Quantitative Methods 2

Tutorial 7

Nhan La

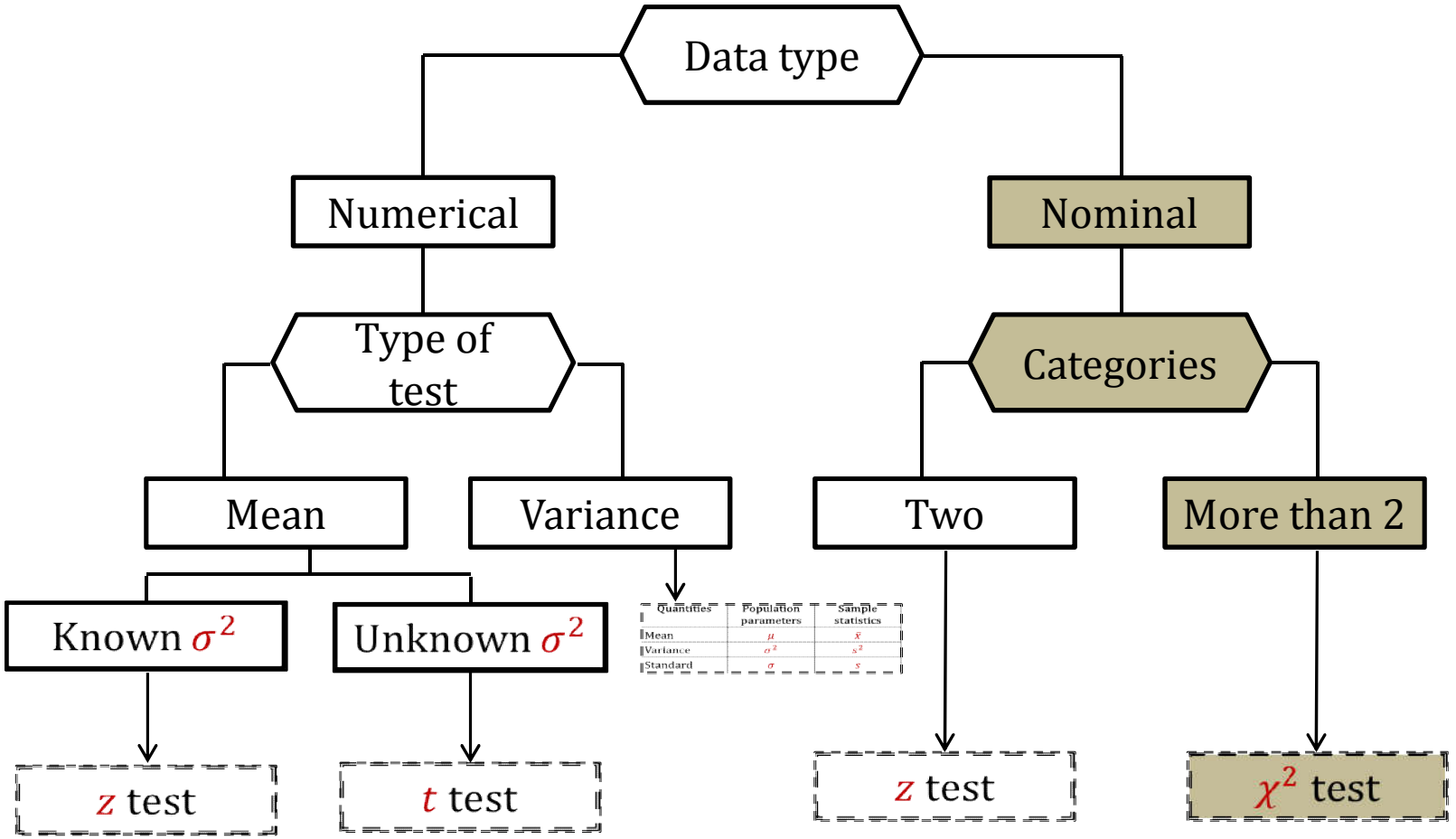
Assignment 1

1. Normality JB test
2. Manual calculation
3. Hypothesis testing
   - Rejection rule: critical value and/or p-value
   - Decision and conclusion

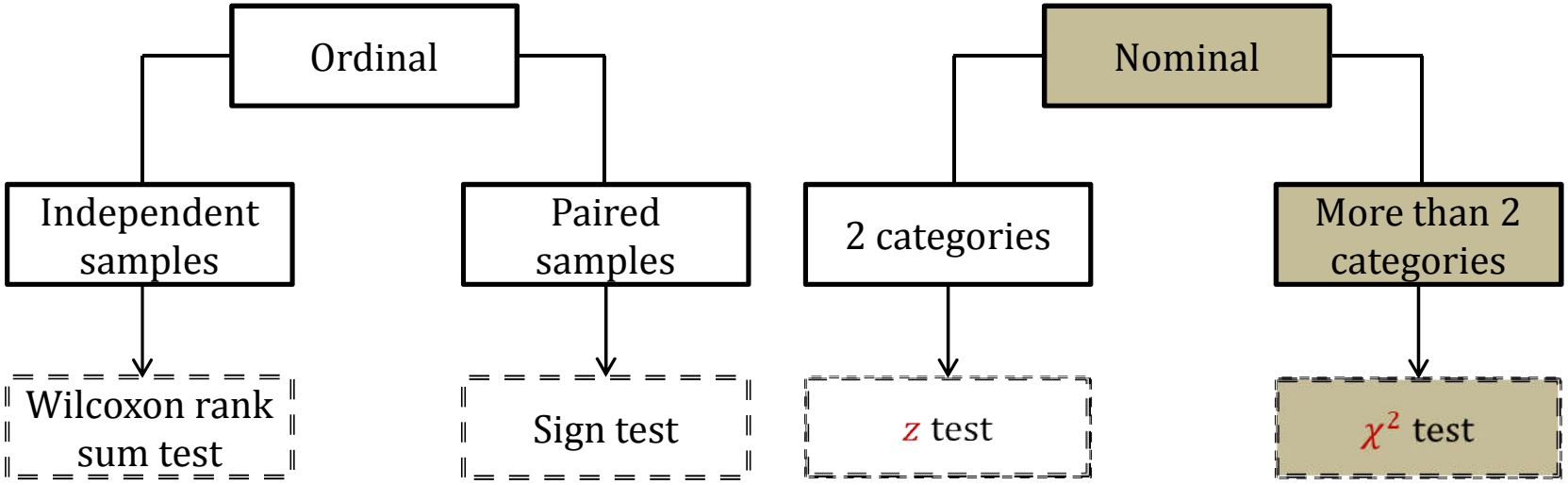
Tutorial 7

1. Goodness of fit test
2. Independence test
3. Homogeneity test
4. Correlation test
   - Pearson correlation
   - Spearman rank correlation

One population



Two populations - Ordinal and Nominal



Goodness of fit test

- Test for the distribution of the sampled population
- Multinomial distributions

$e_i = np_i$; $\sum_{i=1}^{k} e_i = n$

$\hat{p}_i = \frac{o_i}{n}$; $\sum_{i=1}^{k} \hat{p}_i = 1$

$$\chi^2 = n\sum_{i=1}^{k} \frac{(\hat{p}_i - p_i)^2}{p_i} = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} = \left(\sum_{i=1}^{k} \frac{o_i^2}{e_i} - n\right) \sim \chi^2_{k-m}$$

k: number of categories/groups

m: number of constraints on the data + number of coefficients to estimate the expected frequencies.

  - Often times $m = 1$, hence $\chi^2 \sim \chi^2_{k-1}$
  - Constraint: $\sum_{i=1}^{k} e_i = n$; or $\sum_{i=1}^{k} \hat{p}_i = 1$

- Condition: $e_i \geq 5$

Goodness of fit test

Exercise 1

Face	$p_{i,0}$		
1	0.1667	114	100
2	0.1667	92	100
3	0.1667	84	100
4	0.1667	101	100
5	0.1667	107	100
6	0.1667	102	100

Independence vs. homogeneity

- Independence:
  - Random sample is drawn from one population
  - Categorical series/characteristics are used to classify the sample
  - Null hypothesis: Two series are independent, which make their distributions (by categories) independent
- Homogeneity:
  - Random samples are drawn from each population
  - Null hypothesis: Categories, or class intervals, of a qualitative variable follow the same distribution in these populations

Independence vs. homogeneity

- Independence

In a 2008 research study across Melbourne 448 grocery shoppers exiting four of Australia's biggest supermarket retailers were interviewed regarding their views on green issues…

- Homogeneity

Whether there is a difference between the uses of online check-in at three of its ports, Sydney, Singapore and Jakarta…

Independence vs. homogeneity

- Independence

$$e_{lk} = \frac{1}{n}\sum_{j=1}^{c} o_{lj} \times \sum_{i=1}^{r} o_{ik}$$

- Homogeneity

$$e_{lk} = p_{ik} \times \sum_{j=1}^{c} o_{lj} = \frac{1}{n}\sum_{j=1}^{c} o_{lj} \times \sum_{i=1}^{r} o_{ik}$$

$$\chi^2 = \sum_{i-1}^{r}\sum_{j=1}^{c} \frac{(o_{ij}-e_{ij})^2}{e_{ij}} = \sum_{i-1}^{r}\sum_{j=1}^{c} \frac{o_i^2}{e_i} - n \sim \chi^2_{df}$$

$$df = (r-1)(c-1)$$

- Condition: $e_i \geq 5$

Corrections

- Exercise 1 – Page 3: expected probabilities must add up to one

- Exercise 2(b) – Page 6: *aussie* = 2