**Testing means**

**Inference procedures for one mean vector**

Univariate case

The univariate test for a mean taught in an introductory statistics class involves the hypotheses:

H0:μ = μ0

Ha:μ ≠ μ0

for some constant μ0. The test statistic is



where  is the sample mean, s is the sample standard deviation, and N is the sample size. If a random sample comes from a normal distribution with mean μ0 (H0 is then true) and a variance σ2, T has a t-distribution with N – 1 degrees of freedom (T ~ tN-1). If the observed value of T is unusual in size (i.e., |T| > tN-1,1-α/2), we reject the null hypothesis.

Multivariate case

The multivariate extension of the univariate test involves testing

H0:**μ** = **μ**0

Ha:**μ** ≠ **μ**0

for some constant p×1 vector **μ**0. The statistic used to perform the test involves the Hotelling’s T2 statistic:



When p = 1,



which is the square of test statistic used for the univariate test.

To find the distribution of T2, let **x**1, …, **x**N be i.i.d. Np(**μ**,**Σ**) where **μ** and **Σ** are unknown. One can show that

 ~ Fp, N-p

if the null hypothesis is true. Using this result, hypothesis tests and confidence regions for **μ** can be constructed. Thus, we can reject the null hypothesis if

 > F1-α, p, N-p

where F1-α,p, N-p is the 1-α quantile of a F-distribution. Also, a (1-α)100% confidence region for **μ** is the set of **μ** such that

T2 ≤ F1-α,p, N-p

Example: Bivariate normal distribution (HotellingSim.R)

Suppose **x** ~  and 20 observations are simulated from a population characterized by this distribution. Below is the R code for the test of H0:**μ** = [15, 20]′ vs. Ha:**μ** ≠ [15, 20]′. Note that H0 would really be true here!

> library(mvtnorm)

> p <- 2

> mu <- c(15, 20)

> sigma <- matrix(data = c(1, 0.5, 0.5, 1.25), nrow = 2, ncol = 2, byrow = TRUE)

> cov2cor(sigma)

 [,1] [,2]

[1,] 1.0000000 0.4472136

[2,] 0.4472136 1.0000000

> N <- 20

> set.seed(7812)

> x <- rmvnorm(n = N, mean = mu, sigma = sigma)

> head(x)

 [,1] [,2]

[1,] 15.98996 20.40392

[2,] 15.49028 18.35249

[3,] 15.38903 21.06337

[4,] 16.03633 20.03042

[5,] 14.70052 18.29860

[6,] 14.40092 20.51145

> mu.hat <- colMeans(x)

> sigma.hat <- cov(x)

> R <- cor(x)

> mu.hat

[1] 15.28776 20.02926

> sigma.hat

 [,1] [,2]

[1,] 0.9288198 0.5473159

[2,] 0.5473159 1.1268952

> R

 [,1] [,2]

[1,] 1.0000000 0.5349714

[2,] 0.5349714 1.0000000

> #Hypothesis test: Ho:mu=[15,20], Ha:mu<>[15,20]

> mu.Ho <- c(15,20)

> T.sq <- N\*t(mu.hat-mu.Ho)%\*%solve(sigma.hat)%\*%(mu.hat-

 mu.Ho)

> test.stat <- (N-p)/(p\*(N-1))\*T.sq

> crit.val <- qf(0.95, p, N-p)

> p.value <- 1-pf((N-p)/(p\*(N-1))\*T.sq, p, N-p)

> round(data.frame(T.sq, test.stat, crit.val, p.value), 2)

 T.sq test.stat crit.val p.value

1 2.27 1.08 3.55 0.36

Because the p-value is large, the null hypothesis is not rejected. There is not sufficient evidence that the mean vector is different from [15, 20]′. Of course, this result is to be expected because we simulated the data with settings that made the null hypothesis true!

Below is a plot showing why the results of the test make sense (see program for code):



The estimated mean vector is relatively close to the hypothesized mean vector, so this is why the null hypothesis was not rejected.

Examine what happens when you run the code several times with different seednumbers. How often would you expect to reject H0? Examine the plot each time after you run the code and compare it to the hypothesis test result.

Below is an example where I essentially re-run the same code 20 times:

> set.seed(7812)

> #Save results here

> save.results <- matrix(data = NA, nrow = 20, ncol = 4)

> dev.new(width = 10)

> par(mfrow = c(4,5), mar = c(2,2,2,2)) #mar controls the

 margins in each plot

> for(i in 1:20) {

 x <- rmvnorm(n = N, mean = mu, sigma = sigma)

 mu.hat <- colMeans(x)

 sigma.hat <- cov(x)

 T.sq <- N\*t(mu.hat-mu.Ho)%\*%solve(sigma.hat)%\*%(mu.hat-

 mu.Ho)

 test.stat <- (N-p)/(p\*(N-1))\*T.sq

 crit.val <- qf(0.95, p, N-p)

 p.value <- 1-pf((N-p)/(p\*(N-1))\*T.sq, p, N-p)

 save.results[i,] <- c(T.sq, test.stat, crit.val, p.value)

 eval.fx <- dmvnorm(x = all.x, mean = mu, sigma = sigma)

 fx <- matrix(data = eval.fx, nrow = length(x1), ncol =

 length(x2), byrow = FALSE)

 contour(x = x1, y = x2, z = fx,

 xlab = expression(x[1]), ylab = expression(x[2]), xlim

 = c(10,20), ylim = c(15, 25), levels = c(0.001, 0.02))

 points(x = x[,1], y = x[,2], col = "red")

 points(x = mu.hat[1], mu.hat[2], pch = 3, col =

 "black", lwd = 2)

 points(x = mu[1], mu[2], pch = 4, col = "blue", lwd =

 2)

 }

> round(save.results,4)

 [,1] [,2] [,3] [,4]

 [1,] 2.2724 1.0764 3.5546 0.3618

 [2,] 3.2248 1.5275 3.5546 0.2439

 [3,] 4.3798 2.0746 3.5546 0.1546

 [4,] 0.0817 0.0387 3.5546 0.9621

 [5,] 0.5704 0.2702 3.5546 0.7663

 [6,] 3.2307 1.5303 3.5546 0.2433

 [7,] 1.8530 0.8777 3.5546 0.4328

 [8,] 0.2077 0.0984 3.5546 0.9068

 [9,] 2.6390 1.2500 3.5546 0.3102

[10,] 1.2316 0.5834 3.5546 0.5682

[11,] 1.4495 0.6866 3.5546 0.5160

[12,] 0.2764 0.1309 3.5546 0.8781

[13,] 10.6583 5.0487 3.5546 0.0182

[14,] 3.6146 1.7122 3.5546 0.2086

[15,] 0.2824 0.1338 3.5546 0.8757

[16,] 1.5912 0.7537 3.5546 0.4849

[17,] 0.7932 0.3757 3.5546 0.6921

[18,] 0.5071 0.2402 3.5546 0.7890

[19,] 4.1620 1.9715 3.5546 0.1682

[20,] 2.8867 1.3674 3.5546 0.2800

> mean(save.results[,4] < 0.05)

[1] 0.05



When using α = 0.05, 1 out of 20 of the samples resulted in a rejection of the null hypothesis. Examine the plot where this rejection occurred.

Extensions for one mean vector

The testing procedure can be generalized to test linear combinations of the individual means in **μ**. Let **H** be a q×p matrix that forms the linear combinations of the means.

For example, suppose p = 3. Then linear combinations of interest may be μ1 – μ2 = 0 and μ1 – μ3 = 0. In this case, the linear combinations can be formed with the following matrix multiplications:



where



We can now specify a specific vector of hypothesized values for **Hμ**, say **h**.

For example, suppose **h** = [0, 0]′. Notice this implies μ1 – μ2 = 0 and μ1 – μ3 = 0 along with μ2 – μ3 = 0. Thus, μ1 = μ2 = μ3.

A test of

H0:**Hμ** = **h**

Ha:**Hμ** ≠ **h**

can be tested using:



Under the null hypothesis and multivariate normality for **x**, the test statistic has an  distribution.

Another commonly used form for **H** is to specify a linear trend among the means. For example, a test for linear trend involves:

H0: μ1 – μ2 = μ2 – μ3,

μ2 – μ3 = μ3 – μ4,

…,

μp-2 – μp-1 = μp-1 – μp

Ha: At least one ≠

Why is this linear trend?

Equivalently,

H0: μ1 – 2μ2 + μ3 = 0,

μ2 – 2μ3 + μ4 = 0,

…,

μp-2 – 2μp-1 + μp = 0

Ha: At least one ≠ to 0

This leads to an **H** matrix of



where  and .

Tests can also be constructed for quadratic, cubic, or other trends using the correct contrasts of the μ’s.

**Inference procedures for two mean vectors**

Independent samples with **Σ**1 = **Σ**2

Let **x**i1, …,  is i.i.d. Np(**μ**i,**Σ**i) for i = 1, 2, where **μ**i are **Σ**iunknown. We cantest

H0:**μ**1 = **μ**2

Ha:**μ**1 ≠ **μ**2

using



where



(assumes **Σ**1 = **Σ**2). Under the null hypothesis, multivariate normality, and **Σ**1 = **Σ**2, the test statistic has a  distribution. Large values of the test statistic result in rejection of the null hypothesis.

Linear combinations of the individual means can also be found to construct tests of interest. The test statistic becomes



Under the null hypothesis, multivariate normality, and **Σ**1 = **Σ**2, the test statistic has a  distribution.

Independent samples with **Σ**1 ≠ **Σ**2

To test H0:**μ**1 = **μ**2 vs. Ha:**μ**1 ≠ **μ**2, use the statistic



For a large sample, this statistic can be approximated by a χ2 random variable with p degrees of freedom.

#### Paired samples

Similar to how this problem is approached for the univariate setting, one can convert the two sample multivariate problem into a one sample problem!

Let **d**r = **x**1r – **x**2r for r = 1, …, N. Then the **d**r’s can be considered a sample from a population with mean **μ** = **μ**1 – **μ**2. To test H0:**μ** = 0 vs. Ha:**μ** ≠ 0, use the same T2 statistic as given at the beginning of this set of notes.

**Multivariate Analysis of Variance (MANOVA)**

MANOVA is the multivariate generalization of univariate analysis of variance

Univariate ANOVA

Test the following hypotheses:

H0:μ1 = μ2 = = μm
Ha:Not all equal

where μi is the mean for population i. Population “i” can be thought of as treatment “i”.

Consider a completely randomized design (CRD) with only 1 factor. The means model is

xir = μi + εir

where xir is the response of the rth experimental unit for treatment i, μi is the population mean of treatment i, and εir is the error term with εir ~ i.i.d. N(0,σ2).

The ANOVA Table:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Source**  | **d.f.** | **SS** | **MS** | **F** |
| Treatments | m-1 | SST | MST | F |
| Error | N-m | SSE | MSE |  |
| Total | N-1 | SS(total) |  |  |

Notes:

* “Source” means the source of variation
* “Error” means the within treatment variation
* “Treatments” means the between treatment variation
* SST = Sum of squares for treatments
* SSE = Sum of squared errors
* SS(total) = total sum of squares = SST + SSE
* MST = Mean sum of squares for treatments
* MSE = Mean sum of squared errors
* F is the test statistic for H0:μ1 = μ2 = … = μm vs. Ha:Not all equal

Formulas:

* Average variation between the treatments:

MST = SST/(m-1) = 

where , , and 

* Average variation within the treatments

MSE = SSE/(n-m) = 

* Total variation

SS(total) = 

* F = MST/MSE which has a F-distribution with m – 1 and N – m degrees of freedom in the numerator and denominator, respectively, provided the null hypothesis is true.

### MANOVA

Test the following hypotheses:

H0:**μ**1 = **μ**2 = = **μ**m

Ha:Not all equal

where **μ**i = (μi1,…,μip)′.

Notation:

* Let μij = mean response for variable j in treatment i for i = 1, …, m and j = 1, …, p.
* Let xirj be the observed value of the jth response variable on the rth experimental unit from the ith treatment. These values can be put into a vector for the rth experimental unit in the ith population: **x**ir = (xir1,…,xirp)′.
* Let r = 1, …, Ni.

The multivariate means model is

**x**ir = **μ**i + **ε**ir

where **ε**ir = (εir1,…,εirp)′ ~ independent Np(**0**,**Σ**)

Note that dependency is allowed for within an experimental unit. If the responses for the rth experimental unit were independent (**Σ** = σ2**I**), then ANOVA methods could be used on each of the p variables.

The error sums of squares and cross products matrix **E** plays the role of SSE in ANOVA. This matrix is often called the “within sum of squares” matrix. The matrix is



where  and Ni = # of experimental units assigned to treatment i.

The “between sums of squares” matrix **H** plays the role of SST in ANOVA. The matrix is:



where .

The “total sums squares” matrix is **H** + **E**:

 

The MANOVA table is

|  |  |  |  |
| --- | --- | --- | --- |
| **Source**  | **d.f.** | **SS** | **Λ** |
| Treatments | m-1 | H | |**E**|/|**H**+**E**| |
| Error | N-m | E |  |
| Total | N-1 | **H**+**E** |  |

The Λ statistic tests H0: **μ**1 = **μ**2 = = **μ**m vs. Ha: Not all equal. This can be seen to be similar to the F test in ANOVA by noting the following:

Because F = MST/MSE and the null hypothesis is rejected when F is large, this is similar to saying reject the null hypothesis when SST/SSE is large. Equivalently, reject the null hypothesis when 1+SST/SSE = (SSE+SST)/SSE is large. Taking the reciprocal produces a test of SSE/(SSE+SST) and reject the null hypothesis when this is small.

Note that Λ is called Wilk’s lambda. It is actually a likelihood ratio test statistic where the main part of the statistic depends upon |**E**|/|**H**+**E**| (so this is why it is expressed this way instead of –2log(lik. ratio)). In the end, a somewhat complicated F-distribution approximation is used; see Johnson and Wichern’s textbook for details if you are interested.

Questions:

* What if H0 is rejected? Examine which means are different by examining the variables one at a time using ANOVA methods.
* What if H0 is not rejected? There is not a significant difference between the mean vectors. Johnson’s textbook recommends a conservative approach to still look for differences between means of variables. He says to use the Bonferroni procedure when looking for differences using ANOVA methods (i.e., use α/p as the level of significance).

Other testing procedures

Below are other testing procedures for H0: **μ**1 = **μ**2 = = **μ**m vs. Ha: Not all equal

* Roy’s test: Based on the largest λi of **HE**-1
* Lawley and Hotelling’s test: T = tr(**HE**-1)
* Pillai’s test: V = tr[**H**(**H**+**E**)-1]

Notes:

* In ANOVA, the uniformly most powerful unbiased (UMPU; see a mathematical statistics textbook for details) test for H0:μ1 = μ2 = = μm vs. Ha:Not all equal is the F test. Unfortunately, no one testing procedure is UMPU in MANOVA.
* Johnson recommends using Wilk’s likelihood ratio test, so I will only focus on this one.
* When p = 1, all these tests and Wilk’s test are equivalent.

Example: CPT (CPT.R)

A pharmaceutical company is conducting safety clinical trials on a new drug used to treat schizophrenia patients. Healthy male volunteers were given 0, 3, 9, 18, 36, or 72mg of the drug. Before the drug was administered (time = 0) and at 1, 2, 3, 4 hours after, a psychometric test called the Continuous Performance Test (CPT) was administered. The CPT involves the following:

* A subject sits in front a computer screen.
* Randomly generated numbers from 0 to 9 appear on a computer screen.
* Each image is slightly blurred.
* One number appears every second for 480 seconds.
* Subjects are required to press a button whenever the number 0 appears.
* The response variable is the number of hits (i.e., the number of correct responses).

Does the number of hits change after the drug is administered? If it does, this could mean:

* drug causes drowsiness
* drug causes blurred vision
* Some other effect

In data sets like this, one usually will see it in the following format:

|  |  |  |
| --- | --- | --- |
|  |  | **Hits at time** |
| **Patient** | **Dose** | **0** | **1** | **2** |  **3**  | **4** |
| 101 | 0 | 98 | 101 | 100 | 98 | 101 |
|  |  |  |  |  |  |  |
| 504 | 72 | 97 | 96 | 90 | 86 | 89 |

Because the data is owned by the company, I cannot use the actual data in the clinical trial. Instead, I simulated the data with the following R code:

> mu.dose0 <- c(100, 100, 100, 100, 100)

> mu.dose3 <- c(100, 100, 98, 96, 96)

> mu.dose9 <- c(100, 98, 96, 95, 94)

> mu.dose18 <- c(100, 97, 95, 94, 93)

> mu.dose36 <- c(100, 95, 92, 91, 90)

> mu.dose72 <- c(100, 94, 90, 89, 89)

> #Set the covariance matrix - same for each group assume

> rho <- 0.5

> var.common <- 9

> sigma <- var.common\*matrix(data =

 c( 1, rho, rho^2, rho^3, rho^4,

 rho, 1, rho, rho^2, rho^3,

 rho^2, rho, 1, rho, rho^2,

 rho^3, rho^2, rho, 1, rho,

 rho^4, rho^3, rho^2, rho, 1),

 nrow = 5, ncol = 5)> N <- 10

> p <- 5

> library(mvtnorm)

> set.seed(1710)

> dose0 <- round(rmvnorm(n = N, mean = mu.dose0, sigma =

 sigma),0)

> dose3 <- round(rmvnorm(n = N, mean = mu.dose3, sigma =

 sigma),0)

> dose9 <- round(rmvnorm(n = N, mean = mu.dose9, sigma =

 sigma),0)

> dose18 <- round(rmvnorm(n = N, mean = mu.dose18, sigma =

 sigma),0)

> dose36 <- round(rmvnorm(n = N, mean = mu.dose36, sigma =

 sigma),0)

> dose72 <- round(rmvnorm(n = N, mean = mu.dose72, sigma =

 sigma),0)

> temp1 <- rbind(dose0, dose3, dose9, dose18, dose36, dose72)

> patient.numb <- 1:60

> dose.levels <- c(0,3,9,18,36,72)

> dose <- rep(x = dose.levels, times = 1, each = 10)

> cpt <- data.frame(patient = patient.numb, dose = dose,

 time0 = temp1[,1], time1 = temp1[,2],

 time2 = temp1[,3], time3 = temp1[,4],

 time4 = temp1[,5])

> head(cpt)

 patient dose time0 time1 time2 time3 time4

1 1 0 100 101 103 104 104

2 2 0 100 99 98 97 101

3 3 0 104 103 105 105 104

4 4 0 101 100 103 99 100

5 5 0 100 99 98 101 99

6 6 0 96 99 99 101 102

The purpose here is to determine if there are differences in the means hits for the treatment groups:

Ho:**μ**0 = **μ**3 = **μ**9 = **μ**18 = **μ**36 = **μ**72
Ha:Not all equal

where **μ**i = (μi0, μi1, μi2, μi3, μi4)′ and μij = mean hits at time j for dose group i.

Below is the R code and output.

> save <- manova(formula = cbind(time0, time1, time2, time3, time4) ~ factor(dose), data = cpt)

> summary(save, test = "Wilks")

 Df Wilks approx F num Df den Df Pr(>F)

factor(dose) 5 0.13998 5.2258 25 187.24 9.423e-12 \*\*\*

Residuals 54

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> save.means <- aggregate(formula = cbind(time0, time1,

 time2, time3, time4) ~ dose, data = cpt, FUN = mean)

> save.means

 dose time0 time1 time2 time3 time4

1 0 100.1 98.7 98.8 99.7 100.8

2 3 99.3 99.9 96.9 96.3 96.6

3 9 99.8 98.6 97.1 94.3 92.8

4 18 100.2 97.3 94.5 93.7 93.2

5 36 100.1 95.4 91.2 91.2 90.1

6 72 98.8 93.2 88.6 86.0 87.6

> plot(x = 0:4, save.means[1,-1], main = "Means by

 treatment over time", ylim = c(min(save.means[,-1]),

 max(save.means[,-1])), panel.first = grid(), type =

 "o", col = "black", xlab = "Time", ylab = "Mean hits")

> lines(x = 0:4, save.means[2,-1], type = "o", col = "red")

> lines(x = 0:4, save.means[3,-1], type = "o", col =

 "blue")

> lines(x = 0:4, save.means[4,-1], type = "o", col =

 "green")

> lines(x = 0:4, save.means[5,-1], type = "o", col =

 "purple")

> lines(x = 0:4, save.means[6,-1], type = "o", col =

 "orange")

> legend(x = 0, y = 94, legend =

 levels(as.factor(cpt$dose)), col = c("black", "red",

 "blue", "green", "purple", "orange"), lty = 1, bty = "n")



> mod.fit0 <- aov(formula = time0 ~ factor(dose), data = cpt)

> summary(mod.fit0)

 Df Sum Sq Mean Sq F value Pr(>F)

factor(dose) 5 15.5 3.097 0.39 0.853

Residuals 54 428.7 7.939

> mod.fit1 <- aov(formula = time1 ~ factor(dose), data = cpt)

> summary(mod.fit1)

 Df Sum Sq Mean Sq F value Pr(>F)

factor(dose) 5 307.5 61.50 8.797 3.74e-06 \*\*\*

Residuals 54 377.5 6.99

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> mod.fit2 <- aov(formula = time2 ~ factor(dose), data = cpt)

> summary(mod.fit2)

 Df Sum Sq Mean Sq F value Pr(>F)

factor(dose) 5 767.1 153.42 13.45 1.59e-08 \*\*\*

Residuals 54 615.9 11.41

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> mod.fit3 <- aov(formula = time3 ~ factor(dose), data = cpt)

> summary(mod.fit3)

 Df Sum Sq Mean Sq F value Pr(>F)

factor(dose) 5 1085 216.99 22.71 3.42e-12 \*\*\*

Residuals 54 516 9.56

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> mod.fit4 <- aov(formula = time4 ~ factor(dose), data = cpt)

> summary(mod.fit4)

 Df Sum Sq Mean Sq F value Pr(>F)

factor(dose) 5 1098.5 219.70 28.08 6.76e-14 \*\*\*

Residuals 54 422.5 7.82

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Notes:

* The test for the equality of the mean vectors has a very small p-value. Thus, there is sufficient evidence to indicate a difference among the mean hits for the dosage levels.
* Because the null hypothesis is rejected, it is of interest to determine what caused the rejection (i.e., which means are different). ANOVA methods can be used for each time level to examine if there is a difference between means. For this data set, time 0 does not have a significant difference between mean hits. The remaining times do have significant differences.
* Remember that the means used to generate the data were:

mu.dose0 <- c(100, 100, 100, 100, 100)

mu.dose3 <- c(100, 100, 98, 96, 96)

mu.dose9 <- c(100, 98, 96, 95, 94)

mu.dose18 <- c(100, 97, 95, 94, 93)

mu.dose36 <- c(100, 95, 92, 91, 90)

mu.dose72 <- c(100, 94, 90, 89, 89)

* If the MANOVA null hypothesis of equality of mean vectors was NOT rejected, many people would suggest to stop the analysis there. Johnson suggests to go ahead and look at the individual means using a Bonferroni adjustment to the level of significance. If α = 0.05, then to examine for differences between the individual means using ANOVA, a level of significance of 0.05/5 = 0.01 could be used.

The above example is for a one-way fixed effects MANOVA model. These type of models can be extended to other situations, such as for a two-way fixed effects MANOVA model.