Multiple Regression

Idea: Predict numerical variable using several independent variables

Examples

- university performance dependent on general intelligence, high school grades, education of parents,...
- income dependent on years of schooling, school performance, general intelligence, income of parents,...
- level of language ability of immigrants depending on
  - leisure contact with natives
  - age at immigration
  - employment-related contact with natives
  - professional qualification
  - duration of stay
  - accommodation

Regression Techniques Attractive

- allow prediction of one variable value based on one **or more** others
- allow an **estimation of the importance** of various independent factors (cf. ANOVA)

- Normally, dependent variable is numeric. If dependent variable is categorical, multiple LOGISTIC regression is possible.
- Additional point: we'll also examine what happens when one variable is not in a linear scale (transformation is needed).

Not very popular in linguistics, but perhaps it should be.

Models

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \epsilon$$

We'll focus on the case where $n = 2$, others similar.

Questions: which x_i contribute to the explanation of y?

Some answers are arbitrary, viz., those where x_i, x_j **compete** in the explanation of y. There may be no single model which explains the facts best.

We need to examine models with this in mind.

Models for Two Independent Variables

$$y = \epsilon$$
$$y = \beta_0 + \epsilon$$
$$y = \beta_0 + \beta_1 x_1 + \epsilon$$
$$y = \beta_0 + \beta_2 x_2 + \epsilon$$
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

What independent factors, taken together or separately, explain the dependent variable the best?

Interactions

Multiple regression is logically more complicated than simple regression applied several times.

COLLINEARITY: Independent variables may correlate themselves, competing in their explanation. Result: fewer variables are useful in combined models.

SUPPRESSION: An independent variable may appear **not** to be explanatory until it is applied only to the residuals of another variable. Result: initially insignificant variable becomes significant in combined model.

RuG

An example

Peter Trudgill suggests that language varieties may be subject to a “gravity law”, being attracted to one another in a way like the way planets are attracted to the sun.

$$F = G\frac{m_1 m_2}{r^2}$$

F is the force due to gravity,

m_1, m_2 the masses of the two objects attracting each other,

r the distance between them, and

G is a “universal gravitational constant.”

Linguistic Cohesion via Gravity

$$F = G\frac{m_1 m_2}{r^2} = G\frac{p_1 p_2}{r^2}$$

F is the attractive force,

m_1, m_2 the populations of the two settlements,

r the distance between them, and

G won't be speculated on

Idea: social contact promotes linguistic accommodation and linguistic similarity.

Chance of social contact should be

- proportional to the product of settlement size and
- (if travel is random) inversely proportional to squared distance

Measuring Linguistic Distance

Nerbonne, Heeringa et al. have developed a string distance measure that applies to dialect pronunciations.

- numerical, therefore can be summed, averaged
- validated against consensus expert opinion, also against lay dialect speakers impression of dissimilarity
- very reliable when applied to > 100 words

Idea: use distance to test the gravity hypothesis. Distance should be inversely related to the "attraction" postulated by Trudgill.

Segment Distance

- Sum feature distances in feature vectors to obtain segment distances.
  Example: d([i],[e]) $\ll$ d([i],[u])

	i	e	u	i-e	i-u
advancement	2(front)	2(front)	6(back)	0	4
high	4(high)	3(mid high)	4(high)	1	0
long	3(short)	3(short)	3(short)	0	0
rounded	0(not rounded)	0(not rounded)	1(rounded)	0	1
				1	5

- Diacritics [ĩ,e:,ə^r] can also be taken into account
- Different feature systems employed: Vieregge-Cucchiarini and also Almeida-Braun (both developed to measure accuracy of transcribers)
- Theoretical Chomsky-Halle (SPE) system less useful (clever features for making rules compact)

Levenshtein Distance

Cost of least costly set of operations mapping one string into another.

	Operation	Cost
æəf t ən ʉn		
æf t ən ʉn	delete ə	d(ə,[])=0.3
æf t ər n ʉn	insert r	d([],r)=0.2
æf t ər n u n	replace [ʉ] with u	d([ʉ],[u])=0.1
Total		0.6

Computing Levenshtein Distance

		æ	f	t	ə	r	n	u	n
	0	1	2	3	4	5	6	7	8
æ	1	0	1	2	3	4	5	6	7
ə	2	1	2	3	2	3	4	5	6
f	3	2	1	2	3	4	5	6	7
t	4	3	2	1	2	3	4	5	6
ə	5	4	3	2	1	2	3	4	5
n	6	5	4	3	2	3	2	3	4
ʉ	7	6	5	4	3	4	3	4	5
n	8	7	6	5	4	5	4	5	4

Here we simplify costs (everything $= 1$) for illustration.

Dialect Material

We apply the distance measure to dialect pronunciations of the same words, collected over a range of sites (settlements).

lopen — [lopə] vs. [lopm̩] vs. ...

Material originally collected for dialect atlases.

150 words in 52 places throughout the Saxon dialect area of the Netherlands.

Geographic distances and population sizes (1815) also collected.

Linguistic Cohesion via Gravity

$$F = G\frac{p_1 p_2}{r^2}$$

F is the attractive force,

p_1, p_2 the populations of the two settlements, and

r the distance between them

Notate bene: we measure linguistic dissimilarity, which we postulate stands in inverse relation to the attractive force of social contact.

Predictions of Linguistic Cohesion via Gravity

$$F = G\frac{p_1 p_2}{r^2} = 1/D$$

$$D \propto 1/G\frac{r^2}{p_1 p_2}$$

F is ling. attraction, which should produce similarity

D is ling. dissimilarity

p_1, p_2 the populations of the two settlements, and

r the distance between them

Linguistic Cohesion via Gravity

$$D \propto 1/G \frac{r^2}{p_1 p_2} \propto \frac{r^2}{p_1 p_2}$$

$$D \propto r^2 \text{ AND } D \propto -p_1 p_2$$

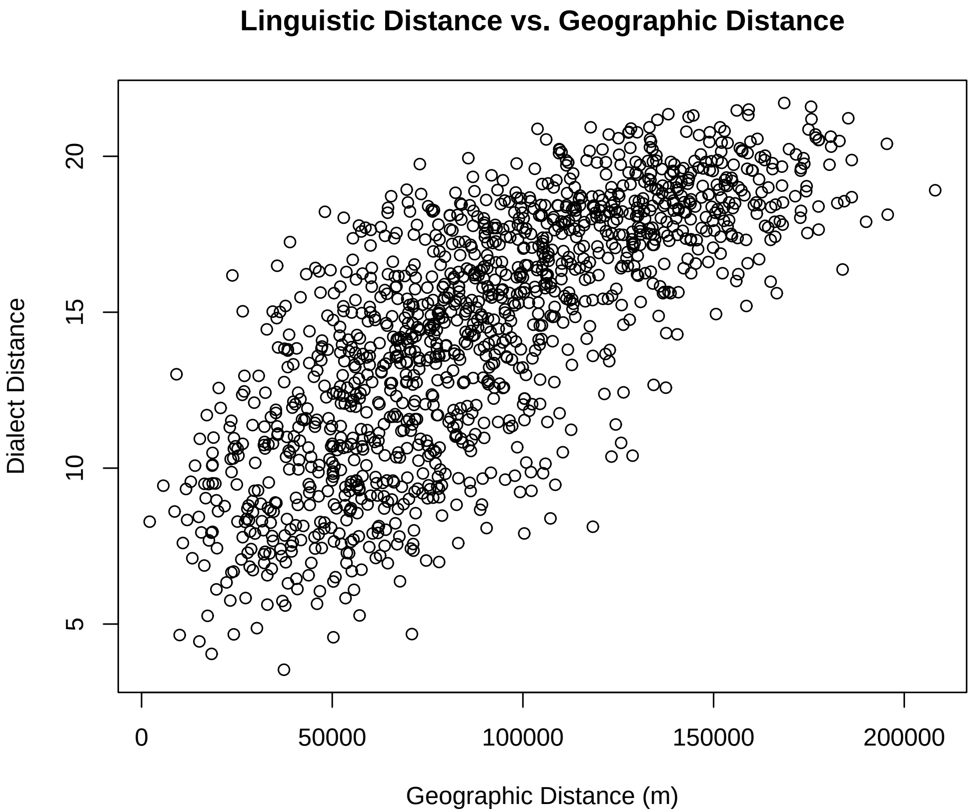
D is linguistic distance,

p_1, p_2 the populations of the two settlements, and

r the distance between them

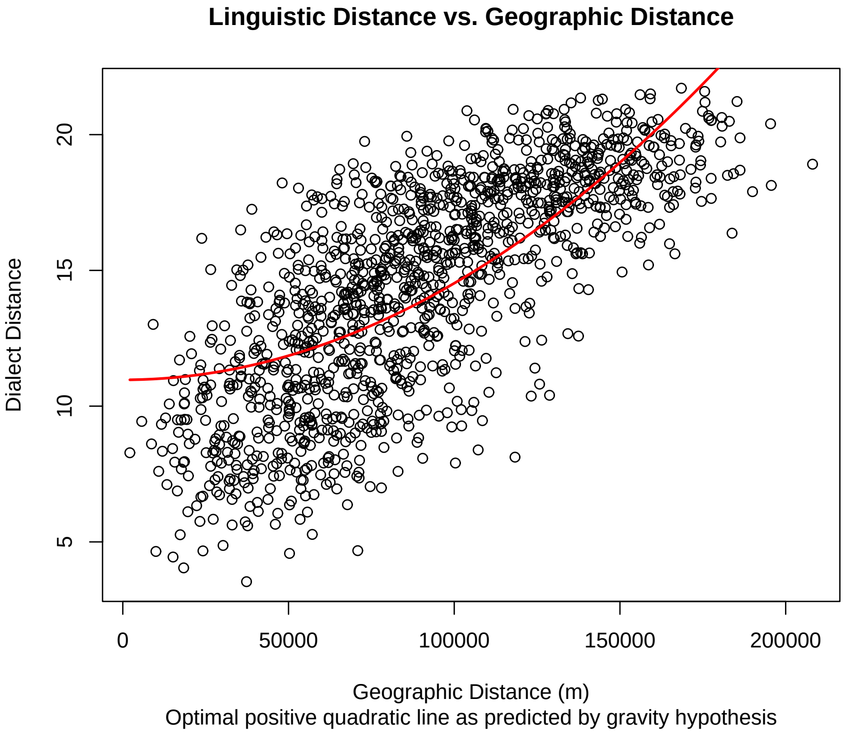
Notate bene: we measure linguistic dissimilarity, which we postulate stands in inverse relation to the attractive force of social contact.

Look at Data



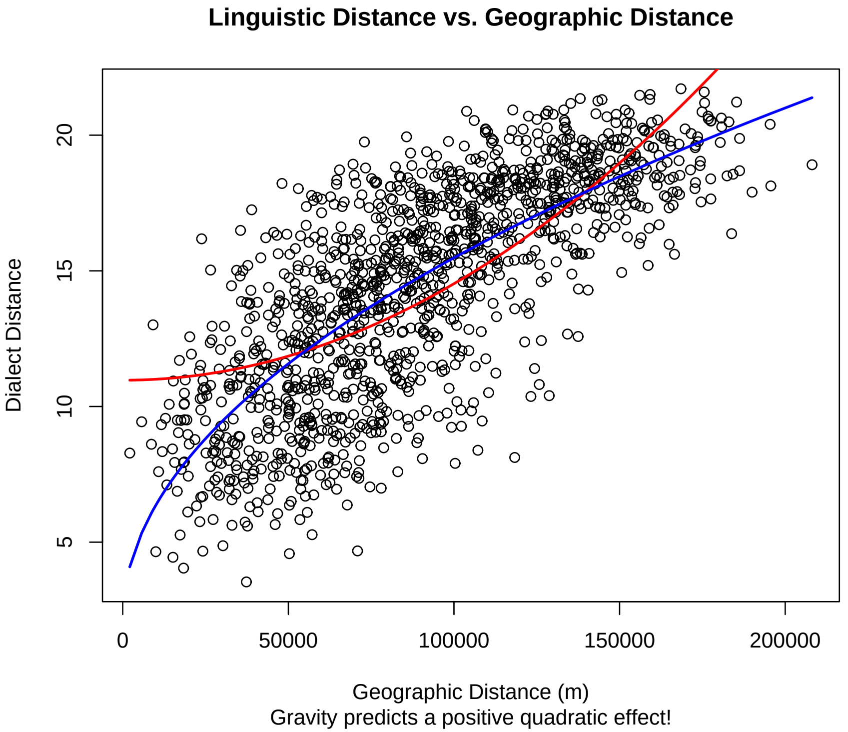
Gravity predicts a positive quadratic effect!

Quadratic?



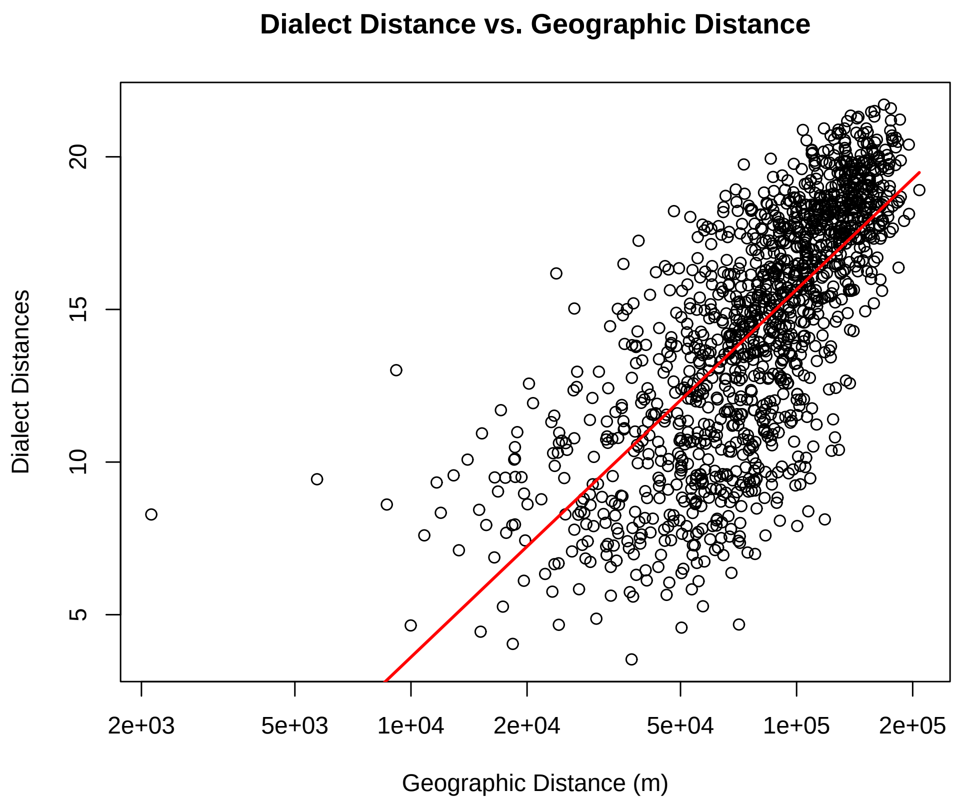
Shape? Zero? ($r^2 = 0.5$)

Function of $\sqrt{x}$?



Shape? Zero? ($r^2 = 0.57$)

Alternative view—logarithmic x-Axis



Pronunciation vs. Geography for 1300 Dialect Pairs

Interpreting Results

Trudgill's gravity model predicts that atttraction is relatively stronger over short distances. This implies that linguistic distances should be relatively smaller over these short distances.

Linguistic distance indeed increases positively with geographic distance, as Trudgill predicts, but the effect is proportionately **greater** over short distances rather than proportionately smaller, as gravity predicts.

Note that this is what one would expect if the fundamental force were not attraction, as Trudgill postulates, but rather repulsion/fission/differentiation. It would be natural to see this grow realtively weaker over long distances.

Effect of Population

$$D \propto \frac{r^2}{p_1 p_2}$$

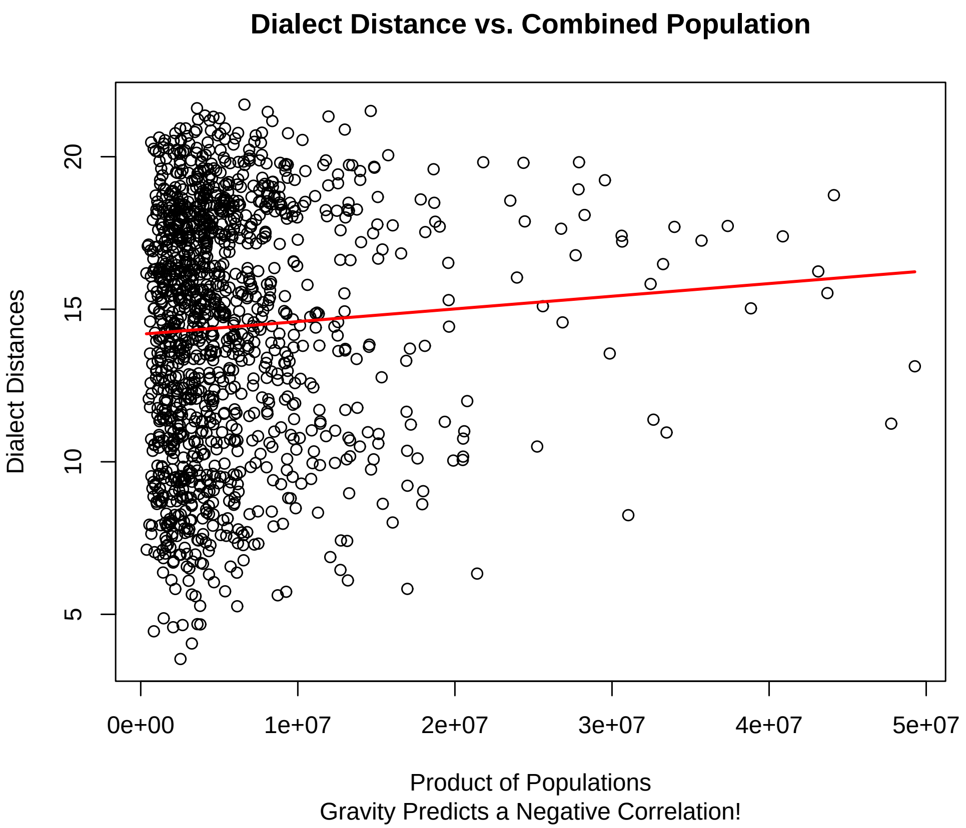
$$D \propto r^2 \text{ AND } D \propto -p_1 p_2$$

Prediction: negative correlation of linguistic distance with product of population sizes.

First view of data: possibly influential points (extreme x values). We examine data with and without these points. Little difference.

Dialect Distance vs. Population

Inf. Stats



Dialect Distance vs. Population

- uneven spread in x direction (population)
- unexpected **positive** correlation with linguistic distance
- $r = 0.06, r^2 = 0.0036$ —little explanatory power
- positive correlation could be interpreted as an indication of fundamental **repelling** forces. Cf. effect of distance.

Summarizing Individual Effects

Gravity predicts:

$$D \propto \frac{r^2}{p_1 p_2}$$

$$D \propto r^2 \text{ AND } D \propto -p_1 p_2$$

Results

- positive correlation between D and r (dialect distance and geographic distance); but $D \propto \sqrt{r}$
  —we use $\sqrt{r}$ for geographic distance below
- unexpected **positive** correlation between linguistic distance and population size

Combined model?

Toward a Combined Model

Check on possible collinearity:

- correlation among explanatory variables?
  —conceptually unlikely, but test!
- calculate correlation

		geo.-dist.	1800 pop. product
geo.-dist.	Pearson r	1,0	,056(*)
Sig. (2-tailed)			,041
1800 pop. prod.	Pearson r	,056(*)	1,0
Sig. (2-tailed)		,041	

Surprise! This could reduce the effectiveness of the second variable in the model.

Combined Model


```
Variables Entered/Removed(a)
```

Model	Variables Entered	Variables Removed
1	geographic distance	.
2	1800 populations’ product	.

```
a Dependent Variable: phonetic distance

Stepwise (Criteria: Probability-of-F-to-enter <= ,050,
                    Probability-of-F-to-remove >= ,100).
```

We use SPSS **stepwise** in order to compare increasingly complex models.

enter builds the complex model all at once.
(forward) stepwise builds the model, one variable at a time
backward (stepwise) builds the complex model, then eliminates one variable at a time

RuG

Two Models

Model Summary

Model	R	R Square	Std. Error of Estimate
1	,755(a)	,571	2,60688
2	,758(b)	,574	2,59702

a Predictors: (Constant), geographic distance
b Predictors: (Constant), geographic distance, 1800 populations' product

No SUPPRESSION effect. That is, population is not more significant that we expected based on viewing it separately.

ANOVA Tables in Multiple Regression

Source	Degrees Freedom	Sum of Squares	Mean Squares	F
Model	p	$\sum(\hat{y}_i - \bar{y})^2$	SSM/DFM	MSM/MSE
Error	$n-p-1$	$\sum(y_i - \hat{y}_i)^2$	SSE/DFE	
Total	$n-1$	$\sum(y_i - \bar{y})^2$	SST/DFT	

where p is the number of variables in the model.

If the mean residue in the model is large wrt the mean residue with no model (large F), then the model is doing worthwhile work. Note that we're comparing the model to a "dumb model" in which the expected value is just the mean value of all y's.

Combined Model

ANOVA(c)

Model		Sum Sqrs	df	Mean Sqr	F	Sig.
1	Regr	11955,8	1	11955,8	1759	,000(a)
	Resi	8997,6	1324	6,8		
	Tota	20953,5	1325			
2	Regr	12030,5	2	6015,3	891	,000(b)
	Resi	8923,0	1323	6,7		
	Tota	20953,5	1325			

a Predictors: (Constant), geo-dist
b Predictors: (Constant), geo-dist, 1800 populations' product
c Dependent Variable: phonetic distance

Regression Equation

Coefficients(a)

Model		Coefficients		Stnd. Coeff.	t	Sig.
		B	Std. Err	Beta		
1	(Constant)	2,148	,302		7,1	,000
	geo-dist	,042	,001	,755	41,9	,000
2	(Constant)	2,018	,303		6,7	,000
	geo-dist	,042	,001	,756	42,1	,000
	1800 pop	1,6E-08	,000	,060	3,3	,001

a Dependent Variable: phonetic distance

t values reflect how likely the coefficients calculated for the sample would be if the coefficients in the population were 0.

Combined Model

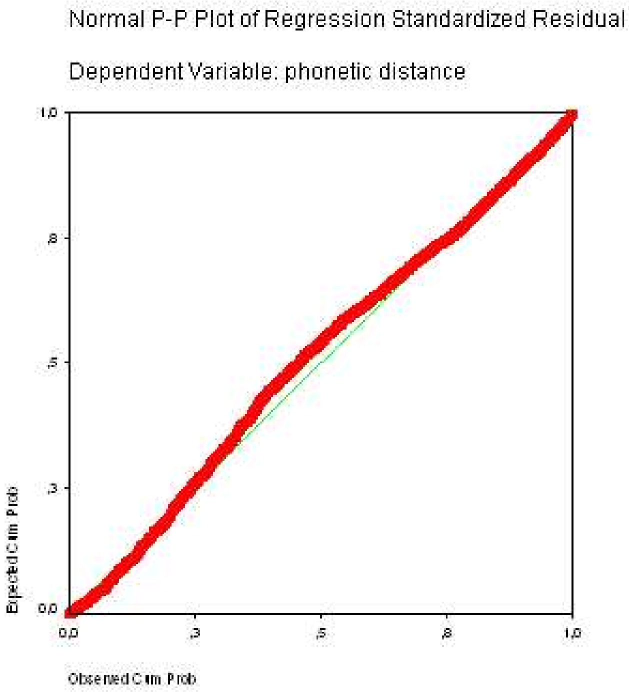
Excluded Variables(b)

Model		Beta In	t	Sig.	Part. Corr	Collinearity Tolerance
1	1800 pop	,060(a)	3,327	,001	,09	1,000

a Predictors in the Model: (Constant), geo-dist

b Dependent Variable: phonetic distance

Normally Distributed Residuals?



In multiple regression, we must check that residuals are roughly normally distributed.

Speculation about Repulsion

Coulomb formulated a law about the attraction and repulsion of charged particles.

$$F = k\frac{q_1 q_2}{r^2}$$

F is the attractive/repellent force due to electrical charge,

q_1, q_2 the charge of the particles attracting/repelling each other,

r the distance between them, and

k is a “constant.”

Where like charges are involved, repulsion obtains:

$$D \propto \frac{q_1 q_2}{r^2}$$

Speculation

$$D \propto \frac{q_1 q_2}{r^2}$$

$$D \propto 1/r^2 \text{ AND } D \propto p_1 p_2$$

This model also doesn't work well—D does not correlate negatively with r or r^2. Furthermore, but the contribution of population size remains minimal.

Real estate agents claim that there are three factors determining the value of a house: "Location, location, and location."