Part VII

GOODNESS-OF-FIT TESTS

The aim of *goodness-of-fit* tests is to test if a sample originates from a specific distribution. One of the oldest tests is the χ^2 goodness-of-fit test (Test 12.2.1). The principle of this test is to divide the sample into classes and compare observed and expected values under the null distribution. The test is suitable for continuous and discrete distributions. However, due to dividing the sample into arbitrary classes the test is not very powerful when testing for a continuous distribution; in this case tests are to be preferred which are customized to specific distributions.

In Chapter 11 we present tests on normality with respect to the outstanding nature of this distribution. Chapter 12 deals with goodness-of-fit tests on distributions other than normal. Most of the tests can be adapted to both cases.

A rough classification of these tests gives two types of goodness-of-fit tests. The first type are tests which employ the empirical distribution function (EDF). Here, the EDF is compared with the theoretical distribution function of the null distribution. One of the famous tests in this class is the Kolmogorov–Smirnov test. The second type are not based on the EDF but compare observed with expected values, such as the above-mentioned χ^2-test.

Statistical Hypothesis Testing with SAS and R, First Edition. Dirk Taeger and Sonja Kuhnt.

11

Tests on normality

In this chapter we present goodness-of-fit tests for the Gaussian distribution. In Section 11.1 tests based on the empirical distribution function (EDF) are treated. A good resource for this kind of test is Stephens (1986). We start with the Kolmogorov–Smirnov test. It evaluates the greatest vertical distance between the EDF and the theoretical cumulative distribution function (CDF). If both, or one parameter are estimated from the sample the distribution of the test statistic changes and the test is called the Lilliefors test on normality.

Section 11.2 deals with tests not based on the EDF such as the Jarque–Bera test which compares observed and expected moments of the normal distribution.

11.1 Tests based on the EDF

11.1.1 Kolmogorov–Smirnov test (Lilliefors test for normality)

Description: Tests if a sample is sampled from a normal distribution with parameter μ and σ^2.

Assumptions:
- Data are measured at least on an ordinal scale.
- The sample random variables $X_1, \ldots, X_n$ are identically, independently distributed with observations $x_1, \ldots, x_n$ and a continuous distribution function $F(x)$.

Hypotheses:

(A) $H_0 : F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) \quad \forall x$ vs
$H_1 : F(x) \neq \Phi\left(\frac{x-\mu}{\sigma}\right)$ for at least one x

(B) $H_0 : F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) \quad \forall x$ vs
$H_1 : F(x) \geq \Phi\left(\frac{x-\mu}{\sigma}\right)$ with $F(x) \neq \Phi\left(\frac{x-\mu}{\sigma}\right)$ for at least one x

(C) $H_0 : F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) \quad \forall x$ vs
$H_1 : F(x) \leq \Phi\left(\frac{x-\mu}{\sigma}\right)$ with $F(x) \neq \Phi\left(\frac{x-\mu}{\sigma}\right)$ for at least one x

Statistical Hypothesis Testing with SAS and R, First Edition. Dirk Taeger and Sonja Kuhnt.

Test statistic:
(A) $D = \sup_x |\Phi(\frac{x-\mu}{\sigma}) - F_n(x)|$
(B) $D^+ = \sup_x (F_n(x) - \Phi(\frac{x-\mu}{\sigma}))$
(C) $D^- = \sup_x (\Phi(\frac{x-\mu}{\sigma}) - F_n(x))$
$F_n(x)$ is the EDF of the sample and
Φ is the CDF of the standard normal distribution

Test decision:
Reject H_0 if for the observed value d of D
(A) $d \geq d_{1-\alpha}$
(B) $d^+ \geq d^+_{1-\alpha}$
(C) $d^- \geq d^-_{1-\alpha}$
The critical values $d_{1-\alpha}, d^+_{1-\alpha}, d^-_{1-\alpha}$ can be found, for example, in Miller (1956).

p-values:
(A) $p = P(D \geq d)$
(B) $p = P(D^+ \geq d^+)$
(C) $p = P(D^- \geq d^-)$

Annotations:

- This test evaluates the greatest vertical distance between the EDF and the CDF of the standard normal distribution.
- The test statistic D is the maximum of D^+ and D^-: $D = \max(D^+, D^-)$.
- If the sample mean and variance are estimated from the sample the distribution of the test statistic changes and different critical values are needed. Lilliefors published tables with corrected values (Lilliefors 1967) and the test is also known as the *Lilliefors test for normality*.
- SAS and R use different methods to calculate p-values. Hence, results may differ.

Example: To test the hypothesis that the systolic blood pressure of a certain population is distributed according to a normal distribution. A dataset of 55 subjects is sampled (dataset in Table A.1).

SAS code

```
*** Variant 1 ***;
proc univariate data=blood_pressure normal;
  var mmhg;
run;

*** Variant 2 ***;
proc univariate data=blood_pressure;
  histogram mmhg /normal(mu=130 sigma=19.16691);
run;
```

SAS output

```
*** Variant 1 ****
                    Tests for Normality

Test                  --Statistic---    -----p Value------
Kolmogorov-Smirnov    D     0.117254    Pr > D      0.0587

*** Variant 2 ****
         Fitted Normal Distribution for mmhg
         Parameters for Normal Distribution

           Parameter   Symbol   Estimate
           Mean        Mu            130
           Std Dev     Sigma    19.16691

      Goodness-of-Fit Tests for Normal Distribution

Test                ----Statistic-----   ------p Value-----
Kolmogorov-Smirnov  D       0.11725352   Pr > D      >0.250
```

Remarks:

- SAS only calculates $D = \max(D^+, D^-)$ as test statistic.
- Variant 1 calculates the *Lilliefors test for normality* by using the sample mean and sample variance for standardizing the sample. The keyword `normal` enables this test.
- With the variant 2 the original Kolmogorov–Smirnov test with the option `normal` of the `histogram` statement can be calculated; values for the mean and variance have to be provided. Here $\mu = 130$ and $\sigma = 19.16691$ are chosen.
- The syntax is `normal(`*normal-options*`)`. If *normal-options* is not given or `normal(mu=EST sigma=EST)` is given the same test is calculated as with variant 1. The following *normal-options* are valid: `mu=`*value* where *value* is the mean μ of the normal distribution and `sigma=`*value* where *value* is the standard deviation σ of the normal distribution. Note, these values are the true parameters of the normal distribution to test against not the sample parameters. This can be seen in the above example. In both variants the same D-statistic is calculated but the p-values are different.

R code

```
# Calculate mean and standard deviation
m<-mean(blood_pressure$mmhg)
s<-sd(blood_pressure$mmhg)

ks.test(blood_pressure$mmhg,pnorm,mean=m,sd=s,
                      alternative="two.sided",exact=FALSE)
```

R output

```
  One-sample Kolmogorov-Smirnov test

data:  z
D = 0.1173, p-value = 0.4361
alternative hypothesis: two-sided
```

Remarks:

- R only computes the Kolmogorov–Smirnov test, so if the parameters are estimated from the sample as in the above example the p-values are incorrect.
- In the case of ties a warning is prompted that the reported p-values may be incorrect.
- `pnorm` indicates that it is tested for the normal distribution.
- `mean=`*value* is optional. The *value* specifies the mean of the normal distribution to test for. The default is 0, if `mean=`*value* is not specified.
- `sd=`*value* is optional. The *value* specifies the standard deviation of the normal distribution to test for. The default is 1, if `sd=`*value* is not specified.
- `alternative=`*"value"* is optional and defines the type of alternative hypothesis: "two.sided"=the CDFs of $F(x)$ and $\Phi(\frac{x-\mu}{\sigma})$ differ (A); "greater"=the CDF of $F(x)$ lies above that of $\Phi(\frac{x-\mu}{\sigma})$ (B); "less"=the CDF of $F(x)$ lies below that of $\Phi(\frac{x-\mu}{\sigma})$ (C). Default is "two.sided".
- `exact=`*value* is optional. If *value* is TRUE, no ties are present and the sample size is less than 100 an exact p-value is calculated. The default is NULL, that is, no exact p-values.

11.1.2 Anderson–Darling test

Description: Tests if a sample is sampled from a normal distribution with parameter μ and σ^2.

Assumptions:
- Data are measured at least on an ordinal scale.
- The random variables $X_1, \ldots, X_n$ are identically, independently distributed with observations $x_1, \ldots, x_n$ and a continuous distribution function $F(x)$.

Hypotheses: $H_0 : F(x) = \Phi(\frac{x-\mu}{\sigma}) \quad \forall x$ vs
$H_1 : F(x) \neq \Phi(\frac{x-\mu}{\sigma})$ for at least one x

Test statistic: $A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)[\ln\ (p_i) + \ln\ (1 - p_{n-i+1})]$

where $p_i = \Phi\left(\frac{X_{(i)}-\overline{X}}{S}\right), i = 1, \ldots, n,$

and $X_{(1)}, \ldots, X_{(n)}$ the sample in ascending order.

Test decision: Reject H_0 if for the observed value a^2 of A^2
$a^2 \geq a_\alpha$
Critical values a_α can be found, for example, in table 4.2 of Stephens (1986).

p-values: $p = P(A^2 \geq a^2)$

Annotations:

- The test statistic A^2 was proposed by Anderson and Darling (1952).
- Stephens (1986) also treats the case that either μ or σ or both are unknown. They are estimated by $\overline{X}$ and $s^2 = \sum_i^n (X_i - \overline{X})^2/(n-1)$. For the most common case that both are unknown the test statistic is modified as $A^{2*} = (1.0 + 0.75/n + 2.25/n^2)A^2$. For the modified test statistic A^{2*} critical values are given in table 4.7 of Stephens (1986).
- Formulas of approximate p-values can also be found in Stephens (1986).

Example: To test the hypothesis that the systolic blood pressure of a certain population is distributed according to a normal distribution. A dataset of 55 subjects is sampled (dataset in Table A.1).

SAS code

```
*** Variant 1 ***;
proc univariate data=blood_pressure normal;
  var mmhg;
run;

*** Variant 2 ***;
proc univariate data=blood_pressure;
  histogram mmhg /normal(mu=130 sigma=19.16691);
run;
```

SAS output

```
*** Variant 1 ****
                    Tests for Normality

Test                  --Statistic---    -----p Value------
Anderson-Darling      A-Sq  0.888742    Pr > A-Sq   0.0224

*** Variant 2 ****
         Fitted Normal Distribution for mmhg
         Parameters for Normal Distribution

           Parameter   Symbol   Estimate
           Mean        Mu            130
           Std Dev     Sigma    19.16691

      Goodness-of-Fit Tests for Normal Distribution
```

```
Test                 ----Statistic-----   ------p Value-----
Anderson-Darling     A-Sq    0.88874206   Pr > A-Sq   >0.250
```

Remarks:

- SAS computes A^2 and not A^{2*}.
- Variant 1 calculates the Anderson–Darling test using the sample mean and sample variance to standardize the sample. The keyword `normal` enables this test.
- With the variant 2 using the option `normal` of the `histogram` statement the test with known theoretical μ and σ is computed.
- The syntax is `normal(`*normal-options*`)`. If *normal-options* is not given the same test is calculated as with variant 1. The following *normal-options* are valid: `mu=`*value* where *value* is the mean μ and `sigma=`*value* where *value* is the standard deviation σ. Thereby, versions of the test are available for μ or σ or both known. Note, these values are the true parameters of the normal distribution to test for not the sample parameters. In all variants the same A^2 statistic is calculated but the p-values are different. This can be seen in the above example.

R code

```
# Get number of observations
n<-length(blood_pressure$mmhg)

# Standardize the blood pressure
m<-mean(blood_pressure$mmhg)
s<-sd(blood_pressure$mmhg)
z<-(blood_pressure$mmhg-m)/s

# z1 is the array of the ascending sorted values
z1<-sort(z)

# z2 is the array of the descending sorted values
z2<-sort(z,decreasing=TRUE)

# Calculate the test statistic
AD<-(1/n)*sum((1-2*seq(1:n))*(log(pnorm(z1))+
                              log(1-pnorm(z2))))-n

# Calculate modified test statistic
AD_mod<-(1.0+0.75/n+2.25/n^2)*AD

# Calculate approximative p-values according table 4.9
# from Stephens (1986)
if (AD_mod<=0.200)
      p_value=1-exp(-13.436+101.140*AD_mod-223.73*AD_mod^2)
if (AD_mod>0.200 && AD_mod<=0.340 )
      p_value=1-exp(-8.318+42.796*Ad_mod-59.938*AD_mod^2)
if (AD_mod>0.340 && AD_mod<=0.600 )
```

```
      p_value=exp(0.9177-4.279*AD_mod-1.38*AD_mod^2)
if (AD_mod>0.600)
      p_value=exp(0.12937-5.709*AD_mod+0.0186*AD_mod^2)

# Output results
cat("Anderson-Darling test \n\n","AD^2    ","AD^2*    ",
"p-value","\n","-------------------------",
"\n",format(AD,digits=6),format(AD_mod,digits=6),
format(p_value,digits=4),"\n")
```

R output

```
 Anderson-Darling test

 AD^2     AD^2*     p-value
 --------------------------
 0.888742 0.901523 0.006722
```

Remarks:

- This example uses sample moments for μ and σ and the modified test statistic A^{2*}. The approximate p-value is calculated according to Stephens (1986). The approximation can be used for samples of size $n \geq 8$.

11.1.3 Cramér–von Mises test

Description: Tests if a sample is sampled from a normal distribution with parameter μ and σ^2.

Assumptions:
- Data are measured at least on an ordinal scale.
- The random variables $X_1, \ldots, X_n$ are identically, independently distributed with observations $x_1, \ldots, x_n$ and a continuous distribution function $F(x)$.

Hypotheses: $H_0 : F(x) = \Phi(\frac{x-\mu}{\sigma}) \quad \forall x$ vs
$H_1 : F(x) \neq \Phi(\frac{x-\mu}{\sigma})$ for at least one x

Test statistic: $W^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left(p_i - \frac{2i-1}{2n}\right)^2$
where $p_i = \Phi\left(\frac{X_{(i)}-\overline{X}}{S}\right)$, $i = 1, \ldots, n$,
and $X_{(1)}, \ldots, X_{(n)}$ the sample in ascending order.

Test decision: Reject H_0 if for the observed value w^2 of W^2
$w^2 \geq w_{1-\alpha}$
Critical values $w_{1-\alpha}$ can be found, for example, in Pearson and Hartley (1972).

p-values: $p = P(W^2 \geq w^2)$

Annotations:

- The test was independently introduced by Cramér (1928) and von Mises (1931).
- Stephens (1986) also treats the case that either μ or σ or both are unknown. They are estimated by $\overline{X}$ and $s^2 = \sum_i^n (X_i - \overline{X})^2/(n-1)$. For the most common case that both are unknown the test statistic is modified as $W^{2*} = (1 + 0.5/n) * W^2$. For the modified test statistic W^{2*} critical values are given in table 4.7 of Stephens (1986).
- Formulas of approximate p-values can also be found in Stephens (1986).

Example: To test the hypothesis that the systolic blood pressure of a certain population is distributed according to a normal distribution. A dataset of 55 subjects is sampled (dataset in Table A.1).

SAS code

```
*** Variant 1 ***;
proc univariate data=blood_pressure normal;
  var mmhg;
run;

*** Variant 2 ***;
proc univariate data=blood_pressure;
  histogram mmhg /normal(mu=130 sigma=19.16691);
run;
```

SAS output

```
*** Variant 1 ****
                    Tests for Normality

Test                  --Statistic---    -----p Value------
Cramer-von Mises      W-Sq  0.165825    Pr > W-Sq   0.0153

*** Variant 2 ****
         Fitted Normal Distribution for mmhg
         Parameters for Normal Distribution

           Parameter   Symbol   Estimate
           Mean        Mu            130
           Std Dev     Sigma    19.16691

      Goodness-of-Fit Tests for Normal Distribution

Test                ----Statistic-----   ------p Value-----
Cramer-von Mises    W-Sq    0.16582503   Pr > W-Sq   >0.250
```

Remarks:

- SAS computes W^2 and not W^{2*}.
- Variant 1 calculates the Cramér–von Mises test by using the sample mean and sample variance to standardize the sample. The keyword `normal` enables this test.
- With the variant 2 using the option `normal` of the `histogram` statement the test with known theoretical μ and σ is computed.
- The syntax is `normal(`*normal-options*`)`. If *normal-options* is not given the same test is calculated as with variant 1. The following *normal-options* are valid: `mu=`*value* where *value* is the mean μ and `sigma=`*value* where *value* is the standard deviation σ. Thereby, versions of the test are available for μ or σ or both known. Note, these values are the true parameters of the normal distribution to test for not the sample parameters. In all variants the same W^2 statistic is calculated but the p-values are different. This can be seen in the above example.

R code

```
# Get number of observations
n<-length(blood_pressure$mmhg)

# Standardize the blood pressure
m<-mean(blood_pressure$mmhg)
s<-sd(blood_pressure$mmhg)
z<-(blood_pressure$mmhg-m)/s

# Sort the sample
z<-sort(z)

# Calculate the test statistic
W_sq<-1/(12*n)+sum((pnorm(z)-(2*seq(1:n)-1)/(2*n))^2)

# Calculate approximative p-values according to table 4.9
# from Stephens (1986)
W<-(1 + 0.5/n) * W_sq
if (W<0.0275)
             p_value=1-exp(-13.953+775.500*W-12542.610*W^2)
if (W>=0.0275 && W<0.0510)
             p_value=1-exp(-5.9030+179.546*W-1515.290*W^2)
if (W>=0.0510 && W<0.092)
             p_value=exp(0.886-31.620*W+10.897*W^2)
if (W>=0.092)
             p_value=exp(1.111-34.242*W+12.832*W^2)

# Output results
cat("Cramer-von Mises test \n\n","W^2        ","W^2*       ",
"p-value",
```

```
"\n","--------- ---------- ----------",
"\n",W_sq,W,p_value,"\n")
```

R output

```
Cramer-von Mises test

W^2        W^2*       p-value
--------- ---------- ----------
0.1658251 0.1673326  0.01412931
```

Remarks:

- This example uses sample moments for μ and σ and the modified test statistic W^{2*}. The approximate p-value is calculated according to Stephens (1986). The approximation can be used for sample sizes ≥ 7.

11.2 Tests not based on the EDF

11.2.1 Shapiro–Wilk test

Description: Tests if a sample is sampled from a normal distribution.

Assumptions:
- Data are measured on a metric scale.
- The random variables $X_1, \ldots, X_n$ are identically, independently distributed with observations $x_1, \ldots, x_n$ and a continuous distribution function $F(x)$.
- The mean μ and variance σ are unknown.

Hypotheses: $H_0 : F(x) = \Phi(\frac{x-\mu}{\sigma}) \quad \forall x$ vs
$H_1 : F(x) \neq \Phi(\frac{x-\mu}{\sigma})$ for at least one x

Test statistic:
$$W = \frac{\left(\sum_{i=1}^{n} a_i X_{(i)}\right)^2}{\sum_{i=1}^{n} (X_{(i)} - \overline{X})^2}$$
with coefficients $(a_1, \ldots, a_n) = \frac{m'V^{-1}}{\sqrt{m'V^{-1}V^{-1}m}}$,
where $m' = (m_1, \ldots, m_n)$ is the mean vector and V is the covariance matrix of standard normal order statistics
and $X_{(1)}, \ldots, X_{(n)}$ is the sample in ascending order.

Test decision: Reject H_0 if for the observed value w of W
$w \leq w_\alpha$
Critical values w_α for $n \leq 50$ can be found, for example, in Shapiro and Wilk (1965).

p-values: $p = P(W \leq w)$

Annotations:

- The test statistic W was proposed by Shapiro and Wilk (1965).
- For the test statistic it holds that $0 < W \leq 1$.
- The distribution of the test statistic W depends on the sample size. Shapiro and Wilk (1965) derived approximate values of the coefficients as well as percentage points of the null distribution of the test statistic for sample sizes up to $n = 50$. Royston (1982, 1992) developed approximations of these values for sample sizes up to $n = 5000$.
- The Shapiro–Wilk test is a powerful test, especially in samples with small sample sizes Shapiro *et al*. (1968).

Example: To test the hypothesis that the systolic blood pressure of a certain population is distributed according to a normal distribution. A dataset of 55 subjects is sampled (dataset in Table A.1).

SAS code

```
proc univariate data=blood_pressure normal;
  var mmhg;
run;
```

SAS output

```
                    Tests for Normality

Test                  --Statistic---    -----p Value------
Shapiro-Wilk          W     0.960775    Pr < W      0.0701
```

Remarks:

- The keyword `normal` enables the Shapiro–Wilk test.
- SAS calculates the Shapiro–Wilk test only for sample sizes ≤ 2000.
- For sample sizes ≥ 4 the p-values are calculated from the standard normal distribution based on a normalizing transformation.

R code

```
shapiro.test(blood_pressure$mmhg)
```

R output

```
Shapiro-Wilk normality test

data:  blood_pressure$mmhg
W = 0.9608, p-value = 0.07012
```

Remarks:

- R calculates the Shapiro–Wilk test only for sample sizes ≤ 5000.
- For sample sizes ≥ 4 the p-values are calculated based on Royston (1995).

11.2.2 Jarque–Bera test

Description: Tests if a sample is sampled from a normal distribution.

Assumptions:
- Data are measured on a metric scale.
- The random variables $X_1, \ldots, X_n$ are identically, independently distributed with observations $x_1, \ldots, x_n$ and a continuous distribution function $F(x)$.
- The mean μ and variance σ are unknown.

Hypotheses: $H_0 : F(x) = \Phi(\frac{x-\mu}{\sigma}) \quad \forall x$ vs
$H_1 : F(x) \neq \Phi(\frac{x-\mu}{\sigma})$ for at least one x

Test statistic:
$$JB = n\left(\frac{\gamma_1^2}{6} + \frac{(\gamma_2-3)^2}{24}\right)$$
$$\text{with } \gamma_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2\right)^{3/2}}$$
$$\text{and } \gamma_2 = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2\right)^{2}}$$

Test decision: Reject H_0 if for the observed value jb of JB
$jb \geq \chi^2_{1-\alpha;2}$

p-values: $p = 1 - P(JB \leq jb)$

Annotations:
- This test was introduced by Jarque and Bera (1987) as a Lagrange multiplier test with the alternative hypothesis covering any other distribution from the Pearson family of distributions.
- For the calculation of the test statistic the sample skewness γ_1 and sample kurtosis γ_2 are used. If the data are normally distributed the skewness is zero and the kurtosis is three, so large values of the test statistic JB are arguing against the null hypothesis.

- The test statistic *JB* is asymptotically χ^2-distributed with two degrees of freedom.
- Critical values, which are obtained by Monte Carlo simulations and should be used for small sample sizes, can be found in Jarque and Bera (1987) or Dep and Sefton (1996).

Example: To test the hypothesis that the systolic blood pressure of a certain population is distributed according to a normal distribution. A dataset of 55 subjects is sampled (dataset in Table A.1).

SAS code

```
proc autoreg data=blood_pressure;
 model mmhg= /normal;
run;
```

SAS output

```
              The AUTOREG Procedure

            Miscellaneous Statistics

Statistic          Value      Prob           Label

Normal Test        2.6279    0.2688      Pr > ChiSq
```

Remarks:

- The option `normal` after the `model` statement in `PROC AUTOREG` enables the Jarque–Bera test for normality.
- The p-value is calculated from a χ^2-distribution with two degrees of freedom. For low sample sizes the p-value is only a rough approximation.

R code

```
# Calculate sample size
n<-length(blood_pressure$mmhg)

# Calculate sample skewness and sample kurtosis
x<-blood_pressure$mmhg
skewness<-(sum((x-mean(x))^3)/n)/
                               (sum((x-mean(x))^2)/n)^(3/2)
kurtosis<-(sum((x-mean(x))^4)/n)/((sum((x-mean(x))^2)/n)^2)

# Calculate test statistic
jb<-n*(skewness^2/6+(kurtosis-3)^2/24)
```

```
# Calculate asymptotic p-value
p_value<-1-pchisq(jb,2)

# Output results
cat("Jarque-Bera Test \n\n",
"JB          ","p-value",
"\n","---------------------",
"\n",jb,"  ",p_value,"\n")
```

R output

```
   Jarque-Bera Test

 JB           p-value
 ---------------------
 2.627909    0.2687552
```

Remarks:

- There is no R function to calculate the Jarque–Bera test directly.
- This implementation of the test uses the χ^2-distribution with two degrees of freedom to calculate the p-values. Because this is the asymptotic distribution of the test statistic, the p-value is for low sample sizes only a rough approximation.

References

Anderson T.W. and Darling D.A. 1952 Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes. *Annals of Mathematical Statistics* **23**, 193–212.

Cramér H. 1928 On the composition of elementary errors: II. Statistical applications. *Skandinavisk Aktuarietidskrift* **11**, 141–180.

Dep P. and Sefton M. 1996 The distribution of a Lagrange multiplier test of normality. *Economics Letters* **51**, 123–130.

Jarque C.M. and Bera A.K. 1987 A test for normality of observations and regression residuals. *International Statistical Review* **55**, 163–172.

Lilliefors H.W. 1967 On the Kolmogorov–Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association* **62**, 399–402.

Miller L.H. 1956 Table of percentage points of Kolmogorov statistics. *Journal of the American Statistical Association* **51**, 111–121.

Pearson E.S. and Hartley H.O. 1972 *Biometrika Tables for Statisticians*, Vol. 2. Cambridge University Press.

Royston P. 1982 An extension of Shapiro and Wilk's W test for normality to large samples. *Applied Statistics* **31**, 115–124.

Royston P. 1992 Approximating the Shapiro–Wilks W test for nonnormality. *Statistics and Computing* **2**, 117–119.

Royston P. 1995 AS R94: a remark on algorithm AS 181: The W-test for normality. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **44**, 547–551.

Shapiro S.S. and Wilk M.B. 1965 An analysis of variance test for normality (complete sample). *Biometrika* **52**, 591–611.

Shapiro S.S., Wilk M.B. and Chen H.J. 1968 A comparative study of various tests for normality. *Journal of the American Statistical Association* **63**, 1343–1372.

Stephens M.A. 1986 Tests based on the EDF statistics. In *Goodness-of-Fit Techniques* (eds D'Agostino R.B. and Stephens M.A.), pp. 97–193. Marcel Dekker.

von Mises R. 1931 *Wahrscheinlichkeitsrechnung und Ihre Anwendung in der Statistik und Theoretischen Physik*. Deutike.

12

Tests on other distributions

In this chapter we present goodness-of-fit tests for distributions other than Gaussian. Tests in Section 12.1 are based on the empirical distribution function (EDF). Section 12.2 deals with Pearson's χ^2-test, which is an omnibus test for goodness-of-fit but not so powerful if alternative tests on specific distributions are available.

12.1 Tests based on the EDF

12.1.1 Kolmogorov–Smirnov test

Description: Tests if a sample is sampled from a specific distribution function $F_0(x)$.

Assumptions:
- Data are at least measured on an ordinal scale.
- The random variables $X_1, \ldots, X_n$ are independently distributed with observations $x_1, \ldots, x_n$ and a continuous distribution function $F(x)$.

Hypotheses:
(A) $H_0 : F(x) = F_0(x) \quad \forall x$ vs
$H_1 : F(x) \neq F_0(x)$ for at least one x
(B) $H_0 : F(x) = F_0(x) \quad \forall x$ vs
$H_1 : F(x) \geq F_0(x)$ with $F(x) \neq F_0(x)$ for at least one x
(C) $H_0 : F(x) = F_0(x) \quad \forall x$ vs
$H_1 : F(x) \leq F_0(x)$ with $F(x) \neq F_0(x)$ for at least one x

Test statistic:
(A) $D = \sup_x |F_0(x) - F_n(x)|$
(B) $D^+ = \sup_x (F_n(x) - F_0(x))$
(C) $D^- = \sup_x (F_0(x) - F_n(x))$
$F_n(x)$ is the EDF of the sample and

Statistical Hypothesis Testing with SAS and R, First Edition. Dirk Taeger and Sonja Kuhnt.

$F_0(x)$ is the cumulative distribution function (CDF) of the distribution to test against.

Test decision: Reject H_0 if for the observed value d of D
(A) $d \geq d_{1-\alpha}$
(B) $d^+ \geq d^+_{1-\alpha}$
(C) $d^- \geq d^-_{1-\alpha}$
The critical values $d_{1-\alpha}$, $d^+_{1-\alpha}$, $d^-_{1-\alpha}$ can be found, for example, in Miller (1956).

p-values: (A) $p = P(D \geq d)$
(B) $p = P(D^+ \geq d^+)$
(C) $p = P(D^- \geq d^-)$

Annotations:

- This test evaluates the greatest vertical distance between the EDF and the CDF of the distribution to test against.
- The test statistic D is the maximum of D^+ and D^-: $D = \max(D^+, D^-)$.
- The distribution $F_0(x)$ must be fully specified. If parameters have to be estimated the distribution of the test statistic may change.
- SAS and R use different methods to calculate p-values. Hence, results may differ.

Example: To test the hypotheses that the waiting time at a ticket machine follows an exponential distribution. A sample of 10 waiting times in minutes are taken (dataset in Table A.10).

SAS code

```
proc univariate data=waiting;
 histogram time /exponential;
run;
```

SAS output

```
            Parameters for Exponential Distribution

                Parameter   Symbol   Estimate
                Threshold   Theta           0
                Scale       Sigma         6.8
                Mean                      6.8
                Std Dev                   6.8

      Goodness-of-Fit Tests for Exponential Distribution

Test                ----Statistic-----  ------p Value------
Kolmogorov-Smirnov  D       0.26318966  Pr > D        0.197
```

Remarks:

- SAS only calculates $D = max(D^+, D^-)$ as test statistic.
- The option of the `histogram` that enables the test for an exponential distribution is `exponential` (*exponential-options*). If *exponential-options* is not given the parameters of the exponential distribution (the threshold θ and the scale σ) are estimated from the sample. The following *exponential-options* are valid: `theta=`*value* where *value* is the threshold value θ of the exponential distribution and `sigma=`*value* where *value* is the scale parameter σ of the exponential. Note, these values are the true parameters of the exponential distribution to test against not the sample parameters.
- Besides the normal (see Test 11.1.1) and the exponential distribution, the following distributions can also be used as null distributions: beta distribution [keyword `beta` (*beta-options*)], gamma distribution [keyword `gamma` (*gamma-options*)], lognormal distribution [keyword `lognormal` (*lognormal-options*)], Johnson S_B distribution [keyword `SB` (S_B-*options*)], Johnson S_U distribution [keyword `SU` (S_U-*options*)], and Weibull distribution [keyword `Weibull` (*weibull-options*)]. As of SAS 9.3 the following additional distributions can be used as null distributions: Gumbel distribution [keyword `gumbel` (*Gumbel-options*)], inverse Gaussian distribution [keyword `iGauss` (*inverse Gaussian-options*)], generalized Pareto distribution [keyword `pareto` (*Pareto-options*)], power function distribution [keyword `power` (*Power-options*)], and Rayleigh distribution [Rayleigh `beta` (*Rayleigh*)]. Without the specific options, parameters of the distributions are estimated from the samples.

R code

```
# Calculate the rate lambda
lambda<-1/mean(waiting$time)

# Calculate the test
ks.test(waiting$time,pexp,rate=lambda,
                       alternative="less",exact=FALSE)
```

R output

```
    One-sample Kolmogorov-Smirnov test

data:  waiting$time
D^- = 0.2632, p-value = 0.2502
alternative hypothesis: the CDF of x lies below
the null hypothesis
```

Remarks:

- `pexp` indicates that it is tested against the exponential distribution. Other CDFs such as `pnorm` (normal distribution), `plnorm` (lognormal distribution), and `pweibull` (Weibull distribution) can be specified as well.

- The `ks.test` needs explicit given parameters of the null distribution. For the exponential distribution this is the rate parameter *lambda*. In this example it is estimated from the sample. If *lambda* is not given the default value for the rate 1 is used. Unlike SAS the threshold parameter is assumed to be zero in R and must not be given. The parameters must be named as defined in the CDFs of R.
- In the case of ties a warning is prompted that the reported p-values may be incorrect.
- `alternative=`*"value"* is optional and defines the type of alternative hypothesis: "two.sided"=the CDFs of $F(x)$ and $F_0(x)$ differ (A); "greater"=the CDF of $F(x)$ lies above that of $F_0(x)$ (B); "less"=the CDF of $F(x)$ lies below that of $F_0(x)$ (C). Default is "two.sided".
- `exact=`*value* is optional. If *value* is TRUE, no ties are present and the sample size is less than 100 an exact p-value is calculated. The default is NULL, that is, no exact p-values.

12.1.2 Anderson–Darling test

Description: Tests if a sample is sampled from a specific distribution function $F_0(x)$.

Assumptions:

- Data are at least measured on an ordinal scale.
- The random variables $X_1, \dots, X_n$ are independently distributed with observations $x_1, \dots, x_n$ and a continuous distribution function $F(x)$.
- $F_0(x) = F_0(x; \Theta)$ is the CDF of the null distribution with fully specified parameter vector Θ.

Hypotheses: $H_0 : F(x) = F_0(x) \quad \forall x$ vs
$H_1 : F(x) \neq F_0(x)$ for at least one x

Test statistic: $A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)[\ln(p_i) + \ln(1-p_{n-i+1})]$
with $p_i = F_0(X_{(i)}; \Theta)$, $X_{(i)}, \dots, X_{(n)}$ the ascending ordered random sample, and Θ a vector of parameters of the CDF.

Test decision: Reject H_0 if for the observed value a^2 of A^2
$a^2 \geq a_{1-\alpha}$
Critical values $a_{1-\alpha}$ can be found, for example, in Stephens (1986).

p-values: $p = P(A^2 \geq a^2)$

Annotations:

- The test statistic A^2 was proposed by Anderson and Darling (1952).

- Stephens (1986) suggests modified test statistics for the case that θ is partly or completely unknown for a number of common distributions. As an example for the two-parameter exponential distribution $F(x,(\theta,\sigma)) = 1 - \exp\ (-(x-\theta)/\sigma)$ with, $\theta = 0$, σ unknown, the Anderson–Darling test statistic is modified by $A^{2*} = (1.0 + 0.6/n)A^2$. Note that for this distribution $E(X) = \theta + \sigma$ and $Var(X) = \sigma^2$.
- Formulas of approximate p-values are also provided in Stephens (1986).
- SAS uses A^2 for calculating p-values and internal tables for probability levels. In the R example below we use the modified statistic A^{2*}, so the resulting p-values differ.

Example: To test the hypotheses that the waiting time at a ticket machine follows an exponential distribution. A sample of 10 waiting times in minutes are taken (dataset in Table A.10).

SAS code

```
proc univariate data=waiting;
 histogram time /exponential;
run;
```

SAS output

```
            Parameters for Exponential Distribution

                Parameter   Symbol   Estimate
                Threshold   Theta           0
                Scale       Sigma         6.8
                Mean                      6.8
                Std Dev                   6.8

      Goodness-of-Fit Tests for Exponential Distribution

Test                ----Statistic-----  ------p Value------
Anderson-Darling    A-Sq    0.59849387  Pr > A-Sq    >0.250
```

Remarks:

- SAS computes A^2 and not A^{2*}.
- The option of the procedure `histogram` that enables the test for an exponential distribution is `exponential (`*exponential-options*`)`. If *exponential-options* is not given the parameters of the exponential distribution (the threshold θ and the scale σ) are estimated from the sample. Following *exponential-options* are valid: `theta=`*value* where *value* is the threshold value θ of the exponential distribution

and `sigma=`*value* where *value* is the scale parameter σ of the exponential. Note, these values are the true parameters of the exponential distribution to test against not the sample parameters.

- Besides the normal (see Test 11.1.2) and the exponential distribution, the following distributions can also be used as null distributions: beta distribution [keyword `beta` (*beta-options*)], gamma distribution [keyword `gamma` (*gamma-options*)], lognormal distribution [keyword `lognormal` (*lognormal-options*)], Johnson S_B distribution [keyword `SB` (S_B-*options*)], Johnson S_U distribution [keyword `SU` (S_U-*options*)], and Weibull distribution [keyword `Weilbull` (*weibull-options*)]. As of SAS 9.3 the following additional distributions can be used as null distributions: Gumbel distribution [keyword `gumbel` (*Gumbel-options*)], inverse Gaussian distribution [keyword `iGauss` (*inverse Gaussian-options*)], generalized Pareto distribution [keyword `pareto` (*Pareto-options*)], power function distribution [keyword `power` (*Power-options*)], and Rayleigh distribution [Rayleigh `beta` (*Rayleigh*)]. Without the specific options the parameters of the distributions are estimated from the samples.

R code

```
# Get number of observations
n<-length(waiting$time)

# Calculate the rate lambda and standardize
# the waiting times
lambda<-1/mean(waiting$time)
z<-lambda*waiting$time

# z1 is the array of the ascending sorted values
z1<-sort(z)

# z2 is the array of the descending sorted values
z2<-sort(z,decreasing=TRUE)

# Calculate the test statistic
AD<-(1/n)*sum((1-2*seq(1:n))*
                    (log(pexp(z1))+log(1-pexp(z2))))-n

# Calculate modified test statistic
AD_mod<-(1.0+0.6/n)*AD

# Calculate approximative p-values according Table 4.12
# from Stephens (1986)
if (AD_mod<=0.260)
      p_value=1-exp(-12.2204+67.459*AD_mod-110.30*AD_mod^2)
if (AD_mod>0.260 && AD_mod<=0.510 )
      p_value=1-exp(-6.1327+20.218*Ad_mod-18.663*AD_mod^2)
if (AD_mod>0.510 && AD_mod<=0.950 )
      p_value=exp(0.9209-3.353*AD_mod-0.300*AD_mod^2)
```

```
if (AD_mod>0.950)
      p_value=exp(0.7310-3.009*AD_mod+0.150*AD_mod^2)

# Output results
cat("Anderson-Darling test \n\n","AD^2    ","AD^2*    "
                                              ,"p-value",
"\n","-------------------------",
"\n",format(AD,digits=6),format(AD_mod,digits=6),
                          format(p_value,digits=4),"\n")
```

R output

```
   Anderson-Darling test

AD^2     AD^2*     p-value
--------------------------
0.598494 0.634403  0.2653
```

Remarks:

- pexp indicates that it is tested against the exponential distribution. Other CDFs such as pnorm (normal distribution), plnorm (lognormal distribution), and pweibull (Weibull distribution) can be specified as well.
- The parameters of the null distribution must be explicitly given. For the exponential distribution this is the rate parameter *lambda*. In this example it is estimated from the sample. Unlike SAS the threshold parameter is assumed to be zero in R and must not be given. If *lambda* is not given the default value 1 for the rate is used. The parameters must be named as defined in the CDFs of R.
- This example uses the approximate p-values for the modified test A^{2*} and for the case that the rate λ of the exponential distribution is estimated from the sample and the threshold parameter is known to be zero.
- For an exponential distribution with unknown threshold parameter see Stephens (1986) for details on how to calculate the p-values. Formulas for p-value calculation for other distributions are also given in Stephens (1986).

12.1.3 Cramér–von Mises test

Description: Tests if a sample is sampled from a specific distribution function $F_0(x)$.

Assumptions:

- Data are at least measured on an ordinal scale.
- The random variables $X_1, \dots, X_n$ are independently distributed with observations $x_1, \dots, x_n$ and a continuous distribution function $F(x)$.
- $F_0(x) = F_0(x; \Theta)$ is the CDF of the null distribution with fully specified parameter vector Θ.

Hypotheses: $H_0 : F(x) = F_0(x) \quad \forall x$ vs
$H_1 : F(x) \neq F_0(x)$ for at least one x

Test statistic: $W^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left(p_i - \frac{2i-1}{2n}\right)^2$
with $p_i = F_0(X_{(i)}; \Theta)$, $X_{(i)}, \dots, X_{(n)}$ the ascending ordered random sample, and Θ a vector of parameters of the CDF.

Test decision: Reject H_0 if for the observed value w^2 of W^2
$w^2 \geq w_{1-\alpha}$
Critical values $w_{1-\alpha}$ can be found, for example, in Stephens (1986).

p-values: $p = P(W^2 \geq w^2)$

Annotations:

- The test was independently introduced by Cramér (1928) and von Mises (1931).
- Stephens (1986) suggests modified test statistics for the case that θ is partly or completely unknown for a number of common distributions. For the exponential distribution as null distribution with known threshold parameter $\theta = 0$ and unknown scale parameter σ the modified Anderson–Darling test statistic is $W^{2*} = (1.0 + 0.16/n)W^2$. Note that $F(x, (\theta, \sigma)) = 1 - \exp(-(x - \theta)/\sigma)$ with $E(X) = \theta + \sigma$ and $Var(X) = \sigma^2$.
- Formulas of approximate p-values can also be found in Stephens (1986).
- SAS uses W^2 for calculating p-values and internal tables for probability levels. In the R example below we use the modified statistic W^{2*}, so the resulting p-values differ.

Example: To test the hypotheses that the waiting time at a ticket machine follows an exponential distribution. A sample of 10 waiting times in minutes are taken (dataset in Table A.10).

SAS code

```
proc univariate data=waiting;
 histogram time /exponential;
run;
```

SAS output

```
           Parameters for Exponential Distribution

                Parameter   Symbol   Estimate
                Threshold   Theta           0
```

```
                Scale       Sigma          6.8
                Mean                       6.8
                Std Dev                    6.8

      Goodness-of-Fit Tests for Exponential Distribution

Test                ----Statistic-----  ------p Value------
Cramér-von Mises    W-Sq    0.10524683  Pr > W-Sq    >0.250
```

Remarks:

- SAS computes W^2 and not W^{2*}.
- The option of the procedure `histogram` that enables the test for an exponential distribution is `exponential (`*exponential-options*`)`. If *exponential-options* is not given the parameters of the exponential distribution (the threshold θ and the scale σ) are estimated from the sample. The following *exponential-options* are valid: `theta=`*value* where *value* is the threshold value θ of the exponential distribution and `sigma=`*value* where *value* is the scale parameter σ of the exponential. Note, these values are the true parameters of the exponential distribution to test against not the sample parameters.
- Besides the normal (Test 11.1.3) and the exponential distribution, the following distributions can also be used as null distributions: beta distribution [keyword `beta (`*beta-options*`)`], gamma distribution [keyword `gamma (`*gamma-options*`)`], lognormal distribution [keyword `lognormal (`*lognormal-options*`)`], Johnson S_B distribution [keyword `SB (`S_B*-options*`)`], Johnson S_U distribution [keyword `SU (`S_U*-options*`)`], and Weibull distribution [keyword `Weilbull (`*weibull-options*`)`]. As of SAS 9.3 the following additional distributions can be used as null distributions: Gumbel distribution [keyword `gumbel (`*Gumbel-options*`)`], inverse Gaussian distribution [keyword `iGauss (`*inverse Gaussian-options*`)`], generalized Pareto distribution [keyword `pareto (`*Pareto-options*`)`], power function distribution [keyword `power (`*Power-options*`)`], and Rayleigh distribution [Rayleigh `beta (`*Rayleigh*`)`]. Without the specific options the parameters of the distributions are estimated from the samples.

R code

```
# Get number of observations
n<-length(waiting$time)

# Calculate the rate lambda and standardize
# the waiting times
lambda<-1/mean(waiting$time)
z<-lambda*waiting$time

# Sort the sample
z<-sort(z)
```

```
# Calculate the test statistic
W_sq<-1/(12*n)+sum((pexp(z)-(2*seq(1:n)-1)/(2*n))^2)

# Calculate approximative p-values according to table 4.12
# from Stephens (1986)
W<-(1.0 + 0.16/n) * W_sq
if (W<0.035)
     p_value=1-exp(-11.334+459.098*W-5652.100*W^2)
if (W>=0.035 && W<0.074)
     p_value=1-exp(-5.779+132.89*W-866.58*W^2)
if (W>=0.074 && W<0.160)
     p_value=exp(0.586-17.87*W+7.417*W^2)
if (W>=0.160)
     p_value=exp(0.447-16.592*W+4.849*W^2)

# Output results
cat("Cramér-von Mises test \n\n","W^2       ",
"W^2*      ","p-value",
"\n","--------- ---------- ----------",
"\n",W_sq,W,p_value,"\n")
```

R output

```
Cramér-von Mises test

W^2        W^2*       p-value
--------- ---------- ----------
0.1052468 0.1069308   0.289371
```

Remarks:

- `pexp` indicates that it is tested against the exponential distribution. Other CDFs such as `pnorm` (normal distribution), `plnorm` (lognormal distribution), and `pweibull` (Weibull distribution) can be specified as well.
- The parameters of the null distribution must be explicitly given. For the exponential distribution this is the rate parameter *lambda*. In this example it is estimated from the sample. If *lambda* is not given the default value 1 for the rate is used. Unlike in SAS the threshold parameter is assumed to be zero in R and must not be given. The parameters must be named as defined in the CDFs of R.
- This example uses the approximative p-values for the modified test W^{2*} and for the case that the rate λ of the exponential distribution is estimated from the sample and the threshold parameter is known to be zero.
- For an exponential distribution with unknown threshold parameter see Stephens (1986) for details on how to calculate the p-values. Formulas for p-value calculation for other distributions are also given in Stephens (1986).

12.2 Tests not based on the EDF

12.2.1 χ^2 Goodness-of-fit test

Description: Tests if a sample is sampled from a distribution function $F_0(x)$.

Assumptions:

- Data are at least measured on a nominal scale.
- The random variables $X_1, \ldots, X_n$ are independently distributed with observations $x_1, \ldots, x_n$ and a distribution function $F(x)$.
- The distribution with distribution function $F_0(x)$ to test against is completely specified.

Hypotheses: $H_0 : F(x) = F_0(x) \quad \forall x$ vs
$H_1 : F(x) \neq F_0(x)$ for at least one x

Test statistic:

$$X^2 = \sum_{j=1}^{k} \frac{(n_j - np_j)^2}{np_j}$$

in which the data are grouped into k classes, n_j is the number of the elements of the sample in class j and p_j is the probability of an observation to be in class j under the null hypothesis.

Test decision: Reject H_0 if for the observed value χ^2 of X^2
$\chi^2 \geq \chi^2_{1-\alpha;k-1}$
Critical values $\chi^2_{1-\alpha;k-1}$ be found in Table B.3 of Appendix B.

p-values: $p = 1 - P(X^2 \leq \chi^2_{1-\alpha})$

Annotations:

- The test statistic X^2 was proposed by Pearson (1900) and is asymptotically χ^2-distributed with $k-1$ degrees of freedom.
- To conduct the test data are grouped into k disjunct classes and the absolute frequencies n_j are compared with the expected values np_j.
- The number of expected observations in each cell should be at least 5 to ensure the approximate χ^2-distribution.
- If only the class of distribution to test against is specified and parameters are estimated, the test can also be applied. However, then the number of degrees of freedom of the χ^2-distribution must be reduced by the number of estimated parameters.
- A special case of this test is Test 4.3.1 for the binomial distribution.
- If specific goodness-of-fit tests for distributions are available, such as for the Gaussian distribution (Chapter 11), they are usually to be preferred. However, this χ^2-test is very suitable for distributions of discrete random variables.

Example: Suppose a dice is thrown 60 times, with the following results: 10 times 1-pip, 12 times 2-pips, 7 times 3-pips, 11 times 4-pips, 9 times 5-pips, 11 times 6-pips. We want to test the hypothesis that the underlying distribution is uniform. We assume that the dice is fair so the expected number is always $60 \times 1/6 = 10$ and the probability for each side is $1/6$.

SAS code

```
data dice;
 input pips counts;
 datalines;
 1 10
 2 12
 3  7
 4 11
 5  9
 6 11
 ;
run;

proc freq data=dice;
 tables pips /chisq
       testp=(0.166667 0.166667 0.166667 0.166667
                                0.166667 0.166667);
 weight counts;
run;
```

SAS output

```
     Chi-Square Test
for Specified Proportions
--------------------------
Chi-Square         1.6000
DF                      5
Pr > ChiSq         0.9012

     Sample Size = 60
```

Remarks:

- The data are ordered in that way, that for each pip number the observed counts are given in the variable *counts*.
- The option `/chisq` of the `table` statement enables the χ^2-test for the variable *pips*.
- The second option `testp=(`*probabilities*`)` is optional. If it is not given equal probabilities for each category are assumed, otherwise *probabilities* contains the probabilities corresponding to the observed numbers, separated by blanks or commas.
- As the null hypothesis is rejected if $\chi^2 \geq \chi^2_{1-\alpha;2}$ the p-value is calculated as `1- probchi(1.6,5)`.

R code

```
obs<-c(10,12,7,11,9,11)
probs<-c(1/6,1/6,1/6,1/6,1/6,1/6)

chisq.test(obs,p=probs)
```

R output

```
Chi-squared test for given probabilities

data:  obs
X-squared = 1.6, df = 5, p-value = 0.9012
```

Remarks:

- The first parameter of the `chisq.test` function holds the vector of the observed numbers. One figure stands for each category.
- The second parameter `p=`*probabilities* is optional. If it is not given equal probabilities for each category are assumed, otherwise *probabilities* contains the vector of the probabilities corresponding to the vector of observed numbers.
- As the null hypothesis is rejected if $\chi^2 \geq \chi^2_{1-\alpha;2}$ the p-value is calculated as `1-pchisq(1.6,5)`.

References

Anderson T.W. and Darling D.A. 1952 Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes. *Annals of Mathematical Statistics* **23**, 193–212.

Cramér H. 1928 On the composition of elementary errors: II. Statistical applications. *Skandinavisk Aktuarietidskrift* **11**, 141–180.

Miller L.H. 1956 Table of percentage points of Kolmogorov statistics. *Journal of the American Statistical Association* **51**, 111–121.

Pearson K. 1900 On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine* **50**, 157–175.

Stephens M.A. 1986 Tests based on the EDF statistics. In *Goodness-of-Fit Techniques* (eds D'Agostino R.B. and Stephens M.A.), pp. 97–193. Marcel Dekker.

von Mises R. 1931 *Wahrscheinlichkeitsrechnung und Ihre Anwendung in der Statistik und Theoretischen Physik*. Deutike.