One-Sample Binomial Test (test_binomial_os)

a.k.a. exact binomial test, one sample binomial test

Author: P. Stikker
Website: https://peterStatistics.com
YouTube: https://www.youtube.com/stikpet
Version: 0.2 (2023-01-06)

Introduction

The *test_binomial_os* function (and *test_binomial_os_arr* in VBA) can perform a one-sample binomial test. 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).

  - **twoSidedMethod** (default is "eqdist")
    Choice for method to calculate the two-sided significance value. Either:
    - "eqdist": equal distance method
    - "smallp": small p method
    - "double": double one-sided probability

1.2 Output:

- The **p-value** and the **name of the test** used. Except for the non-array version in VBA (Excel) which will only show the p-value.

- The array version in VBA (test_binomial_os_arr) requires two rows and two columns.

1.3 Dependencies

- **Excel**
  The VBA code will make use of the worksheetfunctions BinomDist and CountIf.
  You can run the **ts_binomial_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.
  - *scipy*
    the scipy.stats binom function (specifically binom.cdf and binom.pdf) are used in the calculations.

- **R**
  No other libraries required. The pbinom and dbinom functions are used, but are available in R itself.

2 Examples

2.1 Excel

	A	B	C	D	E	F	G	H
1	data		categories					
2	1		1		p0	0,4		
3	2		2					
4	2							
5	1							
6	2		method	p-value				
7	2		eqdist	0,64059	=ts_binomial_os($A$2:$A$20;$C$2:$C$3;$F$2;C7)			
8	1		smallp	0,49416	=ts_binomial_os($A$2:$A$20;$C$2:$C$3;$F$2;C8)			
9	1		double	0,61614	=ts_binomial_os($A$2:$A$20;$C$2:$C$3;$F$2;C9)			
10	2							
11	2			p-value	test			
12	2			0,64059	exact binomial, equal distance			
13	2							
14	2			D11:E12	=ts_binomial_os_arr($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 scipy.stats import binom
```

```
[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_binomial_os(data, codes, p0=0.4)
```

[4]:

	p-value (2-sided)	test
0	0.640588	one-sample binomial, with equal-distance method

```
[5]: ts_binomial_os(data, codes, p0=0.4, twoSidedMethod = "smallp")
```

[5]:

	p-value (2-sided)	test
0	0.494161	one-sample binomial, with small p method

```
[6]: ts_binomial_os(data, codes, p0=0.4, twoSidedMethod = "double")
```

[6]:

	p-value (2-sided)	test
0	0.616139	one-sample binomial, with double one-sided method

2.3 R

```
> ts_binomial_os(data, c("Female", "Male"), p0 = 0.5, twoSidedMethod="eqdist")
       sig2                                     testUsed
1 0.1670685 one-sample binomial, with equal-distance method
> ts_binomial_os(data, c("Female", "Male"), p0 = 0.5, twoSidedMethod="smallp")
       sig2                              testUsed
1 0.1670685 one-sample binomial, with small p method
> ts_binomial_os(data, c("Female", "Male"), p0 = 0.5, twoSidedMethod="double")
       sig2                                       testUsed
1 0.1670685 one-sample binomial, with double one-sided method
> 
```

3 Details of Calculations

The one-sided p-value is calculated using:

$$p_{one-sided} = B(n, n_{min}, p_0^*)$$

With:

$$n_{min} = min\{n_s, n_f\}$$

$$p_0^* = \begin{cases} p_0, & n_{min} = n_s \\ 1 - p_0, & n_{min} = n_f \end{cases}$$

Symbols:

- n is the number of cases in the analysis
- n_s is the number of successes.
- n_f is the number of failures
- p_0 is the probability for a 'success' according to the null-hypothesis
- p_0^* is the probability adjusted in case failures is used
- $B(\dots)$ the binomial cumulative distribution function

For a two-sided test three variations are possible.

3.1 Equal Distance method

The two-sided p-value using the equal distance method is calculated using:

$$p_{eq.dist} = B(n, n_{min}, p_0^*) + 1 - B(n, \lfloor 2 \times n_0 \rfloor - n_{min} - 1, p_0^*)$$

With:

- $n_0 = \lfloor n \times p_0 \rfloor$

Explanation

This method looks at the number of cases. In a sample of n people, we'd then expect $n_0 = \lfloor n \times p_0 \rfloor$ successes (we round the result down to the nearest integer. We only had n_{min}, so a difference of $n_0 - n_{min}$. The 'equal distance method' now means to look for the chance of having k or less, and $n_0 + n_0 - n_{min} = 2 \times n_0 - n_{min}$ or more. Each of these two probabilities can be found using a binomial distribution. Adding these two together than gives the two-sided significance.

3.2 Small p method

The two-sided p-value using the small p method is calculated using:

$$p_{small\ p} = B(n, n_{min}, p_0^*) + \sum_{i=n_{min}+1}^{n} \begin{cases} 0, & b(n, i, p_0^*) > b(n, n_{min}, p_0^*) \\ b(n, i, p_0^*), & b(n, i, p_0^*) \leq b(n, n_{min}, p_0^*) \end{cases}$$

Symbols

- $b(\ldots)$ binomial probability mass function

Explanation

This method looks at the probabilities itself. $b(n, n_{min}, p_0^*)$ is the probability of having exactly n_{min} out of a group of n, with a chance p_0^* each time. The method of small p-values now considers 'or more extreme' any number between 0 and n (the sample size) that has a probability less or equal to this. This means we need to go over each option, determine the probability and check if it is lower or equal. So, the probability of 0 successes, the probability of 1 success, etc. The sum for all of those will be the two-sided significance. We can reduce the work a little since any value below n_{min}, will also have a lower probability, so we only need to sum over the ones above it and add the one-sided significance to the sum of those.

3.3 Double Single

The two-sided p-value using the double method is calculated using:

$$p_{double} = 2 \times p_{one-sided}$$

Explanation

Fairly straight forward. Just double the one-sided significance.