Thyroid Hormone Levels in Males and Females: Does Age Matter and are the Two Genders Different?

By

Aditya Kumar

Department of Civil & Environmental Engineering
Michigan Technological University, Houghton, MI, USA

Table of Contents

Introduction

The thyroid gland, or simply the thyroid is an endocrine gland in our body. It influences almost all of the metabolic processes in our body by producing thyroid hormones, the principal ones being triiodothyronine (T_3) and thyroxine (also referred to as tetraiodothyronine (T_4)).Thyroid disorders can range from being inconsequential to our health to being causes of cancer and hence life threatening.

Thyroid function tests (TFTs) are blood tests used to check the function of the thyroid in individuals. They involve measurement of thyroid hormones such as thyroid-stimulating hormone (TSH), thyroxine (T4), triiodothyronine (T3), thyroid utilization rate (T4U) and free thyroxine index (FTI). All of them have units of concentration (mass/volume) and in this study, these are referred to as thyroid measurement parameters.

There are a number of causes of thyroid disorders (e.g. iodine deficiency, pregnancy). However, of late stress has been increasingly related to thyroid disorders, which implies that healthy individuals with none of the common causes of thyroid disorders could suffer from them due to stress alone. Another interesting question would be how do thyroid levels differ across age groups of healthy individuals of a given gender and what are the differences across genders within the same age group.

This study uses thyroid measurement records from 1984 to 1987, filtered to consider healthy individuals and employs multivariate data analysis techniques to determine the variations in thyroid measurement parameters across age and gender.

Data

This study employs Thyroid disease records supplied by the Garavan Institute and J. Ross Quinlan, New South Wales Institute, Sydney, Australia (http://archive.ics.uci.edu/ml/datasets/Thyroid+Disease).

The dataset contains the latest version of an archive of thyroid diagnoses obtained from the Garvan Institute, consisting of 9172 records from 1984 to early 1987. Table 1 contains a list of the variables and their measurement type. A detailed description of all the variables is provided in the database. This document will only consider the variables of interest in the current study (Highlighted in Table 1).

Variable Description

1.) Age: Age is a continuous variable in the dataset. It generally varies from 1 year to 97 years in the data. However, there were multiple absurd values (e.g. 6000) which were removed in the data processing.

2.) Sex: Sex is a categorical variable in the dataset with two levels (Male and Female abbreviated as M and F respectively).

3.) Thyroid Stimulating Hormone (TSH): TSH levels are continuous.

4.) Triiodothyronine (T_3): T_3 levels are continuous.

5.) Total Thyroxine (TT4): TT4 levels are continuous.

6.) Thyroid Utilization Rate (T4U): T4U levels are continuous.

7.) Free Thyroxine Index (FTI): FTI levels are continuous.

In addition to the above variables, Thyroxine-binding globulin (TBG) levels were also of interest. However, it had a very high percentage of missing data which when removed resulted in a trivial dataset with 10 data points. Hence this variable had to be eliminated from the analysis.

The raw dataset was processed by eliminating cases of missing values in either of the 7 variables listed above. An additional filter was applied by eliminating cases having a history of thyroid disorders, on current/prior thyroid medication or having other non-normal physical/mental

Table 1: Variables in the dataset and their measurement type

Variable	Measurement Type
age:	**Continuous.**
sex:	**M, F.**
on thyroxine:	f, t.
query on thyroxine:	f, t.
on antithyroid medication:	f, t.
sick:	f, t.
pregnant:	f, t.
thyroid surgery:	f, t.
I131 treatment:	f, t.
query hypothyroid:	f, t.
query hyperthyroid:	f, t.
lithium:	f, t.
goitre:	f, t.
tumor:	f, t.
hypopituitary:	f, t.
psych:	f, t.
TSH measured:	f, t.
TSH:	**Continuous.**
T3 measured:	f, t.
T3:	**Continuous.**
TT4 measured:	f, t.
TT4:	**Continuous.**
T4U measured:	f, t.
T4U:	**Continuous.**
FTI measured:	f, t.
FTI:	**Continuous.**
TBG measured:	f, t.
referral source:	WEST, STMW, SVHC, SVI, SVHD, other.

[Note: t: True, f: False]

conditions (e.g. pregnancy, sickness, psychological problems). This was done in order to remove the effects of extreme cases and consider thyroid levels of healthy persons in the population.

Age was converted to a categorical variable by dividing it into 5 categories. Table 2 lists the categories and their names.

Table 2: Age categories

Age Range (Years)	Age Category	Name
1-20	1	Kids & Teenagers
21-40	2	Young Adults
41-60	3	Middle Aged/Adult
61-80	4	Senior Citizen
81-100	5	Old

Figures 1 and 2 show scatter plot matrices for all the variables for males and females respectively. Age and sex are categorical variables and their levels can be seen in the plots. The others are continuous variables and show some correlation among themselves which could be expected as they are all dependent on the thyroid gland functioning.

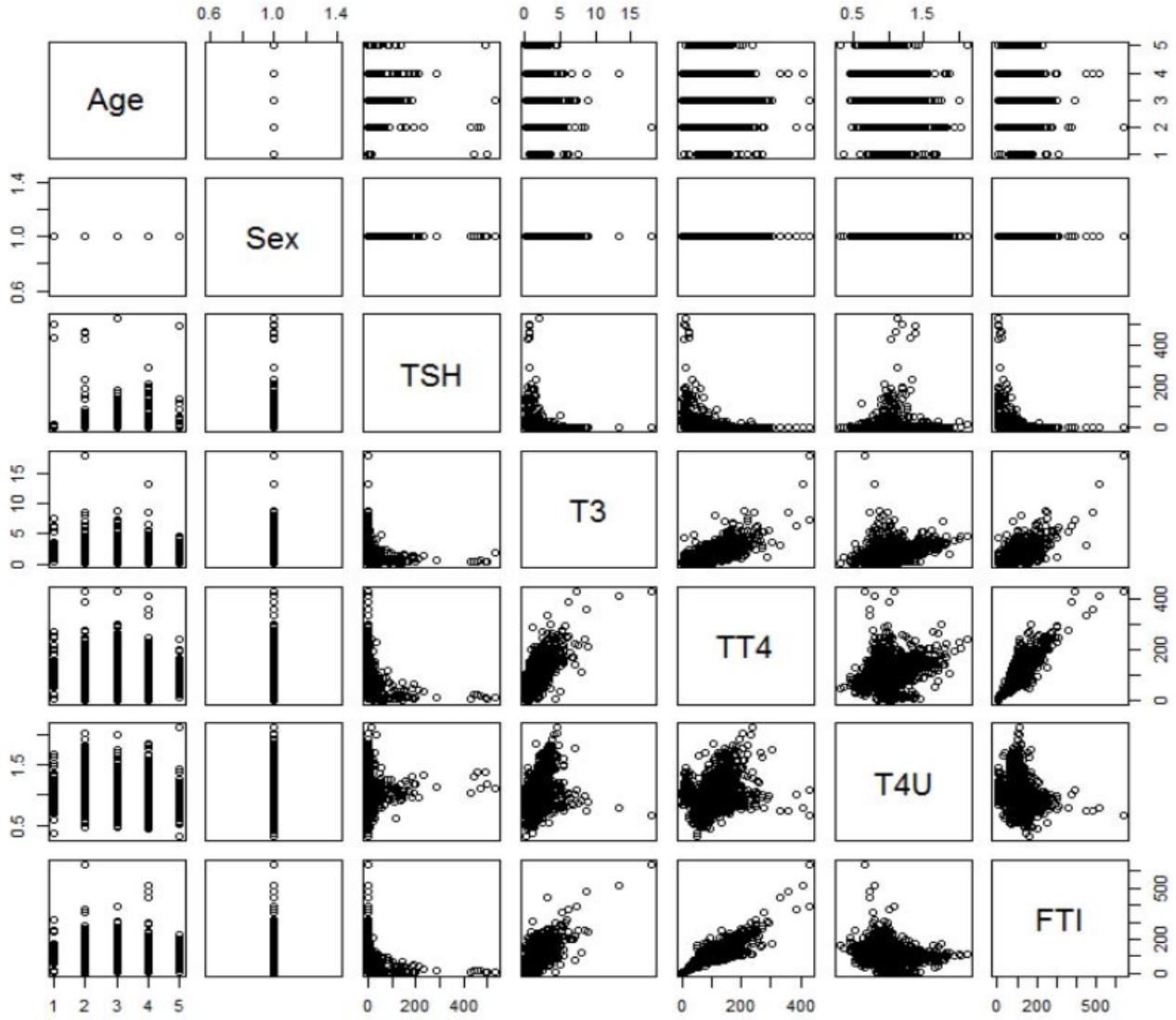

Figure 1: Scatter plot matrix for males

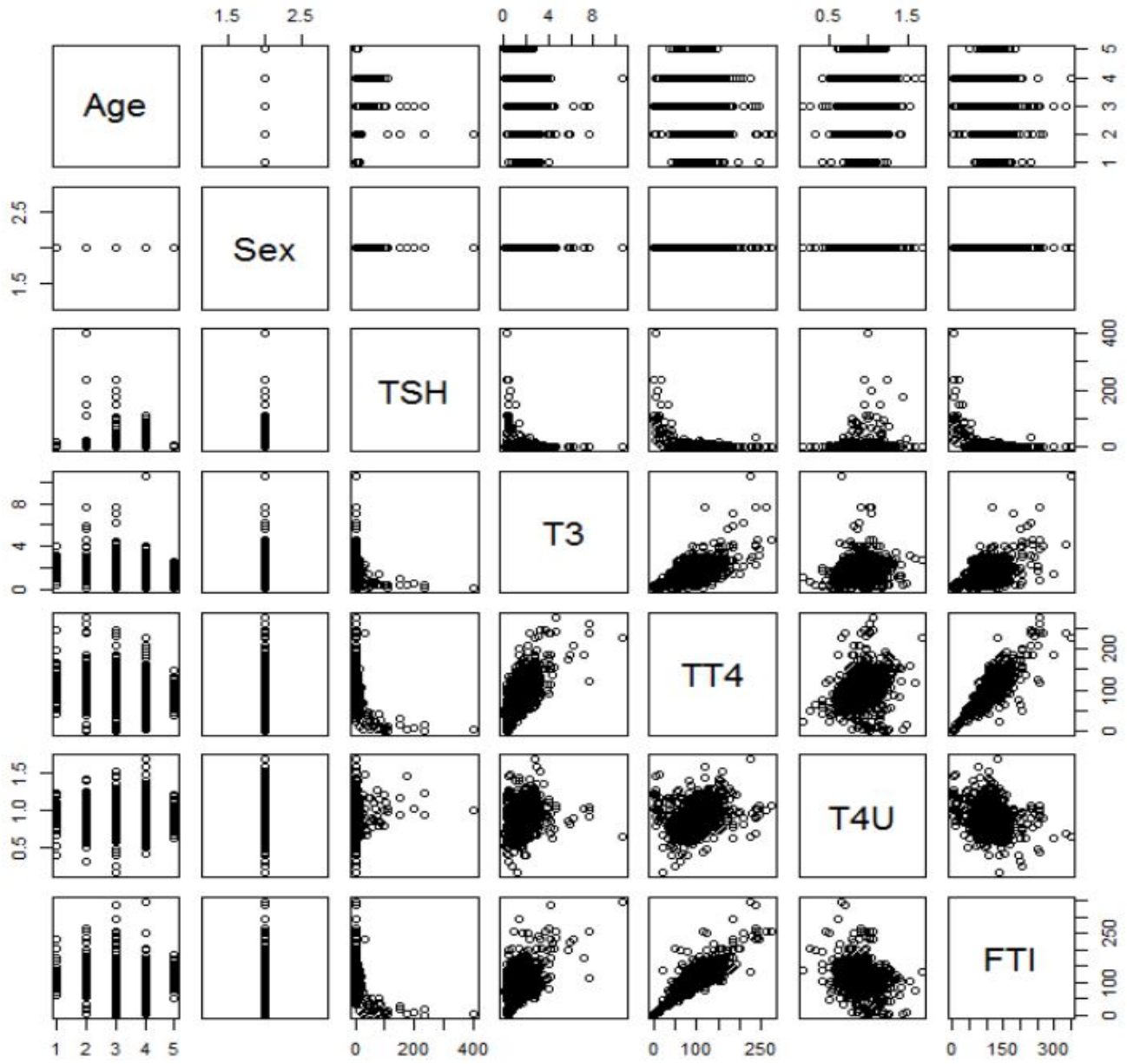

Figure 2: Scatter plot matrix for females

Methods

1.) Multivariate Two Sample T^2 Test

The multivariate two sample T^2 test is the multivariate analog of the two sample t^2 test in the univariate case. It is used to test the equality of the means of two samples. In the multivariate case, the comparison is carried out for vectors rather than scalars as in the univariate case. For a sample where p variables are measured, the hypotheses can be written as:

$$H_0: \mu_1 = \mu_2 \text{ vs } H_1: \mu_1 \neq \mu_2$$

Where

μ_1: Mean vector (p x 1) of observed variables from group 1

μ_2: Mean vector (p x 1) of observed variables from group 2

The Hotelling's T^2 statistic is computed as:

$$\mathbf{T^2 = \left(\frac{n_1 n_2}{n_1 + n_2}\right)(\overline{y_1} - \overline{y_2})' S_{pl}^{-1}(\overline{y_1} - \overline{y_2})}$$

Where:

n_1, n_2 : Sample sizes of group 1 and 2 respectively

$\overline{y_1}, \overline{y_2}$: Mean vectors for group 1 and 2 respectively

S_{pl}: Pooled covariance matrix $\frac{1}{n_1 + n_2 - 2}[(n_1 - 1)S_1 + (n_2 - 1)S_2]$

S_1, S_2: Covariance matrices for group 1 and 2 respectively

This statistic is distributed as T^2_{p,n_1+n_2-2} when H_0 is true. H_0 is rejected when $T^2 \geq T^2_{\alpha,p,n_1+n_2-2}$

The test is based on the following assumptions:

1.) The two samples follow $N_p(\mu_1, \Sigma_1)$and $N_p(\mu_2, \Sigma_2)$ and are independent.

2.) The population covariance matrices for the two samples are equal and are unknown.

The above assumptions are necessary for the T^2 statistic to follow T^2 distribution.

The multivariate two sample T^2 test was employed to test the mean vectors for TSH, T3, TT4, T4U, FTI amongst the age groups in both genders. This test would help us determine whether these thyroid measurement parameters differ significantly among individuals of different age groups for a given sex group and if so, are the results different for both sex groups.

So, the hypotheses were formulated as:

$$\mathbf{H_0}: \begin{pmatrix} \mu_{TSH_i} \\ \mu_{T3_i} \\ \mu_{TT4_i} \\ \mu_{T4U_i} \\ \mu_{FTI_i} \end{pmatrix}_j = \begin{pmatrix} \mu_{TSH_i} \\ \mu_{T3_i} \\ \mu_{TT4_i} \\ \mu_{T4U_i} \\ \mu_{FTI_i} \end{pmatrix}_j, \mathbf{H_1}: \begin{pmatrix} \mu_{TSH_i} \\ \mu_{T3_i} \\ \mu_{TT4_i} \\ \mu_{T4U_i} \\ \mu_{FTI_i} \end{pmatrix}_j \neq \begin{pmatrix} \mu_{TSH_i} \\ \mu_{T3_i} \\ \mu_{TT4_i} \\ \mu_{T4U_i} \\ \mu_{FTI_i} \end{pmatrix}_j$$

i: sex group, μ_k: mean for the test measure k, j: age group

Rejection of H_0 would imply that means of thyroid measurement parameters are different between the age groups being considered.

2.) Testing Individual Variables Conditioned on Rejection of H_0 by the Multivariate Two Sample T^2 Test

This test can be used to determine which variables contributed the most to the rejection of H_0 by the multivariate two sample T^2 Test. It would be important to determine which thyroid parameters differ most in the different age groups for a given gender that result in the groups being different.

The test involves computation of the discriminant function (z).

If $H_0: \mu_1 = \mu_2$ was rejected by T^2, the discriminant function leads to the rejection of $H_0: a'\mu_1 = a'\mu_2$ with the test statistic given as:

$$\mathbf{t(a)} = \frac{\mathbf{a'\overline{y_1} - a'\overline{y_2}}}{\sqrt{\left[\frac{\mathbf{n_1 + n_2}}{\mathbf{n_1 n_2}}\right]\mathbf{a' S_{pl} a}}}$$

Since t(a) is negative $t^2(a)$ is generally used.

The value of a that maximizes $t^2(a)$ and hence increases the likelihood of rejection of H_0 is $a = S_{pl}^{-1}(\overline{y_1} - \overline{y_2})$

Subsequently, the coefficients of the y's in the function $z = a'y$ where $y = \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix}$ can be examined to determine the contribution of the corresponding y_j's to the rejection of $H_0: \mu_1 = \mu_2$.

This test has the advantage of taking into account the effect of each variable on T^2 in the presence of other variables. This is not the case with univariate t-tests. If the variables are not similar in scale and variance then the coefficients in z should be standardized before comparing.

In the current dataset, TT4 and FTI had values which were higher than the other variables, the coefficients were standardized before carrying out the comparisons.

3.)Two Sample Profile Analysis

Two sample profile analysis can be used to compare the profiles obtained by connecting the points $(j, \mu_{1j})(j = 1,2, \cdots p)$ with $(j, \mu_{2j})(j = 1,2, \cdots p)$. One of the commonly used hypotheses is whether the profiles are parallel. In the case of the profiles being parallel, levels in one group are constantly above/below those in the other group. The hypotheses can be expressed as:

$$H_0: \begin{pmatrix} \mu_{1,2} - \mu_{1,1} \\ \vdots \\ \mu_{1,p} - \mu_{1,p-1} \end{pmatrix} = \begin{pmatrix} \mu_{2,2} - \mu_{2,1} \\ \vdots \\ \mu_{2,p} - \mu_{2,p-1} \end{pmatrix}$$

$H_0: C\mu_1 = C\mu_2$ where C is the contrast matrix

$$C = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

The test statistic is:

$$\mathbf{T^2 = (C\overline{y_1} - C\overline{y_2})' \left[\left[\frac{1}{n_1} + \frac{1}{n_2}\right] CS_{pl}C'\right]^{-1} (C\overline{y_1} - C\overline{y_2})}$$

T^2 is distributed as T_{p-1,n_1+n_2-2}

In order to study the thyroid test parameters' variation among the two genders for a given age group, a profile analysis was carried out. The means of all the thyroid test parameters for a given age group of males were compared with the corresponding ones for females. The hypothesis of parallelism was employed to compare the mean vectors. Parallel profiles would indicate the thyroid parameters of one age group in a particular gender were consistently above/below the corresponding age group in the other gender.

$$\mathbf{H_0}: \begin{pmatrix} \mu_{T3_i} - \mu_{TSH_i} \\ \mu_{TT4_i} - \mu_{T3_i} \\ \mu_{T4U_i} - \mu_{TT4_i} \\ \mu_{FTI_i} - \mu_{T4U_i} \end{pmatrix}_j = \begin{pmatrix} \mu_{T3_i} - \mu_{TSH_i} \\ \mu_{TT4_i} - \mu_{T3_i} \\ \mu_{T4U_i} - \mu_{TT4_i} \\ \mu_{FTI_i} - \mu_{T4U_i} \end{pmatrix}_k$$

i: age group, μ_l: mean for the test measure l, j: sex group 1, k: sex group 2

The test is based on the following assumptions:

a.) The two samples follow $N_p(\mu_1, \Sigma_1)$and $N_p(\mu_2, \Sigma_2)$ and are independent.

b.) The population covariance matrices for the two samples are equal and are unknown.

c.) Cy_{1i} and Cy_{2i} are distributed as $N_{p-1}(C\mu_1, C\Sigma C')$and $N_{p-1}(C\mu_2, C\Sigma C')$respectively.

d.) Under H_0, the vector $(C\overline{y_1} - C\overline{y_2})$ is distributed as $N_{p-1}\left(0, C\Sigma C'\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$.

Checking for Assumptions

The assumption of multivariate normality for each of the age groups in both gender groups was checked using the Mardia test. The test results showed that the thyroid variables (y_i's) in none of the groups followed the multivariate normal distribution. This indicates that other assumptions such as $(C\overline{y_1} - C\overline{y_2})$ is distributed as $N_{p-1}\left(0, C\Sigma C'\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$ would also not be satisfied.

This would be expected and is a common outcome for real world datasets that would have outliers which would result in deviation from normality even though the sampled population follows the multivariate normal distribution. Since application of outlier removal techniques without altering the significant properties of the dataset was not the major motivation for this work, and the author's knowledge was limited in this regard due to these techniques being outside the scope of the class, the analysis was carried out in the presence of outliers. Also, considering the large sample size at hand (> 4000 data points), and that the y_i's were independent it was felt that deviations from normality may not have a significant impact on the results of the analysis. Therefore, the analysis was carried out in the presence of the violation of assumption of normality.

Results and Discussion

1.)Multivariate Two Sample T^2 Test

Tables 3 and 4 show the results of the multivariate two sample T^2 test for males and females respectively. The results showed that the means of thyroid parameters were not significantly different between the group of kids & teenagers and young adults and middle aged/adult. However, the levels differed between kids & teenagers and the older aged groups of senior citizens and old individuals. Comparison between all other groups yielded differences in means of the thyroid measurement parameters.

Table 3: Results for multivariate two sample T^2 test (males)

Age Groups (Males)	Mean Vectors
Kids & Teenagers v/s Young Adults	Same
Kids & Teenagers v/s Middle Aged/Adult	Same
Kids & Teenagers v/s Senior Citizen	Different
Kids & Teenagers v/s Old	Different
Young Adults v/s Middle Aged/Adult	Different
Young Adults v/s Senior Citizen	Different
Young Adults v/s Old	Different
Middle Aged/Adult v/s Senior Citizen	Different
Middle Aged/Adult v/s Old	Different
Senior Citizen v/s Old	Different

For females, the result was similar when the group of kids & teenagers was compared with young adults. However for all other groups, the means were different. Thus, female mean thyroid parameter levels showed more cases of difference than males, although the difference was of one case only.

Overall, it could be concluded that thyroid parameter levels in both genders differ between age groups.

Table 4: Results for multivariate two sample T^2 test (females)

Age Groups (Females)	Mean Vectors
Kids & Teenagers v/s Young Adults	Same
Kids & Teenagers v/s Middle Aged/Adult	Different
Kids & Teenagers v/s Senior Citizen	Different
Kids & Teenagers v/s Old	Different
Young Adults v/s Middle Aged/Adult	Different
Young Adults v/s Senior Citizen	Different
Young Adults v/s Old	Different
Middle Aged/Adult v/s Senior Citizen	Different
Middle Aged/Adult v/s Old	Different
Senior Citizen v/s Old	Different

2.) Testing Individual Variables Conditioned on Rejection of H_0 by the Multivariate Two Sample T^2 Test

The rejection of null hypothesis of equality of means of thyroid parameters indicates that atleast one of the parameters differs between two age groups in a particular gender group. The major parameters that are different were ascertained by testing the individual variables. Tables 5 and 6 show the results for males and females respectively. For all age groups between which the mean vectors differed in both genders, it was observed for both genders that the larger the separation between the groups, more the number of variables that contributed to the difference.

In females, more instances of TSH contributions were observed as compared to males. For both genders, young adults vs middle aged/adult showed contributions from all variables even though the two age groups were adjacent groups.

Table 5: Major variables contributing to differences in means across age groups (males)

Age Groups (Males)	Major Variables
Kids & Teenagers v/s Young Adults	
Kids & Teenagers v/s Middle Aged/Adult	
Kids & Teenagers v/s Senior Citizen	T3, FTI, TSH,TT4,T4U
Kids & Teenagers v/s Old	T3,FTI,TSH
Young Adults v/s Middle Aged/Adult	T4U,TT4, T3, FTI
Young Adults v/s Senior Citizen	T3,FTI,TT4
Young Adults v/s Old	FTI,TT4,T3, T4U
Middle Aged/Adult v/s Senior Citizen	FTI, T3,T4U,TT4
Middle Aged/Adult v/s Old	FTI, T3, TT4,T4U,TSH
Senior Citizen v/s Old	FTI, T3,TT4,TSH

Table 6: Major variables contributing to differences in means across age groups (females)

Age Groups (Females)	Major Variables
Kids & Teenagers v/s Young Adults	
Kids & Teenagers v/s Middle Aged/Adult	T3,FTI,T4U
Kids & Teenagers v/s Senior Citizen	T3, FTI,T4U
Kids & Teenagers v/s Old	T3,T4U,TSH,TT4,FTI
Young Adults v/s Middle Aged/Adult	TT4,FTI,T4U,T3, TSH
Young Adults v/s Senior Citizen	FTI,TT4,T3,T4U,TSH
Young Adults v/s Old	TT4,T3,FTI,T4U,TSH
Middle Aged/Adult v/s Senior Citizen	T3,FTI,T4U,TT4,TSH
Middle Aged/Adult v/s Old	T3,TT4,FTI,T4U,TSH
Senior Citizen v/s Old	TT4,T3,T4U,FTI,TSH

3.)Two Sample Profile Analysis

Table 7 shows the profile analysis results. Profiles for the two genders were parallel for the two extreme age groups of kids & teenagers and old. This indicates that all the thyroid parameters have a consistent level of difference (greater/lower) for these age groups between males and females. For other age groups, the variation would not be consistent.

Table 7: Profile analysis results for comparison between male and female of a particular age group

Age Group	Profile Analysis Result
Kids & Teenagers	Profiles are parallel
Young Adults	Profiles are not parallel
Middle Aged/Adult	Profiles are not parallel
Senior Citizen	Profiles are not parallel
Old	Profiles are parallel

In addition to the above tests a two way MANOVA test showed statistically significant effects for age (Pillai statistic: 0.078, p-value< 0.001), sex (Pillai statistic: 0.052, p-value< 0.001)and the interaction between the two variables (Pillai statistic: 0.006, p-value< 0.001) on the thyroid parameters. Similar results were obtained from the MANCOVA analysis (using age as a continuous variable). Thus, it can be concluded that thyroid parameter levels among healthy individuals for a given gender and across genders would vary with age.

Conclusions

This study used a thyroid measurement parameter dataset from 1984-1987 and employed multivariate techniques to determine the differences in the thyroid measurement parameters between age groups and genders for healthy individuals with no thyroid disorder history or deficiencies that could cause thyroid disorders.

The results of the analyses showed that age and gender have a significant influence on the levels of thyroid measurement parameters in healthy individuals. Multivariate T^2 tests showed that both for males and females, mean thyroid measurement parameter levels were significantly different between age groups. The youngest group of kids & teenagers when compared with the adjacent group of young adults showed no significant difference for both males and females. For males, kids & teenagers when compared to middle aged/adult individuals also did not show any significant difference whereas for females all other age groups showed significant differences.

The parameters which contributed the most to the differences among age groups for a given gender group also depended on the age gap between the age groups. For groups having a large gap (e.g. kids and teenagers vs senior citizen), almost all parameters contributed to the difference. However, this pattern was much more consistent in females than in males. Also, in females the TSH parameter contributed more to differences than in males.

Profile analysis of age groups showed that thyroid measurement parameter levels would be consistently higher/lower for kids & teenagers and old individuals from the male group when compared with the female group as the profiles are parallel for only these two age groups.

References

1.) Rencher, A. C. (2012). Methods of multivariate analysis. John Wiley & Sons.

2.) Marden, J. I. (2012). Multivariate Statistics Old School. Department of statistics, University of Illinois at Urbana-Champaign, 168-172.

3.) Fox, J., Friendly, M., & Weisberg, S. (2013). Hypothesis tests for multivariate linear models using the car package. R J, 5(1), 39-52.

4.) Johnson, R. A., & Wichern, D. W. (1992). Applied multivariate statistical analysis (Vol. 4). Englewood Cliffs, NJ: Prentice hall.

5.) Crawley, M. J. (2014). Statistics: an introduction using R. John Wiley & Sons.

Appendix: R Code

```
setwd("H:/Desktop/Degrees/Mathematics/Multivariate_Statistics/Final_Project")
orig_data = read.csv("thyroid0387data.csv",
na.strings="?")
colnames(orig_data)=c("Age","Sex","On_Thyroxine","Query_on_Thyroxine", "On_Antithyroid_Medication",
                      "Sick", "Pregnant",
"Thyroid_Surgery", "I131_Treatment",
"Query_Hypothyroid",
                      "Query_Hyperthyroid", "Lithium",
"Goitre", "Tumor", "Hypopituary","Psych",

"TSH_Measured","TSH","T3_Measured","T3","TT4_Measured",
"TT4","T4U_Measured",

"T4U","FTI_Measured","FTI","TBG_Measured",
"TBG","Referral_Source", "Extra")
orig_data = orig_data[(orig_data$Age < 150), 1:29]
count =0
count1 = 0

#1<=age<=20
#21<=age<=40
#41<=age<=60
#61<=age<=80
#81<=age<=100

Age = rep(NA,length(orig_data$Age))
#Age = rep(NA,length(orig_data$Age))
#Age = rep(NA,length(orig_data$Age))
#Age = rep(NA,length(orig_data$Age))
#Age = rep(NA,length(orig_data$Age))
Sex = rep(NA, length(orig_data$Age))
TSH = rep(NA, length(orig_data$Age))
```

```
T3 =  rep(NA, length(orig_data$Age))
TT4 = rep(NA, length(orig_data$Age))
T4U = rep(NA, length(orig_data$Age))
FTI = rep(NA, length(orig_data$Age))

age_cont = rep(NA,length(orig_data$Age))

male_age = rep(NA,length(orig_data$Age))
male_sex = rep(NA,length(orig_data$Age))
male_TSH = rep(NA, length(orig_data$Age))
male_T3 =  rep(NA, length(orig_data$Age))
male_TT4 = rep(NA, length(orig_data$Age))
male_T4U = rep(NA, length(orig_data$Age))
male_FTI = rep(NA, length(orig_data$Age))


female_age = rep(NA,length(orig_data$Age))
female_sex = rep(NA,length(orig_data$Age))
female_TSH = rep(NA, length(orig_data$Age))
female_T3 =  rep(NA, length(orig_data$Age))
female_TT4 = rep(NA, length(orig_data$Age))
female_T4U = rep(NA, length(orig_data$Age))
female_FTI = rep(NA, length(orig_data$Age))


typeof(orig_data$On_Thyroxine[1])
count_age=0
for (i in 1:length(orig_data$Age)){

  if (!is.na(orig_data$Age[i]) &&
!is.na(orig_data$Sex[i]) &&
!is.na(orig_data$On_Thyroxine[i]) &&
        !is.na(orig_data$On_Antithyroid_Medication[i])
&& !is.na(orig_data$Sick[i]) &&
!is.na(orig_data$Pregnant[i]) &&
```

```
        !is.na(orig_data$Thyroid_Surgery[i]) &&
!is.na(orig_data$I131_Treatment[i])  &&
!is.na(orig_data$Query_Hypothyroid) &&
        !is.na(orig_data$Query_Hyperthyroid) &&
         !is.na(orig_data$Lithium[i]) &&
!is.na(orig_data$Goitre[i]) &&
        !is.na(orig_data$Tumor[i]) &&
!is.na(orig_data$Hypopituary[i]) &&
!is.na(orig_data$Psych[i]) && !is.na(orig_data$TSH[i])
&&
        !is.na(orig_data$T3[i]) &&
!is.na(orig_data$TT4[i]) && !is.na(orig_data$T4U[i]) &&
!is.na(orig_data$FTI[i])){

    #print("MATCHING DATA FOUND")
    count = count + 1


  if ((as.character(orig_data$On_Thyroxine[i])== 'f')
&&
(as.character(orig_data$On_Antithyroid_Medication[i])==
'f') &&
        (as.character(orig_data$Sick[i])== 'f') &&
(as.character(orig_data$Pregnant[i])== 'f') &&
(as.character(orig_data$Thyroid_Surgery[i])== 'f') &&
        (as.character(orig_data$I131_Treatment[i]) ==
'f') && (as.character(orig_data$Query_Hypothyroid) ==
'f') &&
          (as.character(orig_data$Query_Hyperthyroid)
== 'f') &&
        (as.character(orig_data$Lithium[i]) == 'f') &&
(as.character(orig_data$Goitre[i]) == 'f') &&
          (as.character(orig_data$Tumor[i]) == 'f') &&
(as.character(orig_data$Hypopituary[i]) == 'f') &&
        (as.character(orig_data$Psych[i]) == 'f')){
    count_age = count_age + 1
    #print(orig_data$On_Thyroxine[i])
```

```
    count1 = count1 + 1
    age_cont[count_age]=orig_data$Age[i]
    if (orig_data$Age[i] >= 1 && orig_data$Age[i]
<=20){
      Age[count_age] = 1
      Sex[count_age] = orig_data$Sex[i]
      TSH[count_age] = orig_data$TSH[i]
      T3[count_age] = orig_data$T3[i]
      TT4[count_age] = orig_data$TT4[i]
      T4U[count_age] = orig_data$T4U[i]
      FTI[count_age] = orig_data$FTI[i]

      } else if (orig_data$Age[i] >= 21 &&
orig_data$Age[i] <=40){
         Age[count_age] = 2
         Sex[count_age] = orig_data$Sex[i]
         TSH[count_age] = orig_data$TSH[i]
         T3[count_age] = orig_data$T3[i]
         TT4[count_age] = orig_data$TT4[i]
         T4U[count_age] = orig_data$T4U[i]
         FTI[count_age] = orig_data$FTI[i]

      } else if (orig_data$Age[i] >= 41 &&
orig_data$Age[i] <=60){
        Age[count_age] = 3
        Sex[count_age] = orig_data$Sex[i]
        TSH[count_age] = orig_data$TSH[i]
        T3[count_age] = orig_data$T3[i]
        TT4[count_age] = orig_data$TT4[i]
        T4U[count_age] = orig_data$T4U[i]
        FTI[count_age] = orig_data$FTI[i]

      } else if (orig_data$Age[i] >= 61 &&
orig_data$Age[i] <=80){
        Age[count_age] = 4
        Sex[count_age] = orig_data$Sex[i]
        TSH[count_age] = orig_data$TSH[i]
```

```
        T3[count_age] = orig_data$T3[i]
        TT4[count_age] = orig_data$TT4[i]
        T4U[count_age] = orig_data$T4U[i]
        FTI[count_age] = orig_data$FTI[i]

      } else if (orig_data$Age[i] >= 81 &&
orig_data$Age[i] <=100){
        Age[count_age] = 5
        Sex[count_age] = orig_data$Sex[i]
        TSH[count_age] = orig_data$TSH[i]
        T3[count_age] = orig_data$T3[i]
        TT4[count_age] = orig_data$TT4[i]
        T4U[count_age] = orig_data$T4U[i]
        FTI[count_age] = orig_data$FTI[i]

      }

    }


  }


}
print(count1)
y = Age[(complete.cases(Age))]
print(min(y))
print(length(y))
print(y)

#MANOVA analysis
model1 = manova((cbind(TSH[(complete.cases(TSH))],
T3[(complete.cases(T3))], TT4[(complete.cases(TT4))],
T4U[(complete.cases(T4U))],FTI[(complete.cases(FTI))])~
as.factor(Age[(complete.cases(Age))])*as.factor(Sex[(co
mplete.cases(Sex))])))
summary(model1)
```

```
#MANCOVA analysis
model2= manova((cbind(TSH[(complete.cases(TSH))],
T3[(complete.cases(T3))], TT4[(complete.cases(TT4))],
T4U[(complete.cases(T4U))],FTI[(complete.cases(FTI))])~
age_cont[(complete.cases(age_cont))]*as.factor(Sex[(com
plete.cases(Sex))])))

summary(model2)

data_frame_comb =
data.frame(cbind(Age[(complete.cases(Age))],Sex[(comple
te.cases(Sex))],
TSH[(complete.cases(TSH))],T3[(complete.cases(T3))],
TT4[(complete.cases(TT4))], T4U[(complete.cases(T4U))],
FTI[(complete.cases(FTI))]))

#test = is.na(data_frame)
#print(test[test==TRUE])
#male_data =
colnames(data_frame_comb)=c("Age",
"Sex","TSH","T3","TT4","T4U","FTI")
head(data_frame_comb)
print(length(data_frame_comb$Sex))
male_data_frame =
data_frame_comb[data_frame_comb$Sex=="1",]
print(length(male_data_frame$Sex))
#print(length(male_data_frame$Age))

female_data_frame =
data_frame_comb[data_frame_comb$Sex=="2",]
print(length(female_data_frame$Sex))


# MALE DATA SET
male_model =
manova(cbind(male_data_frame$TSH,male_data_frame$T3,
```

```
male_data_frame$TT4, male_data_frame$T4U,
male_data_frame$FTI)~as.factor(male_data_frame$Age))
summary(male_model)


# FEMALE DATA SET
female_model =
manova(cbind(female_data_frame$TSH,female_data_frame$T3
, female_data_frame$TT4, female_data_frame$T4U,
female_data_frame$FTI)~as.factor(female_data_frame$Age)
)
summary(female_model)


#scatter plots for male and female
pairs(male_data_frame)
pairs(female_data_frame)


print(mean(male_data_frame$TSH))
print(mean(male_data_frame$T3))
print(mean(male_data_frame$TT4))
print(mean(male_data_frame$T4U))
print(mean(male_data_frame$FTI))

male_1 = male_data_frame[male_data_frame$Age=="1",3:7]
male_2 = male_data_frame[male_data_frame$Age=="2",3:7]
male_3 = male_data_frame[male_data_frame$Age=="3",3:7]
male_4 = male_data_frame[male_data_frame$Age=="4",3:7]
male_5 = male_data_frame[male_data_frame$Age=="5",3:7]

female_1 =
female_data_frame[female_data_frame$Age=="1",3:7]
female_2 =
female_data_frame[female_data_frame$Age=="2",3:7]
female_3 =
female_data_frame[female_data_frame$Age=="3",3:7]
```

```
female_4 =
female_data_frame[female_data_frame$Age=="4",3:7]
female_5 =
female_data_frame[female_data_frame$Age=="5",3:7]


#Hotellingst2 males
male_1=male_data_frame[male_data_frame$Age=="1",3:7]
male_2=male_data_frame[male_data_frame$Age=="2",3:7]
male_3=male_data_frame[male_data_frame$Age=="3",3:7]
male_4=male_data_frame[male_data_frame$Age=="4",3:7]
male_5=male_data_frame[male_data_frame$Age=="5",3:7]



library(ICSNP)
HotellingsT2(X=male_1,Y=male_2)
HotellingsT2(X=male_1,Y=male_3)
HotellingsT2(X=male_1,Y=male_4)
HotellingsT2(X=male_1,Y=male_5)
HotellingsT2(X=male_2,Y=male_3)
HotellingsT2(X=male_2,Y=male_4)
HotellingsT2(X=male_2,Y=male_5)
HotellingsT2(X=male_3,Y=male_4)
HotellingsT2(X=male_3,Y=male_5)
HotellingsT2(X=male_4,Y=male_5)

#mardia test males

library(MVN)

mardiaTest(as.matrix(male_1))
mardiaTest(as.matrix(male_2))
mardiaTest(as.matrix(male_3))
mardiaTest(as.matrix(male_4))
mardiaTest(as.matrix(male_5))

#Hotellingst2 females
female_1=female_data_frame[female_data_frame$Age=="1",3
:7]
```

```
female_2=female_data_frame[female_data_frame$Age=="2",3
:7]
female_3=female_data_frame[female_data_frame$Age=="3",3
:7]
female_4=female_data_frame[female_data_frame$Age=="4",3
:7]
female_5=female_data_frame[female_data_frame$Age=="5",3
:7]



mardiaTest(as.matrix(female_1))
mardiaTest(as.matrix(female_2))
mardiaTest(as.matrix(female_3))
mardiaTest(as.matrix(female_4))
mardiaTest(as.matrix(female_5))


library(ICSNP)
HotellingsT2(X=female_1,Y=female_2)
HotellingsT2(X=female_1,Y=female_3)
HotellingsT2(X=female_1,Y=female_4)
HotellingsT2(X=female_1,Y=female_5)
HotellingsT2(X=female_2,Y=female_3)
HotellingsT2(X=female_2,Y=female_4)
HotellingsT2(X=female_2,Y=female_5)
HotellingsT2(X=female_3,Y=female_4)
HotellingsT2(X=female_3,Y=female_5)
HotellingsT2(X=female_4,Y=female_5)

#mshapiro.test(as.matrix(female_1))
#mshapiro.test(as.matrix(female_2))
#mshapiro.test(as.matrix(female_3))
#mshapiro.test(as.matrix(female_4))
#mshapiro.test(as.matrix(female_5))



#Profile Analysis
```

```
d_m_1 = cbind(male_1[,2]-male_1[,1],male_1[,3]-
male_1[,2],male_1[,4]-male_1[,3],male_1[,5]-male_1[,4])
d_f_1 = cbind(female_1[,2]-female_1[,1],female_1[,3]-
female_1[,2],female_1[,4]-female_1[,3],female_1[,5]-
female_1[,4])
HotellingsT2(X=d_m_1,Y=d_f_1)

d_m_2 = cbind(male_2[,2]-male_2[,1],male_2[,3]-
male_2[,2],male_2[,4]-male_2[,3],male_2[,5]-male_2[,4])
d_f_2 = cbind(female_2[,2]-female_2[,1],female_2[,3]-
female_2[,2],female_2[,4]-female_2[,3],female_2[,5]-
female_2[,4])
HotellingsT2(X=d_m_2,Y=d_f_2)

d_m_3 = cbind(male_3[,2]-male_3[,1],male_3[,3]-
male_3[,2],male_3[,4]-male_3[,3],male_3[,5]-male_3[,4])
d_f_3 = cbind(female_3[,2]-female_3[,1],female_3[,3]-
female_3[,2],female_3[,4]-female_3[,3],female_3[,5]-
female_3[,4])
HotellingsT2(X=d_m_3,Y=d_f_3)


d_m_4 = cbind(male_4[,2]-male_4[,1],male_4[,3]-
male_4[,2],male_4[,4]-male_4[,3],male_4[,5]-male_4[,4])
d_f_4 = cbind(female_4[,2]-female_4[,1],female_4[,3]-
female_4[,2],female_4[,4]-female_4[,3],female_4[,5]-
female_4[,4])
HotellingsT2(X=d_m_4,Y=d_f_4)

d_m_5 = cbind(male_5[,2]-male_5[,1],male_5[,3]-
male_5[,2],male_5[,4]-male_5[,3],male_5[,5]-male_5[,4])
d_f_5 = cbind(female_5[,2]-female_5[,1],female_5[,3]-
female_5[,2],female_5[,4]-female_5[,3],female_5[,5]-
female_5[,4])
HotellingsT2(X=d_m_5,Y=d_f_5)
```

```
#DISCRIMINANT FUNCTION ANALYSIS for males


#covariance matrices for age groups
SM1 = cov(as.matrix(male_1))
SM2 = cov(as.matrix(male_2))
SM3 = cov(as.matrix(male_3))
SM4 = cov(as.matrix(male_4))
SM5 = cov(as.matrix(male_5))

#sample lengths
nm1 = length(male_1$TSH)
nm2 = length(male_2$TSH)
nm3 = length(male_3$TSH)
nm4 = length(male_4$TSH)
nm5 = length(male_5$TSH)


#mean vectors
ym1bar = c(mean(male_1$TSH),mean(male_1$T3),
mean(male_1$TT4), mean(male_1$T4U), mean(male_1$FTI))
ym2bar = c(mean(male_2$TSH),mean(male_2$T3),
mean(male_2$TT4), mean(male_2$T4U), mean(male_2$FTI))
ym3bar = c(mean(male_3$TSH),mean(male_3$T3),
mean(male_3$TT4), mean(male_3$T4U), mean(male_3$FTI))
ym4bar = c(mean(male_4$TSH),mean(male_4$T3),
mean(male_4$TT4), mean(male_4$T4U), mean(male_4$FTI))
ym5bar = c(mean(male_5$TSH),mean(male_5$T3),
mean(male_5$TT4), mean(male_5$T4U), mean(male_5$FTI))

#case1: male_1 and male_4
Spooledm1 = ((nm1/(nm1+nm4))*SM1)+((nm4/(nm1+nm4))*SM4)
zeem1 = t(ym1bar-ym4bar)%*%solve(Spooledm1)
zeem1
zeem1*sqrt(diag(Spooledm1))
#sqrt(diag(Spooledm1))
#case2: male_1 and male_5
```

```
Spooledm2 = ((nm1/(nm1+nm5))*SM1)+((nm5/(nm1+nm5))*SM5)
zeem2 = t(ym1bar-ym5bar)%*%solve(Spooledm2)
zeem2
zeem2*sqrt(diag(Spooledm2))
#case3: male_2 and male_3
Spooledm3 = ((nm2/(nm2+nm3))*SM2)+((nm3/(nm2+nm3))*SM3)
zeem3 = t(ym2bar-ym3bar)%*%solve(Spooledm3)
zeem3
zeem3*sqrt(diag(Spooledm3))
#case4: male_2 and male_4
Spooledm4 = ((nm2/(nm2+nm4))*SM2)+((nm4/(nm2+nm4))*SM4)
zeem4 = t(ym2bar-ym4bar)%*%solve(Spooledm4)
zeem4
zeem4*sqrt(diag(Spooledm4))
#case5: male_2 and male_5
Spooledm5 = ((nm2/(nm2+nm5))*SM2)+((nm5/(nm2+nm5))*SM5)
zeem5 = t(ym2bar-ym5bar)%*%solve(Spooledm5)
zeem5
zeem5*sqrt(diag(Spooledm5))
#case6: male_3 and male_4
Spooledm6 = ((nm3/(nm3+nm4))*SM3)+((nm4/(nm3+nm4))*SM4)
zeem6 = t(ym3bar-ym4bar)%*%solve(Spooledm6)
zeem6
zeem6*sqrt(diag(Spooledm6))
#case7: male_3 and male_5
Spooledm7 = ((nm3/(nm3+nm5))*SM3)+((nm5/(nm3+nm5))*SM5)
zeem7 = t(ym3bar-ym5bar)%*%solve(Spooledm7)
zeem7
zeem7*sqrt(diag(Spooledm7))
#case8: male_4 and male_5
Spooledm8 = ((nm4/(nm4+nm5))*SM4)+((nm5/(nm4+nm5))*SM5)
zeem8 = t(ym4bar-ym5bar)%*%solve(Spooledm8)
zeem8
zeem8*sqrt(diag(Spooledm8))
```

```
#DISCRIMINANT ANALYSIS for females


#covariance matrices for age groups
SF1 = cov(scale(female_1))
SF2 = cov(as.matrix(female_2))
SF3 = cov(as.matrix(female_3))
SF4 = cov(as.matrix(female_4))
SF5 = cov(as.matrix(female_5))

#sample lengths
nf1 = length(female_1$TSH)
nf2 = length(female_2$TSH)
nf3 = length(female_3$TSH)
nf4 = length(female_4$TSH)
nf5 = length(female_5$TSH)


#mean vectors
yf1bar = c(mean(female_1$TSH),mean(female_1$T3),
mean(female_1$TT4), mean(female_1$T4U),
mean(female_1$FTI))
yf2bar = c(mean(female_2$TSH),mean(female_2$T3),
mean(female_2$TT4), mean(female_2$T4U),
mean(female_2$FTI))
yf3bar = c(mean(female_3$TSH),mean(female_3$T3),
mean(female_3$TT4), mean(female_3$T4U),
mean(female_3$FTI))
yf4bar = c(mean(female_4$TSH),mean(female_4$T3),
mean(female_4$TT4), mean(female_4$T4U),
mean(female_4$FTI))
yf5bar = c(mean(female_5$TSH),mean(female_5$T3),
mean(female_5$TT4), mean(female_5$T4U),
mean(female_5$FTI))
```

```
#case1: female_1 and female_3
Spooledf1 = ((nf1/(nf1+nf3))*SF1)+((nf3/(nf1+nf3))*SF3)
zeef1 = t(yf1bar-yf3bar)%*%solve(Spooledf1)
zeef1
zeef1*sqrt(diag(Spooledf1))
#case2: female_1 and female_4
Spooledf2 = ((nf1/(nf1+nf4))*SF1)+((nf4/(nf1+nf4))*SF4)
zeef2 = t(yf1bar-yf4bar)%*%solve(Spooledf2)
zeef2
zeef2*sqrt(diag(Spooledf2))
#case3: female_1 and female_5
Spooledf3 = ((nf1/(nf1+nf5))*SF1)+((nf5/(nf1+nf5))*SF5)
zeef3 = t(yf1bar-yf5bar)%*%solve(Spooledf3)
zeef3
zeef3*sqrt(diag(Spooledf3))
#case4: female_2 and female_3
Spooledf4 = ((nf2/(nf2+nf3))*SF2)+((nf3/(nf2+nf3))*SF3)
zeef4 = t(yf2bar-yf3bar)%*%solve(Spooledf4)
zeef4
zeef4*sqrt(diag(Spooledf4))
#case5: female_2 and female_4
Spooledf5 = ((nf2/(nf2+nf4))*SF2)+((nf4/(nf2+nf4))*SF4)
zeef5 = t(yf2bar-yf4bar)%*%solve(Spooledf5)
zeef5
zeef5*sqrt(diag(Spooledf5))
#case6: female_2 and female_5
Spooledf6 = ((nf2/(nf2+nf5))*SF2)+((nf5/(nf2+nf5))*SF5)
zeef6 = t(yf2bar-yf5bar)%*%solve(Spooledf6)
zeef6
zeef6*sqrt(diag(Spooledf6))
#case7: female_3 and male_4
Spooledf7 = ((nf3/(nf3+nf4))*SF3)+((nf4/(nf3+nf4))*SF4)
zeef7 = t(yf3bar-yf4bar)%*%solve(Spooledf7)
zeef7
zeef7*sqrt(diag(Spooledf7))
#case8: female_3 and female_5
Spooledf8 = ((nf3/(nf3+nf5))*SF3)+((nf5/(nf3+nf5))*SF5)
```

```
zeef8 = t(yf3bar-yf5bar)%*%solve(Spooledf8)
zeef8
zeef8*sqrt(diag(Spooledf8))
#case9: female_4 and female_5
Spooledf9 = ((nf4/(nf4+nf5))*SF4)+((nf5/(nf4+nf5))*SF5)
zeef9 = t(yf4bar-yf5bar)%*%solve(Spooledf9)
zeef9
zeef9*sqrt(diag(Spooledf9))
```