

1. In this question we study the relationship between the concepts ‘uncorrelated’ and ‘independent’.

(a) Suppose that $(X, Y) \sim N_2(\mu, \Sigma)$, where

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}.$$

Show that X and Y are independent. Hint: consider the joint density of X and Y , and show that it can be written as the product of the marginal density of X and the marginal density of Y .

- (b) In part a) we saw that if (X, Y) have a bivariate normal distribution, and the correlation between X and Y is zero, then it follows that X and Y are independent. This does not hold in general for other distributions. Give a simple example where X and Y are uncorrelated, but not independent.
- (c) (Optional) Suppose that X has a univariate normal distribution, Y has a univariate normal distribution, and the correlation between X and Y is zero. Can we conclude that X and Y are independent? Give a counterexample or a proof.
2. (a) Proof Corollary 3.2.1.1 of the MKB book (see slide 13 on the Multivariate normal distribution).
- (b) Proof Corollaries 3.4.1.1-3.4.1.3 and 3.4.2.1 of the MKB book (see slide 18 on the Multivariate normal distribution).
3. In this question we perform a two-sample T^2 test for difference in means for Bumpus’ sparrow data, comparing the birds who died to the birds who survived. Recall that birds 1-21 survived the storm, and that birds 22-49 died.

(a) Load the data, using:

```
bumpus <-
read.table("http://www.ndsu.nodak.edu/ndsu/doetkott/introsas/rawdata/bumpus.html",
           skip=20, nrows=49, col.names=c("id","total","alar","head","humerus",
           "sternum"))
bumpus <- bumpus[,-1]
attach(bumpus)
```

- (b) Write down the assumptions you need to make in order for the two-sample T^2 test to be valid.
- (c) Write down the null hypothesis and the alternative hypothesis for the test.
- (d) Write down the test statistic, and give the distribution of this test statistic under the null hypothesis. Compute the value of the test statistic for the given data. Hint: look at “Rscript on testing for means”.
- (e) Give the definition of a p-value. Compute the p-value for your test, and state your conclusions.
- (f) Check if your p-value corresponds to the one given by `manova`:

```
Y <- cbind(total, alar, head, humerus, sternum)
Type <- c(rep("surv",21),rep("dead",28))
res <- manova(Y ~ factor(Type))
summary(res)
```

4. In this exercise we apply Fisher’s linear discriminant analysis to the iris data set, which contains measurements on sepal width, sepal length, petal width and petal length for a total of 150 flowers of 3 different types of irises.

(a) We will only look at the two species that are most difficult to separate: ‘versicolor’ and ‘virginica’. Load the data using the following code:

```
data(iris)

# only consider the species ‘versicolor’ and ‘virginica’
```

```
dat <- iris[c(51:150),]  
# re-factorize the last column to get rid of the empty class  
dat[,5] <- factor(dat[,5])
```

- (b) Determine the vector a in Fisher's linear discriminant analysis for these data. Hint: use the following code:

```
library(MASS)  
res <- lda(Species ~ ., data=dat)  
# the '.' means that we want to use all 4 available measurements
```

- (c) Suppose we have a new flower with the following measurements: sepal length=6, sepal width=3, petal length=4, and petal width=1. In which group will this flower be classified?
- (d) Evaluate the performance of this classification method, using the leave-one-out method. Determine the misclassification rate. Hint: see "Rscript on linear discriminant analysis".