13. Bias in research

Albulene Grajcevci

edutech
SAARLAND UNIVERSITY

Agenda

1. Understanding bias
2. Violations of assumptions
3. Additivity and linearity
4. Normality
5. Examples about normality
6. Homoscedasticity/homogeneity of variance
7. Independence
8. How to reduce bias

1. Understanding bias

- **What is bias?**

1. The outliers

- **Imagine a student that never attends class**

2. Violations of assumptions

- **Additivity and linearity**
- **Normality**
- **Homoscedasticity/homogeneity of variance**
- **Independence**

3. Additivity and linearity

Additivity and linearity

….this assumptions means that the relationships between the outcome variable (DV) and the predictors is accurately described by the equation

….. it also means that if you include several predictors their combined effect is best described by adding their effects together

Parametric tests are based on linear models

If you cannot ensure linearity nothing else matters

4. Normality

Normality

….the parametric tests we use expect distributions related to the normal distribution.

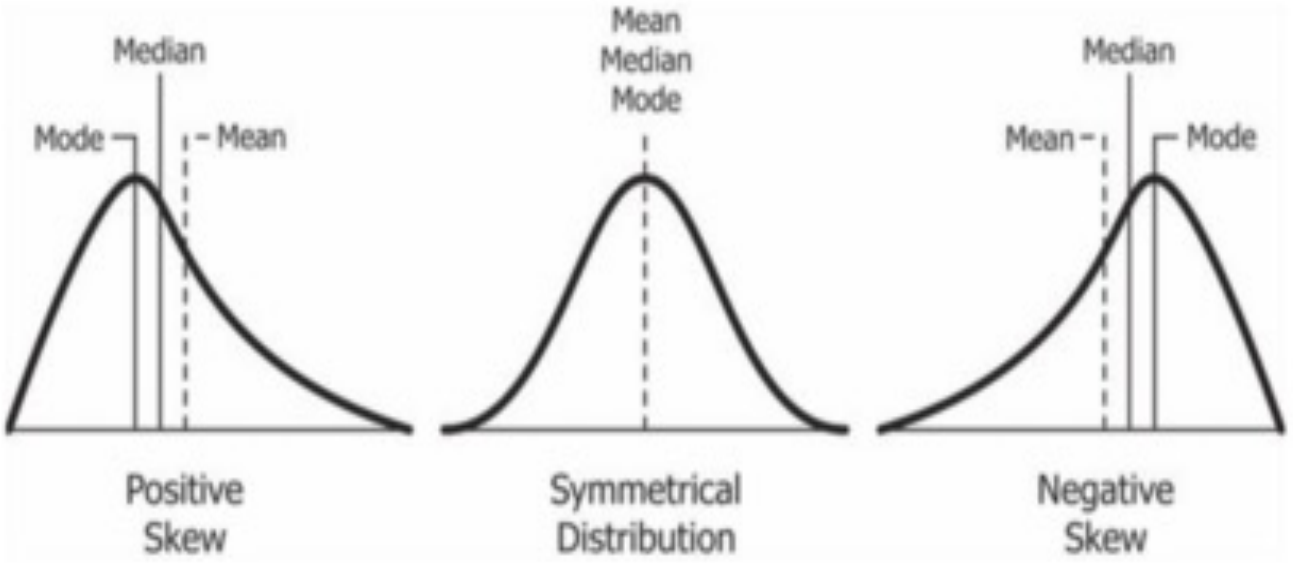
This means that if the parameter estimate is normally distributed then these test statistics and the p-values will be accurate

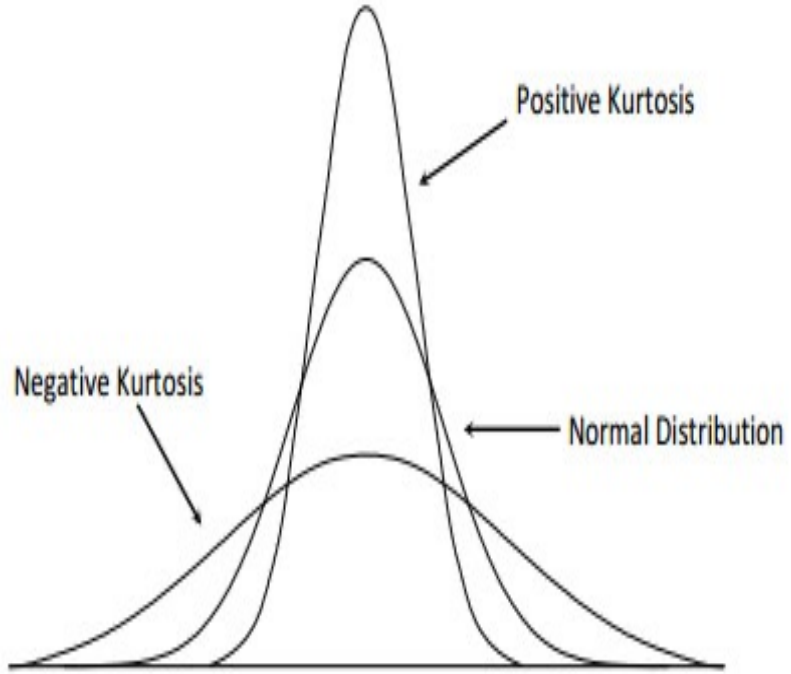
Exp. The mean is also a parameter

Particularly important if you claim that what you are measuring is normal

Conduct the Kolmogorov-Smirnov and the Shapiro-Wilk tests

4. Normality


Median
Mode
Mean
Positive Skew
Mean
Median
Mode
Symmetrical Distribution
Median
Mean
Mode
Negative Skew


Positive Kurtosis
Negative Kurtosis
Normal Distribution

5. Examples about normality

What happens if you have a parameter that is not normally distributed?

What does this mean for your tests?

What happens if in your sample everyone is a high achiever?

Questions: How did you get to the data?

The central limit theorem:

AS YOUR SAMPLE GETS LARGER- THE NEED TO WORRY DECREASES

NEGATIVE Correlation- Remember this?

6. Homoscedasticity/homogeneity of variance

This assumption means that the groups you have come from the populations with the same variance

The variance of the outcome variable should be stable at all levels of the predictor variable

It impacts confidence intervals and significance tests

Conduct the Levine`s test

7. Independence

This assumption means that the errors in your model are not related to one another.

Use the Chi-square test of independence

8. How to reduce bias

1. Trim the data: Delete a certain quantity of scores from the extremes.
2. Winsorizing: Substitute outliers with the highest value that isn't an outlier.
3. Apply a robust estimation method: A common approach is to use bootstrapping.
4. Transform the data: Apply a mathematical function to scores to correct problems.

8. How to reduce bias

Transforming data:

1. Positive skew, positive kurtosis, unequal variances, lack of linearity.

Log transformation: it addresses positive skew and linearity problems. NEGATIVE NUMBERS ARE A PROBLEM

Square root transformation: taking the square root of values. NEGATIVE NUMBERS ARE A PROBLEM

8. How to reduce bias

Transforming data:

2. Positive skew, positive kurtosis, unequal variances

Reciprocal transformation 1/variable: Dividing 1 by each score will reduce the impact of large scores

3. Negative skew

Reverse score transformations

THANK YOU

Questions are welcome :)