Package ltable 2.0.3. Part 2.

Ocheredko Oleksandr

Content:

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. Molds table into acceptable for log-linear modeling data.frame, co.
4. Performs log-linear modeling.
5. Performs power analysis.

- This version is coded in R language exclusively to support across-systems portability.
- Log-linear and power analyses are enhanced with ability to model risks (rates) and relative risks (standardized ratios). Modelling survival data with interval censoring is also supported.

Power analysis

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

Use function MCPower():

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

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 ll 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", "yl*y2*xl*x2", "yl*y2", etc.

offset

Permits to model risks, relative risks, interval censored survival data

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)
require(ltable); data(tdata)
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")
```

```
                                          242              0.88    0.90   0.92    0.94    0.94    0.9
Effect:  contraceptiveYes:trombTrombol    261              0.92    0.94   0.94    0.96    0.96    0.9
                                          *Sample size in models with offset indicates number of case
```

Sample size*:	Coefficients	Errors
70	2.38983	0.91973
89	2.41292	0.85772
108	2.44167	0.81783
127	2.41904	0.78866
146	2.41614	0.76497
165	2.39464	0.74916
184	2.38299	0.73449
204	2.42021	0.72228
223	2.39745	0.71187
242	2.41702	0.70326
261	2.45393	0.69661

Test statistic Z: Quantiles

Sample size*:	Q0.025	Q0.05	Q0.1	Q0.2	Q0.3	Q0.4	Q0.5
70	0.584	0.954	1.072	1.535	1.888	2.140	2.427
89	1.065	1.246	1.427	1.887	2.068	2.416	2.600
108	1.062	1.479	1.681	2.125	2.346	2.723	3.027
127	1.160	1.459	1.758	2.158	2.434	2.832	3.024
146	1.044	1.501	1.827	2.325	2.670	2.858	3.116
165	1.618	1.811	2.061	2.492	2.848	3.083	3.258
184	1.071	1.183	1.537	2.206	2.578	2.939	3.203
204	1.403	1.674	2.248	2.643	2.929	3.199	3.488
223	1.795	2.034	2.367	2.634	2.913	3.072	3.328
242	1.693	1.984	2.205	2.699	3.134	3.325	3.577
261	1.800	2.150	2.580	2.967	3.255	3.534	3.760

Power: Quantiles

Sample size*:	Q0.025	Q0.05	Q0.1	Q0.2	Q0.3	Q0.4	Q0.5
70	0.72	0.74	0.78	0.80	0.81	0.82	0.84
89	0.78	0.80	0.82	0.82	0.86	0.86	0.88
108	0.82	0.84	0.86	0.88	0.89	0.90	0.92
127	0.84	0.84	0.86	0.90	0.90	0.92	0.94
146	0.86	0.88	0.88	0.90	0.92	0.92	0.94
165	0.88	0.88	0.90	0.92	0.92	0.94	0.94
184	0.88	0.90	0.90	0.92	0.92	0.94	0.96
204	0.89	0.90	0.92	0.92	0.94	0.95	0.96
223	0.90	0.90	0.92	0.94	0.94	0.96	0.96

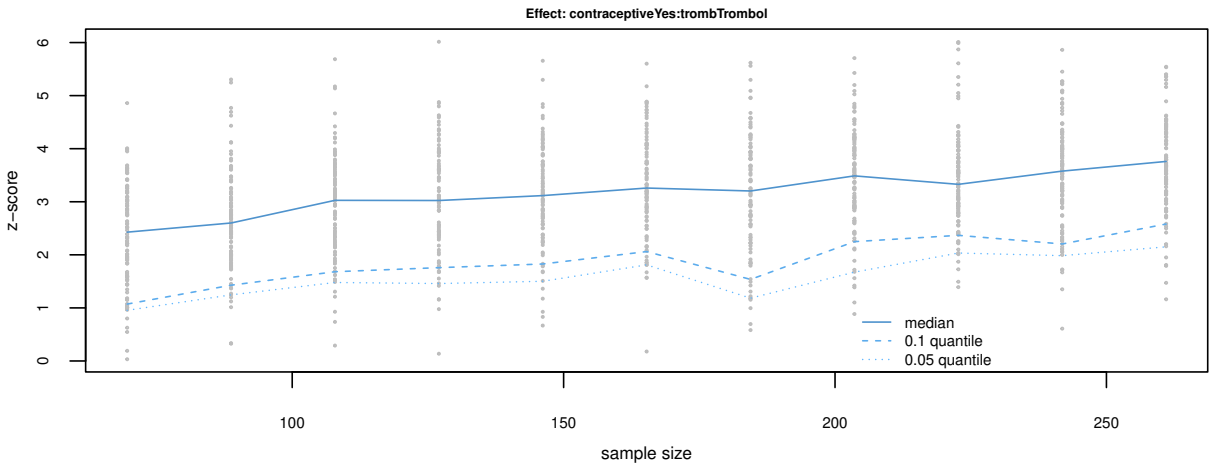
What we can deduce from the result is that 235 total counts is enough to secure *alpha* and *beta* errors. I suggest the most secure Q0.025 quantile to weight decision on. So 235 secures Z=1.96 and power 0.9 given Q0.025 estimates. Results of power analysis backed up with MCMC delivered approach, see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

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.

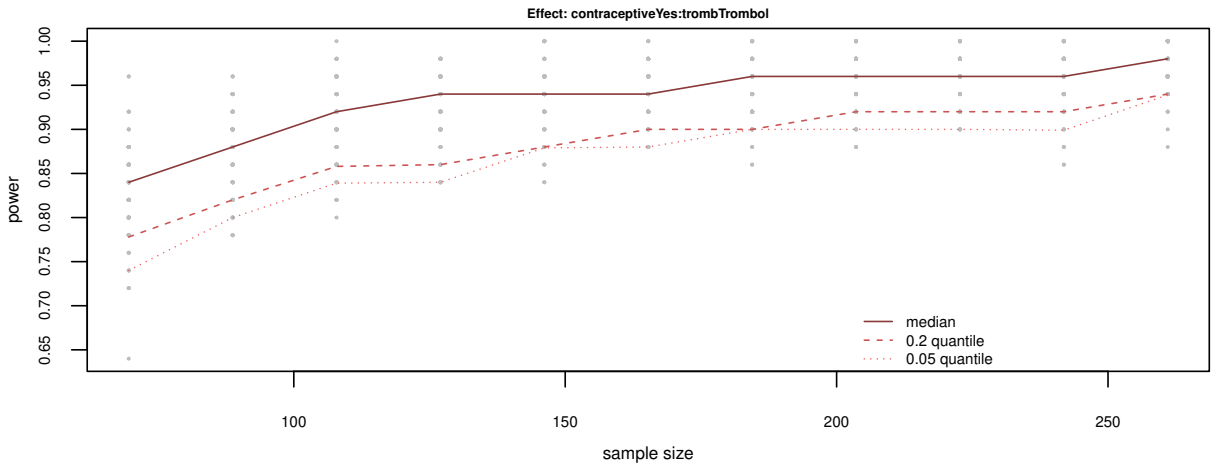
The original data is one of the random snapshots of reality and we have to put as much credit as sensible to it. Not all snapshots of size 174 guarantee a 95% CI with zero excluded. Using BUGS MCMC realization it was indicated that the sample size of 260 affords enough power to assure the significance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

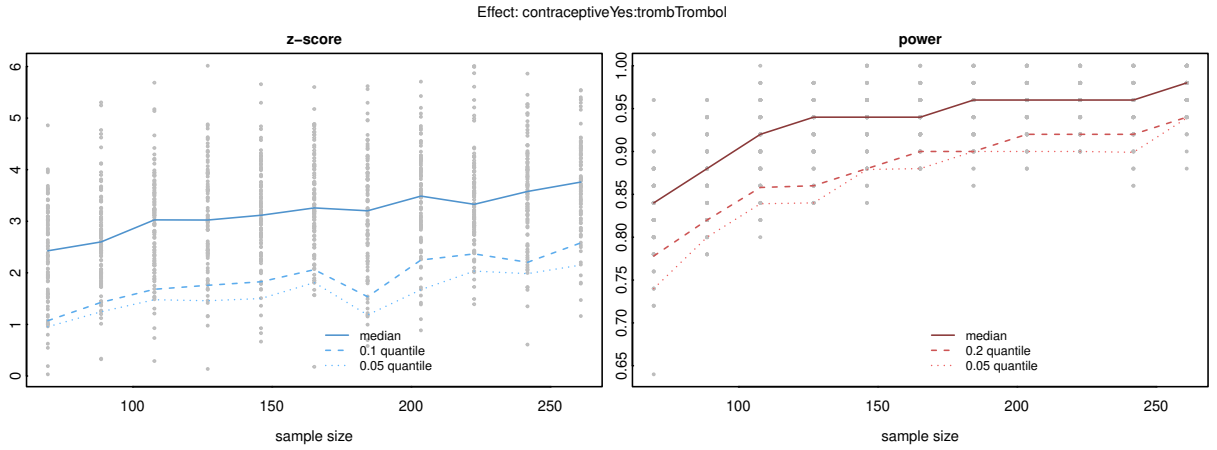
```
ltable::plot(pres, stencil=1)
```




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



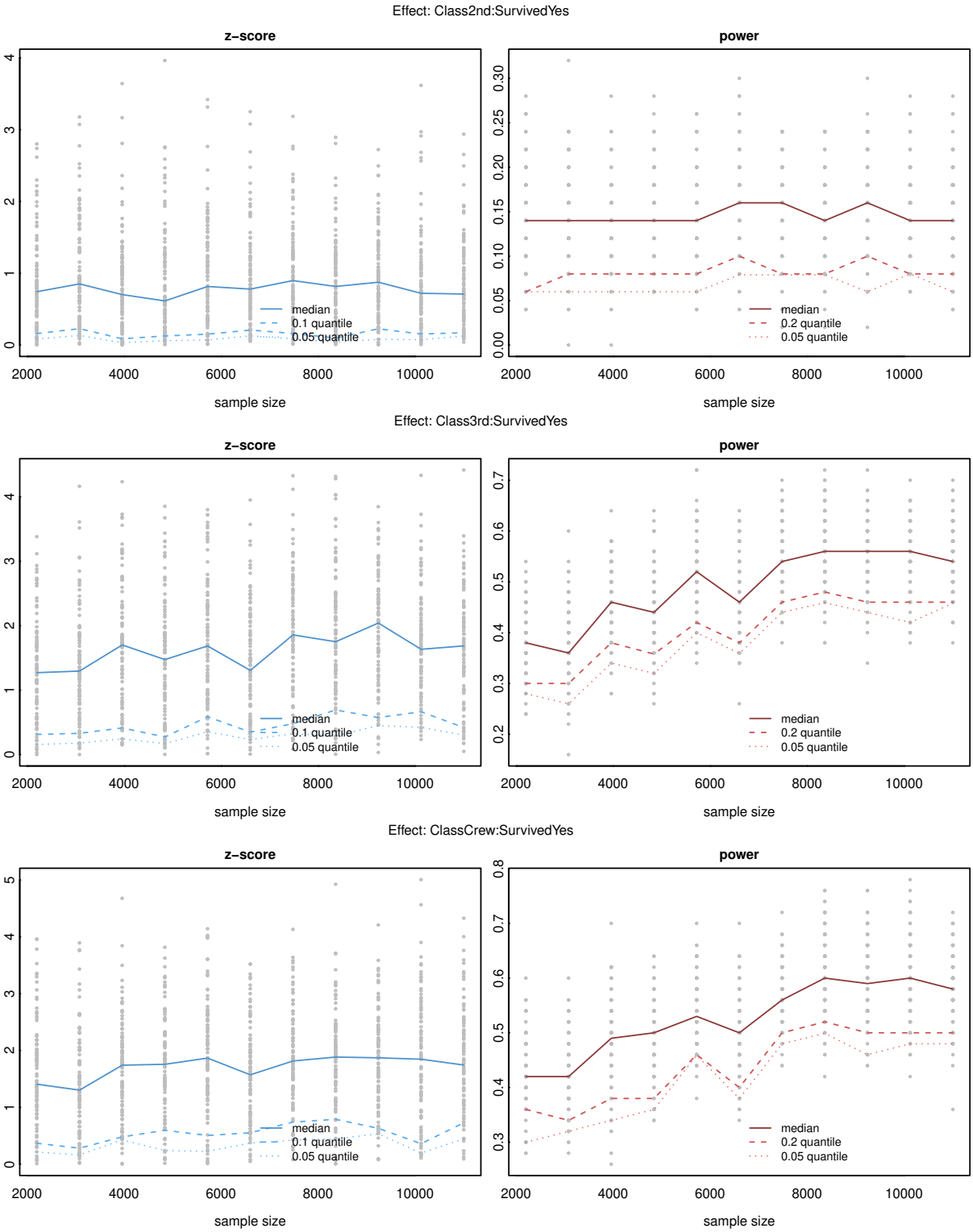

```
ltable::plot(pres, stencil=3)
```



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 (lst, 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 lst class passengers as well as between 3rd class passengers and crew. In particular illustrative is 3rd class and lst 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.

How overdispersion influences power curve?

Roughly overdispersion is in effect with par ψ less than 10. With ψ less than 1 it strongly influences p-values. Overdispersion is caused by data heterogeneity and by unbalanced designs. It takes place almost in all real world data sets. ***Overdispersed data are more required as to sample size.*** Two effects of ovedispersion are:

- It's obvious from formula, that with decrease in ψ errors of regression effects increase.
- Furthermore, the larger overdispersion, the slower increase of power curve and more rapidly it levels off.

Overall, power analyses that don't consider overdispersion are too optimistic. The magnitude of effect in question should be sufficient to overcome overdispersion. Due to overdispersion weak effects engender power curves that never reach predestined levels of power . To elicit such one needs design refinement.

Overview of approaches to power calculus of tabulated data

Two approaches regularly suggested are:
1. Logistic regression approach with effect size log odds ratio (*ad-hoc* power analysis).
2. Contingency table approach with effect size based on noncentrality parameter for chi-square distribution (*post-hoc* power analysis).

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)[1] 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[2]
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-

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

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

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.