

stmh — Calculate rate ratios with the Mantel–Haenszel method

[Description](#)
[Options](#)
[References](#)

[Quick start](#)
[Remarks and examples](#)
[Also see](#)

[Menu](#)
[Stored results](#)

[Syntax](#)
[Acknowledgments](#)

Description

`stmh` calculates stratified rate ratios and significance tests by using a Mantel–Haenszel-type method. You can also use `stmh` to carry out a trend test for a metric explanatory variable.

Quick start

Failure-rate ratio comparing category 1 with 0 in binary variable `a`, computed using the Mantel–Haenszel method

```
stmh a
```

Same as above, but controlling for `catvar`

```
stmh a catvar
```

Failure-rate ratios for each level of `catvar`, with test for unequal rate ratios

```
stmh a, by(catvar)
```

Same as above, but compare 4 with 3 in multivalued `b` at each level of `catvar`

```
stmh b, compare(4,3) by(catvar)
```

Test for a trend of failure rates with `x1` controlling for `b`

```
stmh x1 b
```

Menu

Statistics > Survival analysis > Summary statistics, tests, and tables > Tabulate Mantel–Haenszel rate ratios

Syntax

```
stmh varname [varlist] [if] [in] [, options]
```

<i>options</i>	Description
Main	
<code>by(<i>varlist</i>)</code>	tabulate rate ratio on <i>varlist</i>
<code>compare(<i>num1</i>,<i>den2</i>)</code>	compare categories of exposure variable
<code>missing</code>	include missing values as extra categories
<code>level(#)</code>	set confidence level; default is <code>level(95)</code>

You must `stset` your data before using `stmh`; see [ST] [stset](#).

`by` and `collect` are allowed; see [U] [11.1.10 Prefix commands](#).

`fweights`, `iwweights`, and `pweights` may be specified using `stset`; see [ST] [stset](#).

Options

Main

`by(varlist)` specifies categorical variables by which the rate ratio is to be tabulated.

A separate rate ratio is produced for each category or combination of categories of *varlist*, and a test for unequal rate ratios (effect modification) is displayed.

`compare(num1,den2)` specifies the categories of the exposure variable to be compared. The first code defines the numerator categories, and the second code defines the denominator categories.

When `compare()` is not specified and there are only two categories, the larger category is compared with the smaller one; when `compare()` is not specified and there are more than two categories, `stmh` analyzes log-linear trend.

`missing` specifies that missing values of the explanatory variables be treated as extra categories. The default is to exclude such observations.

`level(#)` specifies the confidence level, as a percentage, for confidence intervals. The default is `level(95)` or as set by `set level`; see [U] [20.8 Specifying the width of confidence intervals](#).

Remarks and examples

[stata.com](http://www.stata.com)

The `stmh` command is used for estimating rate ratios, controlled for confounding, using stratification. You can use it to estimate the ratio of the rates of failure for two categories of the explanatory variable *varname*. Categories to be compared may be defined by specifying the codes of the levels with `compare()`. Rate ratios and significance tests are calculated using a Mantel–Haenszel-type method.

The first variable listed on the command line after `stmh` is the explanatory variable used in comparing rates, and the remaining variables, if any, are categorical variables that are to be “controlled for” by using stratification. Strata are defined by cross-classification of these variables.

`stmh` may also be used to carry out trend tests for a metric explanatory variable. A one-step Newton approximation to the log-linear Poisson regression coefficient is also computed.

▷ Example 1: Stratified rate ratios

We want to examine the rate ratio of coronary heart disease (CHD) for individuals with high and low levels of caloric intake, while controlling for age. We will use data from a heart disease and diet survey that was analyzed in Clayton and Hills (1993). In this dataset, events such as cancer and CHD indicate failure, but we were interested only in CHD, which has been coded as `fail = 1, 3, or 13`. The variable `hienergy` is coded 1 if the total energy consumption is more than 2.75 Mcal and 0 otherwise.

First, we `stset` the data: `origin` is set to date of birth, making `age` the analysis time; and the scale is set to 365.25, so analysis time is measured in years. We also specify the failure codes for CHD.

```
. use https://www.stata-press.com/data/r18/diet, clear
(Diet data with dates)
. stset dox, origin(time dob) enter(time doe) id(id) scale(365.25)
> fail(fail==1 3 13)
(output omitted)
```

Below, we `stsplit` the data into 10-year age bands. The new variable, `ageband`, identifies each observation's age group. We then use `stmh` to compare the rate for `hienergy` level 1 with the rate for level 0, controlled for `ageband`.

```
. stsplit ageband, at(40(10)70) after(time=dob) trim
(26 observations trimmed because of lower bound)
(418 observations (episodes) created)
. stmh hienergy ageband
      Failure _d: fail==1 3 13
  Analysis time _t: (dox-origin)/365.25
      Origin: time dob
Enter on or after: time doe
      ID variable: id
      Note: ageband<=40 trimmed
```

```
Mantel-Haenszel estimate of the rate ratio
comparing hienergy==1 vs. hienergy==0
controlling for ageband
```

Rate ratio	chi2	P>chi2	[95% conf. interval]	
0.534	4.36	0.0369	0.293	0.972

We see that the rate of CHD for individuals with high levels of caloric intake is about half the rate for individuals with low levels of caloric intake, controlling for age.

We can also compare the effect of `hienergy` between jobs, controlling for `ageband`.

```
. stmh hienergy ageband, by(job)
      Failure _d: fail==1 3 13
      Analysis time _t: (dox-origin)/365.25
      Origin: time dob
      Enter on or after: time doe
      ID variable: id
      Note: ageband<=40 trimmed

Mantel-Haenszel estimates of the rate ratio
comparing hienergy==1 vs. hienergy==0
controlling for ageband
by job
```

job	Rate ratio	Lower	Upper
0	0.42	0.13	1.33
1	0.64	0.22	1.87
2	0.51	0.21	1.26

Note: Lower and Upper are bounds of 95% confidence intervals.

Overall Mantel-Haenszel estimate, controlling for **ageband** and **job**

Rate ratio	chi2	P>chi2	[95% conf. interval]	
0.521	4.88	0.0271	0.289	0.939

```
Approx. test for unequal RRs (effect modification): chi2(2) = 0.28
                                                    Pr>chi2 = 0.8695
```

Because the rate-ratio estimates are approximate, the test for unequal rate ratios is also approximate.

◀

► Example 2: Log-linear trend test for metric explanatory variables

`diet.dta` contains the height for each patient recorded in the variable `height`. We can test for a trend of heart disease rates with `height` controlling for `ageband` by typing

```
. stmh height ageband
      Failure _d: fail==1 3 13
      Analysis time _t: (dox-origin)/365.25
      Origin: time dob
      Enter on or after: time doe
      ID variable: id
      Note: ageband<=40 trimmed

Score test for trend of rates with height
Overall Mantel-Haenszel estimate, controlling for ageband
```

Rate ratio	chi2	P>chi2	[95% conf. interval]	
0.906	18.60	0.0000	0.866	0.948

```
Note: The Rate ratio estimate is an approximation to the rate ratio
for a one-unit increase in height.
```

`stmh` tested for trend of heart disease rates with `height` within age bands and provided a rough estimate of the rate ratio for a 1 centimeter increase in height—this estimate is a one-step Newton approximation to the maximum likelihood estimate. It is not consistent, but it does provide a useful indication of the size of the effect.

The rate ratio is less than 1, so there is evidence for a decreasing rate with increasing height (about a 9% decrease in rate per centimeter increase in height).



Stored results

`stmh` stores the following in `r()`:

Scalars

<code>r(rratio)</code>	overall rate ratio
<code>r(chi2)</code>	χ^2
<code>r(p)</code>	p -value
<code>r(level)</code>	confidence level
<code>r(lb)</code>	lower bound of confidence interval
<code>r(ub)</code>	upper bound of confidence interval
<code>r(chi2_unequal)</code>	χ^2 for test of unequal rate ratios with <code>by()</code>
<code>r(p_unequal)</code>	p -value for test of unequal rate ratios with <code>by()</code>

Macros

<code>r(expvar)</code>	explanatory variable
<code>r(explevels)</code>	levels of binary explanatory variable
<code>r(controlvars)</code>	control variables
<code>r(byvars)</code>	<code>by()</code> variables
<code>r(test)</code>	type of test

Matrices

<code>r(table)</code>	group-specific rate ratios
-----------------------	----------------------------

Acknowledgments

The original version of `stmh` was written by David Clayton (retired) of the Cambridge Institute for Medical Research and Michael Hills (1934–2021) of the London School of Hygiene and Tropical Medicine.

Nathan Mantel (1919–2002) was an American biostatistician who grew up in New York. He served in the U.S. Army from 1942 to 1946 and then joined the National Cancer Institute as a statistical consultant in 1947. He would remain there for over 26 years. During this time, he was presented with problems from many fields, which led to his diverse publications and contributions. He is remembered for his outstanding problem-solving skills and brilliance.

Mantel is best known for his 1959 collaboration with William M. Haenszel, which introduced the Mantel–Haenszel method for obtaining odds ratios for stratified data and a test of equality across strata. In 1966, Mantel developed the log-rank test and, through a collaboration in 1974, extended this to data involving transient states. In another collaboration, he helped develop a procedure for testing carcinogenic agents; this paper defined a safe dosage level that was temporarily adopted by the Food and Drug Administration. Mantel also contributed to the development of the polychotomous logistic regression model and methods used to explore temporal and spatial clustering of diseases.

Mantel received many honors recognizing his contributions. He was elected an honorary fellow of the Royal Statistical Society and received the Superior Service Award from the National Institutes of Health. He was also a fellow of the American Statistical Association, which in 2002 created a lifetime achievement award in his name.

William M. Haenszel (1910–1998) was an American biostatistician and epidemiologist. In 1932, he obtained a master's degree in statistics from the University of Buffalo. He then worked at several leading institutions of health, making a lasting impact worldwide.

From 1934 to 1947, he worked for the New York State Department of Health, after which he became the director of the Bureau of Vital Statistics in the Connecticut State Health Department. In 1952, he joined the National Institutes of Health, where he would serve as the chief of the Biometry Branch and head of the Biometric Section. During his time there, Haenszel collaborated with Nathan Mantel to develop the Mantel–Haenszel method. He also initiated a series of studies of migrant populations that explored the causes of diseases. Another major contribution was the instrumental role he played in developing SEER, the first national system for tracking cases of cancer and the potential causes.

Haenszel was a fellow of the American Public Health Association, the American Association for the Advancement of Science, and the American Statistical Association. He also served as chair of the Biometrics Section of the American Statistical Association. Toward the end of his career, Haenszel joined the Illinois Cancer Council and provided consulting services to the World Health Organization. Despite having Parkinson's disease, Haenszel helped tutor a research group that was developing a brain tumor registry.

References

- Clayton, D. G., and M. Hills. 1993. *Statistical Models in Epidemiology*. Oxford: Oxford University Press.
- Gail, M. H. 1997. A conversation with Nathan Mantel. *Statistical Science* 12: 88–97.
<https://doi.org/10.1214/ss/1029963424>.
- Hankey, B. 1997. A conversation with William M. Haenszel. *Statistical Science* 12: 108–112.
<https://doi.org/10.1214/ss/1029963427>.

Also see

[ST] **stci** — Confidence intervals for means and percentiles of survival time

[ST] **stir** — Report incidence-rate comparison

[ST] **stmc** — Calculate rate ratios with the Mantel–Cox method

[ST] **stptime** — Calculate person-time, incidence rates, and SMR

[ST] **strate** — Tabulate failure rates and rate ratios

[ST] **stset** — Declare data to be survival-time data

Stata, Stata Press, and Mata are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. Other brand and product names are registered trademarks or trademarks of their respective companies. Copyright © 1985–2023 StataCorp LLC, College Station, TX, USA. All rights reserved.

