# Automatic regrouping of strata in the goodness-of-fit chi-square test

Vicente Núñez-Antón, Juan Manuel Pérez-Salamero González,
Marta Regúlez-Castillo, Manuel Ventura-Marco
and Carlos Vidal-Meliá

June 2019

The material contained herein is supplementary to the article named in the title and published in SORT-Statistics and Operations Research Transactions Volume 43 (1).

# 1. Pearson's goodness-of-fit test in the software analyzed

**Table A1.** *Pearson's goodness-of-fit test in the software analyzed*

| Software | Chi-square Test | Regroup |
|---|---|---|
| **EXCEL** *MicroSoft Corporation* | CHISQ.TEST Returns the p-value obtained from Pearson's chi-square statistic [1.]. If any expected value is zero the error message ♯`DIV/0! division by zero`♯ is displayed. | NO |
| **GraphPad** *GraphPad Software, Inc* | QuickCalcs. On-line. It computes the statistic keeping the original observed and expected values, but only warns about the violation of the requirement of a minimum size of 5. If any of the expected values is zero the test is not carried out. | NO |
| **JMP** *SAS Institute Inc. Cary* | If any expected value is zero, the chi-square statistic is computed without taking this into account. It reports the error but does not compute the p-value. If the expected values or expected frequencies do not add up to 1 it allows them to be rescaled. | NO |
| **Mathematica** *Wolfram Research, Inc.* | PearsonChiSquareTest is a function that computes Pearson´s chi-square statistic but based on a method due to D`Agostino and Stephens. In this method the histograms of the observed and expected values are compared, so it does not calculate the statistic in the same way as Pearson. (Ross, 2015). | NO |
| **MATLAB** *The MathWorks Inc.* | chi2gof: It computes the goodness-of-fit statistic. It regroups the strata at the extreme of the tails but not the intermediate ones. It allows the minimum size requirement to be set by using the EMin option.It returns the statistic value, the regrouped strata values, and any other information required about the test. | YES but only at the extreme of the tails. |
| **Minitab** *Minitab Inc.* | Mac: Statistics > Tables > Chi-Square Goodness-of-Fit; PC: STATISTICS > Chi-Square Goodness-of-Fit. It warns that results may not be accurate when the strata of expected values have sizes lower than 5 and 1. It also gives information about the percentage of strata that do not satisfy the requirement. The expected frequencies must be entered in place of the expected values. | NO |

continue on the next page

| Software | Chi-square Test | Regroup |
|---|---|---|
| **NCSS**<br>*NCSS, LLC* | It does not inform about the minimum size reached by some categories and does not display any error message if it detects a zero expected value, given that it does not consider such groups in carrying out its calculations. | NO |
| **PH-Stat**<br>*Pearson Education, Inc.* | Add-in in Excel for statistical analysis based on functions written in Excel, so it has the same characteristics. In the chi-square goodness-of-fit tests it displays an error message about the violation of the assumption on the minimum expected frequency if it does not reach a minimum of 1 or 5 (chi-square test about difference between proportions), depending on the cases. | NO |
| **PSPPIRE**<br>*Free Software Foundation, Inc.* | Free software similar to SPSS, but it does not display any minimum size requirement error message as SPSS does. | NO |
| **R**<br>*The R Foundation for Statistical Computer* | chisq.test: If there are strata with expected values lower than 5, it reports that the results are not correct: 'Chi-squared approximation may be incorrect'. It there is a zero expected value, it does not compute the statistic because of division by zero. The statistic given by gofTest of the ENVStats is based on chisq.tets. | NO |
| **SAS-STAT**<br>*SAS Institute Inc.* | The statement TABLES given in the procedure PROC FREQ does not allow for zero expected values in the option TESTF. It calculates the statistic showing the percentage of bins with expected values lower than 5 and it warns that the chi-square test results are not valid. | NO |
| **S-PLUS**<br>*TIBCO Software Inc.* | Chisq.test works as in R, given that both use the same language and many of the functions of S-PLUS. The function chisq.gof in S-PLUS computes the goodness-of-fit test, but for theoretical expected values given by the usual statistical distributions. | NO |

continue on the next page

| Software | Chi-square Test | Regroup |
|---|---|---|
| **SPSS** *IBM Corporation and others* | If any expected frequency is zero the test procedure stops and reports this by telling the user that the expected values in each category must be at least 1 and no more than 20 % of the categories may be lower than 5. If any observed value is zero but the corresponding expected value is not, the test is not computed because it considers that there are fewer observed categories than expected. It does not allow an expected value for frequencies of less than 0.0001 | NO |
| **STATA** *Estima Inc.* | To compute the statistic it deletes the zero expected values. To use the frequency tables (tabi, tab2, tabulate), it demands integer values, because it requires counts (absolute frequency), not relative frequencies. | NO |
| **Statistica** *Dell Inc.* | It reports a problem if the total sum of observed values does not coincide with that of the expected values, but it does not warn about the problem of the violation of the minimum size requirement. | NO |
| **Stochastic Simulation in Java. SSJ 3.2.0** *L´Ecuyer et al. University of Montreal* | GofStat: It conducts Pearson´s goodness-of-fit test with the possibility of regrouping the strata given a minimum size required value to be set by the user. This is done in two steps using the observed and expected values. It combines the functions OutcomeCategoriesChi2 and regroupCategories. | YES |
| **ViSta** *F.W. Young.* **XLISP-STAT** *L. Tierney* **XLISP version** *D. Betz* | LispStat is no longer developed, because its creator is now a member of the R core team of programmers. There is software based on XLisp-Stat, such as ViSta, that warns about the existence of expected frequencies lower than 6 and that the chi-square test will not be valid. | NO |
| **XLStatistics** *R. Carr* | XLStatistics is and add-in in Excel for statistical analysis. It works with the functions given in Excel, so it has the same problems: it does not warn about the violation of the minimum size requirement and it does not allow for zero expected values. | NO |

## 2. Case 2. Mathematical procedure

$$\max_{n_i^{SUB}} \left\{ n^{SUB} = \sum_{i=1}^{k} n_i^{SUB} \right\} \tag{1}$$

subject to:

$$\chi^2(n_1^{SUB}, \cdots, n_k^{SUB}) = \sum_{j=1}^{cat_{reg}} \frac{(\overline{\overline{n}}_j^{SUB} - \overline{\overline{n}}_j^{exp})^2}{\overline{\overline{n}}_j^{exp}} \leq \chi^2_{(\alpha,r)} \tag{2}$$

$$n_i^{exp} = \frac{N_i}{N} n^{SUB} = \frac{N_i}{N} \sum_{i=1}^{k} n_i^{SUB} \tag{3}$$

$$0 \leq n_i^{SUB} \leq N_i \tag{4}$$

$$0 \leq n_i^{SUB} \leq n_i^{RS} \tag{5}$$

$$n_i^{SUB} \in Z; \forall i = 1, \ldots, k \tag{6}$$

with

$n^{SUB}$: Subsample size.
$n_i^{SUB}$: Size of category i from the subsample (observed values).
$k$: Number of strata on the variable of interest from which the stratification is made.
$\chi^2(n_1^{SUB}, \ldots, n_k^{SUB})$: Chi-square goodness-of-fit test statistic. Its value depends on the size of the regrouped strata.
$n_i^{exp}$: Expected value size of category i from the subsample. It depends on the relative frequency of the population and the size of the subsample.
$\chi^2_{(\alpha,r)}$: Critical value from the chi-square distribution with r degrees of freedom and a given statistical significance level $\alpha$ fixed at 5%.
$N_i$: Size of stratum i from male pensioners classified as permanently disabled, given by INSS (2014).
$\overline{\overline{n}}_j^{exp}$: Expected size of the regrouped category j from the sub-sample. It depends on the relative frequency from the population and from the size of the subsample.
$\overline{\overline{n}}_j^{SUB}$: Proposed observed size for the regrouped stratum j from the subsample.
$N$: Total number of male pensioners classified as permanently disabled in the population of pensioners given by INSS (2014).
$cat_{reg}$: Number of regrouped strata.
$r = cat_{reg} - 1$: Degrees of freedom for the test, equal to the number of regrouped strata minus 1, given that in this case there are no parameters to be estimated because the population distribution is known.
$n_i^{RS}$: Size of category i from the post-stratification of the CSWL (Random Sample).
$Z$: Set of integer numbers.

Constraint (2) is intended to achieve a better fit of the extracted subsample than the original (CSWL), given that it provides a value for the goodness-of-fit statistic that does not reject the null hypothesis. Using the functions shown in Appendix 2 the statistic and the degrees of freedom are calculated from the values of the automatically regrouped categories.

Rule (3) establishes that the regrouped expected value of each category or stratum, in each iteration, automatically adapts to the new size that the subsample can take.

Constraint (4) is set to prevent the outliers found in the CSWL. Given the procedure for obtaining the CSWL, and given that it comes from administrative records, the processing date of the CSWL is later than the one on which the Spanish Social Security Institute drawn up its statistics (INSS 2014). Therefore, there might be strata in the CSWL with pensioners who do not belong to the population because their benefits have been awarded retroactively. This constraint can be ignored if the sample is really contained in the population.

Constraint (5) implies that the subsample must be contained in the CSWL and (6) requires that the number of pensioners to be included in each stratum of the subsample be a natural number (non-negative integer).

## 3. Supplementary material. Codes in Excel (VBA) and Mathematica

**Listing 1: VBA in Excel code. Chi-square statistic with regrouped strata**

```
Function ChiSquaredTestChi(Observed As Range, Expected As Range, _
Minimum As Integer) As Double
'The observed and expected variables are declared as cell ranges
'and we use the Minimum variable to set the minimum number of
'elements for each strata.

'Statement of chi variable and observed and expected frequency
'matrices.

Dim chi As Double
Dim FreqExp() As Double
Dim FreqObs() As Double

'Correct Selection of Observed and Expected Data Range

    'By rows
    ObservedRows = Observed.Rows.Count
    ExpectedRows = Expected.Rows.Count

    'By columns
    ObservedColumns = Observed.Columns.Count
    ExpectedColumns = Expected.Columns.Count

    If ObservedColumns <> ExpectedColumns Then
    MsgBox "Incorrect selection"
    GoTo final
    End If

    If ObservedRows <> ExpectedRows Then
    MsgBox "Incorrect selection"
    GoTo final
    End If

'Resize the matrices from their actual dimension according
'to user data selection

ReDim FreqExp(ExpectedRows)
ReDim FreqObs(ObservedRows)

'Initialization of variables to count strata of minimum size

Dim StrataRegrouped As Integer
Dim last As Integer

StrataRegrouped = 0
last = 0
accumulator = 0

'Initial statement of arrays of regrouped frequencies for observed
'and expected values. Frequencies that will result from regrouping
'strata by requeriment of minimum size.

Dim FreqExpReg() As Double
Dim FreqObsReg() As Double

'Initial assignment of Excel values to frequency arrays (expected
'and observed).

    'Expected
ne = 0
For Each x In Expected
ne = ne + 1
FreqExp(ne) = x.Value
Next x
    'Observed
no = 0
For Each x In Observed
no = no + 1
FreqObs(no) = x.Value
Next x

'Grouping in strata of size greater than or equal to Minimum.
'Prepare the array of values for subsequent assignment to the
'frequency matrix of the strata of valid size.

'add to the minor nearest. Start: first

If FreqExp(1) < Minimum Then
   FreqExp(2) = FreqExp(2) + FreqExp(1)
   FreqObs(2) = FreqObs(2) + FreqObs(1)
   FreqExp(1) = 0
   FreqObs(1) = 0
Else
   StrataRegrouped = StrataRegrouped + 1
   last = 1
End If

For i = 2 To ObservedRows - 1
```

```
If last > 0 Then
    If (FreqExp(i) < Minimum) And (FreqExp(i - 1) = 0) And (FreqExp(i + 1) > FreqExp(last)) Then
      FreqExp(last) = FreqExp(last) + FreqExp(i)
      FreqObs(last) = FreqObs(last) + FreqObs(i)
      FreqExp(i) = 0
      FreqObs(i) = 0
    ElseIf (FreqExp(i) < Minimum) And (FreqExp(i - 1) = 0) And (FreqExp(i + 1) <= FreqExp(last)) Then
      FreqExp(i + 1) = FreqExp(i + 1) + FreqExp(i)
      FreqObs(i + 1) = FreqObs(i + 1) + FreqObs(i)
      FreqExp(i) = 0
      FreqObs(i) = 0
    End If
End If
If (FreqExp(i) < Minimum) And (FreqExp(i - 1) = 0) And (last = 0) Then
      FreqExp(i + 1) = FreqExp(i + 1) + FreqExp(i)
      FreqObs(i + 1) = FreqObs(i + 1) + FreqObs(i)
      FreqExp(i) = 0
      FreqObs(i) = 0
End If
If (FreqExp(i) < Minimum) And (FreqExp(i - 1) > 0) And (FreqExp(i - 1) < FreqExp(i + 1)) Then
      FreqExp(i - 1) = FreqExp(i - 1) + FreqExp(i)
      FreqObs(i - 1) = FreqObs(i - 1) + FreqObs(i)
      FreqExp(i) = 0
      FreqObs(i) = 0
      last = i - 1
ElseIf (FreqExp(i) < Minimum) And (FreqExp(i - 1) > 0) And (FreqExp(i - 1) >= FreqExp(i + 1)) Then
      FreqExp(i + 1) = FreqExp(i + 1) + FreqExp(i)
      FreqObs(i + 1) = FreqObs(i + 1) + FreqObs(i)
      FreqExp(i) = 0
      FreqObs(i) = 0
End If
If FreqExp(i) >= Minimum Then
    StrataRegrouped = StrataRegrouped + 1
    last = i
End If
Next i

If FreqExp(ObservedRows) < Minimum Then
   FreqExp(last) = FreqExp(last) + FreqExp(ObservedRows)
   FreqObs(last) = FreqObs(last) + FreqObs(ObservedRows)
   FreqExp(ObservedRows) = 0
   FreqObs(ObservedRows) = 0
Else
 StrataRegrouped = StrataRegrouped + 1
End If

'Control of sample size less than Minimum. No strata of minimum size.

If StrataRegrouped = 0 Then
MsgBox "No strata of minimum size"
GoTo final
End If

'Redimension of the matrices of regrouped frequencies,
'after knowing the number of strata of size greater or equal to Minimum.

ReDim FreqExpReg(StrataRegrouped)
ReDim FreqObsReg(StrataRegrouped)

'Assignment of frequency values to the frequency matrices.

For i = 1 To ObservedRows
If FreqExp(i) <> 0 Then
accumulator = accumulator + 1
FreqExpReg(accumulator) = FreqExp(i)
FreqObsReg(accumulator) = FreqObs(i)
End If
Next i

'Calculation of the chi-square statistic

   'Control of null values in the matrix of expected regrouped frequencies
    For j = 1 To StrataRegrouped
    If FreqExpReg(j) = 0 Then
    GoTo final
    Else
   'Summative in chi-square formula.
    chi = chi + (FreqObsReg(j) - FreqExpReg(j)) ^ 2 / FreqExpReg(j)
    End If
    Next j

'Assignment of the value of the calculated statistic to the
'ChiSquaredTestChi function.

    ChiSquaredTestChi = chi

' ****************** Omitted for iterative use (Solver) ******************
'Verification: Total of observed values is equal to the total of
'expected values.
'
'Dim totalobs As Double
'Dim totalexp As Double
'     For Each x In Observed
'          totalobs = totalobs + x.Value
'     Next x
```

```
'
'     For Each x In Expected
'         totalexp = totalexp + x.Value
'     Next x
'     If totalobs <> totalexp Then
'         MsgBox "Total observed does not match total expected"
'     Else
'     End If
'*************************************************************************

'Initialization of frequency arrays.
Erase FreqExp()
Erase FreqObs()
Erase FreqExpReg()
Erase FreqObsReg()

final:

End Function
```

**Listing 2: Mathematica code. Chi-square statistic with regrouped strata**

```
PearsonTestChi2[obs_List, exp_List, min_Integer] /;
  Length[obs] == Length[exp] :=
 Module[{strata, last, i}, {strata = Length[exp];
   obsSUB = obs;
   expSUB = exp;
   last = 0;
   {{If[expSUB[[1]] < min, {expSUB[[2]] = expSUB[[2]] + expSUB[[1]],
       obsSUB[[2]] = obsSUB[[2]] + obsSUB[[1]], expSUB[[1]] = 0,
       obsSUB[[1]] = 0}, last = 1]}, {For[i = 2, i <= strata - 1, i++,
      {{If[
         last > 0, {{If[(expSUB[[i]] < min) \[And] (expSUB[[i - 1]] ==
                0) \[And] (expSUB[[i + 1]] >
               expSUB[[last]]), {expSUB[[last]] =
              expSUB[[last]] + expSUB[[i]],
             obsSUB[[last]] = obsSUB[[last]] + obsSUB[[i]],
             expSUB[[i]] = 0,
             obsSUB[[i]] =
              0}]}, {If[(expSUB[[i]] < min) \[And] (expSUB[[i - 1]] ==
                0) \[And] (expSUB[[i + 1]] <=
               expSUB[[last]]), {expSUB[[i + 1]] =
              expSUB[[i + 1]] + expSUB[[i]],
             obsSUB[[i + 1]] = obsSUB[[i + 1]] + obsSUB[[i]],
             expSUB[[i]] = 0, obsSUB[[i]] = 0}]}}]}, {If[(expSUB[[
             i]] < min) \[And] (expSUB[[i - 1]] == 0) \[And] (last ==
            0), {expSUB[[i + 1]] = expSUB[[i + 1]] + expSUB[[i]],
          obsSUB[[i + 1]] = obsSUB[[i + 1]] + obsSUB[[i]],
          expSUB[[i]] = 0,
          obsSUB[[i]] =
           0}]}, {If[(expSUB[[i]] < min) \[And] (expSUB[[i - 1]] >
            0) \[And] (expSUB[[i - 1]] < expSUB[[i + 1]]), {expSUB[[
            i - 1]] = expSUB[[i - 1]] + expSUB[[i]],
          obsSUB[[i - 1]] = obsSUB[[i - 1]] + obsSUB[[i]],
          expSUB[[i]] = 0, obsSUB[[i]] = 0,
          last = i - 1}]}, {If[(expSUB[[i]] <
            min) \[And] (expSUB[[i - 1]] >
            0) \[And] (expSUB[[i - 1]] >= expSUB[[i + 1]]), {expSUB[[
            i + 1]] = expSUB[[i + 1]] + expSUB[[i]],
          obsSUB[[i + 1]] = obsSUB[[i + 1]] + obsSUB[[i]],
          expSUB[[i]] = 0, obsSUB[[i]] = 0}]}, {If[expSUB[[i]] >= min,
          last = i]}}]}, {If[
      expSUB[[strata]] <
       min, {expSUB[[last]] = expSUB[[last]] + expSUB[[strata]],
       obsSUB[[last]] = obsSUB[[last]] + obsSUB[[strata]],
       expSUB[[strata]] = 0, obsSUB[[strata]] = 0}]}};
   expreg = {}; obsreg = {};
   For[i = 1, i <= strata, i++,
    If[expSUB[[i]] >= min, {AppendTo[expreg, expSUB[[i]]],
      AppendTo[obsreg, obsSUB[[i]]]}]];
   chi = N[Total[(obsreg - expreg)^2/expreg]]};
  If[Total[exp] != Total[obs],
   Print["Total observed does not match total expected. ",
    " Chi statistic ", chi], Return[chi]]]
```

**Listing 3: VBA in Excel code. Number of regrouped strata**

```
Function ChiSquaredTestStrataReg(Observed As Range, Expected As Range, _
Minimum As Integer) As Double
'The observed and expected variables are declared as cell ranges
'and we use the Minimum variable to set the minimum number of
'elements for each strata.

'Statement of chi variable and observed and expected frequency
'matrices.
Dim chi As Double
Dim FreqExp() As Double
Dim FreqObs() As Double

'Correct Selection of Observed and Expected Data Range
```

```
     'By rows
    ObservedRows = Observed.Rows.Count
    ExpectedRows = Expected.Rows.Count

     'By columnas
    ObservedColumns = Observed.Columns.Count
    ExpectedColumns = Expected.Columns.Count

     If ObservedColumns <> ExpectedColumns Then
    MsgBox "Incorrect selection"
    GoTo final
    End If

     If ObservedRows <> ExpectedRows Then
    MsgBox "Incorrect selection"
    GoTo final
    End If

'Resize the matrices from their actual dimension according to user data
'selection

ReDim FreqExp(ObservedRows)
ReDim FreqObs(ExpectedRows)

'Initialization of variables to count strata of minimum size.

Dim StrataRegrouped As Integer
Dim last As Integer

StrataRegrouped = 0
last = 0
accumulator = 0

'Initial statement of arrays of regrouped frequencies for observed and
'expected values. Frequencies that will result from regrouping strata by
'requeriment of minimum size.

Dim FreqExpReg() As Double
Dim FreqObsReg() As Double

'Initial assignment of Excel values to frequency arrays (expected and
'observed).

     'Expected
ne = 0
For Each x In Expected
ne = ne + 1
FreqExp(ne) = x.Value
Next x
     'Observed
no = 0
For Each x In Observed
no = no + 1
FreqObs(no) = x.Value
Next x

'Grouping in strata of size greater than or equal to Minimum.
'Prepare the array of values for subsequent assignment to the
'frequency matrix of the strata of valid size.

'add to the minor nearest. Start: first

If FreqExp(1) < Minimum Then
   FreqExp(2) = FreqExp(2) + FreqExp(1)
   FreqObs(2) = FreqObs(2) + FreqObs(1)
   FreqExp(1) = 0
   FreqObs(1) = 0
Else
   StrataRegrouped = StrataRegrouped + 1
   last = 1
End If

For i = 2 To ObservedRows - 1

If last > 0 Then
     If (FreqExp(i) < Minimum) And (FreqExp(i - 1) = 0) And (FreqExp(i + 1) > FreqExp(last)) Then
       FreqExp(last) = FreqExp(last) + FreqExp(i)
       FreqObs(last) = FreqObs(last) + FreqObs(i)
       FreqExp(i) = 0
       FreqObs(i) = 0
     ElseIf (FreqExp(i) < Minimum) And (FreqExp(i - 1) = 0) And (FreqExp(i + 1) <= FreqExp(last)) Then
       FreqExp(i + 1) = FreqExp(i + 1) + FreqExp(i)
       FreqObs(i + 1) = FreqObs(i + 1) + FreqObs(i)
       FreqExp(i) = 0
       FreqObs(i) = 0
    End If
End If
If (FreqExp(i) < Minimum) And (FreqExp(i - 1) = 0) And (last = 0) Then
       FreqExp(i + 1) = FreqExp(i + 1) + FreqExp(i)
       FreqObs(i + 1) = FreqObs(i + 1) + FreqObs(i)
       FreqExp(i) = 0
       FreqObs(i) = 0
End If
If (FreqExp(i) < Minimum) And (FreqExp(i - 1) > 0) And (FreqExp(i - 1) < FreqExp(i + 1)) Then
       FreqExp(i - 1) = FreqExp(i - 1) + FreqExp(i)
```

```
       FreqObs(i - 1) = FreqObs(i - 1) + FreqObs(i)
       FreqExp(i) = 0
       FreqObs(i) = 0
       last = i - 1
ElseIf (FreqExp(i) < Minimum) And (FreqExp(i - 1) > 0) And (FreqExp(i - 1) >= FreqExp(i + 1)) Then
       FreqExp(i + 1) = FreqExp(i + 1) + FreqExp(i)
       FreqObs(i + 1) = FreqObs(i + 1) + FreqObs(i)
       FreqExp(i) = 0
       FreqObs(i) = 0
End If
If FreqExp(i) >= Minimum Then
     StrataRegrouped = StrataRegrouped + 1
     last = i
End If
Next i

If FreqExp(ObservedRows) < Minimum Then
   FreqExp(last) = FreqExp(last) + FreqExp(ObservedRows)
   FreqObs(last) = FreqObs(last) + FreqObs(ObservedRows)
   FreqExp(ObservedRows) = 0
   FreqObs(ObservedRows) = 0
Else
 StrataRegrouped = StrataRegrouped + 1
End If

'Control of sample size less than Minimum. No strata of minimum size.

If StrataRegrouped = 0 Then
MsgBox "No strata of minimum size"
GoTo final
End If

'Assignment of the number of the regrouped strata to the
'ChiSquaredTestStrataReg function.

ChiSquaredTestStrataReg = StrataRegrouped

' ****************** Omitted for iterative use (Solver) ******************
'Verification: Total of observed values is equal to the total of
'expected values.
'
'Dim totalobs As Double
'Dim totalexp As Double
'     For Each x In Observed
'         totalobs = totalobs + x.Value
'     Next x
'
'     For Each x In Expected
'         totalexp = totalexp + x.Value
'     Next x
'     If totalobs <> totalexp Then
'         MsgBox "Total observed does not match total expected"
'     Else
'     End If
'************************************************************************

'Initialization of frequency arrays.
Erase FreqExp()
Erase FreqObs()
Erase FreqExpReg()
Erase FreqObsReg()

final:

End Function
```

**Listing 4: VBA in Excel code. Observed and expected values**

```
Function ChiSquaredTesObsExpVal(Observed As Range, Expected As Range, _
Minimum As Integer) As Double()
'The observed and expected variables are declared as cell ranges
'and we use the Minimum variable to set the minimum number of
'elements for each strata.

'Statement of chi variable and observed and expected frequency
'matrices.

Dim values() As Double
Dim FreqExp() As Double
Dim FreqObs() As Double
'Correct Selection of Observed and Expected Data Range

     'By rows
     ObservedRows = Observed.Rows.Count
     ExpectedRows = Expected.Rows.Count

     'By columns
     ObservedColumns = Observed.Columns.Count
     ExpectedColumns = Expected.Columns.Count

     If ObservedColumns <> ExpectedColumns Then
     MsgBox "Incorrect selection"
     GoTo final
```

```
    End If

    If ObservedRows <> ExpectedRows Then
    MsgBox "Incorrect selection"
    GoTo final
    End If

'Resize the matrices from their actual dimension according to user data
'selection

ReDim FreqExp(ObservedRows)
ReDim FreqObs(ExpectedRows)
ReDim values(ObservedRows, 2)

'Initialization of variables to count strata of minimum size.

Dim StrataRegrouped As Integer
Dim last As Integer

StrataRegrouped = 0
last = 0
accumulator = 0

'Initial statement of arrays of regrouped frequencies for observed and
'expected values. Frequencies that will result from regrouping strata by
'requeriment of minimum size.

    'Expected
ne = 0
For Each x In Expected
ne = ne + 1
FreqExp(ne) = x.Value
Next x
    'Observed
no = 0
For Each x In Observed
no = no + 1
FreqObs(no) = x.Value
Next x

'Grouping in strata of size greater than or equal to Minimum.
'Prepare the array of values for subsequent assignment to the
'frequency matrix of the strata of valid size.

'add to the minor nearest. Start: first

If FreqExp(1) < Minimum Then
   FreqExp(2) = FreqExp(2) + FreqExp(1)
   FreqObs(2) = FreqObs(2) + FreqObs(1)
   FreqExp(1) = 0
   FreqObs(1) = 0
Else
   StrataRegrouped = StrataRegrouped + 1
   last = 1
End If

For i = 2 To ObservedRows - 1

If last > 0 Then
    If (FreqExp(i) < Minimum) And (FreqExp(i - 1) = 0) And (FreqExp(i + 1) > FreqExp(last)) Then
      FreqExp(last) = FreqExp(last) + FreqExp(i)
      FreqObs(last) = FreqObs(last) + FreqObs(i)
      FreqExp(i) = 0
      FreqObs(i) = 0
    ElseIf (FreqExp(i) < Minimum) And (FreqExp(i - 1) = 0) And (FreqExp(i + 1) <= FreqExp(last)) Then
      FreqExp(i + 1) = FreqExp(i + 1) + FreqExp(i)
      FreqObs(i + 1) = FreqObs(i + 1) + FreqObs(i)
      FreqExp(i) = 0
      FreqObs(i) = 0
    End If
End If
If (FreqExp(i) < Minimum) And (FreqExp(i - 1) = 0) And (last = 0) Then
      FreqExp(i + 1) = FreqExp(i + 1) + FreqExp(i)
      FreqObs(i + 1) = FreqObs(i + 1) + FreqObs(i)
      FreqExp(i) = 0
      FreqObs(i) = 0
End If
If (FreqExp(i) < Minimum) And (FreqExp(i - 1) > 0) And (FreqExp(i - 1) < FreqExp(i + 1)) Then
      FreqExp(i - 1) = FreqExp(i - 1) + FreqExp(i)
      FreqObs(i - 1) = FreqObs(i - 1) + FreqObs(i)
      FreqExp(i) = 0
      FreqObs(i) = 0
      last = i - 1
ElseIf (FreqExp(i) < Minimum) And (FreqExp(i - 1) > 0) And (FreqExp(i - 1) >= FreqExp(i + 1)) Then
      FreqExp(i + 1) = FreqExp(i + 1) + FreqExp(i)
      FreqObs(i + 1) = FreqObs(i + 1) + FreqObs(i)
      FreqExp(i) = 0
      FreqObs(i) = 0
End If
If FreqExp(i) >= Minimum Then
    StrataRegrouped = StrataRegrouped + 1
    last = i
End If
Next i

If FreqExp(ObservedRows) < Minimum Then
```

```
    FreqExp(last) = FreqExp(last) + FreqExp(ObservedRows)
    FreqObs(last) = FreqObs(last) + FreqObs(ObservedRows)
    FreqExp(ObservedRows) = 0
    FreqObs(ObservedRows) = 0
Else
 StrataRegrouped = StrataRegrouped + 1
End If

'Control of sample size less than Minimum. No strata of minimum size.

If StrataRegrouped = 0 Then
MsgBox "No strata of minimum size"
GoTo final
End If

'Assignment of the regrouped strata values to the "values" array.

For i = 1 To ObservedRows
values(i, 1) = FreqObs(i)
values(i, 2) = FreqExp(i)
Next i

'Assignment of the "values" array to the ChiSquaredTesObsExpVal function.

ChiSquaredTesObsExpVal = values

' ***************** Omitted for iterative use (Solver) *****************
'Verification: Total of observed values is equal to the total of
'expected values.
'
'Dim totalobs As Double
'Dim totalexp As Double
'    For Each x In Observed
'        totalobs = totalobs + x.Value
'    Next x
'
'    For Each x In Expected
'        totalexp = totalexp + x.Value
'    Next x
'    If totalobs <> totalexp Then
'        MsgBox "Total observed does not match total expected"
'    Else
'    End If
'**********************************************************************

'Initialization of frequency arrays.

Erase FreqExp()
Erase FreqObs()
Erase values()
final:

End Function
```

## Listing 5: Mathematica code. Number of regrouped strata

```
PearsonTestStrata[exp_List, min_Integer] :=
  Module[{last, strata, i, stratareg}, {strata = Length[exp];
    expSUB = exp;
    last = 0;
    {{If[expSUB[[1]] < min, {expSUB[[2]] = expSUB[[2]] + expSUB[[1]],
        expSUB[[1]] = 0}, last = 1]}, {For[i = 2, i <= strata - 1, i++,
       {{If[
          last > 0, {{If[(expSUB[[i]] <
                min) \[And] (expSUB[[i - 1]] ==
                0) \[And] (expSUB[[i + 1]] >
                expSUB[[last]]), {expSUB[[last]] =
               expSUB[[last]] + expSUB[[i]],
              expSUB[[i]] =
               0}]}, {If[(expSUB[[i]] <
                min) \[And] (expSUB[[i - 1]] ==
                0) \[And] (expSUB[[i + 1]] <=
                expSUB[[last]]), {expSUB[[i + 1]] =
               expSUB[[i + 1]] + expSUB[[i]],
              expSUB[[i]] = 0}]}}], {If[(expSUB[[i]] <
              min) \[And] (expSUB[[i - 1]] == 0) \[And] (last ==
              0), {expSUB[[i + 1]] = expSUB[[i + 1]] + expSUB[[i]],
            expSUB[[i]] =
             0}]}, {If[(expSUB[[i]] < min) \[And] (expSUB[[i - 1]] >
              0) \[And] (expSUB[[i - 1]] <
              expSUB[[i + 1]]), {expSUB[[i - 1]] =
             expSUB[[i - 1]] + expSUB[[i]], expSUB[[i]] = 0,
            last = i - 1}]}, {If[(expSUB[[i]] <
              min) \[And] (expSUB[[i - 1]] >
              0) \[And] (expSUB[[i - 1]] >=
              expSUB[[i + 1]]), {expSUB[[i + 1]] =
             expSUB[[i + 1]] + expSUB[[i]], expSUB[[i]] = 0}]}, {If[
           expSUB[[i]] >= min, last = i ]}}]}, {If[
        expSUB[[strata]] <
         min, {expSUB[[last]] = expSUB[[last]] + expSUB[[strata]],
         expSUB[[strata]] = 0}]}};
    expreg = {};
    For[i = 1, i <= strata, i++,
```

```
    If[expSUB[[i]] >= min, AppendTo[expreg, expSUB[[i]]]]];
   stratareg = Length[expreg]}; Label[End];
  Return[stratareg]];
```

## Listing 6: Mathematica code. P-value

```
PearsonTestpValue[obs_List, exp_List, min_Integer] /;
  Length[obs] == Length[exp] :=
 Module[{last, chi, strata, i, stratareg,
   pvalue}, {strata = Length[exp];
   obsSUB = obs;
   expSUB = exp;
   last = 0;
   {{If[expSUB[[1]] < min, {expSUB[[2]] = expSUB[[2]] + expSUB[[1]],
       obsSUB[[2]] = obsSUB[[2]] + obsSUB[[1]], expSUB[[1]] = 0,
       obsSUB[[1]] = 0}, last = 1]}, {For[i = 2, i <= strata - 1, i++,
      {{If[
          last > 0, {{If[(expSUB[[i]] < min) \[And] (expSUB[[i - 1]] ==
                 0) \[And] (expSUB[[i + 1]] >
                expSUB[[last]]), {expSUB[[last]] =
               expSUB[[last]] + expSUB[[i]],
              obsSUB[[last]] = obsSUB[[last]] + obsSUB[[i]],
              expSUB[[i]] = 0,
              obsSUB[[i]] =
               0}]}, {If[(expSUB[[i]] < min) \[And] (expSUB[[i - 1]] ==
                 0) \[And] (expSUB[[i + 1]] <=
                expSUB[[last]]), {expSUB[[i + 1]] =
               expSUB[[i + 1]] + expSUB[[i]],
              obsSUB[[i + 1]] = obsSUB[[i + 1]] + obsSUB[[i]],
              expSUB[[i]] = 0,
              obsSUB[[i]] = 0}]}}]}, {If[(expSUB[[i]] <
             min) \[And] (expSUB[[i - 1]] == 0) \[And] (last ==
             0), {expSUB[[i + 1]] = expSUB[[i + 1]] + expSUB[[i]],
           obsSUB[[i + 1]] = obsSUB[[i + 1]] + obsSUB[[i]],
           expSUB[[i]] = 0,
           obsSUB[[i]] =
            0}]}, {If[(expSUB[[i]] < min) \[And] (expSUB[[i - 1]] >
             0) \[And] (expSUB[[i - 1]] <
             expSUB[[i + 1]]), {expSUB[[i - 1]] =
            expSUB[[i - 1]] + expSUB[[i]],
           obsSUB[[i - 1]] = obsSUB[[i - 1]] + obsSUB[[i]],
           expSUB[[i]] = 0, obsSUB[[i]] = 0,
           last = i - 1}]}, {If[(expSUB[[i]] <
             min) \[And] (expSUB[[i - 1]] >
             0) \[And] (expSUB[[i - 1]] >=
             expSUB[[i + 1]]), {expSUB[[i + 1]] =
            expSUB[[i + 1]] + expSUB[[i]],
           obsSUB[[i + 1]] = obsSUB[[i + 1]] + obsSUB[[i]],
           expSUB[[i]] = 0, obsSUB[[i]] = 0}]}, {If[expSUB[[i]] >= min,
           last = i]}}]}, {If[
      expSUB[[strata]] <
       min, {expSUB[[last]] = expSUB[[last]] + expSUB[[strata]],
       obsSUB[[last]] = obsSUB[[last]] + obsSUB[[strata]],
       expSUB[[strata]] = 0, obsSUB[[strata]] = 0}]}};
   expreg = {}; obsreg = {};
   For[i = 1, i <= strata, i++,
    If[expSUB[[i]] >= min, {AppendTo[expreg, expSUB[[i]]],
      AppendTo[obsreg, obsSUB[[i]]]}]];
   chi = N[Total[(obsreg - expreg)^2/expreg]];
   stratareg = Length[expreg];
   pvalue =
    SurvivalFunction[ChiSquareDistribution[stratareg - 1], chi]};
  If[Total[exp] != Total[obs],
   Print["Total observed does not match total expected. ",
    " p-value ", pvalue], Return[pvalue]]]
```

## Listing 7: Mathematica code. Summary of Chi-square Test results

```
PearsonTest[obs_List, exp_List, min_Integer] /;
  Length[obs] == Length[exp] :=
 Module[{last, chi, strata, i, stratareg,
   pvalue}, {strata = Length[exp];
   obsSUB = obs;
   expSUB = exp;
   last = 0;
   {{If[expSUB[[1]] < min, {expSUB[[2]] = expSUB[[2]] + expSUB[[1]],
       obsSUB[[2]] = obsSUB[[2]] + obsSUB[[1]], expSUB[[1]] = 0,
       obsSUB[[1]] = 0}, last = 1]}, {For[i = 2, i <= strata - 1, i++,
      {{If[
          last > 0, {{If[(expSUB[[i]] < min) \[And] (expSUB[[i - 1]] ==
                 0) \[And] (expSUB[[i + 1]] >
                expSUB[[last]]), {expSUB[[last]] =
               expSUB[[last]] + expSUB[[i]],
              obsSUB[[last]] = obsSUB[[last]] + obsSUB[[i]],
              expSUB[[i]] = 0,
              obsSUB[[i]] =
               0}]}, {If[(expSUB[[i]] < min) \[And] (expSUB[[i - 1]] ==
                 0) \[And] (expSUB[[i + 1]] <=
                expSUB[[last]]), {expSUB[[i + 1]] =
               expSUB[[i + 1]] + expSUB[[i]],
```

```
            obsSUB[[i + 1]] = obsSUB[[i + 1]] + obsSUB[[i]],
            expSUB[[i]] = 0,
            obsSUB[[i]] = 0}]}}]}, {If[(expSUB[[i]] <
           min) \[And] (expSUB[[i - 1]] == 0) \[And] (last ==
           0), {expSUB[[i + 1]] = expSUB[[i + 1]] + expSUB[[i]],
         obsSUB[[i + 1]] = obsSUB[[i + 1]] + obsSUB[[i]],
         expSUB[[i]] = 0,
         obsSUB[[i]] =
          0}]}, {If[(expSUB[[i]] < min) \[And] (expSUB[[i - 1]] >
           0) \[And] (expSUB[[i - 1]] <
           expSUB[[i + 1]]), {expSUB[[i - 1]] =
          expSUB[[i - 1]] + expSUB[[i]],
         obsSUB[[i - 1]] = obsSUB[[i - 1]] + obsSUB[[i]],
         expSUB[[i]] = 0, obsSUB[[i]] = 0,
         last = i - 1}]}, {If[(expSUB[[i]] <
           min) \[And] (expSUB[[i - 1]] >
           0) \[And] (expSUB[[i - 1]] >=
           expSUB[[i + 1]]), {expSUB[[i + 1]] =
          expSUB[[i + 1]] + expSUB[[i]],
         obsSUB[[i + 1]] = obsSUB[[i + 1]] + obsSUB[[i]],
         expSUB[[i]] = 0, obsSUB[[i]] = 0}]}, {If[expSUB[[i]] >= min,
         last = i ]}}]}, {If[
     expSUB[[strata]] <
      min, {expSUB[[last]] = expSUB[[last]] + expSUB[[strata]],
      obsSUB[[last]] = obsSUB[[last]] + obsSUB[[strata]],
      expSUB[[strata]] = 0, obsSUB[[strata]] = 0}]}};
  expreg = {}; obsreg = {};
  For[i = 1, i <= strata, i++,
   If[expSUB[[i]] >= min, {AppendTo[expreg, expSUB[[i]]],
     AppendTo[obsreg, obsSUB[[i]]]}]];
  chi = N[Total[(obsreg - expreg)^2/expreg]];
  stratareg = Length[expreg];
  pvalue =
   SurvivalFunction[ChiSquareDistribution[stratareg - 1], chi]};
 If[Total[exp] != Total[obs],
  Print["Total observed does not match total expected. ",
   " Chi statistic ", chi, " p-value ", pvalue,
   " Observed values regrouped ", obsreg,
   " Expected values regrouped ", expreg],
  Rule[{"Chi statistic ", " p-value ", " Observed values regrouped ",
     " Expected values regrouped "}, {chi, pvalue, obsreg, expreg}]]]
```

**Table A2.** *Results reported for 10 different multinomial populations, where the theoretical probabilities under the null hypothesis are described in the top part of the table for the different $k = 10$ classes considered for the simulation study. In this case, 5000 simulations from each population were simulated for $N = 50$ and $k = 10$, and three different nominal significance levels considered ($\alpha$=0.10, 0.05 and 0.01). Significance levels attained by using the procedure without regrouping and those attained using the regrouping procedure proposed here are reported at the bottom of the table for each nominal significance level in the study for the fully specified chi-square goodness-of-fit test.*

| **population distribution** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **k** | **popul. 1** | **popul. 3** | **popul. 4** | **popul. 8** | **popul. 9** | **popul. 12** | **popul. 14** | **popul. 15** | **popul. 16** | **popul. 18** |
| 1 | 0.067722412 | 0.093438111 | 0.099092428 | 0.055167980 | 0.126712503 | 0.158331638 | 0.129522785 | 0.155141773 | 0.129108184 | 0.154908851 |
| 2 | 0.108251949 | 0.039193314 | 0.143654215 | 0.170135074 | 0.175667913 | 0.054474180 | 0.101468142 | 0.117873535 | 0.134484690 | 0.064089325 |
| 3 | 0.059166912 | 0.088253560 | 0.090975941 | 0.058609862 | 0.086016774 | 0.142045200 | 0.058832579 | 0.163627763 | 0.021375951 | 0.166329069 |
| 4 | 0.161665856 | 0.068734338 | 0.085228675 | 0.037666222 | 0.011613805 | 0.104115288 | 0.100833318 | 0.175329030 | 0.018903249 | 0.132811973 |
| 5 | 0.118838198 | 0.145124529 | 0.168663482 | 0.034227447 | 0.101429894 | 0.077007347 | 0.067901927 | 0.029351531 | 0.055381409 | 0.077241285 |
| 6 | 0.149585825 | 0.125514219 | 0.070588045 | 0.103546999 | 0.210875342 | 0.089579279 | 0.168068825 | 0.034844052 | 0.182953566 | 0.053626300 |
| 7 | 0.137791080 | 0.158950344 | 0.133376176 | 0.176459534 | 0.021765578 | 0.023434411 | 0.015373940 | 0.133862115 | 0.127970072 | 0.035629231 |
| 8 | 0.059378892 | 0.119973253 | 0.041377815 | 0.159305809 | 0.044490360 | 0.031816071 | 0.158329630 | 0.055992935 | 0.145612819 | 0.153044117 |
| 9 | 0.083716990 | 0.011236121 | 0.003529199 | 0.133811376 | 0.115097525 | 0.160946852 | 0.162414523 | 0.045883570 | 0.055317926 | 0.066055454 |
| 10 | 0.053881886 | 0.149582209 | 0.163514025 | 0.071069698 | 0.106330305 | 0.158249734 | 0.03725433 | 0.088093695 | 0.128892133 | 0.096264395 |
| **nominal significance level. $\alpha$=10%** | | | | | | | | | | |
| do not regroup | 0.097 | 0.102 | 0.103 | 0.099 | 0.099 | 0.097 | 0.097 | 0.092 | 0.102 | 0.099 |
| regroup | 0.103 | 0.096 | 0.100 | 0.097 | 0.099 | 0.101 | 0.095 | 0.090 | 0.099 | 0.096 |
| **nominal significance level. $\alpha$=5%** | | | | | | | | | | |
| do not regroup | 0.048 | 0.054 | 0.055 | 0.053 | 0.050 | 0.054 | 0.050 | 0.048 | 0.050 | 0.051 |
| regroup | 0.054 | 0.046 | 0.048 | 0.049 | 0.048 | 0.052 | 0.047 | 0.041 | 0.051 | 0.050 |
| **nominal significance level. $\alpha$=1%** | | | | | | | | | | |
| do not regroup | 0.012 | 0.013 | 0.018 | 0.018 | 0.015 | 0.017 | 0.014 | 0.013 | 0.013 | 0.013 |
| regroup | 0.011 | 0.011 | 0.009 | 0.012 | 0.009 | 0.012 | 0.009 | 0.008 | 0.011 | 0.009 |

**Table A3.** *Results reported for 10 selected different multinomial populations, where the theoretical probabilities under the null hypothesis are described in the top part of the table for the different $k = 15$ classes considered for the simulation study. In this case, 5000 simulations from each population were simulated for $N = 75$ and $k = 15$, and three different nominal significance levels considered ($\alpha$=0.10, 0.05 and 0.01). Significance levels attained by using the procedure without regrouping and those attained using the regrouping procedure proposed here are reported at the bottom of the table for each nominal significance level in the study for the fully specified chi-square goodness-of-fit test.*

| | population distribution | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **k** | **popul. 1** | **popul. 3** | **popul. 4** | **popul. 8** | **popul. 9** | **popul. 12** | **popul. 14** | **popul. 15** | **popul. 16** | **popul. 18** |
| 1 | 0.127057671 | 0.083861803 | 0.081106551 | 0.072320794 | 0.078080154 | 0.029394665 | 0.118093818 | 0.131915150 | 0.124764161 | 0.101425616 |
| 2 | 0.088104527 | 0.067347127 | 0.063163406 | 0.031832989 | 0.091721146 | 0.057298124 | 0.064924762 | 0.137187332 | 0.084398252 | 0.054144288 |
| 3 | 0.029902404 | 0.085171458 | 0.035443425 | 0.062531848 | 0.002989394 | 0.050403893 | 0.085464849 | 0.092797832 | 0.074048405 | 0.083714805 |
| 4 | 0.008582462 | 0.024179898 | 0.090621898 | 0.006554065 | 0.009238380 | 0.117528851 | 0.088907182 | 0.013390771 | 0.032826123 | 0.033579822 |
| 5 | 0.113134725 | 0.110562508 | 0.043446425 | 0.085630406 | 0.132330601 | 0.000768729 | 0.081224262 | 0.077103919 | 0.031323120 | 0.069563207 |
| 6 | 0.007700507 | 0.042833764 | 0.079690194 | 0.050492072 | 0.045150829 | 0.021549876 | 0.008446190 | 0.154708685 | 0.001610057 | 0.079027830 |
| 7 | 0.130266721 | 0.036978384 | 0.054341777 | 0.005275661 | 0.119958130 | 0.086208961 | 0.044215657 | 0.049446275 | 0.017379203 | 0.058223969 |
| 8 | 0.043228436 | 0.039777167 | 0.055824977 | 0.037389853 | 0.037573870 | 0.058081580 | 0.105879702 | 0.025265570 | 0.032805863 | 0.008089689 |
| 9 | 0.080676242 | 0.112476617 | 0.079526539 | 0.016972967 | 0.095983503 | 0.120659949 | 0.027606531 | 0.027259488 | 0.103095726 | 0.010500826 |
| 10 | 0.094468872 | 0.094139849 | 0.062781085 | 0.078659953 | 0.124772249 | 0.097925781 | 0.064328900 | 0.145884761 | 0.034570874 | 0.105571614 |
| 11 | 0.023806346 | 0.084163565 | 0.108837039 | 0.107596687 | 0.031639086 | 0.028899879 | 0.027285188 | 0.063096567 | 0.106607857 | 0.058341183 |
| 12 | 0.115757204 | 0.029147017 | 0.025956871 | 0.144561350 | 0.037859485 | 0.096476790 | 0.119112740 | 0.011338246 | 0.093203420 | 0.120721080 |
| 13 | 0.107141421 | 0.101775293 | 0.103602266 | 0.054244863 | 0.121044273 | 0.109784646 | 0.027558602 | 0.021700713 | 0.046306240 | 0.124422616 |
| 14 | 0.017980943 | 0.025664489 | 0.077737739 | 0.118268377 | 0.034410514 | 0.091606477 | 0.051464298 | 0.030177488 | 0.096164501 | 0.091153404 |
| 15 | 0.012191519 | 0.061921062 | 0.037919809 | 0.127668115 | 0.037248387 | 0.033411800 | 0.085487321 | 0.018727203 | 0.120896197 | 0.001520050 |
| **nominal significance level. $\alpha$=10%** | | | | | | | | | | |
| do not regroup | 0.110 | 0.098 | 0.099 | 0.109 | 0.107 | 0.123 | 0.095 | 0.107 | 0.108 | 0.114 |
| regroup | 0.091 | 0.093 | 0.096 | 0.102 | 0.096 | 0.105 | 0.092 | 0.101 | 0.102 | 0.090 |
| **nominal significance level. $\alpha$=5%** | | | | | | | | | | |
| do not regroup | 0.059 | 0.050 | 0.047 | 0.061 | 0.061 | 0.084 | 0.053 | 0.057 | 0.059 | 0.065 |
| regroup | 0.045 | 0.044 | 0.046 | 0.050 | 0.048 | 0.053 | 0.046 | 0.049 | 0.048 | 0.042 |
| **nominal significance level. $\alpha$=1%** | | | | | | | | | | |
| do not regroup | 0.017 | 0.014 | 0.009 | 0.017 | 0.019 | 0.030 | 0.011 | 0.014 | 0.015 | 0.020 |
| regroup | 0.009 | 0.012 | 0.010 | 0.009 | 0.010 | 0.013 | 0.010 | 0.001 | 0.009 | 0.008 |

**Table A4.** *Results reported for 10 selected different multinomial populations, where the theoretical probabilities under the null hypothesis are described in the top part of the table for the different $k = 20$ classes considered for the simulation study. In this case, 5000 simulations from each population were simulated for $N = 100$ and $k = 20$, and three different nominal significance levels considered ($\alpha$=0.10, 0.05 and 0.01). Significance levels attained by using the procedure without regrouping and those attained using the regrouping procedure proposed here are reported at the bottom of the table for each nominal significance levels in the study for the fully specified chi-square goodness-of-fit test.*

| **population distribution** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **k** | **popul. 1** | **popul. 3** | **popul. 4** | **popul. 8** | **popul. 9** | **popul. 12** | **popul. 14** | **popul. 15** | **popul. 16** | **popul. 18** |
| 1 | 0.009906949 | 0.042585801 | 0.036322397 | 0.023803358 | 0.056050361 | 0.048540316 | 0.048978727 | 0.014696213 | 0.004618549 | 0.020337923 |
| 2 | 0.043334958 | 0.084595131 | 0.073420820 | 0.064449053 | 0.043706663 | 0.085501927 | 0.054249217 | 0.049354784 | 0.045174755 | 0.012077438 |
| 3 | 0.068751733 | 0.055949091 | 0.065878707 | 0.037395221 | 0.035956904 | 0.056738810 | 0.062717404 | 0.038945982 | 0.006326356 | 0.012472135 |
| 4 | 0.075322010 | 0.088903934 | 0.093919822 | 0.080488103 | 0.012855218 | 0.082229679 | 0.057512461 | 0.087900507 | 0.104546576 | 0.089286075 |
| 5 | 0.084464071 | 0.089215586 | 0.091398883 | 0.018016210 | 0.085894992 | 0.021013929 | 0.067902855 | 0.082246443 | 0.054052148 | 0.007489974 |
| 6 | 0.055848915 | 0.063276437 | 0.021928939 | 0.036960583 | 0.000994794 | 0.011044924 | 0.037607372 | 0.064129728 | 0.077098867 | 0.000477913 |
| 7 | 0.033849691 | 0.007263521 | 0.064953627 | 0.069011621 | 0.055694252 | 0.093010187 | 0.041731769 | 0.019967383 | 0.110322256 | 0.059307357 |
| 8 | 0.041782157 | 0.015192684 | 0.004878198 | 0.006500386 | 0.038241324 | 0.008823266 | 0.066635437 | 0.037496205 | 0.044073167 | 0.092347960 |
| 9 | 0.049615947 | 0.009841069 | 0.009040455 | 0.094543419 | 0.028413965 | 0.099002066 | 0.074037397 | 0.062557226 | 0.094012295 | 0.081521459 |
| 10 | 0.054796141 | 0.001109167 | 0.090830640 | 0.047526405 | 0.072759348 | 0.039090212 | 0.057435627 | 0.033694948 | 0.027255678 | 0.012003245 |
| 11 | 0.049573297 | 0.045490837 | 0.070485338 | 0.022710569 | 0.092939405 | 0.070205677 | 0.054643564 | 0.035830706 | 0.102879776 | 0.064882703 |
| 12 | 0.034362080 | 0.011503208 | 0.012250549 | 0.077123015 | 0.087869027 | 0.051263015 | 0.046862814 | 0.061074007 | 0.002815851 | 0.065966882 |
| 13 | 0.003321141 | 0.082512233 | 0.036884669 | 0.047694596 | 0.094458297 | 0.053229198 | 0.030201345 | 0.056240124 | 0.006559463 | 0.006896735 |
| 14 | 0.035231977 | 0.028957298 | 0.021443815 | 0.071919102 | 0.009911000 | 0.005160299 | 0.016883433 | 0.011740128 | 0.002543542 | 0.055721211 |
| 15 | 0.096938721 | 0.058782500 | 0.060608656 | 0.064842833 | 0.042211503 | 0.051071775 | 0.044675117 | 0.057298982 | 0.035805897 | 0.044729954 |
| 16 | 0.079250776 | 0.010929522 | 0.058052934 | 0.006038138 | 0.073256785 | 0.003184016 | 0.031371077 | 0.075481374 | 0.070503856 | 0.082763910 |
| 17 | 0.032532973 | 0.090748245 | 0.089832763 | 0.087521877 | 0.023712180 | 0.069816537 | 0.072661166 | 0.038469940 | 0.089473563 | 0.068822077 |
| 18 | 0.068855512 | 0.033207256 | 0.002387938 | 0.054802997 | 0.083944166 | 0.088130478 | 0.057917558 | 0.093425379 | 0.013317647 | 0.086962221 |
| 19 | 0.019344094 | 0.104381792 | 0.043173925 | 0.007385808 | 0.009079374 | 0.014168915 | 0.028620055 | 0.050048690 | 0.046279136 | 0.089193171 |
| 20 | 0.062916859 | 0.075554687 | 0.052306925 | 0.081266705 | 0.052050441 | 0.048774773 | 0.047355605 | 0.029401252 | 0.062340620 | 0.046739658 |
| **nominal significance level. $\alpha$=10%** | | | | | | | | | | |
| not regroup | 0.093 | 0.110 | 0.112 | 0.112 | 0.106 | 0.103 | 0.093 | 0.103 | 0.122 | 0.124 |
| regroup | 0.097 | 0.094 | 0.095 | 0.098 | 0.099 | 0.097 | 0.095 | 0.093 | 0.100 | 0.092 |
| **nominal significance level. $\alpha$=5%** | | | | | | | | | | |
| not regroup | 0.050 | 0.068 | 0.060 | 0.061 | 0.064 | 0.055 | 0.047 | 0.051 | 0.068 | 0.082 |
| regroup | 0.048 | 0.048 | 0.047 | 0.050 | 0.050 | 0.050 | 0.046 | 0.043 | 0.049 | 0.043 |
| **nominal significance level. $\alpha$=1%** | | | | | | | | | | |
| not regroup | 0.013 | 0.023 | 0.019 | 0.018 | 0.020 | 0.020 | 0.010 | 0.011 | 0.027 | 0.035 |
| regroup | 0.009 | 0.009 | 0.009 | 0.010 | 0.010 | 0.010 | 0.008 | 0.009 | 0.012 | 0.007 |

**Table A5.** *Results reported for 10 selected different multinomial populations, where the theoretical probabilities under the null hypothesis are described in the top part of the table for the different $k = 20$ classes considered for the simulation study. In this case, 5000 simulations from each population were simulated for $N = 250$ and $k = 20$, and three different nominal significance levels considered ($\alpha$=0.10, 0.05 and 0.01). Significance levels attained by using the procedure without regrouping and those attained using the regrouping procedure proposed here are reported at the bottom of the table for each nominal significance level in the study for the fully specified chi-square goodness-of-fit test.*

| | population distribution | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **k** | **assig. 1** | **assig. 3** | **popul. 4** | **popul. 8** | **popul. 9** | **popul. 12** | **popul. 14** | **popul. 15** | **popul. 16** | **popul. 18** |
| 1 | 0.047643128 | 0.054666723 | 0.056038758 | 0.090599483 | 0.025317013 | 0.023099063 | 0.064584061 | 0.136478476 | 0.043123333 | 0.054066459 |
| 2 | 0.024791212 | 0.044261848 | 0.032569245 | 0.044497446 | 0.014610403 | 0.084334948 | 0.009032926 | 0.014084696 | 0.072955298 | 0.053573431 |
| 3 | 0.054639637 | 0.071340360 | 0.073471657 | 0.049842681 | 0.020683555 | 0.037648781 | 0.066675056 | 0.006032668 | 0.034107228 | 0.013291318 |
| 4 | 0.011551957 | 0.035790188 | 0.017892204 | 0.048986190 | 0.015986321 | 0.061877989 | 0.023052570 | 0.051462288 | 0.069213717 | 0.045963934 |
| 5 | 0.043209785 | 0.068393517 | 0.060880791 | 0.038016856 | 0.078291674 | 0.015100046 | 0.085352639 | 0.015104595 | 0.055712777 | 0.094207824 |
| 6 | 0.058607724 | 0.084149275 | 0.040842165 | 0.082739121 | 0.059420758 | 0.012071383 | 0.011090706 | 0.077633225 | 0.038688100 | 0.023326840 |
| 7 | 0.004077781 | 0.036377709 | 0.062850108 | 0.075899205 | 0.050698562 | 0.072435772 | 0.049065314 | 0.072825503 | 0.083729668 | 0.047299352 |
| 8 | 0.085955934 | 0.066886372 | 0.073347443 | 0.035307555 | 0.045219825 | 0.044208957 | 0.055207226 | 0.019632440 | 0.056600081 | 0.018773053 |
| 9 | 0.023369288 | 0.015630083 | 0.032843147 | 0.026278327 | 0.058397084 | 0.032659614 | 0.083467899 | 0.016586480 | 0.010340800 | 0.050770889 |
| 10 | 0.087652522 | 0.007532102 | 0.054125222 | 0.034259425 | 0.044017966 | 0.061446558 | 0.039048538 | 0.069347828 | 0.000467925 | 0.095008505 |
| 11 | 0.095771275 | 0.057965331 | 0.055877398 | 0.011887980 | 0.088882200 | 0.029313393 | 0.063182257 | 0.045379877 | 0.079491665 | 0.001551221 |
| 12 | 0.096306521 | 0.030974889 | 0.036689952 | 0.006101936 | 0.044860170 | 0.068419459 | 0.062127872 | 0.119150362 | 0.038817979 | 0.013723146 |
| 13 | 0.080897006 | 0.087201505 | 0.004498861 | 0.029781475 | 0.094173221 | 0.009146832 | 0.018611673 | 0.017167618 | 0.063579053 | 0.040856797 |
| 14 | 0.013705065 | 0.070556907 | 0.100690100 | 0.059337269 | 0.017035620 | 0.083251916 | 0.051065504 | 0.033760507 | 0.011793975 | 0.071820279 |
| 15 | 0.024205251 | 0.014659583 | 0.060392354 | 0.087250033 | 0.067126421 | 0.059621249 | 0.048932496 | 0.001487572 | 0.063205970 | 0.098332826 |
| 16 | 0.004839018 | 0.087276954 | 0.007352012 | 0.074132698 | 0.079880521 | 0.091801986 | 0.023883012 | 0.075759621 | 0.016191521 | 0.021901263 |
| 17 | 0.060014816 | 0.068505786 | 0.040949294 | 0.063520955 | 0.093936062 | 0.032076674 | 0.029721620 | 0.096471797 | 0.075764151 | 0.060809674 |
| 18 | 0.037409444 | 0.050921813 | 0.080223679 | 0.007014358 | 0.004676127 | 0.060818519 | 0.056226443 | 0.055834461 | 0.092658366 | 0.034348079 |
| 19 | 0.071116566 | 0.043564240 | 0.034035218 | 0.082594193 | 0.036804493 | 0.082172574 | 0.077824913 | 0.006284488 | 0.082035648 | 0.102590301 |
| 20 | 0.074236070 | 0.003344816 | 0.074430393 | 0.051952812 | 0.059982005 | 0.038494286 | 0.081847276 | 0.069515496 | 0.011522743 | 0.057784808 |
| | **nominal significance level. $\alpha$=10%** | | | | | | | | | |
| do not regroup | 0.099 | 0.101 | 0.106 | 0.096 | 0.097 | 0.095 | 0.105 | 0.102 | 0.109 | 0.104 |
| regroup | 0.093 | 0.101 | 0.104 | 0.097 | 0.097 | 0.098 | 0.100 | 0.098 | 0.104 | 0.102 |
| | **nominal significance level. $\alpha$=5%** | | | | | | | | | |
| do not regroup | 0.054 | 0.055 | 0.052 | 0.050 | 0.051 | 0.052 | 0.052 | 0.053 | 0.059 | 0.057 |
| regroup | 0.050 | 0.053 | 0.051 | 0.050 | 0.046 | 0.048 | 0.049 | 0.051 | 0.054 | 0.049 |
| | **nominal significance level. $\alpha$=1%** | | | | | | | | | |
| do not regroup | 0.013 | 0.014 | 0.013 | 0.011 | 0.011 | 0.011 | 0.011 | 0.017 | 0.019 | 0.015 |
| regroup | 0.01 | 0.013 | 0.011 | 0.010 | 0.010 | 0.009 | 0.009 | 0.010 | 0.010 | 0.013 |

**Table A6.** *Results reported for 10 selected different multinomial populations, where the theoretical probabilities under the null hypothesis are described in the top part of the table for the different $k = 20$ classes considered for the simulation study. In this case, 5000 simulations from each population were simulated for $N = 500$ and $k = 20$, and three different nominal significance levels considered ($\alpha$=0.10, 0.05 and 0.01). Significance levels attained by using the procedure without regrouping and those attained using the regrouping procedure proposed here are reported at the bottom of the table for each nominal significance level in the study for the fully specified chi-square goodness-of-fit test.*

| | **population distribution** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **k** | **popul. 1** | **popul. 3** | **popul. 4** | **popul. 8** | **popul. 9** | **popul. 12** | **popul. 14** | **popul. 15** | **popul. 16** | **popul. 18** |
| 1 | 0.009323228 | 0.021842441 | 0.073053675 | 0.040873644 | 0.034371818 | 0.004343310 | 0.077050114 | 0.009333736 | 0.089174555 | 0.077053926 |
| 2 | 0.073419610 | 0.084240115 | 0.043107814 | 0.042528243 | 0.090126015 | 0.092113419 | 0.078204289 | 0.067545345 | 0.093289019 | 0.032029821 |
| 3 | 0.064476239 | 0.067584721 | 0.053896718 | 0.030820872 | 0.011727517 | 0.084304086 | 0.054005211 | 0.063428357 | 0.053021839 | 0.082081999 |
| 4 | 0.011049151 | 0.088536877 | 0.024211296 | 0.027384842 | 0.069805333 | 0.028693635 | 0.092724143 | 0.067470149 | 0.019683922 | 0.055045204 |
| 5 | 0.086457614 | 0.055072657 | 0.071720992 | 0.089111120 | 0.069667082 | 0.011893892 | 0.055181544 | 0.006367421 | 0.059474204 | 0.063753881 |
| 6 | 0.022503868 | 0.016163729 | 0.060949072 | 0.091777222 | 0.062247990 | 0.067694023 | 0.016575229 | 0.097107691 | 0.099925118 | 0.039857245 |
| 7 | 0.000016554 | 0.068682807 | 0.070000441 | 0.085376495 | 0.071095031 | 0.090696516 | 0.026952117 | 0.002052394 | 0.068436458 | 0.059934452 |
| 8 | 0.076067786 | 0.058131169 | 0.047627730 | 0.052631881 | 0.014714001 | 0.062019278 | 0.085712216 | 0.018122689 | 0.005818399 | 0.050251930 |
| 9 | 0.044093494 | 0.002919132 | 0.013261422 | 0.017399972 | 0.058380580 | 0.051731317 | 0.009912796 | 0.023110315 | 0.021480560 | 0.023071720 |
| 10 | 0.008782610 | 0.062675614 | 0.060953727 | 0.035420538 | 0.088536843 | 0.060798804 | 0.017801328 | 0.085190935 | 0.083247679 | 0.054008372 |
| 11 | 0.100593141 | 0.100396950 | 0.044992803 | 0.069148810 | 0.055528229 | 0.045118418 | 0.056649822 | 0.087104011 | 0.016148816 | 0.006887184 |
| 12 | 0.097205062 | 0.007259403 | 0.051243758 | 0.101523827 | 0.044649792 | 0.040196511 | 0.069997087 | 0.019445299 | 0.007289295 | 0.077411088 |
| 13 | 0.052323470 | 0.037830917 | 0.031323124 | 0.046548434 | 0.093038962 | 0.049782978 | 0.021150533 | 0.104840295 | 0.063517778 | 0.013079696 |
| 14 | 0.018032175 | 0.025567586 | 0.055357326 | 0.003155478 | 0.057561203 | 0.076744301 | 0.039391817 | 0.003105487 | 0.090649719 | 0.052608559 |
| 15 | 0.068130807 | 0.108793483 | 0.017934527 | 0.031492695 | 0.062140420 | 0.000983061 | 0.067977053 | 0.052423998 | 0.000170263 | 0.076748389 |
| 16 | 0.025913698 | 0.045968927 | 0.082220315 | 0.013776843 | 0.033848726 | 0.009857283 | 0.018884426 | 0.067531926 | 0.058071836 | 0.030397273 |
| 17 | 0.098468059 | 0.006937858 | 0.059284345 | 0.007003944 | 0.016713281 | 0.019919444 | 0.096114478 | 0.083564882 | 0.030034772 | 0.087243466 |
| 18 | 0.043785991 | 0.047600874 | 0.049773597 | 0.092735132 | 0.024615841 | 0.077555439 | 0.002287264 | 0.011137469 | 0.041730468 | 0.020318015 |
| 19 | 0.079254312 | 0.073378990 | 0.080533882 | 0.075053157 | 0.039922652 | 0.089693744 | 0.038659620 | 0.045296185 | 0.057519408 | 0.069982288 |
| 20 | 0.020103130 | 0.020415749 | 0.008553436 | 0.046236850 | 0.001308684 | 0.035860540 | 0.074768913 | 0.085821415 | 0.041315892 | 0.028235492 |
| | **nominal significance level. $\alpha$=10%** | | | | | | | | | |
| do not regroup | 0.089 | 0.109 | 0.104 | 0.106 | 0.103 | 0.106 | 0.097 | 0.095 | 0.107 | 0.103 |
| regroup | 0.100 | 0.105 | 0.102 | 0.099 | 0.102 | 0.101 | 0.095 | 0.102 | 0.098 | 0.101 |
| | **nominal significance level. $\alpha$=5%** | | | | | | | | | |
| do not regroup | 0.048 | 0.057 | 0.055 | 0.053 | 0.058 | 0.060 | 0.052 | 0.053 | 0.060 | 0.054 |
| regroup | 0.054 | 0.053 | 0.053 | 0.048 | 0.056 | 0.053 | 0.047 | 0.047 | 0.047 | 0.050 |
| | **nominal significance level. $\alpha$=1%** | | | | | | | | | |
| do not regroup | 0.018 | 0.015 | 0.009 | 0.011 | 0.014 | 0.016 | 0.011 | 0.014 | 0.019 | 0.011 |
| regroup | 0.013 | 0.013 | 0.009 | 0.012 | 0.012 | 0.013 | 0.011 | 0.011 | 0.011 | 0.011 |

**Table A7.** *Results reported for 10 selected different multinomial populations, where the theoretical probabilities under the null hypothesis are described in the top part of the table for the different $k = 20$ classes considered for the simulation study. In this case, 5000 simulations from each population were simulated for $N = 1000$ and $k = 20$, and three different nominal significance levels considered ($\alpha$=0.10, 0.05 and 0.01). Significance levels attained by using the procedure without regrouping and those attained using the regrouping procedure proposed here are reported at the bottom of the table for each nominal significance level in the study for the fully specified chi-square goodness-of-fit test.*

| | **population distribution** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **k** | **popul. 1** | **popul. 3** | **popul. 4** | **popul. 8** | **popul. 9** | **popul. 12** | **popul. 14** | **popul. 15** | **popul. 16** | **popul. 18** |
| 1 | 0.086403829 | 0.027831969 | 0.034647847 | 0.066427337 | 0.011941121 | 0.036515589 | 0.027178041 | 0.006167233 | 0.102965702 | 0.072385838 |
| 2 | 0.126115131 | 0.033022721 | 0.068331204 | 0.064753732 | 0.030317750 | 0.044022950 | 0.081276230 | 0.058690477 | 0.039857006 | 0.041797540 |
| 3 | 0.003701857 | 0.069020808 | 0.086562978 | 0.014322857 | 0.039400867 | 0.031380261 | 0.002324866 | 0.009757145 | 0.096094633 | 0.044012553 |
| 4 | 0.087334122 | 0.057751497 | 0.001515899 | 0.000061967 | 0.033475410 | 0.023410240 | 0.092569838 | 0.091735551 | 0.053230597 | 0.056106000 |
| 5 | 0.088466246 | 0.083659666 | 0.099684674 | 0.034414442 | 0.070580693 | 0.013220949 | 0.004018535 | 0.070794259 | 0.068660196 | 0.067488478 |
| 6 | 0.040139121 | 0.072053283 | 0.004248836 | 0.039202054 | 0.092060032 | 0.000755158 | 0.062244097 | 0.001818666 | 0.006023243 | 0.045762370 |
| 7 | 0.035496055 | 0.026718019 | 0.034590801 | 0.046239080 | 0.048305619 | 0.063675307 | 0.030445736 | 0.071872580 | 0.039035047 | 0.064274476 |
| 8 | 0.066676696 | 0.034312791 | 0.061627901 | 0.081551238 | 0.003049453 | 0.093602008 | 0.022935046 | 0.044630722 | 0.021955132 | 0.015453954 |
| 9 | 0.126981250 | 0.074811112 | 0.007994219 | 0.089815281 | 0.047053740 | 0.049033194 | 0.017939102 | 0.064169616 | 0.096198491 | 0.073257160 |
| 10 | 0.005244943 | 0.038052516 | 0.058487468 | 0.030237968 | 0.037043015 | 0.110811279 | 0.064454119 | 0.053393883 | 0.057142848 | 0.014064999 |
| 11 | 0.025807180 | 0.007324624 | 0.029073335 | 0.065088621 | 0.086657108 | 0.046797729 | 0.066025531 | 0.106982482 | 0.084322875 | 0.076318917 |
| 12 | 0.048868600 | 0.003362205 | 0.045905763 | 0.059651340 | 0.006491148 | 0.072503067 | 0.074450794 | 0.116293555 | 0.040732458 | 0.076540555 |
| 13 | 0.013164045 | 0.072596073 | 0.098217793 | 0.084083663 | 0.039637897 | 0.002011858 | 0.067179949 | 0.018291986 | 0.039725583 | 0.044254924 |
| 14 | 0.022029874 | 0.073439034 | 0.042148633 | 0.074453016 | 0.096966438 | 0.091853613 | 0.043152479 | 0.070921712 | 0.005461318 | 0.045592138 |
| 15 | 0.063183192 | 0.074696494 | 0.019899462 | 0.088186646 | 0.032838526 | 0.078491318 | 0.075872526 | 0.006711885 | 0.006451768 | 0.036205940 |
| 16 | 0.007570864 | 0.040819714 | 0.090912651 | 0.017936627 | 0.086937269 | 0.011212680 | 0.065368287 | 0.040466816 | 0.052657694 | 0.073639777 |
| 17 | 0.020325254 | 0.082820695 | 0.052077725 | 0.027287750 | 0.090745181 | 0.081551839 | 0.003134611 | 0.006534136 | 0.075001996 | 0.051141004 |
| 18 | 0.033462668 | 0.050321400 | 0.079905443 | 0.012287602 | 0.040715073 | 0.071053400 | 0.091362328 | 0.004076579 | 0.098089330 | 0.060214286 |
| 19 | 0.049686185 | 0.012519013 | 0.022614239 | 0.015349585 | 0.025401215 | 0.008021889 | 0.011371395 | 0.099912643 | 0.011415523 | 0.002374922 |
| 20 | 0.049342888 | 0.064866366 | 0.061553130 | 0.088649193 | 0.080382444 | 0.070075672 | 0.096696489 | 0.056778073 | 0.004978558 | 0.039114168 |
| | **nominal significance level. $\alpha$=10%** | | | | | | | | | |
| do not regroup | 0.106 | 0.104 | 0.108 | 0.110 | 0.099 | 0.111 | 0.101 | 0.102 | 0.105 | 0.096 |
| regroup | 0.100 | 0.106 | 0.106 | 0.092 | 0.098 | 0.106 | 0.099 | 0.102 | 0.104 | 0.094 |
| | **nominal significance level. $\alpha$=5%** | | | | | | | | | |
| do not regroup | 0.053 | 0.055 | 0.058 | 0.062 | 0.050 | 0.055 | 0.055 | 0.053 | 0.051 | 0.050 |
| regroup | 0.051 | 0.055 | 0.056 | 0.043 | 0.051 | 0.056 | 0.051 | 0.053 | 0.050 | 0.050 |
| | **nominal significance level. $\alpha$=1%** | | | | | | | | | |
| do not regroup | 0.010 | 0.010 | 0.013 | 0.017 | 0.011 | 0.014 | 0.013 | 0.011 | 0.011 | 0.011 |
| regroup | 0.009 | 0.009 | 0.013 | 0.009 | 0.013 | 0.010 | 0.011 | 0.009 | 0.011 | 0.010 |

**Table A8.** *Results reported for 10 selected different multinomial population distributions, with the theoretical class probabilities under the null hypothesis of a normal distribution are described in the top part of the table for the different $k = 10$ classes considered for the simulation study. In this case, 5000 simulations from each population were simulated for $N = 50$ and $k = 10$, and three different nominal significance levels considered ($\alpha$=0.10, 0.05 and 0.01). Significance levels attained by using the procedure without regrouping and those attained using the regrouping procedure proposed here are reported at the bottom of the table for each nominal significance level in the study for the partially specified chi-square goodness-of-fit test.*

| | **Probability assignment** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **k** | **assig. 1** | **assig. 3** | **assig. 4** | **assig. 8** | **assig. 9** | **assig. 12** | **assig. 14** | **assig. 15** | **assig. 16** | **assig. 18** |
| 1 | 0.067722412 | 0.093438111 | 0.099092428 | 0.055167980 | 0.126712503 | 0.158331638 | 0.129522785 | 0.155141773 | 0.129108184 | 0.154908851 |
| 2 | 0.108251949 | 0.039193314 | 0.143654215 | 0.170135074 | 0.175667913 | 0.054474180 | 0.101468142 | 0.117873535 | 0.134484690 | 0.064089325 |
| 3 | 0.059166912 | 0.088253560 | 0.090975941 | 0.058609862 | 0.086016774 | 0.142045200 | 0.058832579 | 0.163627763 | 0.021375951 | 0.166329069 |
| 4 | 0.161665856 | 0.068734338 | 0.085228675 | 0.037666222 | 0.011613805 | 0.104115288 | 0.100833318 | 0.175329030 | 0.018903249 | 0.132811973 |
| 5 | 0.118838198 | 0.145124529 | 0.168663482 | 0.034227447 | 0.101429894 | 0.077007347 | 0.067901927 | 0.029351531 | 0.055381409 | 0.077241285 |
| 6 | 0.149585825 | 0.125514219 | 0.070588045 | 0.103546999 | 0.210875342 | 0.089579279 | 0.168068825 | 0.034844052 | 0.182953566 | 0.053626300 |
| 7 | 0.137791080 | 0.158950344 | 0.133376176 | 0.176459534 | 0.021765578 | 0.023434411 | 0.015373940 | 0.133862115 | 0.127970072 | 0.035629231 |
| 8 | 0.059378892 | 0.119973253 | 0.041377815 | 0.159305809 | 0.044490360 | 0.031816071 | 0.158329630 | 0.055992935 | 0.145612819 | 0.153044117 |
| 9 | 0.083716990 | 0.011236121 | 0.003529199 | 0.133811376 | 0.115097525 | 0.160946852 | 0.162414523 | 0.045883570 | 0.055317926 | 0.066055454 |
| 10 | 0.053881886 | 0.149582209 | 0.163514025 | 0.071069698 | 0.106330305 | 0.158249734 | 0.037254330 | 0.088093695 | 0.128892133 | 0.096264395 |
| | **nominal significance level. $\alpha$=10%** | | | | | | | | | |
| do not regroup | 0.113 | 0.115 | 0.121 | 0.108 | 0.112 | 0.119 | 0.108 | 0.112 | 0.108 | 0.114 |
| regroup | 0.133 | 0.124 | 0.158 | 0.135 | 0.126 | 0.128 | 0.126 | 0.135 | 0.127 | 0.144 |
| | **nominal significance level. $\alpha$=5%** | | | | | | | | | |
| do not regroup | 0.051 | 0.068 | 0.061 | 0.052 | 0.061 | 0.061 | 0.058 | 0.058 | 0.061 | 0.056 |
| regroup | 0.068 | 0.062 | 0.076 | 0.067 | 0.063 | 0.061 | 0.064 | 0.063 | 0.065 | 0.073 |
| | **nominal significance level. $\alpha$=1%** | | | | | | | | | |
| do not regroup | 0.010 | 0.018 | 0.024 | 0.012 | 0.016 | 0.013 | 0.015 | 0.014 | 0.018 | 0.012 |
| regroup | 0.013 | 0.012 | 0.012 | 0.014 | 0.012 | 0.013 | 0.014 | 0.011 | 0.015 | 0.014 |

**Table A9.** *Results reported for 10 selected different multinomial population distributions, with the theoretical class probabilities under the null hypothesis of a normal distribution are described in the top part of the table for the different $k = 15$ classes considered for the simulation study. In this case, 5000 simulations from each population were simulated for $N = 75$ and $k = 15$, and three different nominal significance levels considered ($\alpha$=0.10, 0.05 and 0.01). Significance levels attained by using the procedure without regrouping and those attained using the regrouping procedure proposed here are reported at the bottom of the table for each nominal significance level in the study for the partially specified chi-square goodness-of-fit test.*

| **Probability assignment** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **k** | **assig. 1** | **assig. 3** | **assig. 4** | **assig. 8** | **assig. 9** | **assig. 12** | **assig. 14** | **assig. 15** | **assig. 16** | **assig. 18** |
| 1 | 0.127057671 | 0.083861803 | 0.081106551 | 0.072320794 | 0.078080154 | 0.029394665 | 0.118093818 | 0.131915150 | 0.124764161 | 0.101425616 |
| 2 | 0.088104527 | 0.067347127 | 0.063163406 | 0.031832989 | 0.091721146 | 0.057298124 | 0.064924762 | 0.137187332 | 0.084398252 | 0.054144288 |
| 3 | 0.029902404 | 0.085171458 | 0.035443425 | 0.062531848 | 0.002989394 | 0.050403893 | 0.085464849 | 0.092797832 | 0.074048405 | 0.083714805 |
| 4 | 0.008582462 | 0.024179898 | 0.090621898 | 0.006554065 | 0.009238380 | 0.117528851 | 0.088907182 | 0.013390771 | 0.032826123 | 0.033579822 |
| 5 | 0.113134725 | 0.110562508 | 0.043446425 | 0.085630406 | 0.132330601 | 0.000768729 | 0.081224262 | 0.077103919 | 0.031323120 | 0.069563207 |
| 6 | 0.007700507 | 0.042833764 | 0.079690194 | 0.050492072 | 0.045150829 | 0.021549876 | 0.008446190 | 0.154708685 | 0.001610057 | 0.079027830 |
| 7 | 0.130266721 | 0.036978384 | 0.054341777 | 0.005275661 | 0.119958130 | 0.086208961 | 0.044215657 | 0.049446275 | 0.017379203 | 0.058223969 |
| 8 | 0.043228436 | 0.039777167 | 0.055824977 | 0.037389853 | 0.037573870 | 0.058081580 | 0.105879702 | 0.025265570 | 0.032805863 | 0.008089689 |
| 9 | 0.080676242 | 0.112476617 | 0.079526539 | 0.016972967 | 0.095983503 | 0.120659949 | 0.027606531 | 0.027259488 | 0.103095726 | 0.010500826 |
| 10 | 0.094468872 | 0.094139849 | 0.062781085 | 0.078659953 | 0.124772249 | 0.097925781 | 0.064328900 | 0.145884761 | 0.034570874 | 0.105571614 |
| 11 | 0.023806346 | 0.084163565 | 0.108837039 | 0.107596687 | 0.031639086 | 0.028899879 | 0.027285188 | 0.063096567 | 0.106607857 | 0.058341183 |
| 12 | 0.115757204 | 0.029147017 | 0.025956871 | 0.144561350 | 0.037859485 | 0.096476790 | 0.119112740 | 0.011338246 | 0.093203420 | 0.120721080 |
| 13 | 0.107141421 | 0.101775293 | 0.103602266 | 0.054244863 | 0.121044273 | 0.109784646 | 0.027558602 | 0.021700713 | 0.046306240 | 0.124422616 |
| 14 | 0.017980943 | 0.025664489 | 0.077737739 | 0.118268377 | 0.034410514 | 0.091606477 | 0.051464298 | 0.030177488 | 0.096164501 | 0.091153404 |
| 15 | 0.012191519 | 0.061921062 | 0.037919809 | 0.127668115 | 0.037248387 | 0.033411800 | 0.085487321 | 0.018727203 | 0.120896197 | 0.001520050 |
| **nominal significance level. $\alpha$=10%** | | | | | | | | | | |
| do not regroup | 0.132 | 0.127 | 0.126 | 0.136 | 0.142 | 0.152 | 0.140 | 0.132 | 0.141 | 0.140 |
| regroup | 0.144 | 0.135 | 0.139 | 0.134 | 0.139 | 0.147 | 0.141 | 0.133 | 0.142 | 0.142 |
| **nominal significance level. $\alpha$=5%** | | | | | | | | | | |
| do not regroup | 0.078 | 0.078 | 0.082 | 0.092 | 0.093 | 0.111 | 0.086 | 0.084 | 0.089 | 0.094 |
| regroup | 0.090 | 0.083 | 0.085 | 0.083 | 0.085 | 0.086 | 0.089 | 0.084 | 0.086 | 0.085 |
| **nominal significance level. $\alpha$=1%** | | | | | | | | | | |
| do not regroup | 0.044 | 0.040 | 0.041 | 0.047 | 0.049 | 0.062 | 0.045 | 0.041 | 0.049 | 0.052 |
| regroup | 0.041 | 0.040 | 0.041 | 0.040 | 0.039 | 0.041 | 0.040 | 0.040 | 0.041 | 0.041 |

**Table A10.** *Results reported for 10 selected different multinomial population distributions, with the theoretical class probabilities under the null hypothesis of a normal distribution are described in the top part of the table for the different $k = 20$ classes considered for the simulation study. In this case, 5000 simulations from each population were simulated for $N = 100$ and $k = 20$, and three different nominal significance levels considered ($\alpha$=0.10, 0.05 and 0.01). Significance levels attained by using the procedure without regrouping and those attained using the regrouping procedure proposed here are reported at the bottom of the table for each nominal significance level in the study for the partially specified chi-square goodness-of-fit test.*

| | **Probability assignment** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **k** | **assig. 1** | **assig. 3** | **assig. 4** | **assig. 8** | **assig. 9** | **assig. 12** | **assig. 14** | **assig. 15** | **assig. 16** | **assig. 18** |
| 1 | 0.009906949 | 0.042585801 | 0.036322397 | 0.023803358 | 0.056050361 | 0.048540316 | 0.048978727 | 0.014696213 | 0.004618549 | 0.020337923 |
| 2 | 0.043334958 | 0.084595131 | 0.073420820 | 0.064449053 | 0.043706663 | 0.085501927 | 0.054249217 | 0.049354784 | 0.045174755 | 0.012077438 |
| 3 | 0.068751733 | 0.055949091 | 0.065878707 | 0.037395221 | 0.035956904 | 0.056738810 | 0.062717404 | 0.038945982 | 0.006326356 | 0.012472135 |
| 4 | 0.075322010 | 0.088903934 | 0.093919822 | 0.080488103 | 0.012855218 | 0.082229679 | 0.057512461 | 0.087900507 | 0.104546576 | 0.089286075 |
| 5 | 0.084464071 | 0.089215586 | 0.091398883 | 0.018016210 | 0.085894992 | 0.021013929 | 0.067902855 | 0.082246443 | 0.054052148 | 0.007489974 |
| 6 | 0.055848915 | 0.063276437 | 0.021928939 | 0.036960583 | 0.000994794 | 0.011044924 | 0.037607372 | 0.064129728 | 0.077098867 | 0.000477913 |
| 7 | 0.033849691 | 0.007263521 | 0.064953627 | 0.069011621 | 0.055694252 | 0.093010187 | 0.041731769 | 0.019967383 | 0.110322256 | 0.059307357 |
| 8 | 0.041782157 | 0.015192684 | 0.004878198 | 0.006500386 | 0.038241324 | 0.008823266 | 0.066635437 | 0.037496205 | 0.044073167 | 0.092347960 |
| 9 | 0.049615947 | 0.009841069 | 0.009040455 | 0.094543419 | 0.028413965 | 0.099002066 | 0.074037397 | 0.062557226 | 0.094012295 | 0.081521459 |
| 10 | 0.054796141 | 0.001109167 | 0.090830640 | 0.047526405 | 0.072759348 | 0.039090212 | 0.057435627 | 0.033694948 | 0.027255678 | 0.012003245 |
| 11 | 0.049573297 | 0.045490837 | 0.070485338 | 0.022710569 | 0.092939405 | 0.070205677 | 0.054643564 | 0.035830706 | 0.102879776 | 0.064882703 |
| 12 | 0.034362080 | 0.011503208 | 0.012250549 | 0.077123015 | 0.087869027 | 0.051263015 | 0.046862814 | 0.061074007 | 0.002815851 | 0.065966882 |
| 13 | 0.003321141 | 0.082512233 | 0.036884669 | 0.047694596 | 0.094458297 | 0.053229198 | 0.030201345 | 0.056240124 | 0.006559463 | 0.006896735 |
| 14 | 0.035231977 | 0.028957298 | 0.021443815 | 0.071919102 | 0.009911000 | 0.005160299 | 0.016883433 | 0.011740128 | 0.002543542 | 0.055721211 |
| 15 | 0.096938721 | 0.058782500 | 0.060608656 | 0.064842833 | 0.042211503 | 0.051071775 | 0.044675117 | 0.057298982 | 0.035805897 | 0.044729954 |
| 16 | 0.079250776 | 0.010929522 | 0.058052934 | 0.006038138 | 0.073256785 | 0.003184016 | 0.031371077 | 0.075481374 | 0.070503856 | 0.082763910 |
| 17 | 0.032532973 | 0.090748245 | 0.089832763 | 0.087521877 | 0.023712180 | 0.069816537 | 0.072661166 | 0.038469940 | 0.089473563 | 0.068822077 |
| 18 | 0.068855512 | 0.033207256 | 0.002387938 | 0.054802997 | 0.083944166 | 0.088130478 | 0.057917558 | 0.093425379 | 0.013317647 | 0.086962221 |
| 19 | 0.019344094 | 0.104381792 | 0.043173925 | 0.007385808 | 0.009079374 | 0.014168915 | 0.028620055 | 0.050048690 | 0.046279136 | 0.089193171 |
| 20 | 0.062916859 | 0.075554687 | 0.052306925 | 0.081266705 | 0.052050441 | 0.048774773 | 0.047355605 | 0.029401252 | 0.062340620 | 0.046739658 |
| | **nominal significance level. $\alpha$=10%** | | | | | | | | | |
| do not regroup | 0.106 | 0.110 | 0.113 | 0.110 | 0.116 | 0.113 | 0.101 | 0.100 | 0.120 | 0.129 |
| regroup | 0.104 | 0.112 | 0.111 | 0.109 | 0.112 | 0.106 | 0.105 | 0.109 | 0.111 | 0.120 |
| | **nominal significance level. $\alpha$=5%** | | | | | | | | | |
| do not regroup | 0.053 | 0.060 | 0.066 | 0.057 | 0.065 | 0.064 | 0.055 | 0.048 | 0.069 | 0.087 |
| regroup | 0.050 | 0.056 | 0.054 | 0.054 | 0.058 | 0.055 | 0.056 | 0.052 | 0.054 | 0.060 |
| | **nominal significance level. $\alpha$=1%** | | | | | | | | | |
| do not regroup | 0.013 | 0.019 | 0.019 | 0.017 | 0.020 | 0.022 | 0.013 | 0.011 | 0.023 | 0.037 |
| regroup | 0.011 | 0.010 | 0.012 | 0.009 | 0.010 | 0.012 | 0.010 | 0.008 | 0.011 | 0.013 |

**Table A11.** *Results reported for 10 selected different multinomial population distributions, with the theoretical class probabilities under the null hypothesis of a normal distribution are described in the top part of the table for the different $k = 20$ classes considered for the simulation study. In this case, 5000 simulations from each population were simulated for $N = 250$ and $k = 20$, and three different nominal significance levels considered ($\alpha$=0.10, 0.05 and 0.01). Significance levels attained by using the procedure without regrouping and those attained using the regrouping procedure proposed here are reported at the bottom of the table for each nominal significance level in the study for the partially specified chi-square goodness-of-fit test.*

| | Probability assignment | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **k** | **assig. 1** | **assig. 3** | **assig. 4** | **assig. 8** | **assig. 9** | **assig. 12** | **assig. 14** | **assig. 15** | **assig. 16** | **assig. 18** |
| 1 | 0.047643128 | 0.054666723 | 0.056038758 | 0.090599483 | 0.025317013 | 0.023099063 | 0.064584061 | 0.136478476 | 0.043123333 | 0.054066459 |
| 2 | 0.024791212 | 0.044261848 | 0.032569245 | 0.044497446 | 0.014610403 | 0.084334948 | 0.009032926 | 0.014084696 | 0.072955298 | 0.053573431 |
| 3 | 0.054639637 | 0.071340360 | 0.073471657 | 0.049842681 | 0.020683555 | 0.037648781 | 0.066675056 | 0.006032668 | 0.034107228 | 0.013291318 |
| 4 | 0.011551957 | 0.035790188 | 0.017892204 | 0.048986190 | 0.015986321 | 0.061877989 | 0.023052570 | 0.051462288 | 0.069213717 | 0.045963934 |
| 5 | 0.043209785 | 0.068393517 | 0.060880791 | 0.038016856 | 0.078291674 | 0.015100046 | 0.085352639 | 0.015104595 | 0.055712777 | 0.094207824 |
| 6 | 0.058607724 | 0.084149275 | 0.040842165 | 0.082739121 | 0.059420758 | 0.012071383 | 0.011090706 | 0.077633225 | 0.038688100 | 0.023326840 |
| 7 | 0.004077781 | 0.036377709 | 0.062850108 | 0.075899205 | 0.050698562 | 0.072435772 | 0.049065314 | 0.072825503 | 0.083729668 | 0.047299352 |
| 8 | 0.085955934 | 0.066886372 | 0.073347443 | 0.035307555 | 0.045219825 | 0.044208957 | 0.055207226 | 0.019632440 | 0.056600081 | 0.018773053 |
| 9 | 0.023369288 | 0.015630083 | 0.032843147 | 0.026278327 | 0.058397084 | 0.032659614 | 0.083467899 | 0.016586480 | 0.010340800 | 0.050770889 |
| 10 | 0.087652522 | 0.007532102 | 0.054125222 | 0.034259425 | 0.044017966 | 0.061446558 | 0.039048538 | 0.069347828 | 0.000467925 | 0.095008505 |
| 11 | 0.095771275 | 0.057965331 | 0.055877398 | 0.011887980 | 0.088882200 | 0.029313393 | 0.063182257 | 0.045379877 | 0.079491665 | 0.001551221 |
| 12 | 0.096306521 | 0.030974889 | 0.036689952 | 0.006101936 | 0.044860170 | 0.068419459 | 0.062127872 | 0.119150362 | 0.038817979 | 0.013723146 |
| 13 | 0.080897006 | 0.087201505 | 0.004498861 | 0.029781475 | 0.094173221 | 0.009146832 | 0.018611673 | 0.017167618 | 0.063579053 | 0.040856797 |
| 14 | 0.013705065 | 0.070556907 | 0.100690100 | 0.059337269 | 0.017035620 | 0.083251916 | 0.051065504 | 0.033760507 | 0.011793975 | 0.071820279 |
| 15 | 0.024205251 | 0.014659583 | 0.060392354 | 0.087250033 | 0.067126421 | 0.059621249 | 0.048932496 | 0.001487572 | 0.063205970 | 0.098332826 |
| 16 | 0.004839018 | 0.087276954 | 0.007352012 | 0.074132698 | 0.079880521 | 0.091801986 | 0.023883012 | 0.075759621 | 0.016191521 | 0.021901263 |
| 17 | 0.060014816 | 0.068505786 | 0.040949294 | 0.063520955 | 0.093936062 | 0.032076674 | 0.029721620 | 0.096471797 | 0.075764151 | 0.060809674 |
| 18 | 0.037409444 | 0.050921813 | 0.080223679 | 0.007014358 | 0.004676127 | 0.060818519 | 0.056226443 | 0.055834461 | 0.092658366 | 0.034348079 |
| 19 | 0.071116566 | 0.043564240 | 0.034035218 | 0.082594193 | 0.036804493 | 0.082172574 | 0.077824913 | 0.006284488 | 0.082035648 | 0.102590301 |
| 20 | 0.074236070 | 0.003344816 | 0.074430393 | 0.051952812 | 0.059982005 | 0.038494286 | 0.081847276 | 0.069515496 | 0.011522743 | 0.057784808 |
| **nominal significance level. $\alpha$=10%** | | | | | | | | | | |
| do not regroup | 0.106 | 0.106 | 0.106 | 0.105 | 0.107 | 0.106 | 0.103 | 0.116 | 0.110 | 0.109 |
| regroup | 0.103 | 0.103 | 0.101 | 0.106 | 0.105 | 0.102 | 0.102 | 0.104 | 0.100 | 0.098 |
| **nominal significance level. $\alpha$=5%** | | | | | | | | | | |
| do not regroup | 0.055 | 0.057 | 0.056 | 0.054 | 0.053 | 0.054 | 0.047 | 0.064 | 0.060 | 0.056 |
| regroup | 0.047 | 0.055 | 0.051 | 0.052 | 0.051 | 0.051 | 0.053 | 0.055 | 0.053 | 0.054 |
| **nominal significance level. $\alpha$=1%** | | | | | | | | | | |
| do not regroup | 0.010 | 0.012 | 0.014 | 0.012 | 0.013 | 0.012 | 0.010 | 0.018 | 0.019 | 0.015 |
| regroup | 0.009 | 0.013 | 0.010 | 0.013 | 0.008 | 0.010 | 0.009 | 0.013 | 0.013 | 0.012 |

**Table A12.** *Results reported for 10 selected different multinomial population distributions, with the theoretical class probabilities under the null hypothesis of a normal distribution are described in the top part of the table for the different $k = 20$ classes considered for the simulation study. In this case, 5000 simulations from each population were simulated for $N = 500$ and $k = 20$, and three different nominal significance levels considered (α=0.10, 0.05 and 0.01). Significance levels attained by using the procedure without regrouping and those using the regrouping procedure proposed here are reported at the bottom of the table for each of the nominal significance level in the study for the partially specified chi-square goodness-of-fit test.*

| | **Probability assignment** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **k** | **assig. 1** | **assig. 3** | **assig. 4** | **assig. 8** | **assig. 9** | **assig. 12** | **assig. 14** | **assig. 15** | **assig. 16** | **assig. 18** |
| 1 | 0.009323228 | 0.021842441 | 0.073053675 | 0.040873644 | 0.034371818 | 0.004343310 | 0.077050114 | 0.009333736 | 0.089174555 | 0.077053926 |
| 2 | 0.073419610 | 0.084240115 | 0.043107814 | 0.042528243 | 0.090126015 | 0.092113419 | 0.078204289 | 0.067545345 | 0.093289019 | 0.032029821 |
| 3 | 0.064476239 | 0.067584721 | 0.053896718 | 0.030820872 | 0.011727517 | 0.084304086 | 0.054005211 | 0.063428357 | 0.053021839 | 0.082081999 |
| 4 | 0.011049151 | 0.088536877 | 0.024211296 | 0.027384842 | 0.069805333 | 0.028693635 | 0.092724143 | 0.067470149 | 0.019683922 | 0.055045204 |
| 5 | 0.086457614 | 0.055072657 | 0.071720992 | 0.089111120 | 0.069667082 | 0.011893892 | 0.055181544 | 0.006367421 | 0.059474204 | 0.063753881 |
| 6 | 0.022503868 | 0.016163729 | 0.060949072 | 0.091777222 | 0.062247990 | 0.067694023 | 0.016575229 | 0.097107691 | 0.099925118 | 0.039857245 |
| 7 | 0.000016554 | 0.068682807 | 0.070000441 | 0.085376495 | 0.071095031 | 0.090696516 | 0.026952117 | 0.002052394 | 0.068436458 | 0.059934452 |
| 8 | 0.076067786 | 0.058131169 | 0.047627730 | 0.052631881 | 0.014714001 | 0.062019278 | 0.085712216 | 0.018122689 | 0.005818399 | 0.050251930 |
| 9 | 0.044093494 | 0.002919132 | 0.013261422 | 0.017399972 | 0.058380580 | 0.051731317 | 0.009912796 | 0.023110315 | 0.021480560 | 0.023071720 |
| 10 | 0.008782610 | 0.062675614 | 0.060953727 | 0.035420538 | 0.088536843 | 0.060798804 | 0.017801328 | 0.085190935 | 0.083247679 | 0.054008372 |
| 11 | 0.100593141 | 0.100396950 | 0.044992803 | 0.069148810 | 0.055528229 | 0.045118418 | 0.056649822 | 0.087104011 | 0.016148816 | 0.006887184 |
| 12 | 0.097205062 | 0.007259403 | 0.051243758 | 0.101523827 | 0.044649792 | 0.040196511 | 0.069997087 | 0.019445299 | 0.007289295 | 0.077411088 |
| 13 | 0.052323470 | 0.037830917 | 0.031323124 | 0.046548434 | 0.093038962 | 0.049782978 | 0.021150533 | 0.104840295 | 0.063517778 | 0.013079696 |
| 14 | 0.018032175 | 0.025567586 | 0.055357326 | 0.003155478 | 0.057561203 | 0.076744301 | 0.039391817 | 0.003105487 | 0.090649719 | 0.052608559 |
| 15 | 0.068130807 | 0.108793483 | 0.017934527 | 0.031492695 | 0.062140420 | 0.000983061 | 0.067977053 | 0.052423998 | 0.000170263 | 0.076748389 |
| 16 | 0.025913698 | 0.045968927 | 0.082220315 | 0.013776843 | 0.033848726 | 0.009857283 | 0.018884426 | 0.067531926 | 0.058071836 | 0.030397273 |
| 17 | 0.098468059 | 0.006937858 | 0.059284345 | 0.007003944 | 0.016713281 | 0.019919444 | 0.096114478 | 0.083564882 | 0.030034772 | 0.087243466 |
| 18 | 0.043785991 | 0.047600874 | 0.049773597 | 0.092735132 | 0.024615841 | 0.077555439 | 0.002287264 | 0.011137469 | 0.041730468 | 0.020318015 |
| 19 | 0.079254312 | 0.073378990 | 0.080533882 | 0.075053157 | 0.039922652 | 0.089693744 | 0.038659620 | 0.045296185 | 0.057519408 | 0.069982288 |
| 20 | 0.020103130 | 0.020415749 | 0.008553436 | 0.046236850 | 0.001308684 | 0.035860540 | 0.074768913 | 0.085821415 | 0.041315892 | 0.028235492 |
| | **nominal significance level. $\alpha$=10%** | | | | | | | | | |
| do not regroup | 0.085 | 0.107 | 0.113 | 0.112 | 0.104 | 0.109 | 0.109 | 0.109 | 0.116 | 0.103 |
| regroup | 0.106 | 0.111 | 0.114 | 0.105 | 0.108 | 0.110 | 0.107 | 0.107 | 0.103 | 0.104 |
| | **nominal significance level. $\alpha$=5%** | | | | | | | | | |
| do not regroup | 0.048 | 0.056 | 0.055 | 0.058 | 0.055 | 0.061 | 0.059 | 0.062 | 0.069 | 0.052 |
| regroup | 0.053 | 0.056 | 0.057 | 0.057 | 0.053 | 0.057 | 0.053 | 0.058 | 0.054 | 0.050 |
| | **nominal significance level. $\alpha$=1%** | | | | | | | | | |
| do not regroup | 0.016 | 0.014 | 0.014 | 0.016 | 0.011 | 0.017 | 0.017 | 0.016 | 0.020 | 0.012 |
| regroup | 0.011 | 0.011 | 0.014 | 0.013 | 0.013 | 0.015 | 0.015 | 0.013 | 0.012 | 0.010 |

**Table A13.** *Results reported for 10 selected different multinomial population distributions, with the theoretical class probabilities under the null hypothesis of a normal distribution are described in the top part of the table for the different $k = 20$ classes considered for the simulation study. In this case, 5000 simulations from each population were simulated for $N = 1000$ and $k = 20$, and three different nominal significance levels considered ($\alpha$=0.10, 0.05 and 0.01). Significance levels attained by using the procedure without regrouping and those attained using the regrouping procedure proposed here are reported at the bottom of the table for each nominal significance level in the study for the partially specified chi-square goodness-of-fit test.*

| Probability assignment | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **k** | **assig. 1** | **assig. 3** | **assig. 4** | **assig. 8** | **assig. 9** | **assig. 12** | **assig. 14** | **assig. 15** | **assig. 16** | **assig. 18** |
| 1 | 0.086403829 | 0.027831969 | 0.034647847 | 0.066427337 | 0.011941121 | 0.036515589 | 0.027178041 | 0.006167233 | 0.102965702 | 0.072385838 |
| 2 | 0.126115131 | 0.033022721 | 0.068331204 | 0.064753732 | 0.030317750 | 0.044022950 | 0.081276230 | 0.058690477 | 0.039857006 | 0.041797540 |
| 3 | 0.003701857 | 0.069020808 | 0.086562978 | 0.014322857 | 0.039400867 | 0.031380261 | 0.002324866 | 0.009757145 | 0.096094633 | 0.044012553 |
| 4 | 0.087334122 | 0.057751497 | 0.001515899 | 0.000061967 | 0.033475410 | 0.023410240 | 0.092569838 | 0.091735551 | 0.053230597 | 0.056106000 |
| 5 | 0.088466246 | 0.083659666 | 0.099684674 | 0.034414442 | 0.070580693 | 0.013220949 | 0.004018535 | 0.070794259 | 0.068660196 | 0.067488478 |
| 6 | 0.040139121 | 0.072053283 | 0.004248836 | 0.039202054 | 0.092060032 | 0.000755158 | 0.062244097 | 0.001818666 | 0.006023243 | 0.045762370 |
| 7 | 0.035496055 | 0.026718019 | 0.034590801 | 0.046239080 | 0.048305619 | 0.063675307 | 0.030445736 | 0.071872580 | 0.039035047 | 0.064274476 |
| 8 | 0.066676696 | 0.034312791 | 0.061627901 | 0.081551238 | 0.003049453 | 0.093602008 | 0.022935046 | 0.044630722 | 0.021955132 | 0.015453954 |
| 9 | 0.126981250 | 0.074811112 | 0.007994219 | 0.089815281 | 0.047053740 | 0.049033194 | 0.017939102 | 0.064169616 | 0.096198491 | 0.073257160 |
| 10 | 0.005244943 | 0.038052516 | 0.058487468 | 0.030237968 | 0.037043015 | 0.110811279 | 0.064454119 | 0.053393883 | 0.057142848 | 0.014064999 |
| 11 | 0.025807180 | 0.007324624 | 0.029073335 | 0.065088621 | 0.086657108 | 0.046797729 | 0.066025531 | 0.106982482 | 0.084322875 | 0.076318917 |
| 12 | 0.048868600 | 0.003362205 | 0.045905763 | 0.059651340 | 0.006491148 | 0.072503067 | 0.074450794 | 0.116293555 | 0.040732458 | 0.076540555 |
| 13 | 0.013164045 | 0.072596073 | 0.098217793 | 0.084083663 | 0.039637897 | 0.002011858 | 0.067179949 | 0.018291986 | 0.039725583 | 0.044254924 |
| 14 | 0.022029874 | 0.073439034 | 0.042148633 | 0.074453016 | 0.096966438 | 0.091853613 | 0.043152479 | 0.070921712 | 0.005461318 | 0.045592138 |
| 15 | 0.063183192 | 0.074696494 | 0.019899462 | 0.088186646 | 0.032838526 | 0.078491318 | 0.075872526 | 0.006711885 | 0.006451768 | 0.036205940 |
| 16 | 0.007570864 | 0.040819714 | 0.090912651 | 0.017936627 | 0.086937269 | 0.011212680 | 0.065368287 | 0.040466816 | 0.052657694 | 0.073639777 |
| 17 | 0.020325254 | 0.082820695 | 0.052077725 | 0.027287750 | 0.090745181 | 0.081551839 | 0.003134611 | 0.006534136 | 0.075001996 | 0.051141004 |
| 18 | 0.033462668 | 0.050321400 | 0.079905443 | 0.012287602 | 0.040715073 | 0.071053400 | 0.091362328 | 0.004076579 | 0.098089330 | 0.060214286 |
| 19 | 0.049686185 | 0.012519013 | 0.022614239 | 0.015349585 | 0.025401215 | 0.008021889 | 0.011371395 | 0.099912643 | 0.011415523 | 0.002374922 |
| 20 | 0.049342888 | 0.064866366 | 0.061553130 | 0.088649193 | 0.080382444 | 0.070075672 | 0.096696489 | 0.056778073 | 0.004978558 | 0.039114168 |
| **nominal significance level. $\alpha$=10%** | | | | | | | | | | |
| do not regroup | 0.104 | 0.106 | 0.101 | 0.131 | 0.108 | 0.105 | 0.107 | 0.103 | 0.108 | 0.114 |
| regroup | 0.102 | 0.107 | 0.105 | 0.110 | 0.108 | 0.105 | 0.102 | 0.107 | 0.112 | 0.115 |
| **nominal significance level. $\alpha$=5%** | | | | | | | | | | |
| do not regroup | 0.058 | 0.058 | 0.054 | 0.080 | 0.057 | 0.056 | 0.059 | 0.051 | 0.052 | 0.054 |
| regroup | 0.057 | 0.056 | 0.049 | 0.057 | 0.056 | 0.054 | 0.056 | 0.051 | 0.055 | 0.055 |
| **nominal significance level. $\alpha$=1%** | | | | | | | | | | |
| do not regroup | 0.010 | 0.013 | 0.010 | 0.026 | 0.011 | 0.014 | 0.014 | 0.012 | 0.011 | 0.009 |
| regroup | 0.011 | 0.014 | 0.009 | 0.014 | 0.012 | 0.012 | 0.012 | 0.010 | 0.011 | 0.010 |

**Table A14.** *Case 2. Sample values ($N = 50$) generated for a standard normal distribution with the rnorm function in R.*

| | | | | |
|---|---|---|---|---|
| −0.65618448 | −1.54766098 | 0.17835732 | 0.00138515 | −1.40268582 |
| −0.20245233 | 0.484569688 | 0.66742264 | 0.39006195 | 0.82936868 |
| −0.79846811 | 1.56794024 | 0.17461046 | 0.17167498 | 2.17726899 |
| −0.55908895 | 0.735762715 | 0.56947521 | −0.71881009 | −0.81875671 |
| 0.2486506 | 0.042690025 | −0.96268114 | −0.44881241 | −0.40618589 |
| 0.50259586 | −0.88656895 | −0.00530998 | 0.82506957 | 0.14533868 |
| −0.80817805 | −0.10141299 | −0.56441779 | −1.10274085 | −0.10099009 |
| 0.00394035 | −0.15807742 | 0.59004733 | 1.33483171 | −0.06117801 |
| 2.75651035 | 0.73222822 | −0.7226212 | 0.06114264 | −1.01208944 |
| 0.18956141 | 0.766926096 | 0.91360816 | −0.53365171 | 0.3428841 |