

The University of Georgia
Department of Political Science
POLS 2000 Introduction to Political Science


Data Analysis:
Assessing Relationships Between Variables

6/4/2006 1

POLS 2000 Introduction to Political Science 


- **Overview**
 - **Overview of Empirical Analysis**
 - **Associations**
 - **Based on Levels of Measurement**
 - **Nominal, Ordinal, Continuous**

6/4/2006 2


POLS 2000 Introduction to Political Science 

- **Overview of Empirical Data Analysis**
 - **Test Hypothesis Empirically**

6/4/2006 3


POLS 2000 Introduction to Political Science 

Hypothesis Tests Usually About Relationships

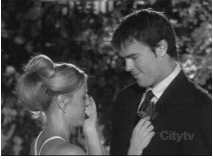

Dr. Phil

(But Not This Kind ...)

6/4/2006 4


POLS 2000 Introduction to Political Science 

Or This ...




Bachelorette

6/4/2006 5


POLS 2000 Introduction to Political Science 

Hypothesis Tests Are About Relationships

Independent Variable Dependent Variable




6/4/2006 6

POLS 2000 Introduction to Political Science 

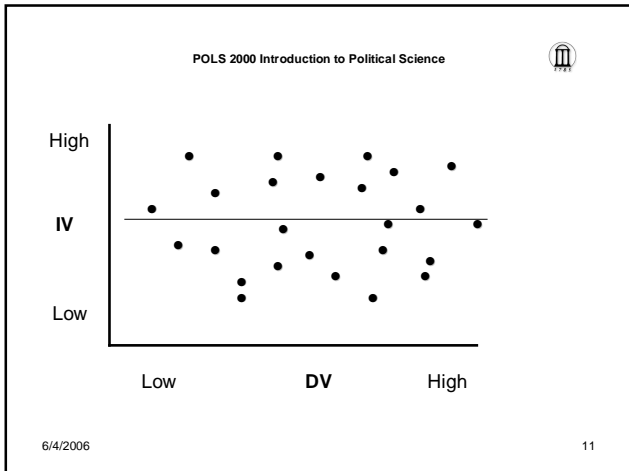
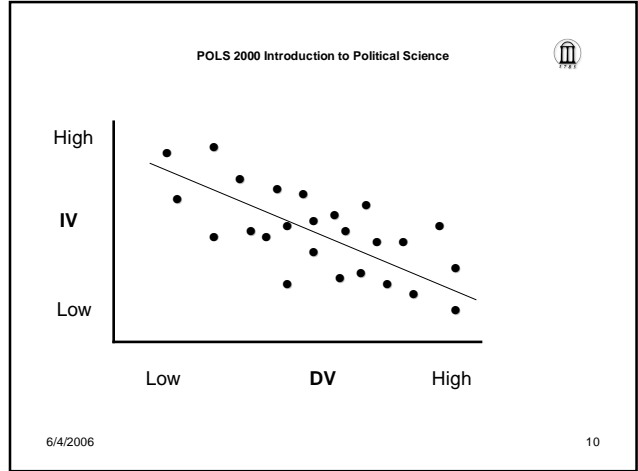
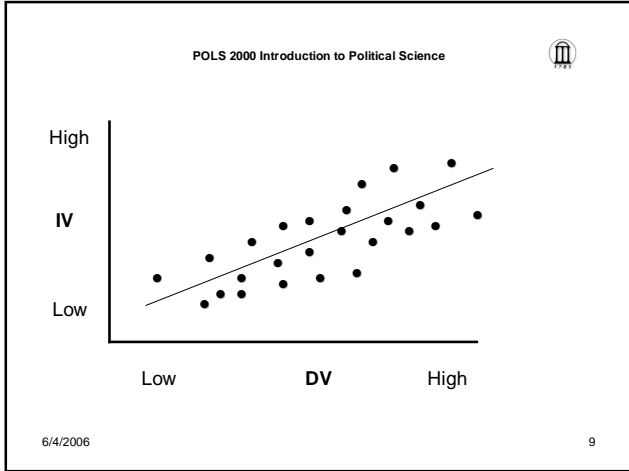
- Overview of Empirical Data Analysis
 - Hypotheses Test *Statistical* Relationship
 - Differences in IV Associated with Differences in DV
 - Direction
 - Strength
 - Likelihood Sample Relationship Occurred by Chance (Statistical Significance)

6/4/2006 7


POLS 2000 Introduction to Political Science 

- Empirical Analysis Overview, Cont.
 - Statistical Associations, Cont.
 - IV Associated with DV
 - Direction
 - Positive: Increase in IV Associated with Increase in DV
 - Negative: Increase in IV Associated with DEcrease IN DV
 - No Association: Change in IV Unrelated to Change in DV

6/4/2006 8




- POLS 2000 Introduction to Political Science
- **Associations: Overview**
 - **Bivariate Associations (DV – IV)**
 - **Nominal/Ordinal (N/O) – N/O**
 - **Crosstabulation**
 - **Continuous – N/O**
 - **Difference of Means/ANOVA**
 - **Continuous – Continuous**
 - **Correlation**
 - **Continuous – N/O/Continuous**
 - **OLS Regression**
- 6/4/2006 12

POLS 2000 Introduction to Political Science 

- **Bivariate Associations**
 - N/O DV and N/O IV
 - **Crosstabulation**
 - **Present Two or More Variables in a Table**
 - **Explore Relationship Between the Variables**

6/4/2006 13


POLS 2000 Introduction to Political Science 

- **Crosstabulation, Cont.**
- **e.g., War In Iraq & Presidential Vote Choice in 2004**

Taking everything into account, do you think the war in Iraq has been WORTH THE COST or NOT?

1. Worth it
5. Not worth it

6/4/2006 14

POLS 2000 Introduction to Political Science 


```
. tab iraq presvote, row all
```

Iraq war worth cost?, 0-1	vote for president, 0-1		Total
	Bush	Kerry	
war not worth cost	98 21.12	366 78.88	464 100.00
war worth cost	304 91.57	28 8.43	332 100.00
Total	402 50.50	394 49.50	796 100.00

```


Pearson chi2(1) = 384.1973 Pr = 0.000
likelihood-ratio chi2(1) = 432.9249 Pr = 0.000
Cramer's V = -0.6947
gamma = -0.9519 ASE = 0.011
Kendall's tau-b = -0.6947 ASE = 0.024
    
```

6/4/2006 15

POLS 2000 Introduction to Political Science 

- **Bivariate Associations, Cont.**
 - **Continuous DV and N/O IV**
 - **Dichotomous IV**
 - **Difference of Means ('t') Test**
 - **Dichotomous OR MORE Categories of IV**
 - **ANOVA**

6/4/2006 16


POLS 2000 Introduction to Political Science 

- **Difference of Means (t-Test), Cont.**
 - **e.g., South & Religiosity (2004 NES)**

South: Respondent Resides in 1 of the 11 States of Old Confederacy (Dichotomous, 0-1)

(Old South States: Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Texas, Virginia)

6/4/2006 17


POLS 2000 Introduction to Political Science 

- **Difference of Means (t-Test) , Cont.**
 - **e.g., South & Religiosity (2004 NES)**

Religiosity Scale (Based on Combined 3 items, Continuous Variable, Coded to Range from 0 to 1)

1. Church Attendance
2. Importance of religion in R's life
3. Frequency of Prayer

6/4/2006 18

POLS 2000 Introduction to Political Science 


- **Difference of Means (t-Test) , Cont.**
 - **e.g., South & Religiosity (2004 NES)**

```

.ttest relig, by(south) unequal
Two-sample t test with unequal variances
-----
Group | Obs      Mean      Std. Err.   Std. Dev.   [95% Conf. Interval]
-----+-----
otherwis | 830   .5726908   .0114927   .3311013   .5501326   .595249
Old Sout | 382   .6901178   .0149927   .2930304   .6606389   .7195967
-----+-----
combined | 1212  .6097016   .0093095   .3240997   .591437   .6279662
-----+-----
diff |          -.117427   .0188909   -.1545066   -.0803475
-----+-----
Satterthwaite's degrees of freedom: 828.785


Ho: mean(otherwis) - mean(Old Sout) = diff = 0
Ha: diff < 0          Ha: diff = 0          Ha: diff > 0
t = -6.2161           t = -6.2161           t = -6.2161
P < t = 0.0000       P > |t| = 0.0000     P > t = 1.0000
    
```

6/4/2006 19

POLS 2000 Introduction to Political Science 

- **Bivariate Associations, Cont.**
 - **Continuous DV and Continuous IV**
 - **Pearson Correlation ("r")**
 - **Positive or Negative**
 - **Ranges from -1 to 1**

6/4/2006 20


POLS 2000 Introduction to Political Science 

- Correlation, Cont.
- e.g., Moral Values & Age in 2004

```
. pwcorr mtrad age, sig star(.05)
```

	mtrad	age
mtrad	1.0000	
age	0.2308* 0.0000	1.0000

6/4/2006 21


POLS 2000 Introduction to Political Science 

- Correlation, Cont.
- e.g., Moral Values & Religiosity in 2004

```
. pwcorr mtrad relig, sig star(.05)
```

	mtrad	relig
mtrad	1.0000	
relig	0.4231* 0.0000	1.0000

6/4/2006 22

POLS 2000 Introduction to Political Science 

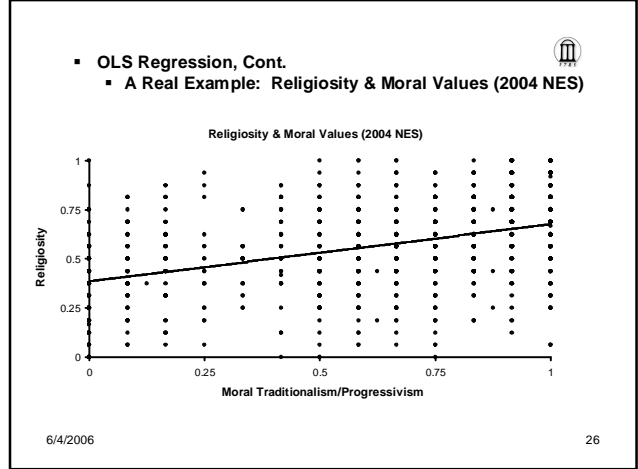
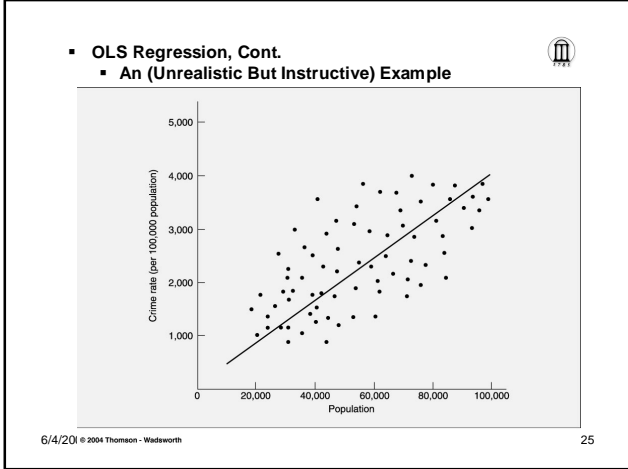
- Pearson Correlation, Cont.
 - What's High? What's Low?
 - Depends on Context
 - No Agreed Upon Standard
 - But, In General
 - $r < .2$ is 'Low' & $r > .5$ is 'High'

6/4/2006 23

POLS 2000 Introduction to Political Science 

- Bivariate Associations, Cont.
 - Continuous DV and N/O/Cont. IV
 - OLS Regression
 - Fit Straight line to Scatterplot
 - e.g., ...

6/4/2006 24



POLS 2000 Introduction to Political Science

- OLS Regression, Cont.
 - Fit 'Best' Straight Line to Scatterplot of DV & IV
 - Best Fit Minimizes Squared Distances
 - $Y = \alpha + \beta X + \epsilon$

6/4/2006 27

- OLS Regression, Cont.
 - e.g., Regress Moral Traditionalism on Religiosity

$$Y = \alpha + \beta X + \epsilon$$

```

. regress mtrad relig
-----+-----
Source |      SS      df    MS              Number of obs = 1065
-----+-----
Model |  9.61174618      1   9.61174618          F(1, 1063) = 231.95
Residual | 44.0694019    1063   .041457575          Prob > F      = 0.0000
Total | 53.6811481    1064   .050452207          R-squared     = 0.1791
                                           Adj R-squared = 0.1783
                                           Root MSE    = .20361
-----+-----
mtrad |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----+-----
relig |   .2906087   .0190858    15.23  0.000   .2531586   .3280587
_cons |   .3855054   .0132222    29.16  0.000   .3595609   .4114499
    
```

6/4/2006 28



- OLS Regression, Cont.
- Another Example
 - Feeling Toward Bush & PID in 2004



I'd like to get your feelings toward some of our political leaders and other people who are in the news these days.

I'll read the name of a person and I'd like you to rate that person using something we call the feeling thermometer.

Ratings between 50 degrees and 100 degrees mean that you feel favorable and warm toward the person. Ratings between 0 degrees and 50 degrees mean that you don't feel favorable toward the person and that you don't care too much for that person. You would rate the person at the 50 degree mark if you don't feel particularly warm or cold toward the person.

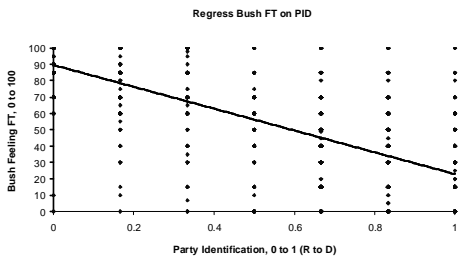
If we come to a person whose name you don't recognize, you don't need to rate that person. Just tell me and we'll move on to the next one.

(How would you rate:)

GEORGE W. BUSH ... HILLARY CLINTON ... JOHN EDWARDS



- OLS Regression, Cont.
- Bush FT & PID in 2004




- OLS Regression, Cont.
 - e.g., Regress Bush Feeling Thermometer on Party Identification
 - $Y = \alpha + \beta X + \epsilon$
 - $y = .90 - 67x + e$

```
. regress bushft pid
```

Source	SS	df	MS	Number of obs = 1190
Model	644551.11	1	644551.11	F(1, 1188) = 1098.15
Residual	697289.083	1188	586.943672	Prob > F = 0.0000
Total	1341840.19	1189	1128.54516	R-squared = 0.4803
				Adj R-squared = 0.4799
				Root MSE = 24.227

bushft	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
pid	-66.75802	2.014527	-33.14	0.000	-70.71045 -62.8056
_cons	89.59767	1.263346	70.92	0.000	87.11903 92.07631

POLS 2000 Introduction to Political Science 


- OLS Regression, Cont.
 - e.g., Regress Hillary Clinton FT on PID
 - $Y = \alpha + \beta X + \epsilon$
 - $y = .25 + 58x + e$

```

. regress hclintonft pid
-----+-----
Source |      SS      df       MS              Number of obs =   1182
-----+-----
Model | 487431.319    1 487431.319          F( 1, 1180) = 792.07
Residual | 726162.109 1180 615.391618          Prob > F      = 0.0000
-----+-----
Total | 1213593.43 1181 1027.59816          R-squared     = 0.4016
                                          Adj R-squared = 0.4011
                                          Root MSE    = 24.807


-----+-----
hclintonft |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----+-----
pid | 58.17064    2.066917    28.14  0.000    54.1154    62.22588
_cons | 25.3648    1.296753    19.56  0.000    22.8206    27.909
-----+-----
    
```

6/4/2006 33

POLS 2000 Introduction to Political Science 


- OLS Regression, Cont.
 - Measuring the Fit of a Regression Line
 - R^2
 - Explained Variance

6/4/2006 34

POLS 2000 Introduction to Political Science 

- OLS Regression, Cont.
 - Multiple Regression
 - Problem of Potential Confounds
 - Statistical Control
 - $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \dots + \epsilon$

6/4/2006 35

POLS 2000 Introduction to Political Science 

- OLS Multiple Regression, Cont.
 - e.g., Regress Bush Feeling Thermometer on Party Identification, Controlling for Ideology, Sex, Race, and Age
 - $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \dots + \epsilon$
 - $y = .86 - 57x_1 - 14x_2 + .76x_3 - 1.20x_4 + 9.05x_5 + e$

```

. regress bushft pid id female black age
-----+-----
Source |      SS      df       MS              Number of obs =   1141
-----+-----
Model | 667969.467    5 133593.893          F( 5, 1135) = 242.80
Residual | 624506.701 1135 550.226169          Prob > F      = 0.0000
-----+-----
Total | 1292476.17 1140 1133.75102          R-squared     = 0.5168
                                          Adj R-squared = 0.5147
                                          Root MSE    = 23.457

-----+-----
bushft |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----+-----
pid | -57.10519    2.421478   -23.58  0.000   -61.85626   -52.35411
id | -14.12498    1.722495   -8.20  0.000   -17.50461   -10.74534
female | .7654903    1.400711    0.55  0.585   -1.982783    3.513763
black | -1.199076    2.086294   -0.57  0.566   -5.292501    2.89435
age | 9.051962    3.009013    3.01  0.003    3.148109    14.95582
_cons | 86.36562    1.862337   46.37  0.000    82.71161    90.01963
-----+-----
    
```

6/4/2006 36