EnvSys_final

VajiraLasantha

10/12/2021

Environmental System Analysis

Autumn 2021

Final Report

Exercises in this report use base datasets bundled with R in ***datasets*** package.

1. Influence of feed type on the weight of chickens

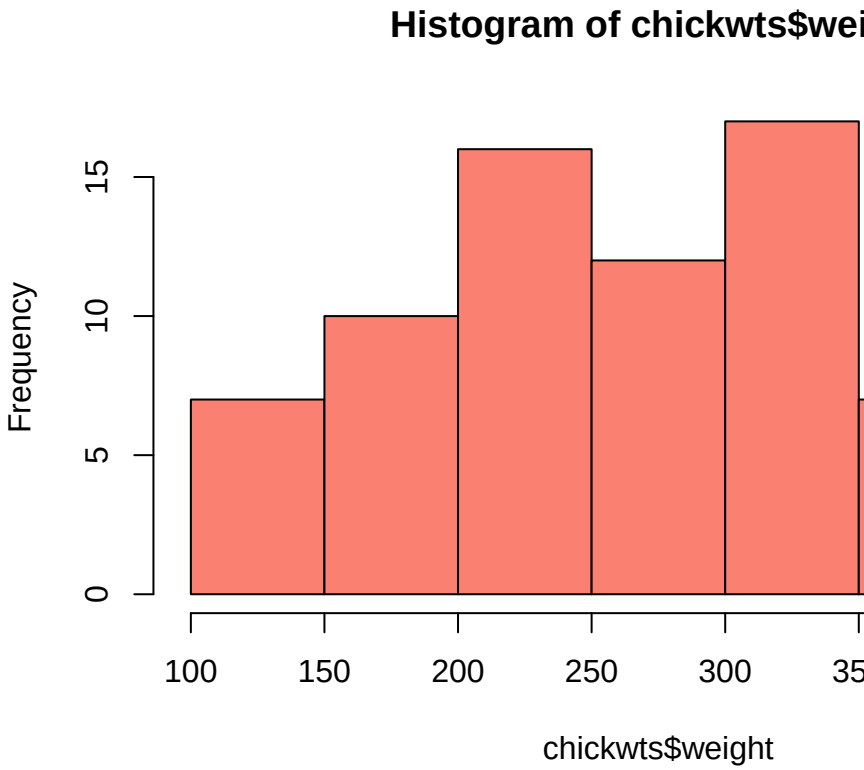
An experiment was conducted to measure and compare the effectiveness of various feed supplements on the growth rate of chickens. `chickwts` dataset from R ***datasets*** package provides the data from this study.

It has 71 observations on 2 variables; Weight, a numerical variable and feed, a categorical variable.

```
summary(chickwts)
```

```
##      weight             feed
##  Min.   :108.0   casein   :12
##  1st Qu.:204.5   horsebean:10
##  Median :258.0   linseed  :12
##  Mean   :261.3   meatmeal :11
##  3rd Qu.:323.5   soybean  :14
##  Max.   :423.0   sunflower:12
```

```
hist(chickwts$weight, breaks=6, col='salmon')
```



1.1 Histogram of the weight of the chicken

```
shapiro.test(chickwts$weight)
```

1.2 Testing for normality

```
##
##  Shapiro-Wilk normality test
##
## data:  chickwts$weight
## W = 0.97674, p-value = 0.2101
```

Shapiro-Wilks test for the wright data gives p-value of 0.21. Because p-value in greater than 0.05, the null hypothesis that the data follows normal distribution, cannot be rejected. Weight data are normally distributed.

1.4 Comparison of means Between two types of feeds

Alternative hypothesis: There is a significant difference in weight between chickens fed with Soybean feed and Sunflower feed.

Null Hypothesis: There is no significant difference in weight.

First, subset the data for Soybean and Sunflower from the original `chickwts` dataset

```
chick2 <- chickwts[chickwts$feed=='soybean' | chickwts$feed=='sunflower', ]
```

Then the data is checked for equal variance.

```
bartlett.test(chick2$weight~chick2$feed)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  chick2$weight by chick2$feed
## Bartlett's K-squared = 0.12018, df = 1, p-value = 0.7288
```

In Bartlett test for equal variance, p-value is 0.72 (> 0.05). Therefore, weight data show homogeneity of variance between different feed types. Because of this, student-t test is used for comparison of means.

```
t.test(chick2$weight~chick2$feed, var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  chick2$weight by chick2$feed
## t = -4.0502, df = 24, p-value = 0.0004641
## alternative hypothesis: true difference in means between group soybean and group sunflower is not eq
## 95 percent confidence interval:
##  -124.52226  -40.45393
## sample estimates:
##   mean in group soybean mean in group sunflower
##                246.4286                328.9167
```

Null hypothesis is rejected.

There is a significant difference in means between chicken fed with soybean feed and sunflower feed.

Between all six types of feed

```
bartlett.test(chickwts$weight~chickwts$feed)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  chickwts$weight by chickwts$feed
## Bartlett's K-squared = 3.2597, df = 5, p-value = 0.66
```

In Bartlett test for equal variance, p-value is 0.66 (> 0.05). Therefore, weight data show homogeneity of variance between different feed types. One-way ANOVA is used for comparison of means.

Null hypothesis: Means of the weights for six types of feed are equal.

```
oneway.test(weight~feed, data = chickwts, var.equal = TRUE )
```

```
##
##  One-way analysis of means
##
## data:  weight and feed
## F = 15.365, num df = 5, denom df = 65, p-value = 5.936e-10
```

With the extremely small p-value, null hypothesis can be rejected. There is a significant difference between the weight of chicken fed with different types of feed.

2. Relationship between hair colour and eye colour

Hair and eye colour and sex of 592 students are recorded the `HairEyeColor` dataset from R *datasets* package.

```
summary(HairEyeColor)
```

```
## Number of cases in table: 592
## Number of factors: 3
## Test for independence of all factors:
##  Chisq = 164.92, df = 24, p-value = 5.321e-23
##  Chi-squared approximation may be incorrect
```

Data are in seperate tables by sexes. They are combined to a single table.

```
h <- margin.table(HairEyeColor, margin = c(1,2))
h
```

```
##        Eye
## Hair    Brown Blue Hazel Green
##   Black    68   20    15     5
##   Brown   119   84    54    29
##   Red      26   17    14    14
##   Blond     7   94    10    16
```

Chi-square test for hair and eye colour

```
chisq_h <- chisq.test(h)
chisq_h
```

```
##
##  Pearson's Chi-squared test
##
## data:  h
## X-squared = 138.29, df = 9, p-value < 2.2e-16
```

p-value $<$ 2.2e-16, therefore there is significant difference between actual and expected distributions.

```
chisq_h$residuals
```

```
##        Eye
## Hair          Brown        Blue       Hazel       Green
##   Black  4.39839852 -3.06937747 -0.47735203 -1.95368354
##   Brown  1.23345810 -1.94947682  1.35328398 -0.34509961
##   Red   -0.07497794 -1.73012546  0.85225273  2.28273672
##   Blond -5.85099741  7.04959022 -2.22784430  0.61269815
```

Individual chi-squared residuals show strong associations/disassociations between some hair and eye colours.

There are strong associations between

- Blond hair and Blue eyes
- Black hair and Brown eyes

There are strong negative associations between

- Blond hair and Brown eyes
- Black hair and Blue eyes

3. Relationship between speed of cars and stopping distance

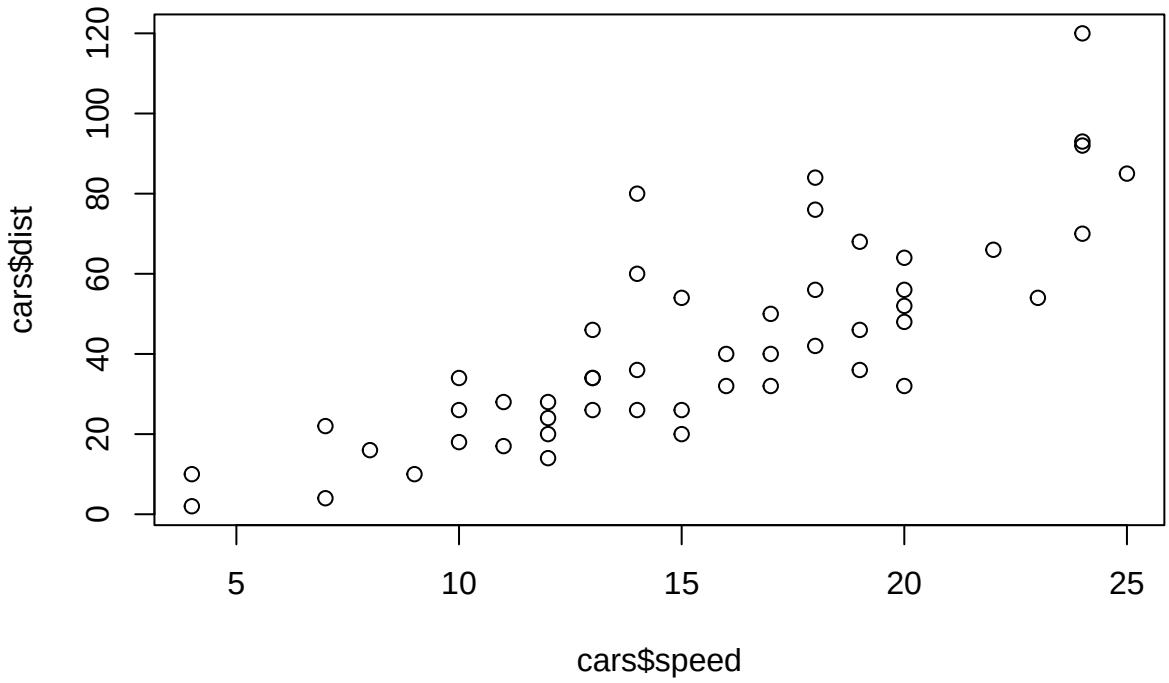
Data is from `cars` dataset. It has two variables, speed (mph) and stopping distance (ft).

```
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

3.1 Scatter plot

```
plot(cars$speed, cars$dist)
```



3.2 Correlation test

```
cor.test(cars$speed, cars$dist)
```

```
##
##  Pearson's product-moment correlation
##
## data:  cars$speed and cars$dist
## t = 9.464, df = 48, p-value = 1.49e-12
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.6816422 0.8862036
## sample estimates:
##       cor
## 0.8068949
```

There is significant correlation between speed and stopping distance of cars.

3.3 Regression Analysis

Above data are fitted to a linear model using linear regression.

```
SR <- lm(cars$speed~cars$dist)
summary(SR)
```

```
##
```

```
## Call:
## lm(formula = cars$speed ~ cars$dist)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.5293 -2.1550  0.3615  2.4377  6.4179
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.28391    0.87438   9.474 1.44e-12 ***
## cars$dist    0.16557    0.01749   9.464 1.49e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.156 on 48 degrees of freedom
## Multiple R-squared:  0.6511, Adjusted R-squared:  0.6438
## F-statistic: 89.57 on 1 and 48 DF,  p-value: 1.49e-12
```

4. Employment and economic indicators

`Longley` dataset presents 7 variables from 1947 to 1962 that can be used for predicting the number of employed persons in the USA.

A multiple regression model is fitted to this data.

4.1 Multiple regression analysis

```
summary(longley)
```

```
##   GNP.deflator          GNP          Unemployed     Armed.Forces
##  Min.   : 83.00   Min.   :234.3   Min.   :187.0   Min.   :145.6
##  1st Qu.: 94.53   1st Qu.:317.9   1st Qu.:234.8   1st Qu.:229.8
##  Median :100.60   Median :381.4   Median :314.4   Median :271.8
##  Mean   :101.68   Mean   :387.7   Mean   :319.3   Mean   :260.7
##  3rd Qu.:111.25   3rd Qu.:454.1   3rd Qu.:384.2   3rd Qu.:306.1
##  Max.   :116.90   Max.   :554.9   Max.   :480.6   Max.   :359.4
##    Population         Year         Employed
##  Min.   :107.6   Min.   :1947   Min.   :60.17
##  1st Qu.:111.8   1st Qu.:1951   1st Qu.:62.71
##  Median :116.8   Median :1954   Median :65.50
##  Mean   :117.4   Mean   :1954   Mean   :65.32
##  3rd Qu.:122.3   3rd Qu.:1958   3rd Qu.:68.29
##  Max.   :130.1   Max.   :1962   Max.   :70.55
```

```
#Scaling the data
scl <- scale(longley)
scl_df <- data.frame(scl)
mlr <- lm(Employed~.,data= scl_df)
summary(mlr)
```

```
##
## Call:
## lm(formula = Employed ~ ., data = scl_df)
```

```
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.116776 -0.044896 -0.008019  0.028916  0.129669
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.752e-15  2.170e-02   0.000 1.000000
## GNP.deflator  4.628e-02  2.609e-01   0.177 0.863141
## GNP          -1.014e+00  9.479e-01  -1.070 0.312681
## Unemployed   -5.375e-01  1.300e-01  -4.136 0.002535 **
## Armed.Forces -2.047e-01  4.246e-02  -4.822 0.000944 ***
## Population   -1.012e-01  4.478e-01  -0.226 0.826212
## Year          2.480e+00  6.175e-01   4.016 0.003037 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0868 on 9 degrees of freedom
## Multiple R-squared:  0.9955, Adjusted R-squared:  0.9925
## F-statistic: 330.3 on 6 and 9 DF,  p-value: 4.984e-10
```

Multiple linear regression model is fitted to the data with $R^2 > 99.5\%$.

4.2 Multicolliniarity

```
library(car)
```

```
## Warning: package 'car' was built under R version 4.1.1
```

```
## Loading required package: carData
```

```
vif(mlr)
```

```
## GNP.deflator          GNP   Unemployed Armed.Forces   Population         Year
##    135.53244   1788.51348     33.61889      3.58893    399.15102    758.98060
```

```
cor(scl_df)
```

```
##              GNP.deflator       GNP Unemployed Armed.Forces Population
## GNP.deflator    1.0000000 0.9915892  0.6206334    0.4647442  0.9791634
## GNP             0.9915892 1.0000000  0.6042609    0.4464368  0.9910901
## Unemployed      0.6206334 0.6042609  1.0000000   -0.1774206  0.6865515
## Armed.Forces    0.4647442 0.4464368 -0.1774206    1.0000000  0.3644163
## Population      0.9791634 0.9910901  0.6865515    0.3644163  1.0000000
## Year            0.9911492 0.9952735  0.6682566    0.4172451  0.9939528
## Employed        0.9708985 0.9835516  0.5024981    0.4573074  0.9603906
##                   Year  Employed
## GNP.deflator 0.9911492 0.9708985
## GNP          0.9952735 0.9835516
## Unemployed   0.6682566 0.5024981
## Armed.Forces 0.4172451 0.4573074
## Population   0.9939528 0.9603906
## Year         1.0000000 0.9713295
## Employed     0.9713295 1.0000000
```

VIF > 10 and high correlation (>0.95) between GNP and GNP-deflator, GNP and Population, GNP and Year, Population and Unemployed, Year and Unemployed, and Population and Year.

Some variables should be removed to reduce the multicolliniarity.

4.3 Step-wise method

```
library(MASS)
stp_mr <- stepAIC(mlr,direction = 'both')
```

```
## Start:  AIC=-73.42
## Employed ~ GNP.deflator + GNP + Unemployed + Armed.Forces + Population +
##     Year
##
##                Df Sum of Sq      RSS     AIC
## - GNP.deflator  1  0.000237 0.068052 -75.361
## - Population    1  0.000385 0.068200 -75.326
## - GNP           1  0.008619 0.076434 -73.503
## <none>                      0.067815 -73.417
## - Year          1  0.121520 0.189335 -58.989
## - Unemployed    1  0.128924 0.196739 -58.375
## - Armed.Forces  1  0.175200 0.243015 -54.996
##
## Step:  AIC=-75.36
## Employed ~ GNP + Unemployed + Armed.Forces + Population + Year
##
##                Df Sum of Sq      RSS     AIC
## - Population    1  0.001567 0.069619 -76.997
## <none>                      0.068052 -75.361
## - GNP           1  0.011867 0.079919 -74.789
## + GNP.deflator  1  0.000237 0.067815 -73.417
## - Year          1  0.123825 0.191877 -60.776
## - Unemployed    1  0.177550 0.245602 -56.826
## - Armed.Forces  1  0.194384 0.262436 -55.765
##
## Step:  AIC=-77
## Employed ~ GNP + Unemployed + Armed.Forces + Year
##
##                Df Sum of Sq     RSS     AIC
## <none>                      0.06962 -76.997
## + Population    1   0.00157 0.06805 -75.361
## + GNP.deflator  1   0.00142 0.06820 -75.326
## - GNP           1   0.03767 0.10729 -72.076
## - Year          1   0.15389 0.22351 -60.334
## - Armed.Forces  1   0.19301 0.26263 -57.754
## - Unemployed    1   0.32829 0.39791 -51.106
```

```
summary(stp_mr)
```

```
##
## Call:
## lm(formula = Employed ~ GNP + Unemployed + Armed.Forces + Year,
##     data = scl_df)
##
```

```
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.120061 -0.035470 -0.006878  0.023829  0.128895
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.879e-15  1.989e-02   0.000 1.000000
## GNP          -1.137e+00  4.662e-01  -2.440 0.032833 *
## Unemployed   -5.557e-01  7.716e-02  -7.202 1.75e-05 ***
## Armed.Forces -2.011e-01  3.641e-02  -5.522 0.000180 ***
## Year          2.559e+00  5.189e-01   4.931 0.000449 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07956 on 11 degrees of freedom
## Multiple R-squared:  0.9954, Adjusted R-squared:  0.9937
## F-statistic: 589.8 on 4 and 11 DF,  p-value: 9.5e-13
```

Best model is selected. Employment is best predicted by GNP, Unemployment, Armed Forces and Year.

4.4 Principal Component Analysis

Principal component analysis is used on the same data, to reduce dimentionality.

```
rpca = prcomp(x=longley, scale=T)
summary(rpca)
```

```
## Importance of components:
##                           PC1    PC2     PC3     PC4     PC5     PC6     PC7
## Standard deviation     2.3522 1.0897 0.50221 0.12344 0.10313 0.03206 0.01608
## Proportion of Variance 0.7904 0.1696 0.03603 0.00218 0.00152 0.00015 0.00004
## Cumulative Proportion  0.7904 0.9601 0.99612 0.99830 0.99982 0.99996 1.00000
```

```
rpca$x[,1]
```

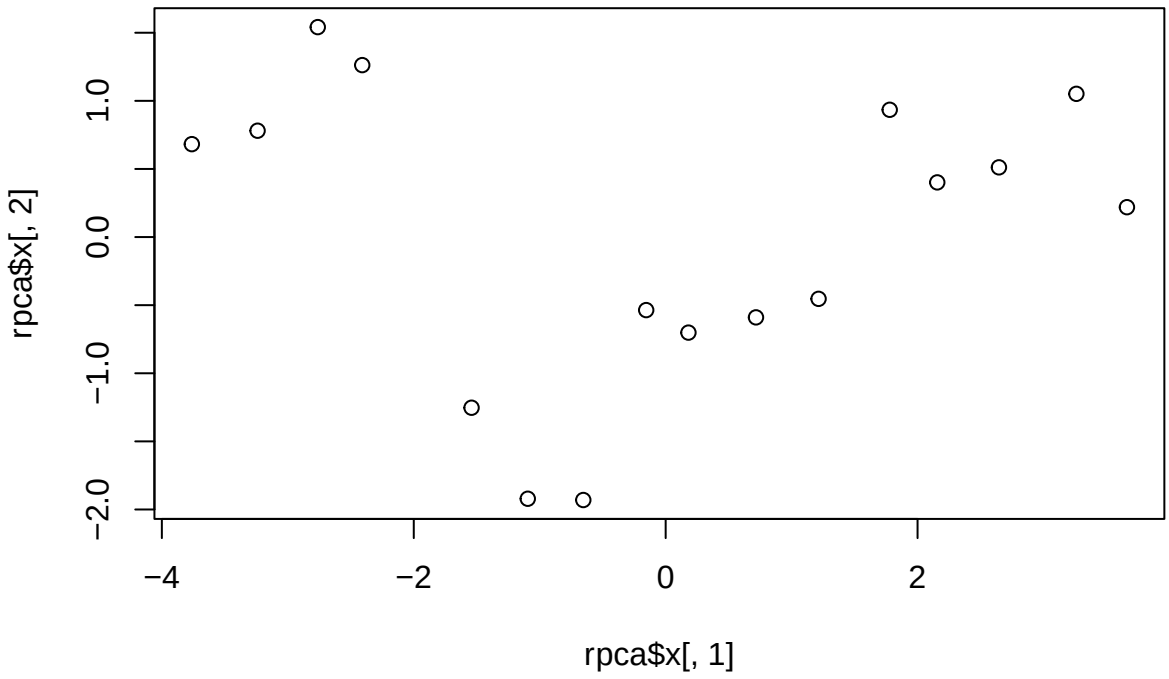
```
##       1947       1948       1949       1950       1951       1952       1953
## -3.7619812 -3.2399564 -2.7621211 -2.4091075 -1.5410767 -1.0947545 -0.6539954
##       1954       1955       1956       1957       1958       1959       1960
## -0.1545799  0.1809465  0.7170900  1.2140437  1.7790878  2.1565376  2.6463217
##       1961       1962
##  3.2609692  3.6625764
```

```
rpca$x[,2]
```

```
##       1947       1948       1949       1950       1951       1952       1953
##  0.6821378  0.7807091  1.5412760  1.2623474 -1.2529165 -1.9210578 -1.9303682
##       1954       1955       1956       1957       1958       1959       1960
## -0.5361761 -0.7014064 -0.5895881 -0.4535338  0.9344479  0.4014410  0.5117680
##       1961       1962
##  1.0515298  0.2193898
```

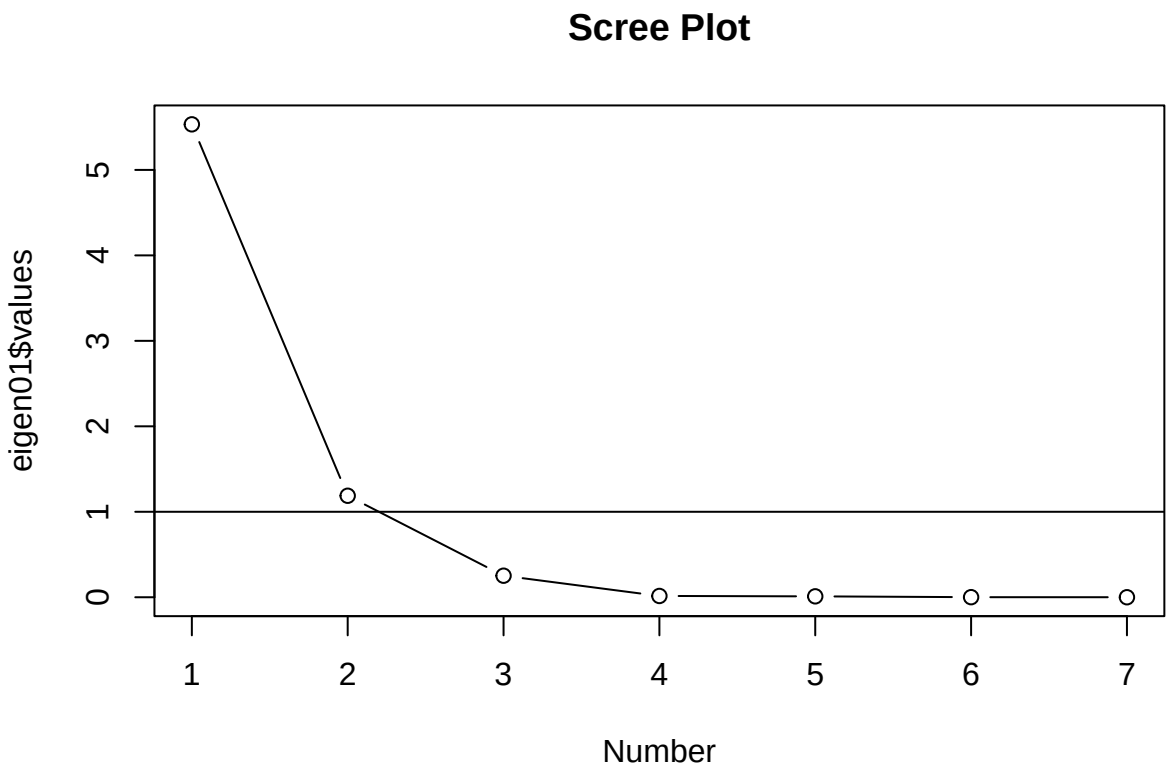
Plot of first two components

```
plot(rpca$x[,1],rpca$x[,2])
```



Selection of the number of principal components

```
cor01 <- cor(longley)
eigen01 <- eigen(cor01)
plot(eigen01$values, type = 'b', main = 'Scree Plot', xlab = 'Number')
abline(h=1, untf = FALSE)
```



Based on the elbow on scree plot and the Kaiser criterion, first two PCs are selected.

4.5 Factor analysis

```
rfa <- factanal(x=longley, factors = 2, rotation = 'promax', scores = 'Bartlett')
rfa
```

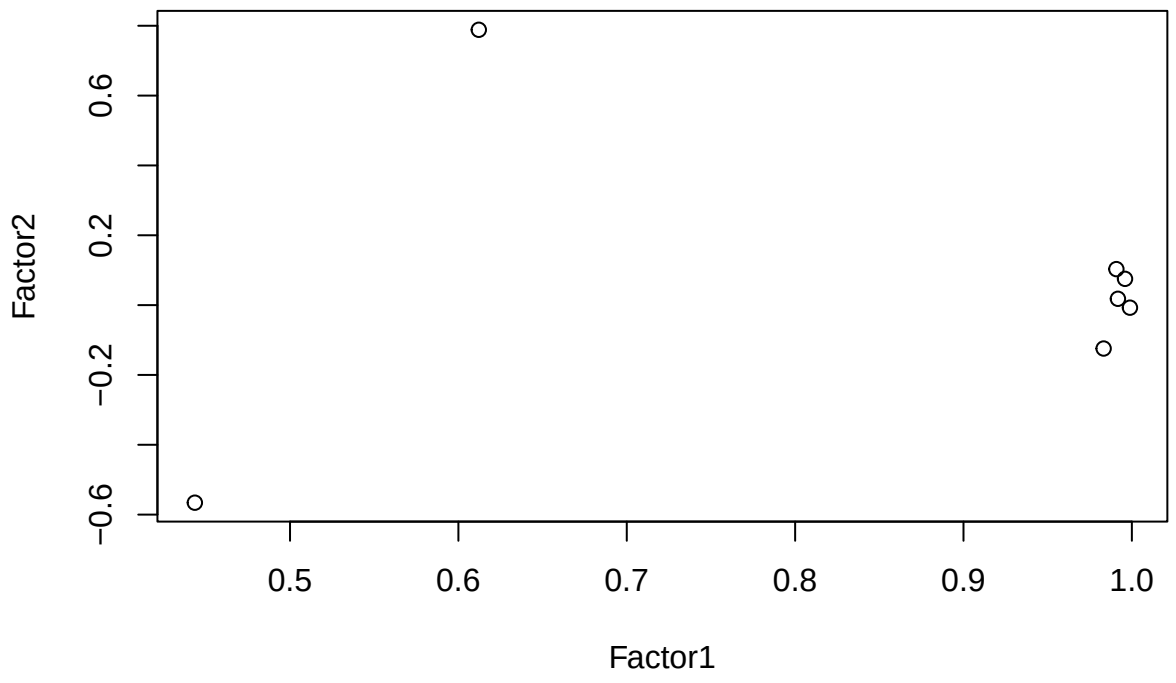
```
##
## Call:
## factanal(x = longley, factors = 2, scores = "Bartlett", rotation = "promax")
##
## Uniquenesses:
## GNP.deflator          GNP   Unemployed Armed.Forces   Population         Year
##        0.016        0.005        0.005        0.483        0.008        0.005
##     Employed
##        0.017
##
## Loadings:
##              Factor1 Factor2
## GNP.deflator  0.992
## GNP           0.999
## Unemployed    0.612   0.788
## Armed.Forces  0.443  -0.566
## Population    0.991   0.103
## Year          0.996
## Employed      0.983  -0.124
##
```

```
##                Factor1 Factor2
## SS loadings      5.493   0.974
## Proportion Var   0.785   0.139
## Cumulative Var   0.785   0.924
##
## Factor Correlations:
##          Factor1  Factor2
## Factor1  1.00000 -0.00128
## Factor2 -0.00128  1.00000
##
## Test of the hypothesis that 2 factors are sufficient.
## The chi square statistic is 41.68 on 8 degrees of freedom.
## The p-value is 1.55e-06
```

All variables are present in Factor 1. Variables that increase annually i.e, GNP,GNP deflator, Population, Year and Employed have the largest values. Only four variables form the factor 2. Unemployment and Armed forces have the largest magnitudes.

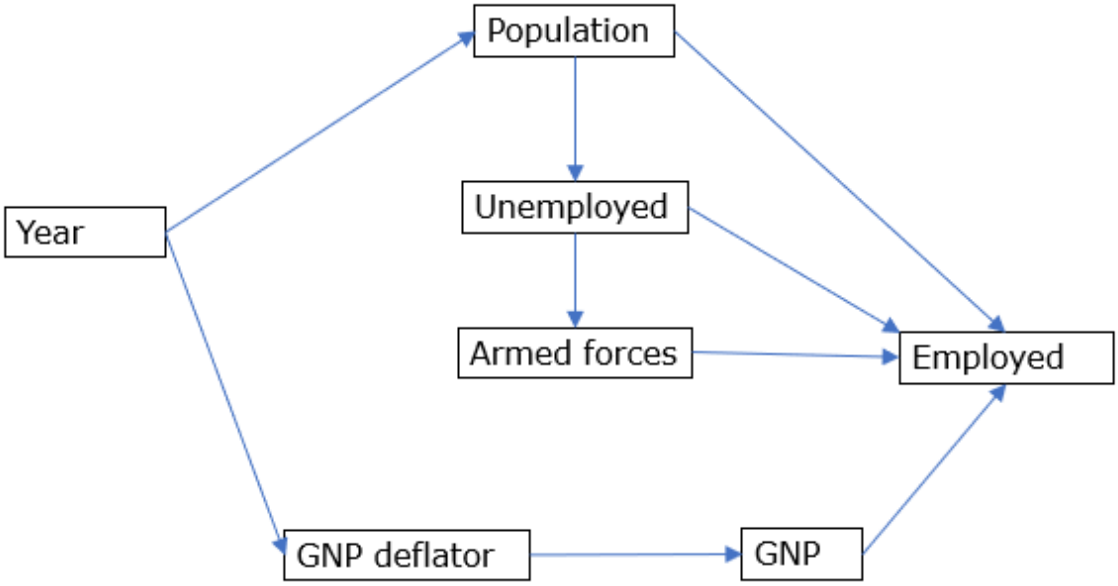
Factor Mapping

```
plot(rfa$loadings)
```



4.6 Path Analysis

Relationships described in the following figure in expected to exist among the variables in `longley` dataset.




```
shp.model <- 'Employed~Population+Unemployed+Armed.Forces+GNP
GNP~GNP.deflator
GNP.deflator~Year
Population~Year
Unemployed~Population
Armed.Forces~Unemployed'
library(lavaan)
```

```
## Warning: package 'lavaan' was built under R version 4.1.1
```

```
## This is lavaan 0.6-9
## lavaan is FREE software! Please report any bugs.
```

```
shp.fit <- sem(shp.model, data=longley)
summary(shp.fit, standardized=TRUE, rsquare=TRUE, fit.measures=TRUE)
```

```
## lavaan 0.6-9 ended normally after 109 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of model parameters                        15
##
##   Number of observations                            16
##
## Model Test User Model:
##
##   Test statistic                                92.612
```

```
##   Degrees of freedom                                12
##   P-value (Chi-square)                           0.000
##
## Model Test Baseline Model:
##
##   Test statistic                               373.800
##   Degrees of freedom                            21
##   P-value                                        0.000
##
## User Model versus Baseline Model:
##
##   Comparative Fit Index (CFI)                    0.772
##   Tucker-Lewis Index (TLI)                       0.600
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)               -295.785
##   Loglikelihood unrestricted model (H1)       -249.479
##
##   Akaike (AIC)                                 621.571
##   Bayesian (BIC)                               633.160
##   Sample-size adjusted Bayesian (BIC)          587.256
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.648
##   90 Percent confidence interval - lower         0.529
##   90 Percent confidence interval - upper         0.774
##   P-value RMSEA <= 0.05                          0.000
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.247
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Regressions:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   Employed ~
##     Population       -0.325    0.068   -4.773    0.000   -0.325   -0.598
##     Unemployed       -0.005    0.001   -3.526    0.000   -0.005   -0.128
##     Armed.Forces     -0.006    0.001   -4.081    0.000   -0.006   -0.109
##     GNP               0.062    0.005   13.292    0.000    0.062    1.632
##   GNP ~
##     GNP.deflator      9.133    0.298   30.646    0.000    9.133    0.992
##   GNP.deflator ~
##     Year              2.247    0.075   29.864    0.000    2.247    0.991
##   Population ~
##     Year              1.452    0.040   36.207    0.000    1.452    0.994
##   Unemployed ~
```

```
##     Population        9.223    2.442    3.777    0.000    9.223    0.687
##   Armed.Forces ~
##     Unemployed       -0.132    0.183   -0.721    0.471   -0.132   -0.177
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .Employed          0.148    0.052    2.828    0.005    0.148    0.011
##    .GNP             155.145   54.852    2.828    0.005  155.145    0.017
##    .GNP.deflator      1.924    0.680    2.828    0.005    1.924    0.018
##    .Population        0.547    0.193    2.828    0.005    0.547    0.012
##    .Unemployed     4327.760 1530.094    2.828    0.005 4327.760    0.529
##    .Armed.Forces   4397.440 1554.730    2.828    0.005 4397.440    0.969
##
## R-Square:
##                    Estimate
##     Employed          0.989
##     GNP               0.983
##     GNP.deflator      0.982
##     Population        0.988
##     Unemployed        0.471
##     Armed.Forces      0.031
```

Comparative Fit Index (CFI) is 0.772

Akaike (AIC) is 621.571

RMSEA is 0.648

Modification indices of the model are investigated for further improve the model.

```
modificationindices(shp.fit)
```

```
##             lhs op          rhs     mi       epc   sepc.lv sepc.all sepc.nox
## 16         Year ~~         Year  0.000     0.000     0.000    0.000    0.000
## 17     Employed ~~          GNP  0.150     0.583     0.583    0.122    0.122
## 18     Employed ~~ GNP.deflator  2.279    -0.255    -0.255   -0.479   -0.479
## 19     Employed ~~   Population  1.568    -0.104    -0.104   -0.365   -0.365
## 22          GNP ~~ GNP.deflator  8.415   -12.642   -12.642   -0.732   -0.732
## 23          GNP ~~   Population  4.791     5.041     5.041    0.547    0.547
## 24          GNP ~~   Unemployed  1.129  -217.645  -217.645   -0.266   -0.266
## 25          GNP ~~ Armed.Forces  0.265  -106.202  -106.202   -0.129   -0.129
## 26 GNP.deflator ~~   Population  2.704    -0.422    -0.422   -0.411   -0.411
## 27 GNP.deflator ~~   Unemployed  2.427   -35.542   -35.542   -0.389   -0.389
## 28 GNP.deflator ~~ Armed.Forces  1.798    30.832    30.832    0.335    0.335
## 29   Population ~~   Unemployed  0.502     8.671     8.671    0.178    0.178
## 30   Population ~~ Armed.Forces  2.959   -21.152   -21.152   -0.431   -0.431
## 31   Unemployed ~~ Armed.Forces  7.388 -4317.751 -4317.751   -0.990   -0.990
## 32     Employed  ~ GNP.deflator  0.150    -0.034    -0.034   -0.098   -0.098
## 33     Employed  ~         Year  1.568     0.276     0.276    0.347    0.075
## 34          GNP  ~     Employed  1.607    -5.522    -5.522   -0.210   -0.210
## 35          GNP  ~   Population 13.177     9.776     9.776    0.684    0.684
## 36          GNP  ~   Unemployed  0.219    -0.022    -0.022   -0.021   -0.021
## 37          GNP  ~ Armed.Forces  0.201    -0.021    -0.021   -0.015   -0.015
## 38          GNP  ~         Year  8.416    14.761    14.761    0.707    0.153
## 39 GNP.deflator  ~     Employed  5.385    -0.808    -0.808   -0.283   -0.283
```

```
## 40 GNP.deflator  ~          GNP  8.416    -0.081    -0.081   -0.751   -0.751
## 41 GNP.deflator  ~   Population  2.704    -0.771    -0.771   -0.497   -0.497
## 42 GNP.deflator  ~   Unemployed  2.956    -0.009    -0.009   -0.078   -0.078
## 43 GNP.deflator  ~ Armed.Forces  2.415     0.008     0.008    0.052    0.052
## 44   Population  ~     Employed  0.132     0.054     0.054    0.029    0.029
## 45   Population  ~          GNP  0.131     0.004     0.004    0.054    0.054
## 46   Population  ~ GNP.deflator  2.704    -0.219    -0.219   -0.340   -0.340
## 47   Population  ~   Unemployed  0.502     0.002     0.002    0.027    0.027
## 48   Population  ~ Armed.Forces  3.231    -0.005    -0.005   -0.050   -0.050
## 49   Unemployed  ~     Employed  1.841   -17.087   -17.087   -0.692   -0.692
## 50   Unemployed  ~          GNP  3.841    -1.566    -1.566   -1.666   -1.666
## 51   Unemployed  ~ GNP.deflator  2.736   -15.166   -15.166   -1.751   -1.751
## 52   Unemployed  ~ Armed.Forces  7.388    -0.982    -0.982   -0.731   -0.731
## 53   Unemployed  ~         Year  0.502   -23.023   -23.023   -1.173   -0.254
## 54 Armed.Forces  ~     Employed  9.596    17.439    17.439    0.948    0.948
## 55 Armed.Forces  ~          GNP  9.203     0.705     0.705    1.006    1.006
## 56 Armed.Forces  ~ GNP.deflator 10.062     6.833     6.833    1.060    1.060
## 57 Armed.Forces  ~   Population  7.388     9.202     9.202    0.920    0.920
## 58 Armed.Forces  ~         Year  8.876    14.658    14.658    1.003    0.218
## 59         Year  ~     Employed  0.000     0.000     0.000    0.000    0.000
## 60         Year  ~          GNP  0.000     0.000     0.000    0.000    0.000
## 61         Year  ~ GNP.deflator  0.000     0.000     0.000    0.000    0.000
## 62         Year  ~   Population  0.000     0.000     0.000    0.000    0.000
## 63         Year  ~   Unemployed  0.000     0.000     0.000    0.000    0.000
## 64         Year  ~ Armed.Forces  0.017     0.000     0.000    0.002    0.002
```

Two new paths, GNP~Population and Armed.Forces~GNP.deflation are added to the model as suggested by the high modification indices.

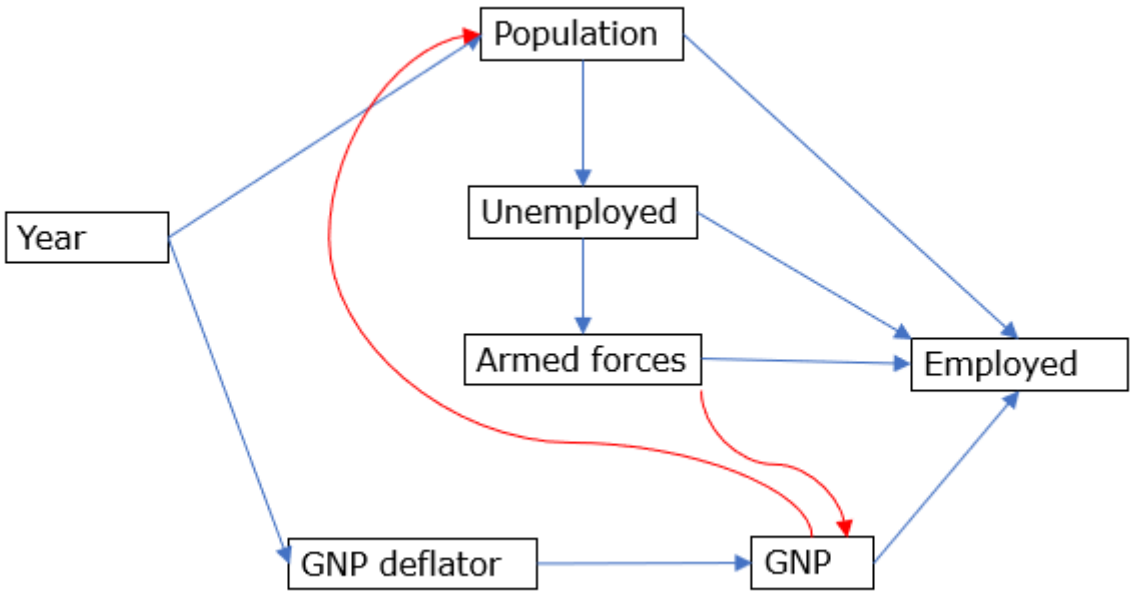


Figure 1: Modified model

```
shp.model <- 'Employed~Population+Unemployed+Armed.Forces+GNP
GNP~GNP.deflator+Population
GNP.deflator~Year
Population~Year
Unemployed~Population
Armed.Forces~Unemployed+GNP.deflator'
library(lavaan)
shp.fit <- sem(shp.model, data=longley)
summary(shp.fit, standardized=TRUE, rsquare=TRUE, fit.measures=TRUE)
```

```
## lavaan 0.6-9 ended normally after 146 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of model parameters                        17
##
##   Number of observations                            16
##
## Model Test User Model:
##
##   Test statistic                                65.452
##   Degrees of freedom                                10
##   P-value (Chi-square)                           0.000
##
## Model Test Baseline Model:
##
##   Test statistic                               373.800
##   Degrees of freedom                                21
##   P-value                                        0.000
##
## User Model versus Baseline Model:
##
##   Comparative Fit Index (CFI)                    0.843
##   Tucker-Lewis Index (TLI)                       0.670
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)               -282.205
##   Loglikelihood unrestricted model (H1)       -249.479
##
##   Akaike (AIC)                                 598.411
##   Bayesian (BIC)                               611.545
##   Sample-size adjusted Bayesian (BIC)          559.520
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.589
##   90 Percent confidence interval - lower         0.458
##   90 Percent confidence interval - upper         0.728
##   P-value RMSEA <= 0.05                          0.000
##
## Standardized Root Mean Square Residual:
##
```

```
##   SRMR                                           0.041
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Regressions:
##                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   Employed ~
##     Population       -0.325    0.119   -2.734    0.006   -0.325   -0.649
##     Unemployed       -0.005    0.002   -2.740    0.006   -0.005   -0.139
##     Armed.Forces     -0.006    0.002   -2.761    0.006   -0.006   -0.113
##     GNP               0.062    0.008    7.409    0.000    0.062    1.774
##   GNP ~
##     GNP.deflator      4.724    1.114    4.242    0.000    4.724    0.512
##     Population        6.986    1.728    4.043    0.000    6.986    0.488
##   GNP.deflator ~
##     Year              2.247    0.075   29.865    0.000    2.247    0.991
##   Population ~
##     Year              1.452    0.040   36.207    0.000    1.452    0.994
##   Unemployed ~
##     Population        9.223    2.442    3.777    0.000    9.223    0.687
##   Armed.Forces ~
##     Unemployed       -0.564    0.166   -3.400    0.001   -0.564   -0.790
##     GNP.deflator      6.030    1.437    4.196    0.000    6.030    0.974
##
## Variances:
##                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .Employed          0.148    0.052    2.828    0.005    0.148    0.013
##    .GNP              63.846   22.573    2.828    0.005   63.846    0.007
##    .GNP.deflator      1.924    0.680    2.828    0.005    1.924    0.018
##    .Population        0.547    0.193    2.828    0.005    0.547    0.012
##    .Unemployed     4327.749 1530.090    2.828    0.005 4327.749    0.529
##    .Armed.Forces   1957.004  691.905    2.828    0.005 1957.004    0.468
##
## R-Square:
##                   Estimate
##     Employed         0.987
##     GNP              0.993
##     GNP.deflator     0.982
##     Population       0.988
##     Unemployed       0.471
##     Armed.Forces     0.532
```

Comparative Fit Index (CFI) is 0.843

Akaike (AIC) is 598.411

RMSEA is 0.589

Shown by the increased fit and reduced AIC and RMSEA, the model has been improved by the modifications.