

# PLSC 473: American Judicial Behavior

Christopher Zorn

September 24, 2015

# Measuring Association (Bivariate Plots / Statistics)

- Pictures
  - Boxplots & Density Plots (redux)
  - Scatterplots
- Numbers
  - Crosstables
  - $t$ -tests
  - Correlation
  - Bivariate regression

# Africa Data (Again)

```
> summary(Africa)
```

ccode		cabbr		country	population
Min.	:404	AGO	: 1	Angola	: 1 Min. : 470000
1st Qu.:	:452	BDI	: 1	Benin	: 1 1st Qu.: 3446000
Median	:510	BEN	: 1	Botswana	: 1 Median : 9662000
Mean	:510	BWA	: 1	Burundi	: 1 Mean : 17388558
3rd Qu.:	:556	CAF	: 1	Cameroon	: 1 3rd Qu.: 19150000
Max.	:651	CIV	: 1	Central African Republic:	1 Max. :117000000
		(Other):37	(Other)		:37

popthou		popden		polity	gdppppd	tradegdp			
Min.	: 470	Min.	:0.0022	Min.	:-9.000	Min.	: 0.500	Min.	: 4.03
1st Qu.:	3446	1st Qu.:	:0.0134	1st Qu.:	:-4.500	1st Qu.:	0.855	1st Qu.:	7.64
Median	: 9662	Median	:0.0357	Median	: 0.000	Median	: 1.200	Median	: 13.56
Mean	: 17390	Mean	:0.0643	Mean	: 0.512	Mean	: 2.159	Mean	: 30.49
3rd Qu.:	19189	3rd Qu.:	:0.0683	3rd Qu.:	5.500	3rd Qu.:	2.040	3rd Qu.:	30.01
Max.	:116929	Max.	:0.5740	Max.	:10.000	Max.	:10.800	Max.	:272.69

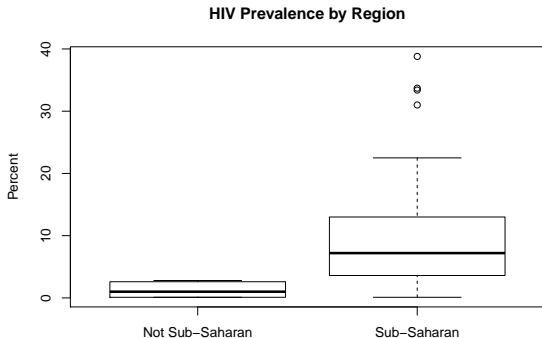
war		adrate		healthexp		subsaharan		muslperc	
Min.	:0.000	Min.	: 0.10	Min.	:2.00	Not Sub-Saharan: 6		Min.	: 0.0
1st Qu.:	:0.000	1st Qu.:	2.70	1st Qu.:	:3.45	Sub-Saharan :37		1st Qu.:	10.0
Median	:0.000	Median	: 6.00	Median	:4.40			Median	: 20.0
Mean	:0.116	Mean	: 9.37	Mean	:4.60			Mean	: 36.0
3rd Qu.:	:0.000	3rd Qu.:	:12.90	3rd Qu.:	:5.80			3rd Qu.:	55.5
Max.	:1.000	Max.	:38.80	Max.	:8.60			Max.	:100.0

literacy		internalwar		intensity	
Min.	:17.0	Min.	:0.000	Min.	:0.000
1st Qu.:	:43.0	1st Qu.:	:0.000	1st Qu.:	:0.000
Median	:61.0	Median	:0.000	Median	:0.000
Mean	:60.1	Mean	:0.302	Mean	:0.581
3rd Qu.:	:78.5	3rd Qu.:	:1.000	3rd Qu.:	:1.000
Max.	:89.0	Max.	:1.000	Max.	:3.000

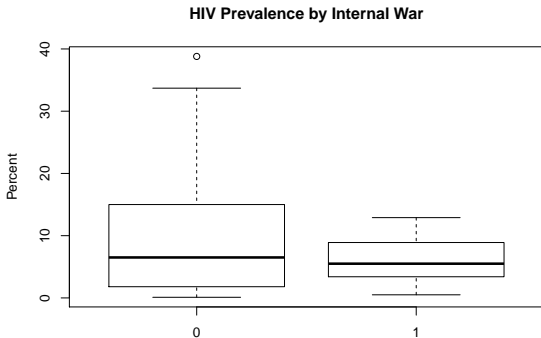
# Comparing Boxplots

```
> with(Africa, boxplot(adrate~subsaharan,ylab="Percent",  
  main="HIV Prevalence by Region"))
```



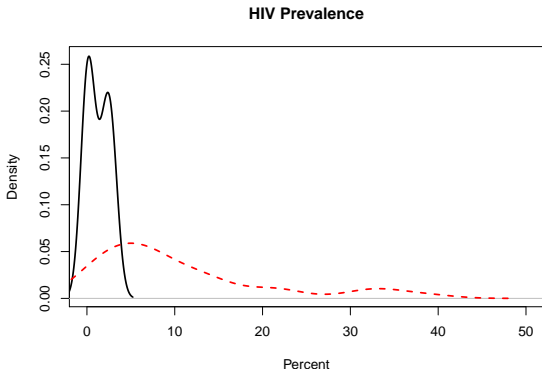
# Comparing Boxplots (cont'd)

```
> with(Africa, boxplot(adrate~internalwar,ylab="Percent",  
  main="HIV Prevalence by Internal War"))
```



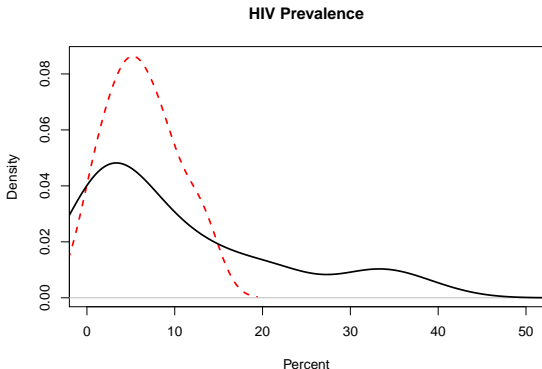
# Comparing Density Plots

```
> with(Africa[Africa$subsaharan=="Not Sub-Saharan",],  
      plot(density(adrate),lwd=2,col="black",  
          main="HIV Prevalence",xlim=c(0,50),xlab="Percent"))  
> with(Africa[Africa$subsaharan=="Sub-Saharan",],  
      lines(density(adrate),lwd=2,col="red",lty=2))
```



# Comparing Density Plots

```
> with(Africa[Africa$internalwar==1,],  
      plot(density(adrate),lwd=2,col="red",lty=2,  
          main="HIV Prevalence",xlim=c(0,50),xlab="Percent"))  
> with(Africa[Africa$internalwar==0,],  
      lines(density(adrate),lwd=2,col="black"))
```

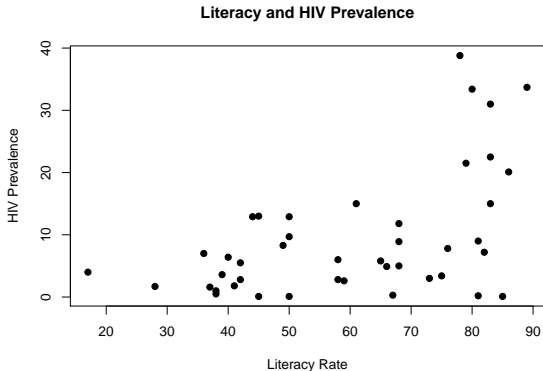


- Plotting two continuous variables.
- Horizontal (“ $X$ ”) axis = “independent” variable.
- Vertical (“ $Y$ ”) axis = “dependent” variable.
- Illustrates degree of association between  $X$  and  $Y$ .



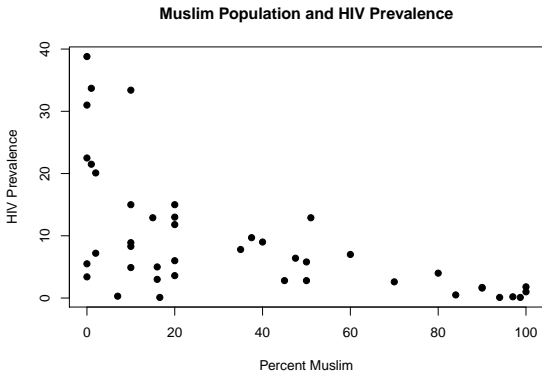
# Scatterplots: Examples

```
> with(Africa, plot(literacy, adrate, pch=19,  
  xlab="Literacy Rate", ylab="HIV Prevalence",  
  main="Literacy and HIV Prevalence"))
```



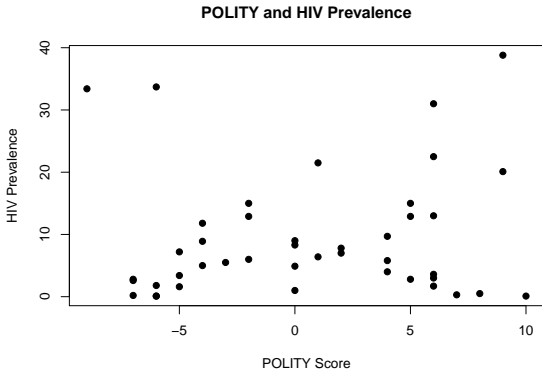
# Scatterplots: Examples

```
> with(Africa, plot(muslperc, adrate, pch=19,  
  xlab="Percent Muslim", ylab="HIV Prevalence",  
  main="Muslim Population and HIV Prevalence"))
```



# Scatterplots: Examples

```
> with(Africa, plot(polity, adrate, pch=19,  
  xlab="POLITY Score", ylab="HIV Prevalence",  
  main="POLITY and HIV Prevalence"))
```



# Crosstables (“Contingency Tables”)

- “Two-way” frequency tables
- Reports the number of observations that have a given *pair* of values on two variables (say  $X$  and  $Y$ ).
- Can show frequencies or percentages.
- Best for *discrete* / *categorical* variables

# Crosstab Example

Basic crosstab:

```
> with(Africa, xtabs(~internalwar+subsaharan))
```

	subsaharan	
internalwar	Not Sub-Saharan	Sub-Saharan
0	5	25
1	1	12

Cell proportions:

```
> with(Africa, prop.table(xtabs(~internalwar+subsaharan)))
```

	subsaharan	
internalwar	Not Sub-Saharan	Sub-Saharan
0	0.11628	0.58140
1	0.02326	0.27907

**Interpretation: “11.6 percent of all African countries were both not sub-saharan and had no internal conflict in 2001. 58.1 percent of all African countries were both sub-Saharan and had no internal conflict in 2001....” etc.**

# Crosstab Example (continued)

“Row marginals”:

```
> with(Africa, prop.table(xtabs(~internalwar+subsaharan),1))
```

	subsaharan	
internalwar	Not Sub-Saharan	Sub-Saharan
0	0.16667	0.83333
1	0.07692	0.92308

**Interpretation:** “In 2001, 16.7 percent of all African countries that did not have an internal conflict were Saharan, and 83.3 percent were sub-Saharan. In contrast, only 7.7 percent of all African countries that experienced an internal conflict were Saharan, while 92.3 percent were sub-Saharan.”

# Crosstab Example (continued)

“Column marginals”:

```
> with(Africa, prop.table(xtabs(~internalwar+subsaharan),2))
```

	subsaharan	
internalwar	Not Sub-Saharan	Sub-Saharan
0	0.8333	0.6757
1	0.1667	0.3243

**Interpretation:** “In 2001, 16.7 percent of all Saharan African countries experienced an internal conflict. By contrast, 32.4 percent of all sub-Saharan countries experienced an internal conflict. Sub-Saharan countries were therefore about twice as likely to experience an internal conflict than were Saharan countries.”

# Differences of Means

Suppose  $Y$  is our variable of interest, and that there are two groups,  $X = 0$  and  $X = 1$ , with  $n_0$  and  $n_1$  observations in each group. We want to know if  $\bar{Y}$  is different for  $X = 0$  vs.  $X = 1$ .

*Difference of means:*

$$\bar{Y}_1 - \bar{Y}_0 = \frac{1}{n_1} \sum_{i=1}^{n_1} Y_{1i} - \frac{1}{n_0} \sum_{i=1}^{n_0} Y_{0i}$$

*Variation*

$$s_{\bar{Y}_1 - \bar{Y}_0} = \sqrt{\frac{s_0^2}{n_0} + \frac{s_1^2}{n_1}}.$$

“ $T$ -score”:

$$t = \frac{(\bar{Y}_1 - \bar{Y}_0)}{s_{\bar{Y}_1 - \bar{Y}_0}}$$



Table: Rough Values of  $t$  You'll Want To Get To Know

Absolute Value of $t$	P-Value
$\approx 1.3$	0.20
$\approx 1.65$	0.10
$\approx 2$	0.05
$\approx 2.4$	0.02
$\approx 2.6$	0.01
$> 3$	$< 0.002$

# T-test: Examples

```
> with(Africa, t.test(adrate~subsaharan))
```

Welch Two Sample t-test

```
data:  adrate by subsaharan
t = -5.4, df = 41, p-value = 0.000003
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -12.942  -5.881
sample estimates:
mean in group Not Sub-Saharan      mean in group Sub-Saharan
                1.267                  10.678
```

**Interpretation: “The mean HIV prevalence rate in 2001 was 1.27 percent in Saharan Africa, and 10.68 percent in sub-Saharan Africa. This difference is statistically important / significant ( $t = -5.4$ ).”**

# T-test: Examples

```
> with(Africa, t.test(internalwar~subsaharan))
```

Welch Two Sample t-test

data: internalwar by subsaharan

t = -0.86, df = 7.4, p-value = 0.4

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.5883 0.2730

sample estimates:

mean in group Not Sub-Saharan

0.1667

mean in group Sub-Saharan

0.3243

**Interpretation: “The incidence of internal conflict in 2001 was 16.7 percent in Saharan Africa and 32.4 percent in sub-Saharan Africa. This difference is not statistically important / significant ( $t = -0.86$ ).”**

# Continuous Variables: Correlation

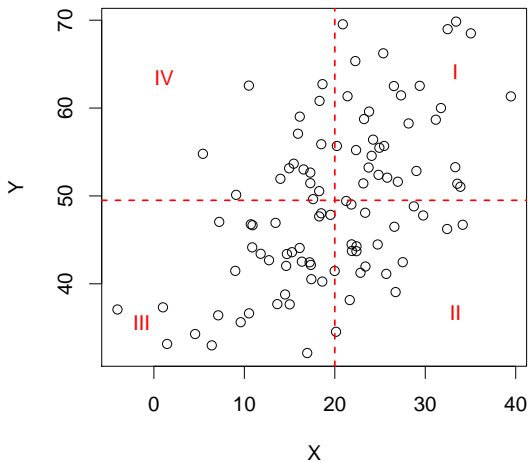
*Linear* relationship:

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

“Pearson’s  $r$ ”:

$$r = \frac{\sum_{i=1}^N \left( \frac{X_i - \bar{X}}{s_X} \right) \left( \frac{Y_i - \bar{Y}}{s_Y} \right)}{N - 1}$$

# Pearson's $r$ : Intuition



# The Meaning of $r$

Table: Rough Values of  $r$  You'll Want To Get To Know

Absolute Value of $t$	Meaning
$r = -1.0$	<i>Perfect negative</i> linear association
$r \approx -0.7$	<i>Strong negative</i> linear association
$r \approx -0.3$	<i>Weak negative</i> linear association
$r \approx 0$	No linear association
$r \approx 0.3$	<i>Weak positive</i> linear association
$r \approx 0.7$	<i>Strong positive</i> linear association
$r = 1.0$	<i>Perfect positive</i> linear association

# Correlation: Examples

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

```
[1] 0.5149
```

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

```
[1] -0.5709
```

```
> with(Africa, cor(adrate,popden))
```

```
[1] -0.1681
```

**Interpretation: “In 2001, there was a moderate, positive correlation between HIV prevalence rates and literacy rates in Africa. That same year, there was a moderate, negative correlation between HIV prevalence and the percentage of the population who self-identified as Muslim, and a small, negative correlation between HIV prevalence rates and population density.”**

# (Bivariate) Linear Regression

- Related to correlation: *linear* association.
- “Fit” a (straight) line through the points of a scatterplot:

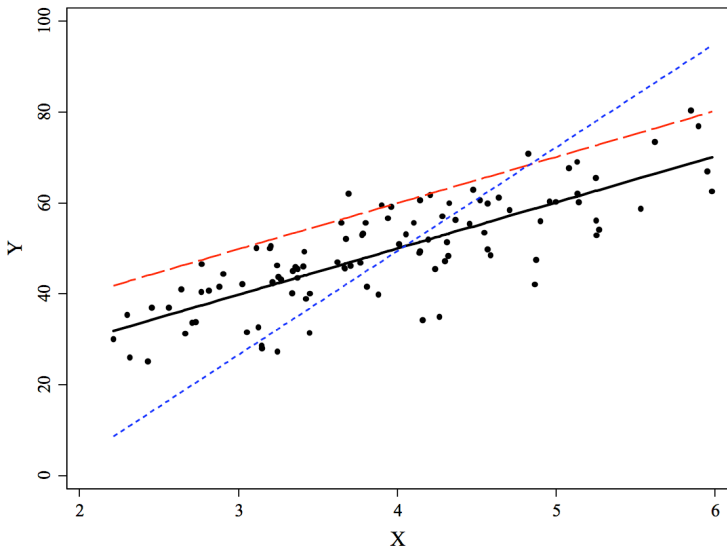
$$Y_i = \beta_0 + \beta_1 X_i + u_i \quad (1)$$

- $\beta_0$  is the “intercept”
  - $\beta_1$  is the “slope”
- Based on principle of “least squares”:

**Choose  $\beta_0$  and  $\beta_1$  so that they make  $u_i^2 = (Y_i - \beta_0 - \beta_1 X_i)^2$  as small as possible.**



# Linear Regression: Intuition



# Linear Regression: Examples

```
> fit<-with(Africa, lm(adrate~literacy))  
> summary(fit)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-6.9012	4.4301	-1.56	0.12696
literacy	0.2708	0.0704	3.85	0.00041 ***

---

Signif. codes:

0 \*\*\* 0.001 \*\* 0.01 \* 0.05 . 0.1 1

Residual standard error: 8.64 on 41 degrees of freedom

Multiple R-squared: 0.265, Adjusted R-squared: 0.247

F-statistic: 14.8 on 1 and 41 DF, p-value: 0.000411

**Interpretation: "In the 2001 Africa data, the intercept is -6.9 and the slope is 0.27. This means that a one percentage point increase in literacy is associated with an increase in the HIV prevalence rate of 0.27 percentage points."**

# Linear Regression Illustrated

```
> with(Africa, plot(literacy, adrate, pch=19,  
  xlab="Literacy Rate", ylab="HIV Prevalence",  
  main="Literacy and HIV Prevalence"))  
> abline(fit, lwd=3, col="red")
```

