Tools for looking at DISCRETE variable values

1. Relative Frequencies:

$$\frac{f(x)}{n} = \frac{\text{number of cases with the value of x}}{\text{total number of cases}}$$

2. **Pie Charts** show relative frequencies for one outcome
With mutually exclusive and exhaustive values

3. **Bar Charts** show relative frequencies
Can be one outcome or more:

- Same outcome, multiple values (mutually exclusive and exhaustive)
- Same outcome over time or different sub-samples (so not mutually exclusive)
- Multiple outcomes for same sample (so not mutually exclusive)

4. **Tables for discrete variables** show frequencies or relative frequencies

Statistics

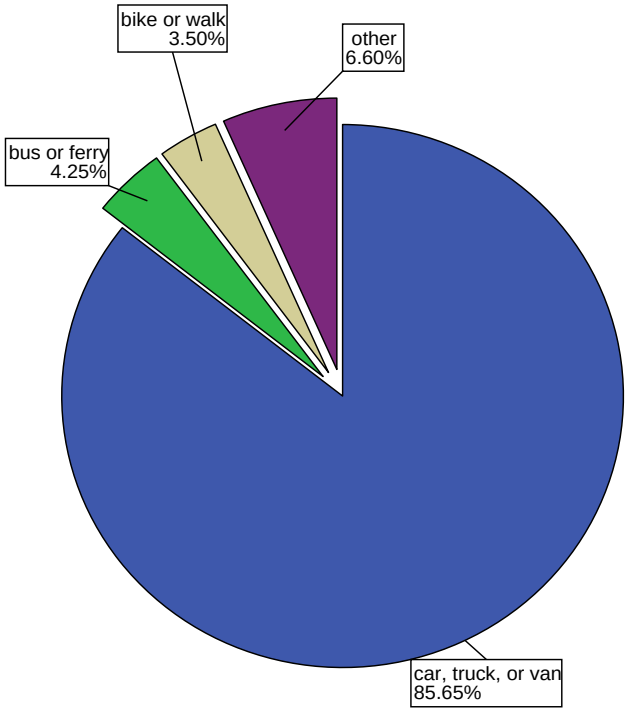
USUAL WAY GETTING TO WORK/SCHOOL LAST WK

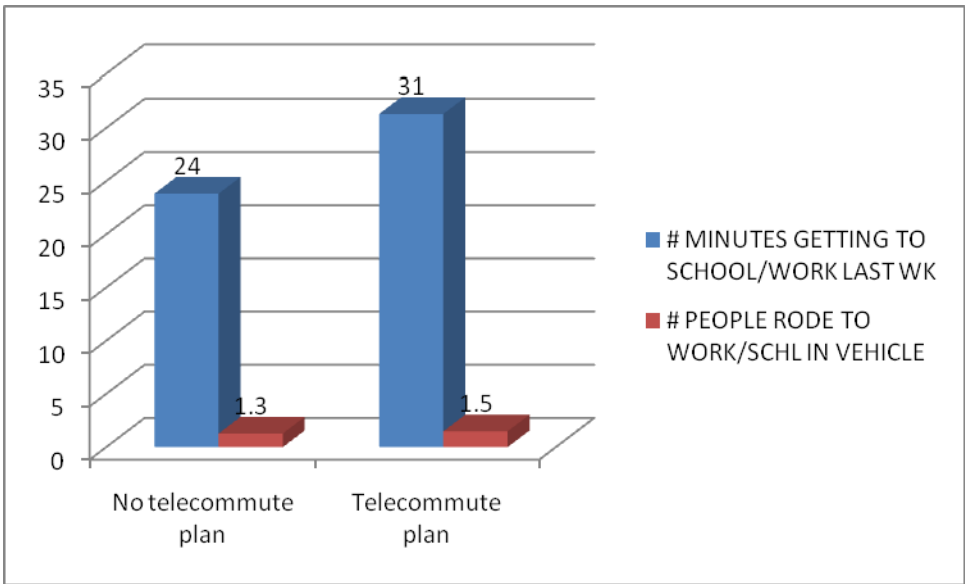
N	Valid	2801
	Missing	9031

USUAL WAY GETTING TO WORK/SCHOOL LAST WK

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.CAR	1922	16.2	68.6	68.6
	2.TRUCK	355	3.0	12.7	81.3
	3.VAN	122	1.0	4.4	85.6
	4.BUS	117	1.0	4.2	89.8
	5.FERRY	2	.0	.1	89.9
	7.MOTORCYCLE	7	.1	.2	90.1
	8.BICYCLE	17	.1	.6	90.8
	9.WALKED	81	.7	2.9	93.6
	10.WRK/STUDY HOME	107	.9	3.8	97.5
	11.EQUAL COMB MODES	36	.3	1.3	98.8
	12.OTHER (SPECIFY)	2	.0	.1	98.8
	14.AIRPLANE (+)	26	.2	.9	99.8
	15.TRAIN (+)	3	.0	.1	99.9
	16.CARPOOL/PUBLIC TRANSPORT (+)	4	.0	.1	100.0
	Total	2801	23.7	100.0	
Missing	System	9031	76.3		
Total		11832	100.0		

Usual way of getting to work or school





Tools for looking at CONTINUOUS variable values

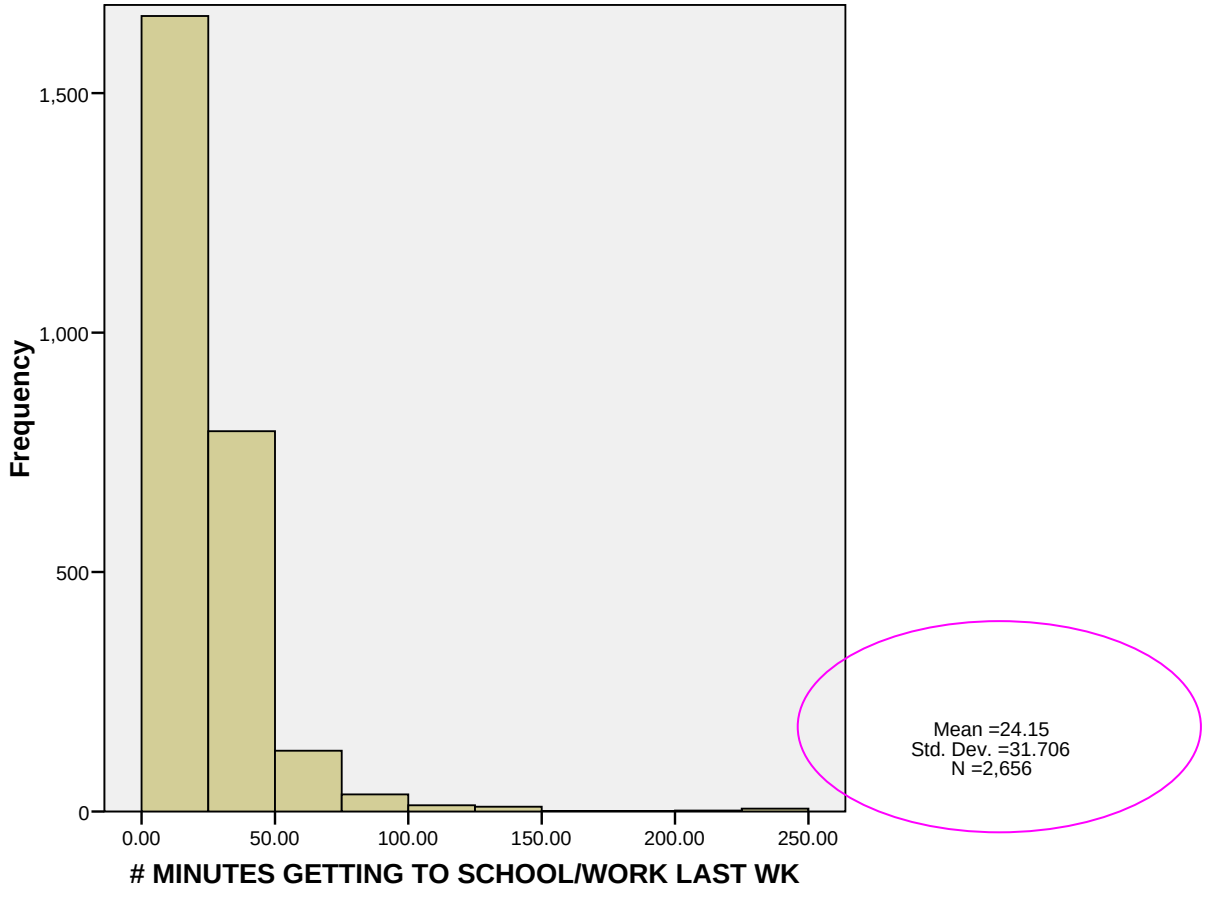
5. Histograms

Picture of frequencies or relative frequencies

- data are grouped into intervals (usually of equal interval size)
- each column shows the frequency or the relative frequency of the values

Histograms tell us about the variable value distribution:

- Shape—peaked, uniformly flat, or messy!
- Symmetric—one side of the distribution mirrors the other
- Center (rough location)
- Outliers
- Skewness



6. Measures of Central Tendency

a) Mean --Arithmetic average

Sample Mean

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \left(x_1 + x_2 + x_i .... + x_n\right)/n$$

where *n* is the number in the sample and *x* is the variable of interest.

Symbols:

$\bar{x}$ = mean for a sample ("x bar")

μ = mean for a population (also called expected value) ("mu")

n = sample size

N= population size (the denominator for a population mean)

b) Median

The middle value—50th percentile. Half of cases are above this and half are below.

c) Mode(s)

The most frequent value(s).

Statistics

MINUTES GETTING TO SCHOOL/WORK LAST WI

N	Valid	2656
	Missing	9176
Mean		24.1476
Std. Error of Mean		.61521
Median		20.0000
Mode		20.00
Std. Deviation		31.70582
Variance		1005.259
Range		999.00
Minimum		.00
Maximum		999.00
Percentiles	25	10.0000
	50	20.0000
	75	30.0000

7. Measures of Variability (spread)

a) Range

Distance between the smallest and the largest values. Check for "outliers"

b) Variance

The average of the squared deviations from the mean. (Symbols: s^2 for sample, and σ^2 for a population.)

Sample Variance

$$s^2 = \frac{\sum_{i=1}^{n} \left(x_i - \bar{x} \right)^2}{n-1}$$

where n=sample size and
x=the variable of interest

c) Standard Deviation

(Symbols: s (for a sample), σ (for a population))
How far, on average, an observation is from mean.

Sample Standard Deviation

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} \left(x - \bar{x} \right)^2}$$

SPSS Descriptive Statistics Output:

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
# MINUTES GETTING TO SCHOOL/WORK LAST WK	2656	.00	999.00	24.1476	31.70582
PARTICIPATE IN TELECOMMUTE PLAN	2485	.00	1.00	.0555	.22906
NUMBER OF DAYS TELECOMMUTED LAST WEEK	137	.00	7.00	2.2482	2.13094
# PEOPLE RODE TO WORK/SCHL IN VEHICLE	2373	1.00	50.00	1.2971	1.33344
USUAL WAY GETTING TO WORK/SCHOOL LAST WK	2801	1.00	16.00	2.2681	2.80838
Valid N (listwise)	95				

8. Measures of Relative Standing

a) Percentiles

The *pth* percentile is the value at which *p* percent of the values (when ordered from smallest to largest) are below this number.

b) z-scores

This measure of relative standing tells how far a particular value is from the mean—in standard deviation units.

- If z is positive, then the value is above the mean
- If z is negative, then the value is less than the mean.
- If z is 0 or near 0, then it is near the mean.

For a measurement x:

Sample z-score

$$z = \frac{x - \bar{x}}{s}$$

Population z-score

$$z = \frac{x - \mu}{\sigma}$$