

# Describing Bivariate Relationships

17.871

Spring 2004

# Measures of association

- Discrete data
  - $\chi^2$
  - Gamma, Beta, etc.
- Continuous or discrete data
  - (Pearson) correlation coefficient
  - (Spearman) rank-order correlation coefficient

# Example I

- What is the relationship between religion and abortion sentiments?
- The abortion scale:
  1. BY LAW, ABORTION SHOULD NEVER BE PERMITTED.
  2. THE LAW SHOULD PERMIT ABORTION ONLY IN CASE OF RAPE, INCEST, OR WHEN THE WOMAN'S LIFE IS IN DANGER.
  3. THE LAW SHOULD PERMIT ABORTION FOR REASONS OTHER THAN RAPE, INCEST, OR DANGER TO THE WOMAN'S LIFE, BUT ONLY AFTER THE NEED FOR THE ABORTION HAS BEEN CLEARLY ESTABLISHED.
  4. BY LAW, A WOMAN SHOULD ALWAYS BE ABLE TO OBTAIN AN ABORTION AS A MATTER OF PERSONAL CHOICE.

# Theoretical distribution of cells with abortion and religion independent

	Abortion opinion				
Religion	1	2	3	4	Total
Protestant	.0796	.2096	.1214	.2298	.6640
Catholic	.0384	.1012	.0586	.1109	.3092
Jewish	.0015	.0039	.0022	.0042	.0118
Orthodox	.0006	.0015	.0009	.0017	.0047
Non-Xn/Jewish	.0015	.0039	.0022	.0042	.0118
Other	.0004	.0010	.0006	.0011	.0031
Total	.1243	.3273	.1896	.3588	1.000

# Actual and theoretical distribution

Religion	Abortion opinion				Total
	1	2	3	4	
Protestant	(101.2) <b>118</b>	(226.4) 274	(154.3) 160	(292.0) 292	844
Catholic	(48.8) <b>38</b>	(128.6) 133	(74.5) 75	(141.0) 141	387
Jewish	(1.9) <b>0</b>	(4.9) <b>0</b>	(2.8) <b>0</b>	(5.4) <b>15</b>	15
Orthodox	(0.7) <b>0</b>	(2.0) <b>3</b>	(1.1) <b>1</b>	(2.1) 2	6
Non-Xn/Non-Jewish	(1.9) 2	(4.9) <b>3</b>	(2.8) <b>4</b>	(5.4) <b>6</b>	15
Other	(0.5) <b>0</b>	(1.3) <b>3</b>	(0.7) <b>1</b>	(1.4) <b>0</b>	4
Total	158	416	241	456	1271

$$\chi^2=38.2$$

# Example II

- What is the relationship between income and newspaper reading?

# Stylized relationship if newspaper reading increases with income

	Income			
Readership	Low	Med.	High	
Never				
Sometimes				
Daily				
Total				

# Actual relationship between newspaper reading and income

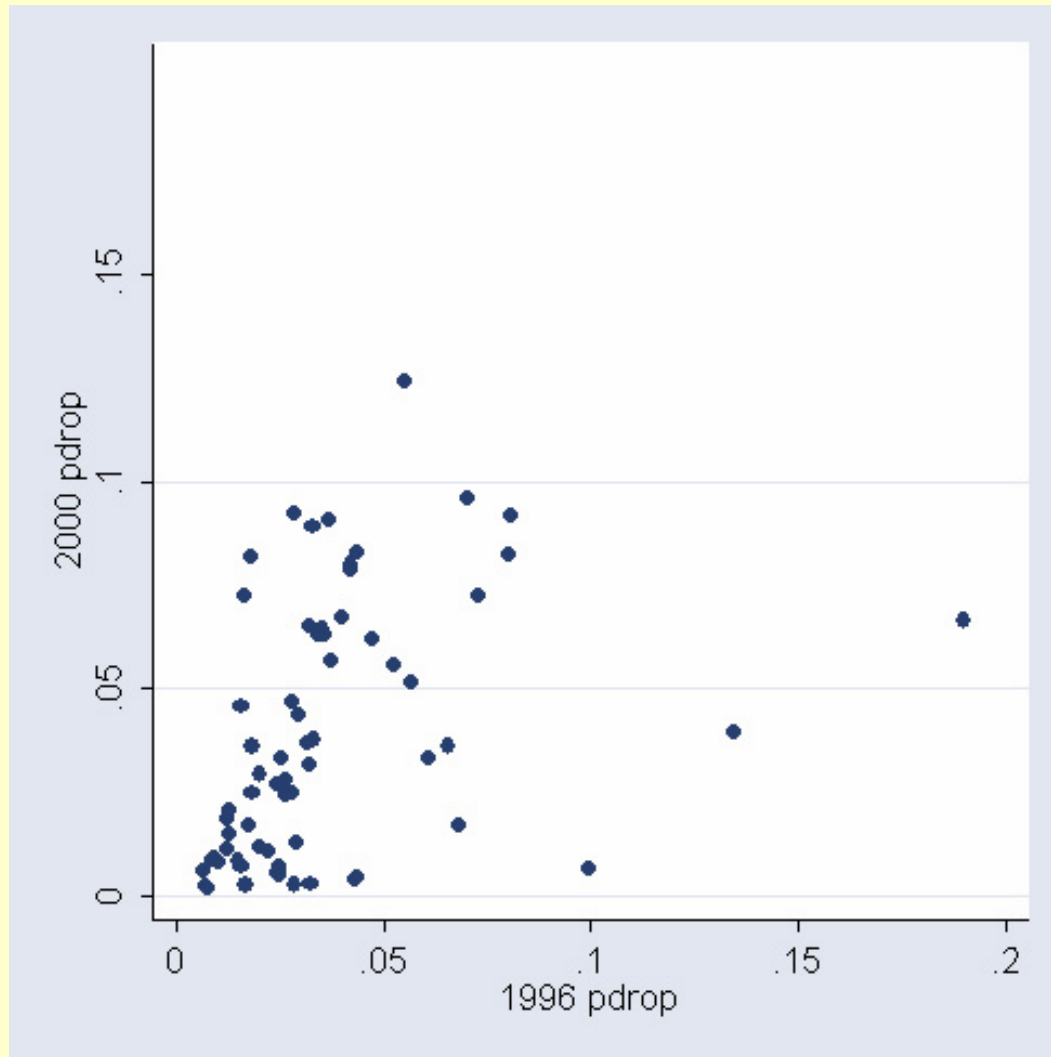
	Income			
Readership	≤\$65k	\$65K-\$125K	>\$125K	Total
0-1/week	440	126	118	684
2-6/week	291	124	98	513
Daily	311	154	145	610
Total	1042	404	361	1807

$\chi^2=25.0$ , Gamma = +.16

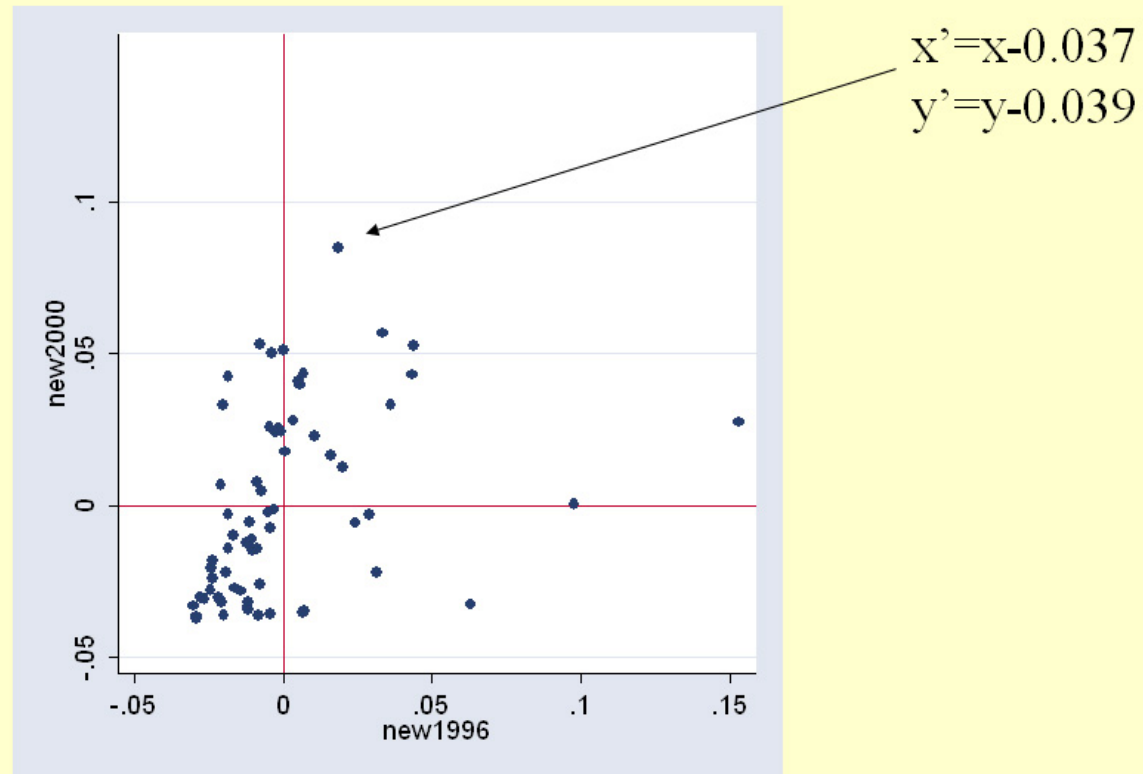
# Example III

- What is the relationship between residual vote in Florida in 1996 vs. 2000?

# 2000 Prez. Residual Vote vs. 1996 Prez. Residual Vote

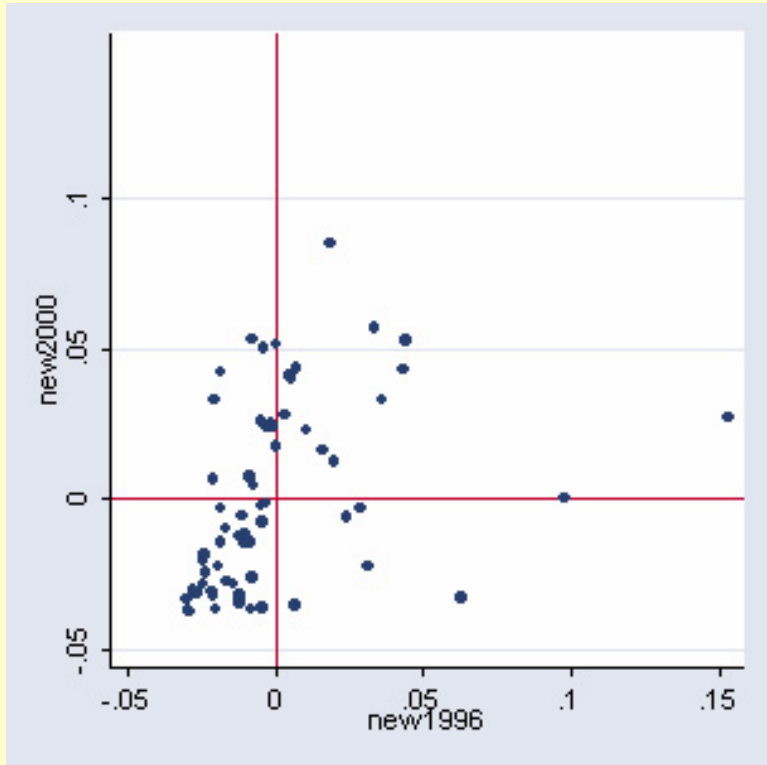


# Subtract each observation from its mean



# Covariance formula

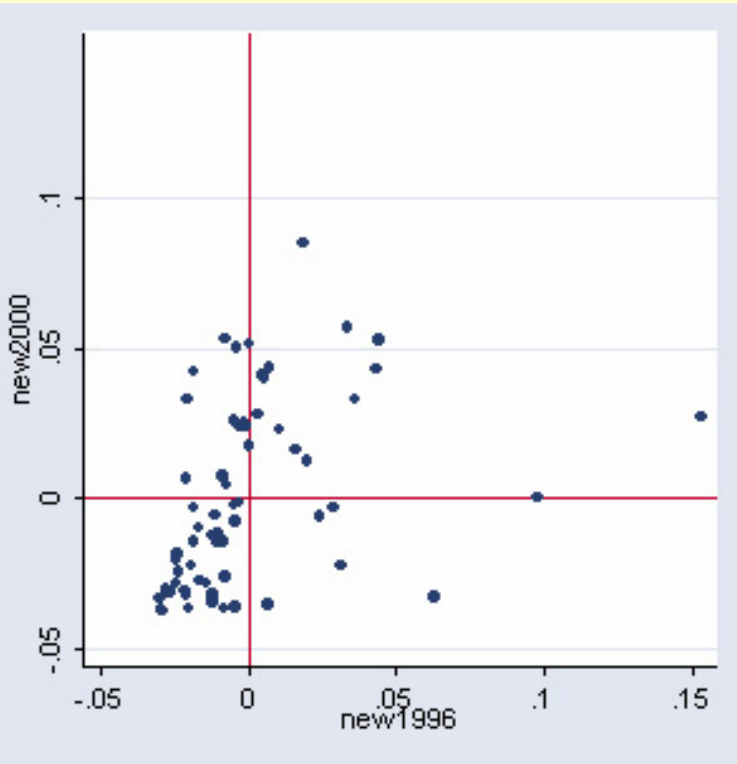
$$\text{Cov}(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n}$$



$$\text{Cov}(\text{RV}_{96}, \text{RV}_{00}) = 0.000346$$

# Correlation formula

$$\text{Corr}(x, y) = \frac{\text{Cov}(x, y)}{\sigma_x \sigma_y} = r$$



$$\text{Corr}(\text{RV}_{96}, \text{RV}_{98}) = .36$$

c.f. Gujarati, pp. 85-86  
Equation 3.5.12

# Playing with P.B. Stark's Correlation Java Script

<http://www.stat.berkeley.edu/users/stark/Java/Html/Correlation.htm>

# Guessing Correlations Applet

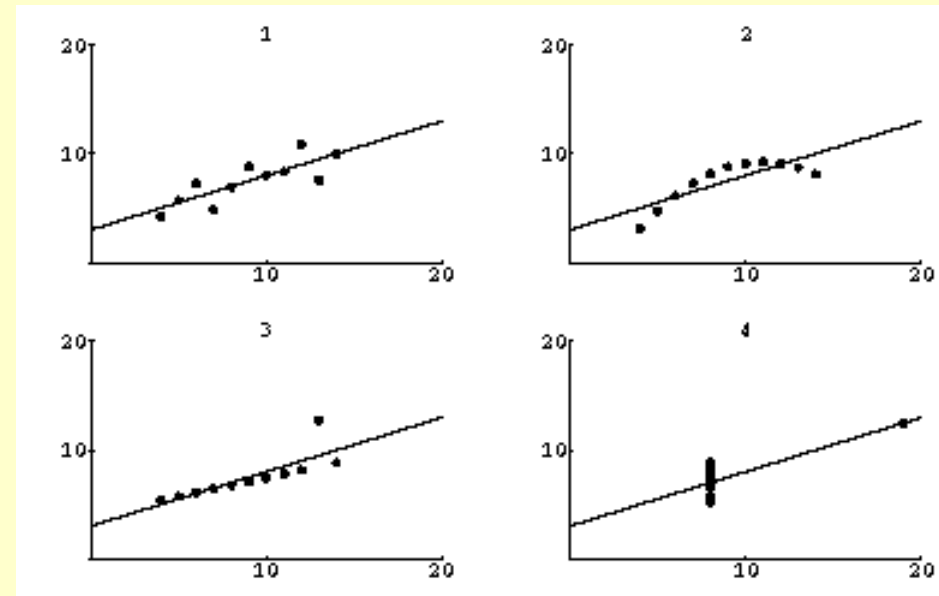
<http://www.stat.uiuc.edu/~stat100/java/guess/GCApplet.html>

# Warning:

- Correlation only measures linear relationship

# Anscombe's Quartet

I		II		III		IV	
x	y	x	y	x	y	x	y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89



$$r = .8164$$