Slides PDF 9/3/14

Sherri Verdugo

September 3, 2014

Contents

Administration Items Lecture 3: 9/3/14

- Office Hours: New
- Schedule coming out next week for all items this semester
- Recap: Textbook Chapter One was discussed last week
- New: Chapter 2 Introduction

New Office Hours

Instructor	Sherri Ann Verdugo, M.S.
Soc 303–03 15668	Room: H512
Day & Time	Mon. & Wed. 2:30 – 3:45PM
Office: CP-923	Hours: Wednesday 1:00 - 2:00PM
Day & Time	Hours: Tuesday 10:00 - 11:00 AM
Phone	657.278.6728
Soc 303–85 16584	Room: IRVC-203
Day & Time	Friday 1:00 – 3:45PM
Office:	Hours: Friday 11:00 - 12:00PM
Google Voice:	562.444.5036

Key Concepts

Term	Pages	Term	Pages
measurement	34	frequencies	45
levels of measurement: scales	35	scores	45
qualitative: Categorical	35	interval level of measurement	46
quantitative: Numerical	35	ratio level of measurement	46
nominal level of measurement	36	absolute zero	47
attributes versus quantities	36	demographic variables	47
individual nominal data	38	individual interval data/ raw score	48
grouped Nominal data	39	ungrouped frequency distribution	49
frequency distributions	39	grouped interval data	51
dichotomies	39	class interval	51
n-category	40	closed-ended intervals	52
ordinal level of measurement	40	open-ended intervals	52
individual ordinal data	40	likert scale	43
grouped ordinal data	41		

Prologue

We are investigating data. . . numerical and sometimes not numerical (outside of the scope for this class).

- Data: information specifically intended to assist decision making or to provide analysis (page 33, textbook).
- Pro tip: numerical data is something that gives us insight into the sociological and social science perspectives we want to study.

Measurement:

Definition:

A very specific process, such as measuring length, or the assignment to a category. (page 34)

N(nominal)**O**(ordinal)**I**(interval)**R**(atio)

Levels:

Measurement can be classified as the four levels of measurement: nominal, ordinal, interval and ratio (noir).

Qualitative

Data assigned to categories that do not imply amounts. (nominal, ordinal)

Quantitative

Data assigned to categories that are involved in amounts. (interval, ratio)

Example

```
library(vcdExtra)
```

```
## Loading required package: vcd
## Loading required package: grid
## Loading required package: gnm
```

```
library(xtable)
data(ICU)
icu.sub <- ICU[, 1:13]
print(xtable(head(icu.sub, 10)))
```

% latex table generated in R 3.1.0 by xtable 1.7-3 package % Wed Sep 3 08:35:48 2014

	died	age	sex	race	service	cancer	renal	infect	cpr	systolic	hrtrate	previcu	admit
8	No	27	Female	White	Medical	No	No	Yes	No	142	88	No	Emergency
12	No	59	Male	White	Medical	No	No	No	No	112	80	Yes	Emergency
14	No	77	Male	White	Surgical	No	No	No	No	100	70	No	Elective
28	No	54	Male	White	Medical	No	No	Yes	No	142	103	No	Emergency
32	No	87	Female	White	Surgical	No	No	Yes	No	110	154	Yes	Emergency
38	No	69	Male	White	Medical	No	No	Yes	No	110	132	No	Emergency
40	No	63	Male	White	Surgical	No	No	No	No	104	66	No	Elective
41	No	30	Female	White	Medical	No	No	No	No	144	110	No	Emergency
42	No	35	Male	Black	Medical	No	No	No	No	108	60	No	Emergency
50	No	70	Female	White	Surgical	Yes	No	No	No	138	103	No	Elective

In this dataset, we have mixed types of variables. In statistics, we are going to learn how to handle this type of information like a pro.

Often, we find ourselves working with mixed data in the social sciences because we are looking at things from the social investigator's point of view.

Qualitative Data

Nominal Levels of Data

Nominal simply means naming. It is a system we use when we classify things. In the last example, you could see the gender and race column as examples of nominal levels. We can see them as individuals or grouped data.

- We can group this data and see frequencies of the data
- Attributes: these are characteristics

Ordinal Levels of Data

Ordinal involves a ranking or ordering of data. For example, if we are looking at class designations (i.e. freshman, sophomore, junior, senior, or grad student)... we can order them in terms of time in school and levels of expertise. Further, we can see them as individuals or grouped data.

- We can group this data and see frequencies of the data
- Attributes: ranked characteristics

Frequency Distribution of Nominal Data

- Dichotomies: two category variable (vote status: yes/no, disease status: infected/not infected)
- n-category: more than two categories (like in our example)
- Frequency distribution: tabulation that lists the variable, it's categories, and a frequency column. Add the numbers together to get the total Σ.
- Grouped nominal data: data that are presented as a category of the variable listed, and the subjects are not named but are counted in the category that each subject falls into.

Political Status	$f =$
Democrat	100
Republican	100
Independent/Other	200
Total n	$\Sigma = 400$

Frequency Distribution of Ordinal Data

- n-category: more than two categories
- Frequency distribution: tabulation that lists the variable, categories, and frequency column.
- Scores: numbers that are used to represent amounts of rankings.
- Frequencies: headcounts or tallies for the number of cases in a particular category.
- Grouped ordinal: presented in ranked categories from highest to lowest or lowest to highest.

Fear of Heights	$f =$
Extremely Mild	100
Mild	100
Neutral	200
High	50
Extremely High	50
Total n	$\Sigma = 500$

Likert Scale

Likert Scale: a scale whose categories are based on the level of agreement with a particular statement or issue.

We are interested in seeing how viewers respond to a new movie. Please select your emotional response for fear using the following question based on the previous commercial.

I felt extreme fear after watching the previous commercial.

1. Agree
2. Somewhat Agree
3. Neutral
4. Somewhat Disagree
5. Disagree

Hint: Select any response above.

Explanation: *All answers are correct.*

- We can measure this information statistically in the future.

Car Data

- contains quantitative data

```
print(xtable(head(mtcars, 10)))
```

% latex table generated in R 3.1.0 by xtable 1.7-3 package % Wed Sep 3 08:35:48 2014

This type of data is mixed, as was the previous version. It contains both qualitative and quantitative data. Can you spot the different types of variables in the presented data?

Qualitative Data

- Interval Levels of Data
- the subject receives a numerical score

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.00	6.00	160.00	110.00	3.90	2.62	16.46	0.00	1.00	4.00	4.00
Mazda RX4 Wag	21.00	6.00	160.00	110.00	3.90	2.88	17.02	0.00	1.00	4.00	4.00
Datsun 710	22.80	4.00	108.00	93.00	3.85	2.32	18.61	1.00	1.00	4.00	1.00
Hornet 4 Drive	21.40	6.00	258.00	110.00	3.08	3.21	19.44	1.00	0.00	3.00	1.00
Hornet Sportabout	18.70	8.00	360.00	175.00	3.15	3.44	17.02	0.00	0.00	3.00	2.00
Valiant	18.10	6.00	225.00	105.00	2.76	3.46	20.22	1.00	0.00	3.00	1.00
Duster 360	14.30	8.00	360.00	245.00	3.21	3.57	15.84	0.00	0.00	3.00	4.00
Merc 240D	24.40	4.00	146.70	62.00	3.69	3.19	20.00	1.00	0.00	4.00	2.00
Merc 230	22.80	4.00	140.80	95.00	3.92	3.15	22.90	1.00	0.00	4.00	2.00
Merc 280	19.20	6.00	167.60	123.00	3.92	3.44	18.30	1.00	0.00	4.00	4.00

- can be positive or negative
- raw score: individual listing
- example: temperature
- Ratio Levels of Data
- similar to interval
- absolute zero: zero is absolute
- can not be negative
- famous example: Absolute Zero, Kelvin Scale

A bit more about data

- Demographic variables: background information on the human subjects studied.
- Ungrouped frequency distribution: every scores listed in a sequence (usually highest to lowest)
- Grouped interval data: interval
- Class interval: interval that indicates the space between two end points
- Closed ended: each class interval must have an upper and lower limit

```
icu.sum <- summary(ICU[,1:6])
print(xtable(icu.sum))
```

% latex table generated in R 3.1.0 by xtable 1.7-3 package % Wed Sep 3 08:35:48 2014

	died	age	sex	race	service	cancer
1	No :160	Min. :16.0	Female: 76	Black: 15	Medical : 93	No :180
2	Yes: 40	1st Qu.:46.8	Male :124	Other: 10	Surgical:107	Yes: 20
3		Median :63.0		White:175		
4		Mean :57.5				
5		3rd Qu.:72.0				
6		Max. :92.0				

Scores vs. Frequencies

- Open ended: class interval that has a lower limit but no upper limit or vice versa
- Closed ended: class interval that has a lower and upper limit
- f = frequency

Example of kids in daycare

Age Range	f
0 to 5	5
5.5 to 10.5	10
———	—-
n	15

Exercises: Your Turn

Class Assignment: Work in groups to perform the following tasks. Send an email with all names attached to the email regarding the assignment labeled with today's date.

- identify at least 5 variable examples for each type of measurement
- nominal
- ordinal
- interval
- ratio
- what subject would you expect to find the different types of variables in data
- nominal
- ordinal
- interval
- ratio

Next week:

- Assignment Schedule
- Chapter 3 and 4
- Homework 1
- Writing Assignment Proposal Discussion