Package ltable 2.0.2

Ocheredko Oleksandr

FUNCTIONALITY

1. Constructs tables of counts and proportions out of data sets.
2. Inserts table into Excel and Word documents using clipboard, into LaTeX, HTML, Markdown and reStructuredText documents by the knitr::kable agency.
3. Moulds table into acceptable for log-linear modeling data.frame.
4. Performs log-linear modeling.
5. Performs power analysis.

This version is coded in R language exclusively to support across system transportability

Construction of tables of counts and proportions out of data sets

Use function table_f():

table_f(data, datavars, type = 1, digits = 2, extended = FALSE, MV = FALSE, cb = FALSE)

Examples:

```
data(sdata, package="ltable")
sdata
```

```
##        a  b      c    d
## 1   TRUE NA   male    A
## 2     NA  1   male    B
## 3  FALSE  1   male    A
## 4   TRUE  1   male <NA>
## 5   TRUE  1   male    A
## 6   TRUE  2 female    B
## 7  FALSE  2 female    A
## 8  FALSE  2 female    B
## 9   TRUE  2 female    A
## 10 FALSE  2 female    B
## 11    NA NA   <NA> <NA>
## 12  TRUE  1   male    A
## 13 FALSE  1   male    B
## 14 FALSE  1   male    A
## 15  TRUE  1   male    B
## 16  TRUE  1   male    A
## 17  TRUE  2 female    B
## 18 FALSE  2 female    A
## 19 FALSE  2 female    B
## 20  TRUE  2 female    A
## 21 FALSE  2 female    B
## 22    NA NA   <NA> <NA>
```

```
lapply(sdata,class)
```

```
## $a
## [1] "logical"
##
## $b
## [1] "numeric"
##
## $c
## [1] "factor"
##
## $d
## [1] "character"
```

I built data.frame *sdata* with fields of different basic classes just for demonstration. No other meaning applies. Let's build a simple table across levels of *a*:

```
ltable::table_f(sdata, "a")
```

```
##   a:FALSE a:TRUE Total, N
## 1       9     10       19
```

One might have interest in *NA* values for there may be quite informative pattern across levels or levels combinations. Use *MV=TRUE*. It's a part of data exploration:

```
ltable::table_f(sdata, "a", MV=TRUE, ext=TRUE)
```

```
##   a:FALSE a:TRUE NA Total, N
## 1       9     10  3       22
```

Unrelated option *extended=TRUE* is used just to demonstrate that abundant args have no effect. If one wants to tabulate numerous factors it's important to arrange them properly in sequence of presentation delimited with comma ",". Sorted levels of all but last variable are rolled out vertically in indicated sequence, the last has its sorted levels spread by columns.

```
ltable::table_f(sdata, "b,c")
```

```
##            b c:female c:male Total, N
## 1          1        0      9        9
## 2          2       10      0       10
## sum Total, N       10      9       19
```

One can also obtain the table of frequencies by choosing arg *type* values { *2*, *3*, *4* } as shown below:

```
ltable::table_f(sdata, "a,c",
                type=2, digits=3)
```

```
##            a c:female c:male Total, p
## 1      FALSE    0.667  0.333         1
## 2       TRUE      0.4    0.6         1
## sum Total, p    0.534  0.466         1
```

```
ltable::table_f(sdata, "a,c",
                type=3, digits=2)
```

```
##            a c:female c:male Total, p
## 1      FALSE      0.6   0.33      0.47
## 2       TRUE      0.4   0.67      0.53
## sum Total, p        1      1         1
```

```
ltable::table_f(sdata, "a,c",
                type=4, digits=3)
```

```
##            a c:female c:male Total, p
## 1      FALSE    0.316  0.158    0.474
## 2       TRUE    0.211  0.316    0.527
## sum Total, p    0.527  0.474    1.001
```

One can include number of fields (variables):

```
options(width=40)
ltable::table_f(sdata, "a,b,c,d",
        type=2, digits=3)
```

```
##            a        b        c   d:A
## 1      FALSE        1   female     0
## 2      FALSE        1     male 0.667
## 3      FALSE        2   female 0.333
## 4      FALSE        2     male     0
## 5       TRUE        1   female     0
## 6       TRUE        1     male  0.75
## 7       TRUE        2   female   0.5
## 8       TRUE        2     male     0
## sum Total, p Total, p Total, p 0.562
##       d:B Total, p
## 1       0        0
## 2   0.333        1
## 3   0.667        1
## 4       0        0
## 5       0        0
## 6    0.25        1
## 7     0.5        1
## 8       0        0
## sum 0.438        1
```

arg value *extended=TRUE* adds margins of counts, applied only for proportions and frequencies, value is *FALSE* by default. In last two examples options(width) was used to accommodate tables:

```
options(width=40)
ltable::table_f(sdata, "b,c,a,d", type=2,
                digits=3, extended=TRUE)
```

```
##            b        c        a   d:A
## 1          1   female    FALSE     0
## 2          1   female     TRUE     0
## 3          1     male    FALSE 0.667
## 4          1     male     TRUE  0.75
## 5          2   female    FALSE 0.333
## 6          2   female     TRUE   0.5
## 7          2     male    FALSE     0
## 8          2     male     TRUE     0
## sum Total, p Total, p Total, p 0.562
##     Total, N Total, N Total, N     9
```

```
##       d:B Total, p Total, N
## 1       0         0         0
## 2       0         0         0
## 3   0.333         1         3
## 4    0.25         1         4
## 5   0.667         1         6
## 6     0.5         1         4
## 7       0         0         0
## 8       0         0         0
## sum 0.438         1        17
##         8        17        17
```

Transporting table into documents

One can paste table into clipboard by using arg *cb=TRUE*. To insert table into Word document one should first open Excel, choose left high corner of placement by mouse click and use copy and paste key combinations or click on the Copy and Paste icons (the clipboard), then open Word document, use Copy icon to place the table. $table\_f(sdata,\ "a,c",\ type = 2,\ digits = 3,\ cb = TRUE)$

Use knitr::kable() to import table to other available formats through .Rmd or other engines:

$t <- table\_f(sdata,\ "a,c",\ type = 2,\ digits = 3)$

$knitr :: kable(t)$

	a	c:female	c:male	Total, p
1	FALSE	0.667	0.333	1
2	TRUE	0.4	0.6	1
sum	Total, p	0.534	0.466	1

Transforming table into acceptable for log-linear modelling data.frame.

Use function tableToData():

$tableToData(\ tname,\ numerictype = "",\ orderedtype = ""\ )$

Example:

```
data(sdata, package="ltable")
stab<-ltable::table_f(sdata, "a,b,c")
```

```
sdat<-ltable::tableToData(stab,
                          numerictype ="b",
                          orderedtype="a,c")
sdat
```

```
##       a b      c Counts
## 1 FALSE 1 female      0
## 2 FALSE 2 female      6
## 3  TRUE 1 female      0
## 4  TRUE 2 female      4
## 5 FALSE 1   male      3
## 6 FALSE 2   male      0
## 7  TRUE 1   male      5
## 8  TRUE 2   male      0
```

```
lapply(sdat,class)
```

```
## $a
## [1] "ordered" "factor"
##
## $b
## [1] "numeric"
##
## $c
## [1] "ordered" "factor"
##
## $Counts
## [1] "numeric"
```

Arg *tname* is the name of table created by function *table_f()*. In both next args *numerictype* and *orderedtype* variable names separated by comma to be transformed to numeric or ordered factor classes. Variable "Counts" shouldn't be listed in both.

Log-linear modeling

Use function MCLogLin():

$MCLogLin(formula,\ data,\ contrasts = NULL,\ XLB = -100,\ XUB = 100,\ a = 0.1,\ b = 0.1,\ DIC = FALSE,\ pcov = FALSE,\ draw = 10000,\ burnin = 3000)$

Log-linear analysis features some advantages against $glm\{stats\}$, first of all due to stability of GSL IWLS algorithms that insures distinctly less biased covariances estimates, the pivot issue for implemented power analysis. In some instances hypothesis testing of higher order effects disagrees with that of *glm* on account of larger GSL estimated errors. Another though related enhancement is distinct better fit assessed by sum of squared differences between observed and expected counts.

Example

Let's begin with historical example of log-linear modeling with Tromboembolism Data. This case-control data first considered by Worcester, J. (1971). The data y[ijk] cross-classify thromboembolism and control patients (i=1 and 2 respectively) by two risk factors: oral contraceptive user (j =1 for user, j = 2 for non-user) and smoking (k=1 for smokers, k =2 for non-smokers). Test quantifies boosting effect of contraceptive on odds of thromboembolism using log-linear analysis. Reproduced grouped data frame with 8 rows of factors' levels combinations is given below. Factors are: smoking status (Yes, No), contraceptive usage (Yes, No), thromboembolism status (Trombol, Control).

```
data(tdata, package="ltable")
tdata
```

```
  smoker contraceptive   tromb Counts
1    Yes           Yes Trombol     14
2    Yes           Yes Control      2
3    Yes            No Trombol      7
4    Yes            No Control     22
5     No           Yes Trombol     12
6     No           Yes Control      8
7     No            No Trombol     25
8     No            No Control     84
```

Data has been used in subsequent model choice studies, such as Spiegelhalter and Smith (1982), Pettit and Young (1990), Congdon (2005).

Under the potentially informative priors used, the Bayes factor estimate was B_{21} = 23.8, quite strongly in favour of the smaller model with single interaction effect *contraceptive*thromboembolism* that was opted for consideration in example. The fact that the reduced model gives a close fit implies that the use of oral contraceptives indeed instigates the odds of thromboembolism, effect significancy supported by classical and MCMC based log-linear estimates. Further inclusion of third order interaction indicated that the use of oral contraceptives particularly among those who smoke, is a risk for thromboembolism, but for smokers who do not take the pill there is no excess risk.

Let's check hypothesis by compare output of $MCLogLin\{ltable\}$ function with that of $glm\{stats\}$ function:

```
resglm<-glm(Counts~ smoker +contraceptive +
         tromb + contraceptive*tromb,
         family="poisson",
         data=tdata)
```

Results of MCLogLin {ltable} modeling

```
resMCLogLin<-ltable::MCLogLin(Counts~smoker +
      contraceptive +tromb +contraceptive*tromb,
      data=tdata)
```

```
Call:
```

```
ltable::MCLogLin(formula = Counts ~ smoker + contraceptive +
    tromb + contraceptive * tromb, data = tdata)


Coefficients:
                                Estimate  Std.Error  |z-score|   Pr(>|z|)
(Intercept)                    4.425e+00  3.855e-01  1.148e+01  1.715e-30
smokerYes                     -9.759e-01  3.873e-01  2.520e+00  1.174e-02
contraceptiveYes              -2.400e+00  5.796e-01  4.142e+00  3.445e-05
trombTrombol                  -1.199e+00  5.090e-01  2.356e+00  1.846e-02
contraceptiveYes:trombTrombol  2.436e+00  7.894e-01  3.087e+00  2.025e-03
phi                            4.870e+00  7.571e-01  6.433e+00  1.251e-10


Model fit:
MCMC fitting
Samplers : Gibbs for expected counts, Slice for regr. coeff. and inv.var.par. phi
Language: R
Jacobian reciprocal condition number =  0.09201567
chisq/n =  0.02360806
Deviance=  0.0002946049
NULL Deviance=  12.7355
Log.likelihood=  -23.53307
AIC(1) =  57.06613
AIC(n) =  7.133267
BIC =  57.46334


Residuals report:
Row Ovserved Y Predicted Y Raw Residual Pearson Residual Anscombe Residual
 1          14      12.324        1.676            0.254            0.957
 2           2       2.423       -0.423           -0.222           -0.563
 3           7       7.451       -0.451           -0.104           -0.347
 4          22      22.708       -0.708           -0.062           -0.289
 5          12      13.632       -1.632           -0.227           -0.911
 6           8       7.629        0.371            0.084            0.278
 7          25      24.513        0.487            0.040            0.189
 8          84      83.332        0.668            0.017            0.120
```

Output also conveys info:

Jacobian reciprocal condition number measures the inverse sensitivity of the solution to small perturbations in the input data. It tends to zero as J tends to singularity indicating solution instability.")

The value of ch-squared per number of counts (chisq/n) approximately 1 indicates a good fit.) If chisq/n » 1 the error estimates obtained from the covariance matrix will be too small and should be multiplied by square root of chisq/dof .

Poor fit will result from the use of an inappropriate model, and the scaled error estimates may then be outside the range of validity for Gaussian errors.

BEWARE: Poor fit jeopardizes the validity of power analysis.

Results of glm {stats} modeling

```
options(width=80)
summary(resglm)$coefficients
```

```
                              Estimate Std. Error   z value     Pr(>|z|)
(Intercept)                   4.364196  0.1069522 40.805108 0.000000e+00
smokerYes                    -1.053150  0.1731305 -6.082984 1.179660e-09
contraceptiveYes             -2.360854  0.3308080 -7.136628 9.564814e-13
trombTrombol                 -1.197703  0.2017027 -5.937964 2.885828e-09
contraceptiveYes:trombTrombol 2.153215  0.4232558  5.087265 3.632639e-07
```

```
cat("Predicted Y:", sprintf("%.2f", resglm$fitted.values), "\n
Deviance residuals:", sprintf("%.2f", summary(resglm)$deviance.resid))
```

```
Predicted Y: 6.72 2.59 8.28 27.41 19.28 7.41 23.72 78.59

Deviance residuals: 2.45 -0.38 -0.46 -1.07 -1.78 0.21 0.26 0.60
```

Juxtaposing two results we have the same conclusion on effects, specifically on hypothesized second order interaction term *contraceptive*tromb*, though differences are conspicuous on a part of error terms, higher order effect in particular. Checking with other data sets the regularity holds, that is higher order effects estimates feature larger errors against glm $\{stats\}$ counterparts due to handling overdispersion. The same rests with *chisq/n* statistic, predicted counts, and residuals (*deviance residuals* are larger in glm $\{stats\}$ than *Anscombe Residuals*). Repercussion on power analysis is about to be demonstrated.

Power analysis

Outlines of offered power study methodology can be found in ISDSA[1] paper.

Use function MCPower():

$MCPower(formula,\ data,\ contrasts = NULL,\ XLB = -100,\ XUB = 100,\ a = 0.1,\ b = 0.1,\ scale\_min = 1,\ scale\_max = 5,\ effect,\ p\_alpha = 0.05,\ draw = 10000,\ burnin = 3000)$

formula

- Incorporation of formula based approach facilitates extracting true influence of hypothesized effect by catching other intermingled influences. It's up to investigator's acumen and experience in process under study to delineate and separate hypothesized effect by appropriate data collection design and model formulation.
- The issue resolved is contrasts that constitute effect. Mostly investigator is interested in contrasts rather than effect. Say, if one proceeds with clinical trial to test medicines A, B, C, D it's A (new drug) against traditional set that usually implied. If the optimal dosage is under consideration, they are contrasts that help out (average against min, max; max against others, etc.).

scale_min, scale_max

Indicate the range of sample sizes. *scale_min* is the smallest number of sample size scale range, 1 signifies the given data sample size (observed total counts). *scale_max* is maximal sample size considered in power analysis. 5 by default means 5 times observed counts. The inspected sample size range defined by scale_min - scale_max automatically is divided into 11 consecutive values investigated by function. Given the results one can change sample size range, for example to scrutinize some particular interval to ensure power and p-value.

effect

Represents quoted effect tested by hypothesis; it should be one from the model formula, of second or higher order, introduced by * delimiter, i.e., "y*x", "y1*y2*x1*x2", "y1*y2", etc.

p_alpha

Serves to signify Z to check simulated z-scores against in power analysis, 0.05 by default.

contrasts

Serves to choose types of contrasts to study effects of factors, the same with *glm* {*stats*}, orthogonal polynomials by default.

draw

Indicate number of samples to draw (chain length)

burnin

Indicate number of initial samples to discard. *draw* should exceed *burnin* by at least 3000.

Example

Let's begin with Tromboembolism Data.

```
options(width=40)
pres<-ltable::MCPower(Counts~
      smoker +contraceptive +tromb +
      contraceptive*tromb, data=tdata,
      effect="contraceptive*tromb",
      scale_min=0.4, scale_max=1.5)
ltable::print(pres, choice="power")
```

```
Effect:  contraceptiveYes:trombTrombol
```

[1] https://meeting.isdsa.org/index.php/isdsa/2019/paper/viewPaper/3

Test statistic Z: Quantiles

Sample size:	Q0.025	Q0.05	Q0.1	Q0.2	Q0.3	Q0.4	Q0.5
70	0.929	1.203	1.532	1.810	2.102	2.479	2.722
89	1.151	1.264	1.663	1.933	2.200	2.452	2.735
108	0.738	0.901	1.512	2.116	2.367	2.672	2.898
127	1.315	1.403	1.597	2.286	2.589	2.883	3.053
146	1.059	1.509	1.754	2.211	2.658	2.845	2.988
165	1.380	1.615	1.938	2.248	2.528	2.850	3.080
184	1.255	1.592	1.899	2.404	2.759	3.089	3.315
204	1.153	1.365	1.693	2.385	2.694	3.002	3.148
223	0.981	1.695	2.001	2.343	2.718	2.923	3.214
242	1.461	1.633	2.075	2.455	2.908	3.163	3.395
261	1.287	1.498	1.852	2.104	2.685	3.138	3.354

Power: Quantiles

Sample size:	Q0.025	Q0.05	Q0.1	Q0.2	Q0.3	Q0.4	Q0.5
70	0.75	0.78	0.80	0.82	0.83	0.84	0.86
89	0.77	0.80	0.80	0.82	0.84	0.86	0.86
108	0.80	0.82	0.84	0.86	0.88	0.90	0.90
127	0.84	0.84	0.86	0.90	0.90	0.92	0.92
146	0.86	0.88	0.88	0.90	0.92	0.92	0.92
165	0.86	0.88	0.88	0.90	0.92	0.93	0.94
184	0.87	0.88	0.90	0.92	0.92	0.94	0.94
204	0.88	0.90	0.90	0.92	0.92	0.94	0.94
223	0.90	0.90	0.90	0.92	0.94	0.94	0.96
242	0.88	0.88	0.90	0.92	0.94	0.96	0.96
261	0.90	0.92	0.92	0.94	0.94	0.96	0.96

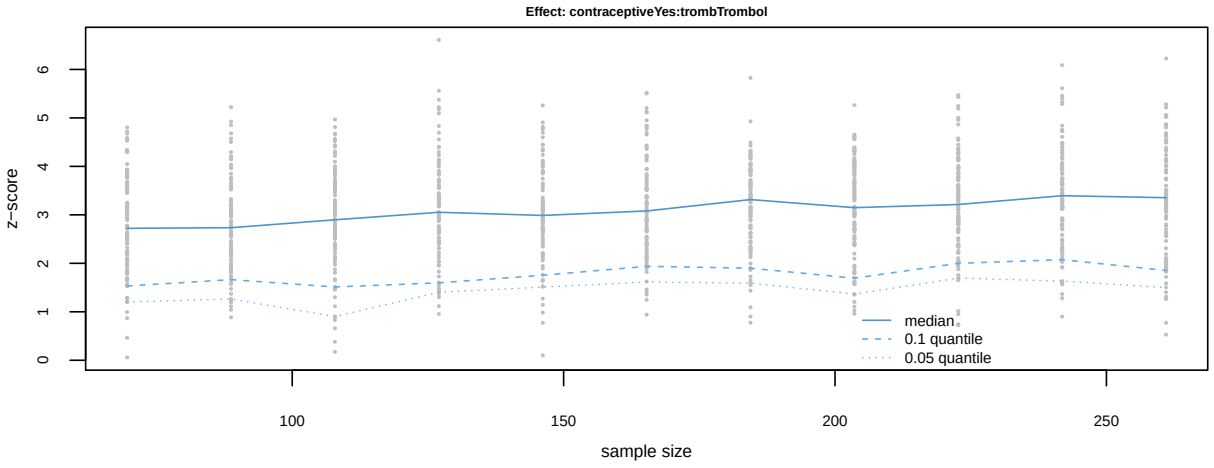
What we can deduce from the result is that 235 total counts is enough to secure *alpha* and *beta* errors. I suggest the most secure Q0.025 quantile to weight decision on. So 235 secures Z=1.96 and power 0.9 given Q0.025 estimates. Results of power analysis backed up with MCMC delivered approach, see Ocheredko O.M. MCMC Bootstrap Based Approach to Power and Sample Size Evaluation.[2].

Discussion
The log-linear estimates of *contraceptiveYes*trombTrombol* effect tested to be significant. Is it not strong enough evidence of association? Why should we collect almost 1.5-fold as many data? The answer of course is related to the specifics of the sample. The basic design itself is a sample, not status quo that represents true frequencies ratios in population. Therefore, we have to secure that the sample data brings in enough information to overpower sample specifics. Of course, the more complex design is the larger sample variation has to be outbalanced by signal, the larger sample size is required.

The original data is one of the random snapshots of reality and we have to put as much credit as sensible to it. Not all snapshots of size 174 guarantee a 95% CI with zero excluded. Using BUGS MCMC realization it was indicated that the sample size of 260 affords enough power to assure the significance of the association in practically all samples. The same logic is behind any application of power analysis.

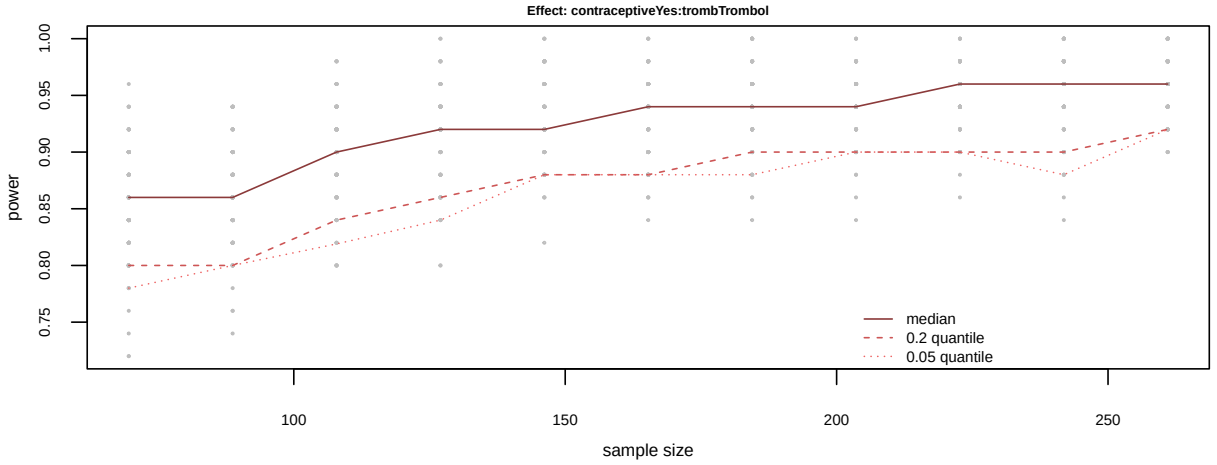
The other lay belief is that with the increase of sample size any association is doomed to be significant. For sure, it is not, and the strength of power analysis is to determine the optimal sample size of hypothesis testing. The power analysis assures that given H_0 is true there is no prospect of decisive augmentation of power and significance following the increase in sample size that will shortly be demonstrated. Before turning to another example the graphic output produced by function *plot* {*ltable*} is paneled:

```
ltable::plot(pres, stencil=1)
```

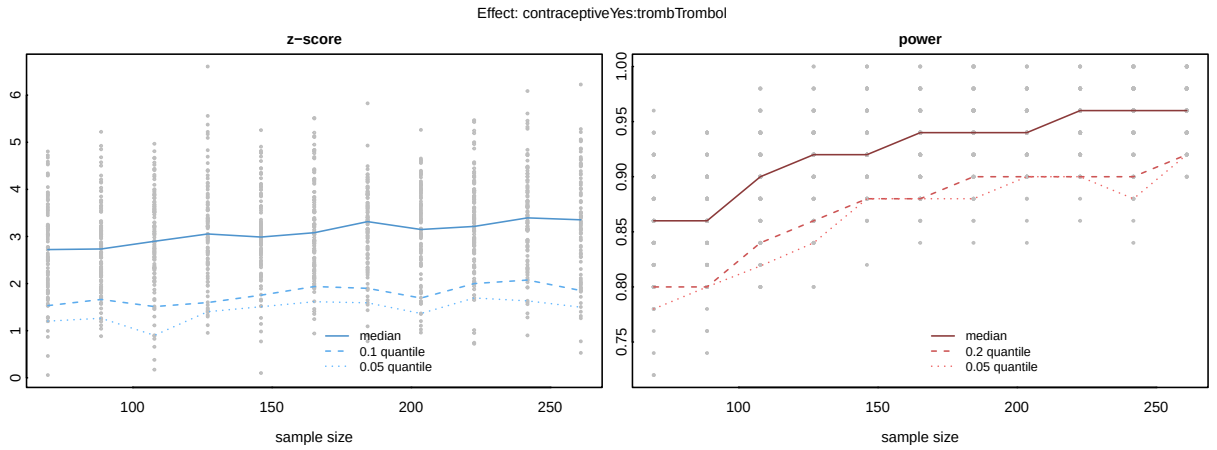


[2]https://www.amazon.com/gp/product/1946728039/

```
ltable::plot(pres, stencil=2)
```



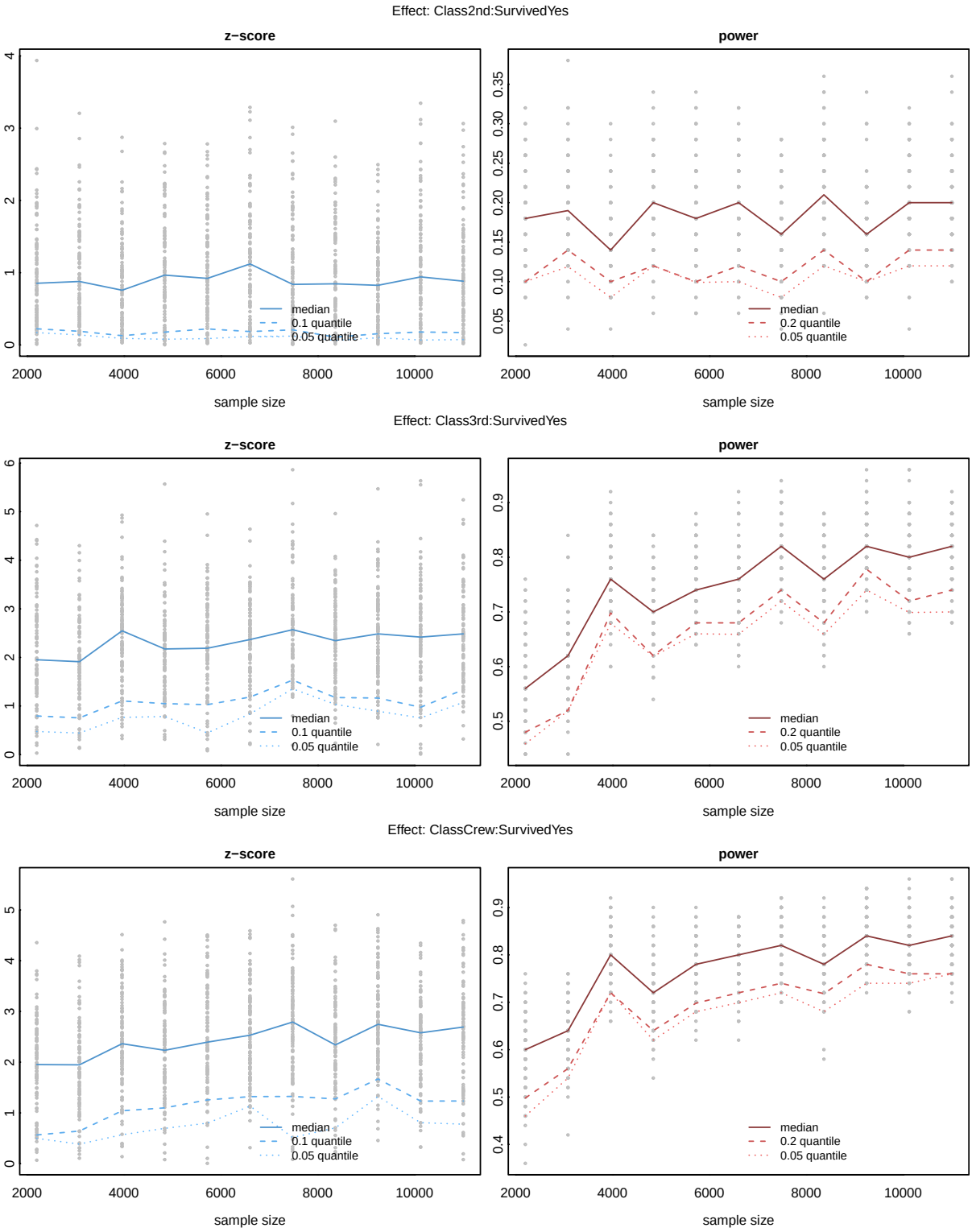

```
ltable::plot(pres, stencil=3)
```



Example

This is example of no observed association

```
TitanicData<-as.data.frame(datasets::Titanic)
names(TitanicData)[5]<-"Counts"
pres<-ltable::MCPower(Counts~Class+Age+Survived+Class*Survived, a=0.1, b=0.1,
draw=10000, data=TitanicData, effect="Class*Survived")
ltable::plot(pres, stencil=3)
```


Effect: Class2nd:SurvivedYes
z−score
power
median
0.1 quantile
0.05 quantile
0.2 quantile
sample size
Effect: Class3rd:SurvivedYes
Effect: ClassCrew:SurvivedYes

Let's consider Titanic data, available in package *datasets* and accessible by *datasets::Titanic*. This data set provides information on the fate of passengers on the fatal maiden voyage of the ocean liner 'Titanic', summarized according to economic status (class), sex, age and survival. Many well-known facts—from the proportions of first-class passengers to the 'women and children first' policy, and the fact that that policy was not entirely successful in saving the women and children in the third class—are reflected in the survival rates for various classes of passenger. Let's conduct power analysis focused on effect of Class (1st, 2nd, 3rd, crew) of passenger on Survival (Yes, No). From the graphical output it's obvious that survival doesn't show significant difference between 3rd and 2nd passengers accommodations and there is *no way* to prove its significance by augmenting the sample. Indeed this is example of impossibility to consider sample size expansion. So why not to put it to rest? Just because absence of significance can be ascribe to small sample size. Having support of power analysis we are perfectly aware that should we have opportunity to enlarge the sample test would not change. The opposite conclusion is driven by power analysis on survival differences between 3rd class and 1st class passengers as well as between 3rd class passengers and crew. In particular illustrative is 3rd class and 1st class passengers difference which non-significance indeed can be explained by sheer paucity of information. Should we be able to expand sample the difference would augment its significance to the point of being significant. As demonstrated by power curve the chance to detect it would be around 80% .

What do we make of it?

1. There is no chance to observe significant association by accumulating data if used tabulated design reproduces natural frequencies that indicate no natural relationship.
2. There is no increase in both significance and power with sample size growth given H_0 is true.
3. Power and significance may behave differently with sample size dynamic, so that we can't play one against the other as classical power methodology implies. Usually one is less responsive than another and it is former that defines necessary data load.

What is there under the hood?

The clue is Hessian estimate that provides error terms (for testing complex effect relevant covariance structure is used). The Hessian decomposition can be shown is the sum of two components. The first is

$$-\frac{\psi * e^{\beta * \boldsymbol{X}}}{(e^{\beta * X}) + \psi}\boldsymbol{X}\boldsymbol{X}^T$$

It helps to understand errors dynamic with growing sample size. The only growing constituent is $e^{\beta * \boldsymbol{X}}$ which substantiates slight (dependent on NB2 inverse dispersion par ψ and sample size) initial decrease and then flatten.

Second component is proportionate to ratio of difference between observed and expected counts to expected counts. Therefore if the model leaves small residuals or constant ratio with growing sample size the addend has no influence on errors dynamic.

If regression effect is influential and significant it grows in magnitude with growing sample size. In such case given stability of error we would have increasing test Z-score. Effect would not gain magnitude in the absence of influence and we would have flatten test curve.

Overview of approaches to power calculus of tabulated data

Two approaches regularly suggested are:
1. Logistic regression approach with effect size log odds ratio
2. Contingency table approach with effect size based on noncentrality parameter for chi-square distribution

1. Logistic regression approach
Formulas for sample size *n* use a guess for $\hat{\pi} = \pi(\overline{x})$ and the distribution of *X*. The effect size is the log odds ratio τ comparing $\pi(\overline{x})$ to $\pi(\overline{x} + s_x)$, the probability at a standard deviation above the mean of x. For a one-sided test when *X* is approximately normal, Hsieh (1989)[3] derived

$$n = [z_\alpha + z_\beta * exp(-\tau^2/4)]^2(1 + 2\hat{\pi}\delta)/(\hat{\pi}\tau^2),$$

where

$$\delta = [1 + (1 + \tau^2)exp(5\tau^2/4)]/[1 + exp(-\tau^2)/4].$$

The value *n* decreases as $\hat{\pi} \to 0.50$ and as $|\tau|$ increases.

Given several predictors first multiple correlation **R** is calculated between the predictor *X* of interest and the others in the model. Then formula for *n* divides by $(1 - \boldsymbol{R}^2)$. In that formula, $\hat{\pi}$ is evaluated at the mean of all the explanatory variables, and the odds ratio refers to the effect of *X* at the mean level of the other predictors.

2. Contingency table approach[4]
When hypotheses are false, squared normal and *chi-square* and G^2 statistics have large-sample noncentral chi-squared distributions. Suppose that H_0 is equivalent to model M for a contingency table. Let π_i for model M converges, where $\sum_i \pi_i = \sum_i \pi_i(M) = 1$. For a multinominal sample of size *n*, the noncentrality parameter for *chi-square* statistic equals

$$\lambda = n \sum_i \frac{[\pi_i - \pi_i(M)]^2}{\pi_i(M)}$$

This has the same form as *chi-square* statistic, with π_i in place of the sample proportion p_i and $\pi_i(M)$ in place of $\hat{\pi}_i$. The noncentrality parameter for G^2 equals

$$\lambda = 2n \sum_i \pi_i log \frac{\pi_i}{\pi_i(M)}$$

When H_0 is true, all $\pi_i = \pi_i(M)$. Then, for either statistic, $\lambda = 0$ and the ordinary (central) chi-squared distribution applies. Finally, power equals

$$P[\chi^2_{\nu,\lambda} > \chi^2_\nu(\alpha)]$$

These two approaches to power calculus of tabulated data suffer from important flaws:
1. No design information incorporated (**XX**)
2. No overdispersion/heterogeneity parameters
3. α and β errors are interchangeable
4. No accommodation of growing magnitude of effect size with growing sample

How to read power/test curves

See-saw dynamic of either power or test curves is caused by Jacobian singularity, that indicates solution instability.
Flat profiles given low test or power values are indicative for insignificance of tested effect.
Flat profiles with z-values above 2 or power values that exceed 0.8 are indicative for significance of tested effect. On such occasions decrease both scale parameters to inspect smaller samples.

[3] Hsieh, F. (1989). Sample size tables for logistic regression. Statistics in Medicine. Volume 8, Issue 7. P. 795-802

[4] Agresti, A. (2013). Categorical Data Analysis. 3rd ed. (Wiley series in prob. and stat.; 792).