

# Package ‘TableHC’

January 8, 2020

**Version** 0.1.2

**Date** 2019-12-21

**Encoding** UTF-8

**Title** Higher Criticism Test of Two Frequency Counts Tables

**Description** Higher Criticism (HC) test between two frequency tables. Test is based on an adaptation of the Tukey-Donoho-Jin HC statistic to testing frequency tables described in Kipnis (2019) <arXiv:1911.01208>.

**Author** Alon Kipnis <kipnisa1@stanford.edu>

**Maintainer** Alon Kipnis <kipnisa1@stanford.edu>

**License** MIT + file LICENSE

**RoxygenNote** 6.1.1

**Suggests** testthat, knitr, rmarkdown

**VignetteBuilder** knitr

**Imports** stats

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2020-01-08 10:50:02 UTC

## R topics documented:

HC.vals . . . . .	2
two.sample.HC . . . . .	3
two.sample.pvals . . . . .	4

<b>Index</b>	<b>6</b>
--------------	----------

---

 HC.vals

*Higher Criticism (HC) test*


---

### Description

Compute the HC statistic and the HC threshold given a list of P-values. Can be used with function [two.sample.pvals](#) to get a list of p-values discriminating each feature between the two tables.

stbl – normalize using expected p-value (stbl==True) or observed (stbl==False) alpha – lower fraction of p-values to use

### Usage

```
HC.vals(pv, alpha = 0.45, stbl = TRUE)
```

### Arguments

pv	A list of numbers between 0 and 1.
alpha	A number between 0 and 1.
stbl	A boolean.

### Value

A list containing the following fields: HC – Higher Criticism (HC) score HC.star – HC score corrected to finite sample p – p-value attaining HC p.star – p-value attaining HC.star

### Examples

```
tb1 = table(c(1,1,1,1,1,1,2,2,2,2,2,3,3,3,3,4,4,4,5,6,6,7,7,7))
tb2 = table(c(1,1,1,1,1,1,1,1,1,2,3,3,3,3,3,4,4,4,5,5,6))
PV = two.sample.pvals(tb1, tb2) # compute P-values
HC.vals(PV$pv) # combine P-values using the HC statistics
```

```
# Can be used to check similarity of word-frequencies in texts:
text1 = "On the day House Democrats opened an impeachment inquiry of
President Trump last week, Pete Buttigieg was being grilled by Iowa
voters on other subjects: how to loosen the grip of the rich on government,
how to restore science to policymaking, how to reduce child poverty. At an
event in eastern Iowa, a woman rose to say that her four adult children
were 'stuck' in life, unable to afford what she had in the 1980s when a
$10-an-hour job paid for rent, utilities and an annual vacation."
```

```
text2 = "How can the federal government help our young people that want to
do whats right and to get to those things that their parents worked so hard
for? the voter asked. This is the conversation Mr. Buttigieg wants to have.
Boasting a huge financial war chest but struggling in the polls, Mr.
Buttigieg is now staking his presidential candidacy on Iowa, and particularly
on connecting with rural white voters who want to talk about personal
concerns more than impeachment. In doing so, Mr. Buttigieg is also trying to
```

show how Democrats can win back counties that flipped from Barack Obama to Donald Trump in 2016 there are more of them in Iowa than any other state by focusing, he said, on “the things that are going to affect folks’ lives in a concrete way.”

```
tb1 = table(strsplit(tolower(text1), ' '))
tb2 = table(strsplit(tolower(text2), ' '))
pv = two.sample.pvals(tb1, tb2)
HC.vals(pv$pv)
```

---

two.sample.HC

*Higher Criticism (HC) Test Between two Tables*


---

### Description

Compute HC statistic directly from two one-way contingency tables. `stbl` – normalize using expected p-value (`stbl==True`) or observed (`stbl==False`) `alpha` – lower fraction of p-values to use

### Usage

```
two.sample.HC(tb1, tb2, alpha = 0.45, stbl = TRUE)
```

### Arguments

<code>tb1</code>	A one-way table with integer counts.
<code>tb2</code>	A one-way table with integer counts.
<code>alpha</code>	A number between 0 and 1.
<code>stbl</code>	A boolean.

### Value

A list containing the following fields: `HC` – Higher Criticism (HC) score `HC.star` – HC score corrected to finite sample `p` – p-value attaining HC `p.star` – p-value attaining `HC.star`

### Examples

```
text1 = "On the day House Democrats opened an impeachment inquiry of President Trump last week, Pete Buttigieg was being grilled by Iowa voters on other subjects: how to loosen the grip of the rich on government, how to restore science to policymaking, how to reduce child poverty. At an event in eastern Iowa, a woman rose to say that her four adult children were 'stuck' in life, unable to afford what she had in the 1980s when a $10-an-hour job paid for rent, utilities and an annual vacation."
```

```
text2 = "How can the federal government help our young people that want to do what's right and to get to those things that their parents worked so hard for? the voter asked. This is the conversation Mr. Buttigieg wants to have."
```

Boasting a huge financial war chest but struggling in the polls, Mr. Buttigieg is now staking his presidential candidacy on Iowa, and particularly on connecting with rural white voters who want to talk about personal concerns more than impeachment. In doing so, Mr. Buttigieg is also trying to show how Democrats can win back counties that flipped from Barack Obama to Donald Trump in 2016 there are more of them in Iowa than any other state by focusing, he said, on “the things that are going to affect folks’ lives in a concrete way.”

```
tb1 = table(strsplit(tolower(text1), ' '))
tb2 = table(strsplit(tolower(text2), ' '))
pv = two.sample.pvals(tb1, tb2)
HC.vals(pv$pv)
```

---

two.sample.pvals      *Feature-by-feature exact binomial test between two tables*

---

### Description

Align tables and use an exact binomial test (`binom.test`) on each feature. Alignment is done using "outer mergeing"; missing values are filled with zeros.

### Usage

```
two.sample.pvals(tb1, tb2)
```

### Arguments

tb1                    A one-way table with integer counts.  
 tb2                    A one-way table with integer counts.

### Value

table of pair of counts per feature and a p-value associated with each pair.

### Examples

```
tb1 = table(c(1,1,1,1,1,1,2,2,2,2,2,3,3,3,3,4,4,4,5,6,6,7,7,7))
tb2 = table(c(1,1,1,1,1,1,1,1,1,2,3,3,3,3,3,4,4,4,5,5,5,6))
PV = two.sample.pvals(tb1, tb2) # compute P-values
HC.vals(PV$pv) # use the Higher-Criticism to combine the P-values
# for a global test
```

# Can be used to check similarity of word-frequencies in texts:

```
text1 = "On the day House Democrats opened an impeachment inquiry of
President Trump last week, Pete Buttigieg was being grilled by Iowa
voters on other subjects: how to loosen the grip of the rich on government,
```

how to restore science to policymaking, how to reduce child poverty. At an event in eastern Iowa, a woman rose to say that her four adult children were 'stuck' in life, unable to afford what she had in the 1980s when a \$10-an-hour job paid for rent, utilities and an annual vacation."

text2 = "How can the federal government help our young people that want to do whats right and to get to those things that their parents worked so hard for? the voter asked. This is the conversation Mr. Buttigieg wants to have. Boasting a huge financial war chest but struggling in the polls, Mr. Buttigieg is now staking his presidential candidacy on Iowa, and particularly on connecting with rural white voters who want to talk about personal concerns more than impeachment. In doing so, Mr. Buttigieg is also trying to show how Democrats can win back counties that flipped from Barack Obama to Donald Trump in 2016 there are more of them in Iowa than any other state by focusing, he said, on "the things that are going to affect folks' lives in a concrete way."

```
tb1 = table(strsplit(tolower(text1),' '))
tb2 = table(strsplit(tolower(text2),' '))
pv = two.sample.pvals(tb1,tb2)
HC.vals(pv$pv)
```

# Index

HC.vals, [2](#)

two.sample.HC, [3](#)

two.sample.pvals, [2](#), [4](#)