Misusing Statistics

Lecturer: John Guttag

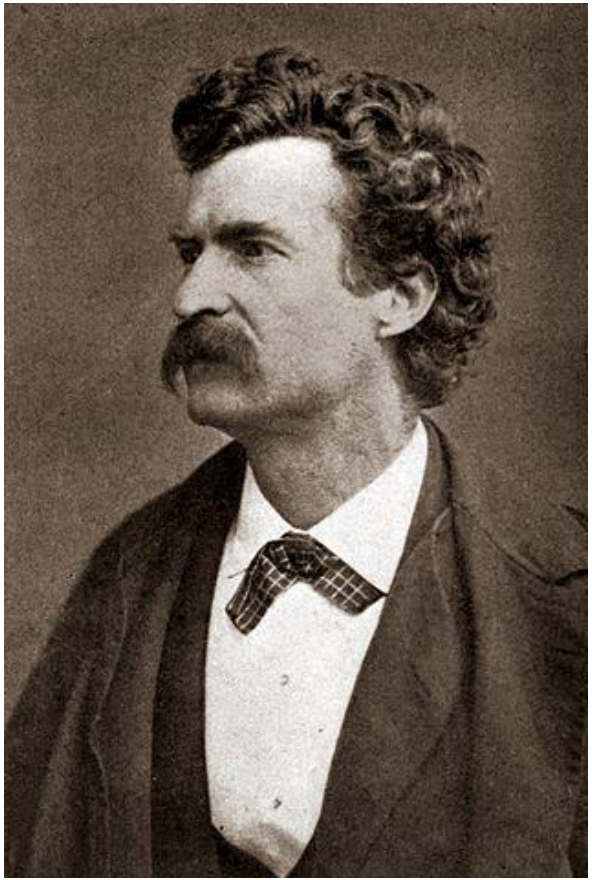
There are three kinds of lies

LIES

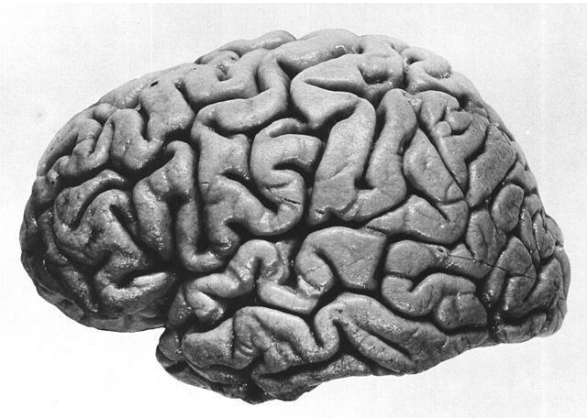
DAMNED LIES

and

STATISTICS



"If you can't prove what you want to prove, demonstrate something else and pretend they are the same thing. In the daze that follows the collision of statistics with the human mind, hardly anyone will notice the difference." – *Darrell Huff*



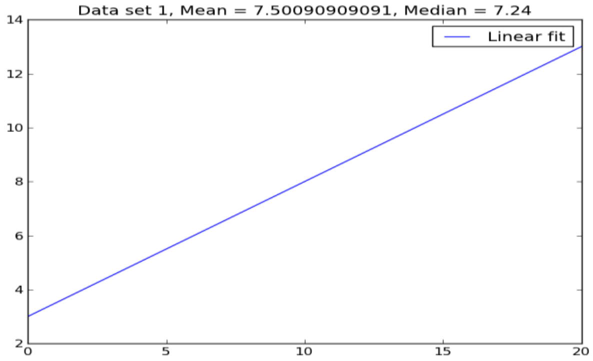
$$
\begin{aligned}
\partial_t S &= \int_0^L K(\frac{\partial_x T}{T})^2\,dx \quad (4),\\
&= \int_0^a K(\frac{\partial_x T}{T})^2\,dx + \int_a^L K(\frac{\partial_x T}{T})^2\,dx \quad (5),\\
&= K\alpha_1^2 \int_0^a K\frac{dx}{T^2} + K\alpha_2^2 \int_a^L \frac{dx}{T^2} \quad (6),\\
&= K\alpha_1 \left[\frac{1}{T_1} - \frac{1}{T_a}\right] + K\alpha_2 \left[\frac{1}{T_a} - \frac{1}{T_2}\right] \quad (7),\\
&= \frac{K}{T_a}\left[\frac{(T_a - T_1)^2}{aT_1} + \frac{(T_2 - T_a)^2}{(L-a)T_2}\right] \quad (8),
\end{aligned}
$$

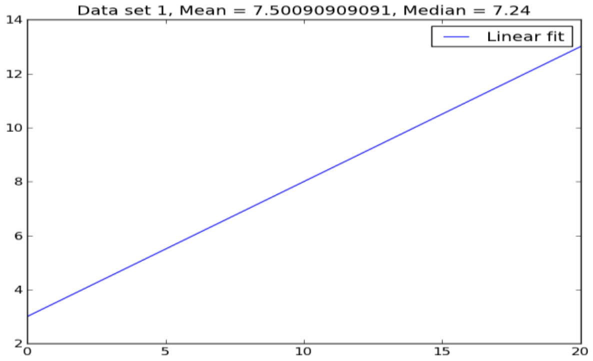
Anscombe's Data

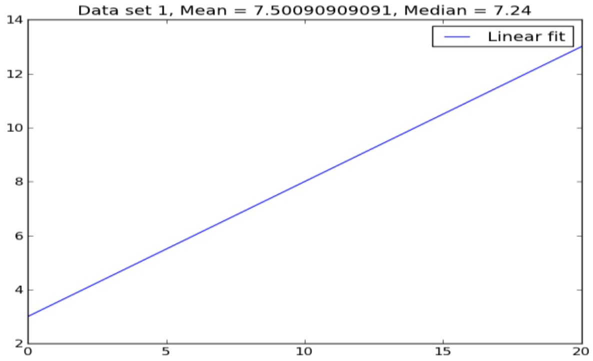
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

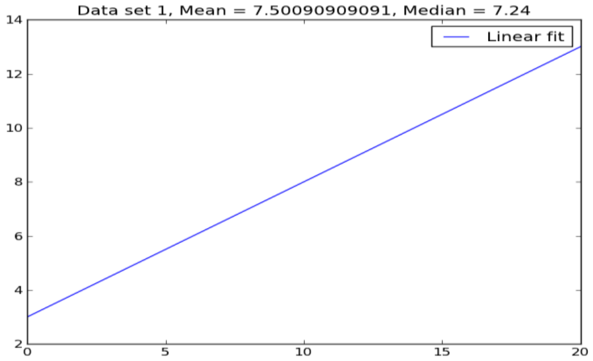
Executive Summary

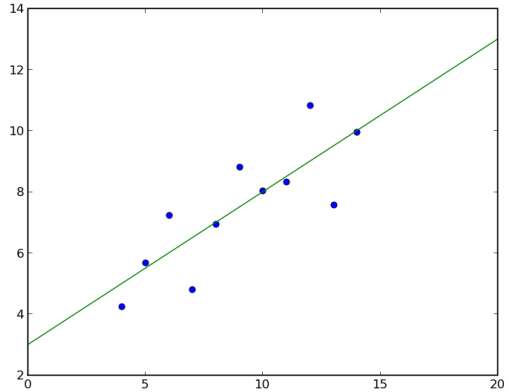
- Same mean value for x (9.0) and y (7.5)
- Same variance for x (9.0) and y (3.75)
- Same correlation between x and y (0.816)
- Linear regression yields same fit (y = 0.5x + 3)

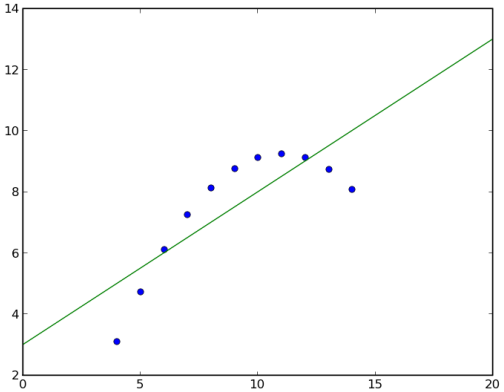

Data set 1, Mean = 7.50090909091, Median = 7.24
Linear fit
14
12
10
8
6
4
2
0
5
10
15
20

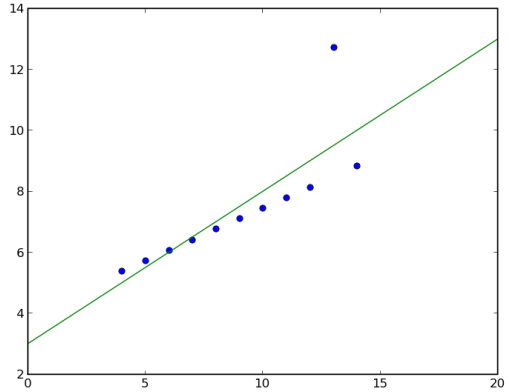

Data set 1, Mean = 7.50090909091, Median = 7.24
Linear fit
14
12
10
8
6
4
2
0
5
10
15
20

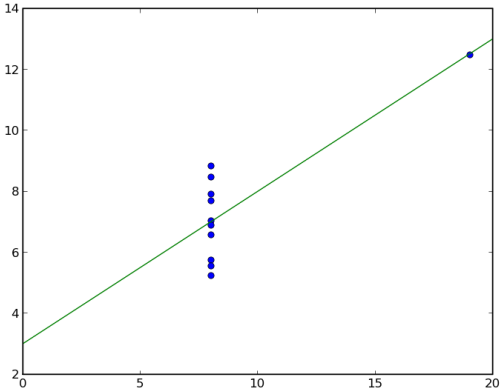

Data set 1, Mean = 7.50090909091, Median = 7.24
Linear fit
14
12
10
8
6
4
2
0
5
10
15
20

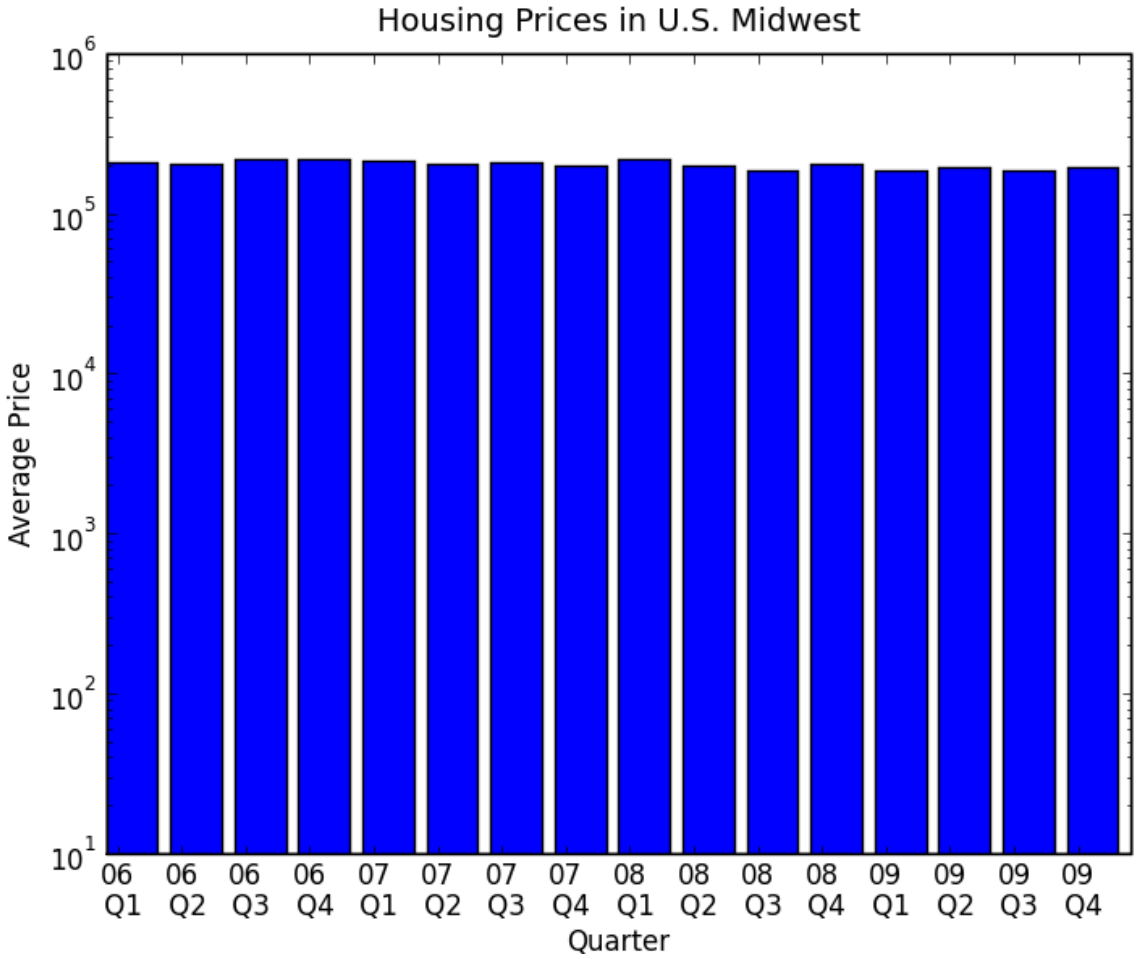

Data set 1, Mean = 7.50090909091, Median = 7.24
Linear fit
14
12
10
8
6
4
2
0
5
10
15
20

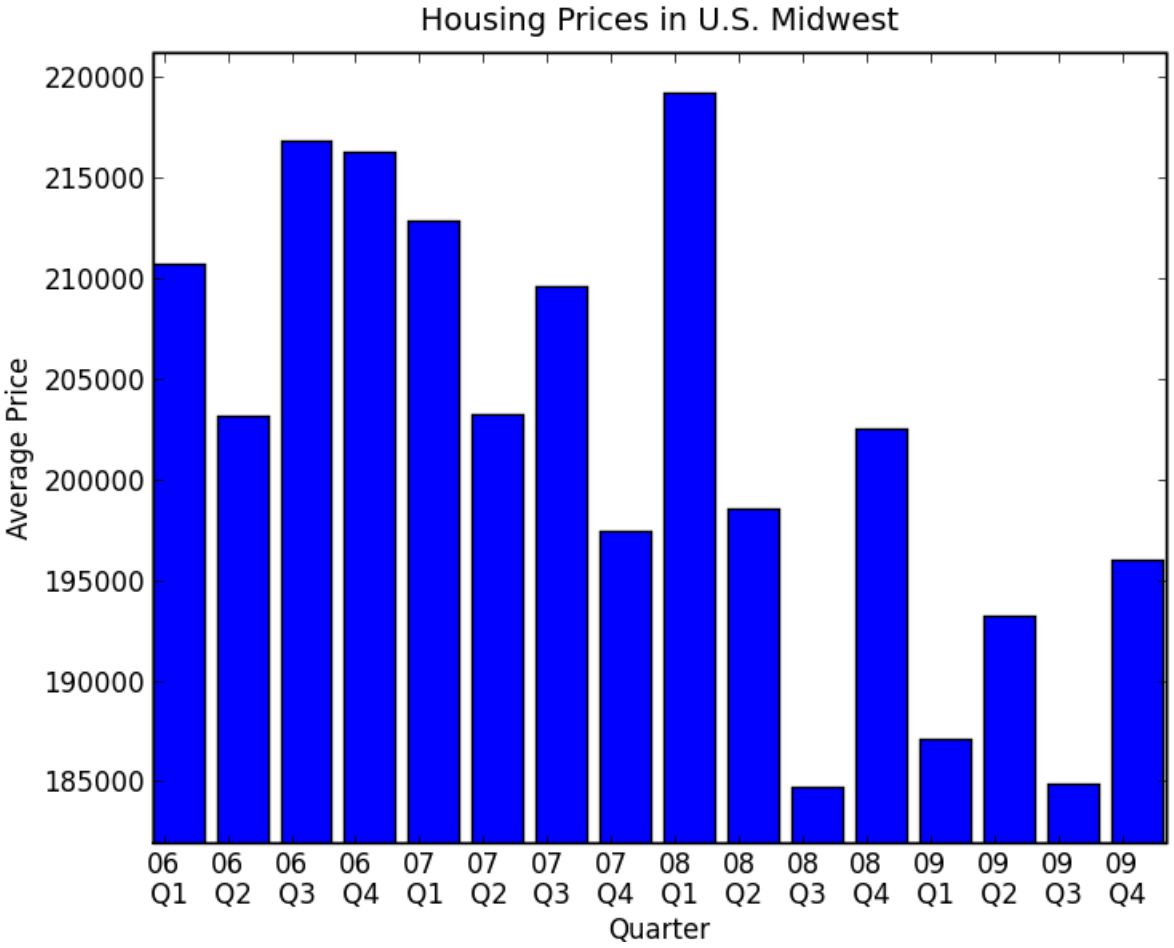

14
12
10
8
6
4
2
0
5
10
15
20

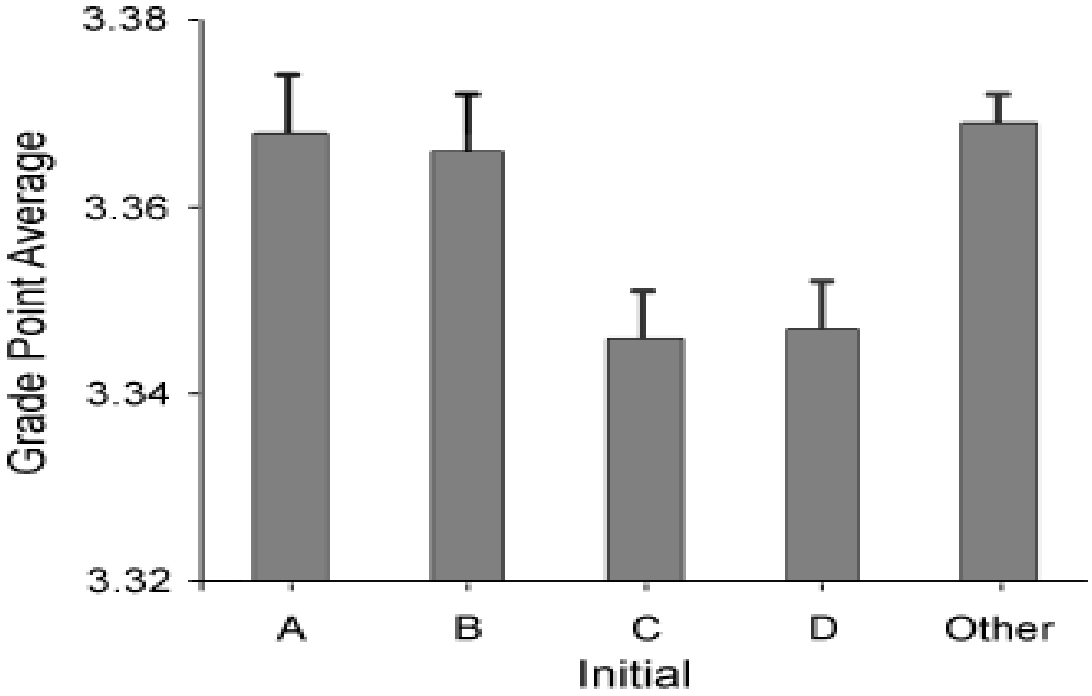

14
12
10
8
6
4
2
0
5
10
15
20


14
12
10
8
6
4
2
0
5
10
15
20


14
12
10
8
6
4
2
0
5
10
15
20


Housing Prices in U.S. Midwest
Average Price
10^6
10^5
10^4
10^3
10^2
10^1
06 Q1
06 Q2
06 Q3
06 Q4
07 Q1
07 Q2
07 Q3
07 Q4
08 Q1
08 Q2
08 Q3
08 Q4
09 Q1
09 Q2
09 Q3
09 Q4
Quarter


Housing Prices in U.S. Midwest
Average Price
220000
215000
210000
205000
200000
195000
190000
185000
06 Q1
06 Q2
06 Q3
06 Q4
07 Q1
07 Q2
07 Q3
07 Q4
08 Q1
08 Q2
08 Q3
08 Q4
09 Q1
09 Q2
09 Q3
09 Q4
Quarter



Nelson and Simmons, “Moniker Maladies,” Psychological Science, Vol. 18, No. 12, 2007.