

Nine Mistakes in Quantitative Data Analysis

Dr. Hazhar Blbas

Statistics Department - MSc at UCF in 2014 Statistics Department - PhD at SUE in 2022

Error 1: Inadequate Configuration

Hypothesis: what is the relationship between two variables, For example, relationship between age and systolic blood pressure

Wrong analysis: using Descriptive Statistics

Correct Analysis: Using Inferential Statistics

Error 2: Not having enough Sample Size

Sample size should be representative and unbiased of the population

A sample sample that's too small may be **unrepresentative** of the population, while a sample that's too large will be more **costly** than necessary.

https://www.calculator.net/sample-size-calculator.html?type=1&cl=95&ci=1&pp=50&ps=600&x=82&y=25

Cochran formula

 $Z^2p(1-p)$

 n_0

- Z value of 95% Confidence is 1.96
- ▶ P estimated proportion of the population for e.g. p=0.5
- ▶ q = 1 P = 0.5
- e is error Margin = 0.05 (5%)
- \blacktriangleright N is population size = 1000
- Sample Size (n) = 278

Error 3: Probability Sampling or Nonprobability Sampling

A. Probability (Random) Sampling



B. Nonprobability Sampling



Error 4: Using a statistical test that is inappropriate for the type of data





Error 5: Avoid using complicated statistical analysis

- Great thesis is not depending on that you have used a complex statistical analysis but it depend on that you have been using the correct analysis that give you the best conclusion depending on your data.
- You don't need to show up how much you are smart and did difficult analysis but important you have to have a correct analysis and result even you have use a simple analysis

Some Analysis



Error 6: Not cleaning your data and believes that it meets with the assumptions

- Remove incomplete cases
- Missing data
- Outliers

Example, age of university students is 18, 19, 20, 21, 18, 85, so mean is (30.17) which is wrong result

- Checking normality of data

- Make transformations of variable as needed

Error 7: Wrong Visualizing Data, graphing to see pattern of data



Line chart is best for showing data change over time,

For example showing sales trends, or global warming temperatures over decades.

Mistakes I have seen involve using line charts with categorical Y axis (e.g. a specific product).

Another thing to avoid using line chart are showing too many variables on one chart.



Pie and Doughnut Charts

Those are nice looking charts on shape of circle, that are meant to represent data that compose one whole. e.g. a survey answers distribution.

Visual mistakes I see a lot is using pie charts with way to many categories (showing max 6 categories)





Horizontal bar charts vs Vertical bar charts





Charts that do not start at zero (misleading)



Too Bad : 3D graphs

Error 8: Seeking a statistically significant outcome

Why researchers do not like to have NON Significant result

- Significant result is easy to compare with other results
- Some of the journals do not like to publish not significant result
- How do you figure out not significant result, Explain???
- Why this is not significant
- Why it is interesting if it is not significant

Error 9: No link between research questions and results

Thank you for your listening

STAT Office

for Statistical Data Analysis and Training

https://www.facebook.com/SOSDAT	لىنكى پەيجى
	فەيس بوک
https://www.youtube.com/channel/UCBzV1G_2a176r	لىنكى چەنالى
<u>OKEcWym-iA</u>	يوتيوب
https://www.sosdat.com	لینکی ویْبسایت

