

# PLSC 502 – Autumn 2016

## Measures of Association

### Interval/Ratio Data

November 8, 2016

$$\frac{\partial Y}{\partial X} = m;$$

$$Y = mX + b$$

Other alternatives:

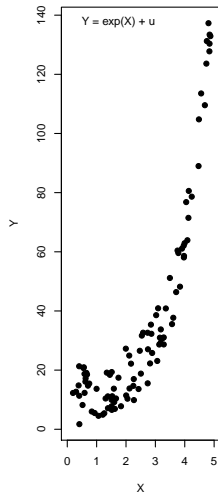
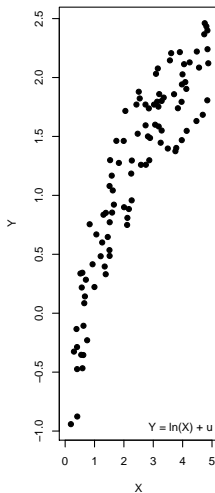
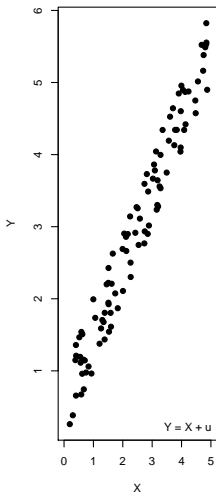
- *Logarithmic:*

$$\frac{\partial^2 Y}{\partial X \partial X} < 0$$

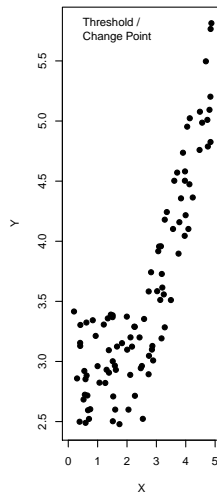
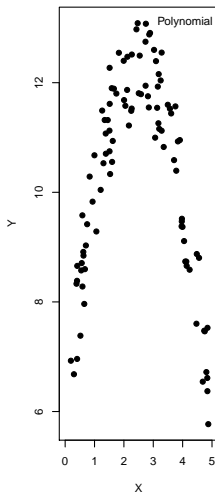
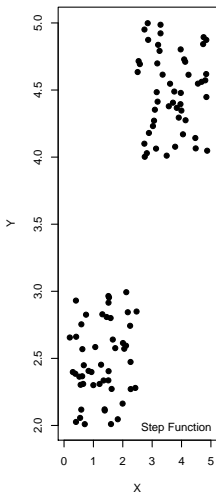
- *Exponential:*

$$\frac{\partial^2 Y}{\partial X \partial X} > 0$$

# Linear, Logarithmic, Exponential



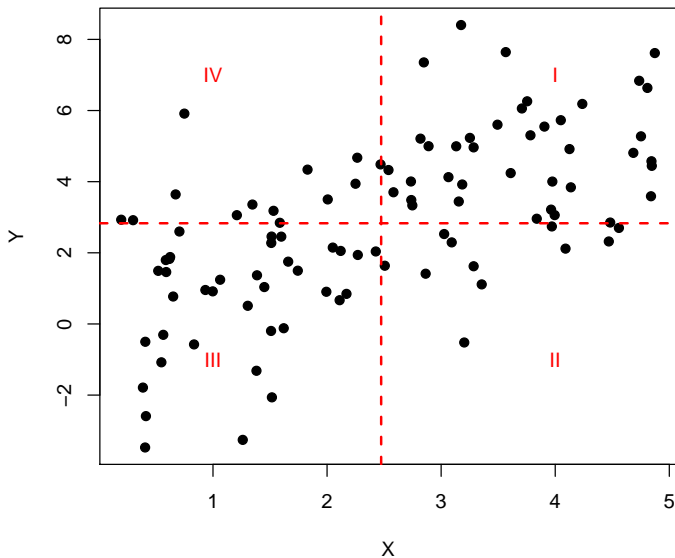
# Other Possibilities



# Linear Association: Pearson's $r$

$$\begin{aligned} r &= \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \\ &= \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^N (Y_i - \bar{Y})^2}} \\ &= \frac{\sum_{i=1}^N \left( \frac{X_i - \bar{X}}{s_X} \right) \left( \frac{Y_i - \bar{Y}}{s_Y} \right)}{N - 1} \end{aligned}$$

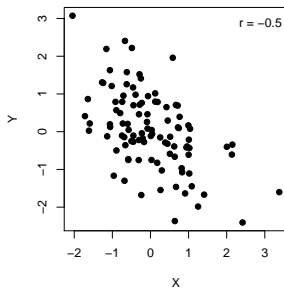
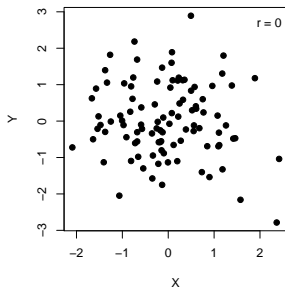
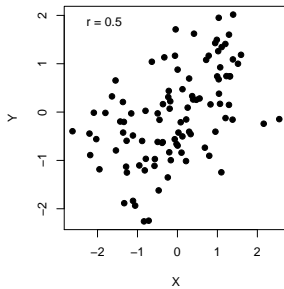
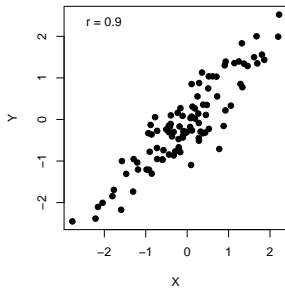
# Pearson's $r$ : Intuition



# Pearson's $r$ : Characteristics

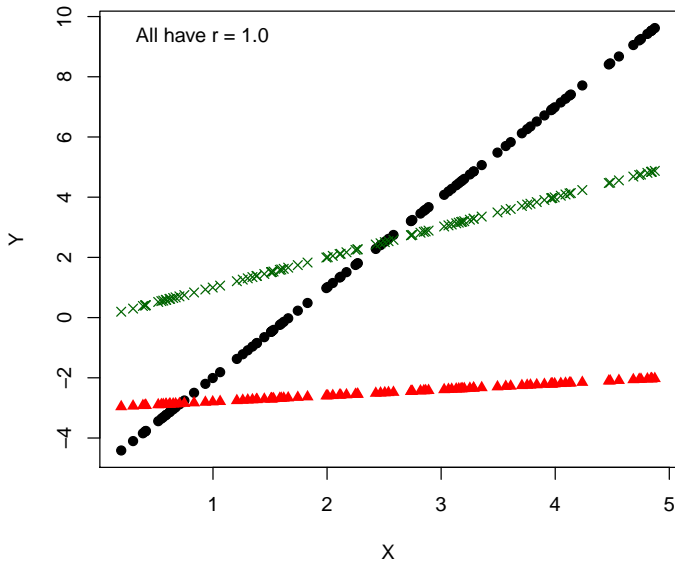
- $r \in [-1, 1]$
- $r = 0 \leftrightarrow$  no association between  $Y$  and  $X$ .
- $\text{Sign}(r) \rightarrow$  “direction” of the *linear* association
- $|r| \rightarrow$  “strength” of the *linear* association
- In general:
  - $|r| < 0.3 \rightarrow$  “weak” association
  - $0.3 < |r| < 0.7 \rightarrow$  “moderate” association
  - $|r| > 0.7 \rightarrow$  “strong” association

# Examples

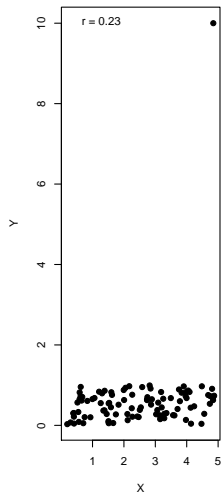
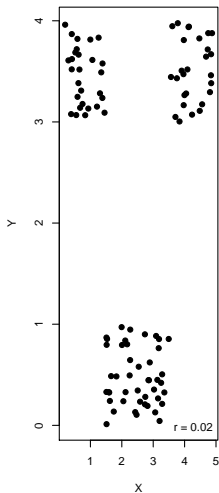
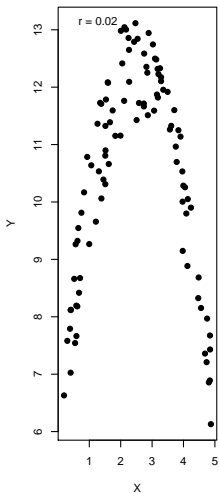




$$r = \pm 1.0 \rightarrow ?$$



# Nonlinearity, etc.



The sampling distribution of  $r$  is:

- complex, and
- skewed as  $|r| \rightarrow 1.0$ .

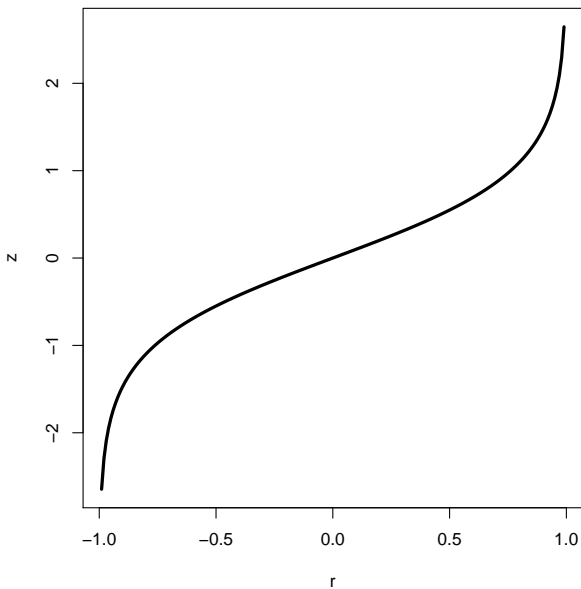
Fisher:

$$\hat{w} \equiv \frac{1}{2} \ln \left( \frac{1 + \hat{r}}{1 - \hat{r}} \right) \sim \mathcal{N} \left[ \frac{1}{2} \ln \left( \frac{1 + \hat{r}}{1 - \hat{r}} \right), \frac{1}{\sqrt{N-3}} \right]$$

implying:

$$z_r = \frac{\frac{1}{2} \ln \left( \frac{1 + \hat{r}}{1 - \hat{r}} \right) - \frac{1}{2} \ln \left( \frac{1 + r}{1 - r} \right)}{\sqrt{\frac{1}{N-3}}} \sim \mathcal{N}(0, 1)$$

# Fisher's $z$ Transformation of $r$



## Alternative Formula ( $t$ )

$$\frac{\hat{r}\sqrt{N-2}}{\sqrt{1-\hat{r}^2}} \sim t_{N-2}.$$

Note that this converges to  $z$  as  $N \rightarrow \infty$ .

## Alternative Measure: Spearman's $\rho$

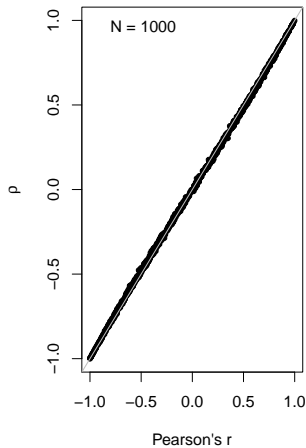
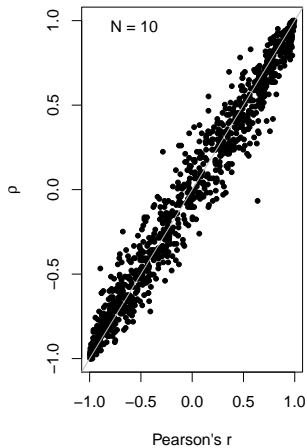
For sorted data on  $X$  and  $Y$ , where  $R_{Y_i}$  and  $R_{X_i}$  are the respective ranks,

$$\rho = 1 - \frac{6 \sum_{i=1}^N (R_{Y_i} - R_{X_i})^2}{N(N^2 - 1)}$$

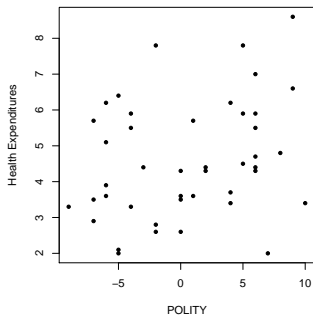
Characteristics:

- $\rho \in [-1, 1]$
- Same interpretation as  $r$ .
- Also appropriate for use with ordinal data; but
- When many “ties” occur, calculate Pearson's  $r$  on the ranks  $R_{Y_i}$  and  $R_{X_i}$ , and assign “partial” (or “half”) ranks to tied individuals.

# $r$ vs. $\rho$ Comparison



# Africa Example: POLITY and Healthcare \$



```
> with(Africa, cor.test(polity,healthexp))
```

Pearson's product-moment correlation

data: polity and healthexp

t = 1.909, df = 41, p-value = 0.06328

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.01601428 0.53973343

sample estimates:

cor

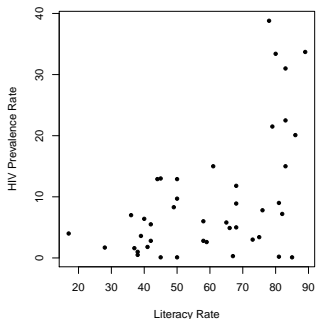
0.2857038

```
> with(Africa, SpearmanRho(polity,healthexp))
```

[1] 0.2809225



# Africa Example: Literacy and HIV \$



```
> with(Africa, cor.test(literacy,adrate))
```

Pearson's product-moment correlation

data: literacy and adrate

t = 3.8459, df = 41, p-value = 0.0004112

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.2537995 0.7060508

sample estimates:

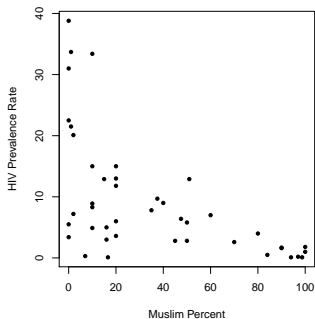
cor

0.5148944

```
> with(Africa, SpearmanRho(literacy,adrate))
```

[1] 0.4570738

# Africa Example: Muslim Percent and HIV \$



```
> with(Africa, cor.test(muslperc,adrate))
```

Pearson's product-moment correlation

data: muslperc and adrate

t = -4.4527, df = 41, p-value = 6.391e-05

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.7437363 -0.3265790

sample estimates:

cor

-0.5709233

```
> with(Africa, SpearmanRho(muslperc,adrate))
```

[1] -0.6313474

# Summary: Measures of Association

		X			
		Nominal	Binary	Ordinal	Interval/Ratio
Y	Nominal	$\chi^2$	$\chi^2$	$\chi^2$	$t$ -test (and $\eta$ )
	Binary	$\chi^2$	$\phi, Q$	$\gamma, \tau_c$	$t$ -test
	Ordinal	$\chi^2$	$\gamma, \tau_c$	$\gamma, \tau_a, \tau_b$	Spearman's $\rho$
	Interval / Ratio	$t$ -test (and $\eta$ )	$t$ -test	Spearman's $\rho$	$r$