

1. The file shelter.dat posted on the web page as

<http://www.public.iastate.edu/stat415/model.free.curve.fitting.html>

contains data on wind speed at various distances behind a windbreak. In this case the windbreak is a row of trees planted along the western boundary of a field. The data were recorded when the wind was directly from the west and hit the windbreak at a 90 degree angle. There are two numbers on each line of the file. The first is distance behind the windbreak (in meters) and the second is the relative velocity of the wind recorded as a percentage of the wind speed just before it encounters the front of the windbreak.

- (a) Describe what is meant by the trade off between "variance" and "bias" in deciding how smooth to make an estimated curve.
- (b) In addition to looking at the how fitted curves pass through the data points on a plot, describe two other methods you could use to determine the value of the span in using the loess procedure to fit a curve. Apply these methods to the selection of the loess curve in part C.
- (c) Try each of the following methods for fitting a curve to these data:
 - (i) Polynomial regression (It may require a high order polynomial)
 - (ii) Loess
 - (iii) Gaussian kernel smoother

For each method provide a graph of the fitted curve and also present a residual plot with a "smoothed curve" to indicate any trend. Indicate the degree of the polynomial or the span or bandwidth you selected for Loess or Gaussian kernel smoothing. Finally estimate the mean relative wind speed at 1, 2, 5, 10, 20 and 30 meters behind the windbreak and report standard errors or confidence intervals for your estimates.

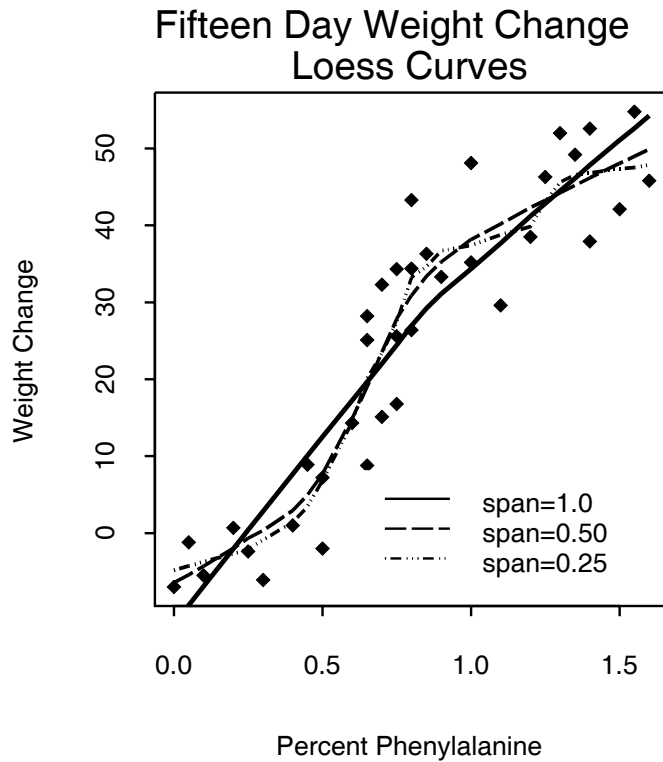
2. In a study of the amino acid phenylalanine, 36 young male rats weighing 50 to 55 grams were fed a diet lacking phenylalanine for two days. Then, the rats were fed new diets with different levels of phenylalanine. The change in weight after 15 days on the new diet was recorded for each rat. The following table show the levels of phenylalanine used in the new diets and the results for each rat. Some levels of phenylalanine were fed to more than one rat. Negative values correspond to weight losses and positive values correspond to weight gains.

Level of Phenylalanine in the Diet (Percent)	Weight Change After Fifteen Days (grams)
0.00	-7.0
0.05	-1.2
0.10	-5.5
0.20	0.7
0.25	-2.4
0.30	-6.1
0.40	1.0
0.45	8.9
0.50	-2.0, 7.2
0.60	14.3
0.65	8.8, 25.1, 28.2
0.70	32.3, 15.1
0.75	25.6, 16.8, 34.3
0.80	43.3, 26.4, 34.4
0.85	36.3
0.90	33.3
1.00	35.2, 48.1
1.10	29.6
1.20	38.5
1.25	46.3
1.30	52.0
1.35	49.2
1.40	37.9, 52.6
1.50	42.1
1.55	54.8
1.60	45.8

The following S-PLUS code was used to enter the data into a data frame and fit three curves to the data using the loess procedure.

```
wtloss <- read.table("palanine.dat",col.names=c("x","y"))
model100 <- loess(formula=y~x,data=wtloss, span=1.00,degree=1)
model50 <- loess(formula=y~x,data=wtloss,span=.50,degree=1)
model25 <- loess(formula=y~x,data=wtloss,span=.25,degree=1)
```

The curves are shown on the following graph.



(a) Describe the basic features of the loess procedure for fitting a smooth curve. Your description should include a definition of the “span”.

(b) The `anova()` function in S-PLUS was applied with the following statement

```
anova(model100, cmodel50, model125)
```

to obtain the following table of sums of squared residuals and “effective” degrees of freedom.

Analysis of Variance Table

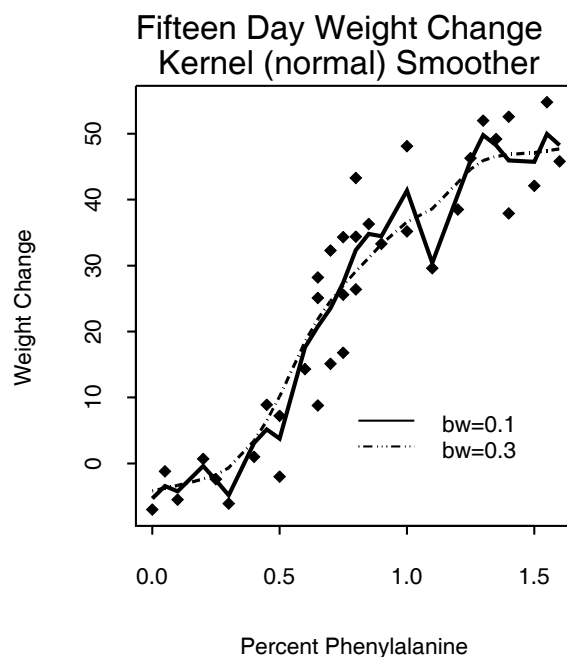
Model	Effective Number of Parameters	Sum of Squared Residuals
	ENP	RSS
1: (span=1.00)	2.4	2009.2
2: (span=0.50)	4.2	1442.0
3: (span=0.25)	7.2	1320.2

Show how to use these results to determine which of the loess curves provides the best description of how the mean fifteen day weight change in rats varies with the level of phenylalanine in the diet. State your conclusions.

- (c) If you do not specify a value of the span option in the `loess()` function in S-PLUS, it will automatically select a span based on a crossvalidation procedure. Briefly describe how this is done.
- (d) The following S-PLUS code was used to fit smooth curves to the data using kernel estimation with a normal kernel and two different bandwidths.

```
ker1<-ksmooth(wtloss$x, wtloss$y, kernel="normal", bandwidth=.1)
ker2<-ksmooth(wtloss$x, wtloss$y, kernel="normal", bandwidth=.3)
```

The curves are shown on the following graph.



Describe the basic features of the kernel method for estimating a smooth curve.

(e) The following two plots are plots of the residuals from the estimated curve obtained from the kernel smoother with bandwidth 0.30. What can you conclude from these plots?

