

Midterm results

Stat 576

11-28-17

$n = 24$ $\bar{x} = 26.92$ out of 36
(74.8%)

median = 27.5 (76.4%)

1x	56
2	1 2 3 3
2x	5 6 6 7 7 7 8 8 8 9 9
3	0 0 2
3x	5 5 6

Imputation Methods

① Deductive: use existing info to deduce what the missing value must have been

② Cell mean: Say you have several categorical

Variable:
ethnicity

	1	2	3
Gender M	y ₁ , y ₂ , y ₃	y ₄ , y ₅	y ₆ , y ₇
F	y ₈ , y ₉ , y ₁₀	y ₁₁ , y ₁₂	y ₁₃ , y ₁₄

Suppose y₁₀ is missing

The imputed value is the average of the nonmissing values in that cell

③ Sequential: Same as ②, but the most recently observed value in that cell is used as the imputed value.

④ Random: Similar to ②, but you randomly select an observed value in that cell, to use as your imputed value

⑤ Nearest neighbor As in cluster analysis, you would create a distance measure for each pair of observations. For the observation with the missing value, copy that value from its nearest neighbor

⑥ Regression Predict the missing value from the other variables, using multiple regression or logistic regression

For methods ② thru ⑤, you could add a stochastic adjustment to the imputed value

For example, use Excel to generate a $U(0,1)$ observation.

Take $\Phi^{-1}(u)$ to get a standard normal observation
↑ inverse of cumulative standard normal distribution

Multiply that by the desired σ to
get an observation from $N(0, \sigma^2)$

(5)

Example of the regression method,
using the judging grid
on the next page:

(6)

	Judges			Avg before imp	Rank before imp	imp. Avg.	
	A	B	C				
1	8	7.2	(6.95)	7.6	(3)	7.383	(3)
2	7.1	(7.275)	6.8	6.95	(4)	7.058	(4)
3	(9.075)	9.1	8.4	8.75	(2)	8.858	(1)
4	9.2	8.7	(8.3)	8.95	(1)	8.733	(2)
5	7.3	(7.225)	6.5	6.9	(5)	7.008	(5)
6	(6.625)	6.6	6.0	6.3	(6)	6.408	(6)
	7.9	7.9	6.925	7.575			

Model $y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$

Using L.S., we can show that

(7)

$$\hat{\mu} = \bar{y}_{..}$$

$$\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}$$

$$\hat{\beta}_j = \bar{y}_{.j} - \bar{y}_{..}$$

$$\begin{aligned}\hat{y}_{ij} &= \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..}) \\ &= \bar{y}_{i.} + \bar{y}_{.j} - \bar{y}_{..}\end{aligned}$$

Survey Design

(8)

- 1 Decide what type of information is needed. If a hypothesis test is to be done, state the null & alternative hypotheses in advance
- 2 Test questions
Either use questions from an established survey, or put together an expert panel to evaluate your questions
- 3 Make the questions simple & clear

- 4 Make the questions as specific as possible, avoiding generalities
- 5 When reporting the answers, let the reader see the exact wording of the question
- 6 Avoid leading questions
- 7 Consider the number of choices in each answer

Likert scale: 5 to 7 ranked choices

1 2 3 4 5

- 8 Pay attention to question order + order of answers