Euclidean distance versus H distance

The MEANS Procedure

Variable	Mean	Std Dev
x1	0.00	2.32
x2	0.50	0.45

x1	x2
0	1
-2	0
-3	1
-3	0
-2	.5
0	.5
3	.5
3	0
3	1
1	0
0	1

y=3 + 2*x1 + 4*x2 + rand('NORMAL',0,2); *B0=3 B1 = 2, B2 = 4 and Sigma=2;

Euclidean distance versus H distance

The REG Procedure
Model: MODEL1
Dependent Variable: y

Number of Observations Read	11
Number of Observations Used	11

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	264.69687	132.34843	40.53	<.0001
Error	8	26.12415	3.26552		
Corrected Total	10	290.82102			

Root MSE	1.80707	R-Square	0.9102
Dependent Mean	5.64744	Adj R-Sq	0.8877
Coeff Var	31.99810		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	3.66678	0.84024	4.36	0.0024
x1	1	2.04225	0.24620	8.30	<.0001
x2	1	3.96133	1.27928	3.10	0.0147

Euclidean distance versus H distance

The REG Procedure
Model: MODEL1
Dependent Variable: y

Output Statistics								
						DFBETAS		
Obs	Residual	RStudent	Hat Diag H	Cov Ratio	DFFITS	Intercept	x1	x2
1	2.7927	2.0753	0.2162	0.4519	1.0899	-0.1733	-0.0399	0.8297
2	2.9898	2.5216	0.2812	0.2988	1.5770	1.3532	-0.7598	-1.0138
3	-1.8321	-1.3770	0.3972	1.2064	-1.1177	0.1577	0.7551	-0.6626
4	-0.0922	-0.0601	0.3693	2.3632	-0.0460	-0.0341	0.0296	0.0253
5	-1.7113	-1.0420	0.1652	1.1601	-0.4635	-0.2116	0.3107	-0.0150
6	-0.0493	-0.0268	0.0909	1.6415	-0.0085	-0.0055	0.0000	0.0000
7	-0.0104	-0.006272	0.2580	2.0116	-0.0037	-0.0015	-0.0030	0.0001
8	-1.4249	-1.0179	0.3972	1.6366	-0.8262	-0.6292	-0.5581	0.4898
9	0.5591	0.3679	0.3693	2.2347	0.2816	-0.0273	0.1815	0.1549
10	-0.5871	-0.3515	0.2394	1.8622	-0.1972	-0.1894	-0.0618	0.1453
11	-0.6342	-0.3745	0.2162	1.7944	-0.1967	0.0313	0.0072	-0.1497

Sum of Residuals	0
Sum of Squared Residuals	26.12415
Predicted Residual SS (PRESS)	51.08065

Euclidean distance versus H distance

Obs	x1	x2	edistance	hdistance
1	0	0.5	0.00000	0.09091
2	0	1.0	0.50000	0.21620
3	0	1.0	0.50000	0.21620
4	1	0.0	1.11803	0.23940
5	-2	0.5	2.00000	0.16516
6	-2	0.0	2.06155	0.28116
7	3	0.5	3.00000	0.25796
8	-3	1.0	3.04138	0.39717
9	-3	0.0	3.04138	0.36933
10	3	0.0	3.04138	0.39717
11	3	1.0	3.04138	0.36933

```
edistance=sqrt(((x1-x1mean)**2) + ((x2-x2mean)**2));

hdistance=hii;
```

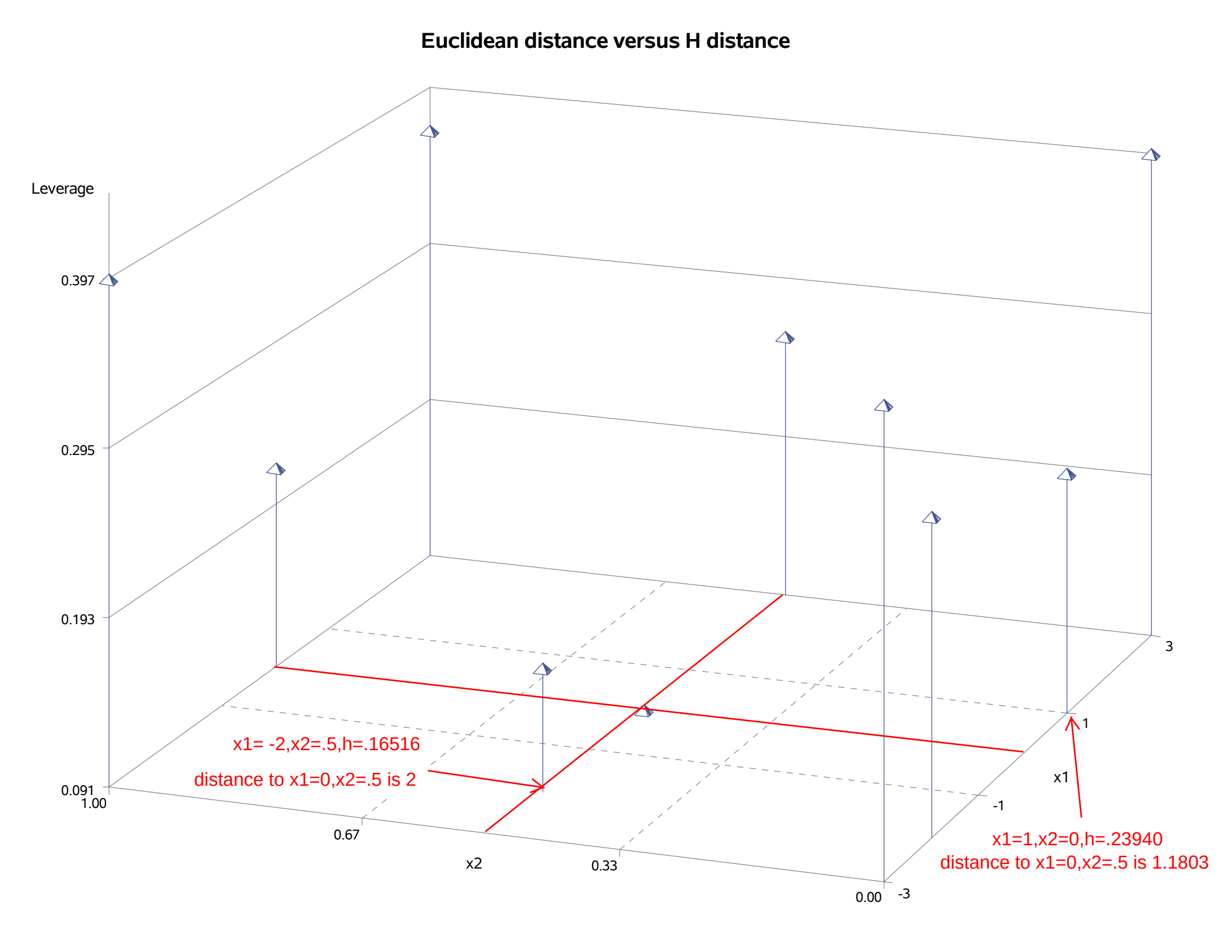
- The **hat matrix** is:

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

- The diagonal elements of the hat matrix are given by

$$h_{ii} = \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i$$

- h_{ii} – standardized measure of the distance of the *i*th observation from the center of the x-space.


Euclidean distance versus H distance
Leverage
0.397
0.295
0.193
0.091
1.00
0.67
x2
0.33
0.00
-3
-1
1
3
x1
x1= -2,x2=.5,h=.16516
distance to x1=0,x2=.5 is 2
x1=1,x2=0,h=.23940
distance to x1=0,x2=.5 is 1.1803