Problem 1

Study 1:

Treatment: different hair colors
Response : the pain tolerance scores (between 0 and 100)

Observational study. We can't choose hair colors for individuals, either control their other personal characteristics that might affect their pain tolerance scores. So, we can't change it into an experiment.

Study 2 :

Treatment: the different baking temperatures
Response : the impact strengths

As before this is also a controlled experiment .

Study 3 :

treatment: the new diagnostic machine.
Response: the time it took for each mechanic to complete the jobs

This is clearly an experiment having treatments imposed .

Study 4 :

For this study treatment/factor is gender of the students. This is an observational study .

An experiment is generally designed to compare different treatments . So if we are interested for instance in finding the effects of some hypothetical "memory pill" on the exam scores separately for male and female students then the experimenter can choose the groups ,split each of them randomly in two distinct groups ,one to which the memory pill would be administered and the other one to which some "placebo" pill (which contains no drug) would be given .Then , after the exam the grades would be compared within the females group between the ones who took the memory pill and the ones who got the placebo (to see if the drug has any real effect) and within the males group similarly .Then the comparison should be made between males and females to asses any score differences between genders . This is one type of experiment which can be designed instead of just observing some grades after an exam took place .

Problem 2

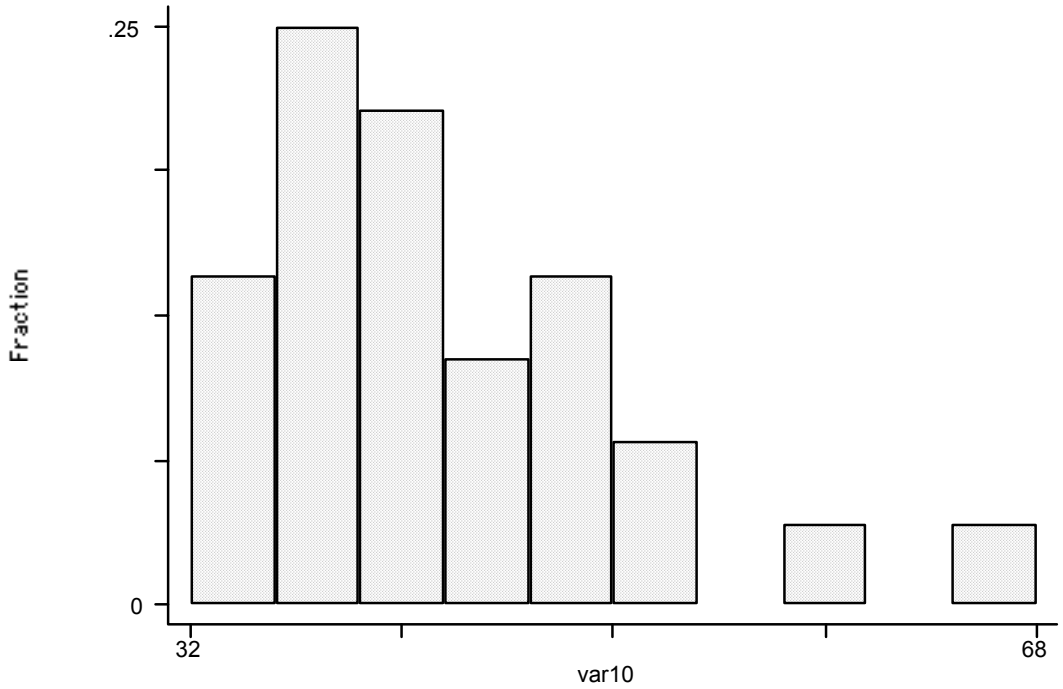
1) Display the data sorted in ascending order.

```
. sort var1

. list

     var1
 1.    32
 2.    34
 3.    34
 4.    35
 5.    36
 6.    37
 7.    37
 8.    38
 9.    38
10.    38
11.    39
12.    40
13.    41
14.    41
15.    42
16.    42
17.    42
18.    43
19.    44
20.    46
21.    47
22.    47
23.    49
24.    49
25.    50
26.    52
27.    58
28.    68
```

The histogram of number of parking tickets is shown below.



2) The data summary is

. summarize, detail

var10

Percentiles		Smallest		
1%	32	32		
5%	34	34		
10%	34	34	Obs	28
25%	37.5	35	Sum of Wgt.	28
50%	41.5		Mean	42.82143
		Largest	Std. Dev.	7.869555
75%	47	50		
90%	52	52	Variance	61.92989
95%	58	58	Skewness	1.315232
99%	68	68	Kurtosis	5.050482

The five number summary is
Min 32
1st quartile 35
median 41.5
3rd quartile 50
maximum 68

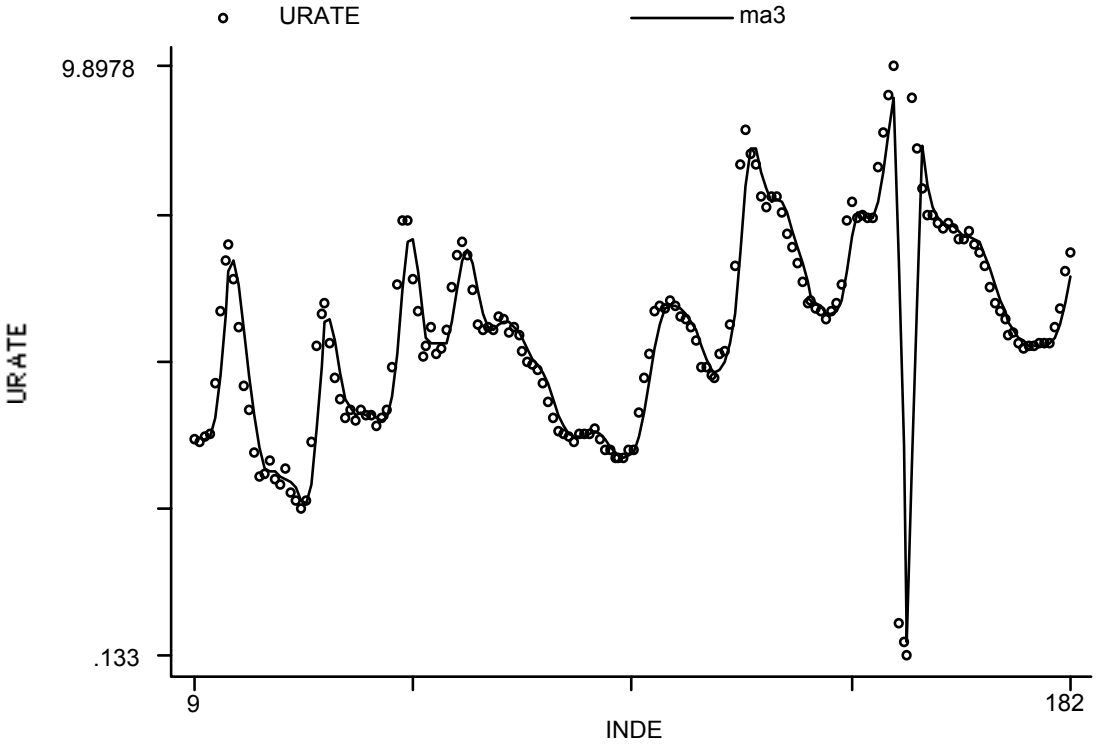
3) The distribution of number of parking tickets is skewed to the right, with many people getting tickets for a few days. For the 11th day of this period, 68 tickets are given, which is the largest value. For most of the days however, the number of tickets are between 32 and 45 or so.

Problem 3

Procedure to calculate 3-term moving average for unemployment rate variable:
The 3-term moving average for time t is $(Y_t + Y_{t-1} + Y_{t-2})/3$. In STATA, after loading data with command:
.insheet using http://www.stat.ucla.edu/~dinov/courses_students.dir/data.dir/unemgnp.dat

we can generate variables for 'urate' at time t-1 and t-2 by the following command:
. gen urate1 = urate[_n-1]
. gen urate2 = urate[_n-2]
. gen ma3=(urate + urate1 +urate2)/3

Plot original data 'urate' and 'ma3' together for a comparison:
. graph urate sum inde, connect(.l) symbol (oi)
As time t increases, the mean goes up, so does variance of the 'urate', therefore, this



process is not stationary.

The moving-average process is smoother than the original data as seen from the plot above, while preserving the non-stationarity. The original data is noisier.