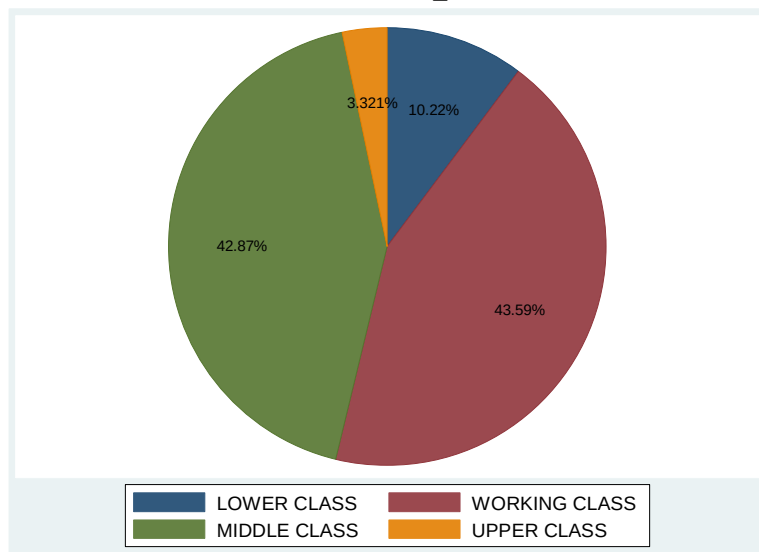


Handout 2: Graphs and Distributions**Commands in Stata**

`graph bar, over(varname) blabel(bar)` – bar graph, with or without percentage labels
`graph pie, over(varname) plabel(_all percent)` – pie graph, with or without percentage labels
`histogram varname [, bins(#) normal]` – histogram; options: number of intervals, normal curve
`stem varname` – stem-and-leaf plot
`graph box varname` – box-and-whisker plot
`graph save filename, replace` – saves a graph into a .gph file (Stata graph format) and replaces it if it exists
`graph use filename` – displays a previously saved graph in .gph format
`graph export filename.png, replace` – exports the graph into a PNG format and replaces it if it exists

Examples in Stata***Pie Chart***

```
graph pie, over(class) plabel(_all percent)
```

***Saving Graphs***

Graphs do not end up in the log file. To save graphs for future use in Stata graph file format, .gph file, we use **graph save** command:

```
graph save class_pie.gph
```

This format is only useful inside Stata (for example, if we may later want to reopen the graph in Stata and perhaps edit).

To display saved graph in Stata, we use **graph use** command:

```
graph use class_pie.gph
```

We can also export graphs into non-Stata formats to use outside of Stata – that’s what you will do to include graphs with your assignments. There are a few different file types that Stata can export to:

suffix	Output format
.ps	PostScript5

```
.eps      EPS (Encapsulated PostScript)
.wmf      Windows Metafile
.emf      Windows Enhanced Metafile
.pdf      PDF
.png      PNG (Portable Network Graphics)
.tif      TIFF
```

Best format for us to use for now: png

```
graph export class_pie.png
```

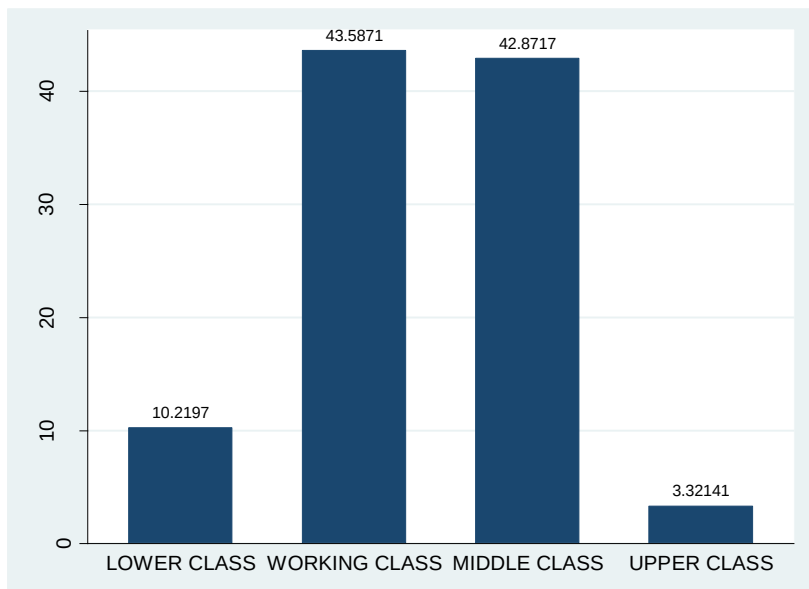
If the file exists already and you want to overwrite it, use replace option:

```
graph export class_pie.png, replace
```

Then, when you are, for example, in Word, use Insert -> Picture to add each graph to the document.

Bar Graph

```
graph bar, over(class) blabel(bar)
```



Stem and Leaf Plot

```
stem tvhours
```

```
0* | 00000000000000000000000000000000000000000000000000000000000000000000 ... (90)
0* | 11111111111111111111111111111111111111111111111111111111111111111111 ... (255)
0* | 22222222222222222222222222222222222222222222222222222222222222222222 ... (325)
0* | 33333333333333333333333333333333333333333333333333333333333333333333 ... (238)
0* | 44444444444444444444444444444444444444444444444444444444444444444444 ... (171)
0* | 55555555555555555555555555555555555555555555555555555555555555555555
0* | 66666666666666666666666666666666666666666666666666666666666666666666
0* | 77777777777777777777777777777777777777777777777777777777777777777777
0* | 88888888888888888888888888888888888888888888888888888888888888888888
0* | 999
1* | 000000000000000000000000
1* |
1* | 222222222222
1* | 33
1* | 4444
1* | 555
1* | 6
1* |
1* | 8
1* |
```

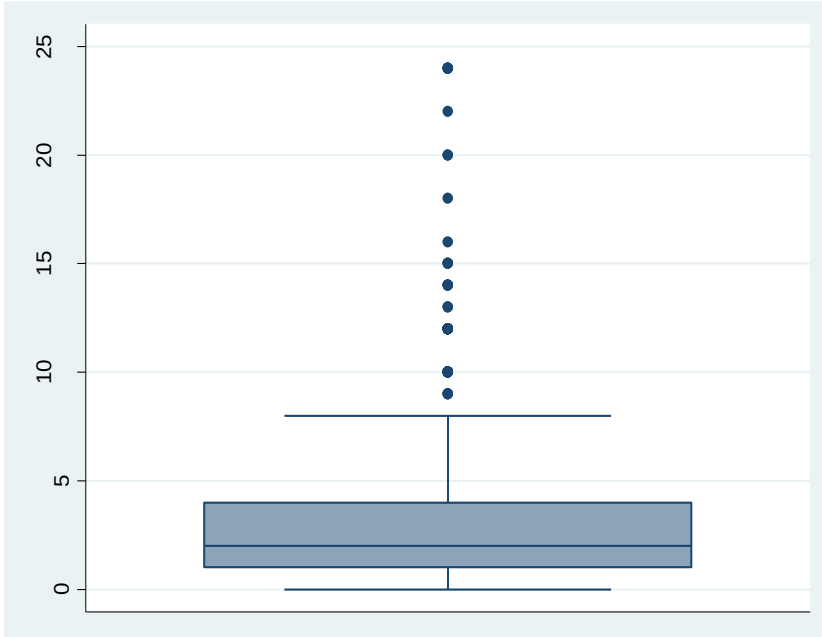
```

2* | 00
2* |
2* | 2
2* |
2* | 44444

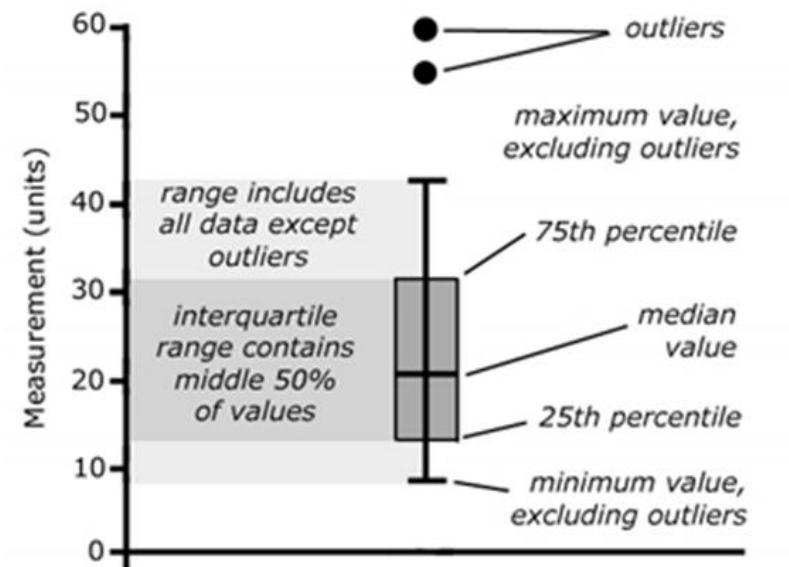
```

Boxplot (or box-and-whisker plot)

graph box tvhours



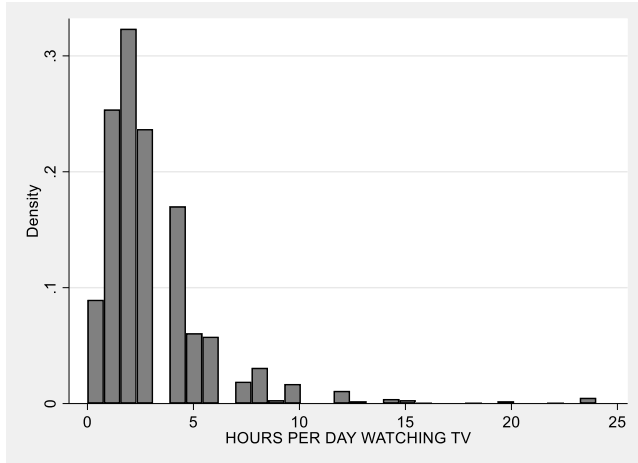
How to interpret it:



Note: Outliers on boxplot = observations that are below $Q1 - 1.5 \times IQR$; above $Q3 + 1.5 \times IQR$ ($Q1$ =first quartile, $Q3$ = third quartile, IQR =interquartile range)

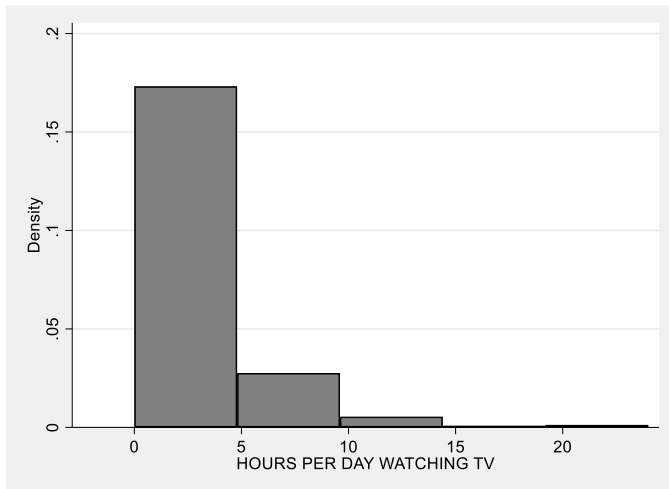
Histogram

```
histogram tvhours
```



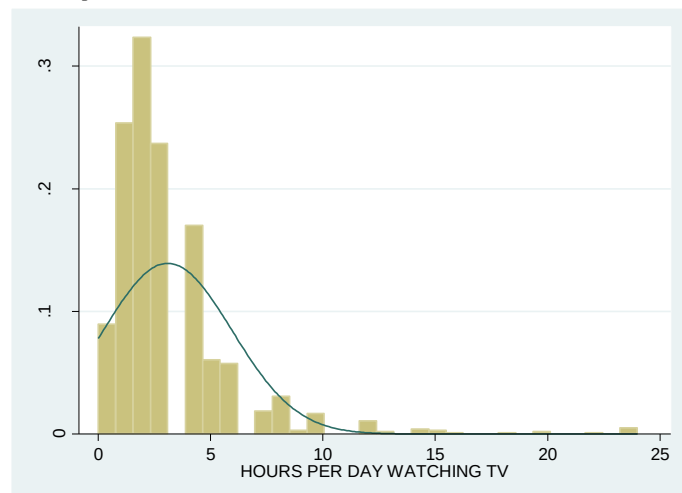
Change the number of bins:

```
histogram tvhours, bin(5)
```



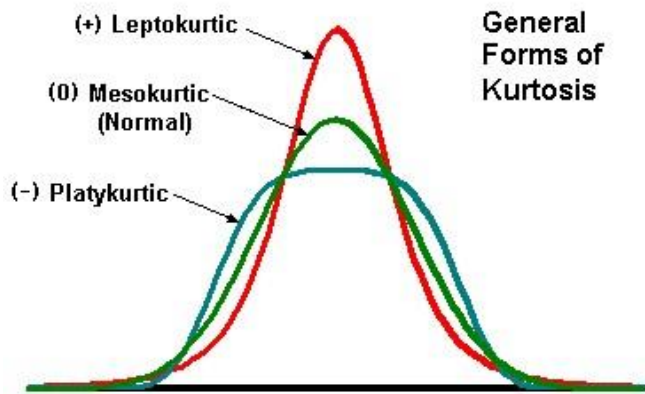
Adding a normal curve with the same mean and standard deviation:

```
histogram tvhours, normal
```

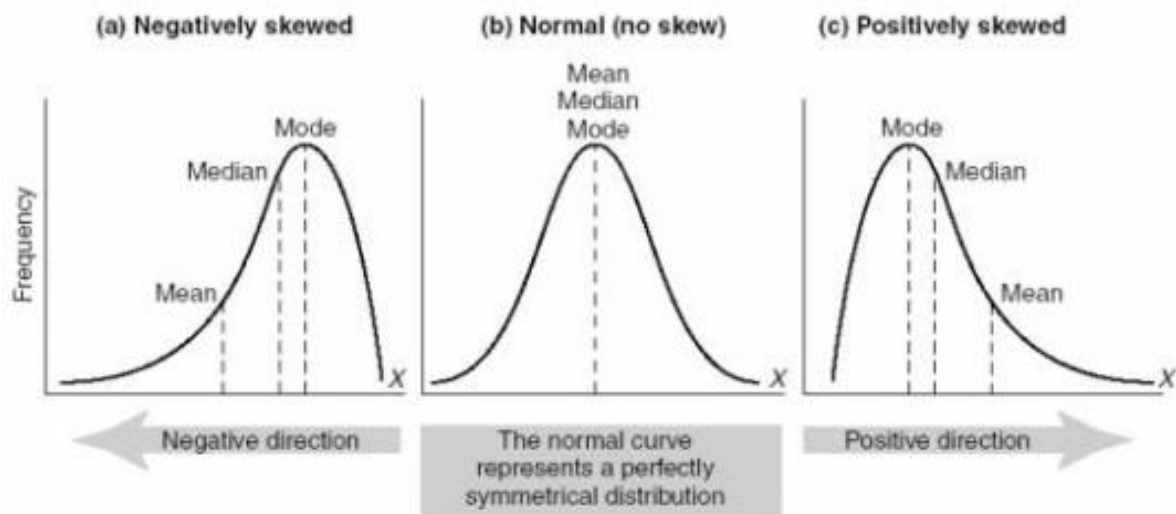


Measures of Distribution Shape

1. Kurtosis:



2. Skewness



Skewness and Kurtosis Statistics in Stata

```
tabstat tvhours, stats(mean skewness kurtosis)
      variable |      mean  skewness  kurtosis
-----+-----
      tvhours |   3.088598   3.123997  18.48296
-----+-----
```

Skewness:

- Near zero (between -2 and 2) = symmetric
- Positive = right, positive skew (>2)
- Negative = left, negative skew (<-2)

Kurtosis:

- **In Stata: subtract 3 before interpreting!**
- Near zero (between -2 and 2) = close to normal
- Positive = leptokurtic (>2)
- Negative = platykurtic (<-2)

Summary Chart: Univariate Descriptive Statistics by Level of Measurement

Variable Type	Central Tendency	Dispersion	Graphs	Distribution Shape
Nominal	mode	variation ratio	bar, pie	---
Ordinal	mode, median (sometimes: mean)	range, interquartile range (sometimes: standard deviation and variance)	bar, pie	---
Interval/ratio	mean, median (median preferred if outliers/skew), mode (rarely useful)	range, interquartile range, standard deviation, variance	histogram, frequency polygon, ogive, boxplot, stem-and-leaf	skewness and kurtosis