DEPARTMENT OF STATISTICS
Course STATS 330: Advanced Statistical Modelling
Tutorial Sheet 4: August 18, 2011

This tutorial is designed to give you practice in the following:

- Checking a model for normality
- Checking for independence
- Subset selection using all possible regressions and stepwise
- Simulating the distribution of a statistic

In this tutorial we will be using the **car data** used in Tutorial 3 and also the **wine data** discussed in Lecture 13.

Task 1: Read in the data

Download the wine data from the web. Make a data frame **`wine.df`**. If you haven't saved the car data, read that in again and recreate the last regression **`recip.no47.lm`** from Tutorial 3.

Task 2: Check normality

Check the car data for normality, using a normal plot and the WB test.

Use

```
plot(recip.no47.lm)
```

to draw the normal plot and

```
WB.test(recip.no47.lm)
```

to do the test. Are there any indications of non-normality?

Task 3: Fit the wine regression

Fit the wine data, using the reciprocal cube root as was done in class in Lecture 13. Call the result **`recip.cube.root.lm`**.

Compare this with using log price as the response. How do these two fits compare?

Task 4: Check for independence

Check the wine data for independence, using the fit **`recip.cube.root.lm`**. Use the acf plot, the residuals versus previous residual plot, and the Durbin-Watson test. Use code

```
res <-residuals(recip.cube.root.lm)
acf(res)  # for the acf plot

n<-length(res)
plot.res<-res[-n]
prev.res<-res[-1]
plot(prev.res,plot.res,
xlab="previous residual",
ylab="residual",
main="Residual versus previous residual \n for the wine
data")
rhohat<-cor(plot.res,prev.res)
DW=2*(1-rhohat)
```

Does this indicate a lack of independence?

Task 5: Select a subset

Using the car data (use the final model **`recip.no47.lm`** derived in Tutorial 3) select a subset of variables. Use the all possible regressions method discussed in class today.

Task 6: Simulating the distribution of a statistic

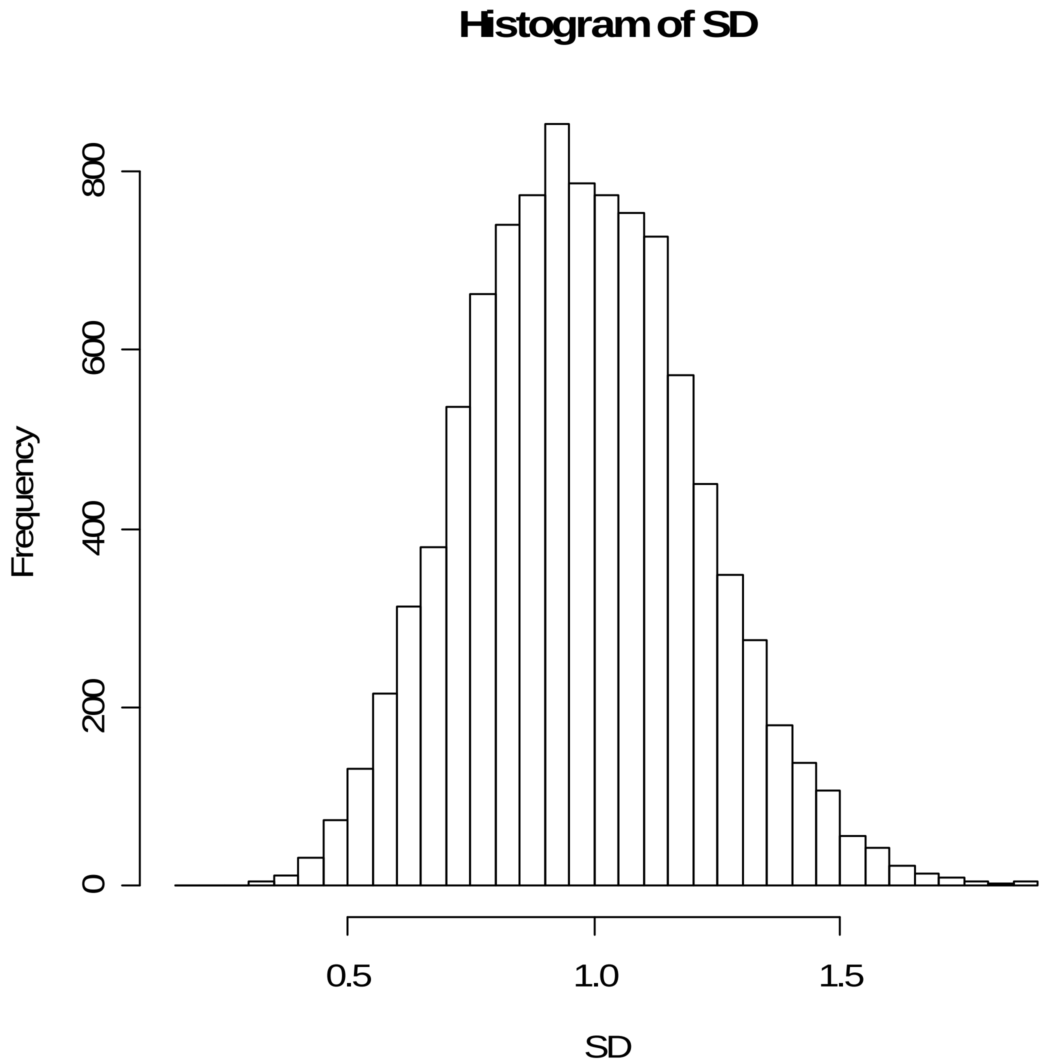
R makes it easy to simulate the distribution of any statistic. For example, suppose we want to study the distribution of the sample standard deviation in samples of size 10 taken from a normal distribution.

We can build up the distribution of the standard deviation by drawing repeated samples, storing the value of the standard deviation each time, and plotting the results. The following code does this 10000 times:

```
n=10  # set sample size
N=10000  # set number of repeats
SD = numeric(N)  # make a place to store the results
# do the simulation
for(i in 1:N){   #  repeat 10000 times
     SD[i] = sd(rnorm(n))
     }

# draw a picture of the results

hist(SD, nclass=50)
```



Histogram of SD
Frequency
0
200
400
600
800
0.5
1.0
1.5
SD

Table of the Durbin-Watson Test (5% points)

k	k=1		k=2		k=3		k=4		k=5	
n	dL	du	dL	du	dL	du	dL	du	dL	du
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
16	1.10	1.37	0.98	1.54	0.86	1.73	0.74	1.93	0.62	2.15
17	1.13	1.38	1.02	1.54	0.90	1.71	0.78	1.90	0.67	2.10
18	1.16	1.39	1.05	1.53	0.93	1.69	0.82	1.87	0.71	2.06
19	1.18	1.40	1.08	1.53	0.97	1.68	0.86	1.85	0.75	2.02
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
21	1.22	1.42	1.13	1.54	1.03	1.67	0.93	1.81	0.83	1.96
22	1.24	1.43	1.15	1.54	1.05	1.66	0.96	1.80	0.86	1.94
23	1.26	1.44	1.17	1.54	1.08	1.66	0.99	1.79	0.90	1.92
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	0.93	1.90
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
26	1.30	1.46	1.22	1.55	1.14	1.65	1.06	1.76	0.98	1.88
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.76	1.01	1.86
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.85
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.84
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.83
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.15	1.81
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80
36	1.41	1.52	1.35	1.59	1.29	1.65	1.24	1.73	1.18	1.80
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.21	1.79
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.79
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.78
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.38	1.77
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77
75	1.60	1.65	1.57	1.68	1.54	1.71	1.51	1.74	1.49	1.77
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
85	1.62	1.67	1.60	1.70	1.57	1.72	1.55	1.75	1.52	1.77
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.78
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78