



One-Sample Score Test (`test_score_os`)

Author: P. Stikker

Website: <https://peterStatistics.com>

YouTube: <https://www.youtube.com/stikpet>

Version: 0.2 (2023-01-06)

Introduction

The `test_score_os` function (and `test_score_os_arr` in VBA) can perform a one-sample score test. An approximation for the one-sample binomial test. This test should only be applied if n is large.

The p-value indicates the probability of a result as in the sample, or more extreme, if the probability of success in the population is indeed the used one for the test. If this probability is small (usually below 0.05) the assumption is rejected, concluding that it is significantly different, otherwise it is not accepted but rather insufficient evidence to reject it.

This document contains the details on how to use the functions, and formulas used in them.

1 About the Function

1.1 Input parameters:

- **data**
The data to be used. Note for Python this needs to be a pandas data series.

- **codes**
Two codes for the two categories to be compared. For example if the data has a list of scores with “national” and “international”, the codes used can be exactly those: “national”, “international”.
This makes it possible to also use a nominal data set (with more than two categories) and then select the two for this test to be used.

- *Optional parameters*
 - **p0** (default is 0.5)
The hypothesized proportion for the first category (as in codes).

 - **cc** (default is “none”)
Choice to use a continuity correction either:
 - “none”: no correction used
 - “yates”: Yates continuity correction is used

 - **out** (default is “pvalue”) – only applies to VBA `test_score_os`
Choice what to show as result. Either:
 - “pvalue”: show the p-value (significance)
 - “statistic”: show the test-statistic used



1.2 Output:

- The **p-value**, the **test statistic**, and the **name of the test** used. Except for the non-array version in VBA (Excel) which will only show the requested output via the 'out' parameter.
- The array version in VBA (test_score_os_arr) requires **two rows** and **three columns**.

1.3 Dependencies

- **Excel**
The VBA code will make use of the worksheetfunctions NormSDist.
You can run the **ts_score_addHelp** macro so that the function will be available with some help in the 'User Defined' category in the functions overview.
- **Python**
The following additional libraries will have to be installed:
 - *pandas*
the data input needs to be a pandas data series, and the output is also a pandas dataframe.
 - *statistics* (from Python)
the NormalDist function from Python's statistics library is needed in the calculations, specifically NormalDist().cdf is used.
- **R**
No other libraries required. The *pnorm* function is used, but is available in R itself.

2 Examples

2.1 Excel

	A	B	C	D	E	F	G
1	data		categories				
2	1		1	p0		0,4	
3	2		2				
4	2						
5	1						
6	2		method	p-value			
7	2		none	0,4537	=ts_score_os(\$A\$2:\$A\$20;\$C\$2:\$C\$3;\$F\$2;C7)		
8	1		yates	0,60647	=ts_score_os(\$A\$2:\$A\$20;\$C\$2:\$C\$3;\$F\$2;C8)		
9	1						
10	2						
11	2			statistic	p-value	test	
12	2			-0,74927	0,4537	Normal approximation	
13	2						
14	2			D11:F12	=ts_score_os_arr(\$A\$2:\$A\$20;\$C\$2:\$C\$3;\$F\$2)		
15	2						
16	1						
17	2						
18	1						
19	2						
20	2						



2.2 Python

```
[2]: import pandas as pd
      from statistics import NormalDist

[3]: #example
      dataList = ['Female', 'Male', 'Male', 'Female', 'Male', 'Male', 'Female', 'Female', 'Male', 'Male',
                  'Male', 'Male', 'Male', 'Male', 'Female', 'Male', 'Female', 'Male', 'Male']
      data = pd.Series(dataList)
      codes = ['Female', 'Male']

[4]: ts_score_os(data, codes, 0.4)

[4]:
```

	statistic	p-value (2-sided)	test
0	-0.749269	0.453695	Normal approximation

```
[5]: ts_score_os(data, codes, p0=0.4, cc="yates")

[5]:
```

	statistic	p-value (2-sided)	test
0	-0.515122	0.606468	Normal approximation with Yates continuity cor...

2.3 R

```
> data <- c("Female", "Male", "Male", "Female", "Male", "Male",
+          "Female", "Female", "Male", "Male", "Male", "Male",
+          "Male", "Male", "Female", "Male", "Female", "Male", "Male")
> ts_score_os(data, c("Female", "Male"), p0 = 0.4)
  testValue      sig2      testUsed
1 -0.7492686 0.4536953 Normal approximation
> ts_score_os(data, c("Female", "Male"), p0 = 0.4, cc="yates")
  testValue      sig2      testUsed
1 -0.5151222 0.6064676 Normal approximation with Yates continuity correction
> |
```

3 Details of Calculations

3.1 The Test

$$z = \frac{x - \mu}{\sigma}$$

$$sig. = 2 \times (1 - Z(|z|))$$

With:

$$\mu = n \times p_0$$

$$\sigma = \sqrt{\mu \times (1 - p_0)}$$

The z value follows a standard normal distribution.

Symbols:

- x the number of successes in the sample
- p_0 the expected proportion (i.e. the proportion according to the null hypothesis)
- $Z(\dots)$ the standard normal cumulative density function.



3.2 Yates Continuity Correction

A Yates continuity correction is often recommended, which can be applied as:

$$z_{Yates} = \frac{|x - \mu| - 0.5}{\sigma}$$

Although most likely better would be:

$$z_{Yates} = \frac{\max(0, |x - \mu| - 0.5)}{\sigma}$$

4 Sources

The term ‘Score test’ and the formula were taken from IBM’s SPSS statistics algorithms documentation.

Score Z test statistic (Agresti, 2019):

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{r - np_0}{\sqrt{np_0(1-p_0)}}$$

(IBM SPSS Statistics Algorithms, 2021, p. 997)

Yates describes this for a 2x2 table:

tribution. This is equivalent to computing the values of χ^2 for deviations half a unit less than the true deviations, 8 successes, for example, being reckoned as $7\frac{1}{2}$, 2 as $2\frac{1}{2}$. This correction may be styled the *correction for continuity*, and the resultant value of χ denoted by χ' .

(Yates, 1934, p. 222)

References

IBM SPSS Statistics Algorithms. (2021). IBM.

Yates, F. (1934). Contingency tables involving small numbers and the chi square test. *Supplement to the Journal of the Royal Statistical Society*, 1(2), 217–235. <https://doi.org/10.2307/2983604>