Chapter 2

Describing Variables

2.4   Measures of Central Tendency

Central Tendency Measures

Measures of central tendency summarize the scores of a frequency distribution into a single number – the average or value most typical of all the scores.

How familiar are you with such "average values" as:

- College GPA

- Baseball batting averages

- Dow-Jones industry stock market index

- Unemployment rate

In statistics, three commonly used central tendency measures are the mode, median, and mean, which can be calculated for grouped and ungrouped data.

Mode for Ungrouped Data

MODE: the single category among the *K* categories in a distribution with the largest number (or highest percentage) of observations

Find the mode of these 12 scores
(**HINT:** Rewrite them in order from lowest to highest):

$Y_1 = 4$ $Y_7 = 6$

$Y_2 = 3$ $Y_8 = 7$

$Y_3 = 8$ $Y_9 = 2$

$Y_4 = 6$ $Y_{10} = 6$

$Y_5 = 5$ $Y_{11} = 4$

Mode = _____

$Y_6 = 4$ $Y_{12} = 6$

The Mode for Grouped Data

One type of grouped frequency distribution shows the lower and upper limits of collapsed ordered scores. In this case, the grouped data mode is the midpoint of the category with the largest number of responses (i.e., highest percentage). Calculate midpoint by summing the limits and dividing by 2.

Find the modal category for AGE20, an ordered discrete variable created by recoding AGE into four 20-year intervals:

AGE	Frequency	Percent
20-39	695	35.2
40-59	752	38.1
60-79	437	22.0
80-99	92	4.7
Total	1976	100.0

Modal value = _____

The Mode For Grouped Data

Another type of grouped data table displays categories without collapsed values. Its mode is the score or label of the category with the most cases (i.e., largest percentage).

Find the modal category for REGION4, a nonordered discrete variable with K = 4 categories:

		Frequency	Percent
Valid	1 NORTH EAST	355	17.5
	2 MIDWEST	435	21.5
	3 SOUTH	731	36.1
	4 WEST	502	24.8
	Total	2023	100.0

Modal region = _____

The Median for Ungrouped Data

MEDIAN (Mdn): the outcome dividing an <u>ordered</u> frequency distribution exactly into halves

CAUTION: Don't compute a median for a nonordered variable!

Find the median of these N = 11 scores:

$Y_1 = 2$ $Y_7 = 19$

$Y_2 = 17$ $Y_8 = 3$

$Y_3 = 17$ $Y_9 = 4$

$Y_4 = 6$ $Y_{10} = 9$

$Y_5 = 15$ $Y_{11} = 6$

$Y_6 = 28$ Mdn = _____

The Median for Grouped Data

Find the median for an ordered variable grouped frequency distribution by examining the cumulative percentages of a table. Report label, score, or midpoint of the category whose cumulative percentage <u>exceeds</u> 50.0% (the 50th percentile).

SEXFREQ	"About how often did you have sex during the last 12 months?"	Frequency	Valid Percent	Cumulative Percent
Valid	0 NOT AT ALL	416	24.7	24.7
	1 ONCE OR TWICE	149	8.8	33.5
	2 ONCE A MONTH	176	10.4	44.0
	3 2-3 TIMES A MONTH	243	14.4	58.4
	4 WEEKLY	285	16.9	75.3
	5 2-3 PER WEEK	309	18.3	93.6
	6 4+ PER WEEK	108	6.4	100.0
	Total	1686	100.0	

Mdn = _____

Mean for Ungrouped Data

MEAN: the arithmetic average of a continuous variable's scores

$$\overline{Y} = \sum_{i=1}^{N} \frac{Y_i}{N}$$

- Variable Y has N total observations (cases)
- The ith case is denoted subscript i
- The ith case's score is Y_i
- Σ: add <u>all</u> the Y_i scores, from $i = 1^{st}$ to N^{th}
- Divide this sum of all the scores by N

Find the mean of these scores:

$Y_1 = 5$

$Y_2 = 3$

$Y_3 = 8$

$Y_4 = 6$ $N =$ _____

$Y_5 = 5$

$\Sigma Y_i =$ _____

$Y_6 = 4$

$$\overline{Y} = \sum_{i=1}^{N} \frac{Y_i}{N} = \underline{\hspace{4cm}}$$

Mean for Grouped Continuous Data

For continuous variables in a grouped frequency distribution, used this formula to calculate a <u>weighted</u> mean for the grouped data:

$$\overline{Y} = \sum_{i=1}^{K} \frac{(f_i Y_i)}{N}$$

K is the number of categories

f_i is the frequency (count) of cases in the ith category

Y_i is the score of the ith category (or its midpoint)

The 2008 GSS asked, "About how often did you have sex during the last 12 months?" Response categories allow SEXFREQ to be treated as a continuous variable.

To find the mean, first recode the original six GSS scores into an approximate number of times per year. (A year has 52 weeks and 12 months; for "3+ per week," I assumed a value of "4 per week")

SEXFREQ	Frequency	Valid Percent	Cumulative Percent
0 NOT AT ALL = 0	416	24.7	24.7
1 ONCE OR TWICE = 2	149	8.8	33.5
2 ONCE A MONTH = 12	176	10.4	44.4
3 2-3 TIMES A MONTH = 36	243	14.4	58.4
4 WEEKLY = 52	285	16.9	75.3
5 2-3 PER WEEK = 156	309	18.3	93.6
6 3+ PER WEEK = 208	108	6.4	100.0
Total	1686	100.0	

Next, use those new Y scores to calculate SEXFREQ mean:

Category	f_i	Y_i	f_iY_i
NOT AT ALL	416	0	_____
ONCE OR TWICE	149	2	_____
ONCE A MONTH	176	12	_____
2-3 TIMES A MONTH	243	36	_____
WEEKLY	285	52	_____
2-3 PER WEEK	309	156	_____
4+ PER WEEK	108	208	_____
Total (N)	Σf_i = **1,686**	$\Sigma (f_iY_i)$ =	_____

$$\overline{Y} = \sum_{i=1}^{K} \frac{(f_iY_i)}{N} = \underline{\hspace{4cm}}$$

Grouped Age Mean

For this table, which has the lower and upper limits of grouped scores, first calculate each interval's midpoint. Then, apply the formula for the mean of grouped data.

AGE	Midpoint	Frequency
20-39	29.5	695
40-59	49.5	752
60-79	69.5	437
80-99	89.5	92
Total		1976

$$\overline{Y} = \sum_{i=1}^{K} \frac{(f_i Y_i)}{N} = $$

Mean of a Dichotomy

You can also apply the grouped data formula to find the mean of a dichotomous variable (coded 0-1), which simplifies to:

$$\overline{Y} = \frac{\Sigma(f_i Y_i)}{N} = \frac{(f_0)(0) + (f_1)(1)}{N} = \frac{f_1}{N}$$

where (f_0 / N) equals the proportion (p_0) of cases with Y_i coded = 0
and (f_1 / N) equals the proportion (p_1) of cases with Y_i coded = 1

Thus, the mean of a dichotomy is simply the proportion of cases in one of the two categories.

Calculate the Mean of POSTLIFE

The 2008 GSS asked, "Do you believe there is a life after death?"

0 NO	332
1 YES	1,455
Total	1,787

$$\overline{Y} = \frac{(f_0)(0) + (f_1)(1)}{N} = \frac{f_1}{N}$$

$$= \frac{(332)(0) + (1,455)(1)}{1,787} = \underline{\hspace{4cm}}$$

What percentage of the 2008 GSS respondents believe in an afterlife?