10.3-10.4: Pairwise t-Test and t-Test on Proportions

- HW6 due Thu
- Please download: Mileage.xls

25 Oct 2011
BUSI275
Dr. Sean Ho

TRINITY WESTERN UNIVERSITY

Outline for today

- Exploratory analysis
  - Preview of statistical tests for your projects
- Repeated measures
  - Overview of Excel TTEST() types
- T-test on paired data
  - Calculating SE
  - Using Excel
- T-test on proportions

Exploratory analysis

- Choosing good research questions:
- Start with the outcome variable (DV)
  - e.g., sales volume
- Research background (prior literature) on the DV to find likely predictors
  - e.g., marketing budget, consumer trends, new products from competitors, etc.
- Select some effect/predictor(s) to examine
  - In your analysis, control for other covariates
- Correlation ≠ causation: look for hidden vars
  - e.g., ice cream correlates with drownings!
    - Why? What are they both correlated with?

Analysis Types by IV/DV

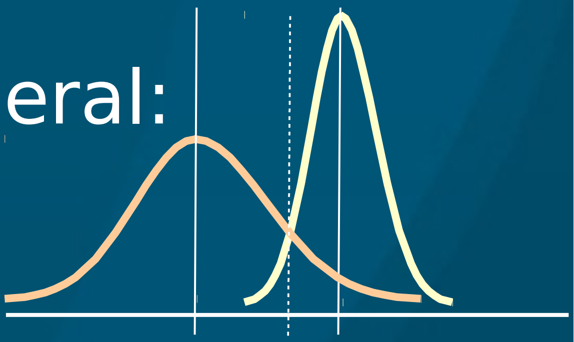
- DV quantitative, IV categorical:
  - IV dichotomous (two groups): t-test
  - IV has many groups: ANOVA
  - Multiple categorical IVs: Factorial ANOVA
    - Controlling for covariates: ANCOVA
- DV quantitative, IV quantitative:
  - One IV: Simple Regression
  - Multiple IVs: Multiple Regression
    - Also if mix of categorical/quant IVs
- DV dichotomous: Logistic Regr. (survival an.)
- DV ordinal: Ordinal Regr.
  - … and much more!

Repeated measures

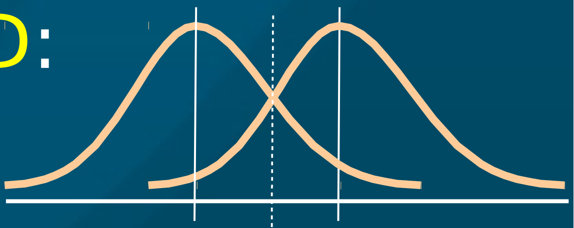
- Apply same measurement to same subjects, but at different points in time:
  - e.g., annual revenue, 2000-2010
  - Time series / longitudinal data
- Or under different conditions:
  - e.g., highway vs. city mileage (on same car!)
  - e.g., wife's income, husband's income
    - (What is the unit of observation?)
- The measurements are linked to each other
  - Not independent
- Paired data is the simplest repeated measure
  - Use a t-test on the pairwise differences

Types of t-test (as in Excel)

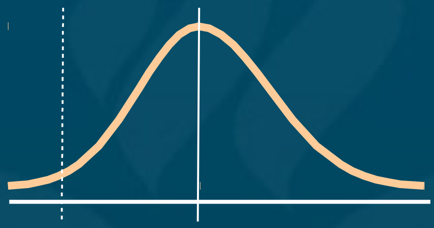
- Type 3: two indep groups, most general:
  - H_A: $\mu_1 - \mu_2 \neq 0$ (or >0)
  - SE = $\sqrt{(SE_1^2 + SE_2^2)}$, df is messy



- Type 2: two indep groups, similar SD:
  - H_A: $\mu_1 - \mu_2 \neq 0$ (or >0)
  - SE = $s_p\sqrt{(1/n_1 + 1/n_2)}$, df = $df_1 + df_2$



- Type 1: paired observations:
  - Form pairwise diffs: n = # pairs
  - H_A: $\mu_d \neq 0$ (or >0)
  - SE = $s_d / \sqrt{n}$, df = $n-1$



Paired data t-test

- e.g., Mileage.xls
- Calculate the pairwise differences: =A2-B2, fill
- Find n, mean ($\bar{d}$), and SD (s_d) of pairs:
  - COUNT(), AVERAGE(), STDEV()
  - SD of diffs is not the same as diff of SDs!
- Calculate standard error: SE = $s_d/\sqrt{n}$
- Find t-score: $(\bar{d} - 0) / SE$
- Use TDIST() to find p-value, compare w/α
  - TDIST(*t*, *n*-1, *tails*)
- All-in-one Excel function:
  TTEST(*before*, *after*, *tails*, 1)

T-test on proportions

- e.g., customer satisfaction vs. bank branch:
  - At Langley, 160/200 customers satisfied
  - At Abbt., 210/300 satisfied
  - Is there a significant difference?
- Use normal approximation to binomial:
  - When n is big enough and p is not extreme
  - Rule of thumb: both np, $nq > 5$ (both groups)
- Normal dist means no worries about df
  - Just need standard error: $SE = \sqrt{SE_1^2 + SE_2^2}$
  - Where each $SE_i = \sqrt{p_i q_i / n_i}$

Proportions: bank example

- Langley: $SE_L = \sqrt{(160*40 / 200^3)} \approx 2.828\%$
- Abbt.: $SE_A = \sqrt{(210*90 / 300^3)} \approx 2.646\%$
- Combined standard error is $SE = \sqrt{(SE_L^2 + SE_A^2)}$
  $= \sqrt{(160*40 / 200^3 + 210*90 / 300^3)} \approx 3.873\%$
- Sample difference of proportions is
  $p_L - p_A = (160/200) - (210/300) = 10\%$
- This means a z-score of
  $z = ((p_L - p_A) - 0) / SE \approx 10\% / 3.873\% \approx 2.582$
- Find the p-value (2-tailed):
  - =2*(1-NORMSDIST(2.582)) → 0.0098
  - Reject H_0: yes, there is a difference

Alternate SE on proportions

- The SE above is used for confidence intervals
- The textbook offers a second form of the SE for binomial proportions:

$$SE = \sqrt{\bar{p}\,\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

- Where $\bar{p}$ is the pooled proportion:

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

- This is equivalent to the "χ^2 test of goodness-of-fit" we will learn in ch13
  - Most stats software uses this method

TODO

- Project proposal due tonight
  - Sample size (unit of observation?)
  - Outcome variable
  - Predictor variables
  - What other predictors might make sense?
- HW6 (ch9-10): due Thu 27 Oct