# PS2030
# Political Research and Analysis
## Unit 1: Fundamentals of Linear Regression
## 4. Multiple Regression

---

Spring 2025, Week 3

WW Posvar Hall 3600

Professor Steven Finkel

# Plan for Session

- Extension of bivariate regression to include additional explanatory variables
- Estimation of partial slope coefficients
- Hypothesis testing in multiple regression
- Assessing the relative importance of explanatory variables
- Testing alternative models
- Regression diagnostics

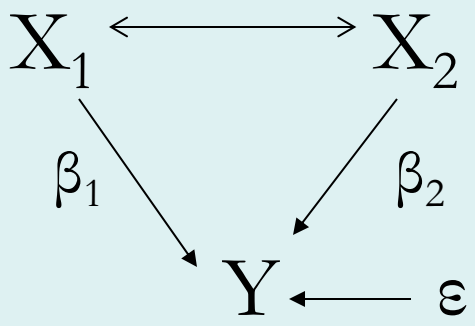
# Why Multiple Regression?

- Essential for Causal Analysis
  - We need to control for third, fourth, etc. variables so that we get the "true" (unbiased) effect of the primary independent variable of interest on the dependent variable
  - Is X truly related to Y or is the relationship "spurious"?
  - Is a treatment or a policy intervention truly responsible for some outcome, or is it because the people or places or units exposed to the treatment or intervention already differed on some important variable that produced the outcome (i.e., the selection problem in non-experimental research)
  - In non-experimental research, we cannot be sure without controlling for as many other variables as we plausibly can (and even then, we cannot be 100% sure because of unmeasured variables that may be relevant!)

- With multiple regression, we obtain a better understanding of *all* (or at least more of) the factors that explain the dependent variable
  - No relationship in social or policy sciences is monocausal, so multivariate explanations are more likely to be correct, i.e., predict the DV better
  - Introducing additional variables may help clarify which ones are the most important predictors of Y
  - Introducing additional variables may help clarify the conditions under which each one has strongest effects on Y
- So: multiple regression is more likely to satisfy regression assumptions, reduce *specification error*, increase R-squared, and produce *unbiased* (or less biased) estimates of the effects of each independent X variable on Y

# Multiple Regression Analysis

- We introduce $X_2$ ($X_3$,$X_4$…) into the process to see whether $X_1$ is truly related to Y, once $X_2$ is controlled, and to see whether $X_1$ and $X_2$, taken together, provide a better explanation of Y than either by itself



# Estimation of Multiple Regression Coefficients

- Logic: Take out the part of $X_1$ that is related to $X_2$, and take out the part of Y that is related to $X_2$, and then regress what is left from $X_1$ on what is left from Y!
- This is then the effect of $X_1$ on Y with no influence of $X_2$ on the process at all, or, "controlling for $X_2$," or, "holding $X_2$ constant"
- These effects are called "partial slopes"

- Partial slope for $X_1$, controlling for $X_2$:

$$X_{1i} = c + dX_{2i} + u_i$$

$$u_i = X_{1i} - \hat{X}_{1i.X_2}$$

and

$$Y_i = f + gX_{2i} + v_i$$

$$v_i = Y_i - \hat{Y}_{i.X_2}$$

- Then regress $v_i$ on $u_i$: $v_i = a + b_{X_1 \text{multivariate}} u_i + e_i$

$$b_{X_1 \text{multivariate}} = \frac{S(X_{1i} - \hat{X}_{1i.X_2})(Y_i - \hat{Y}_{i.X_2})}{S(X_{1i} - \hat{X}_{1i.X_2})^2}$$

- The regression estimate is the partial, or multivariate, slope of $X_1$ on Y, controlling for $X_2$. We've regressed the residuals of Y from an $X_2$ equation against the residuals of $X_1$ from an $X_2$ equation
- Same procedure to find the multivariate slope for the effect of $X_2$

# Computational Formula

- Bivariate Slope:

$$\beta_1 = r_{yx_1}\left(\frac{SD_y}{SD_{x_1}}\right)$$

- Multivariate Slopes:

$$\beta_1(x_1) = \left(\frac{SD_y}{SD_{x_1}}\right)\left(\frac{r_{yx_1} - r_{yx_2}\, r_{x_1x_2}}{1 - r^2_{x_1x_2}}\right)$$

$$\beta_2(x_2) = \left(\frac{SD_y}{SD_{x_2}}\right)\left(\frac{r_{yx_2} - r_{yx_1}\, r_{x_1x_2}}{1 - r^2_{x_1x_2}}\right)$$

- What is the difference? Multivariate slope subtracts out the joint correlation of $X_1$ and Y with $X_2$!! That is what it means to "control" for $X_2$ (or to control for $X_1$ in the equation for $\beta_2$)!
- If all variables are positively related with each other, the multivariate slope will be *smaller* than the bivariate slope

# Example

- South Africa Civic Education Data

| | St.Dev. | Correlation Matrix | | |
|---|---|---|---|---|
| Political Knowledge | 1.94 | 1.00 | | |
| Civic Education Workshops | 1.18 | .216 | 1.00 | |
| Education | 1.37 | .562 | .122 | 1.00 |

- Bivariate slope for Civic Education on Knowledge:

$$b_{1\text{bivariate}} = \frac{1.94}{1.18} * .216 = .355$$

- Bivariate slope for Education on Knowledge:

$$b_{2\text{bivariate}} = \frac{1.94}{1.37} * .562 = .796$$

- Multivariate slopes:

$$b_{1\text{multivariate}} = \frac{1.94}{1.18} * \frac{.216 - .562(.122)}{1 - (.122)^2} = \frac{1.94}{1.18} * \frac{.147}{.985} = .245$$

$$b_{2\text{multivariate}} = \frac{1.94}{1.37} * \frac{.562 - .216(.122)}{1 - (.122)^2} = \frac{1.94}{1.37} * \frac{.536}{.985} = .771$$

- Why the differences? The bivariate slopes overestimated the unique effects of each variable, misattributing the joint correlated effect of $X_1$ and $X_2$ on Y to the separate variables. This was especially the case with the effect of civic education.

# R-Squared in Multiple Regression

$$R^2 = \frac{\Sigma(\hat{Y}_i - \bar{Y})^2}{\Sigma(Y_i - \bar{Y})^2} \text{ or } 1 - \frac{\Sigma(Y_i - \hat{Y}_i)^2}{\Sigma(Y_i - \bar{Y})^2}$$

- "Explained" or Regression Sum of Squares, divided by Total Sum of Squares, or 1 Minus (Error Sum of Squares divided by Total Sum of Squares)
- It is **not** the case that $R^2$ (multivariate) = $(r_{yx1})^2 + (r_{yx2})^2$ as in the bivariate case. Why not? Some of the individual correlations with Y are "joint sums of squares" due to the interrelationship between $X_1$ and $X_2$, so they would be "double-counted" by simply adding the bivariate correlation coefficients together!
- But:

$$R^2_{y.x_1x_2} = \frac{r^2_{yx_1} + r^2_{yx2} - 2r_{yx_1} r_{yx_2} r_{x_1x_2}}{1 - r^2_{x_1x_2}}$$

- This implies that we cannot partition the variance of Y to $X_1$ and $X_2$ only; there will always be "joint correlated effects" so long as $X_1$ and $X_2$ are related
- It **is** the case, though, that multiple R, squared = multivariate R-squared

$$R^2_{y.x_1x_2} = (R_{y.x_1x_2})^2$$

# Adjusted R-squared

- Since OLS maximizes $R^2$ by construction, it is always possible to improve (or at least not decrease) $R^2$ by adding new variables, whether or not they are relevant. If you add enough variables, you will always get some improvement in $R^2$
- Logically, if you have 2 cases, **any** one variable will perfectly predict the outcome for those 2 cases; if you have 3 cases, any two variables will perfectly predict; 4 cases, 3 variables will perfectly predict, etc.
- We can adjust $R^2$ to take into account the number of independent variables, relative to the number of cases. Adjusted R-squared, or "R-bar-squared" is:

$$\bar{R}^2 = (R^2 - \frac{k}{N-1})(\frac{N-1}{N-k-1})$$

- where *k* is the number of independent variables, and N is the number of cases
- You can see that in small samples, as *k* increases, the adjustment could be substantial
- R-bar squared can decrease when adding new variables, and it can also be <0

# Example

- South Africa Data

| | St.Dev. | Correlation Matrix | | |
|---|---|---|---|---|
| Political Knowledge | 1.94 | 1.00 | | |
| Civic Education | 1.18 | .216 | 1.00 | |
| Education | 1.37 | .562 | .122 | 1.00 |

- R-squared (multivariate):

$$R^2_{y.x_1x_2} = \frac{(.216)^2 + (.562)^2 - 2(.216 * .562 * .122)}{1 - .122^2} = .338$$

- Adjusted R-squared:

$$\bar{R}^2 = (.338 - \frac{2}{939})(\frac{939}{937}) = .336$$

- Not much difference in this example. Why? The adjustment factor is small because of only 2 independent variables and the large number of cases. This of course will not always be the case!

# The Relative "Importance" of Variables

- How do we assess which variables are most "important" in explaining or accounting for the dependent variable?
- In general, it is difficult to compare the regression coefficients of variables that are measured on different scales. How does the coefficient for civic education exposure of .25 – meaning that each workshop an individual attends is associated with a .25 change in "correct" political knowledge – compare to the coefficient for education of .77 – meaning that each year of education is associated with a .77 change in "correct" political knowledge? Which is the "more important" effect?
- This can be even more difficult in other cases: how can we compare the effect of, say, one dollar of GDP on a country's Freedom House score to one additional NGO or one additional average group membership to one additional point in a Gini coefficient? One scale of the IV is in dollars, the other is in groups, the other in the Gini scale based on income concentration. How can they be compared?
- One way is just to accept the scale difference and interpret the unstandardized βs that you obtain; that is, do what we did in the bivariate case for each variable separately and come to a conclusion about which variable has the biggest theoretical impact. We can call this the **"theoretical importance"** of a variable
-

- Another way to compare variables is to *standardize* them all to have a mean of zero and a standard deviation of 1, (i.e. "z-scores"), and then run a regression of ***Standardized Y*** (z-score Y) against the ***Standardized X*** (z-score X) variables. The resultant regression coefficients are called ***Standardized Beta*** coefficents.
- They can be obtained, as above, by manually re-expressing all variables to be in z-scores and then running a regression; or, equivalently, through the formula:

$$\text{Beta}(X_1)=\beta_1 * \frac{\text{S.D.}(X_1)}{\text{S.D.}(Y)}$$

- Interpretation: "As the standard deviation of X changes by 1 unit, the standard deviation of Y changes by **Beta** units".
- Since all variables are on the same scale, the Beta coefficients can be compared: ones with higher Betas are "more important" than those with lower Betas
- Our example:

$$\text{Beta(Civic Ed)}=\frac{.25(1.18)}{1.94}=.15$$

$$\text{Beta(Education)}=\frac{.77(1.37)}{1.94}=.54$$

- So a standard deviation change in education brings about a .54 standard deviation change in knowledge, while a standard deviation change in *civic* education brings about only a .15 standard deviation change in knowledge. Education has a stronger effect in standardized terms than civic education; it is more "important" in this sense

# Comments on Beta

- In a bivariate model, Beta=r $b = r * \frac{S.D.(Y_1)}{S.D.(X_1)} = r = Beta$ when X, Y are standardized
- One argument against the use of Betas in general is that, because they conflate the "true" effect along with the sample standard deviation in the independent variable, they are not useful for explaining Y. This debate has not been resolved in the literature. We need a way to compare relative importance of variables, and Beta is an intuitively appealing measure that makes use of standardized scales for all variables
- But wherever you come down on this argument, it is important not to use Beta for comparing effect sizes for the same variable **across samples**; in that case you absolutely won't know whether the difference reflects a difference in the "true" effect or differences in the variances of the explanatory variables in those samples.
- So compare Betas ***within*** samples and compare Unstandardized β ***across*** samples to see where effects are strongest
- Betas with dummy variables (0,1) are not meaningful. Use what is called the "effect coefficient" or "Y-standardized" coefficient instead: β/S.D.(Y) and say that moving from 0 to 1 on the variable changes the SD of Y by a certain amount

- Betas provide what can be called **"dispersion importance"**; the extent to which standardized changes in X are associated with strong or weak standardized changes in Y. Beta thus gives you something akin to an "R-squared or explained variance" importance for each variable
- In fact, you can imagine adding each variable to an equation that has *all other explanatory variables* already included and noting how much R-squared changes by adding the variable in question. This "increment to R-squared by adding X at the last step" is what portion of R-squared we can attribute *uniquely* to X. Beta-squared is (nearly) identical to this value:

$$(\text{Beta}(X_1))^2 = (b_1 * \frac{\text{S.D.}(X_1)}{\text{S.D.}(Y)})^2 = \frac{b_1^2 Var(X_1)}{Var(Y)} = DR^2 \text{ for } X_1$$

- Rule of Thumb: This value should be at least .01 (corresponding to a Beta of at least .1) for $X_1$ to have **"substantive importance"** in explaining Y
- See Gross (2015) for more on substantive versus statistical significance

# Hypothesis Testing in Multiple Regression

- Individual Coefficients: T-test

$$t = \frac{b_1 - \beta_{1_0}}{s.e._{b_1}} = \frac{b_1 - \beta_{1_0}}{\hat{\sigma}_\beta} \; with \text{ n-k-1 df}$$

where

$$\hat{\sigma}_{b_1} = \sqrt{\frac{\frac{\Sigma(Y_i - \hat{Y}_i)^2}{N-k-1}}{\Sigma(X_i - \bar{X})^2(1-r^2_{x_1x_2})}}$$

  - Same procedures as in bivariate case, but the estimated standard error of the slope is adjusted to take into account the degree of correlation between $X_1$ and $X_2$. As that correlation increases, the standard error associated with the estimated slope coefficient gets bigger (i.e., more uncertainty). When the correlation is extremely high, we have **multicollinearity** and the standard errors explode or, when r=1, become impossible to calculate. This is logical!!
  - Otherwise, the same factors that decrease standard errors are in play for the multivariate as bivariate case: greater explanatory power of the model, larger N, and larger variance in X

- The formula for standard errors of the partial slopes can be generalized to the *k* independent variable case as:

$$\hat{\sigma}_{b_1} = \sqrt{\frac{\frac{\Sigma(Y_i - \hat{Y}_i)^2}{N-k-1}}{\Sigma(X_i - \bar{X})^2(1-R^2_{1.k})}}$$

- where the last term in the denominator is the R-squared of the given variable in an equation where it is regressed against **all other independent variables**
- So smaller standard errors
  - when model explanatory power is high
  - when variance in X is high
  - when the joint correlation between a variable and all the other IVs is smaller
  - when the number of cases is higher
  - when the number of IVs, relative to the number of cases, is smaller

    (this follows the logic of "adjusted R-squared")

# Example

- Civic Education on Knowledge
  - Bivariate slope= .355
  - Multivariate slope, adding education: .245
- Standard errors:
  - Bivariate:

$$\hat{s}_{b_1} = \sqrt{\frac{\frac{\sum(Y_i - \hat{Y}_i)^2}{N - 2}}{\sum(X_i - \bar{X})^2}} = \frac{RMSE}{\sqrt{\sum(X_i - \bar{X})^2}} = \frac{1.898}{\sqrt{1303.36}} = \frac{1.898}{36.10} = .053$$

  - Multivariate

$$\hat{s}_{b_1} = \sqrt{\frac{\frac{\sum(Y_i - \hat{Y}_i)^2}{N - k - 1}}{\sum(X_i - \bar{X})^2(1 - r^2_{x_1x_2})}} = \frac{RMSE}{\sqrt{\sum(X_i - \bar{X})^2(1 - r^2_{x_1x_2})}} = \frac{1.58}{\sqrt{1303.36 * (.985)}} = \frac{1.58}{35.83} = .044$$

  - Why the difference? We obtain a *smaller* multivariate standard error because of:
    - greater explanatory power of the multivariate model (i.e. lower RMSE)
    - low intercorrelation between education and civic education exposure (r=.122)
  - In other instances, standard errors will be larger in the multivariate case, esp. if intercorrelation between IVs is higher

# Hypothesis Testing, Continued: The F Test

- Do all of the explanatory variables, taken together, account for a significant amount of variance in Y? Or, can we reject the hypothesis that the slopes of all variables are equal to 0 in the population
  - $H_0$: $\beta_1 = \beta_2 = \beta_j = 0$
  - $H_0$: $R^2 = 0$
  - Specify alternative hypothesis and alpha level
  - Calculate test statistic F and make a decision

$$F = \frac{\frac{\Sigma(\hat{Y}_i - \bar{Y})^2}{k}}{\frac{\Sigma(Y_i - \hat{Y}_i)^2}{N-k-1}} \quad \text{with df(k,N-k-1)}$$

or

$$F = \frac{\frac{R^2}{k}}{\frac{1-R^2}{N-k-1}} \text{ with df(k,N-k-1)}$$

- F is the ratio of two variances, the regression variance and the error variance. Under the null, F=1, so you are testing how much larger than 1 is the ratio in your case, and whether that difference from 1 is large enough to rule out sampling error as having been responsible (at a particular alpha level or level of significance)
- Our example:

| Source | SS | df | MS |
|---|---|---|---|
| Model | 1199.10829 | 2 | 599.554145 |
| Residual | 2346.48745 | 937 | 2.50425555 |
| Total | 3545.59574 | 939 | 3.77592731 |

| | |
|---|---|
| Number of obs | = 940 |
| F( 2, 937) | = 239.41 |
| Prob > F | = 0.0000 |
| R-squared | = 0.3382 |
| Adj R-squared | = 0.3368 |
| Root MSE | = 1.5825 |

So

$$F = \frac{\frac{1199.1}{2}}{\frac{2346.5}{940-2-1}} = \frac{599.55}{2.50} = 239.4 \text{ with df}(2,937) = \frac{\frac{.3382}{2}}{\frac{1-.3382}{937}} = \frac{.1691}{.000706} = 239.4$$

- Critical value of F for those df is 3.00. So we reject $H_0$
- **Important: F can be significant even if all T-tests are insignificant (due to possible multicollinearity)**. Need to examine all significance tests as you interpret the models

# The F* test and Model Building

- Logic of F can be extended to testing the significance of *sets of independent variables*. Do the variables, taken together, explain a significant amount of variation in Y, once all other variables are taken into account? This is useful information for several reasons:
  - There may be high intercorrelations between the variables, so that individual t-tests are insignificant but there is a lot of explained variance from the variables taken jointly
  - You can use the results of these tests to cast light on the importance of groups of variables from different theories in accounting for some outcome
  - You can use the results of these tests to inform how you build and report the different models from your analysis
- Think about this test as comparing the explanatory power of two models:
  - Full Model: One that has all variables included
  - Reduced Model: One that has all variables except for a given set of variables included

- Does the full model explain a significantly greater sum of squares in the dependent variable than the reduced model? (Equivalently, does the full model reduce the error sum of squares compared to the reduced model?) Taken together, does the set of IVs lead to a significant marginal improvement in R-squared?

$$F^* = \frac{\frac{\Sigma(Y_i - \hat{Y}_i)^2_{\text{Reduced}} - \Sigma(Y_i - \hat{Y}_i)^2_{\text{Full}}}{df_{\text{Reduced}} - df_{\text{Full}}}}{\frac{\Sigma(Y_i - \hat{Y}_i)^2_{\text{Full}}}{df_{\text{Full}}}}$$

or

$$F^* = \frac{\frac{R^2_{\text{Full}} - R^2_{\text{Reduced}}}{k \text{ tested variables}}}{\frac{1 - R^2_{\text{Full}}}{df_{\text{Full}}}}$$

- You could say that F* is a test of the significance of the increment to R-squared that a set of independent variable would contribute at the last step of model building, i.e., after all other variables are included
- This increment is based on what is called the **"Extra Sums of Squares"** for the set of IVs, i.e., the reduction in error sums of squares that occurs after the set of IVs are included
- The Reduced model can have any number of variables deleted from the Full model that you desire. You can test the effects of variables 1,2, and 3 by comparing a Full model that includes variables 1,2,3,4,5, and 6, with a Reduced model that includes variables 4,5 and 6 only.
- This is a useful way to test sets of variables that belong to different theories
- You can accomplish this in STATA by estimating a full model and then entering

  test *var1 var2 var3*

  which returns the results of the F* test for the set of variables (*var1*, *var2*, *var3*)
- In R, you estimate the full and reduced model and then run an Analysis of Variance to compare the Fs from the two models:

  anova(name_reduced_model, name_full_model)

- Example: Which set of factors are better at explaining political knowledge -- social background characteristics (resources), or motivational factors?

Full Model: Age, education, church attendance, interest, media exposure, efficacy ($R^2$=.43)

Reduced Model (1): Interest, media exposure, efficacy only ($R^2$=.29)

Reduced Model (2): Age, education, church attendance only ($R^2$=.32)

| | |
|---|---|
| Increment to R-squared attributable to Social Background Factors: | **.14** |
| Increment to R-squared attributable to Motivational Factors: | **.11** |
| Total R-squared | **.43** |

$$F^*(\text{Social Background}) = \frac{\frac{.43-.29}{3}}{\frac{.57}{933}} = \frac{.046667}{.00061} = 76.3 \text{ with df}(3,933)$$

$$F^*(\text{Motivational}) = \frac{\frac{.43-.32}{3}}{\frac{.57}{933}} = \frac{.036667}{.00061} = 58.33 \text{with df}(3,933)$$

So both sets of IVs are "important", taken together

- Notes on F and F*
  - F***≠F since F*** tests a subset of variables while F tests them all
  - $F \neq t^2$ in multiple regression as it does in bivariate regression, but the F* test for a single variable *does* equal $t^2$ for that variable
  - We can partition the total Explained or Regression Sum of Squares in a model as:

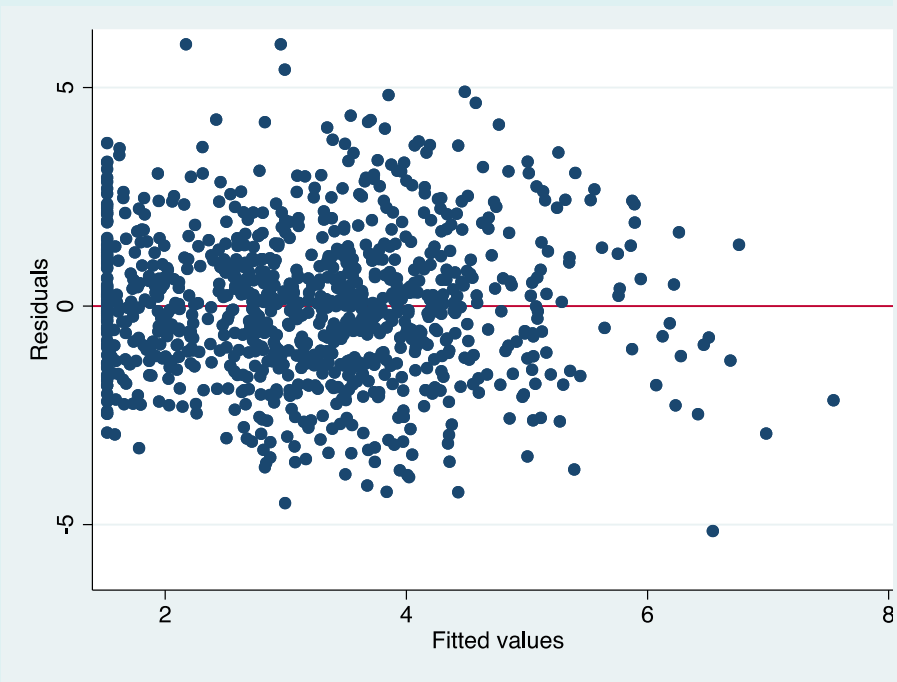
    Extra SS($X_1$)+Extra SS ($X_2$) + Joint SS($X_1$,$X_2$)+Extra SS($X_j$)….
  - You can see that F* for each variable tests the significance of its respective Extra SS
  - With **Multicollinearity**, *all* or nearly all of the Explained SS is joint SS – there is no unique Extra SS for the individual variables. That is why the F (or F*) test can be significant while the individual t-tests are not. In fact this is one of main diagnostic tests for the presence of multicollinearity in the model
  - When you misspecify a model by excluding, for example, $X_2$, OLS takes all of the joint SS and assigns it to $X_1$. This is why you get misleading (biased and inconsistent) estimates of its causal effect. The Joint SS does not belong to either $X_1$ or $X_2$ *uniquely*

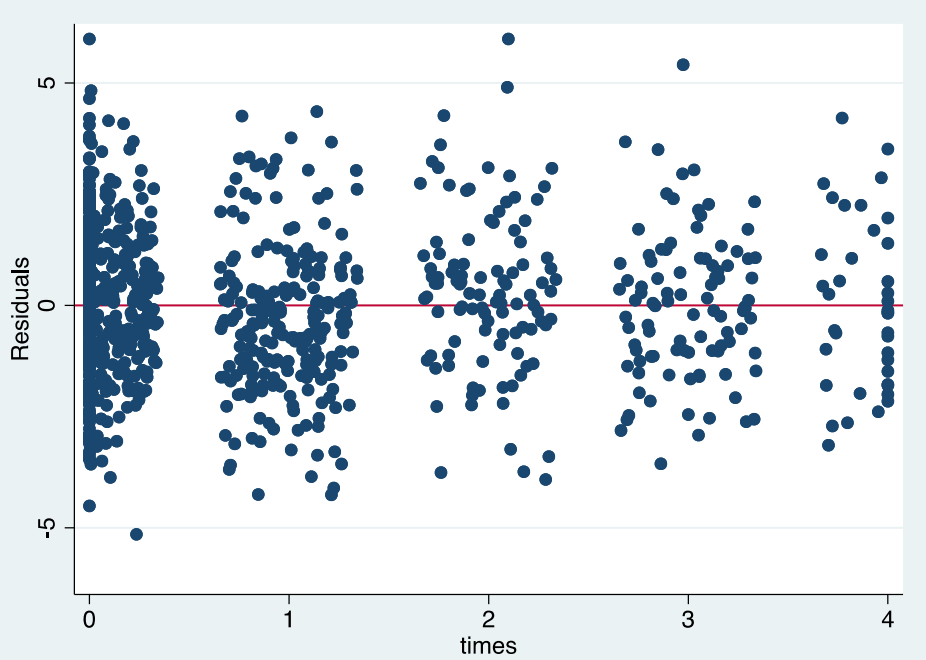
- Notes on Model Comparisons
  - Models are "better" than other models if:
    - They include relevant variables that other models omitted
    - Adjusted R-squared improves (by at least .01) but not simply because of including theoretically meaningless variables
    - The model makes sense theoretically. You must report and explain anomalous findings!!!
    - Do not let the statistical software program find your model for you!!!
  - You can compare the explanatory power of different sets of variables from different theories, show the increment to R-squared of those variables at the last step, show the F* tests, and make assessments of the relative "importance" of the variables based on the totality of the evidence

# Regression Diagnostics and Residual Analysis

- Assessment of the results of OLS models also depends on how well we can justify the OLS assumptions. Some of this is based on theory (in particular, specifying the appropriate model), but we **can** use the empirical results to shed some light on some of the assumptions, in particular those concerning the error term or residuals ε
- Although we do not observe ε, we do observe the OLS residuals "e", and we can use the sample errors to shed some light on what is likely to be the case for the population ε
- We can see, for example, whether the errors look to be:
  - Heteroskedastic
  - Autocorrelated
  - Non-normally distributed
- We can also use the OLS residuals to see whether there are significant "outliers" that are distorting the analysis or observations that are exerting undue "leverage" on the results
- Finally, we can use the residuals to shed some light on functional form issues

# The Basic Plot: Residuals With Predicted Y or With X



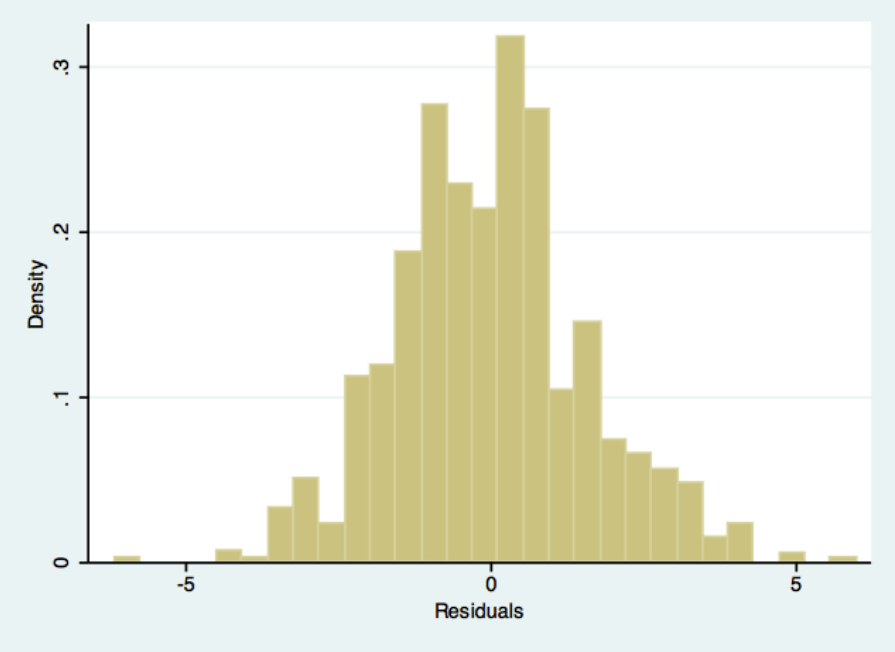




Any pattern here aside from rectangular, even distribution indicates a possible problem.

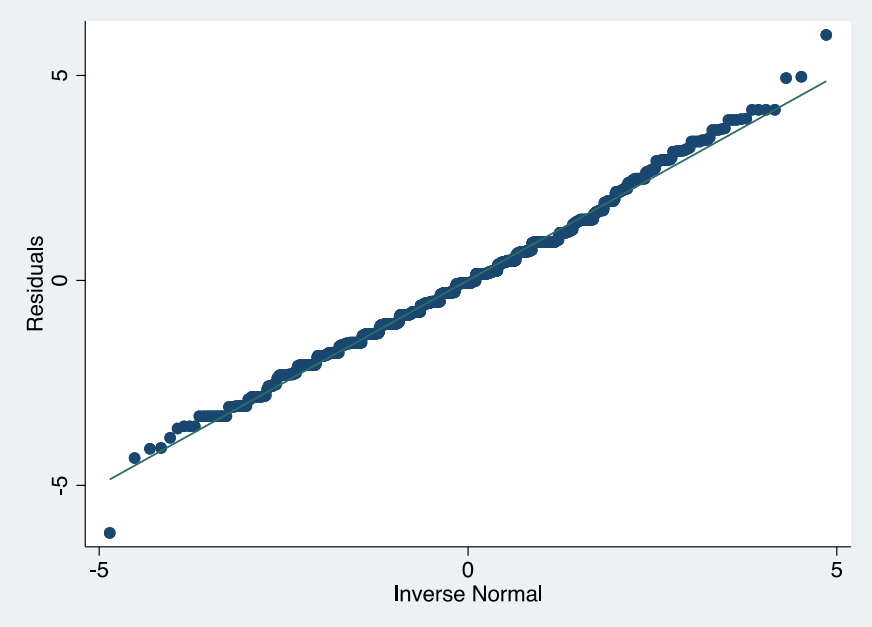
Here: heteroskedasticity possible, and possible outlier

With time-series data, plot resids against time to see autocorrelation possibility

# Normality of Errors




Histogram

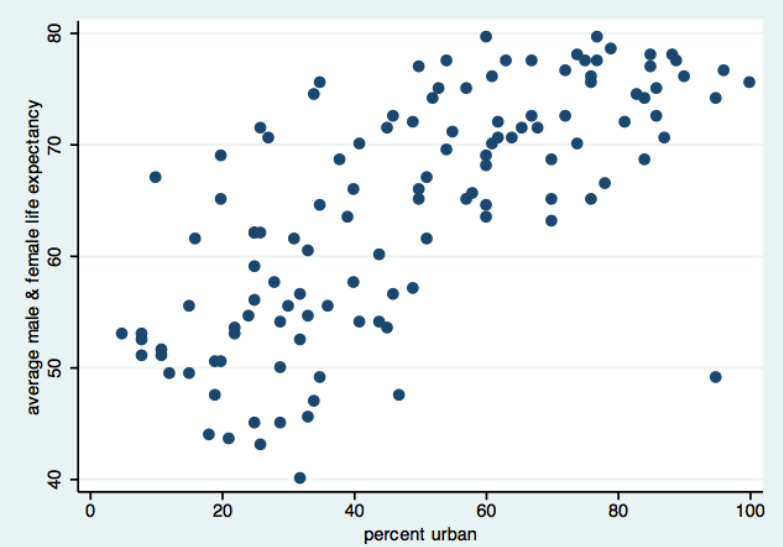


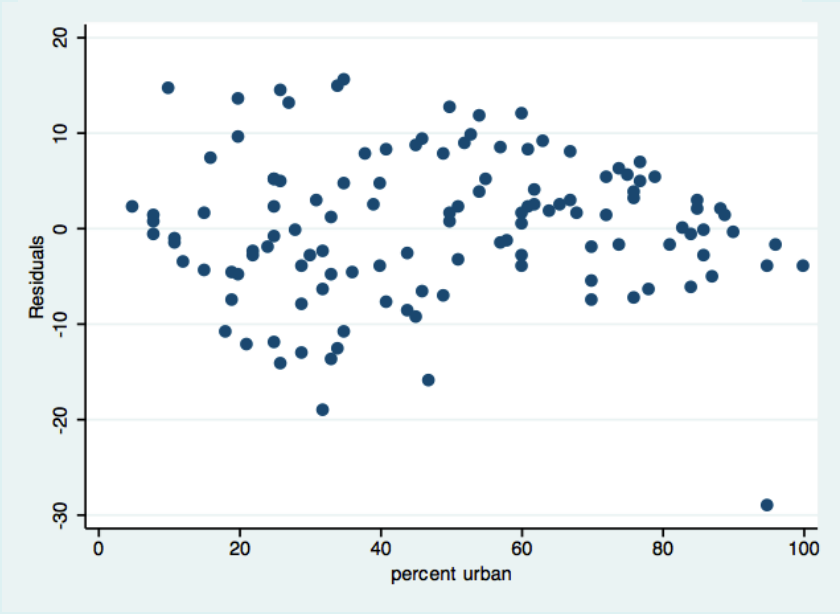

Normal probability plot

https://www.youtube.com/watch?v=okjYjClSjOg

# Outliers and Leverage

- In the "countries data set, we run a scatterplot of life expectancy against per cent urban. Nice relationship but one point looks "out of whack"
- After regression, we generate the "rvfplot" against urban. Then we can see that the one case has a huge negative residual. This is an "outlier". What country is it?
- Bhutan: Urbanization 95%, Life Expectancy 49, some 29 years lower than expected based on urbanization
- Should we drop it from the analysis? Is it theoretically so distinctive that it should not be governed by the same processes as all the other countries?



- Leverage: How much would the estimates change if the i*th* observation were eliminated from analysis? If a lot, the case has "leverage" and we *might* consider dropping it – but at least understand what it is doing to the estimates.
- Cases can have lots of leverage even if they have small residuals. For example, a case that is very far from the mean of X (or the probability mass of X) has the potential to change the OLS line by a lot, and there may or may not be a large residuals associated with that case.
- Stata statistics associated with leverage and residuals:
  - DFBETA: the distance the regression coefficient would shift when the i*th* observation is included or excluded, measured in estimated standard errors of the slope. If greater in absolute value than $2/\sqrt{N}$, then look into it further
  - DFITS: a scaled difference between predicted values for that case when it is included and when it is excluded from the regression. Will be high when either the residual is very high or when leverage is very high. Suggested cut-point for concern: absolute value greater than $2/\sqrt{kN}$