Normalizing Survey Strangeness

Considerations for Normalization in Correction of Survey Biases

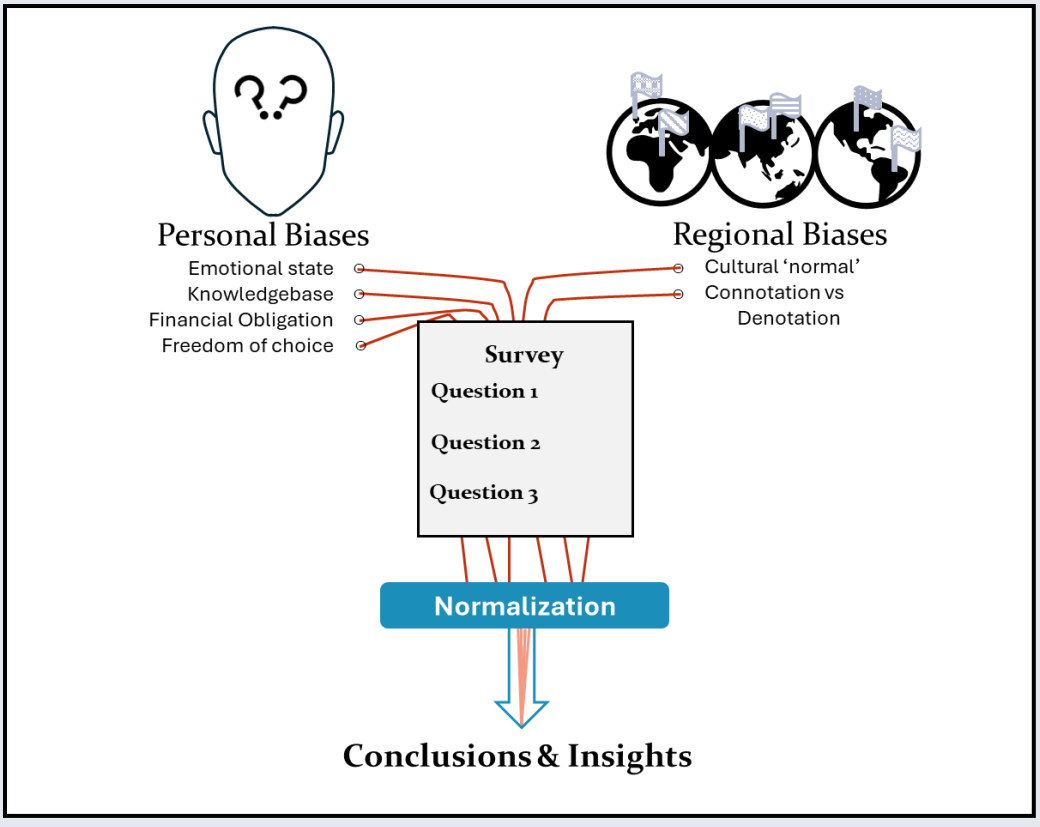
Josef Grey

Normalizing Survey Strangeness

Considerations for Normalization in Correction of Survey Biases

Abstract

Surveys measure respondent's perceptions, affections, emotions, and biases. These biases can be personally or regionally generalized, and alter answers in unfavorable ways. Normalization, a term referring to any method of transforming the data by changing the scale, can present a solution to complex survey datasets that yield few clear insights. By stretching or altering a scale, such as a numerical satisfaction score 1-5, differences or similarities can be enlarged for easier identification – or for misidentification. Normalization should be clearly labeled and defined for observer trust.



Relevancy

Market Research, Marketing, Sales, Product Development, Medical Affairs, Consumer Insights

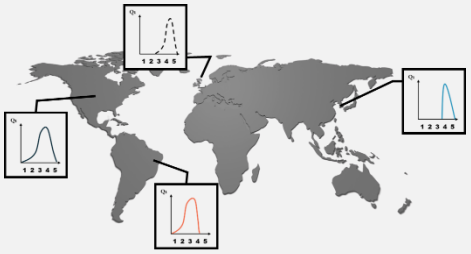
Grey THEOREM

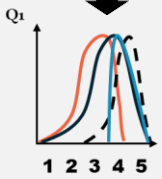
Translating data to insights is challenging. Translating hundreds of responses, across dozens of questions, various competitive brands, and respondents across the world to insights, is downright strange. To combat strange, normalization is nearly any method of transforming the data's scale, such as a satisfaction score 1-5, into something more *logical*. Normalization is useful for insight generation, partly by correcting for personal and regional biases. Survey bias is a long-studied concept impacting data analysis, conclusions, and general survey effectiveness. Condensing these to the personal and regional level offers a simplified framework in survey design and data analysis for presentation. Furthermore, normalization can be applied across multiple axes or questions to correct for multiple biases, or enable population characterization and classifications.

Normalization is a common statistical and data science method, referring to any transformation of an axis of data to a theoretical state of "normal", i.e. making a satisfaction scale log-based, or transforming a competitor's numerical scores to a relative scale. The obvious joke, "normal is in the eye of the beholder," leads to the Golden Rule of data science/mathematics/leadership/life – be honest and define everything. Setting the scale's minimum or maximum is the primary challenge, as stretching any axis can distort the picture. The simplest, and most practical form of normalization is referential normalization using the high-score as the denominator, thus making everyone a percent of the high-score. A common rule of thumb is to look at the data through the question, never the conclusion. Any chosen method of normalization should be defined alongside the conclusion and narrative evidence.

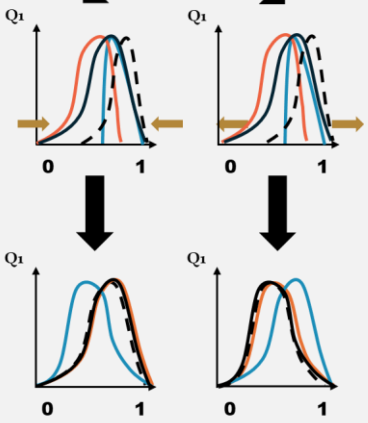
Answer the question; it is easy to alter a dataset into a desired picture, or to lose sight of the question.

Consider a survey with respondents from around the world. Answers for a region may group together, yet distinctly from other regions.





Comparative Analysis of the raw dataset may provide several different answers to a question where only one can be chosen, e.g. new product or treatment indication.

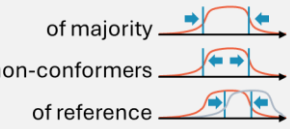


Alternative Normalization methods, such as pushing the data together, or pulling the data apart, on a new scale can be used to get alternative results.

Set the method to the Question. While the pressure to identify "key" differences is omnipresent, key differences aren't.

If lost, answer the question.

What is...



Label & Define Methods

Clearly label modified numbers with how. Transparency breeds trust, but the observer deserves to decide.

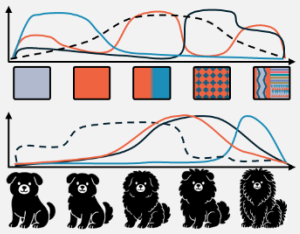
A	Absolute data values
B	Basic data manipulation *Definition – "Normalized to Maximum"*
C	Complex data manipulation *Definition – "Normalized by grouping (region, customer type, etc.)"*
P	Proprietary data manipulation *Definition – "Nonlinear referential normalization with grouping by trait xyz"*

While simplicity might be the most desirable, complex questions and answers can have more impactful results. Analysis isolated to a common group of respondents, from a similar region with similar backgrounds may not need normalization; complexity is driven by the question length **and** quantity of differences in answers, or variance. Simply, as the question gets stranger, normalization gets more complex. Fluff is as arbitrary a scale as satisfaction, as each data question relies on respondent's translation of survey question, the answer gets stranger.

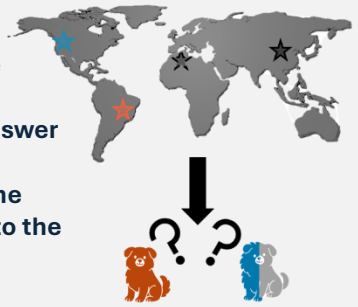
Skew of scales leads to the first bias, cultural 'normal'. Cultural stereotypes offer the blandest example, the angry Northerner vs the friendly Southerner. Culinary spiciness, grooming behaviors, even acceptance of corruption fall on regional scales of 'normal'. Specific regions, especially in an international survey, may have greatly altered internal scales, and interact with question options in strange ways. Normalizing regional scales, either stretching, cutting, or transforming them, allows for greater level of detail in comparing different regions or groups.

At time of writing, Merriam-Webster lists 160 synonyms for strange (surprisingly, skibidi wasn't one). Simple is strenuous, especially when questioning complex internal perceptions or medical conditions, *especially* when doing it across the globe. "Hang low" means different yet the same thing in different parts of the US, and means nothing much outside of it. Regional connotation and denotation are silent killers of survey impact, with misinterpreted words leading to incorrect conclusions or post facto removal of population. Normalizing with a secondary or paired question, i.e. familiarity or personal measurement of a common reference, or stretching the differences between viable responses can yield constrained insights.

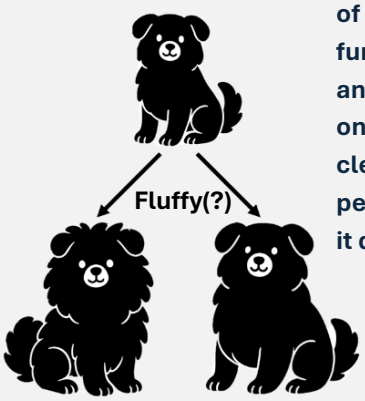
Normalization is a straight forward tool, but the ability to make strange look normal means any question can have a normal-looking answer without it being a right, or logical, answer.



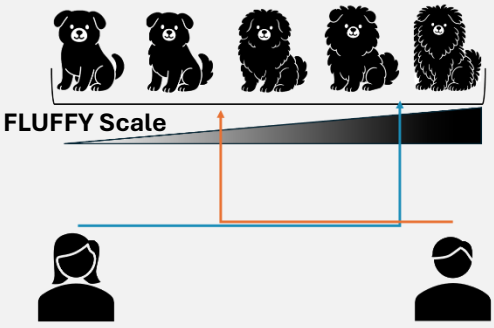
Focus on 3 key questions:
(1) Does the question make sense?
(2) Does the answer make sense?
(3) How does the answer relate to the question?



Words can carry different meanings, both to the individual and the region. Semantics is the code of surveys as any function is to data analysis. Even where one word may seem clear, a region or person may interpret it different.



Where possible, be absolute. Lacking that, use a reference. Simple scales can increase respondent (and stakeholder) accuracy and precision.

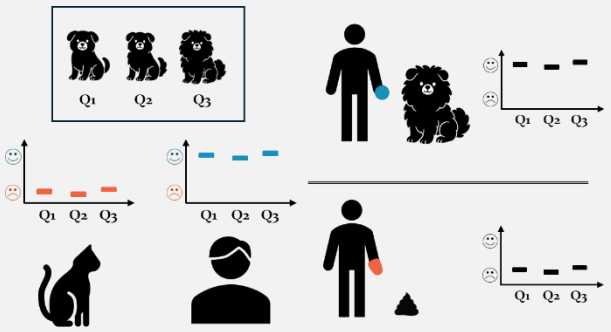


Regional interpretation of words is one hindrance, respondent-specific interpretation is another, or **personal emotions**. Respondent emotional attachment to the survey theme or question element can limit conclusions in (un)predictable ways, even the respondent's base emotional state, or how they are feeling while answering the survey, alters answer choice. Product or treatment satisfaction can change based on the emotional state of the last time getting a treatment or using the item, or if the respondent received traumatic news a week before, or if they just won compensation for answering a brief survey. Inferring between a generally happy or sad person from one who had a bad day can be difficult, referential normalization can be a superior method in either case. By using referential questioning, either with multiple brands or conditions or treatment variants, the base scale of each respondent can be stretched to identify the differences between choices.

Respondents are traditionally chosen by a select set of demographic criteria, then filtered with interest, familiarity, and population-fit questions before full recruitment. Ultimately, the opinion of the customer is all that matters, but familiarity is a spectrum **and** filter for question depth. Emotional attachment, enjoyment, satisfaction, and even quantity or history of use lack inherent qualitative understanding – a machine may do the same thing a million times and never know why. Understanding is the most difficult metric to quantify, but even basic mechanism of action or instructional questions can measure respondents fit for inclusion on higher level themes.

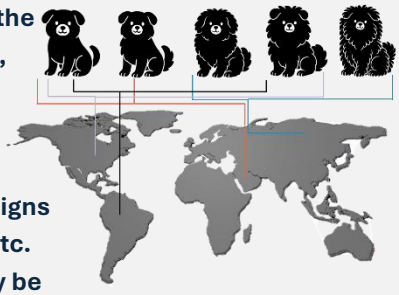
The second to last bias is also the principal reason for respondent participation, financial reward. While necessary, reward alters population recruitment and responses. Internal pressure to answer favorably

Personal emotions continuously slide on personal scales, and ungrounded responses can be based on latest vs average interaction.



Meanwhile, emotional affinity can be ignored or viewed negatively instead of as an axis for data analysis. The world is full of different people, maximizing insights from surveys can be as simple as paying attention to each group.

Familiarity with subject matter, often unmeasured, always impacts. Whether it's understanding the absolute scope, mechanism of action, disease physiology, recognition of signs or symptoms, etc. While each may be a valid consumer or market participant, alternative qualifications may be needed for specific purposes.

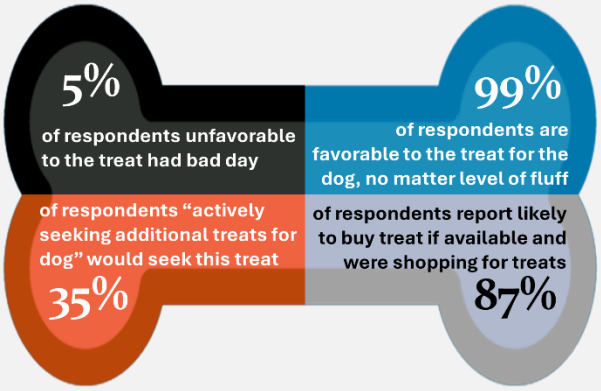


Surveying because of the bone or surveying for the bone, similar but different. Remove reward-weight early.

or knowledgably towards the subject matter at hand for potential reward, reward maximalization (i.e. belief in "correct" responses for follow-up surveys or a greater reward), or affection from compensation will alter responses (un)favorably and deceive stakeholders. Measuring respondent's variance in answers, and /or including an intrinsic control for answer skew, can alleviate this concern. Filtering out respondents for specific data questions, or respondent-internal normalization can help to find differences and prevent false-positive responses. Any method should be defined with the conclusion, preferably with an unmodified score in accompaniment for observers to digest the conclusion appropriately.

The final bias is the most counter-intuitive, and that is choice constraint. Questions seeking to limit answer complexity by limiting answer options or real variation in answer, force respondents between a rock and a hard place. Answer choices may even underlie the worst bias of all, that of seeking a specific conclusion. Common industry practice of adding the question to each conclusion or datapoint should be amended by including the answer choices. To avoid the pitfall of framing, there must be at least one reasonable alternative answer to every question. What is viewed as "noisy" variation can be easily reframed into clustered answers for respondent or insight classification, normalizing the answer internally, or externally with other questions or demographics. Along with the freedom of choice is confidence to provide a different answer. Either internal pressure to align to a cultural norm, or belief that an altered answer will result in loss of compensation – the respondent must feel safe in answering according to their own beliefs and perceptions, not the surveyors or societies.

Learnt responses to reward, or simply not wanting to say no can look like a yes. Carefully measuring the question, and respondent, can provide a clearer answer. Normalization is that tool for the analysis stage.



Freedom of Choice vs Self-justification

99% of Dogs are The Best Boy – All in on Marketing*

*Q17: If you have a dog, are they:
(A1) The Best Boy (A2) The Worst Dog Ever

Providing real answer variation, or adding secondary open-ended questions for confirmation of key principles can be the final driver for organizational action.

Normalization is an efficacious method for data analysis, especially in correcting for survey biases, whether regionally or respondent based. The person's mood, affiliation, familiarity, and cultural environs can all be corrected by pulling and transforming each scale, giving directional insights as long as the transformation is clearly defined. Data analysis is one place to correct for survey biases, but the design phase offers a chance to avoid more complex or widespread mathematics.

- **Reset the respondent with the Introduction**: Utilize the introduction to (re)set the immediate emotional state or tone of the respondent. Providing a sincere note of appreciation for their answers and time can provide a short-term boost in attitude.
- **Define the recruitment, survey, and termination process**: Clearly explain why the respondent was recruited for the survey, the purpose of the survey, how long or how many questions, and then how it will end. Avoid re-recruitment clauses unless there is a clear desire/need in the design phase, and transparently tell the participant answer choice is not saved to their profile and will not affect their recruitment for any other survey.
- **Provide *all* reasonable answers to a question**: Each question must be specifically worded and framed to ensure that there is a set number of possible answers and that these are included in a multiple choice, otherwise open ended or short response are a viable question format with scaling solutions.

The level of strategy put into the design phase is returned in the analysis phase, and every ounce of energy put into the analysis phase has to be put in a fifteen-minute PowerPoint presentation. Consider normalization for strange survey results to provide actionable insights.

Please enjoy these pictures of puppies before answering the following questions:

Something as simple as pictures of puppies can alter someone's emotions, perceptions, and willingness to do something.

Normalizing Survey Strangeness

Contact Advisory@Greytheorem.com

See more at Greytheorem.com