36-401 Class Notes and Exercise for 9/5/00

Howard Seltman

- Concord Water Study

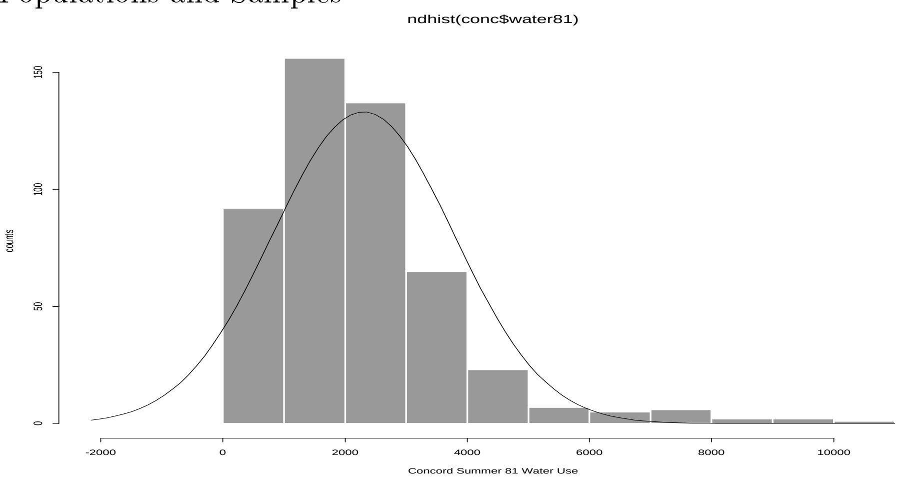
  Concord, New Hampshire (pop 38,000) had a water shortage in 1979 and 1980. An informational campaign was undertaken to convince people to use less water in 1981. An overall 15% decline in water use was seen. Questionnaire results from a random sample of households were matched to Water Department usage records to produce the data found in concord1.dat.

  496 households produced data like this:

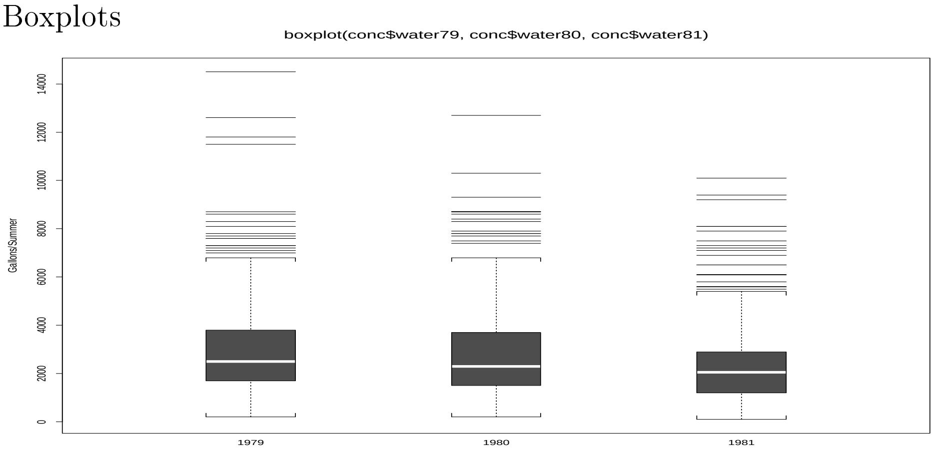
```
> conc[1:5,]
  case water81 water80 water79 income educat retire peop81 cpeop peop80
1    5    2300    3500    3500     35     18     no      3     0      3
2    6    2500    4600    4100     60     20     no      5     0      5
3    7    4000    3800    5500     30     16     no      6     1      5
4    8     500    1500    1700     20     11    yes      1    -1      2
5    9    4400    4300    3700     50     20     no      2     0      2
```

  The water columns give summer water use in gallons for the 2 years before the campaign and the one year after. Household income is in $thousands, education is in years, retire indicates whether the head of household was retired in 1981, peop81 is the number of people resident in 1981, cpeop is the change in number of people from 1980 to 1981, and peop80 is the number of people resident in 1980.

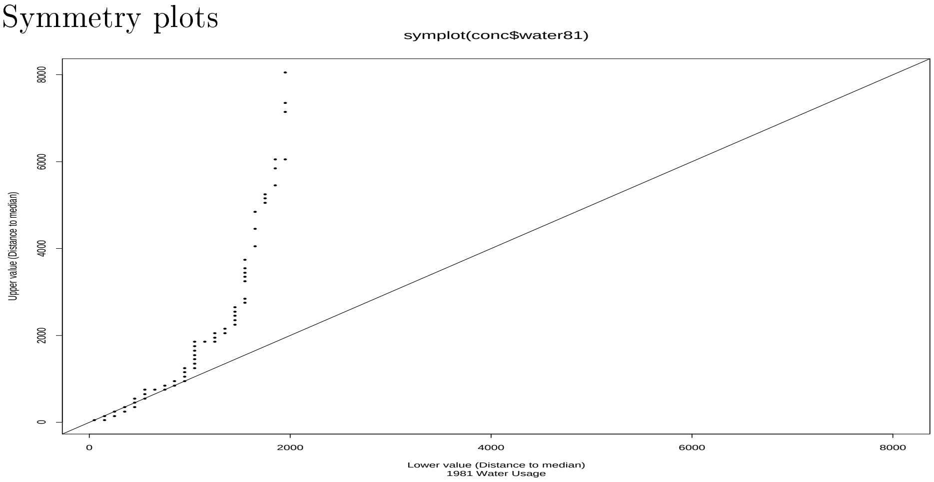
- Populations and Samples



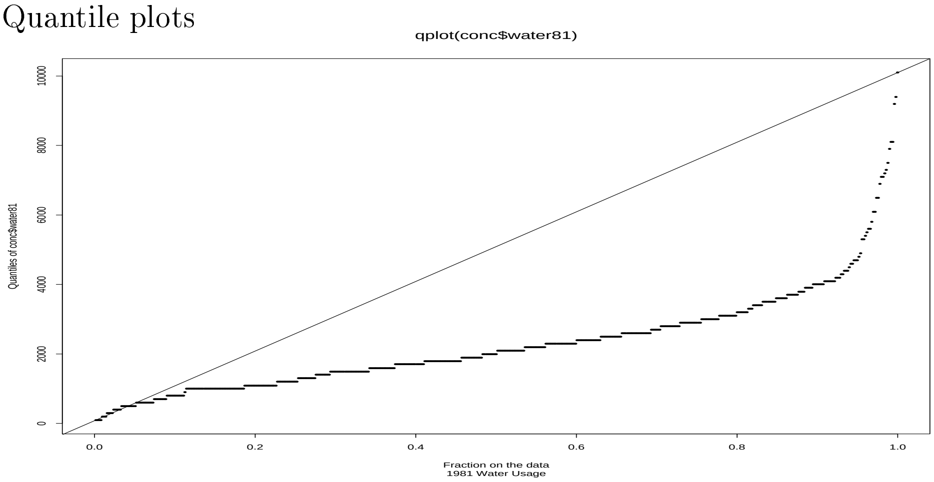
- Boxplots



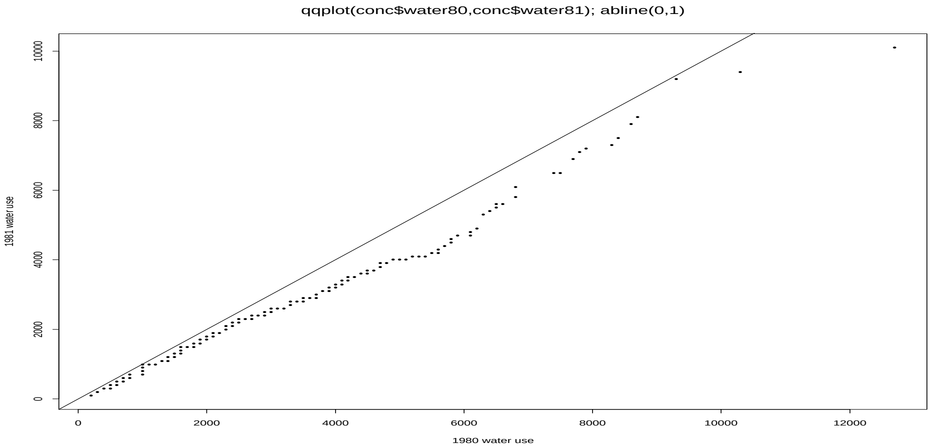
- Symmetry plots



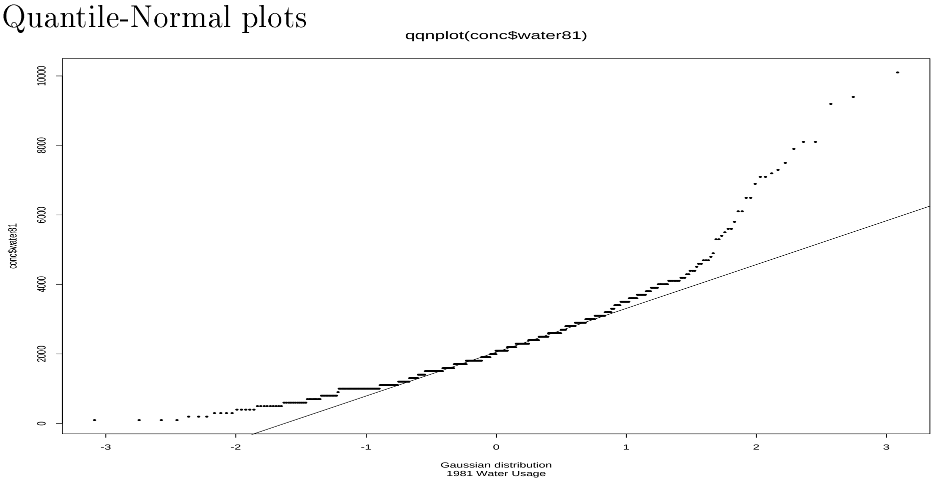
- Quantile plots



- Quantile-Quantile (Q-Q) plots



- Quantile-Normal plots

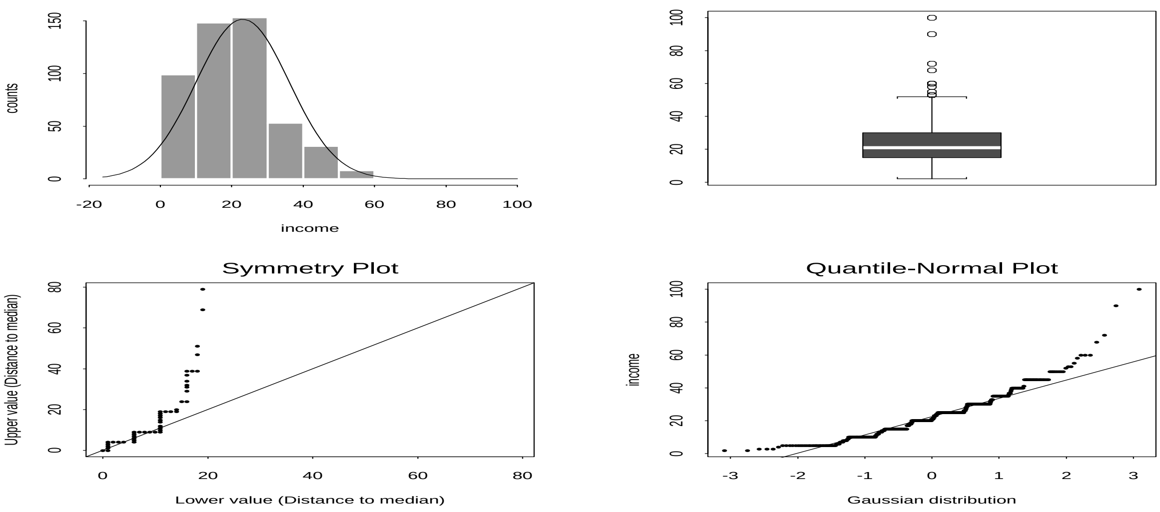


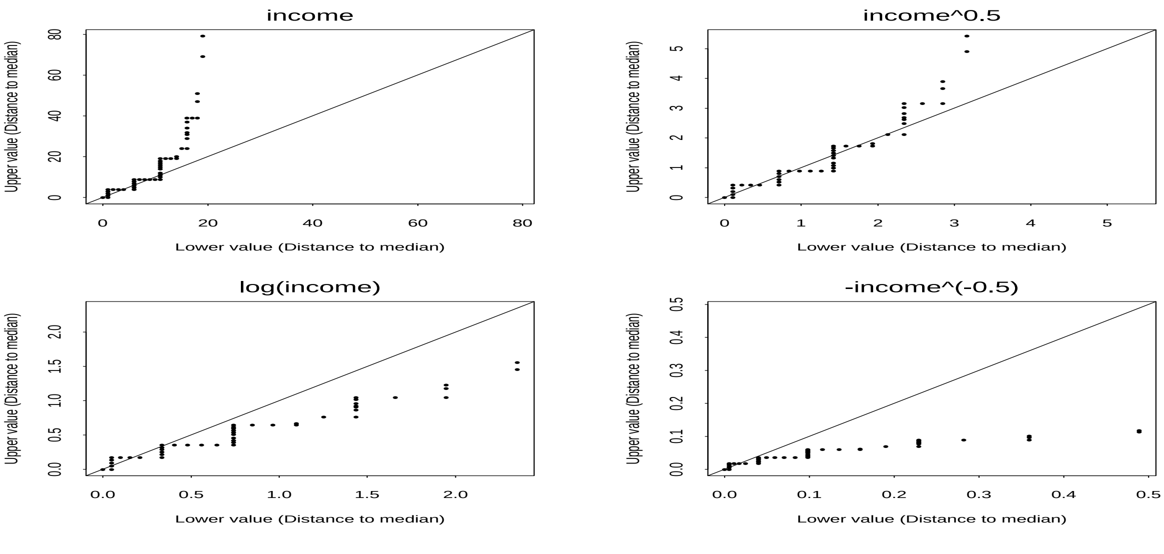
- Power transformations

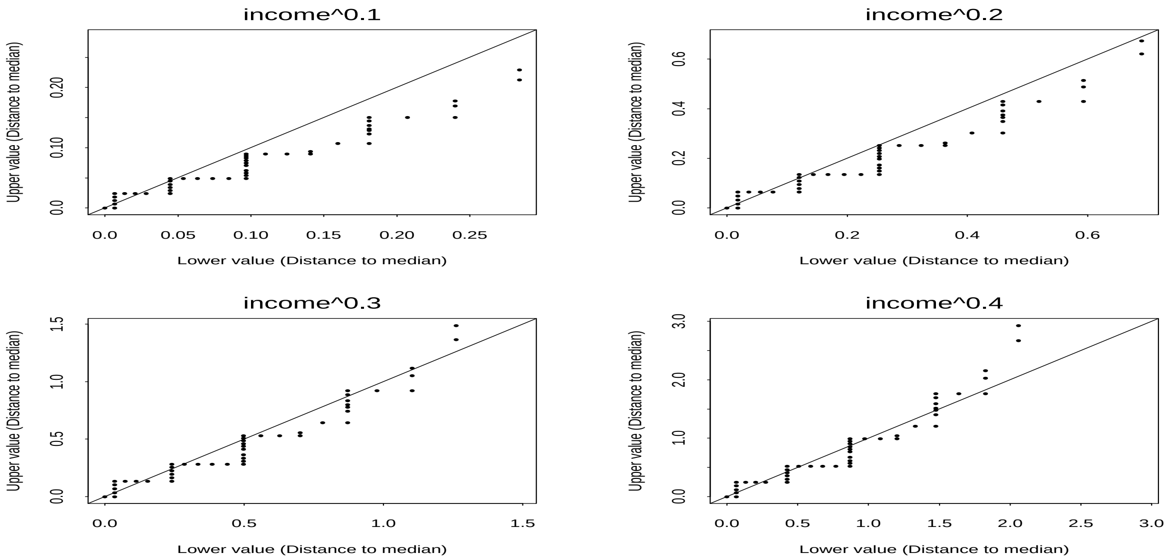
  Y^q $q < 1$ fixes skew right and vice versa

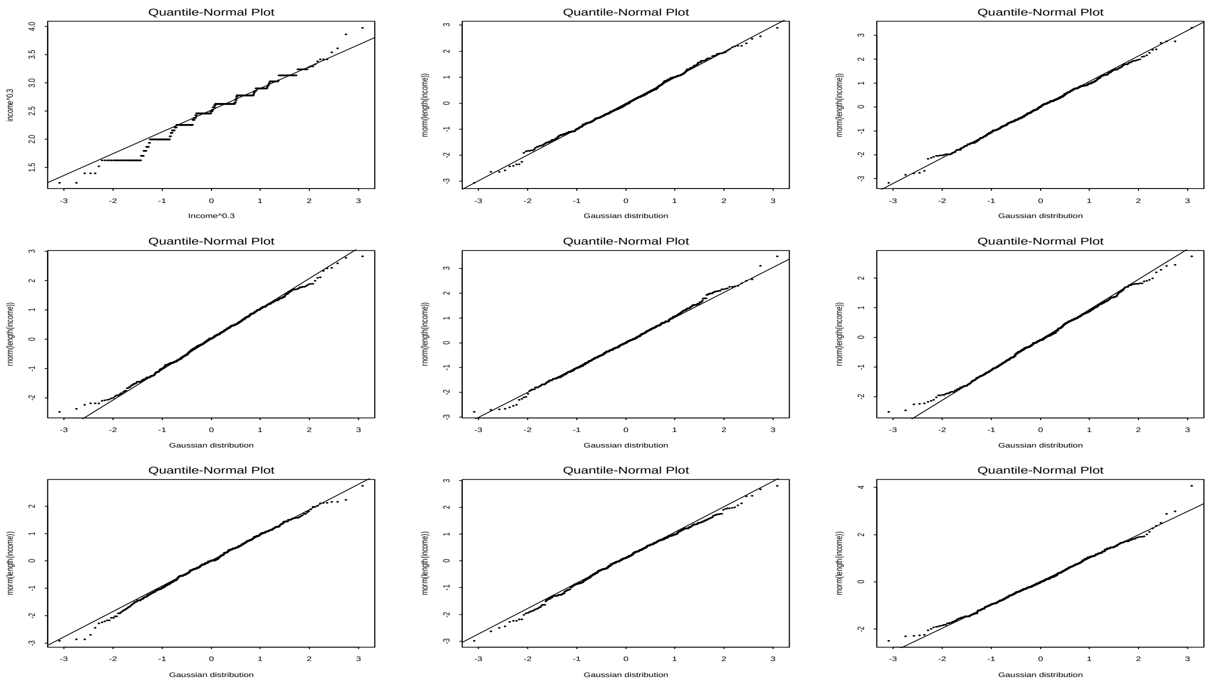
  use log(Y) instead of q=0

Exercise


counts
income
Symmetry Plot
Upper value (Distance to median)
Lower value (Distance to median)
Quantile-Normal Plot
income
Gaussian distribution


income
income^0.5
log(income)
-income^(-0.5)
Upper value (Distance to median)
Lower value (Distance to median)


income^0.1
income^0.2
income^0.3
income^0.4
Upper value (Distance to median)
Lower value (Distance to median)


Quantile-Normal Plot
Income^0.3
Gaussian distribution