Lab work 4

Team: Arsenii Kazymyr, Serhii Matsyshyn, Teodor Muzychuk

```
require(BSDA)
```

```
## Loading required package: BSDA
```

```
## Warning: package 'BSDA' was built under R version 4.2.2
```

```
## Loading required package: lattice
```

```
##
## Attaching package: 'BSDA'
```

```
## The following object is masked from 'package:datasets':
##
##     Orange
```

```
library(BSDA)
require(EnvStats)
```

```
## Loading required package: EnvStats
```

```
## Warning: package 'EnvStats' was built under R version 4.2.2
```

```
##
## Attaching package: 'EnvStats'
```

```
## The following objects are masked from 'package:stats':
##
##     predict, predict.lm
```

```
## The following object is masked from 'package:base':
##
##     print.default
```

```
library(EnvStats)
```

Data generation

```
n<-10
x_k<-seq(1,100)
y_k<-seq(101, 150)
x_k=qnorm((x_k*log(x_k^2*n+pi))%%1)
y_k=qnorm((y_k*log(y_k^2*n+pi))%%1)
```

Problem 1

$H_0 : \mu_1 = \mu_2 \quad \text{vs} \quad H_1 : \mu_1 ! = \mu_2 \quad \sigma_1^2 = \sigma_2^2 = 1$

```
z.test(x_k, y_k, alternative = "t", sigma.x =1, sigma.y =1)
```

```
##
##  Two-sample z-Test
##
## data:  x_k and y_k
## z = -0.38769, p-value = 0.6982
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.4066258  0.2723257
## sample estimates:
##     mean of x     mean of y
## -6.705019e-02  9.986778e-05
```

The results of the two-sample z-test indicate that there is not sufficient evidence to reject the null hypothesis that the means of the two populations, represented by the samples x_k and y_k, are equal. This is supported by the p-value of 0.6982, which is greater than the commonly used significance level of 0.05. This means that there is a 69.82% chance that the observed difference in means between the two samples could have occurred by random chance, and therefore it is not statistically significant.

Additionally, the 95% confidence interval for the difference in means, which is [-0.4066258, 0.2723257], includes 0, further supporting the conclusion that the observed difference in means is not statistically significant. This means that we can be 95% confident that the true difference in means between the two populations lies within this interval.

Overall, these results suggest that there is not a significant difference in means between the two populations represented by the samples x_k and y_k. It is important to note, however, that these conclusions are based on the assumption that the variances of the two populations are equal, as stated in the null hypothesis. If this assumption is not met, the results of the test may not be valid. It may be necessary to use a different statistical test, such as a Welch's t-test, which does not assume equal variances.

Problem 2

$H_0, : \sigma_1 = \sigma_2 \quad \text{vs} \quad H_1, : \sigma_1 > \sigma_2 \quad \mu_1 \, and \, \mu_2 \, are \, unknow$

```
var.test(x_k, y_k, alternative = "g")
```

```
##
##  F test to compare two variances
##
## data:  x_k and y_k
## F = 0.65847, num df = 99, denom df = 49, p-value = 0.9598
## alternative hypothesis: true ratio of variances is greater than 1
## 95 percent confidence interval:
##  0.4300888       Inf
## sample estimates:
## ratio of variances
##          0.6584653
```

The output of the var.test() function in R indicates that the null hypothesis (H0) that the variances of the two groups (x_k and y_k) are equal cannot be rejected. This is based on the p-value of 0.9598, which is greater than the commonly used significance level of 0.05. This means that there is not enough evidence to conclude that the variances are different.

The F statistic and the confidence interval can also be used to evaluate the result. The F statistic is calculated as the ratio of the variances of the two groups, and in this case it is 0.65847. The confidence interval, on the other hand, gives an estimate of the range of values that the true ratio of variances is likely to fall within. In this case, the 95% confidence interval ranges from 0.4300888 to infinity, which includes the value of 1 (which would indicate equal variances). This further supports the conclusion that the variances are not significantly different.

In conclusion, based on the output of the var.test() function, it can be concluded that there is not enough evidence to reject the null hypothesis that the variances of the two groups (x_k and y_k) are equal. This means that the data do not provide strong support for the alternative hypothesis that the variances are different.

Problem 3

A) $\{x_k\}_{k=1}^{100}$ are normally distributed (with parameters calculated from the sample)

```
fractional_x_k <- x_k - trunc(x_k)
fractional_y_k <- y_k - trunc(y_k)

mu_calculated <- mean(fractional_x_k)
sigma_calculated <- sd(fractional_x_k)

ks.test(fractional_x_k, "pnorm", mean=mu_calculated, sd=sigma_calculated)
```

```
##
##  Asymptotic one-sample Kolmogorov-Smirnov test
##
## data:  fractional_x_k
## D = 0.056059, p-value = 0.9118
## alternative hypothesis: two-sided
```

The output of the ks.test() function in R indicates that the null hypothesis (H0) that the data in the group $\{x_k\}_{k=1}^{100}$ are normally distributed cannot be rejected. This is based on the p-value of 0.9118, which is greater than the commonly used significance level of 0.05. This means that there is not enough evidence to conclude that the data are not normally distributed.

The D statistic and the alternative hypothesis can also be used to evaluate the result. The D statistic is a measure of the maximum distance between the empirical cumulative distribution function (ECDF) of the data and the theoretical cumulative distribution function (CDF) of the normal distribution. A small D value indicates that the data are well-described by the normal distribution. In this case, the D value is 0.056059, which is relatively small and suggests that the data are consistent with the normal distribution. The alternative hypothesis is “two-sided”, which means that the test is looking for evidence that the data are not normally distributed in either direction (i.e., either higher or lower than the normal distribution).

In conclusion, based on the output of the ks.test() function, it can be concluded that there is not enough evidence to reject the null hypothesis that the data in the group $\{x_k\}_{k=1}^{100}$ are normally distributed. This means that the data do not provide strong support for the alternative hypothesis that the data are not normally distributed.

B) $\{|x_k|\}_{k=1}^{100}$ are exponentially distributed with λ = 1

```
ks.test(abs(fractional_x_k), "pexp", rate=1)
```

```
##
##  Asymptotic one-sample Kolmogorov-Smirnov test
##
## data:  abs(fractional_x_k)
## D = 0.36972, p-value = 2.679e-12
## alternative hypothesis: two-sided
```

The output of the ks.test() function in R indicates that the null hypothesis (H0) that the data in the group $\{|x_k|\}_{k=1}^{100}$ are exponentially distributed with a rate parameter of 1 can be rejected. This is based on the p-value of 2.679e-12, which is significantly smaller than the commonly used significance level of 0.05. This means that there is strong evidence to conclude that the data are not exponentially distributed with a rate of 1.

The D statistic and the alternative hypothesis can also be used to evaluate the result. The D statistic is a measure of the maximum distance between the empirical cumulative distribution function (ECDF) of the data and the theoretical cumulative distribution function (CDF) of the exponential distribution. A large D value indicates that the data are not well-described by the exponential distribution. In this case, the D value is 0.36972, which is relatively large and suggests that the data are not consistent with the exponential distribution. The alternative hypothesis is “two-sided”, which means that the test is looking for evidence that the data are not exponentially distributed in either direction (i.e., either higher or lower than the exponential distribution).

In conclusion, based on the output of the ks.test() function, it can be concluded that there is strong evidence to reject the null hypothesis that the data in the group $\{|x_k|\}_{k=1}^{100}$ are exponentially distributed with a rate of 1. This means that the data provide strong support for the alternative hypothesis that the data are not exponentially distributed.

C) $\{x_k\}_{k=1}^{100}$ and $\{y_k\}_{k=1}^{100}$ have the same distributions.

```
ks.test(fractional_x_k, fractional_y_k)
```

```
##
##  Exact two-sample Kolmogorov-Smirnov test
##
## data:  fractional_x_k and fractional_y_k
## D = 0.12, p-value = 0.7112
## alternative hypothesis: two-sided
```

The output of the ks.test() function in R indicates that the null hypothesis (H0) that the data in the groups $\{x_k\}_{k=1}^{100}$ and $\{y_k\}_{k=1}^{100}$ come from the same distribution cannot be rejected. This is based on the p-value of 0.7112, which is greater than the commonly used significance level of 0.05. This means that there is not enough evidence to conclude that the data come from different distributions.

The D statistic and the alternative hypothesis can also be used to evaluate the result. The D statistic is a measure of the maximum distance between the empirical cumulative distribution functions (ECDFs) of the two groups of data. A small D value indicates that the data are well-described by the same distribution. In this case, the D value is 0.12, which is relatively small and suggests that the data are consistent with coming from the same distribution. The alternative hypothesis is “two-sided”, which means that the test is looking for evidence that the data come from different distributions in either direction (i.e., either higher or lower than the other group).

In conclusion, based on the output of the ks.test() function, it can be concluded that there is not enough evidence to reject the null hypothesis that the data in the groups $\{x_k\}_{k=1}^{100}$ and $\{y_k\}_{k=1}^{100}$ come from the same distribution. This means that the data do not provide strong support for the alternative hypothesis that the data come from different distributions.