

Technical Supplement for “A Comparison of Methods for Binning Responses to Open-Ended Survey Items About Everyday Events: a New Tailored-Binning Approach”

Derrick M. Angier and John D. Mayer

da1067@wildcats.unh.edu jack.mayer@unh.edu

Personality Lab, Department of Psychology

15 Academic Way, University of New Hampshire, Durham, NH, United States.

Table of Contents

Appendix A: Tailored Binning Rules	3
Binning distributions mentioned in the Introduction section of the paper.	10
Figure 1.	10
Figure 2.	10
Example of possible response alternatives based on binning results.	11
Table 1	11
References	12

Appendix A: Tailored Binning Rules

Follow each Step, progressing downward, and if called for by the Step, jump straight to a specific Part.

General Instructions	
<ul style="list-style-type: none"> Following this system will improve many but not all items. It is possible that a few items will not be optimally binned at the end; that is normal. Researchers should recognize that the content and context of an item may influence binning choices. In such instances, the researcher's discretion should take precedence over these rules. For example, the 'zero' category may be of particular significance to some survey items (i.e., number of times tried heroin), as might be '7' (i.e., brushed teeth every day of the week). In the heroine example, for instance, researchers might wish to treat the answer 'zero' differently from the remainder—and refrain, for example, from merging it into a “1-3” category that might be next upward. This ruleset serves as a tool to assist with binning items with open-ended responses into a set of categories useful for analysis and future survey administration. These rules may be especially useful with skewed distributions (e.g., as often found with biographical data). These rules assume and describe numerical responses, but you could theoretically apply this process to open-ended categorical responses (i.e., Dell, Lenovo, Apple, etc.) with the understanding that the researcher would have to interpret these guidelines in light of that situation 	
Preliminary Preparation	
<ul style="list-style-type: none"> Produce frequency tables and/or histograms for the responses to the item. Ideally you want to observe: the item's range of response values, the frequency of responses, the percent of responses per value, and the cumulative percent representing each value. 	

PART A: Initial Item Modifications		
Step	Rules	Examples
A1	<ul style="list-style-type: none"> To begin, we will use each of the different response values made by participants to the item to define the starting bins for the item. 	<ul style="list-style-type: none"> If the responses to an item are as shown in Table A, then there are 7 bins total to start.
A2A	<ul style="list-style-type: none"> If there is a logical “0” or a floor value that is a possible response but received no responses, add a new bin to account for that value. Do the same in the case of a logical ceiling value (for example, how many days a week did you exercise more than 30 min.?) if the maximum was not included as a response. 	<ul style="list-style-type: none"> See Table A under “Revised Responses”.
A2B	<ul style="list-style-type: none"> If there are no absolute ends to the distribution, add “or less” to the minimum value, and “or more” to the maximum value. 	<ul style="list-style-type: none"> See Table A under “Updated Bins with Modified Ends”.
A3	<ul style="list-style-type: none"> If there are gaps between possible values represented by the bins, adjust the bin intervals to accommodate values that may appear in subsequent data collection. When adjusting bins for gaps, take the lower value bin and bin “up” to the value prior to the next bin's lower value to establish the range of responses within the bin. (see Table A) 	<ul style="list-style-type: none"> See Table A under “Updated Bins with Revised Intervals”.
A4	<ul style="list-style-type: none"> If the number of bins are: <ul style="list-style-type: none"> Fewer than 4, go to Part B. Between 4 and 8, binning is complete. More than 8, go to Part C. 	<ul style="list-style-type: none"> Count the number of bins at this point. NOTE: From this point and forward, count the number of current bins after each step. If an item achieves 4-8 bins, binning is complete. If fewer than 4 or more than 8 bins remain, continue following the steps.
A5	<ul style="list-style-type: none"> Visually inspect the histogram displaying the distribution of responses for this item. If there is minimal skewness and no side of the distribution tapers to a point relative to the other side, proceed to Part C. If the researcher can observe a tapered tail of the distribution (pointed end) and aims to manage skewness, proceed to Part S. 	<ul style="list-style-type: none"> Highly skewed response distributions are likely to show a tendency for fewer responses out to one end, creating a tapering tail. From here and on, this tapered tail (or in some cases, tails) will be referred to as the “pointed end” for simplicity.

Table A				
<i>Part A Steps: Binning Example, Initial Item Modification</i>				
Example: Absolute minimum value “zero”, relative maximum value				
BEFORE		Step A2A	Step A2B	Step A3 AFTER
Participant Responses	Bin Number	Revised Responses	Updated Bins with Modified Ends	Updated Bins with Revised Intervals
		0	0	0
1	1	1	1	1
2	2	2	2	2
3	3	3	3	3
4	4	4	4	4
5	5	5	5	5 to 9
10	6	10	10	10 to 14
15	7	15	15 or more	15 or more
<i>7 totals bins</i>		<i>8 totals bins</i>		<i>8 totals bins</i>

PART B: Fewer than Four Categories		
Step	Rules	Examples
B1	<ul style="list-style-type: none"> If you decide to attempt to meet the four to eight bin goal for your items, proceed to Step B2. If you decide that fewer than four bins is acceptable, than keep the number of starting categories the same; binning is complete. 	
B2	<ul style="list-style-type: none"> If the distribution of responses involves values that are whole numbers and adjacent then add bin(s) at the “pointed end” of the distribution. This additional bin, or bins, should be the next corresponding value(s) that was not represented by a response and should include “or more” or “or less” if it is at either pointed end of response values. 	<ul style="list-style-type: none"> See Table B, under Example 1. Highly skewed response distributions are likely to show a tendency for fewer responses out to one end, creating a tapering tail. From here and on, this tapered tail (or in some cases, tails) will be referred to as the “pointed end” for simplicity.
B3	<ul style="list-style-type: none"> If fewer than four bins remain, despite following the rules so far, the researcher may use their own discretion about how to add more bins and/or they may accept fewer than four bins for that item. Binning is complete. 	<ul style="list-style-type: none"> See Table B, under Example 2.

Table B					
<i>Part B: Binning Example, Fewer than Four Categories</i>					
Example 1: Adjacent values			Example 2: Non-adjacent values		
BEFORE		Step B2 AFTER	BEFORE		Step B3 AFTER
Participant Responses	Bin Number	Revised Responses	Participant Responses	Bin Number	Revised Responses
0	1	0	1	1	1 to 4
1	2	1	5	2	5 to 9
2	3	2	10	3	10 or more
		3 or more			10 or more
<i>3 totals bins</i>		<i>4 totals bins</i>	<i>3 totals bins</i>		<i>4 totals bins</i>

PART C: More than Eight Categories, Reducing to a More Manageable Number		
Step	Rules	Examples
C0	<ul style="list-style-type: none"> Visually and statistically inspect the histogram displaying the distribution of responses for this item. If there is minimal skewness and no side of the distribution tapers to a point relative to the other side, continue to Step C1. If the researcher can observe a tapered tail of the distribution (pointed end) and aims to manage skewness, proceed to Part S. 	<ul style="list-style-type: none"> Highly skewed response distributions are likely to show a tendency for fewer responses out to one end, creating a tapering tail. From here and on, this tapered tail (or in some cases, tails) will be referred to as the “pointed end” for simplicity.
C1	<ul style="list-style-type: none"> In a case where the binning of one or more non-whole numbers (i.e., .25, 1.5, etc.) with low response rates (under 1%) would bring the number of bins to 8 or under, bin that non-whole number value with the next highest value. Complete this step when applicable. 	<ul style="list-style-type: none"> See Table C.
C2	<ul style="list-style-type: none"> This Step is done starting with the bin at the end of the pointed tail (if applicable) or higher response value tail, and may proceed through all other bins. Continue the process of merging categories, now with the focus on the bins with the lowest frequencies of responses. Start by binning extremely low frequencies to the adjacent bin, typically next toward the center of the distribution of bins. The researcher should use discretion in combining bins in seeking to maintain qualitative meaning. 	<ul style="list-style-type: none"> See Table D, Example 1. It is up to the researcher’s discretion regarding combining bins that contain values of particular informative value (i.e., 0, 1, 2, etc.). If qualitatively useful, these bins may be kept as is, even if there were relatively few responses given to these categories.
C3	<ul style="list-style-type: none"> If you complete one pass of binning adjacent categories following the guidelines of C2 and more than 8 bins remain, complete another pass over the remaining new arrangement of bins. Combining bins with the lowest adjusted response frequencies should be prioritized. Repeat this step as needed until 8 or fewer bins remain. When 8 or fewer bins remain, binning is complete. 	<ul style="list-style-type: none"> See Table D, Example 2.

Table C		
<i>Part C1: More than Eight Categories & Non-Whole Numbers</i>		
Example: Ten response values		
BEFORE		Step C1 AFTER
Participant Responses	Bin Number	Revised Responses
0	1	0
.5	2	
1	3	.5 to 1
1.5	4	
2	5	1.5 to 2
3	6	3
4	7	4
5	8	5
6	9	6
7	10	7 or more
<i>10 totals bins</i>		<i>8 totals bins</i>

Table D			
<i>Part C2-C3: Binning Example, Binning Based on Response Frequencies</i>			
Example 1: Combining bins: First pass		Example 2: Combining bins: Follow-up passes	
BEFORE		Step D1 1st Pass Complete	
Revised Bins as established prior to Part D	Percentage of Responses	Revised Responses 1 st pass	Percentage of Responses
0	30%	0	30%
1	18%	1	18%
2	12%	2	12%
3	8%	3	8%
4	7%	4	7%
5 to 6	2.5%	5 to 6	2.5%
7	1%	7 to 9	3%
8 to 9	2%		
10 to 11	10%	10 to 11	10%
12 to 19	6%	12 to 19	6%
20 or more	3.5%	20 or more	3.5%
<i>11 totals bins</i>		<i>10 totals bins</i>	
		Step D2 2nd Pass Complete	
Responses Further Revised from Example 1	Percentage of Responses	Revised Responses 3 rd pass complete	Percentage of Responses
0	30%	0	30%
1	18%	1	18%
2	12%	2	12%
3	8%	3	8%
4	7%	4	7%
5 to 9	5.5%	5 to 9	5.5%
10 to 11	10%	10 to 11	10%
12 to 19	6%	12 or more	9.5%
20 or more	3.5%		
<i>9 totals bins</i>		<i>8 totals bins</i>	

PART S: Skewness Management		
Step	Rules	Examples
S1	<ul style="list-style-type: none"> When applicable, take the approximate 3%-5% of cumulative percent at the pointed end tail and, if those responses span multiple bins, bin the corresponding values together into a single “tail” bin. Reminder: the pointed end is the end that has lower response frequencies and tends to taper to a point. 	<ul style="list-style-type: none"> See Table S. Look at the value that denotes that it, and the values toward the pointed end of the tail, received 3% or more of the cumulative percent of responses at the pointed end, and then bin that value with all others toward the tail. Example 1 for Positive Skew, Example 2 for Negative Skew.
S2	<ul style="list-style-type: none"> <u>Starting with the bin immediately adjacent</u> to the bin at the end of the pointed tail, merge adjacent categories in pairs, proceeding toward the opposite end of the distribution. Refrain from merging, or end the binning process altogether, if one or both of the to-be-merged bins already have high response frequency. Categories uniquely higher in response rate than their adjacent categories may be skipped over to prevent stacking already relatively high response categories. 	<ul style="list-style-type: none"> See Table S, Example 3. E.g., bins of 7% & 9% of responses, compared with bins of 1.2% & .8%, may be skipped over at this Step.
S3	<ul style="list-style-type: none"> If the number of bins are: <ul style="list-style-type: none"> Between 4 and 8, binning is complete. More than 8, proceed to Part C, Step C1. 	

Table S								
<i>Part S: Binning Example, Skewness Management</i>								
Example 1: Capping the pointed end: Positively skewed			Example 2: Capping the pointed end: Negatively skewed			Example 3: Binning adjacent values from the pointed end		
BEFORE		Step S1 AFTER	BEFORE		Step S1 AFTER	BEFORE		Step S2 AFTER
Participant Responses	Cumulative Percentage of Responses	Revised Responses	Participant Responses	Cumulative Percentage of Responses	Revised Responses	Revised Responses (after Step S1)	Cumulative Percentage of Responses	Alternate Revised Responses
0	30%	0	1	1.5%		0	30%	0
1	45%	1	3	2.7%		1	45%	1
2	55%	2	5	5%	1 to 5	2	55%	2
3	65%	3	8	10%	8 to 9	3	65%	3
4	70%	4	10	15%	10 to 12	4	70%	4
5	75%	5	13	25%	13 to 14	5	75%	5 to 9
6	85%	6 to 9	15	30%	15	6 to 9	85%	
10	90%	10 to 11	16	35%	16	10 to 11	90%	10 to 14
12	95%	12 to 14	17	45%	17	12 to 14	95%	
15	97.3%	15 or more	18	55%	18	15 or more	97.3%	15 or more
18	98.5%		19	70%	19			
20	100 %		20	100%	20			
<i>12 totals bins</i>		<i>10 totals bins</i>	<i>12 totals bins</i>		<i>10 totals bins</i>	<i>10 totals bins</i>		<i>8 totals bins</i>

Derrick M. Angier & John D. Mayer

University of New Hampshire

Binning distributions mentioned in the Introduction section of the paper.

Figure 1.

23	Age of Householder	
24	Under 65 years	
25	15 to 24 years	
26	25 to 34 years	
27	35 to 44 years	
28	45 to 54 years	
29	55 to 64 years	
30	65 years and older	

Table A-1. Income summary measures by selected characteristics: 2017 and 2018.

(Semega et al., 2020)

Figure 2.

2020 federal income tax brackets

2020 tax brackets (for taxes due A

TAX RATE	SINGLE	HEAD OF HOUSEHOLD
10%	\$0 to \$9,875	\$0 to \$14,100
12%	\$9,876 to \$40,125	\$14,101 to \$53,700
22%	\$40,126 to \$85,525	\$53,701 to \$85,500
24%	\$85,526 to \$163,300	\$85,501 to \$163,300
32%	\$163,301 to \$207,350	\$163,301 to \$207,350
35%	\$207,351 to \$518,400	\$207,351 to \$518,400
37%	\$518,401 or more	\$518,401 or more

(Current Population Survey: 2019 ASEC Technical Documentation, 2020)

Example of possible response alternatives based on binning results.

Table 1

Example of Closed-Ended Response Alternatives Developed from the Three Binning Methods for Item 7.

Over the past week, how many hours a day did you spend playing FPS (First-Person Shooter) games?								
Tailored	Zero hours in the last week.	1 hour	2 hours	3 hours	4 hours	5 hours	6 to 9 hours	10 or more hours
Over the past week, how many hours a day did you spend playing FPS (First-Person Shooter) games?								
Equal Percentage	Zero hours in the last week.	1 hour	2 hours	3 hours	4 or more hours			
Over the past week, how many hours a day did you spend playing FPS (First-Person Shooter) games?								
Equal Interval	Zero to 2 hours in the last week.	3 to 5 hours	6 to 7 hours	8 to 10 hours	11 to 13 hours	14 to 15 hours	16 to 18 hours	19 or more hours

References

Current Population Survey: 2019 ASEC Technical Documentation. (2020). Washington: U.S.

Census Bureau. <https://www2.census.gov/programs-surveys/cps/techdocs/cpsmar19.pdf>

Semega, J., Kollar, M., Creamer, J., & Mohanty, A. (2020). *Income and Poverty in the United States: 2018*.

<https://www.census.gov/content/dam/Census/library/publications/2019/demo/p60-266.pdf>