What to do today (Apr 11, 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)
Part 4. Further Topics (Selected from Chp 14-16)

To Prepare for the Final Exam

- **Review A. Outline what has been studied**
- **Review B. Work on a few examples**

Some Logistics.

- There won't be any homework.
- Please be referred to the new office hours after today (Apr 11).
- The final exam is scheduled for Apr 22 15:30, at WMC 2532.

Final Exam Preparation

- **Office Hours during the Final Exam Period**
  - Quinn Forzley: Room K10504
    Wed April 12 - 11:30-12:30; Mon April 17 - 10:00-11:00;
    Fri April 21 - 11:00-12:00
  - Joan Hu: Room K10555
    Thu April 13 - 15:30-17:00; Thu April 20 - 15:30-17:00;
    Fri April 21 - 12:00-16:00; Sat April 22 - 12:00-14:00
- **Final Exam**
  - Time and place: Sat Apr 22 15:30, at WMC 2532
  - Material to be covered: Chp 6-13, 15-16
  - Close-book: each student is allowed to use one letter sized (A4) sheet of notes, two sided if needed, and one calculator during the final exam.
- **Time and Place to Review the Final Exam Papers**
  10:00-12:00 Fri May 12, 2022; K10555 (Please have your student ID card handy to show.)

Review A. Outline what has been studied

- Part 1. Introduction and Review (Chp 1-5)
  - basic concepts
  - distribution models and sampling distributions
- Part 2. Basic Statistical Inference (Chp 6-9)
  - point estimation
  - interval estimation
  - hypothesis testing (one-sample problem)
  - two-sample problem

Review A. Outline what has been studied

- Part 3. Important Topics in Statistics (Chp 10-13)
  - one-factor ANOVA
  - two-factor ANOVA
  - simple linear regression
  - nonlinear and multiple regression
- Part 4. Further Topics (Selected from Chp 14-16: 15.1-2; 16.1-3)
  - nonparametrics
  - control charts

Review B. Working on A Few Examples

Quiz 2. Angela Smith with SFU believed that the median of SFU student heights is 170*cm* but her friends thought the median is lower than that. One day, Angela randomly selected 7 SFU students and found out their heights, which are listed in the following table.

observation	x_1	x_2	x_3	x_4	x_5	x_6	x_7
height (cm)	177.8	163.8	152.5	157.5	152.7	167.8	184.2
$x - 170$	7.8	-6.2	-17.5	-12.5	-17.3	-2.2	14.2

- What are the rank statistics of the observations $x_1, \ldots, x_7$ in the table?
- Use two procedures with Angela's data to test the hypotheses $H_0 : m = 170$ vs $H_1 : m < 170$ at the level of significance $\alpha = .05$.

Reminder: For each of the tests, write down (i) your assumptions, (ii) test statistic and its (approximate) distribution, and (iii) your conclusion.

Review B. Working on A Few Examples

Example. Textbook Exercise Section 12.2 #29 gives the following data, in which the variables of interest are $x =$ commuting distance, and $y =$ commuting time.

Data Set	A		B		C	
	x	y	x	y	x	y
	15	42	5	16	5	8
	16	35	10	32	10	16
	17	45	15	44	15	22
	18	43	20	45	20	23
	19	49	25	63	25	31
	20	46	50	115	50	60

Question (i). Suppose rv Y is of the primary interest with the above independent observations on it.

- (i.1) Give the order and rank statistics of Y's observations in Data Set A and the sample median.
- (i.2) Assume Y's observations are $n = 18$ iid from the population.
  - Give the MM estimators of $\mu = E(Y)$ and $\sigma^2 = V(Y)$.
  - If $Y \sim \frac{1}{\theta} e^{-y/\theta}$ for $y > 0$ with $\theta > 0$, derive the MLE of θ with the given data.
  - If $Y \sim N(\mu, \sigma^2)$, give an approximate 95% CI for μ.

- (i.3) How about to test whether the observations on Y in Data Set A and the ones in Data Set B are from the same population?
  - Assume $Y_A \sim (\mu_A, \sigma_A^2)$ and $Y_B \sim (\mu_B, \sigma_B^2)$.
    - If the two sets are indpt?
      To test on $H_0 : \mu_A = \mu_B$ if $\sigma_A = \sigma_B$. or To test on $H_0 : \sigma_A = \sigma_B$ first, and then to test on $H_0 : \mu_A = \mu_B$ if $\sigma_A = \sigma_B$.
    - If not sure about whether the two sets are indpt, but the data are in pairs, $\{(y_{Ai}, y_{Bi}) : i = 1, \ldots, 6\}$?
  - If not to assume Y_A and Y_B's distributions, how to test on whether they have the same distribution?
- (i.4) How about to check whether the observations on Y in the three Data Set are from the same populations? Assume $Y_A \sim (\mu_A, \sigma^2)$, $Y_B \sim (\mu_B, \sigma^2)$, and $Y_C \sim (\mu_C, \sigma^2)$.
  - To test on $H_0 : \mu_A = \mu_B = \mu_C$ vs H_1 : otherwise by one-factor ANOVA?

- (i.5) What if the data were collected in the following way? Assume the observations are inpdt and follow normal distn.

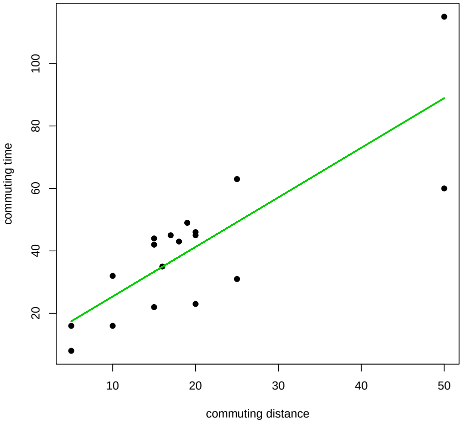
Day Type	A		B		C	
Weather	x	y	x	y	x	y
sunny	15	42	5	16	5	8
	16	35	10	32	10	16
	17	45	15	44	15	22
rainy	18	43	20	45	20	23
	19	49	25	63	25	31
	20	46	50	115	50	60

- Are commuting times associated with weather and/or type of day?
  - Modeling: $Y_{ijk} = \mu_0 + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$ with $\epsilon_{ijk} \sim N(0, \sigma^2)$ iid:
    for $i = 1, 2$ for weather type sunny, rainy, $j = 1, 2, 3$ for day-type A, B, C, and $k = 1, 2, 3$ for the 3 records of the commuting times in each of the weather-type and day-type combinations.
  - Two-factor ANOVA?

Variation Source	Df	Sum Sq	Mean Sq	F value	$Pr(> F)$
day-type	2	2058.3	1029.17	3.1150	0.08135
weather	1	2568.1	2568.06	7.7728	0.01641
interaction	2	1105.4	552.72	1.6729	0.22863
Residuals	12	3964.7	330.39		

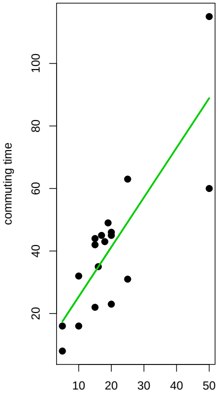
$\hat{\sigma}^2 = 330.39$ and $R^2 = 1 - SS_{res}/SS_{total} = 0.591$.

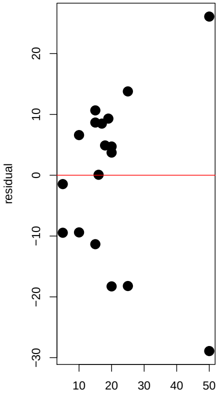
Question (ii). What if to study how Y depends on X, the commuting distance?



- the simple linear regression model: $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ for $i = 1, \ldots, 18$ indpt, and with $E(\epsilon_i) = 0$ and $V(\epsilon_i) = \sigma^2$.
- LSE: $\hat{\beta}_1 = S_{XY} / S_{XX} = 1.588$, $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{X} = 9.517$, and $\hat{\sigma}^2 = 195.8$.

- (ii) cont'd
  - Provided $\epsilon_i \sim N(0, \sigma^2)$,
    - 95% CI of β_1? $1.588 \pm 1.96 * 0.2742, \implies (1.05, 1.85)$.
    - Test on $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 > 0$ with $Z_{obs} = \hat{\beta}_1 / 0.2742 = 5.791, \implies$ strong evidence against H_0.
    - How about to estimate $E(Y|X = 60)$?
    - How about to predict for Y when $X = 60$?
  - Model checking using residuals $e_i = y_i - \hat{y}_i$ for $i = 1, \ldots, 18$? $R^2 = 0.657$
  - Althernative models? ... ...





Good luck on your final exam.

Thank you much for your participation in STAT285!