What to do today (Feb 17, 2023)?

Part 1. Introduction and Review (Chp 1-5)

Part 2. Basic Statistical Inference (Chp 6-9)

Part 3. Important Topics in Statistics (Chp 10-13)

§3.1. Analysis of Variance (ANOVA, Chp 10-11)

§3.1.1 Introduction

§3.1.2 One-Factor ANOVA (Chp 10)

§3.1.3 Multi-Factor ANOVA (Chp 11)

§3.1.4 Further Topics on ANOVA

§3.2 Introduction to Regression Analysis (Chp 12-13)

Part 4. Further Topics (Selected from Chp 14-16)

Some Logistics.

- Homework 5 has been assigned. It's due on Monday of the next week, the reading week.
- There will be no class on Tue and Thu next week.

STAT-285 Outline

- **Part 1. Introduction and Review**
- **Part 2. (Chp 6-9)**
  basic in statistics in general; essential to stat285 in particular
- **Part 3. Important Topics in Statistics (Chp 10-13)**

  §3.1A One-Factor Analysis of Variance (ANOVA) (Chp 10)

  §3.1B Multi-Factor ANOVA (Chp 11)

  §3.2A Simple Linear Regression Analysis (Chp 12)

  §3.2B More on Regression (Chp 13)
- **Part 4. Selected Topics in Statistics (from Chp 14-16)**

§3.1. Analysis of Variance (ANOVA, Chp 10-11)

§3.1.1 Introduction

- multiple comparisons (multiple-sample problems)
- why ANOVA?

§3.1.2 One-Factor ANOVA (Chp 10)

§3.1.3 Two-Factor ANOVA (Chp 11)

§3.1.4 Discussion

§3.1.1 Introduction

Recall ... ...

- *One-Sample Problem:*
  to estimate/assess about θ in a population distn, such as $\mu = E(X)$ or $\sigma^2 = V(X)$
- *Two-Sample Problem:*
  to compare two populations regarding their parameters, such as $\mu_1 = E(X_1)$ vs $\mu_2 = E(X_2)$ ($H_0 : \mu_1 - \mu_2 = \Delta_0$) or $\sigma_1^2 = V(X_1)$ vs $\sigma_2^2 = V(X_2)$ ($H_0 : \sigma_1^2 = \sigma_2^2$)

What is a multi-sample problem?
to compare multiple populations regarding their population parameters, such as their means: $\mu_1 = E(X_1)$ vs $\ldots$ vs $\mu_I = E(X_I)$ ($H_0 : \mu_1 = \ldots = \mu_I$)

How to conduct such a comparison?

For example, testing $H_0 : \mu_1 = \mu_2 = \mu_3$ at level α?

formulating ...

- *Hypotheses.* $H_0 : \mu_1 = \mu_2 = \mu_3$ vs H_1 : *otherwise*
- *Populations.* Independent $X_1 \sim N(\mu_1, \sigma^2)$, $X_2 \sim N(\mu_2, \sigma^2)$, $X_3 \sim N(\mu_3, \sigma^2)$
- *Data.* $\{X_{11}, \ldots, X_{1n_1}\}$, $\{X_{21}, \ldots, X_{2n_2}\}$, $\{X_{31}, \ldots, X_{3n_3}\}$ random samples from the populations, respectively.

How to test H_0 vs H_1 at level of α?

- *A natural approach.*
  - *Re-formuating ...* $H_0 \Leftrightarrow$
    $H_{01}: \mu_1 = \mu_2, H_{02}: \mu_2 = \mu_3,$ and $H_{03}: \mu_3 = \mu_1$ (*Why not* H_{01} *and* $H_{02} \Leftrightarrow H_0$*?*)
    $\Longrightarrow$ pairwise comparisons, each by a t-test, the test statistics:

$$T_1 = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S^2(1/n_1 + 1/n_2)}}, \quad T_2 = \frac{\bar{X}_2 - \bar{X}_3}{\sqrt{S^2(1/n_2 + 1/n_3)}},$$

$$T_3 = \frac{\bar{X}_3 - \bar{X}_1}{\sqrt{S^2(1/n_3 + 1/n_1)}}$$

  - to combine all the three test outcomes to conclude about H_0 and keep significance level (type I error rate) α:
    $P_{H_0}(T_1 > c_1$ or $T_2 > c_2$ or $T_3 > c_3) \leq \alpha$,thus
    $P_{H_{0j}}(T_j > c_j) = \alpha^* < \alpha$ for $j = 1, 2, 3$.

$\Longrightarrow$ "Bonferroni adjustment for multiple comparisons" $\alpha^* = \alpha/3$ (*conservative, why?*)

In general,

$I = 2 \Longrightarrow 1$ test

$I = 3 \Longrightarrow \binom{3}{2} = 3$ tests: each with $\alpha^* = \alpha/3$

$I = 4 \Longrightarrow \binom{4}{2} = 6$ tests

... ...

$I = 10 \Longrightarrow \binom{10}{2} = 45$ tests: each with $\alpha^* = \alpha/45$

... ...

impractical: computing? efficiency?

Can we test on H_0 by a single test?

$\Longrightarrow$ **One-Factor (One-Way) ANOVA!**

- to compare several populations, viewing them as classified following one factor or in one way
  e.g., young vs middle aged vs old adults in weight
  $\Longrightarrow$ **One-Factor (One-Way) ANOVA!**

- what if to compare several populations, classified in two ways?
  e.g., by gender and age
  $\Longrightarrow$ **Two-Factor (Two-Way) ANOVA**

- how about to compare several popluations, classified in multiple ways?
  e.g., by gender, age and education
  $\Longrightarrow$ **Multi-Factor (Multi-Way) ANOVA**

What will we study next?

Part 1. Introduction and Review (Chp 1-5)

Part 2. Basic Statistical Inference (Chp 6-9)

Part 3. Important Topics in Statistics (Chp 10-13)

- **3.1A One-Factor Analysis of Variance (Chp 10)**
- **3.1B Multi-Factor ANOVA (Chp 11)**
- *3.2A Simple Linear Regression Analysis (Chp 12)*
- *3.2B More on Regression (Chp 13)*

Part 4. Further Topics (Selected from Chp 14-16)