Journal of Bibliographical Analyses in Statisctics

Article

How to simulate Rasch data in two scenarios of dichotomous and polytomous responses

Po-Chih Lai[1], Tsair-Wei Chien[2]

[1] Private Deguang Catholic High School, Tainan, Taiwan; 940208963852bob@gmail.com
[2] Department of Medical Research, Chi-Mei Medical Center, Tainan 700, Taiwan; smile@mail.chimei.org.tw
* Correspondence: smile@mail.chimei.org.tw: +8-866-281-2811

Abstract: The cumulative annual number of publications continually grown was frequently applied using a literature logical growth curve and the corresponding inflection point. However, none was found using a scientific approach to verigy the way for determinating the inflection point. We use the second order derivative in calculus to prove the way that is suitable for use in determination of inflection point. The aims of this study include that verifying the efficacy of the exponential growth model on accumulative publications of mobile health research between 1997 and 2017 in literature. We observed that the model accuracy(R2 = 0.99) is higher than the one(R2 = 0.98) in literature, based on identical data The exponential growth model can be applied to other disciplines for helping us predict the oucomes in the future. The inflection curve provides us a deeper insight into the ogive curve that represents the trajectory and trend of interest we concern about in practice.

Keywords: inflection point, exponential growth model**, calculus; ogive curve**

Citation: Lai PC, Chien TW. The determination of inflection curve on a given ogive curve using the second order derivative in calculus. *J Bibliographical Analyses in Statistics* **2021;** *18(3): 31-33*.

Received: 2 July 2021
Accepted: 12 July 2021
Published: 2021/07/21

Publisher's Note: Coding stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Introduction

The study[1] applied Rasch[2] simulated data[3] with known item parameters from published papers to yield person measures when person abilities normally distributed. The question is how to program codes for simulating Rasch data. Although Linacre[3[taught us generate those dichotomous and plolytomous responses in data, no such research on codes with visual basic for application (VAB) in Micro soft Excel. We tend to demonstrate the way to generate Rasch simulated data following the guideline[3] and provide a deeper insights into the core of data simulated in MS Excel.

Methods

2.1 Data source

A virtue test was designed with 20 persons and 7 items, following normal and uniform distribution, respectively, in true scores and item difficulties. Two sets of responses consisting of dichotomous(i.e., binary) and polytomous(i.e., from 0 to 4, the higher means more tendency toward success or satisfaction) were manipulated in this study.

2.2 The procedure of simulating Rasch data

Based on Linachre's guideline on the simulated data, the dichotomous data were come from item difficulies uniformly distributed and person measures noremally distributed. The item mean is set at 0 logits in a range from -3 to 3 logits. Owning to 20 persons and 7 items observed in the data, a total of 140(=20 $\times$ 7) cells are requied to obtain the corresponding probability for each response by a person measure to an item difficulty. The rule of a binary response determined is bases on the following steps:

(1) generate a random number U = uniform [0,1]; (2) compare the probability of failure [= 1/(1 + exp(ability - difficulty))]< U, then X=0. Otherwise. X=1, where X is the simulated observation. The constraind conditions were set at two exterme scenarios in person abilities for a very high ability person (logit = 10) to assight X =1, and for a very low ability (logit = -10): to endore X=-0.

Similarly, the polytomous (rating scale or partial credit) data are obtained by the equations (1) to (5) below:

First, let $X_{ni} = x \in \{0, 1, \ldots, m_i\}$,(1), be an integer random variable where m_i is the maximum score for item i. That is, the variable X_{ni} is a random variable that can take on integer values between 0 and a maximum of m_i. In the polytomous Rasch model[4], the probability of the outcome $X_{ni} = x$ is

$$\Pr\{X_{ni} = x, x > 0\} = \frac{\exp \sum_{k=1}^{x} (\beta_n - \tau_{ki})}{1 + \sum_{j=1}^{m_i} \exp \sum_{k=1}^{j} (\beta_n - \tau_{ki})}$$

$$\Pr\{X_{ni} = 0\} = \frac{1}{1 + \sum_{j=1}^{m_i} \exp \sum_{k=1}^{j} (\beta_n - \tau_{ki})}$$

.(2),(3)

Where τ_{ki} is the kth threshold location of item i on a latent continuum, β_n is the location of person n on the same continuum, and m_i is the maximum score for item i. These equations are the same as the one bilow:

$$\Pr\{X_{ni} = x\} = \frac{\exp \sum_{k=0}^{x} (\beta_n - \tau_{ki})}{\sum_{j=0}^{m_i} \exp \sum_{k=0}^{j} (\beta_n - \tau_{ki})}$$

,(4)

Where the value of τ_{oi} is chosen for computational convenience that is:

$$\sum_{k=0}^{m_i} (\beta_n - \tau_{ki}) \equiv 0.$$

, (5).

According to Linachre's guideline on the simulated data for the polytomous respnses(e.g., 0 to 4 in this study) with 20 persons and 7 items, similar to the dichotomous scenarua. Due to the number of categories is 4(i.e., m=4 in Eq. 5). the higher categories, 2 to m, have Rasch-Andrich threshold values that are usually ascending and sum to zero across all the categories. Accordingly, simulate the threshold values for each response by a person to an item with steps below:

(1)benerate a random number U = uniform [0,1]; (2)compute the cumulative exponential of observing each category:

measure = 0

cumexp(1) = 1

Compute for category j = 2 to m

```
    measure = measure + ability - difficulty - threshold(j)
    cumexp(j) = cumexp(j-1) + exponential(measure)
  Next category
```

Where threshold(j) are assigned as {-1.5, -0.5, 0.5, 1.5}.

(3) identify the simulated observation by the criteria and judgment below:

```
    U = U * cumexp(m)
  For category j = 1 to m
    if U <= cumexp(j) then X = j: exit
  Next category
```

(4) X is the simulated observation.

(5) check this by simulating data for a very high ability person (logit = 10): the data should all be "m" (the top category). Meanwile, simulate data for a very low ability person (logit = -10): the data should all be "0" (the bottom category).

Results

(1) Dichotomous responses:

The codes are listed below:

```
Sheets("a").Range("B2:K100").ClearContents
    y_count = Sheets("a").Range("A225536").End(xlUp).Row - 1
    mrange = "wfd" & 1
    mcol = Sheets("a").Range(mrange).End(xlToLeft).Column - 1

For jk = 2 To y_count + 1
    For j = 2 To mcol + 1
      mrange1 = Sheets("a").Cells(jk, 1)
      mrange2 = Sheets("a").Cells(1, j)
      Sheets("a").Cells(jk, j) = 1 / (1 + Exp(mrange1 - mrange2))
      Randomize
      mrnd = Rnd()
      Sheets("b").Cells(1, 1) = mrnd
     If Sheets("a").Cells(jk, j) >= Sheets("b").Cells(1, 1) Then
         Sheets("a").Cells(jk, j) = 0
     Else
         : Sheets("a").Cells(jk, j) = 1
     End If
    Next j
 Next jk
```

The example of simulated data were generated as shown in Figure 1.

	A	B	C	D	E	F	G	H
1	IQ	3	2	1	0	-1	-2	-3
2	-0.08597	1	1	1	0	0	0	0
3	0.369257	0	1	1	1	0	0	0
4	0.578473	1	1	0	0	0	[illegible]	0
5	1.125195	1	1	1	0	0		0
6	-0.90035	1	1	1	0	1	CommandButton1	0
7	0.710145	1	1	0	1	0		0
8	-0.00356	1	1	1	0	0		0
9	-0.54471	1	1	1	1	1	0	0
10	-1.03919	1	1	1	1	1	0	0

Figure 1 The dichotomous responses

(2) Polytomous response

The codes are listed below:

```
Sheets("c").Range("B2:K100").ClearContents
  y_count = Sheets("a").Range("A225536").End(xlUp).Row - 1
  mrange = "wfd" & 1
  mcol = Sheets("a").Range(mrange).End(xlToLeft).Column - 1
  stept0 = Sheets("b").Cells(4, 2)
  stept1 = Sheets("b").Cells(5, 2)
  stept2 = Sheets("b").Cells(6, 2)
  stept3 = Sheets("b").Cells(7, 2)

 For jk = 2 To y_count + 1
 For j = 2 To mcol + 1
    ability = Sheets("a").Cells(jk, 1)
    item_diff = Sheets("a").Cells(1, j)
   Randomize
     mrnd = Rnd()
     Sheets("b").Cells(1, 1) = mrnd
        a0 = 1
        a1 = Exp(ability - (step0 + item_diff))
        a2 = Exp(2 * ability - (step0 + item_diff) - (step2 + item_diff))
        a3 = Exp(3 * ability - (step0 + item_diff) - (step2 + item_diff) - (step3 + item_diff))
        a4 = Exp(4 * ability - (step0 + item_diff) - (step2 + item_diff) - (step3 + item_diff) - (step4 + item_diff))

         all_p = a0 + a1 + a2 + a3 + a4
         p_a0 = a0 / all_p
         p_a1 = a1 / all_p
         p_a2 = a2 / all_p
         p_a3 = a3 / all_p
         p_a4 = a4 / all_p
         all_p1 = p_a0 + p_a1 + p_a2 + p_a3 + p_a4

         If mrnd <= p_a0 Then
           Sheets("c").Cells(jk, j) = 0
         ElseIf mrnd <= p_a1 + p_a0 Then
```

```
                Sheets("c").Cells(jk, j) = 1
            ElseIf mrnd <= p_a1 + p_a0 + p_a2 Then
                Sheets("c").Cells(jk, j) = 2
            ElseIf mrnd <= p_a1 + p_a0 + p_a2 + p_a3 Then
               Sheets("c").Cells(jk, j) = 3
            Else
               Sheets("c").Cells(jk, j) = 4
            End If
        Next j
    Next jk
```

The example of simulated data were generated as shown in Figure 2.

	A	B	C	D	E	F	G	H
1	IQ							
2	-0.08597	1	1	1	0	0	0	0
3	0.369257	0	1	1	1	0	0	0
4	0.578473	1	1	0	0	0	0	0
5	1.125195	1	1	1	0	0	0	0
6	-0.90035	1	1	1	0	1	0	0
7	0.710145	1	1	0	1	0	0	0

Figure 2 The polyttomous responses

(3) Validation of data fitting to Rasch model's requirement

The simulated responses are put into the website[5] for verifying the validation fitting to the requrement of Rasch model. The option of category is selected at the below combo box. The transformed data is redirected to category. If the option of forest plot is selected, the result appear with a forest plot for the comparison of person measures. The premise is to arrange responses separated by comma. The mean square errors(MSQNs) of Infit and Outfit were provide to examine abnormal responses in items and persons, respectively, by observing them less than 2.0 for persons and 1.5 for items[6].

Discussions and conclusion

We demonstrated the procedure of Rasch simulation data with code of VBA in MS Excel. All responses in either scenario of dichotomy or polytomy have been generated. Whether data fitting to the expectation of Rasch model can be verified via an author-made module[5]. The technique of generating Rasch simulation data can be applied to numerous research in the future. .

References

1. Chien TW, Lin WS. Simulation study of activities of daily living functions using online computerized adaptive testing. BMC Med Inform Decis Mak. 2016 Oct 10;16(1):130. doi: 10.1186/s12911-016-0370-8. PMID: 27724939; PMCID: PMC5057399.
2. Rasch G. Probabilistic Models for Some Intelligence and Attainment Tests. Chicago, Illinois: University of Chicago Press; 1980
3. Linacre JM. How to Simulate Rasch Data. Rasch Measurement Transactions 2007;21(3): 1125
4. Andrich D. A rating formulation for ordered response categories. Psychometrika 1978; 43: 561–73.
5. Chien TW. Rasch model module online. 2021/8/25 at http://www.healthup.org.tw/kpiall/raschcontinous_rsm.asp
6. Linacre JM. Optimizing rating scale category effectiveness. J Appl Meas. 2002;3(1):85–106.