Semester Review

- **Paper** due tomorrow
- **Final** this Sat

6 Dec 2011
BUSI275
Dr. Sean Ho

TRINITY WESTERN UNIVERSITY

Overview: foundation

- Intro: variables, sampling (Ch1)
- Exploring data:
  - Via charts (Ch2), via descriptives (Ch3)
- Probability and independence (Ch4)
- Probability distributions:
  - Discrete: binom, Poisson, hypg (Ch5)
  - Continuous: norm, unif, expon (Ch6)
- Sampling distributions (Ch7, 8)
  - SDSM (norm and t-dist), binomial
  - Types of problems: % area, conf. int., n
- Hypothesis testing (Ch9):
  - H_0/H_A, rej / fail rej, Type-I/II, α/β, p-value

Overview: statistical tests

- *T*-tests (Ch10):
  - 1 sample mean (ch9)
  - Two independent samples (het σ, hom σ)
  - Paired data (Excel type 1)
- Regression (Ch14-15):
  - Linear model, predicted $\hat{y}$, residuals
  - R^2, F-test, t-test on slopes, interaction
- ANOVA (ch12):
  - One-way + Tukey-Kramer
  - Blocking (w/o repl) + Fisher's LSD
  - Two-way (w/repl), interaction
- χ^2 (ch13): contingency tables, O vs. E

Ch2-3: Exploring Data

- For nominal variables:
  - Charts: bar/col, pie
    - Joint distrib of 2 vars: pivot table
  - Stats: frequency distribution
- For scale (quantitative) (interval/ratio) vars:
  - Charts: histogram, ogive (cum), boxplot
    - 2 vars: scatter
    - Time series: line
  - Centre: mean, median, mode, (skew)
  - Quantile: Q_1/Q_3, %ile, IQR
  - Std dev: σ, s, CV, empirical rule, z-score

Ch4: Probability

- Tree diagrams
- P(A) notation, Venn diagrams
  - Sample space, outcome, event
  - ∩, ∪, complement
- Addition rule: $A \cup B = A + B - (A \cap B)$
  - Mutual exclusivity
- Conditional probability
  - What does it mean; how to find it (Bayes)
  - Statistical independence
    - Does $P(A|B) = P(A)$?

Ch5: Discrete distributions

- Binomial: BINOM(x, n, p, cum)
  - x: counting # of successes out of n trials
  - p: probability of success (binom proportion)
- Poisson: POISSON(x, λ, cum)
  - x: # occurrences within the time period
  - λ: mean (expected) # occ w/in the period
- Hypergeometric: HYPGEOMDIST(X, N, x, n)
  - X, N: # successes & tot size of population
    - Binomial p = X/N
  - x, n: # successes & tot size of sample

Ch6: Continuous distributions

- Normal: NORMDIST(x, μ, σ, cum)
  - Also NORMINV(area, μ, σ), NORMSDIST(z), NORMSINV(area)
- Uniform:
  - P(x) = 1/(b-a), μ = (a+b)/2, σ = $\sqrt{((b-a)^2/12)}$
- Exponential: EXPONDIST(x, λ, cum)
  - x: time between occurrences
  - λ: 1 / (mean time between occurrences)
    - λ = expected frequency of occurrences (e.g., occurrences per min)

Ch7-8: Sampling distributions

- Sampling distributions:
  - SDSM, w/σ: NORMDIST(), SE = σ/√n
  - SDSM, w/s: TDIST(), SE = s/√n
  - Binomial proportion: norm, SE = √(pq / n)
- Types of problems: area, μ, thresh, n, σ
  - Area: prob of getting a sample in given range
  - Threshold: e.g., confidence interval
  - n: minimum sample size

Ch9: Hypothesis testing

- Decision making
- H_0 vs. H_A , in words and notation (e.g., $\mu_1 \neq \mu_2$)
- Conclusions: reject H_0 vs. fail to reject H_0
- Risks/errors: Type-I vs. Type-II
  - Level of significance: α
  - Power: $1-\beta$
- p-value: what is it, how do we use it?

Ch10: t-tests

- T-test on 1 sample (ch8-9):
  - SDSM: $SE = s/\sqrt{n}$
  - Binomial proportions: $SE = \sqrt{(pq/n)}$
- T-test on two independent samples, general:
  - $SE = \sqrt{(SE_1^2 + SE_2^2)}$, df = complicated
- T-test on two independent samples, similar σ:
  - $SE = s_p \sqrt{(1/n_1 + 1/n_2)}$, $df = df_1 + df_2$
- T-test on two proportions:
  - $SE = \sqrt{(SE_1^2 + SE_2^2)}$, use z instead of t
- T-test on paired data:
  - $SE = s_d / \sqrt{n}$, df = (#pairs) – 1

Ch14-15: Regression

- Scatter plots and correlation, t-test on r
  - R^2 and % variability explained
- Linear model $Y = b_0 + b_1X + \varepsilon$
  - Finding+interpreting slope+intercept
  - Finding+interpreting s_ε (STEYX)
- Assumptions / diagnostics:
  - Linearity + homoscedasticity (residual plots)
  - Normality of residuals (histogram)
  - *(skip: non-collinearity + indep of resids)*
- Multiple regression + concept of moderation

Ch12-13: Categorical data

- Ch12: ANOVA:
  - H_0 / H_A, global F-test, concept of follow-up
  - One-way ANOVA + Tukey-Kramer
  - Blocking ANOVA + Fisher's LSD
    - F-test for main factor effect
    - F-test for whether blocking is needed
  - Two-way ANOVA
    - F-test for each main effect
    - F-test for interaction
- Ch13: χ^2 (O vs. E)
  - 1 var vs. uniform, normal
  - 2 vars (contingency table): independence