What to do today (Nov 29, 2021)?

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)

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

§4.1 Distribution-Free Procedures (Chp15.1, 15.2)

§4.1.1 Basic Concepts

§4.1.2 Nonparametric Testing Procedures

§4.1.2A Binomial Test (Sign Test)
§4.1.2B Wilcoxon's Signed-Rank Test
§4.1.2C Wilcoxon's Rank-Sum Test
§4.2.2D. Kolmogorov-Smirnov Test*

§4.2 Quality Control Methods (Chp16.1, 16.2, 16.3)

Some Logistics.

- Homework 10 has been assigned. It's due on Monday Apr 3.
- April 11 is the last day of classes in spring 2023.
- The final exam is scheduled for Apr 22 15:30.

§4.1.2B Wilcoxon's Signed-Rank Test

- **setting.** r.v. $X \sim f(\cdot)$ with a random sample $X_1, \ldots, X_n$. To test H_0 : its population median $m = m_0$ vs H_1 : otherwise.

- **test statistic.** For any i, $S_i = 1$ or 0 if $X_i - m_0 \geq 0$ or not; r_i^* is the rank statistic of $|X_i - m_0|$ in the sample of $|X_1 - m_0|, \ldots, |X_n - m_0|$: Consider $S_+ = \sum_{i=1}^n S_i r_i^*$

  (i) critical values with the distribution of S_+ under H_0 are tabulated in Table A.13 of the appendix

  (ii) if $n \gg 1$, $Z = \frac{S_+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}} \sim N(0,1)$ approximately under H_0.

- **making inference.**

  (i) Obtain the two critical values c_1 and c_2 from Table A.13: $P_{H_0}(S_+ > c_2) = \alpha/2$; $P_{H_0}(S_+ < c_1) = \alpha/2$.
  Reject H_0 if $S_{+obs} > c_2$ or $S_{+obs} < c_1$. [the exact approach]

  (ii) If $n \gg 1$, reject H_0 if $|Z_{obs}| > z_{\alpha/2}$. [the approximate approach]

Example 10.2 (p658) **to change the numerical results a bit to make it more interesting.**

- *Study.* a type of steel beam with a compressive strength $\geq$ 50K lb/in^2?
- *Data.* $n = 25$ beams (observations). (Assume they're iid.)
- *Hypotheses.* $H_0 : m = 50K$ vs $H_1 : m < 50K$
- *by the Binomial Test.*

  (i) [the exact approach]
  $S = \sum_{i=1}^{25} S_i \sim B(25, 1/2)$ under H_0. From Table A.1, the critical value c with $\alpha = 0.01$ for $P_{H_0}(S < c) = 0.01$ is between 6 and 7.
  Since $S_{obs} = 6$, $\Longrightarrow$ inconclusive.
  (ii) [the approximate approach]

$$Z = \frac{\sum_{i=1}^{25} S_i - \frac{25}{2}}{\sqrt{25/4}}$$

  approximately under H_0. $Z_{obs} = -2.6 < -z_{0.01} = -2.33$ $\Longrightarrow$ reject H_0.

- *by the Wilcoxon Signed-Rank Test.*

  (i) [the exact approach]
  $S_+ = \sum_{i=1}^{25} S_i r_i^*$ with $n = 25$ – no critical values available from Table A.13.

  (ii) [the approximate approach]

$$Z = \frac{S_+ - \frac{25(25+1)}{4}}{\sqrt{25(26)(50+1)/24}} \sim N(0,1)$$

  approximately under H_0. $Z_{obs} = -2.78 < -z_{0.01} = -2.33$
  $\Longrightarrow$ reject H_0.

§4.1.2C Wilcoxon's Rank-Sum Test

- **setting.** r.v. $X \sim F_x(\cdot)$ with a random sample $X_1, \ldots, X_g$, r.v. $Y \sim F_y(\cdot)$ with a random sample $Y_1, \ldots, Y_n$, and X and Y are independent.
  To test $H_0 : F_x(\cdot) = F_y(\cdot)$ vs H_1 : otherwise.
- *reformulating:* To test $H_0 : m_{X-Y} = 0$ vs H_1 : otherwise? (m_{X-Y} is the median of $X - Y$.)
- **test statistic.** For any X_i, let r_{X_i} be its the rank statistic in the sample of $X_1, \ldots, X_g, Y_1, \ldots, Y_n$:

$$W_{RS} = \sum_{i=1}^{g} r_{X_i}$$

  (i) critical values with the distribution of W_{RS} under H_0 are tabulated in Table A.14 of the appendix
  (ii) if $n \gg 1$,

$$Z = \frac{W_{RS} - g(g+n+1)/2}{\sqrt{gn(g+n+1)/12}} \sim N(0,1)$$

  approximately under H_0.
- **making inference.** ... ...

Remarks:

- How does the Wilcoxon rank-sum test compare to a 2-sample t-test?

  – *normal assumption is not required.*

- What if X and Y are not indpt?

  When the data are paired: $\{(X_i, Y_i) : i = 1, \ldots, n\}$, to apply the sign test, or the Wilcoxon signed-rank test to $\{D_i = X_i - Y_i : i = 1, \ldots, n\}$ with $H_0 : m_D = 0$. How does it compare with the paired t-test?

§4.2.2D. Kolmogorov-Smirnov Test*

§4.2.2D (a) One-Sample Kolmogorov Test

Suppose r.v. $X \sim F(\cdot)$ with a random sample $X_1, \ldots, X_n$.

- $H_0 : F(\cdot) = F_0(\cdot)$ vs H_1 : otherwise
- $D_1 = \sup_{-\infty < x < \infty} |\hat{F}_n(x) - F_0(x)|$
  The limiting distn under H_0 is tabulated.

§4.2.2D (b) Two-Sample Kolmogorov-Smirnov Test

Suppose r.v. $X \sim F_x(\cdot)$ with a random sample $X_1, \ldots, X_n$, and r.v. $Y \sim F_y(\cdot)$ with a random sample $Y_1, \ldots, Y_m$. Suppose X and Y are indpt.

- $H_0 : F_x(\cdot) = F_y(\cdot)$ vs H_1 : otherwise
- $D_2 = \sup_{-\infty < z < \infty} |\hat{F}_{x,n}(z) - \hat{F}_{y,m}(z)|$
  The limiting distn under H_0 is also tabulated.

What will we study/do 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)

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

4.1 Distribution-Free Procedures (Chp15.1, 15.2)

4.2 Quality Control Methods (Chp16.1, 16.2, 16.3)

4.2.1 Introduction

4.2.2 Examples of Control Charts

Review for Final Exam

Review A. Outline What Has Been Studied

Review B. Working on A Few Examples