The Most Common Analytical and Statistical Mistakes
Posted by Sunil Kappal on January 3, 2017 at 7:00am

It is not only about understanding about statistics, it is also about implementing the correct statistical approach or method. In this brief article I will showcase some common statistical blunders that we generally make and how to avoid them.

To make this information simple and consumable I have divided these errors into two parts:

1. Data Visualization Errors (Erroneous Graphs)
2. Statistical Blunders Galore (pun intended)

Data Visualization Errors (Erroneous Graphs): This is one area that can give a nightmare to both the parties the presenter as well as the audience. Incorrect data presentation can skew the inference and can leave the interpretation at the mercy of the audience.

Pie Charts: “Get back to the kitchen and make me some good Pie”

86.4% of people will believe any data you put in a PowerPoint slide, even if you just totally made it up to prove your point.

Pie charts are considered to be the best graph when you want to show how the categorical values are broken. However, they can be seriously deceptive or misleading. Below are some quick points to remember when looking at the Pie Charts:

- Percentages should add up to 100%
- Prefer 3D in VR consoles not in the pie charts
- Thou shall not have Other: Beware of the slices with “others”. If that is larger than the rest of the slices “You have a problem” it makes the pie chart vague
- Show the total number of reported categories to determine “how big is the pie”

Bar Graphs: “Let’s make a Bar Graph not a Bawaaaah Graph”

It’s a great graph to show the categorical data by the number or percent for a particular group. Points to consider when examining a Bar Graph:
- Thou shall have right scale: Scale made very small to make the graph look big or severe
- Consider the units being represented by the height of the bar and what it means as a result in terms of those units

**Time Charts:** *“What time is it?”*
A time charts is used to show how the measurable quantities change by time.
- Thou shall have the right scale and the axis: It is a good practice to check the scale on the vertical axis (usually the quantity) as well as the horizontal axis (timeline) as the results can be made to look very impactful by switching the scales
- Don’t try to answer the “Why it is happening question” using the time charts as they only show “What is happening”
- Ensure that your time charts should show empty spaces for the times when no data was recorded

**Histograms:** *The Binning Man*
It is a good practice to check the scale used for the vertical axis frequency (relative or otherwise), especially when the results are showed down through the use of inappropriate scale
- Ensure that the intervals are not missed on the x or y axis to make the data look smaller
- Ensure the application of histogram is correct as people tend to confuse histograms with a bar graphs

**Statistical Blunders Galore:** This is probably a “no non-sense zone” where one would not want to make false assumptions or erroneous selections. Statistical errors can be a costly affair, if not checked or looked into it carefully.

**Biased Data:**

Bias in statistics can be termed as over or underestimating the true value. Below are some most common sources or reasons for such errors.
• Measurement instruments that are systematically off and causing such bias. Example a scale that adds up 5 pounds each time you weigh.
• Survey participants influenced by the questioning techniques
• A Population sample of individuals who doesn’t represent the population of interest. For example, examining the workout habits by only visiting people in the gymnasiums will introduce a bias.

No Margin of Error: “No there isn’t any margin of error on spelling tests, it is not mathematics” This is a great way to understand the potential miscalculation or change in circumstance that can result in a sampling error and ensures that the result from a sample study is close to the number that can be expected from the entire population. It is a good idea to always look for this statistics to ensure that the audiences are not left to wonder about the accuracy of the study.

Non Random Sample: Nonrandom samples are biased, and their data cannot be used to represent any other population beyond themselves. It is pivotal to ensure that any study is based on the random sample and if it isn’t you are about to get into a big trouble. “Go and hide somewhere!”

Correlation is not Causation:
Besides the above statement correlation is one statistic that has been misused more than being used. Below are the few reasons that makes me believe the misuse part of this statistic.

Correlation applies only to two numerical variables, such as weight and height, call duration and hold time, test scores for a subject and time spent studying that subject etc. So, if you hear someone say, “It appears that the study pattern is correlation with gender,” you know that’s statistically incorrect. Study pattern and gender might have some level of association but they cannot be correlated in the statistical sense.

Correlation helps to measure the strength and the direction of a linear relationship. If the correlation is weak, once can say that there is no linear relationship but that doesn’t mean that there is no other type of relationship that might exist.

Botched Numbers: One should not believe in everything that appears with statistics. As we know error appears all the time (either by design or by error), so look for the below points to ensure that there are no botched numbers.

• Make sure everything adds upto what it is reported to
• “Stich in time saves the nine”, do not hesitate to double-check the numbers and basic of calculations
• Look at the response rates of a survey – Number of people responded divided by the number of people surveyed
• Question the statistic type used to ensure it is the best fit

Being a consumer of the information it is your job to identify shortcomings within the data and analysis presented to avoid that “oops” moment. Statistics are nothing but simple calculations that are smartly used by people who are either ignorant or don’t want you to catch them to make their story interesting. So to be a certified skeptic wear your statistic glasses.