Statistical Data Analysis

Prof. Dr. Nizamettin AYDIN

naydin@yildiz.edu.tr

http://www3.yildiz.edu.tr/~naydin

Course Details

- Course Code: BLM3590
- Course Name: Statistical Data Analysis
- Credit: 3
- Nature of the course: Lecture
- Course web page:
   http://www3.yildiz.edu.tr/~naydin/na_SdA.htm
- Instructors: Nizamettin AYDIN

   Email: naydin@yildiz.edu.tr

Rules of the Conduct

- No eating /drinking in class
  - *except water*
- Cell phones must be kept outside of class or switched-off during class
  - *If your cell-phone rings during class or you use it in any way, you will be asked to leave and counted as unexcused absent.*
- No web surfing and/or unrelated use of computers,
  - when computers are used in class or lab.

Rules of the Conduct

- You are responsible for checking the class web page often for announcements.
  - http://www3.yildiz.edu.tr/~naydin/na_ SdA.htm
- Academic dishonesty and cheating
  - will not be tolerated
  - will be dealt with according to university rules and regulations
    - http://www.yok.gov.tr/content/view/475/
    - Presenting any work that does not belong to you is also considered academic dishonesty.

Attendance Policy

- The requirement for attendance is **70%**
  - **Hospital reports** are **not** accepted to fulfill the requirement for attendance.
  - **The students, who fail to fulfill the attendance requirement, will be excluded from the final exams and the grade of F0 will be given.**

Assesment

Quiz	:	10%
Midterm	:	25%
Homework	:	20%
Final	:	40%
Attendance & participation	:	05%

(The requirement for attendance is 70%)

Objective

- Overall
  - Reinforce your understanding of statistical data analysis
- Specific
  - Concepts of data analysis
  - Some data analysis techniques
  - Some tips for data analysis
- Try to cover every bit and pieces of statistical data analysis techniques

Data analysis – "The Concept"

- Approach to de-synthesizing data, informational, and/or factual elements to answer research questions

- Method of putting together facts and figures to solve research problems

- Systematic process of utilizing data to address research questions

- Breaking down research issues through utilizing controlled data and factual information

Categories of data analysis

- Narrative (e.g. laws, arts)
- Descriptive (e.g. social sciences)
- Statistical/mathematical (pure/applied sciences)
- Audio-Optical (e.g. telecommunication)
- Others
- Most research analyses adopt the first three
- The second and third are most popular in pure, applied, and social sciences

Statistical Methods

- Something to do with "statistics"
  - Statistics
    - meaningful quantities about a sample of objects, things, persons, events, phenomena, etc.
    - Widely used in many fields (social sciences, engineering, etc.)
    - Simple to complex issues. E.g.
      - correlation
      - anova
      - manova
      - regression
      - econometric modelling
- Two main categories:
  - Descriptive statistics
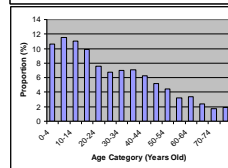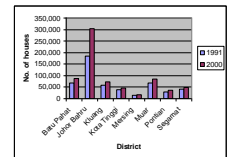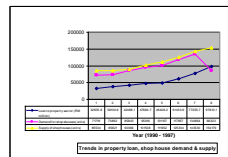 - Inferential statistics

Descriptive statistics

- Use sample information to explain/make abstraction of population "phenomena"
- Common "phenomena":
  - Association (e.g. $\sigma1,2.3 = 0.75$)
  - Tendency (left-skew, right-skew)
  - Causal relationship (e.g. if X, then, Y)
  - Trend, pattern, dispersion, range
- Used in non-parametric analysis
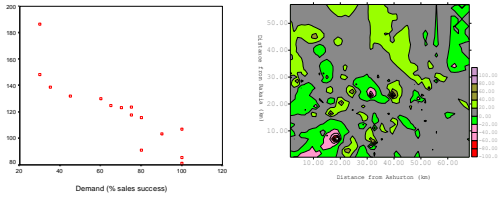 - e.g. chi-square, t-test, 2-way anova)

Examples of "abstraction" of phenomena
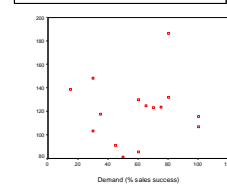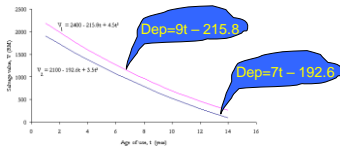
Examples of "abstraction" of phenomena

Inferential statistics

- Using sample statistics to infer some "phenomena" of population parameters
- Common "phenomena": cause-and-effect
- One-way relationship — $Y = f(X)$
- Multi-directional relationship — $Y1 = f(Y2, X, e1)$ / $Y2 = f(Y1, Z, e2)$
- Recursive — $Y1 = f(X, e1)$ / $Y2 = f(Y1, Z, e2)$

- Use parametric analysis

Examples of relationship

Dep=9t – 215.8

Dep=7t – 192.6

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	1993.108	239.632		8.317	.000
	Tanah	-4.472	1.199	-.190	-3.728	.000
	Bangunan	6.938	.619	.705	11.209	.000
	Ansilari	4.393	1.807	.139	2.431	.017
	Umur	-27.893	6.108	-.241	-4.567	.000
	Flo_go	34.895	89.440	.020	.390	.697

Which one to use?

- Nature of research
  - Descriptive in nature?
  - Attempts to infer, predict, find cause-and-effect, influence, relationship?
  - Is it both?
- Research design (including variables involved)
- E.g. outputs/results expected
  - research issue
  - research questions
  - research hypotheses

Common mistakes in data analysis

- Wrong techniques. E.g.

Issue	Data analysis techniques	
	Wrong technique	Correct technique
To study factors that "influence" visitors to come to a recreation site	Likert scaling based on interviews	Data tabulation based on open-ended questionnaire survey
"Effects" of KLIA on the development of Sepang	Likert scaling based on interviews	Descriptive analysis based on ex-ante post-ante experimental investigation

Note: Likert scaling cannot show "cause-and-effect" phenomena!

- Infeasible techniques. E.g.
  - How to design ex-ante effects of KLIA?
    - ex ante: based on forecasts rather than actual results
  - Development occurs "before" and "after"!
  - What is the control treatment?

Common mistakes– "Abuse of statistics"

Issue	Data analysis techniques	
	Example of abuse	Correct technique
Measure the "influence" of a variable on another	Using partial correlation (e.g. Spearman coeff.)	Using a regression parameter
Finding the "relationship" between one variable with another	Multi-dimensional scaling, Likert scaling	Simple regression coefficient
To evaluate whether a model fits data better than the other	Using R^2	Many – a.o.t. Box-Cox χ^2 test for model equivalence
To evaluate accuracy of "prediction"	Using R^2 and/or F-value of a model	Hold-out sample's MAPE
"Compare" whether a group is different from another	Multi-dimensional scaling, Likert scaling	Many – a.o.t. two-way anova, χ^2, Z test
To determine whether a group of factors "significantly influence" the observed phenomenon	Multi-dimensional scaling, Likert scaling	Many – a.o.t. manova, regression

How to avoid mistakes - Useful tips

- Crystalize the research problem
  - operability of it!
- Read literature on data analysis techniques
- Evaluate various techniques that can do similar things w.r.t. research problem
- Know what a technique does and what it doesn't
- Consult people, esp. supervisor
- Pilot-run the data and evaluate results

19

Principles of analysis…

- Goal of an analysis:
  - To explain cause-and-effect phenomena
  - To relate research with real-world event
  - To predict/forecast the real-world phenomena based on research
  - Finding answers to a particular problem
  - Making conclusions about real-world event based on the problem
  - Learning a lesson from the problem

20

…Principles of analysis…

- Data cannot talk
- An analysis contains some aspects of scientific reasoning/argument:
  - Define
  - Interpret
  - Evaluate
  - Illustrate
  - Discuss
  - Explain
  - Clarify
  - Compare
  - Contrast

21

…Principles of analysis

- An analysis must have four elements:
  - Data/information (what)
  - Scientific reasoning/argument
    - what? who? where? how? what happens?
  - Finding
    - what results?
  - Lesson/conclusion
    - so what? so how? therefore, …

22

Principles of data analysis…

- Basic guide to data analysis:
  - Analyze, not narrate
  - Go back to research flowchart
  - Break down into research objectives and research questions
  - Identify phenomena to be investigated
  - Visualize the expected answers
  - Validate the answers with data
  - Do not tell something not supported by data

23

…Principles of data analysis…

Shoppers	Number
Male	
Old	6
Young	4
Female	
Old	10
Young	15

- More female shoppers than male shoppers
- More young female shoppers than young male shoppers

24

4

…Principles of data analysis

- When analyzing:
  – Be objective
  – Be accurate
  – Be true
- Separate facts and opinion
- Avoid "wrong" reasoning/argument.
  – E.g. mistakes in interpretation.

25
26

5