Linear Regression

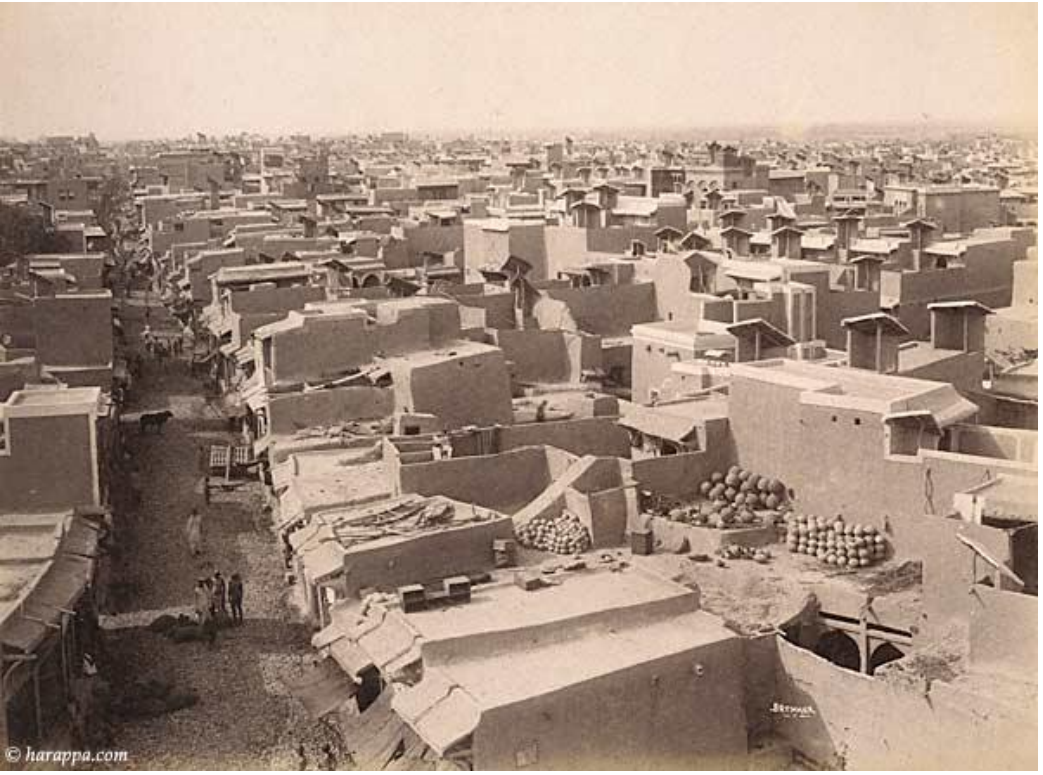
The Descriptive Angle - II

What have we seen thus far?

1. **Choose** X variables
2. **View** scatter plots (careful!)
3. **Fit** a simple/multiple regression model to data. Get estimated model
4. Check **validity** of model assumptions
5. Use estimated model to **test/infer** relationship in the population

Example: Housing prices in MidCity

Housing Prices.xlsx :
128 recent sales of single-family houses in MidCity



Price: Final sale price

SqFt: Floor area in ft^2

Bedrooms: # bedrooms

Bathrooms: # bathrooms

Offers: # offers made on the house prior to sale

Brick: Brick construction? (yes/no)

Neighborhood: East/West/North

Explanatory Objective:

Estimate and interpret the pricing structure of houses in MidCity

Data sample

Home	Price	SqFt	Bedrooms	Bathrooms	Offers	Brick	Neighborhood
1	114300	1790	2	2	2	No	East
2	114200	2030	4	2	3	No	East
3	114800	1740	3	2	1	No	East
4	94700	1980	3	2	3	No	East
5	119800	2130	3	3	3	No	East
6	114600	1780	3	2	2	No	North
7	151600	1830	3	3	3	Yes	West
8	150700	2160	4	2	2	No	West
9	119200	2110	4	2	3	No	East
10	104000	1730	3	3	3	No	East
11	132500	2030	3	2	3	Yes	East
12	123000	1870	2	2	2	Yes	East
13	102600	1910	3	2	4	No	North
14	126300	2150	3	3	5	Yes	North
15	176800	2590	4	3	4	No	West
16	145800	1780	4	2	1	No	West
17	147100	2190	3	3	4	Yes	East
18	83600	1990	3	3	4	No	North
19	111400	1700	2	2	1	Yes	East
20	167200	1920	3	3	2	Yes	West

USE THE MODEL

Interpreting the output

The coefficient of SqFt is 52.99.

What is an **economic interpretation** of this number?

House #39 in the dataset has 1720 sq ft.,
3 bedrooms, 2 bathrooms and 1 offer. It is a non-brick house in the West neighborhood.
Suppose the list price is $131,300.

This house is

1. over-priced
2. under-priced
3. priced just right

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	598.9199219	9552.197266	0.95010996	2.17745E+12
SqFt	52.99374008	5.73424006	0	28036362240
Bedrooms	4246.793945	1597.910767	0.00893895	7992064000
Bathrooms	7883.27832	2117.0354	0.00030041	4272561664
Offers	-8267.48828	1084.776733	0	23712010240
Brick_Yes	17297.34961	1981.616333	0	7782121472
Neighborhood_North	1560.579224	2396.765381	0.51621467	97610600
Neighborhood_West	22241.61719	2531.758301	0	7746974720

Statistical Inference: Effect of house size on price

Is the coefficient of SqFt statistically significant?

What does this mean?

Five years back, each additional sq foot of floor area added approximately $50 to the value of the house, after controlling for the #bedrooms, #bathrooms and #offers made.

Has the coefficient of SqFt increased significantly from its value five years back?

1. Yes
2. No

Measure explanatory power

What is the proportion of the variation in house prices explained jointly by *SqFt, Bedrooms, Bathrooms, Offers, Brick and Neighborhood*?