t_{n-1}(\alpha/2) \cdot \frac{s}{\sqrt{n}} \leq \mu_0 \leq \bar{x} + t_{n-1}(\alpha/2) \cdot \frac{s}{\sqrt{n}}.$$

Thus the confidence interval

$$\left(\bar{x} - t_{n-1}(\alpha/2) \cdot \frac{s}{\sqrt{n}}, \bar{x} + t_{n-1}(\alpha/2) \cdot \frac{s}{\sqrt{n}}\right)$$

contains all values of μ_0 that would not be rejected by the t-test at level α.

Confidence regions: Confidence ellipses

Analogously, the region

$$n(\bar{\mathbf{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu}) \leq \frac{p(n-1)}{n-p} F_{p,n-p}(\alpha)$$

contains the values of $\boldsymbol{\mu_0}$ that would not be rejected by Hotelling's T^2 at level α.

We have that

$$P\Big(n(\bar{\mathbf{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu}) \leq \frac{p(n-1)}{n-p} F_{p,n-p}(\alpha) \Big) = 1 - \alpha.$$

The region above is thus a confidence region for $\boldsymbol{\mu}$. It's an ellipsoid centered at $\bar{\mathbf{x}}$. The axes of the ellipsoid are given by the eigenvectors of $\mathbf{S}$.

See figure on blackboard!

Confidence regions: Simultaneous intervals

Often, we do not only wish to obtain a confidence region in p-space, but also confidence intervals for each μ_i.

More generally, we are interested in simultaneous confidence intervals for various linear combinations $\mathbf{a}'\boldsymbol{\mu}$ of the means.

We would like these intervals to have a *simultaneous confidence level* α, that is, we would like that

$$P(\text{all } p \text{ intervals cover the true parameter value}) = 1 - \alpha.$$

The ordinary one-variable-at-a-time confidence intervals seem hard to use here. For p independent variables

$$P(\text{all } p \text{ intervals cover the true parameter value}) = (1 - \alpha)^p$$

but for dependent variables this probability is difficult or impossible to calculate!

Confidence regions: Simultaneous intervals

Result 5.3. Let $\mathbf{X_1}, \ldots, \mathbf{X_n}$ be a random sample from $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is positive definite. Then

$$\left(\mathbf{a}'\bar{\mathbf{X}} - \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p}(\alpha) \frac{\mathbf{a}'\mathbf{S}\mathbf{a}}{n}}, \mathbf{a}'\bar{\mathbf{X}} + \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p}(\alpha) \frac{\mathbf{a}'\mathbf{S}\mathbf{a}}{n}}\right)$$

contains $\mathbf{a}'\boldsymbol{\mu}$ with probability $1-\alpha$ simultaneously for all $\mathbf{a}$.

The part under the square root comes from the distribution of the T^2 statistic.

Taking $\mathbf{a}' = (1, 0, \ldots, 0)$, $\mathbf{a}' = (0, 1, 0, \ldots, 0)$, $\ldots$, $\mathbf{a}' = (0, 0, \ldots, 0, 1)$ gives us simultaneous intervals for $\mu_1, \ldots, \mu_n$. Taking $\mathbf{a}' = (1, -1, 0 \ldots, 0)$ gives us an interval for $\mu_1 - \mu_2$, and so on.

Proof of Res 5.3: see blackboard!

Confidence regions: One-at-a-time intervals

From the previous slide, the simultaneous confidence interval for μ_i is

$$\left(\bar{x}_i \pm \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p}(\alpha)} \sqrt{\frac{s_{ii}}{n}}\right)$$

How does this compare to the ordinary one-at-a-time confidence interval

$$\left(\bar{x}_i \pm t_{n-1}(\alpha/2) \sqrt{\frac{s_{ii}}{n}}\right)?$$

To compare the intervals, we need only to compare $\sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p}(\alpha)}$ and $t_{n-1}(\alpha/2)$.

Confidence regions: One-at-a-time intervals

Let $a(p) = \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p}(0.05)}$.

n	$t_{n-1}(0.025)$	$a(2)$	$a(5)$	$a(10)$
10	2.262	3.167	6.742	–
20	2.093	2.739	4.287	7.522
100	1.984	2.498	3.470	4.617
∞	1.960	2.448	3.327	4.277

Larger p give wider intervals for fixed n. Larger n give smaller intervals for fixed p.

So what do the simultaneous intervals look like? *See figure on blackboard!*
The intervals cover to wide an area!

Confidence regions: Bonferroni intervals

Bonferroni's inequalities is a set of inequalities for probabilities of unions of events.
Let $C_1, \ldots, C_p$ be confidence intervals, with $P(C_i$ covers the true parameter value$) = 1 - \alpha_i$.
The Bonferroni inequality for confidence intervals is:

$$P(\text{all } C_i \text{ cover the true parameter values}) \geq 1 - (\alpha_1 + \ldots + \alpha_p)$$

Proof: see blackboard.

Typically, $\alpha_i = \alpha/p$ is choosen. Then $P($all C_i cover the true parameter value$) \geq 1 - \alpha$ and the Bonferroni simultaneous confidence interval for μ_i is

$$\Big(\bar{x}_i \pm t_{n-1}\Big(\frac{\alpha}{2p}\Big)\sqrt{\frac{s_{ii}}{n}}\Big).$$

Confidence regions: Bonferroni intervals

Let's study the ratio

$$\frac{\text{length of Bonferroni interval}}{\text{length of } T^2 \text{ interval}} = \frac{t_{n-1}\left(\frac{\alpha}{2p}\right)}{\sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p}(\alpha)}}$$

n	$p = 2$	$p = 4$	$p = 10$
15	0.88	0.69	0.29
25	0.90	0.75	0.48
50	0.91	0.78	0.58
∞	0.91	0.81	0.66

In general, the T^2 intervals are wider.

See figure on blackboard!

Bonferroni inequality for tests

Similarly, a Bonferroni inequality can be stated for tests.
We wish to perform m test with a simultaneous significance level α. Let $P_1, \ldots, P_m$ be the p-values for the m tests.
Then

$$P\Big(\bigcup_{i=1}^{m}(P_i \leq \alpha/m)\Big) \leq \alpha.$$

That is, the probability of rejecting at least one hypothesis when all hypotheses are true is no greater than α. Thus the simultaneous significance level is at most α.

The Bonferroni inequlity for tests is useful when we wish to test hypotheses about different variables simultaneously (for instance when testing for marginal normality).

(An extension of this idea is studied in homework 2.)

Large sample approximations

For large n, the methods we've discussed for normal data can often be used even if the data is non-normal.

For multivariate distributions with finite $\mathbf{\Sigma}$, the multivariate central limit theorem can be used together with *Cramérs lemma* (also known as *Slutsky's lemma*) to show that

$$T^2 \xrightarrow{d} \chi^2_p$$

so that

$$P\Big(T^2 \leq \chi^2_p(\alpha)\Big) = P\Big(n(\bar{\mathbf{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu}) \leq \chi^2_p(\alpha)\Big) \approx 1 - \alpha$$

when n is sufficiently large.

Large sample approximations

Thus, the large sample T^2 test rejects $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu_0}$ if

$$T^2 = n(\bar{\mathbf{X}} - \boldsymbol{\mu_0})'\mathbf{S}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu_0}) > \chi_p^2(\alpha).$$

It also follows that, simultaneously for all $\mathbf{a}$,

$$\left(\mathbf{a'}\bar{\mathbf{X}} \pm \sqrt{\chi_p^2(\alpha)}\sqrt{\frac{\mathbf{a'Sa}}{n}}\right)$$

contains $\mathbf{a'}\boldsymbol{\mu}$ with probability approximately $1 - \alpha$.
Finally, the Bonferroni simultaneous confidence intervals for the μ_i are obtained using the univariate central limit theorem:

$$\left(\bar{x}_i \pm \lambda\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{ii}}{n}}\right).$$

where $\lambda\left(\frac{\alpha}{2p}\right)$ are the quantiles from the standard normal distribution.

Summary

- Testing hypotheses in p dimensions
  - Hypotheses about $\boldsymbol{\mu}$ for the MVN
- Hotelling's T^2
  - Analogue to t-test
- Confidence regions
  - T^2
  - Bonferroni's inequalities
- Large sample approximations