These are intended to illustrate the type of Excel problems that will be on the midterm exam. As here, most of the Midterm problems will be very similar the homework problems. There are 2 main types:

(i)   "fill out the menu", which you do by writing on the menu on the exam what you would ordinarily type into each field. Caution: this requires some familiarity with Excel, as you are expected to know what to put into each field, without being able to click on it to get the explanation.

(ii)   "give an Excel formula", which you do by writing the formula that you would type to get Excel to solve the problem. The idea here is to avoid you doing the work (e.g. arithmetic and other numerical calculations) that is done by Excel, and at the same time it shows that you know how to do it in Excel. To get full credit on a problem, you need to write something which will give the correct numerical answer when typed into Excel (e.g. you need to include the "equals" sign that proceeds functions in an Excel field). Caution: this requires you to know the correct names and fields of the Excel functions that we have used so far in HW assignments, so you may want to put this information on your sheet of formulas. Note that often there are several ways to get the right answer, and any correct one (in the sense of giving the right numerical answer in Excel) will be accepted. For example either "=2*3-1" or "=5" (or even "=2*(4-1.5)") would be an acceptable answer to the question "what is one less than twice three?". But note that your chances of making a mistake are less if you use the first one.

Note: these are only example problems, and not intended to be comprehensive. The best way to study for the exam is to re-work the old HW problems, and while doing that, make sure you understand the concepts behind them.

I.   This excel spreadsheet contains survey scores for 7 people.

	B	C	D
6			
7	132	mean:	122.8571
8	114	s.d.:	23.00311
9	88	median:	129
10	129	IQR:	26
11	157	M-1.5IQR:	90
12	137	M+1.5IQR:	168
13	103		
14			

Sheet1

(a)   Write an Excel command to calculate the mean.

(b)   Write the Excel command that was used to calculate the entry in field D8.

(c)   Write an Excel command to calculate the interquartile range.

(d)   Write the Excel command that was used to calculate the entry in field D12.

(e)   Are there any outliers according to the rule given in class? If so, what are they?

(f)   What do the relative position of the mean and the median suggest about the shape of the distribution?

(g)   Write an Excel command that would give the standardized version of the data point in field B7, in a form that could be "dragged downward" to give the standard versions of all numbers in the list.

(h)   Write an Excel command to find the Z score of the median.

II.   The GRE exams help predict the performance of applicants to graduate schools. The scores X of applicants to a department have mean $\mu = 455$ and standard deviation $\sigma = 87$.

(a) Choose an Excel menu (only one!) below, and fill it out to find the relative frequency of applicants whose score X satisfies X < 480.

NORMDIST

X	= number
Mean	= number
Standard_dev	= number
Cumulative	= logical

=

Returns the normal cumulative distribution for the specified mean and standard deviation.

X is the value for which you want the distribution.

Formula result =   [OK]   [Cancel]

NORMINV

|	
Probability	= number
Mean	= number
Standard_dev	= number

=

Returns the inverse of the normal cumulative distribution for the specified mean and standard deviation.

Probability is a probability corresponding to the normal distribution, a number between 0 and 1 inclusive.

Formula result =   [OK]   [Cancel]

(b) Give an Excel command to find the relative frequency of applicants whose score X satisfies X > 520. (Caution: be sure to put the *entire* formula, including the = sign, for full credit).

(c) Give an Excel command to find the relative frequency of applicants whose score X satisfies 550 > X > 450.

(d) Give an Excel command to find the relative frequency of applicants whose score X satisfies 450 > X > 550.

(e) Give an Excel command to find the score where only 20% of the population are higher.

(f) Give an Excel command to find the first quartile of the population.

III.   The same set of data were analyzed using 3 different bin grids, and gave these histograms:


Histogram, Bin Range 1

Histogram, Bin Range 2

Histogram, Bin Range 3
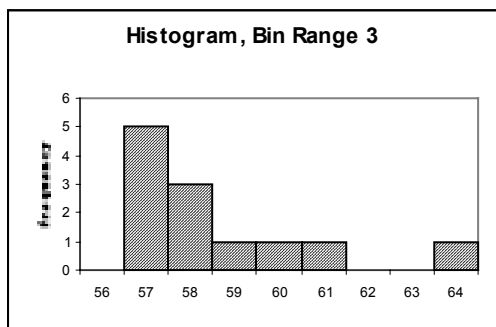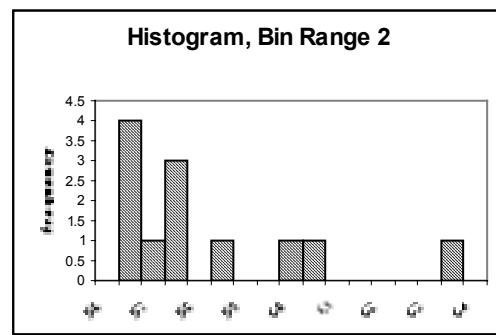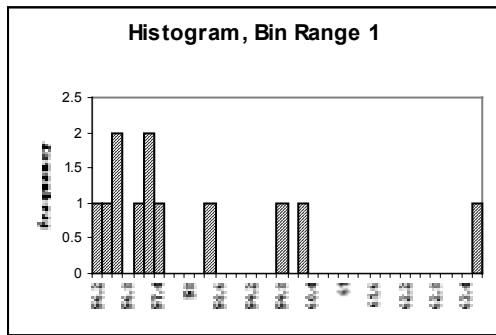
(a)   Which histogram best fills the description "oversmoothed, important detail is gone"?

(b)   Which histogram best fills the description "smoothed about right, important detail is present"?

(c)   Which histogram best fills the description "undersmoothed, sampling noise obscures main features of the population?"

(d)   Does the population shape seem symmetric, left skewed or right skewed?

(e)   Do you expect the median or the mean to be larger for these data?

(f)   Do there appear to be any candidates for potential outliers?

IV.   The time plot of the home runs per year hit by a fictional baseball player is:

Fictional Baseball Player
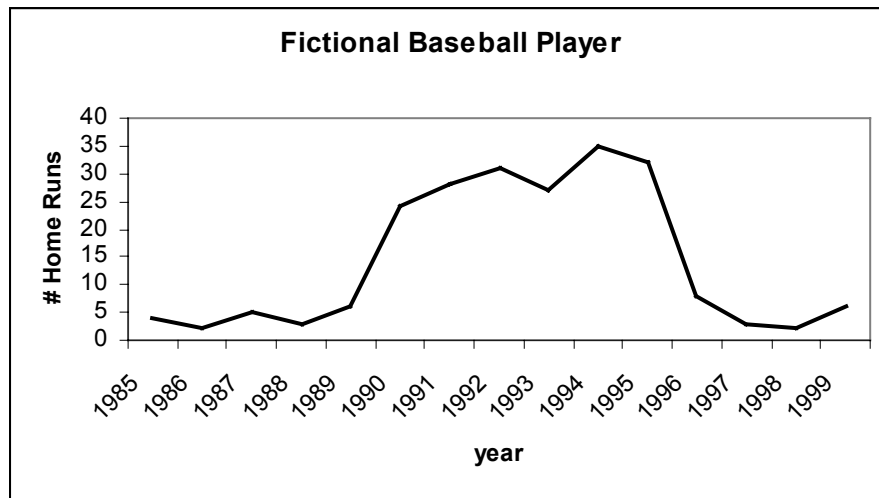
a.   Which of the following best applies to this time plot?
       i.   He was very hot early in his career, and gradually tapered off over time.
       ii.   He showed a steady improving trend through his career.
       iii.   His performance was initially poor, much better in mid-career, then decreased later.

b.   Which explanation of the shape seems most sensible?
       i.   Hitting home runs is a difficult skill that takes a long time to master, but once you do you never lose the skill.
       ii.   In his first few years he didn't play much, then he became a starter, finally he was hampered by serious injuries.
       iii.   Home run hitting requires the strength of a young man, so players gradually hit fewer home runs as their careers progress.

c.   Time plots can be constructed using:
       i.   The Excel function TIMEPLOT.
       ii.   The Excel chart Wizard.
       iii.   The Excel function TIMESERIES

V.   Here is a scatterplot of manatees killed per year, as a function of number of boats registered.
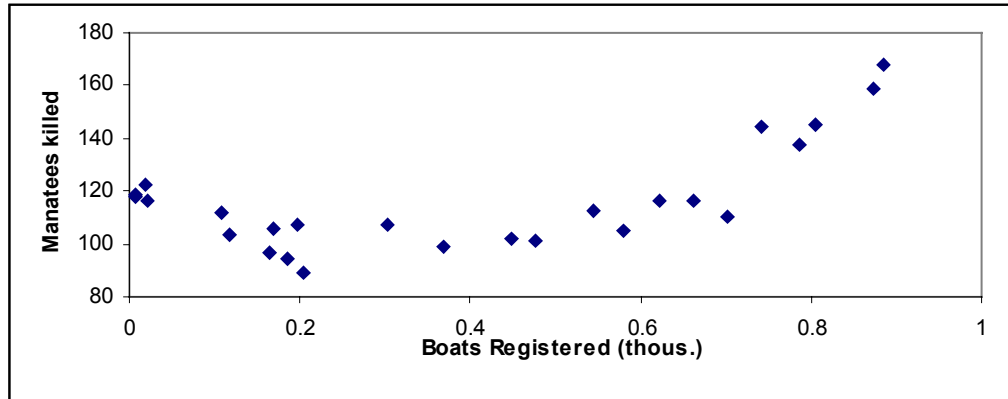
a.   Which best describes the degree of linear fit?
      (i) strongly linear   (ii) roughly linear   (iii) crudely linear, but somewhat curved.

b.   Which one of these could be the correlation between manatees killed and boats reg'd?
      (i) -0.74   (ii) -0.003   (iii) 0.19   (iv) 0.62   (v) 0.97   (vi) 4.3

c.   Name an Excel function that could be used to compute the correlation.

d.   Which one of these could be the slope,intercept of the least squares fit line?
      (i) 41,100   (ii) 47,180   (iii) 120,80   (iv) -20,100

e.   Name Excel functions that can be used to calculate the slope and the intercept.

f.   Which of these could be the residual plot from the least squares fit line?

(iii)
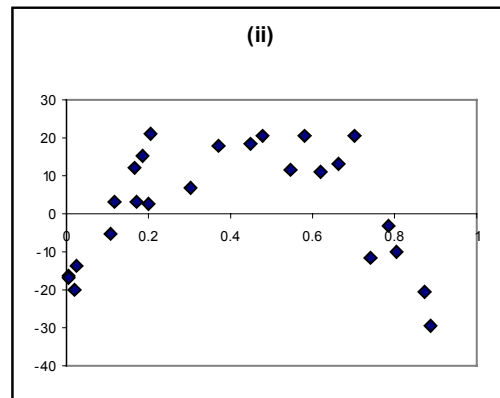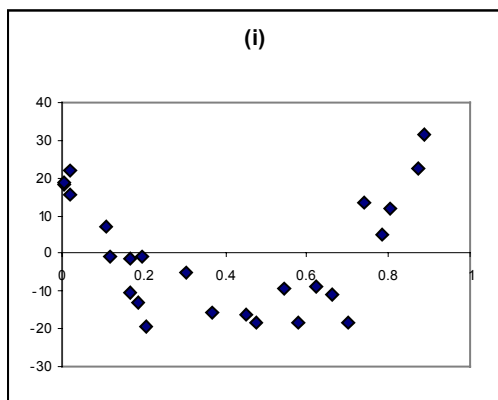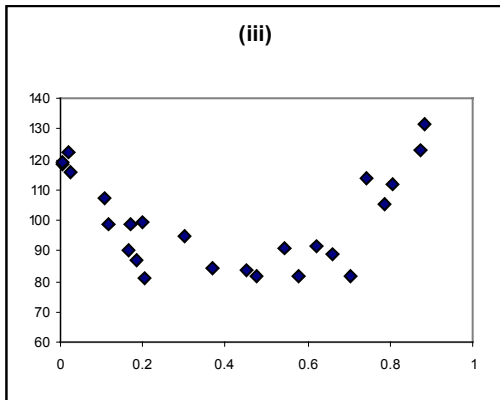
g.   Do any of these residual plots suggest that the line is a good model for this dependence? Explain why or why not in 20 words or less.

h.   Do there appear to be any influential observations, or outliers?

i.   If a future year has 1000 boats registered, which number seems closest to the number of manatees that will be killed?
(i) 74   (ii) 141   (iii) 243   (iv) 1,590

j.   Do you have a lot of confidence in the answer to (ii)? Why or why not? (20 words or less)