Correlation

Kin 304W

Week 7: February 19, 2013

But first…your Project Part I

- Due at beginning of lecture Tuesday, February 26

- Read the Project Part I Instructions carefully (3 page .pdf on the website)

- Review the Project Part I Slides carefully (week #5)

- Refer to the 4 sample articles posted on the website (week #5 readings) to see examples of results sections (text, tables, and figures).

- Refer to lecture slides from week #4 (Tests of Differences II, 2-Way Mixed ANOVA) for examples of text and figures.

- Refer to http://abacus.bates.edu/~ganderso/biology/resources/writing/HTWtoc.html

- Perveen Biln (Writer Marker) has office hours Feb 21 and Feb 25 (see course website for times and location)

- Also, volunteer opportunities for FH Mobile Falls Clinic and HealthySteps

Project Part I

- Objectives of the project are to determine:
    1. The effect of a mobile falls prevention clinic on **fall risk** (PPA score) relative to health education in community-dwelling older adults over 6 months?
    2. The effect of a mobile falls prevention clinic on **physical activity** (daily minutes of moderate-to-vigorous physical activity) relative to health education in community-dwelling older adults over 6 months?

- Components of the writing assignment
    1. Title page
    2. Text of Results Section
    3. Table 1
    4. Figure 1

Project Part I: Text of Results Section

- 200-250 words of text.

- Para 1: The Results section should begin with text and specifically with a description of the study participants.

  - Include reference to the Table.

- Para 2 & 3: Continue by reporting key results from your statistical analysis.

  - Include reference to the Figure.

Project Part I: Table 1

- Display participant characteristics separately for Mobile Falls Clinic, Health Education, and overall.

- Include the total sample size for each group.

- For continuous variables, report means and standard deviations (SD) and p-values.

- For categorical variables, report the number of participants (N) and the corresponding %. You don't need to report p-values.

- Include the following participant characteristics: age, sex, body mass index, number of medications, hypertension, and university education or greater.

- Refer to Table 1 in Liu-Ambrose et al. as an example (see week #5 readings on website)

Table 1. Characteristics of men in the Osteoporotic Fractures in Men (MrOS) study by life-space (LS)

	Life-Space Score			
	LS 0-20	LS 21-40	LS 41-60	*P*
	(N= 37)	(N= 204)	(N= 373)	
Age, y	85.6 (6.5)	83.2 (5.8)	81.7 (5.7)	<0.01
BMI, kg/m^2	29.5 (6.0)	27.3 (5.2)	27.1 (4.4)	0.01
Gait speed, m/s	0.69 (0.25)	0.80 (0.22)	1.0 (0.24)	<0.01
White, non-Hispanic, N (%)	32 (86.5)	178 (87.3)	311 (83.4)	<0.01
Married, N (%)	21 (56.8)	134 (65.7)	253 (67.8)	<0.01
Excellent/good self-rated health, N (%)	25 (67.6)	118 (57.8)	263 (70.9)	<0.01
Any ADL limitation, N (%)	30 (81.1)	174 (85.3)	207 (55.5)	<0.01
Depression, GDS $\geq$ 6, N (%)	11 (29.7)	59 (28.9)	49 (13.2)	<0.01
Number of medical conditions, N (%)	2.9 (1.7)	3.0 (1.9)	2.6 (1.7)	<0.01

Abbreviations: ADL, activities of daily living; GDS, geriatric depression scale.
P values from one-way ANOVA.

Project Part I: Figure 1

- Display mean PPA values and mean MVPA values for both treatments (Mobile Falls Clinic, Health Education)

  – Figure 1 will have two separate plots

- Include error bars to show the standard errors associated with each mean.

- For examples, see lecture slides from week #4 (Tests of Differences II).

- Excel example coming up…

Now, onto Correlation…

Outline

- Pearson Correlation Coefficient, r

- Limitations of r

- Significance of r

- Coefficient of Determination, R^2

Pearson Correlation Coefficient (*r*)

- Pearson Correlation Coefficient (*r*) is a measure of the linear association between two variables

- Varies from -1 to +1

- Conceptually, *r* is a ratio of variability in X_1 to that of X_2.

  ○ 0 = no relationship

  ╱ -1 or 1 = perfect relationship

$$r = \frac{\sum x_1 x_2}{\sqrt{(\sum x_1^2)(\sum x_2^2)}}$$

$$x_1 = X_1 - \overline{X}_1$$
$$x_2 = X_2 - \overline{X}_2$$
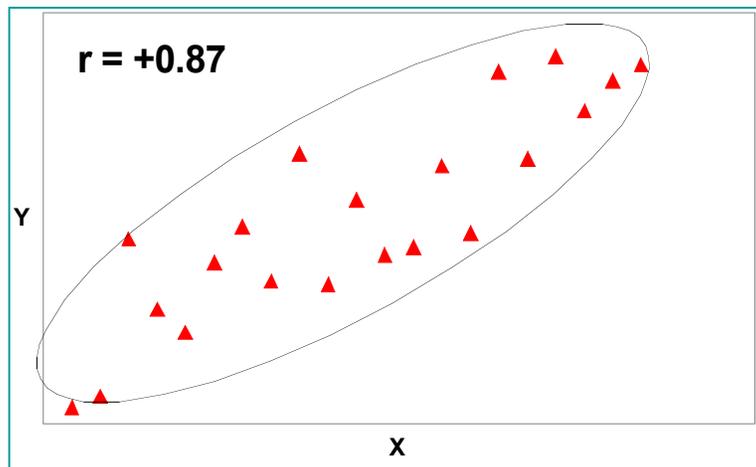
Pearson Correlation Coefficient

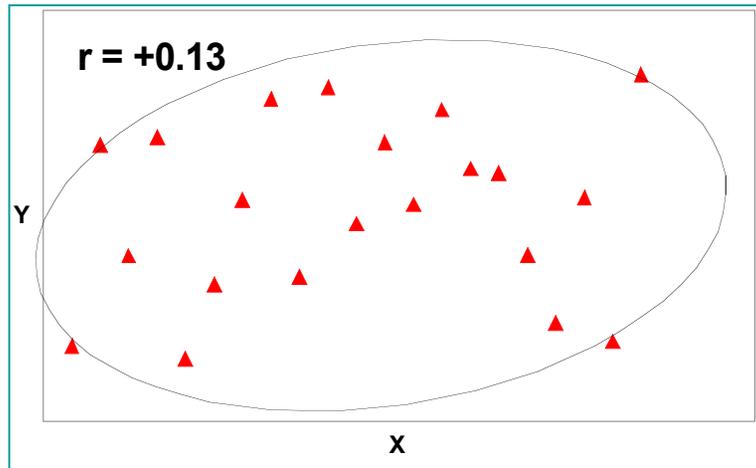## X and Y are perfectly correlated

r = +1

Y

x
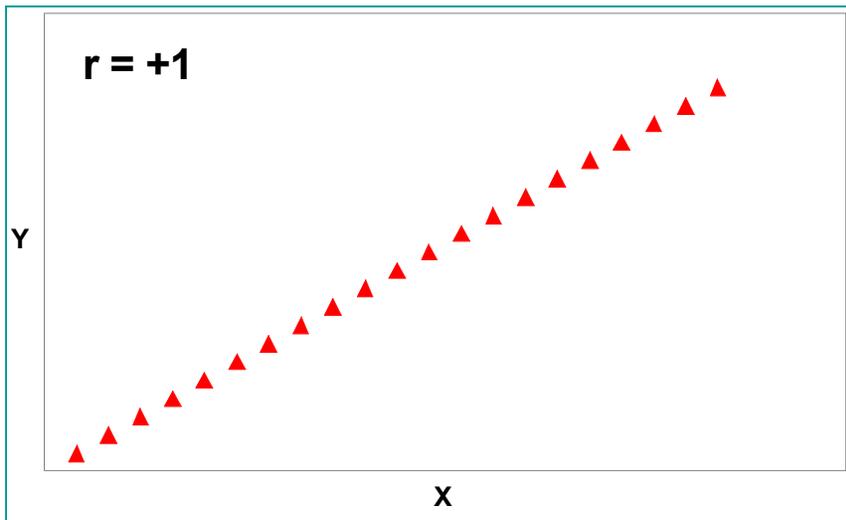
Positive correlation
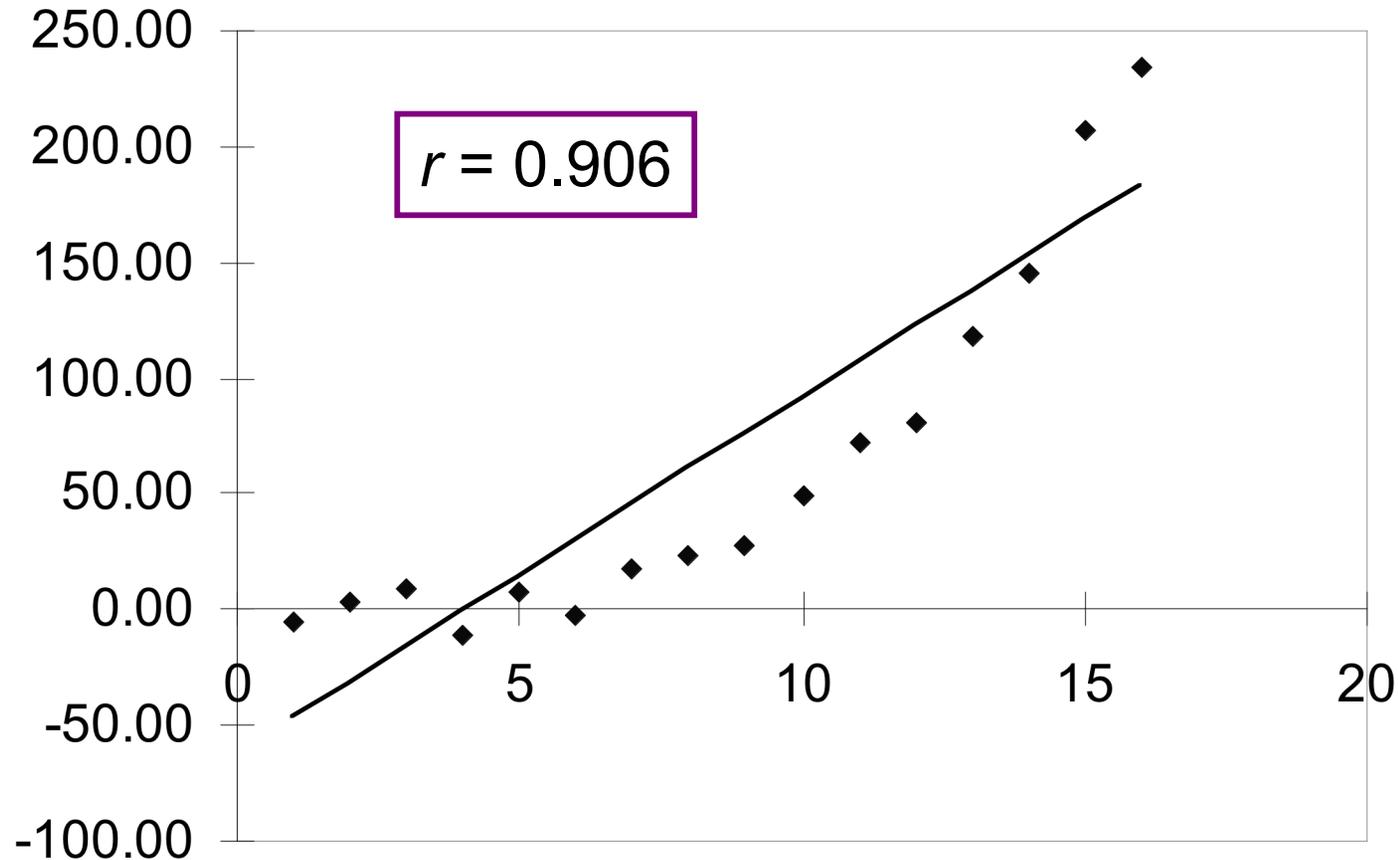
r = -1

Y

x
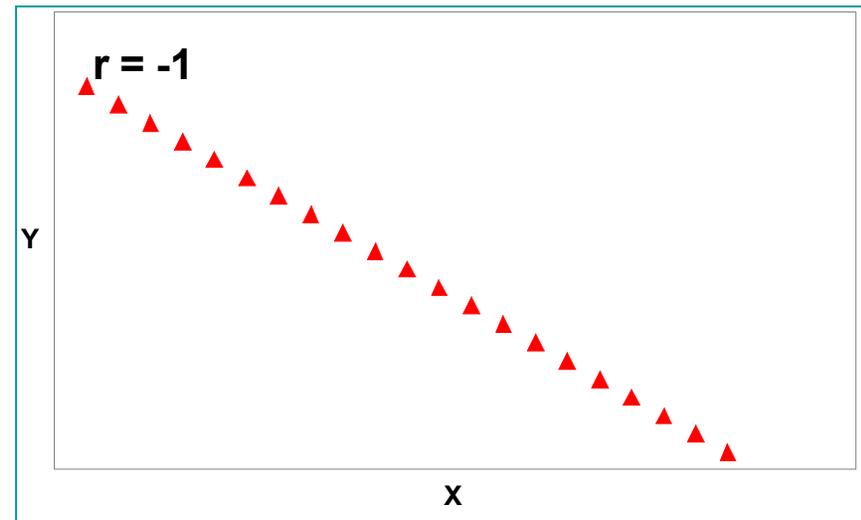
Negative correlation
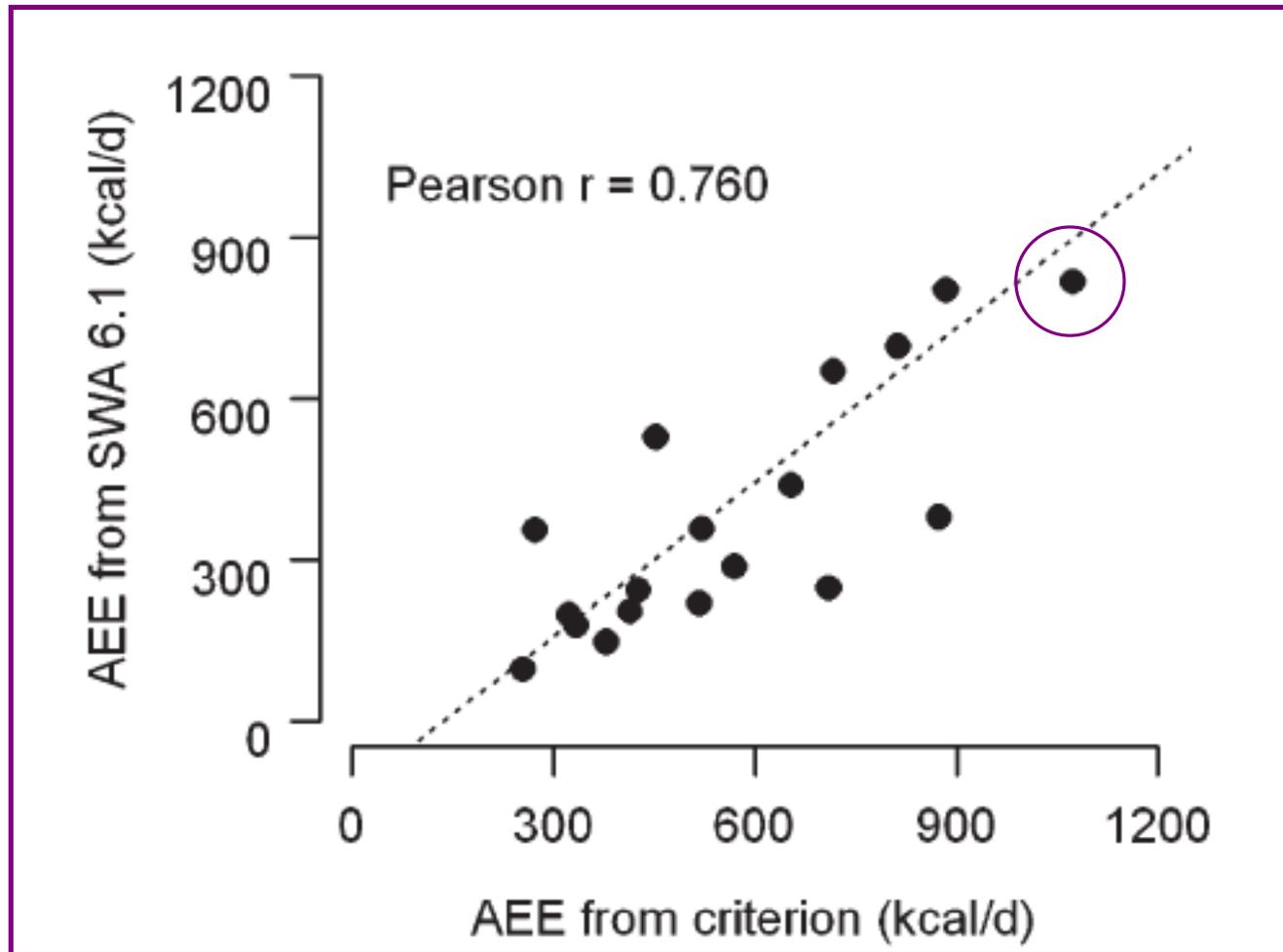
Pearson Correlation Coefficient
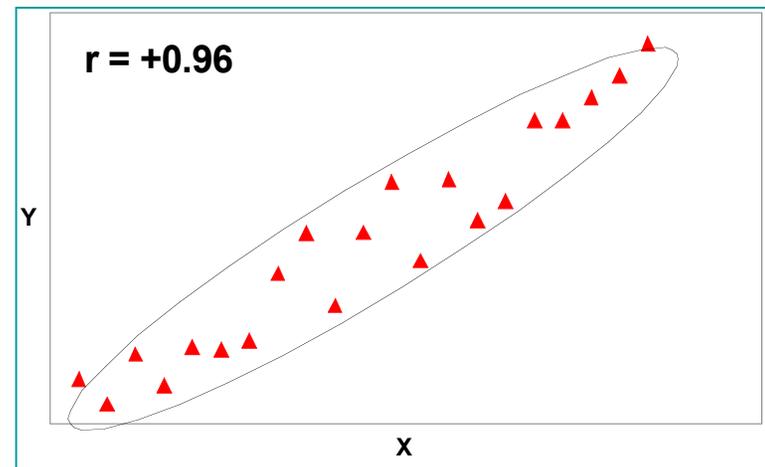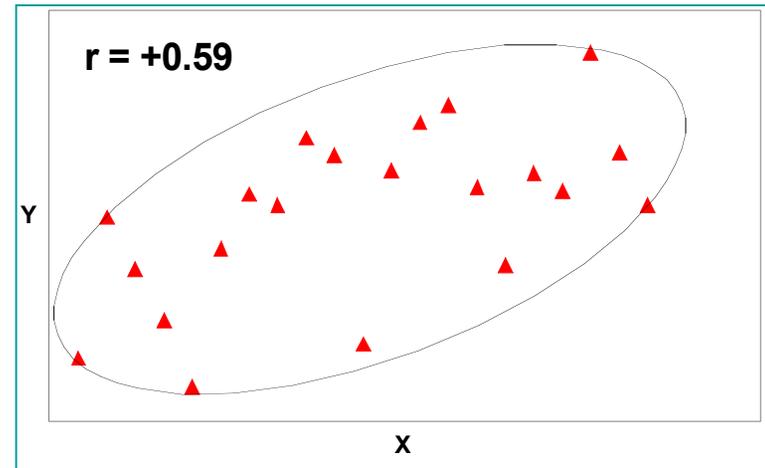
Pearson Correlation Coefficient

Be careful of common pitfalls of the correlation coefficient…

High correlation does not always mean there is a linear association
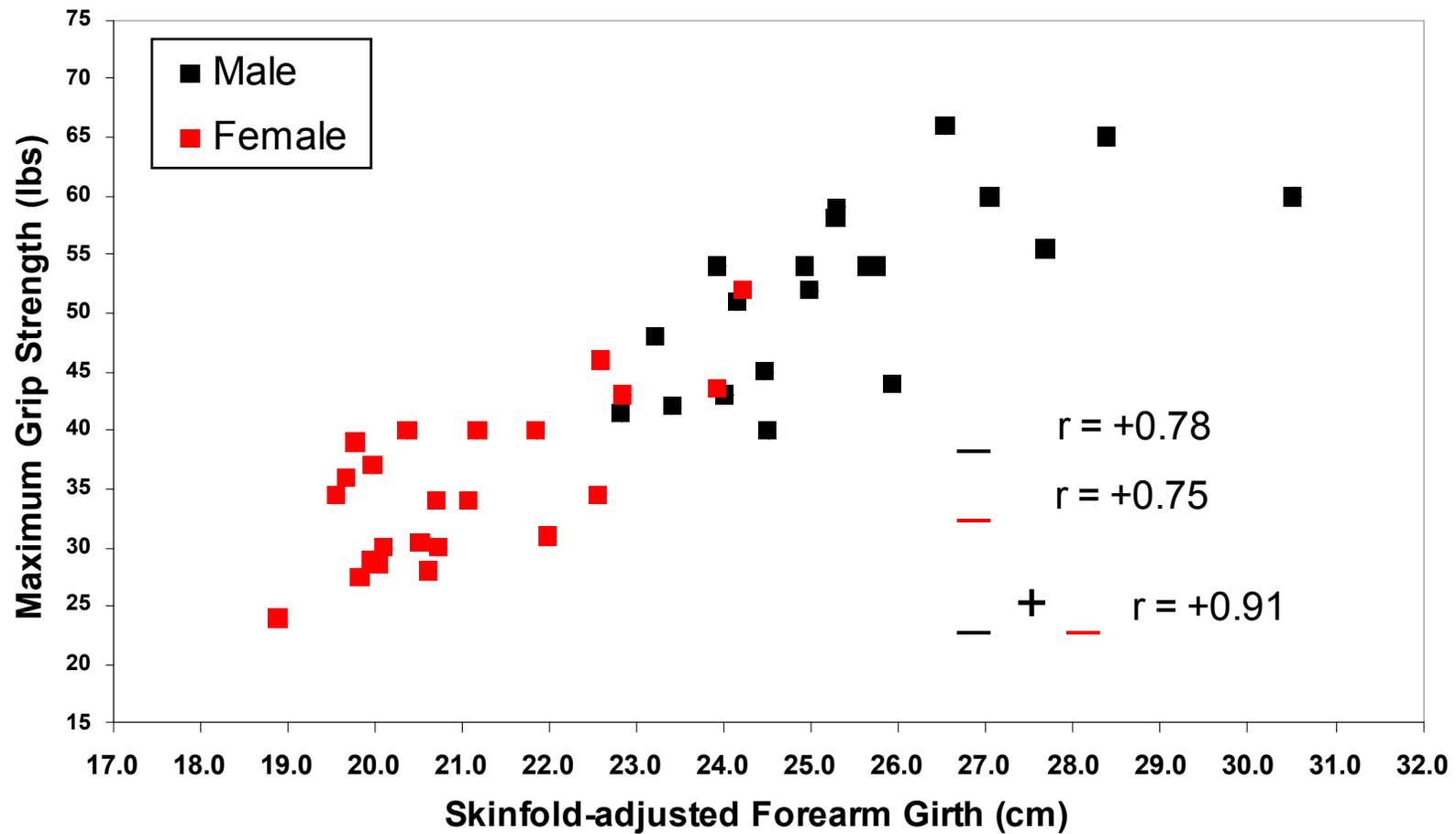

$r = 0.906$

High correlation does not always mean there is high agreement between methods
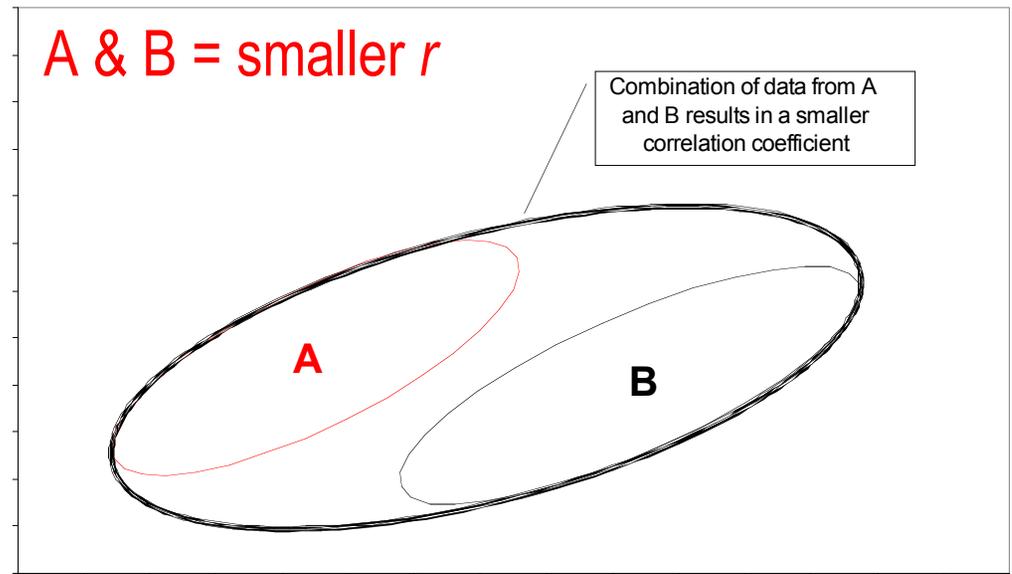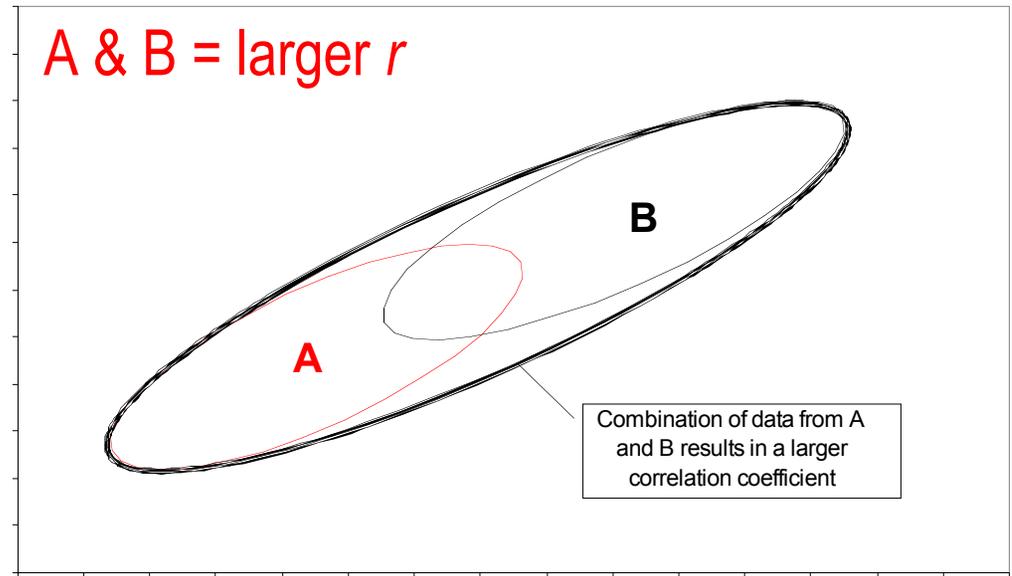

Mackey et al. *Journal of Gerontology: Medical Sciences.* 2011. 66(A)10:1108-1113

Correlation depends on the range of the data

The correlation coefficient depends upon the orientation of the two groups.

A & B = larger *r*

B

A

Combination of data from A and B results in a larger correlation coefficient

A & B = smaller *r*

Combination of data from A and B results in a smaller correlation coefficient

A

B

Correlation does not mean causation!

Fig 6
IS THIS MOUNTAIN RANGE AFFECTING
THE MURDER RATE?

836

Murders in
New York State

866

1965

2010
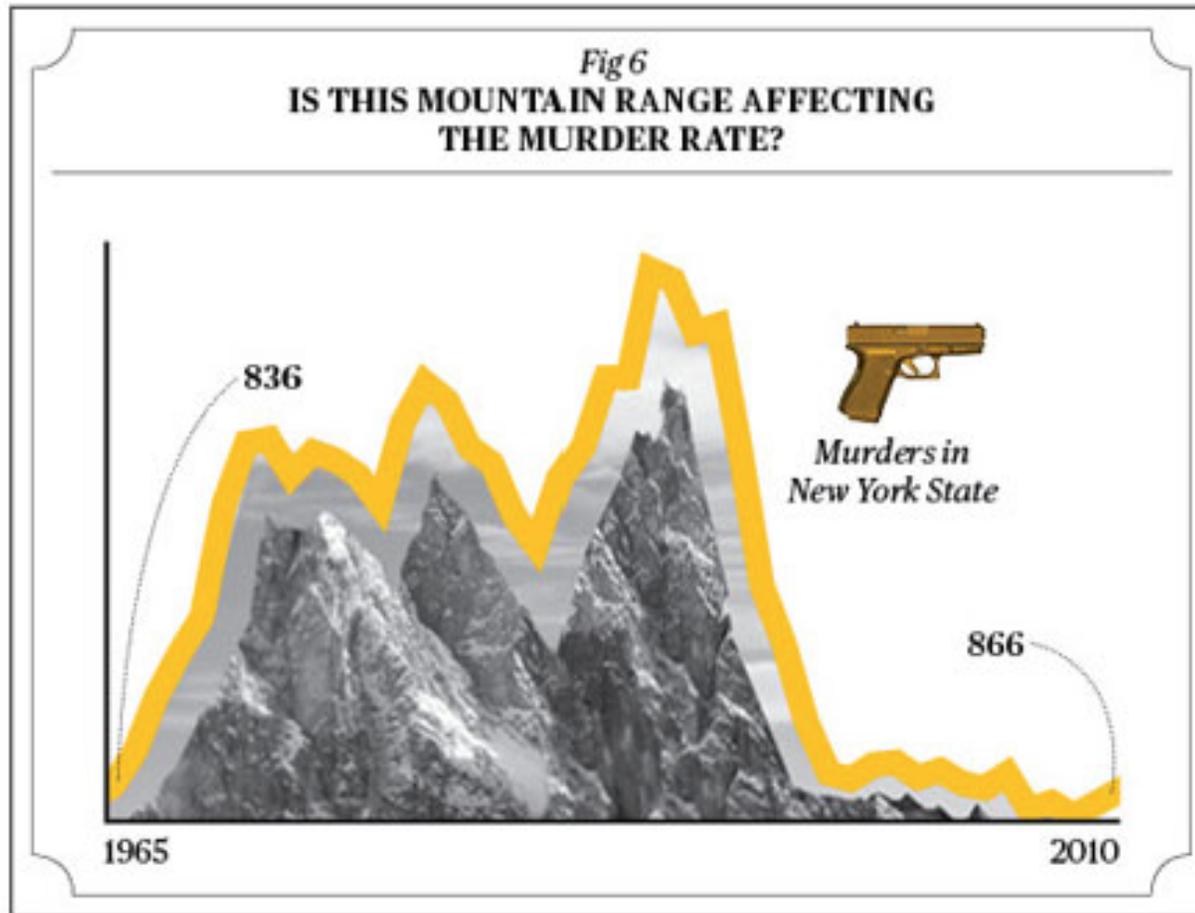
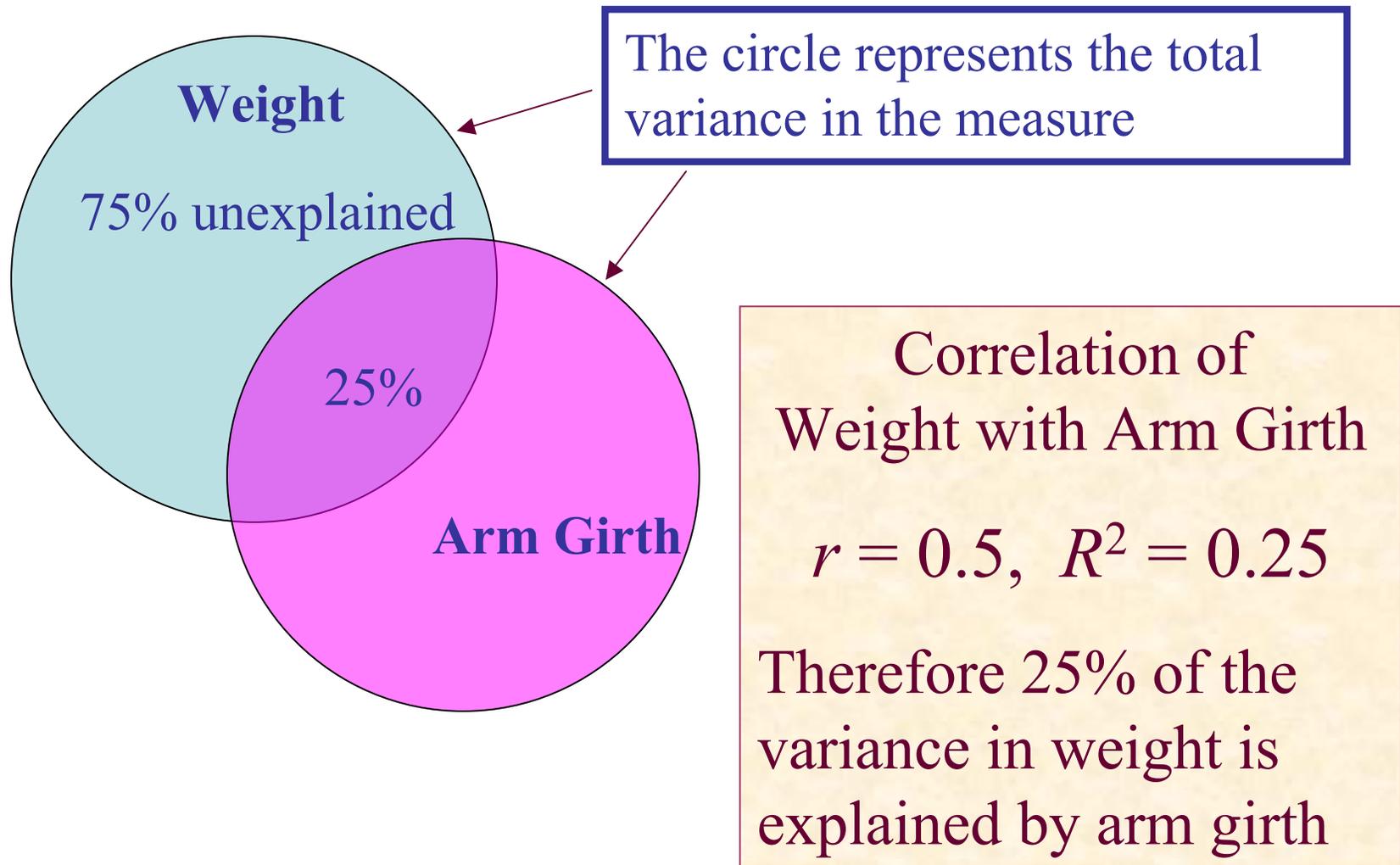# Statistical Significance of the Correlation Coefficient

- The critical value of the correlation coefficient is determined by the sample size and the significance level, $\propto$

- Degrees of freedom (df) = n-1

- Bigger sample size = lower critical value of r

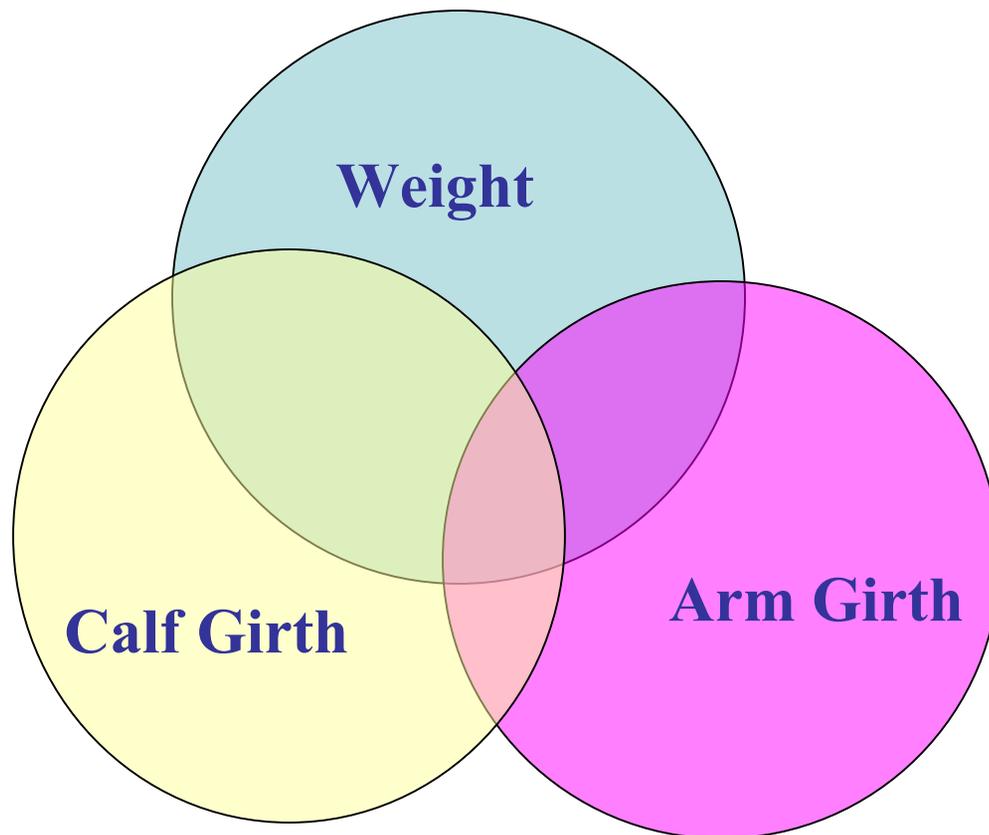- Statistical significance of r does not infer "practical significance."

| Degrees | Probability | | Degrees | Probability | |
of Freedom	0.05	0.01	of Freedom	0.05	0.01
1	.997	1.000	24	.388	.496
2	.950	.990	25	.381	.487
3	.878	.959	26	.374	.478
4	.811	.917	27	.367	.470
5	.754	.874	28	.361	.463
6	.707	.834	29	.355	.456
7	.666	.798	30	.349	.449
8	.632	.765	35	.325	.418
9	.602	.735	40	.304	.393
10	.576	.708	45	.288	.372
11	.553	.684	50	.273	.354
12	.532	.661	60	.250	.325
13	.514	.641	70	.232	.302
14	.497	.623	80	.217	.283
15	.482	.606	90	.205	.267
16	.468	.590	100	.195	.254
17	.456	.575	125	.174	.228
18	.444	.561	150	.159	.208
19	.433	.549	200	.138	.181
20	.423	.537	300	.113	.148
21	.413	.526	400	.098	.128
22	.404	.515	500	.088	.115
23	.396	.505	1,000	.062	.081

Table 2-4.2: Critical Values of the Correlation Coefficient

Coefficient of Determination (R^2)

Weight

75% unexplained

25%

Arm Girth

The circle represents the total variance in the measure

Correlation of Weight with Arm Girth

$r = 0.5, \; R^2 = 0.25$

Therefore 25% of the variance in weight is explained by arm girth