Quantitative Analysis for Research

CLEMENT TWUMASI

Email: twumasic@cardiff.ac.uk
School of Mathematics, Cardiff University, UK

April 10, 2021

Outline of presentation

- Brief introduction to statistics & statistical methods.
- Types of data variables.
- Steps for dealing with Questionnaire data.
- Other data structures for quantitative studies.
- Univariate & multivariate statistical tests.
- Regression analyses and variable selection techniques.
- Statistical software available for data analysis.
- Personal initiative on Advanced Data Analysis.
- Questions and Answers sections.

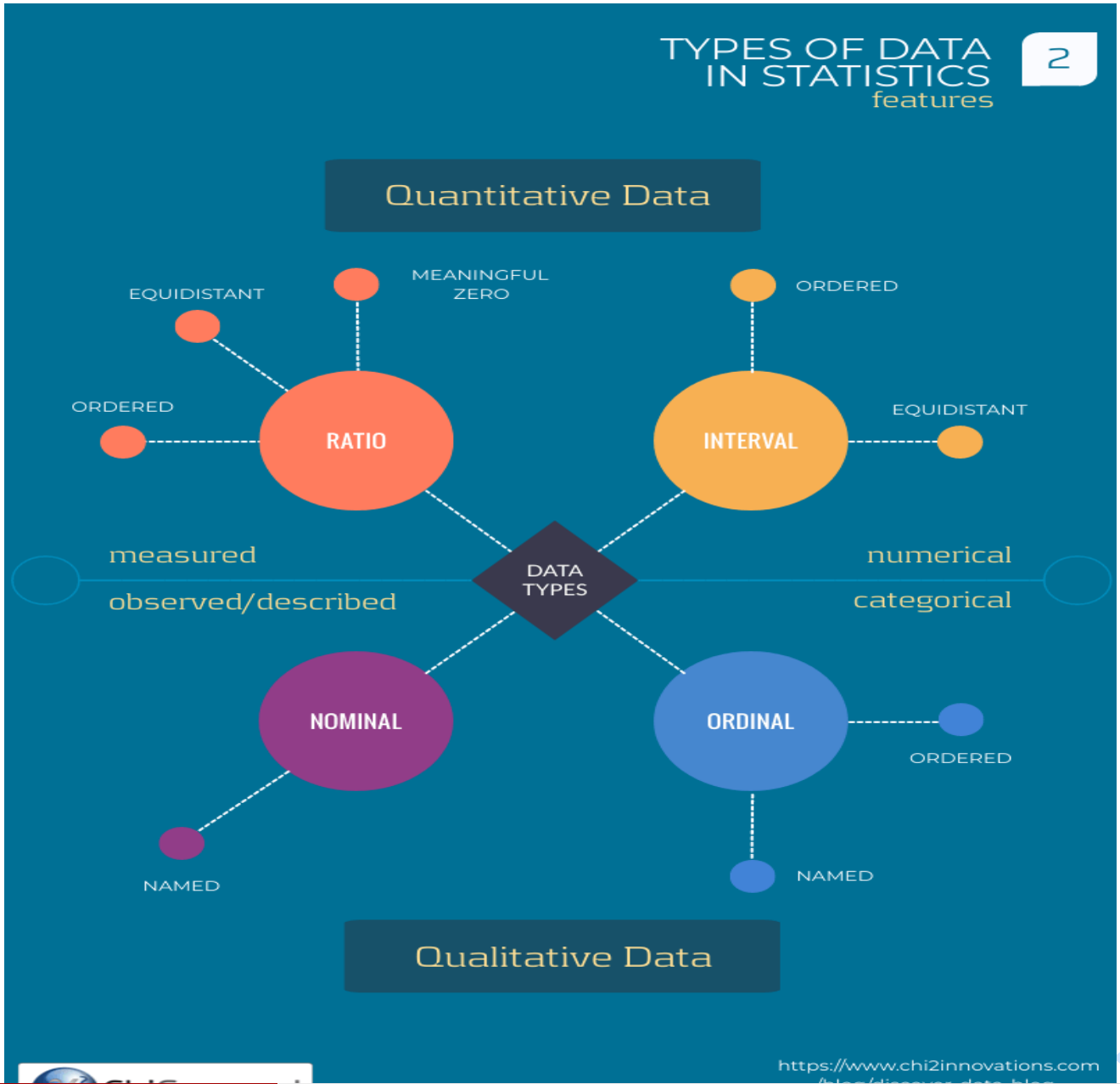
Brief introduction to statistics & statistical methods

- **What is Statistics?:** It basically deals with data collection, analysis, interpretation & making meaningful conclusions.
- **Types of statistics for data analysis:**
  1. Descriptive statistics (eg. central tendencies, measures of dispersion, descriptives plots, etc.)
  2. Inferential statistics (statistical tests and models)
- **Data source:**
  1. Primary data (collected by researchers directly from main sources through interviews, surveys, experiments, etc.)
  2. Secondary data (pre-existing data collected through primary and/or other sources)

Statistical methods for research

- Qualitative methods: including observations, textual or visual analysis (eg. from books or videos) and interviews (individual or focused groups)- Thematic analysis & data transcription are employed for qualitative analysis.
- Quantitative methods: numerical analysis of data collected through polls, questionnaires, and surveys, or by manipulating pre-existing statistical data using computational techniques.
- Mixed methods (Qualitative + Quantitative methods).

Types of data variables



Example of a questionnaire

Gender: ☐ Female ☐Male

Age (In Years): ☐ 15-20 ☐15-20 ☐ 26-30 ☐ Above 30

Education Level: ☐ Matric ☐ Intermediate ☐Under Graduate ☐ Post Graduate

Status: ☐ Single ☐ Married

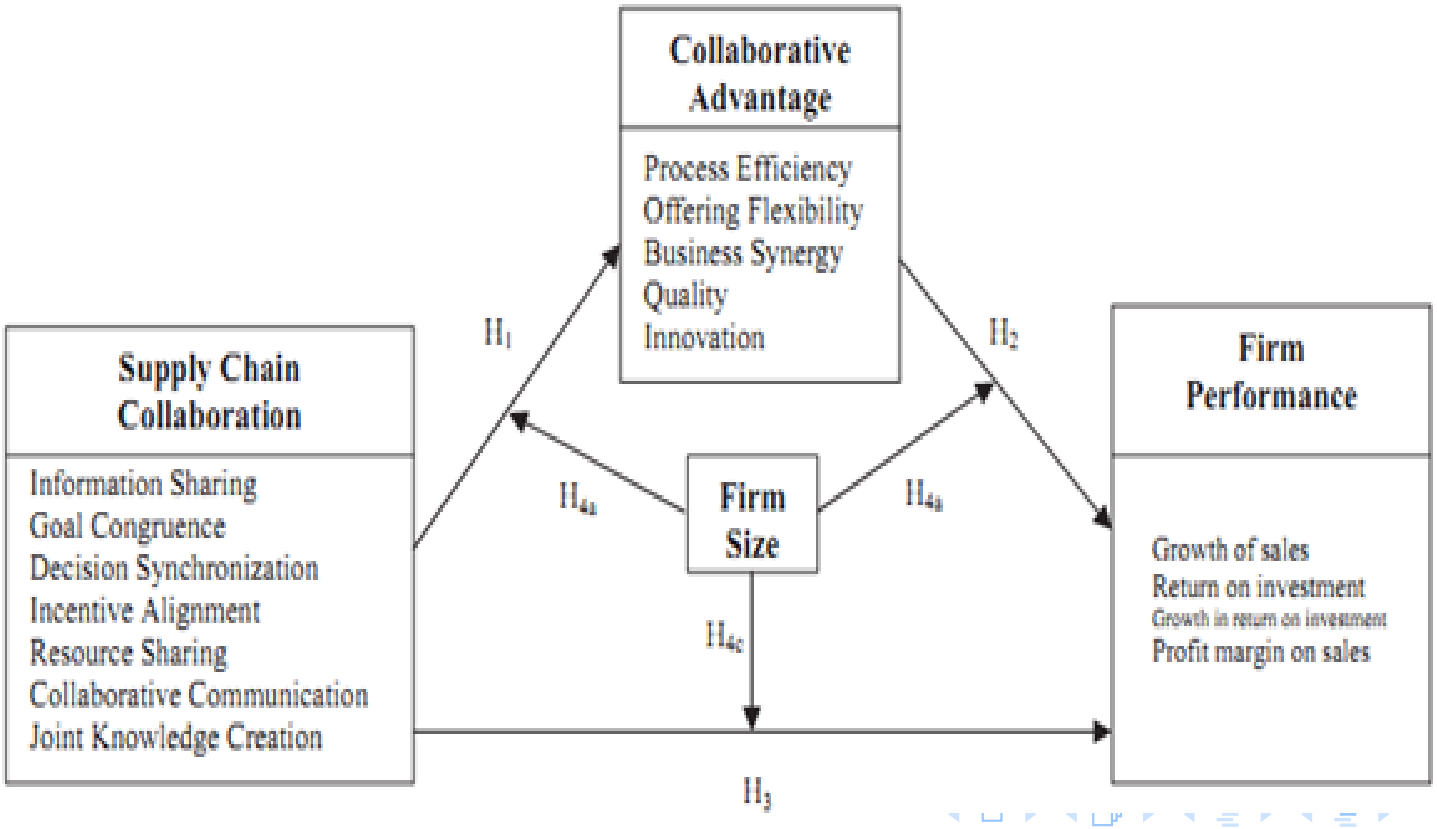
Answer the following questions keeping in mind celebrities in ads that you watch on televisions, radio, magazines, billboards and newspapers

1)Strongly Disagree 2) Disagree 3) Neutral 4) Agree 5) Strongly Agree

S.no	Item	SD	D	N	SA	A
1.	Do you buy a product which is endorsed by a celebrity?					
2.	Do you get attracted to buy a product/brand endorsed by a celebrity?					
2.	Do you think that ads having celebrities are more effective than those which don't?					
3.	Would you switch from you regular products to a new product endorsed by your favorite actor/ actress?					
4.	Does the presence of a celebrity help you recognize a brand?					
5.	Are the claims made in advertisements having celebrities believable?					
6.	Celebrity endorsed ad grab an audience's attention more easily than a standard ad.					
7.	Are you able to identify the product because of the celebrity associated with it?					
8.	I don't believe the celebrities also use those products which they themselves endorse.					
9.	If a celebrity you disliked was endorsing a brand you often purchase from, would this change your consumer interests?					
10.	I find product endorsed by celebrities are trustworthy.					
11.	I find product endorsed by celebrities are informative.					
12.	I find product endorsed by celebrities are interesting.					
13.	Celebrities help me more to remember a brand/product.					
14.	My favorite celebrity give a positive image to the endorsed brand.					
15.	I think that "negative publicity" about a celebrity endorser can influence my decision of buying a brand?					
14.	I will stop buying a brand if my favorite celebrity endorsing it gets involved in scandal					
15.	I buy a product just because the celebrities are using it.					
16.	I think celebrity endorsement is an important factor when I make my decision					
17.	I would buy a brand if my favorite celebrity is endorsing it.					

Steps for dealing with Questionnaire data (Con't)

1. Check for reliability or internal consistency of the responses of the questionnaire items (using Cronbach reliability analysis)- for either standardised or unstandardised questionnaires.
2. In quantifying latent variables (eg. satisfaction, child negative emotionality, etc.) employ Correspondence Factor Analysis (CFA). Hierarchical regression and/or Structural Equation Model (SEM) can be used to explore complex relationships between dependent & independent variables. Eg of latent variables and its constructs:



Other data structures for quantitative studies (cross-sectional data, time series data, panel data, etc)

Table 1. Example of cross sectional data

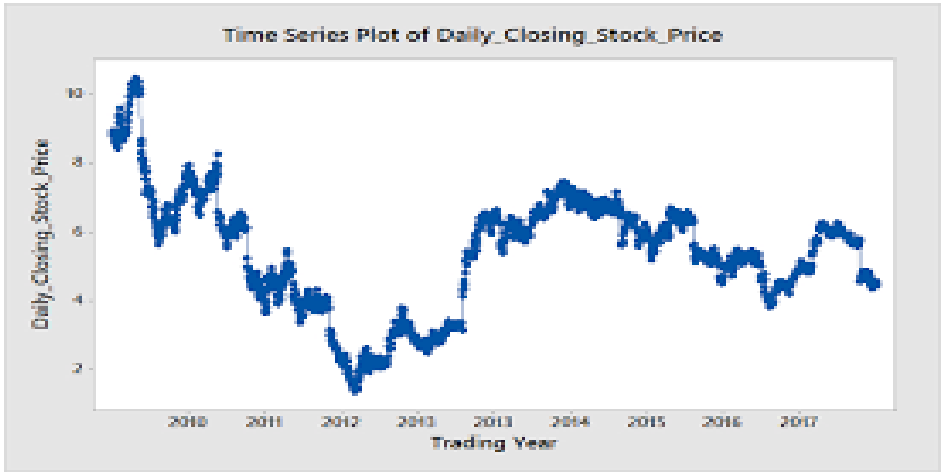
indiv (i)	year	wage	edu	exper	female
1	1990	3.10	11	2	1
2	1990	3.24	12	22	1
.	.	.	.	.	.
100	1990	5.30	12	7	0

Table 2. Example of pooled cross sectional data

house (i)	year (t)	hprice	bdrms	bthrms	sqrft
1	2000	85,500	3	2.0	1600
2	2000	67,300	3	2.5	1400
.	.	.	.	.	.
100	2000	134,000	4	2.5	2000
101	2010	243,000	4	3.0	2600
102	2010	65,000	2	1.0	1250

Table 3. Example of panel data (aka, longitudinal data)

obs.	i	t	murder rate	pop density	police
1	1	2000	9.3	2.24	440
2	1	2001	11.6	2.38	471
3	2	2000	7.6	1.61	75
4	2	2001	10.3	1.73	75
.	.	.	.	.	
199	100	2000	11.1	11.1	520
200	100	2001	17.2	17.2	493



- Mean tests (Parametric tests: Paired t-tests, independent t-test, ANOVA, Bonferroni pairwise comparison tests, Repeated Measures ANOVA, MANOVA,etc).
- Median tests (Non-parametric tests: Wilcoxon sign test, Mann-Whitney, Kruskal-Wallis test, Bonferroni-Dunn's test, Friedman test, Multivariate Kruskal-Wallis test,etc).
- Test of proportions for one or more groups, etc.
- Correlation test/analysis
- etc.

- GLM (OLS regression, binomial & multinomial regression, poisson regression, quasi poisson regression, negative binomial regression, etc) [cross-sectional data]
- Generalized Linear Mixed Models (for all types of dependent variables) [these class of models are for longitudinal data]
- Generalized Additive Models (fixed and mixed-effect types; both cross-sectional & longitudinal data), Panel regression models and FDA [these class of models for longitudinal data]
- Machine learning algorithms (Classification tree, Random forest, Gradient Boosting Machine, etc) [both cross-sectional & longitudinal data]

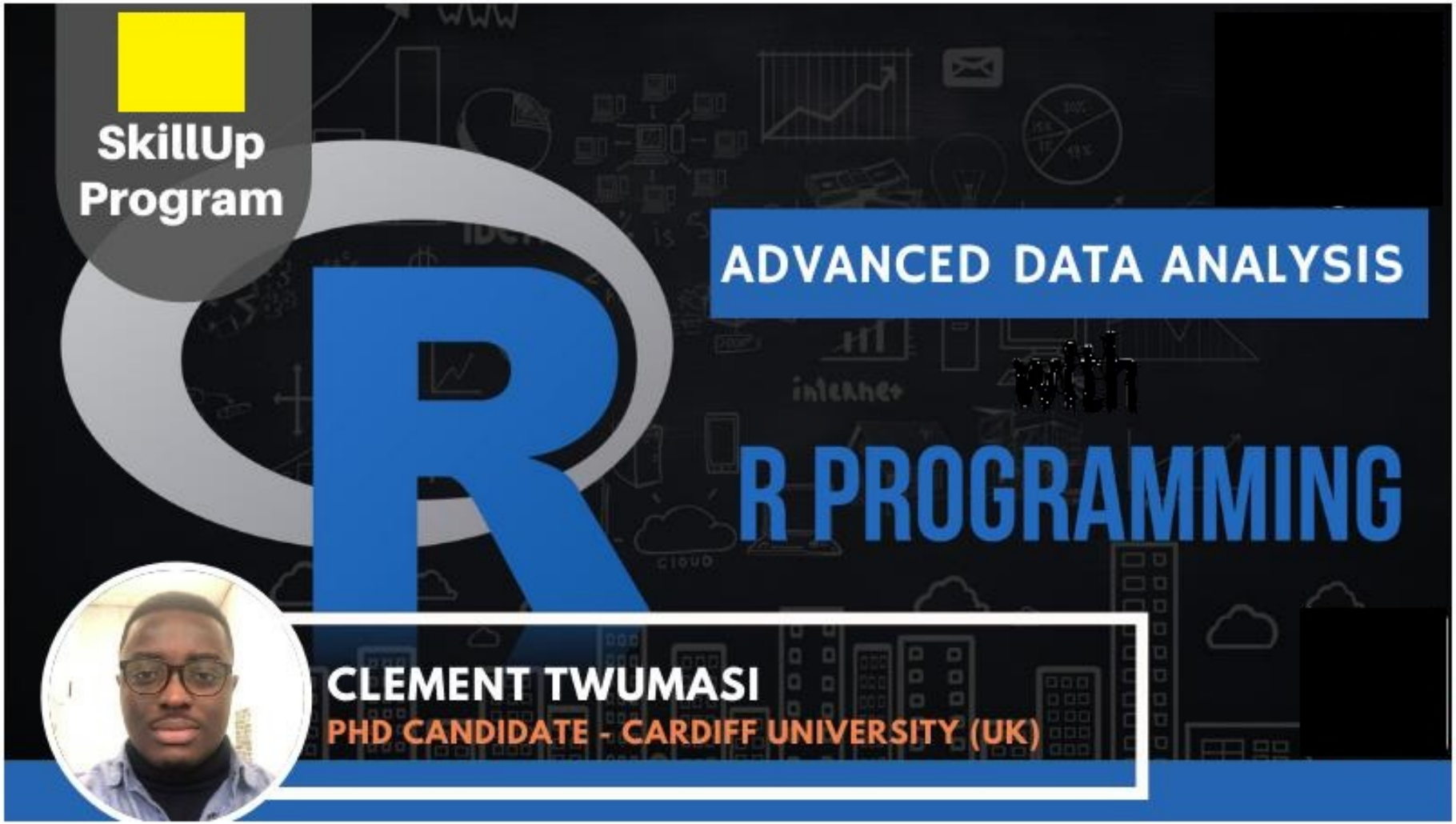
Method of variable selection: Stepwise regression, Penalized regression (Ridge, LASSO and Elastic net), Recursive Feature Selection.

NB: Beware of multicollinearity and its effects.

Statistical software available
(their advantages over the other)

- SPSS, STATA, EVIEWS, MINITAB, GRETL, XLSTAT, etc,
- R
- Python
- Julia
- SAS
- Matlab,
- C language , C++, etc.

Personal initiative on Advanced Data Analysis



Questions and Answers section