DEBRECENI EGYETEM

PRACTICAL EXERCISES FOR THE COURSE OF RESEARCH METHODS, BIOMETRY

Dr. Posta János

University of Debrecen

Faculty of Agricultural and Food Sciences, and Environmental Management

A Debreceni Egyetem fejlesztése a felsőfokú oktatás minőségének és hozzáférhetőségének együttes javítása érdekében
EFOP-3.4.3-16-2016-00021

SZÉCHENYI 2020
MAGYARORSZÁG KORMÁNYA
Európai Unió
Európai Szociális Alap
BEFEKTETÉS A JÖVŐBE

Content

Exercise 1

Design experiments, experimental units, treatments. Experimental designs.

Required knowledge: Ability to create experimental design. Students should be able to compare different experimental designs (e.g.: random design, block design and randomized block design) and create them.

a) Please show the differences between random and randomized block design as create experimental designs using both method for a research to analyse three treatments with five replicates (overall 15 units)!

b) Create an experimental design for a research with four treatments and five replicates! Show the difference between block design and randomized block design!

Exercise 2

Purpose of the exercise: Numerical description of data - Calculation of central and variance indicators

Required knowledge: Ability to compute descriptive statistics using R-software. Students should be able to compute as well as graphically (histogram, boxplot and barplot) illustrate descriptive statistics.

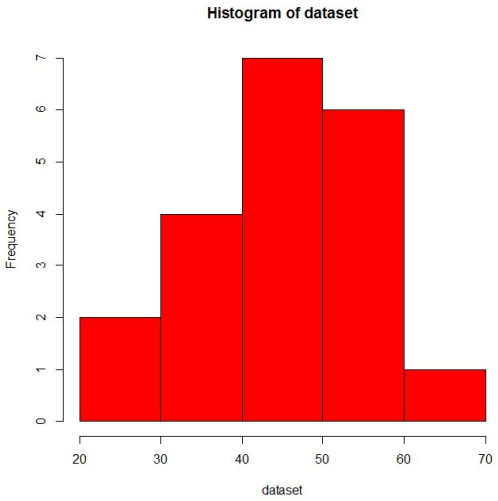
a) Please compute the descriptive statistics of the data given below:

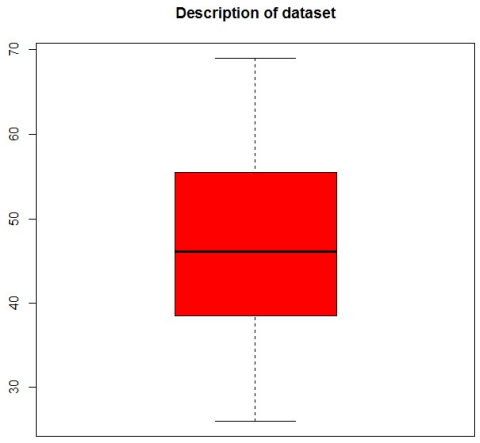
54	59	39	33	46	26	47	60	59	46
49	46	41	44	26	36	38	51	57	69

Please describe it also graphically!

Solution:

```
> dataset <– c(54,59,39,33,46,26,47,60,59,46,49,46,41,44,26,36,38,51,57,69)
> summary(dataset)
   Min.   1st Qu.  Median  Mean    3rd Qu   Max.
  26.00   38.75    46.00    46.30    54.75    69.00
> hist(dataset, col="red")
> boxplot(dataset, col="red", main="Description of dataset")
> sd(dataset)/mean(dataset)          #Compute coefficient of variation (CV%)
[1] 0.2473624
>
```





b) Compute descriptive statistics of geese conformation parameters and also show it graphically using the data (http://alfa.date.hu/posta/dataset.csv)!

c) Compute descriptive statistics of Hungarian Frizzled geese conformation parameters and also show the frequencies graphically using the data (http://alfa.date.hu/posta/dataset4.csv)!

d) Daily gain of some sheep breeds was analysed. Number of lambing was also given. Please calculate descriptive statistics of daily gain group by breed and also for lambing order! Graphical interpretation of these statistics is also important. The measurement data can be downloaded from: http://alfa.date.hu/posta/lamb.csv.

e) Describe the turkey population graphically using descriptive statistics! Please describe the overall sample and after take into account the genotype of the birds.

Genotype	Br	Bl	Br	Bl	Bl	Br	C	Br	Br	Bl	Bl	Br	Bl	Bl
Weight	850	750	820	730	730	1050	1070	1280	970	1060	800	880	770	1070

Genotype	Br	R	R	Bl	C	C	Br	Bl	Bl	Br	Br	C	Br	C
Weight	840	820	690	960	830	1220	980	990	840	840	1040	1250	950	1180

(Abbreviations: Br: bronze, Bl: black, C: copper.)

f) There are some weight of pig below:

105 110 108 107 108 106 107

Please compute the variance and standard deviation of these data! Also give information about the heterogeneity of the sample! (CV%)

g) Compute average body weight of geese (http://alfa.date.hu/posta/dataset.csv) by genotype and show in a barplot figure! Illustrate the descriptive statistics of body weight of geese also graphically!

h) Please calculate descriptive statistics of conformation traits (height at withers, hearth girth, cannon bone circumference) of mare horses were measured during self performance tests! Show the frequencies of the traits graphically! The measurement data can be downloaded from: http://alfa.date.hu/posta/horses.csv.

i) Please calculate the descriptive statistics for conformation traits (length of breast and tight, weight) of 25 weeks old Bronze Turkeys grouping by sex. The measurement data can be downloaded from: http://alfa.date.hu/posta/bronzeturkey.csv.

j) Please calculate the descriptive statistics for egg parameters (egg weight, height, width, yolk weight, white weight, yolk colour, eggshell thickness) of Hungarian Chickens. The measurement data can be downloaded from: http://alfa.date.hu/posta/hen.csv.

Exercise 3

Purpose of the exercise: Comparison of means - Learn how to perform the t-test.
Required knowledge: Students need to make statistical comparisons of two populations or a sample to a given value. They have to be able to make a correct null hypothesis and to choose the right method (e.g. in case of two independent samples they need to perform an F-test to decide whether t-test or Welch test should be used).

a) There are weights of two turkey breeds within the table below. Is there any difference between the two breeds?

Breed	Weight (kg)				
Bronze Turkey	5.00	4.60	4.89	5.44	5.00
Copper Turkey	5.80	4.40	5.73	5.13	5.00

```
Solution:
# Two independent samples
> turkeyW <– c(5.00,4.60,4.89,5.44,5.00,5.80,4.40,5.73,5.13,5.00)
> turkeyG <– c("Bronze","Bronze","Bronze","Bronze","Bronze","Copper","Copper","Copper","Copper","Copper")
> var.test(turkeyW~turkeyG)
        F test to compare two variances
data:  turkeyW by turkeyG
F = 0.2752, num df = 4, denom df = 4, p-value = 0.2393
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.0286577 2.6435869
sample estimates:
ratio of variances
         0.2752438
# P-value is higher then 0.05, so no significant difference between the variances, so independent samples t-test should
be used.
> t.test(turkeyW~turkeyG, alternative="two.sided", var.equal=T)
        Two Sample t-test
data:  turkeyW by turkeyG
t = -0.7775, df = 8, p-value = 0.4592
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.89629  0.44429
sample estimates:
mean in group Bronze mean in group Copper
               4.986               5.212
>
# P-value is higher then 0.05, so no significant difference between the two breeds.
```

b) A machine loads automatically the boxes in a feed company. The expected weight of the boxes is 500gram. Please check the sample below that it meets with the expectation or not?

483 502 498 496 502 483 494 491 505 486

```
Solution:
# One sample t-test
> feed <– c(483,502,498,496,502,483,494,491,505,486)
> t.test(feed, mu=500, alternative="two.sided")
        One Sample t-test
data:  feed
t = -2.3554, df = 9, p-value = 0.04292
alternative hypothesis: true mean is not equal to 500
95 percent confidence interval:
 488.2375 499.7625
sample estimates:
mean of x
      494
```

```
>
# P is lower than 0.05, so the sample is significantly different from the expected (500) value.
```

c) The same turkey's weights were measured in 23rd and 25th weeks of age. Is there any difference between the weights of the birds between the two ages? The measurement data can be downloaded from: http://alfa.date.hu/posta/turkeys.csv.

```
Solution:
#Paired samples t-test.
> turkey <– read.table("http://alfa.date.hu/posta/turkeys.csv", header=T, sep=";")
> t.test(turkey$w23,turkey$w25, paired=T, alternative="two.sided")
        Paired t-test
data:  turkey$w23 and turkey$w25
t = -2.1145, df = 200, p-value = 0.03571
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -6.2688327 -0.2187295
sample estimates:
mean of the differences
             -3.243781
>
# P is lower than 0.05, so there is significant difference between the weights.
```

d) There are two different types of lighting traps to catch butterflies. Please compare the two types of traps!

A	41	34	33	36	40	25	31	37	34	30	38
B	52	57	62	55	64	57	56	55			

e) Is there any effect of sex of turkeys on weights measured in 25th weeks of age? The measurement data can be downloaded from: http://alfa.date.hu/posta/turkeys2.csv.

f) Bronze Turkeys' conformation traits (length of breast and tight, weight) were measured as 25 weeks old Please check the hypothesis, that males performance is higher than those of females!

The measurement data can be downloaded from: http://alfa.date.hu/posta/bronzeturkey.csv.

g) The same Hungarian Sporthorse mares were tested as 3- and 4-years-old. Please check are there any difference within the movement traits (walk, trot, canter) evaluated in different

ages, respectively! The measurement data can be downloaded from: http://alfa.date.hu/posta/movement.csv.

h) The same Hungarian Sporthorse mares were tested as 3- and 4-years-old. Scores of free jumping traits (jumping style, jumping ability and observation during training) received as 4-years-old should exceed those of received as 3-years-old. Please check this hypothesis! The measurement data can be downloaded from: http://alfa.date.hu/posta/jumping.csv.

Exercise 4

Purpose of the exercise: One-way analysis of variance.

Required knowledge: Students need to make correct conclusions in case of having more than two measurement groups. They need to be able to make correct null hypothesis and statistical conclusion. In case of significant difference among groups, they also need to perform post-hoc tests (e.g. Tukey-test, or LSD test).

a) There were several geese genotypes compared within an experiment. Is there any difference in beak length among the genotypes and if any please sign the differences among groups! The measurement data can be downloaded from: http://alfa.date.hu/posta/dataset.csv.

Solution:

```
> geese <– read.table("http://alfa.date.hu/posta/dataset.csv", sep=";", header=T)
> AovRes <– aov(BeakLength~Genotype, data=geese)
> summary(AovRes)
             Df Sum Sq Mean Sq F value Pr(>F)
Genotype      2   1709   854.4   50.43 <2e-16 ***
Residuals   200   3389    16.9
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
# P is smaller then 0.05, so there is difference among genotypes, post-hoc test (e.g. Tukey test) is needed.
> TukeyHSD(AovRes)
  Tukey multiple comparisons of means
    95% family-wise confidence level
Fit: aov(formula = BeakLength ~ Genotype, data = geese)

$Genotype
                          diff        lwr        upr     p adj
Oroshaza-Frizzled   12.7491905  9.712444  15.785937 0.0000000
Parlag-Frizzled      0.9192456 -0.540295  2.378786 0.2992271
```

Parlag-Oroshaza	-11.8299449	-14.771545	-8.888344	0.0000000

>
There is difference between “Oroshaza” and “Frizzled” as well as “Parlag” and “Oroshaza” groups, as P value smaller than 0.05.

b) Different colour varieties of Hungarian Frizzled Geese were compared within an experiment. Is there any difference in beak length and body weight among the genotypes? If any, please sign the differences among groups! The measurement data can be downloaded from: http://alfa.date.hu/posta/dataset4.csv.

c) Three goat sires were compared based on their daughters lactation milk yield (LMY) Is there any difference among the sires based on their daughters’ milking performance? If any, please sign the differences among groups! The measurement data can be downloaded from: http://alfa.date.hu/posta/goat.csv.

d) Charolais fattening bulls were weighted in different ages. Please check whether there is any difference among leptin (LEP) genotypes on weight in different ages? If any, please sign the differences among groups! The measurement data can be downloaded from: http://alfa.date.hu/posta/charolais.csv.

Exercise 5

Purpose of the exercise: Multi-factor variance analysis.
Required knowledge: Students need to able to develop models with at least two influencing factors. They also need to be able to check interaction between effects.

a) Some conformation traits (height at withers, hearth girth, cannon bone circumference) of mare horses were measured during self performance tests. Owners and birth years of mares is known. Please check that owners and/or birth year have a significant effect on conformation traits! The measurement data can be downloaded from: http://alfa.date.hu/posta/horses.csv.

Solution:

```
> horses <– read.table("http://alfa.date.hu/posta/horses.csv", header=T, sep=";")
> AovHorses <- aov(Height~BirthYear + Owner, data=horses)
> summary(AovHorses)
                Df Sum Sq  Mean Sq  F value  Pr(>F)
BirthYear       39  2652   68.01    3.538    2.03e-12 ***
Owner           1154 36249 31.41    1.634    < 2e-16 ***
Residuals       1877 36082 19.22
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
9 observations deleted due to missingness
>
# There is significant effect of birth year and owner of mares on height at withers.
```

b) Charolais fattening bulls were weighted in different ages. Please check whether there is any effect of Stat5A and PSMC1 genes on weight in different ages? If any, please sign the differences among groups! The measurement data can be downloaded from: http://alfa.date.hu/posta/charolais2.csv.

c) Several sheep genotypes were analysed for parasite infection using FAMACHA test. Please check the possible effect of genotype and sex on parasite infection! The measurement data can be downloaded from: http://alfa.date.hu/posta/sheep_parasite.csv.

d) Daily gain of some sheep breeds was analysed. Number of lambing was also given. Is there any difference in daily gain among the breed and lambing order? If any please sign the differences among groups! The measurement data can be downloaded from: http://alfa.date.hu/posta/lamb.csv.

Exercise 6

Purpose of the exercise: Correlation analysis.

Required knowledge: Explanation of correlation, know the different computation methods of correlations depending on the type of the numeric data.

a) Please compute the correlation between the body height and bodyweight for geese (http://alfa.date.hu/posta/dataset.csv)! Please also compute it for beak length and beak width!

```
Solution:
> geese <– read.table("http://alfa.date.hu/posta/dataset.csv", sep=";", header=T)
> cor.test(geese$BeakLength,geese$BeakWidth, method="pearson")
        Pearson's product-moment correlation
data:  geese$BeakLength and geese$BeakWidth
t = 16.9683, df = 201, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.7040901 0.8185960
sample estimates:
      cor
0.7673934
>
# P value is smaller than 0.05, so the estimated correlation value (0.77) is significantly different from 0.
```

b) Please compute the correlation between conformation traits (height at withers, hearth girth, cannon bone circumference) of mare horses were measured during self performance tests!
The measurement data can be downloaded from: http://alfa.date.hu/posta/horses.csv.

c) Please compute the correlation between conformation traits (length of breast and tight, weight) of 25 weeks old Bronze Turkeys.
The measurement data can be downloaded from: http://alfa.date.hu/posta/bronzeturkey.csv.

d) Different colour varieties of Hungarian Frizzled Geese were measured within an experiment. Please estimate the correlations between beak length, beak width, leg length, tight length, breast, body height and weight!
The measurement data can be downloaded from: http://alfa.date.hu/posta/dataset4.csv.

e) Charolais fattening bulls were weighted in different ages. Please estimate the correlation among the weights measured in different ages!
The measurement data can be downloaded from: http://alfa.date.hu/posta/charolais.csv.

f) Hungarian Sporthorse mares were tested as 3-years-old. Estimate the correlation among free jumping traits (jumping style, jumping ability and observation during training)!
The measurement data can be downloaded from: http://alfa.date.hu/posta/mares3y.csv.

g) Hungarian Sporthorse mares were tested as 4-years-old. Estimate the correlation among movement analysis traits (walk, trot, canter)!
The measurement data can be downloaded from: http://alfa.date.hu/posta/mares4y.csv.

h) Please estimate the correlation among egg parameters (egg weight, height, width, yolk weight, white weight, yolk colour, eggshell thickness) of Hungarian Chickens. The measurement data can be downloaded from: http://alfa.date.hu/posta/hen.csv.

Exercise 7

Purpose of the exercise: Regression analysis.
Required knowledge: Explanation of regression. Students need to be familiar regression models and be able to develop and check them.

a) Please estimate a regression equation to predict body weight from breast length using Hungarian geese data (http://alfa.date.hu/posta/dataset.csv)!

```
Solution:
> geese <– read.table("http://alfa.date.hu/posta/dataset.csv", sep=";", header=T)
> LmRes <– lm(Weight~Breast, data=geese)
> summary(LmRes)
Call:
lm(formula = Weight ~ Breast, data = geese)
Residuals:
     Min       1Q   Median       3Q      Max
-1.58634 -0.46380 -0.06057  0.37678  2.91322
Coefficients:
             Estimate  Std. Error  t value  Pr(>|t|)
(Intercept) -5.429922   0.551593   -9.844   <2e-16 ***
Breast       0.055991   0.003083   18.163   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 0.7134 on 201 degrees of freedom
Multiple R-squared:  0.6214,    Adjusted R-squared:  0.6195
F-statistic: 329.9 on 1 and 201 DF,  p-value: < 2.2e-16
>
# Weight = -5.429922 + 0.055991 * Breast
```

b) Charolais fattening bulls were weighted in different ages. Please create a regression equation to estimate 350-d-weight based on weights measured at previous ages!
The measurement data can be downloaded from: http://alfa.date.hu/posta/charolais.csv.

c) Please develop a regression equation to estimate weight based on length of breast and length of tight of 25 weeks old Bronze Turkeys!
The measurement data can be downloaded from: http://alfa.date.hu/posta/bronzeturkey.csv.

d) Please develop a regression equation to estimate egg weight from height and width of the egg of Hungarian Chickens. The measurement data can be downloaded from: http://alfa.date.hu/posta/hen.csv.

Exercise 8

Purpose of the exercise: Non-parametric tests - Fitting, Homogeneity, and Independence Test with $\chi 2$.
Required knowledge: Ability to analyse quality (categorical) traits. Students also need to be able to discover relationship between quality traits.

a) There are collected results of dog cancer. There were 122 dogs int he analysis, three different types of cancer were identified.
Is there any relationship between cancer type and sex of the dogs?

	A	B	C
Male	40	22	10
Female	22	30	15

Solution:

```
> dogs <- matrix(c(40,22,22,30,10,15), nrow=2)
> chisq.test(dogs)
        Pearson's Chi-squared test
data:  dogs
X-squared = 7.2861, df = 2, p-value = 0.02617
>
# The p-value is smaller than 0.05, so there is relationship between sex and cancer type of dogs.
```

b) There are some known Mendelian distributions (9:6:1, 9:3:4, 12:3:1) for dihybrid inheritance. The colour of onion shell is influenced by 2 genes. Homozygote reds and whites were crossed and the first generation (F_1) was also crossed among each other. The distribution was 110 reds, 35 yellows and 45 whites in the second (F_2) generation. To which Mendelian distribution can be fitted this distribution?

```
Solution:
> onion <– c(110,35,45)
> exp1 <– c(9,6,1)
> exp2 <– c(9,3,4)
> exp3 <– c(12,3,1)
> chisq.test(onion, p=exp1, rescale.p=T)
        Chi-squared test for given probabilities
data:  onion
X-squared = 110.9357, df = 2, p-value < 2.2e-16
> chisq.test(onion, p=exp2, rescale.p=T)
        Chi-squared test for given probabilities
data:  onion
X-squared = 0.2339, df = 2, p-value = 0.8896
> chisq.test(onion, p=exp3, rescale.p=T)
        Chi-squared test for given probabilities
data:  onion
X-squared = 99.8246, df = 2, p-value < 2.2e-16
# P is smaller in case of 1st and 3rd expected distribution. The results are fitting to the 2nd (9:3:4) distribution.
```

c) Distribution of horn colours of Hungarian Grey cattle were observed in a herd. Is there any difference in the distribution of horn colours among sexes?

	White	'Cardy'	Green
Males	112	46	23
Females	256	138	37
Steers	33	23	2

(Radácsi, 2008)

d) Distribution of horn colours of within 'Cardy' colour group of Hungarian Grey cattle were observed in a herd. Is there any difference in the distribution of horn colour subgroups between females and steers?

Sub-group within 'Cardy' colour	Females	Steers
WRC	35	7
CMW	54	6
CSW	40	9
GSW	9	1

(Abbreviations: WRC: white, rare cardy, CMW: cardy with much white, CSW: cardy with small white, GSW: green with small white.)

(Radácsi, 2008)

e) Eventing competition results of Kisber Halfblood horses were analysed. Is there difference in the distribution of participating horses and number of starts?

	Number of horses	Number of starts
Geldings	23	102
Females	27	159
Males	31	140

f) Using the previous dataset, please check that distribution of participating Kisber Halfblood horses on eventing sport competitions grouped by sex are equally distributed or not?

Exercise 9

Purpose of the exercise: Principal component analysis.

Required knowledge: Importance of possible way to reduce number of evaluated traits in case of to many. Discover background effects modifying the measured traits.

a) Hungarian Sporthorse mares were tested as 3-years-old. Try to decrease the number of analyzed traits and discover background effects influencing performance! The measurement data can be downloaded from: http://alfa.date.hu/posta/mares3y.csv.

Solution:

```
> Mare3y <- read.table("http://alfa.date.hu/posta/mares3y.csv", header=T, sep=";")
> prcomp(Mare3y, scale=TRUE)
Standard deviations:
 [1] 2.5661498 1.5120696 1.2730520 1.2522320 0.8854254 0.8526937 0.8304374
 [8] 0.7771210 0.7276714 0.6740512 0.6463261 0.5791523 0.5740375 0.5457421
[15] 0.4936565 0.4394677 0.4151955 0.4019841
Rotation:
                PC1         PC2         PC3         PC4         PC5
Type    -0.01663606  0.57644149 -0.15868346  0.189950714 -0.12328452
Head     0.09867797  0.37733175 -0.18446842  0.238410306  0.26055678
Neck     0.29068874 -0.13191997 -0.20274900  0.059924955  0.08121094
Saddle   0.26145970 -0.03932158 -0.26551795  0.097572221  0.31534459
Frame    0.17674267  0.38833151 -0.25095734  0.147183910 -0.08349585
```

```
Forelimbs     0.27419645 -0.25032922 -0.22482879  0.079271108  0.06016137
Hindlimbs     0.26516682 -0.23503451 -0.22541079  0.012825808 -0.06135161
RegularityMovement  0.26864948 -0.27534494 -0.10711702  0.004963883 -0.10567110
ImpulsionMovement   0.30164261 -0.12154859  0.01366762  0.056372171 -0.42428065
OverallImpression     0.31069636  0.11189050 -0.21777439  0.113210811 -0.24560596
JumpingStyle   0.15770228  0.21213331 -0.11054144 -0.563924949  0.18498140
JumpingAbility      0.22381320  0.14541926 -0.02998963 -0.558525736  0.04852075
ObsTraining  0.16522218  0.12817075  0.11322800 -0.389276598 -0.36523206
Walk        0.22864056 -0.08822305  0.25480936  0.137428881  0.18718897
Trot         0.20794451  0.14153245  0.39357897  0.175853847 -0.35038807
Canter       0.24649355  0.12803496  0.39486837  0.070149986  0.01011207
OverallImprMove     0.29443030  0.04537244  0.33511157  0.111738920  0.21852454
TestScore   0.23552115  0.07569562  0.30293152 -0.033477691  0.41156090
> summary(prcomp(Mare3y, scale=TRUE))
Importance of components:
                          PC1    PC2     PC3     PC4     PC5     PC6     PC7
Standard deviation     2.5661 1.5121 1.27305 1.25223 0.88543 0.85269 0.83044
Proportion of Variance 0.3658 0.1270 0.09004 0.08712 0.04355 0.04039 0.03831
Cumulative Proportion  0.3658 0.4929 0.58290 0.67001 0.71357 0.75396 0.79227
                          PC8     PC9    PC10    PC11    PC12    PC13    PC14
Standard deviation     0.77712 0.72767 0.67405 0.64633 0.57915 0.57404 0.54574
Proportion of Variance 0.03355 0.02942 0.02524 0.02321 0.01863 0.01831 0.01655
Cumulative Proportion  0.82582 0.85524 0.88048 0.90369 0.92232 0.94063 0.95718
                          PC15    PC16    PC17    PC18
Standard deviation     0.49366 0.43947 0.41520 0.40198
Proportion of Variance 0.01354 0.01073 0.00958 0.00898
Cumulative Proportion  0.97072 0.98145 0.99102 1.00000
>
```

b) Hungarian Sporthorse mares were tested as 4-years-old. Try to decrease the number of analysed traits and discover background effects influencing performance! The measurement data can be downloaded from: http://alfa.date.hu/posta/mares4y.csv.

c) Egg parameters (egg weight, height, width, yolk weight, white weight, yolk colour, eggshell thickness) of Hungarian Chickens were collected. Try to decrease the number of analysed traits and discover background effects influencing performance! The measurement data can be downloaded from: http://alfa.date.hu/posta/hen.csv.