# TUTORIAL 10
# STA437 WINTER 2015

AL NOSEDAL

Contents



## 1. Canonical Correlation

In this tutorial, we discuss basic tests of significance associated with canonical correlations and relationships of canonical correlation analysis to other multivariate techniques.

### 1.1. Tests of NO relationship between the $y$'s and the $x$'s.

When $\mathbf{\Sigma}_{xy} = \mathbf{0}$, $\mathbf{a}'\mathbf{X}$ and $\mathbf{b}'\mathbf{Y}$ have covariance $\mathbf{a}'\mathbf{\Sigma}_{xy}\mathbf{b} = 0$ for all vectors $\mathbf{a}$ and $\mathbf{b}$. Consequently, all the canonical correlations must be zero, and there is no point in pursuing a canonical correlation analysis. In Chapter 7, we considered the hypothesis of independence, $H_0 : \mathbf{\Sigma}_{yx} = \mathbf{0}$. If $\mathbf{\Sigma}_{yx} = \mathbf{0}$, the covariance of every $y_i$ with every $x_j$ is zero, and all corresponding correlations are likewise zero. Hence, under $H_0$ there is **no** linear relationship between the $y$'s and the $x$'s, and $H_0$ is equivalent to the statement that **all** canonical correlations $r_1,\ r_2, ..., r_s$ are nonsignificant. Thus, the significance of $r_1,\ r_2, ..., r_s$ can be tested by

$$\Lambda_1 = \frac{|\mathbf{S}|}{|\mathbf{S}_{yy}||\mathbf{S}_{xx}|} = \frac{|\mathbf{R}|}{|\mathbf{R}_{yy}||\mathbf{R}_{xx}|},$$

which is distributed as $\Lambda_{p,q,n-1-q}$. We reject $H_0$ if $\Lambda_1 \leq \Lambda_\alpha$. Critical values $\Lambda_\alpha$ are available in Table A.9 using $\nu_H = q$ and $\nu_E = n - 1 - q$. The statistic $\Lambda_1$ is also distributed as $\Lambda_{q,p,n-1-p}$. $\Lambda_1$ is expressible in terms of the squared canonical correlations:

$$\Lambda_1 = \Pi_{i=1}^{s}(1 - r_i^2).$$

If the parameters exceed the range of critical values for Wilks' $\Lambda$ in Table A.9, we can use the $\chi^2$-approximation,

$$\chi^2 = -[n - \frac{1}{2}(p + q + 3)]ln(\Lambda_1),$$

which is approximately distributed as $\chi^2$ with $pq$ degrees of freedom. We reject $H_0$ if $\chi^2 \geq \chi^2_\alpha$.

1.2. **Example 1.** In an investigation of the relation of the Wechsler Adult Intelligence Scale to age. Researchers obtained this matrix of correlations among the digit span and vocabulary subsets, chronological age, and years of formal education:

$$\mathbf{R} = \begin{pmatrix} 1 & 0.45 & -0.19 & 0.43 \\ 0.45 & 1 & -0.02 & 0.62 \\ -0.19 & -0.02 & 1 & -0.29 \\ 0.43 & 0.62 & -0.29 & 1 \end{pmatrix},$$

The sample consisted of $n = 933$ men and women aged 25 to 64. Test the significance of the canonical correlations. Use $\alpha = 0.05$.

**Solution**

As noted at the beginning of this tutorial, the following tests are equivalent:
1. Test of $H_0 : \mathbf{\Sigma_{yx}} = \mathbf{0}$, independence of two sets of variables.
2. Test of significance of the canonical correlations.

First, let us use Wilks' lambda.
$p = q = 2$, $\nu_H = 2$, and $\nu_E = 933 - 1 - 2 = 930$

$|\mathbf{R}| = 0.4015025$

$|\mathbf{R}_{xx}| = 0.7975$

$|\mathbf{R}_{yy}| = 0.9159$

$$\Lambda = \frac{|\mathbf{R}|}{|\mathbf{R}_{yy}||\mathbf{R}_{xx}|} = \frac{0.4015025}{(0.7975)(0.9159)} = 0.5497$$

$$\Lambda_{0.05,2,2,930} \approx 0.9955$$

Since $\Lambda = 0.5497 < \Lambda_{0.05,2,2,930} \approx 0.9955$, we reject the hypothesis of independence. We must conclude that the subtests are dependent upon age and education.

Which is equivalent to saying that at least $r_1^2$ appears to be nonzero.

Now, let us use a $\chi^2$-approximation,

$$\chi^2 = -[n - \frac{1}{2}(p+q+3)]ln(\Lambda_1)$$

$$\chi^2 = -[933 - \frac{1}{2}(2+2+3)]ln(0.5497) = -[933 - 3.5](-0.5984)$$

$$\chi^2 = -[929.5](-0.5984) = 556.2128$$

The critical value is $\chi^2_{0.05,(2)(2)} = \chi^2_{0.05,4} = 9.488$. Since $556.2128 > 9.488$, we reject $H_0$. We must conclude that the subtests are dependent upon age and education. Which is the same as saying that at least $r_1^2$ appears to be nonzero.

1.3. **Test of Significance of Succeeding Canonical Correlations.** If the test $\Lambda_1 = \Pi_{i=1}^{s}(1-r_i^2)$ based on all $s$ canonical correlations rejects $H_0$, we are not sure if the canonical correlations beyond the first are significant. To test the significance of $r_2, r_3, ..., r_s$, we delete $r_1^2$ from $\Lambda_1$ to obtain

$$\Lambda_2 = \Pi_{i=2}^{s}(1 - r_i^2)$$

If this test rejects the hypothesis, we conclude that at least $r_2$ is significantly different from zero. We can continue in this manner, testing each $r_i$ in turn, until a test fails to reject the hypothesis. At the $k$th step, the test statistic is

$$\Lambda_k = \Pi_{i=k}^{s}(1 - r_i^2),$$

which is distributed as $\Lambda_{p-k+1,q-k+1,n-k-q}$ and tests the significance of $r_k, r_{k+1}, ..., r_s$. The $\chi^2$-approximation is given by

$$\chi^2 = -[n - \frac{1}{2}(p+q+3)]ln(\Lambda_k),$$

which is approximately distributed as $\chi^2$ with $(p-k+1)(q-k+1)$ degrees of freedom.

1.4. **Exercise.** Continue your analysis of the canonical correlations in Example 1.

1.5. **Relationships of Canonical Correlation Analysis to other Multivariate Techniques.**

| Subject | $X_1$ | $X_2$ | $Y_1$ | $Y_2$ | $Z_1$ |
|---|---|---|---|---|---|
| 1 | 4 | 2 | 11 | 25 | 1 |
| 2 | 1 | 3 | 14 | 25 | 1 |
| 3 | 3 | 5 | 12 | 26 | 1 |
| 4 | 2 | 2 | 11 | 30 | 1 |
| 5 | 2 | 2 | 13 | 25 | -1 |
| 6 | 4 | 5 | 12 | 30 | -1 |
| 7 | 5 | 4 | 12 | 27 | -1 |
| 8 | 5 | 5 | 15 | 28 | -1 |

a) Conduct a canonical analysis of the relationship between variables $Z_1$, on the one hand, and variables $X_1$, $X_2$, $Y_1$, and $Y_2$. Report both sets of canonical coefficients and the canonical correlation between the two sets of variables.

b) Treating subjects 1-4 as members of one group and subjects 5-8 as members of a second group, conduct a one-way MANOVA on the differences among the groups in mean response vectors.

c) Using vector **b** from part a), calculate for each subject the single score, $V_i = b_1 X_{1i} + b_2 X_{2i} + b_3 Y_{1i} + b_4 Y_{2i}$. Now conduct a t-test on the difference in **V** between the two groups.

d) Using Hotelling's $T^2$, conduct a test of the overall null hypothesis of no difference between the two groups in overall mean response vectors.

e) Compare and contrast the results obtained in parts a)-d).

**Solution**

```
##############################################
###  a) Canonical Correlation
##############################################


z1<-c(1,1,1,1,-1,-1,-1,-1)

x1<-c(4,1,3,2,2,4,5,5)

x2<-c(2,3,5,2,2,5,4,5)

y1<-c(11,14,12,11,13,12,12,15)

y2<-c(25,25,26,30,25,30,27,28)
```

```
data<-cbind(z1,x1,x2,y1,y2)

data


## Using covariance matrix

S<-cov(data)


## product 1

S11<-S[1,1]

S22<-S[2:5,2:5]

S12<-S[1,2:5]

S21<-S[2:5,1]

prod1<-solve(S11)%*%S12%*%solve(S22)%*%S21

prod1

## Finding Eigenvalues and Eigenvectors

e.val.vec.1<-eigen(prod1)

## Canonical correlation

r1<-sqrt(e.val.vec.1$val)

r1

## a1 = vector of coefficients that defines first canonical variate

a1<-e.val.vec.1$vec

a1

## product 2
```

```
prod2<-solve(S22)%*%S21%*%solve(S11)%*%S12

prod2

## Finding Eigenvalues and Eigenvectors

e.val.vec.2<-eigen(prod2)

## Canonical correlation

r2<-sqrt(e.val.vec.2$val[1])

r2

## b1 = vector of coefficients that defines "other" first canonical variate

b1<-e.val.vec.2$vec[ ,1]

b1


#########################
### b) One-Way MANOVA
#########################

groups<-factor(data[ ,1])
groups
Y<-data[ , 2:5]
Y

fit<-manova(Y~groups)

sum.wilks<-summary(fit,test="Wilks")
sum.wilks
## B
B<-sum.wilks$SS[1]

B

B<-matrix(unlist(B),4,4)

## W
```

```
W<-sum.wilks$SS[2]

W

W<-matrix(unlist(W),4,4)


inv.W.B<-solve(W)%*%B

eigen(inv.W.B)

## Is there something "interesting" about
## the first eigenvector?

lambda.1<-eigen(inv.W.B)$values[1]

## Remember Tutorial 4?
## The test statistic for Roy's test is:
## theta = lambda.1/(1+lambda.1)

theta<-lambda.1/(1+lambda.1)

theta



#######################################
## c) t-Test on canonical variate
#######################################

b1<-matrix(b1,ncol=1)

b1

V1<-data[ ,-1]%*%b1

V1

## t-test

T.test<-t.test(V1[1:4],V1[5:8],alternative="two.sided",var.equal=TRUE)

T.test
```

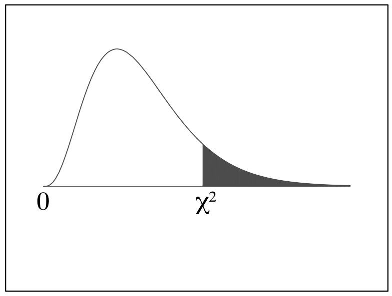
```
## Squaring test statistic

(T.test$statistic)^2


##########################################################
## d) Hotelling's T^2 on difference between two groups
##########################################################

## breaking down data

group1<-data[1:4 ,-1]

group2<-data[5:8,-1]

## sample sizes

n1<-dim(group1)[1]

n2<-dim(group2)[1]

## mean vectors

y.bar.1<-apply(group1,2,FUN=mean)

y.bar.2<-apply(group2,2,FUN=mean)

## covariance matrices

S.1<-cov(group1)

S.2<-cov(group2)

Sp<-(n1+n2-2)^(-1)*((n1-1)*S.1 + (n2-1)*S.2)

## Hotelling?s T^2

T.2<-(n1*n2/(n1+n2))*t(y.bar.1-y.bar.2)%*%solve(Sp)%*%(y.bar.1-y.bar.2)

T.2
```

```
## Is there something "interesting"
## about T.2?
```

e) **Comparisons**

- The canonical coefficients for variables $X_1$, $X_2$, $Y_1$, and $Y_2$ are identical to the coefficients of the MANOVA-derived discriminant function (”biggest” eigenvalue of $\mathbf{W^{-1}B}$). They are also identical with the $T^2$-derived discriminant function, though you were not required to calculate this last function.
- When these coefficients are used to obtain a single combined score for each subject, and a standard univariate t-test is conducted on the difference between the two groups on the resulting new variable, the square of this is the same as the value of Hotelling’s $T^2$ statistic.
- We saw that $r_1^2 = \frac{\lambda_1}{1+\lambda_1}$.

# Chi-Square Distribution Table




The shaded area is equal to $\alpha$ for $\chi^2 = \chi^2_\alpha$.

| *df* | $\chi^2_{.995}$ | $\chi^2_{.990}$ | $\chi^2_{.975}$ | $\chi^2_{.950}$ | $\chi^2_{.900}$ | $\chi^2_{.100}$ | $\chi^2_{.050}$ | $\chi^2_{.025}$ | $\chi^2_{.010}$ | $\chi^2_{.005}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.000 | 0.000 | 0.001 | 0.004 | 0.016 | 2.706 | 3.841 | 5.024 | 6.635 | 7.879 |
| 2 | 0.010 | 0.020 | 0.051 | 0.103 | 0.211 | 4.605 | 5.991 | 7.378 | 9.210 | 10.597 |
| 3 | 0.072 | 0.115 | 0.216 | 0.352 | 0.584 | 6.251 | 7.815 | 9.348 | 11.345 | 12.838 |
| 4 | 0.207 | 0.297 | 0.484 | 0.711 | 1.064 | 7.779 | 9.488 | 11.143 | 13.277 | 14.860 |
| 5 | 0.412 | 0.554 | 0.831 | 1.145 | 1.610 | 9.236 | 11.070 | 12.833 | 15.086 | 16.750 |
| 6 | 0.676 | 0.872 | 1.237 | 1.635 | 2.204 | 10.645 | 12.592 | 14.449 | 16.812 | 18.548 |
| 7 | 0.989 | 1.239 | 1.690 | 2.167 | 2.833 | 12.017 | 14.067 | 16.013 | 18.475 | 20.278 |
| 8 | 1.344 | 1.646 | 2.180 | 2.733 | 3.490 | 13.362 | 15.507 | 17.535 | 20.090 | 21.955 |
| 9 | 1.735 | 2.088 | 2.700 | 3.325 | 4.168 | 14.684 | 16.919 | 19.023 | 21.666 | 23.589 |
| 10 | 2.156 | 2.558 | 3.247 | 3.940 | 4.865 | 15.987 | 18.307 | 20.483 | 23.209 | 25.188 |
| 11 | 2.603 | 3.053 | 3.816 | 4.575 | 5.578 | 17.275 | 19.675 | 21.920 | 24.725 | 26.757 |
| 12 | 3.074 | 3.571 | 4.404 | 5.226 | 6.304 | 18.549 | 21.026 | 23.337 | 26.217 | 28.300 |
| 13 | 3.565 | 4.107 | 5.009 | 5.892 | 7.042 | 19.812 | 22.362 | 24.736 | 27.688 | 29.819 |
| 14 | 4.075 | 4.660 | 5.629 | 6.571 | 7.790 | 21.064 | 23.685 | 26.119 | 29.141 | 31.319 |
| 15 | 4.601 | 5.229 | 6.262 | 7.261 | 8.547 | 22.307 | 24.996 | 27.488 | 30.578 | 32.801 |
| 16 | 5.142 | 5.812 | 6.908 | 7.962 | 9.312 | 23.542 | 26.296 | 28.845 | 32.000 | 34.267 |
| 17 | 5.697 | 6.408 | 7.564 | 8.672 | 10.085 | 24.769 | 27.587 | 30.191 | 33.409 | 35.718 |
| 18 | 6.265 | 7.015 | 8.231 | 9.390 | 10.865 | 25.989 | 28.869 | 31.526 | 34.805 | 37.156 |
| 19 | 6.844 | 7.633 | 8.907 | 10.117 | 11.651 | 27.204 | 30.144 | 32.852 | 36.191 | 38.582 |
| 20 | 7.434 | 8.260 | 9.591 | 10.851 | 12.443 | 28.412 | 31.410 | 34.170 | 37.566 | 39.997 |
| 21 | 8.034 | 8.897 | 10.283 | 11.591 | 13.240 | 29.615 | 32.671 | 35.479 | 38.932 | 41.401 |
| 22 | 8.643 | 9.542 | 10.982 | 12.338 | 14.041 | 30.813 | 33.924 | 36.781 | 40.289 | 42.796 |
| 23 | 9.260 | 10.196 | 11.689 | 13.091 | 14.848 | 32.007 | 35.172 | 38.076 | 41.638 | 44.181 |
| 24 | 9.886 | 10.856 | 12.401 | 13.848 | 15.659 | 33.196 | 36.415 | 39.364 | 42.980 | 45.559 |
| 25 | 10.520 | 11.524 | 13.120 | 14.611 | 16.473 | 34.382 | 37.652 | 40.646 | 44.314 | 46.928 |
| 26 | 11.160 | 12.198 | 13.844 | 15.379 | 17.292 | 35.563 | 38.885 | 41.923 | 45.642 | 48.290 |
| 27 | 11.808 | 12.879 | 14.573 | 16.151 | 18.114 | 36.741 | 40.113 | 43.195 | 46.963 | 49.645 |
| 28 | 12.461 | 13.565 | 15.308 | 16.928 | 18.939 | 37.916 | 41.337 | 44.461 | 48.278 | 50.993 |
| 29 | 13.121 | 14.256 | 16.047 | 17.708 | 19.768 | 39.087 | 42.557 | 45.722 | 49.588 | 52.336 |
| 30 | 13.787 | 14.953 | 16.791 | 18.493 | 20.599 | 40.256 | 43.773 | 46.979 | 50.892 | 53.672 |
| 40 | 20.707 | 22.164 | 24.433 | 26.509 | 29.051 | 51.805 | 55.758 | 59.342 | 63.691 | 66.766 |
| 50 | 27.991 | 29.707 | 32.357 | 34.764 | 37.689 | 63.167 | 67.505 | 71.420 | 76.154 | 79.490 |
| 60 | 35.534 | 37.485 | 40.482 | 43.188 | 46.459 | 74.397 | 79.082 | 83.298 | 88.379 | 91.952 |
| 70 | 43.275 | 45.442 | 48.758 | 51.739 | 55.329 | 85.527 | 90.531 | 95.023 | 100.425 | 104.215 |
| 80 | 51.172 | 53.540 | 57.153 | 60.391 | 64.278 | 96.578 | 101.879 | 106.629 | 112.329 | 116.321 |
| 90 | 59.196 | 61.754 | 65.647 | 69.126 | 73.291 | 107.565 | 113.145 | 118.136 | 124.116 | 128.299 |
| 100 | 67.328 | 70.065 | 74.222 | 77.929 | 82.358 | 118.498 | 124.342 | 129.561 | 135.807 | 140.169 |