

BUSI 275: Business Statistics

8 Sep 2011

Instructor: Dr. Sean Ho

TA: Anne Schober

busi275.seanho.com

- No **food**/drink in the computer lab, please!

- Please pick-up:*
- **Syllabus**

Outline for today

- **Welcome**, devotional, introductions
- Administrative details: **syllabus**, schedule
 - myCourses, Excel, textbook
- **Introduction** to statistics for business:
 - Decision making and asking good **questions**
 - **Population** vs. **sample**
 - **Variables**: levels, IV/DV, cross-sectional
 - Stages/**cycles** in statistical analysis
- Term **Project** and **HW1**

What is statistics?

- Data-driven **decision** making
 - **Evidence**-based, not (only) “gut feeling”
- A way to **answer** vital Q's about business processes
 - “Which **market sgmt** is most **price**-conscious?”
 - “Is **online** advertising more effective than **print**?”
- A way to **ask** more relevant questions
 - “How do we measure **customer satisfaction**?”
- A way to determine **what** questions to ask
 - “What factors have the strongest influence on **employee retention**?”

Basic terms



- **Population:** group of interest
 - e.g., **Canadians aged 18-25**
- **Sample:** participants in our study
 - e.g., **200 students interviewed on TWU campus**
 - **Sampling** is how we draw a sample from a pop
 - **Inferences** are estimates (guess) on larger pop
- **Variable:** measurable of interest
 - e.g., **“\$/mo spent on food”**
- **Observation:** value of a variable for a single participant
 - e.g., Jane spends **\$200/mo** on food

Levels of measurement

- **Nominal** (categorical):
 - Province, colour, store branch, any yes/no
 - “Are you satisfied as a customer?”
- **Ordinal** (has an ordering, $</>$ makes sense):
 - Letter grade, “satisfactory ... unsatisfactory”
 - “Are you very satisfied, somewhat satisfied, ...?”
- **Interval** (+/-/avg makes sense, but 0 is arbitrary):
 - °C/F, **Likert** scale (“on a scale of 1-5”)
- **Ratio** (mult/divide makes sense):
 - Salary, quantity of sales, height in cm

Direction of influence

- We nearly always care about **relationships** amongst variables:
 - “Does **advertising medium** affect **sales**?”
- Often, one variable **drives**/influences another:
 - **Predictor** (independent variable, **IV**) drives
 - **Outcome** (dependent variable, **DV**) is influenced



Cross-sectional vs. time-series

- **Cross-sectional** data look at a **snapshot** in time:
 - e.g., **2010** revenue for different store branches
- **Time-series** data track the same **variables** on the same **participants**, but at **several** points in **time**:
 - Annual revenue for store branches, **2000-2010**
- Time-series data need to worry about
 - **Attrition** (**missing** data)
 - **Sampling** in time (e.g., **monthly** vs. **annual**)
 - **Uneven** time (**2010**, **2009**, and “<2009”)
- We will mostly examine cross-sectional data here

Cycles in statistical analysis

- Formulate **research question** (RQ)
- **Gather** data: sampling, metrics
- **Prep data**: input errors/typos, missing data, obvious outliers
- **Explore** variables: IV, DV, charts
- **Model** building: choose a model based on RQ
- Check **assumptions** of model
 - If not, either clean **data** or change **model**
 - May need to **modify RQ!**
- Run final model and interpret **results**



Research question: example

- RQ: are men taller than women?
 - Is this relationship real? How strong is it?
- What are the **variables**? **IV/DV**? Level of **meas**?
 - Levels of **measurement**: categorical, ordinal, scale (interval, ratio)
 - IV: **gender** (dichot), DV: **height** (scale)
- What type of **test** should we use?
 - Independent samples: **t-test**
- **Limitations**/assumptions of this test?

Model-building process

- Operationally **define** a phenomenon: **variables**
- **Measure** it (collect data): how to do **sampling**?
- **Build** a model: verify data meet **assumptions** and **input** data into model
- Draw **conclusions** in the “**real world**” population
 - e.g., if child A has **2** apples, B has **6**, and C has **1**, how many apples is a child most likely to have?
 - **Individual** vs. **group**
- “Everything that can be **counted** does not necessarily **count**; everything that **counts** cannot necessarily be **counted**.” (Albert Einstein)

Example: Retail duration of stay

- **RQ**: Does **volume** of music affect **duration** of customer stay in retail shops?
- **Population, sample**: how to gather data?
- **Variables**: how to measure?
 - **Volume**, in dBA
 - **Duration of stay**, in seconds
- **Predictor (IV) / outcome (DV)**?
 - Predictor: **volume**. Outcome: **duration of stay**
- **Levels** of measurement?
 - Volume (dBA): **ratio**. Duration (seconds): **ratio**
- **Moving beyond**: what **other** factors affect duration?

Term Project

- A big part of this course is your term **project**:
 - Find suitable **data**: get your own, or use existing
 - **Propose** a statistical analysis of it
 - Get approval by **Research Ethics Board**
 - Go through “spiral” of statistical **analysis**
 - **Write** it up in an MLA-style manuscript
- **Groups** of up to 4 people
 - Email me when you have your group

Project step 1: Finding data

- Can be **existing** data, or you may gather your **own**
 - Collecting data takes time! (and may need REB)
 - **No simulated** (made-up) data
- Minimum **sample size**: 80
- Type of analysis:
 - Distribution fit, time series, multiple **regression** or **ANOVA**
- Possible **sources**: your own data, faculty members, publicly available / government data (StatCan, Chamber Commerce, etc.)

Dataset description: due 4Oct

- Written **description** of the dataset you will be using and the particular **variables** you consider
- Preliminary **explorations** of the data
 - Descriptives, histograms, boxplots, etc.
 - Include as **figures** in your write-up
- Formated **neatly** in a document (Word, etc.)
 - MLA style or similar
- **Upload** your document to myCourses
 - One person can submit for whole group

Project step 2: REB (due 11Oct)

- Approval by TWU **Research Ethics Board** is required **before** any new analysis may be done!
- You are **not** allowed to start your analysis until you get REB approval (expect 3-6 weeks)
 - You may not even **recruit** your study subjects!
- Use either the “**Request for Ethical Review**” form or the “**Analysis of Existing Data**” form
- If using existing non-public data, you need **written permission** from the original owner of the data
- REB approval is not needed for **publicly available** data

Project step 3: Proposal (due 25 Oct)

- Written **proposal** of the particular analysis you plan to do on the dataset
 - Describe **variables** and **why** we should care
 - State specific **research questions**
 - Anticipate possible **problems**, plan
 - Plan for how you will divide the work amongst your **team members**

Project step 4: Present (29Nov-1Dec)

- 10-min in-class presentation
- Assume **target audience** is *not* familiar with statistics
 - e.g., your company's CEO or board
- **Motivate** why we should care about your topic
- Have some **preliminary results** to show
- Every **team member** must participate
- Also complete **feedback forms** for other teams' presentations

Project step 5: Paper (due 7Dec)

- Aim at **non-statistician** (CEO, etc.)
 - But include enough details to **reproduce** study
- Proper, **professional** English
 - Format in **MLA** or similar style
- **Related work** / background research
 - Cite **references**
- Include relevant **figures** / tables
 - Can include more in **appendix** or separate **Excel**

TODO

- **HW1** (ch1-2): due next week Thu **15Sep**
 - Format as a clear, neat **document**
 - Also upload **Excel** spreadsheet with your work
 - HWs are to be **individual** work
- Get to know your classmates and **form teams**
 - **Email me** when you know your team
 - You can come up with a good **name**, too
- Discuss **topics**/variables you are interested in
 - Find **existing** data, or **gather** your own?