Review 2

Test 1 topics

- Descriptive statistics
  - e.g. stem-and-leaf plot, mean, median, mode…
- Probability distribution
  - e.g. discrete vs. continuous probability distribution…
- Hypothesis testing: one-sample inference
  - e.g. p-value, null hypothesis, alternate hypothesis…
- Hypothesis testing: two-sample inference
  - e.g. paired samples hypothesis testing, independent samples hypothesis testing (F-test: check equal variances, t-test)…

Hypothesis testing: categorical data

- **Contingency Table Approach**
  - **Expected number of units** in the (***i,j***) cell (***E_{ij}*):**

    ith row margin X ***jth*** column margin

    grand total
  - **none of the four expected values < 5**
- **Fisher's exact test**
  - **1 of the cells with expected values <=5**
- **McNemar's Test**

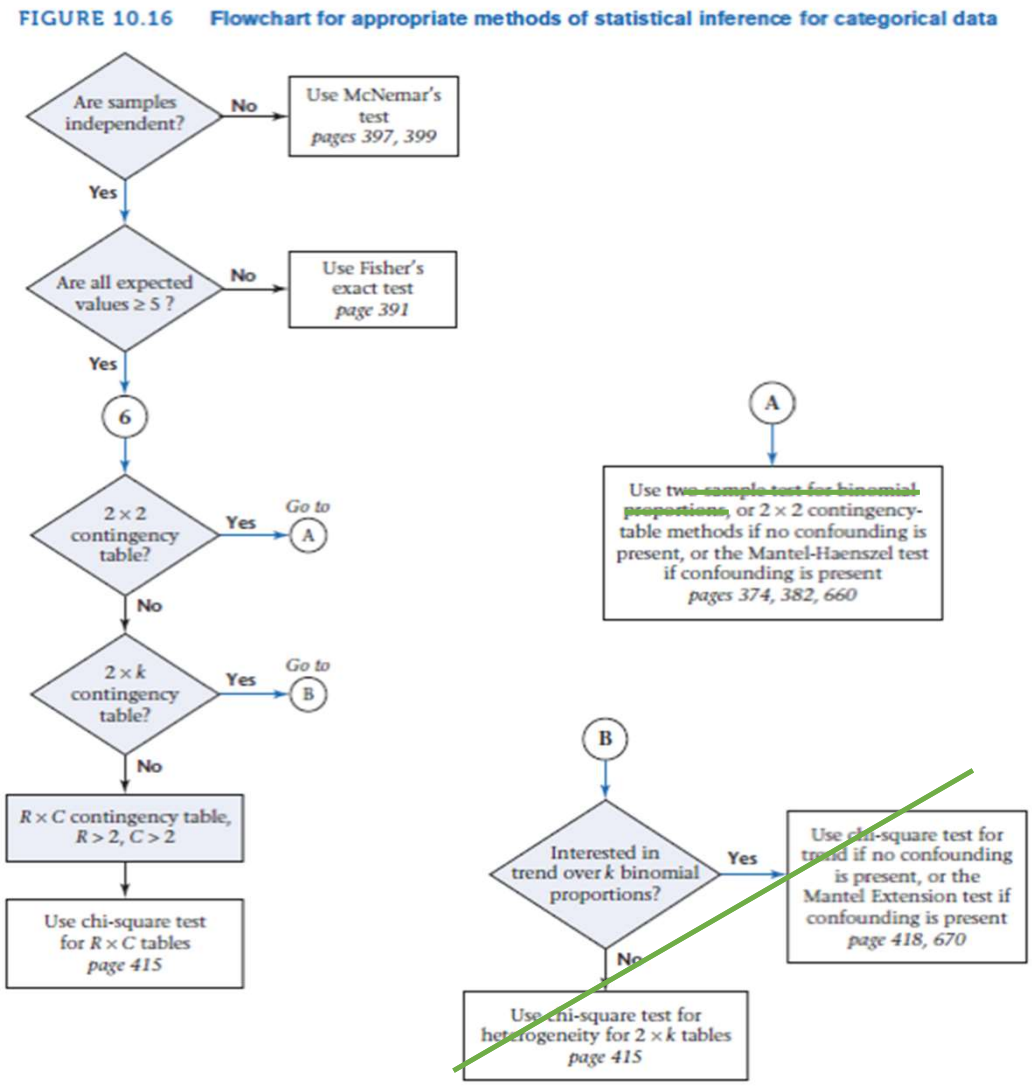
i) Normal Theory test ($n_D \geq 20$)

ii) Exact Method ($n_D < 20$)

- **RxC Contingency Table**
  - **Test statistic:**

$$X^2 = (O_{11} - E_{11})^2 / E_{11} + (O_{12} - E_{12})^2 / E_{12} + \cdots + (O_{RC} - E_{RC})^2 / E_{RC}$$

  - $H_0 \sim \chi^2$ distribution with $(R - 1) \times (C - 1)$ *df*



Regression and Correlation

- **Interpretation of regression line**
- **Correlation (Pearson's vs. Spearman ranks)**
- **Hypothesis testing for multiple regression**
  - **F test:**

    H_0: $\beta_1 = \beta_2 = \ldots = \beta_k = 0$

    vs. H_1: at least one of the $\beta_j \neq 0$ in multiple linear regression

$$\text{Res SS} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

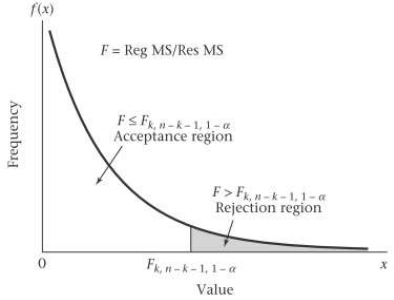
$$\text{Reg SS} = \text{Total SS} - \text{Res SS}$$

$$\text{Total SS} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$\hat{y}_i = a + \sum_{j=1}^{k} b_j x_{ij}$$

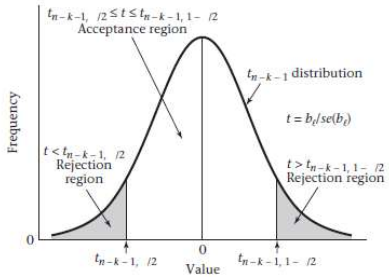
Test statistic:

F = Reg MS/Res MS, df = n-k-1 where n=sample size, k =no of independent



  - **T test:**

    H_0: $\beta_l = 0$, All other $\beta_j \neq 0$ vs. H_1: $\beta_l \neq 0$, all other $\beta_j \neq 0$ in multiple linear regression



- *Statistical output for multiple regression model*

Nonparametric Methods

- Parametric Methods: data of known distribution
- Non-parametric methods: data of unknown distribution, skewed / not normally distributed, ordinal

Analysis Type	Example	Parametric Procedure	Nonparametric Procedure
Compare means between two distinct/independent groups	Is the mean systolic blood pressure (at baseline) for patients assigned to placebo different from the mean for patients assigned to the treatment group?	Two-sample t-test	Wilcoxon rank-sum test
Compare two quantitative measurements taken from the same individual	Was there a significant change in systolic blood pressure between baseline and the six-month follow-up measurement in the treatment group?	Paired t-test	Wilcoxon signed-rank test
Estimate the degree of association between two quantitative variables	Is systolic blood pressure associated with the patient's age?	Pearson coefficient of correlation	Spearman's rank correlation