Continuous Outcomes

Objectives

Review the linear regression model (LRM)

Discuss the idea of identification

Present the method of maximum likelihood

The Linear Regression Model

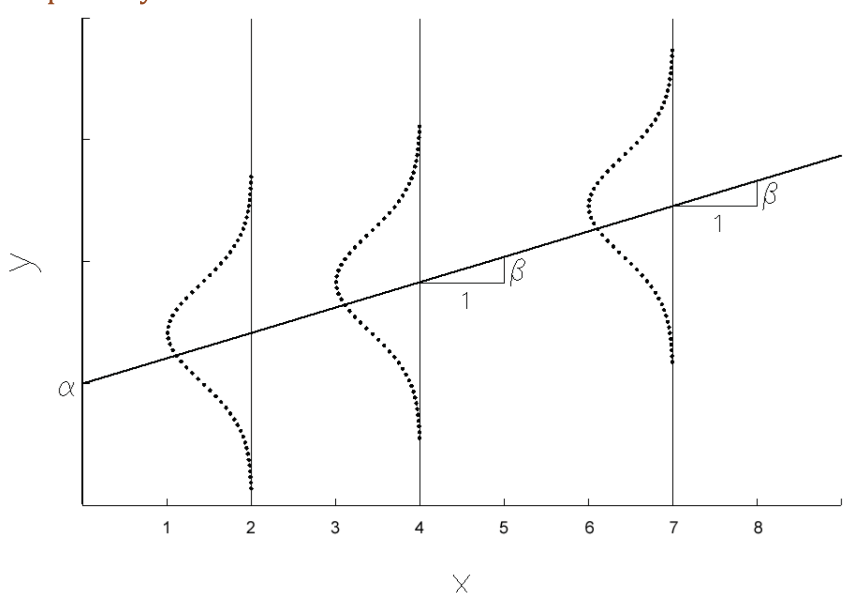
$y_i = \mathbf{x}_i \boldsymbol{\beta} + \varepsilon_i$

If $\quad \mathbf{x}_i = \begin{pmatrix} 1 & x_{i1} & x_{i2} & x_{i3} \end{pmatrix}$,

$$= \begin{bmatrix} 1 & x_{i1} & x_{i2} & x_{i3} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \varepsilon_i$$

$$= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$$

Graphically



Assumptions

Linearity

Linear independence of the x's

Errors:

- Zero conditional mean
- Homoscedastic
- Uncorrelated for any pair of x's
- [Normality]

Linearity

y is linearly related to the x's through the $\beta's$

Collinearity

the x_k's are not *perfectly* collinear

Zero conditional mean

$E(\varepsilon_i \mid \mathbf{x}_i) = 0$

This ***identifying*** assumption implies:

$$\begin{aligned} E(y_i \mid \mathbf{x}_i) &= E(\mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i \mid \mathbf{x}_i) \\ &= \mathbf{x}_i\boldsymbol{\beta} + E(\varepsilon_i \mid \mathbf{x}_i) \\ &= \mathbf{x}_i\boldsymbol{\beta} \end{aligned}$$

Homoscedastic errors

$Var(\varepsilon_i \mid \mathbf{x}_i) = \sigma^2$ for all i

Uncorrelated errors

For two observations *i* and *j*, the covariance between ε_i and ε_j is 0

What common situations violate this assumption?

Estimation by OLS

The OLS estimator of β is that value $\hat{\beta}$ that minimizes the sum of the squared residuals $\sum_{i=1}^{N}\left(y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}\right)^2$:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$$

When the assumptions of the model hold, $\hat{\boldsymbol{\beta}}$ is the best linear unbiased estimator

Estimation by Maximum Likelihood

Instead of minimizing the sum of squared errors…

Maximize the likelihood

The ML estimate is that value of the parameter that makes the observed (sample) data most likely

Requires an assumption about the conditional distribution of y

A simple example

- s be the # of men in the sample
- N be the sample size
- π be the population probability of being male

We know s and N

We want an estimate of π

$L(\pi \mid s, N)$

The Likelihood Function

Binomial formula

$$\Pr(s \mid \pi, N) = \frac{N!}{s!(N-s)!}\pi^{s}(1-\pi)^{N-s}$$

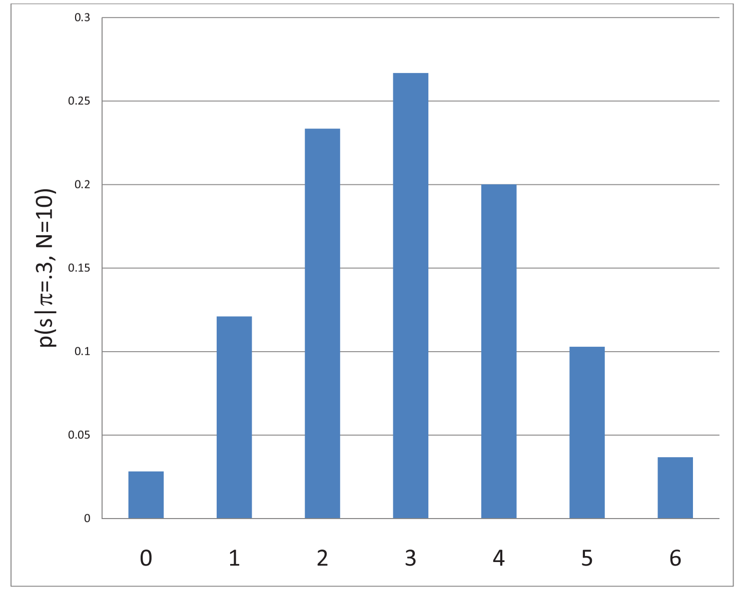
Note that

$k! = k(k-1)\cdots 2\cdot 1$

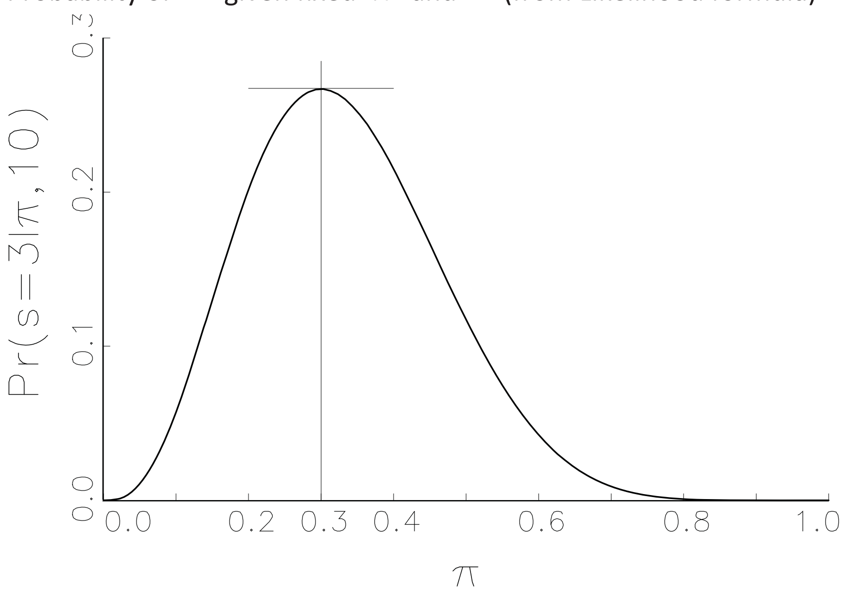
Rewrite as likelihood function for s=3 & N=10

$$L(\pi \mid s=3, N=10) = \frac{10!}{3!7!}\pi^{3}(1-\pi)^{7}$$

Probability of s given fixed N and π (from Binomial formula)



Probability of π given fixed N and s (from Likelihood formula)



Maximize the likelihood

The maximum occurs when the derivative (or gradient) is zero

$$\frac{\partial L(\pi \mid s=3, N=10)}{\partial \pi} = 0$$

The value that maximizes the likelihood function also maximizes the log of the likelihood (which is easier to calculate):

$$\frac{\partial \ln L(\pi \mid s=3, N=10)}{\partial \pi} = 0$$

For our example,

$$\frac{\partial \ln L(\pi \mid s=3, N=10)}{\partial \pi} = \frac{\partial \ln\left[\frac{10!}{3!7!}\pi^3(1-\pi)^7\right]}{\partial \pi}$$

$$= \frac{\partial \ln \frac{10!}{3!7!}}{\partial \pi} + \frac{\partial 3\ln\pi}{\partial \pi} + \frac{\partial 7\ln(1-\pi)}{\partial \pi}$$

$$= 0 + \frac{\partial 3\ln\pi}{\partial \pi} + \frac{\partial 7\ln(1-\pi)}{\partial \pi}$$

$$= \frac{3}{\pi} - \frac{7}{1-\pi} = 0$$

$\hat{\pi} = .3$ maximizes the likelihood

ML estimation of the Sample Mean

PDF for y

$$f(y_i \mid \mu, \sigma=1) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-(y_i-\mu)^2}{2}\right)$$

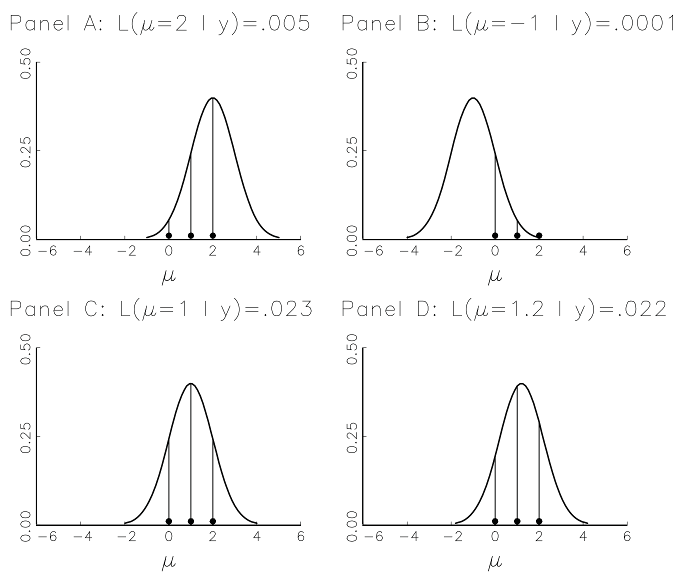
Rewrite in terms of μ

$$L(\mu \mid y_i, \sigma=1) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-(y_i-\mu)^2}{2}\right)$$

For three independent observations, the likelihood is

$$L(\mu \mid \mathbf{y}, \sigma=1) = \prod_i^3 \ln L(\mu \mid y_i, \sigma=1)$$

Graphically



ML estimation for the LRM

The pdf

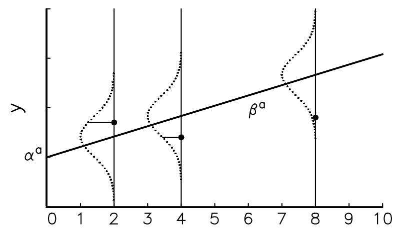
$$f(y_i \mid \alpha+\beta x_i, \sigma) = \frac{1}{\sigma}\varphi\left(\frac{y_i - [\alpha+\beta x_i]}{\sigma}\right)$$

Rewrite as likelihood equation

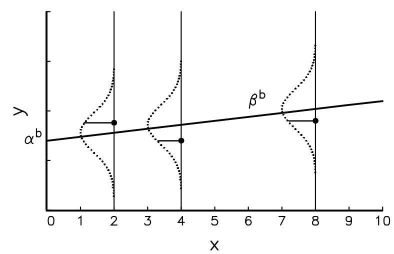
$$L(\alpha, \beta, \sigma \mid \mathbf{y}, \mathbf{X}) = \prod_{i=1}^{N} \frac{1}{\sigma}\varphi\left(\frac{y_i - [\alpha+\beta x_i]}{\sigma}\right)$$

Graphically

Panel A: Worse Fit



Panel B: Better Fit



The Properties of ML Estimators

Under very general conditions, the ML estimator is:

- Consistent
- Asymptotically efficient
- Asymptotically normally distributed

These are asymptotic properties; they describe the ML estimator as the sample size approaches infinity. But, how big must N be to be approximately infinite?

Guidelines

It is risky to use ML for $N < 100$; $N > 500$ seems safe.

These values should be raised depending on characteristics of the model and the data

Some models seem to require more observations – for example, the ordinal regression model

Identification

Occurs before estimation

Demonstration of Identification

In the LRM, the structural model is:

$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K + \varepsilon$ where $\mathrm{E}(\varepsilon \mid \mathbf{x}) = 0$

If we assume:

$\mathrm{E}(\varepsilon \mid \mathbf{x}) = \delta$

The structural equation can be modified to create an error with mean zero:

$$
\begin{aligned}
y &= 0 + \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K + \varepsilon \\
&= (\delta - \delta) + \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K + \varepsilon \\
&= (\beta_0 + \delta) + \beta_1 x_1 + \cdots + \beta_K x_K + (\varepsilon - \delta) \\
&= \beta_0^* + \beta_1 x_1 + \cdots + \beta_K x_K + \varepsilon^*
\end{aligned}
$$

This equation has all of the properties of the LRM, including

$E(\varepsilon^* \mid \mathbf{x}) = 0$

But note:

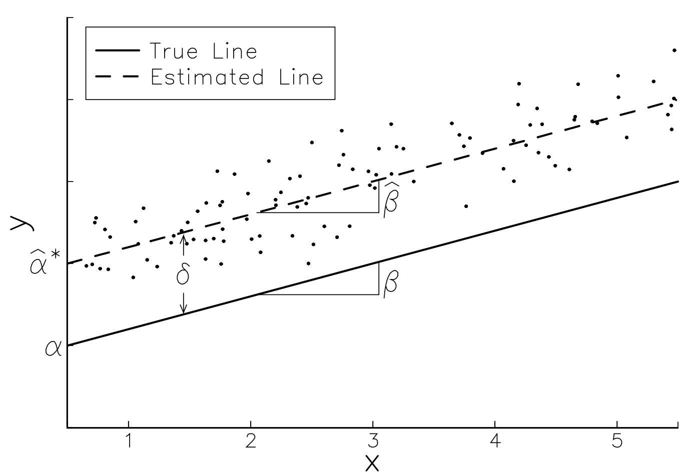
$\mathrm{E}\left(\widehat{\beta}_0^*\right) = \beta_0 + \delta$

No matter how large the sample, it is impossible to disentangle estimates of β_0 and δ

β_0 and δ *are not identified individually*, although their sum $\beta_0 + \delta$ is identified

Graphically

$\mathrm{E}(\varepsilon \mid \mathbf{x}) = \delta$



Basic Ideas about Identification

- A parameter is unidentified when it is impossible to estimate the parameter regardless of the data available
- Models become identified by adding assumptions, not by increasing the sample size. For example $E(\varepsilon) = 0$
- It is possible for some parameters of a model to be identified while others are not. For example, β but not α
- While individual parameters may not be identified, combinations of those parameters may be identified. For example, $\alpha + \delta$ but not δ and α

Interpreting Regression Coefficients

- Slopes as *marginal change* (partial derivative)
- Slopes as *discrete change* (first difference)
- Relationship between discrete and marginal change

Partial or Marginal Change

The *partial derivative* of y with respect to x_k:

$$\frac{\partial E(y \mid \mathbf{x})}{\partial x_k} = \frac{\partial \mathbf{x\beta}}{\partial x_k} = \beta_k$$

Discrete Change

Notation

Before: $E(y \mid \mathbf{x}, x_2)$ is the expected value of y given $\mathbf{x}$, explicitly noting a specific value of x_2

After: $E(y \mid \mathbf{x}, x_2 + 1)$ is the expected value of y given $\mathbf{x}$ when x_2 increases by 1

$$\frac{\Delta E(y \mid \mathbf{x}, x_2)}{\Delta x_2} = \text{After} - \text{Before}$$

$$= E(y \mid \mathbf{x}, x_2 + 1) - E(y \mid \mathbf{x}, x_2)$$

$$= [\beta_0 + \beta_1 x + \beta_2 (x_2 + 1) + \beta_3 x_3] - [\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3]$$

$$= [\beta_0 + \beta_1 x + \beta_2 x_2 + \beta_2 + \beta_3 x_3] - [\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3]$$

$$= \beta_2$$

Equality of Discrete and Partial Change

In the LRM,

$$\frac{\partial E(y \mid \mathbf{x})}{\partial x_k} = \frac{\Delta E(y \mid \mathbf{x}, x_2)}{\Delta x_2} = \beta_k$$

Simple Interpretation

For a unit increase in x_k the expected change in y equals β_k, holding all other variables constant

Having characteristic x_k (as opposed to not having the characteristic) results in an expected change of β_k in y, holding all other variables constant

Data

Career data on biochemists who obtained their Ph.D.s in 1957, 1958, 1962, and 1963 (n=408)

Descriptive Information

Name	Mean	Std Dev	Min	Max	Description
JOB	2.23	0.97	1.00	4.80	Prestige of job (from 1 to 5).
FEM	0.39	0.49	0.00	1.00	1 if female; 0 if male.
PHD	3.20	0.95	1.00	4.80	Prestige of Ph.D. department.
MENT	45.47	65.53	0.00	532.00	Citations received by mentor.
FEL	0.62	0.49	0.00	1.00	1 if held fellowship; else 0.
ART	2.28	2.26	0.00	18.00	Number of articles published.
CIT	21.72	33.06	0.00	203.00	Number of citations received.

```
. spex regjob3,clear
. use "http://www.indiana.edu/~jslsoc/stata/spex_data/regjob3.dta", clear

. codebook job fem phd ment fel art cit , compact

Variable   Obs Unique      Mean  Min       Max  Label
-------------------------------------------------------------------------------
job        408     80  2.233431    1       4.8  Prestige of 1st job on 1 to 5 scale
fem        408      2  .3897059    0         1  Gender: 1=female 0=male
phd        408     89  3.200564    1       4.8  PhD prestige on 1 to 5 scale
ment       408    123  45.47058    0  531.9999  Citations received by mentor
fel        408      2  .6176471    0         1  Fellow: 1=yes 0=no
art        408     14  2.276961    0        18  # of articles published
cit        408     87  21.71569    0       203  # of citations received
job100     408     80  223.3431  100       480  Prestige of 1st job on 100 to 500
scale
phd100     408     89  320.0564  100       480  PhD prestige on 100 to 500 scale
-------------------------------------------------------------------------------
```



```
. summarize job fem phd ment fel art cit

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         job |       408    2.233431    .9736029          1        4.8
         fem |       408    .3897059    .4882823          0          1
         phd |       408    3.200564    .9537509          1        4.8
        ment |       408    45.47058    65.52988          0   531.9999
         fel |       408    .6176471    .4865587          0          1
-------------+--------------------------------------------------------
         art |       408    2.276961    2.256143          0         18
         cit |       408    21.71569    33.05988          0        203
```


Stata: Estimating the LRM

Our LRM is:

$$JOB = \beta_0 + \beta_1 FEM + \beta_2 PHD + \beta_3 MENT + \beta_4 FEL + \beta_5 ART + \beta_6 CIT + \varepsilon$$

In Stata:

```
. regress job fem phd ment fel art cit

      Source |       SS       df       MS              Number of obs =     408
-------------+------------------------------           F(  6,   401) =   17.78
       Model |  81.0584763     6  13.5097461           Prob > F      =  0.0000
    Residual |  304.737915   401  .759944926           R-squared     =  0.2101
-------------+------------------------------           Adj R-squared =  0.1983
       Total |  385.796392   407  .947902683           Root MSE      =  .87175

------------------------------------------------------------------------------
         job |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         fem |  -.1391939   .0902344    -1.54   0.124    -.3165856    .0381977
         phd |   .2726826   .0493183     5.53   0.000     .1757278    .3696375
        ment |   .0011867   .0007012     1.69   0.091    -.0001917    .0025651
         fel |   .2341384   .0948206     2.47   0.014     .0477308    .4205461
         art |   .0228011   .0288843     0.79   0.430    -.0339824    .0795846
         cit |   .0044788   .0019687     2.28   0.023     .0006087     .008349
       _cons |   1.067184   .1661357     6.42   0.000     .7405785     1.39379
------------------------------------------------------------------------------
```

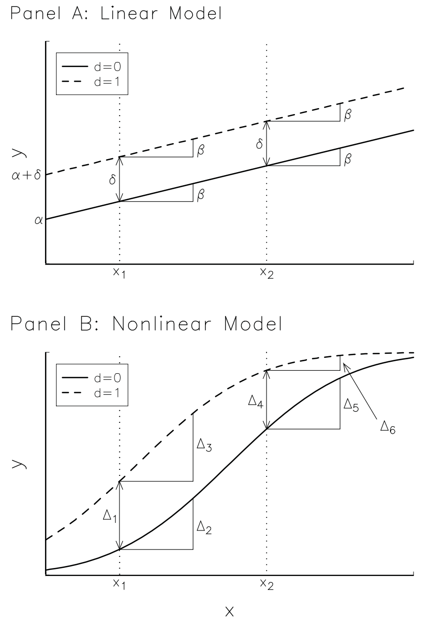

Simple Interpretation

```
------------------------------------------------------------------------------
         job |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         fem |  -.1391939   .0902344    -1.54   0.124    -.3165856    .0381977
         phd |   .2726826   .0493183     5.53   0.000     .1757278    .3696375
        ment |   .0011867   .0007012     1.69   0.091    -.0001917    .0025651
         fel |   .2341384   .0948206     2.47   0.014     .0477308    .4205461
         art |   .0228011   .0288843     0.79   0.430    -.0339824    .0795846
         cit |   .0044788   .0019687     2.28   0.023     .0006087     .008349
       _cons |   1.067184   .1661357     6.42   0.000     .7405785     1.39379
------------------------------------------------------------------------------
```

For every additional citation, the prestige of the first job is expected to increase by *.004* units, holding all other variables constant.

The expected prestige of the first job is *.14* points lower for females as compared to their male counterparts.

Comparison of Linear & Nonlinear Models



In nonlinear models, partial and discrete change *are not* equal:

$$\frac{\partial \mathrm{E}(\cdot)}{\partial x_k} \neq \frac{\Delta \mathrm{E}(\cdot)}{\Delta x_k}$$

In nonlinear models, both *discrete* & *partial* change depend on:

- the value of x_k, and
- the values of the other x's in the model

Standardized and Semi-Standardized Coefficients

y-standardized

x-standardized

fully standardized

y-standardized coefficients

Standardizing y to a unit variance:

$$\frac{y}{\sigma_y} = \frac{\beta_0}{\sigma_y} + \frac{\beta_1}{\sigma_y}x_1 + \frac{\beta_2}{\sigma_y}x_2 + \frac{\beta_3}{\sigma_y}x_3 + \frac{\varepsilon}{\sigma_y}$$

Adding new notation:

$$y^S = \beta_0^{S_y} + \beta_1^{S_y}x_1 + \beta_2^{S_y}x_2 + \beta_3^{S_y}x_3 + \varepsilon^{S_y}$$

Interpretation

For a *continuous* variable:

- For a unit increase in x_k, y is expected to change by $\beta_k^{S_y}$ standard deviations, holding all other variables constant

For a *dummy* variable:

- Having characteristic x_k (as opposed to not having the characteristic) results in an expected change in y of $\beta_k^{S_y}$ standard deviations, holding all other variables constant

x-standardized coefficients

Standardizing the x's to a unit variance:

$$y = \beta_0 + (\sigma_1\beta_1)\frac{x_1}{\sigma_1} + (\sigma_2\beta_2)\frac{x_2}{\sigma_2} + (\sigma_3\beta_3)\frac{x_3}{\sigma_3} + \varepsilon$$

Adding new notation:

$$y = \beta_0 + \beta_1^{S_x} x_1^S + \beta_2^{S_x} x_2^S + \beta_3^{S_x} x_3^S + \varepsilon$$

Interpretation

For a *continuous* variable

- For a standard deviation increase in x_k, y is expected to change by $\beta_k^{S_x}$ units, holding all other variables constant

For a *dummy* variable

- The meaning of a standard deviation change is unclear

Fully standardized coefficients

Standardizing both y and x's:

$$\frac{y}{\sigma_y} = \frac{\beta_0}{\sigma_y} + \left(\frac{\sigma_1\beta_1}{\sigma_y}\right)\frac{x_1}{\sigma_1} + \left(\frac{\sigma_2\beta_2}{\sigma_y}\right)\frac{x_2}{\sigma_2} + \left(\frac{\sigma_3\beta_3}{\sigma_y}\right)\frac{x_3}{\sigma_3} + \frac{\varepsilon}{\sigma_y}$$

Adding new notation:

$$y^S = \beta_0^S + \beta_1^S x_1^S + \beta_2^S x_2^S + \beta_3^S x_3^S + \varepsilon^{S_y}$$

Interpretation

For a *continuous* variable:

- For a standard deviation increase in x_k, y is expected to change by β_k^S standard deviations, holding all other variables constant

For a *dummy* variable:

- The meaning of a standard deviation change is unclear

Stata: Using `listcoef` to compute standardized effects

After running `regress`, the `listcoef` command is used:

```
. listcoef, cons help

regress (N=408): Unstandardized and Standardized Estimates

 Observed SD: .97360294
 SD of Error: .8717482

-------------------------------------------------------------------------------
         job |       b          t     P>|t|    bStdX    bStdY   bStdXY     SDofX
-------------+-----------------------------------------------------------------
         fem |  -0.13919    -1.543   0.124  -0.0680  -0.1430  -0.0698    0.4883
         phd |   0.27268     5.529   0.000   0.2601   0.2801   0.2671    0.9538
        ment |   0.00119     1.692   0.091   0.0778   0.0012   0.0799   65.5299
         fel |   0.23414     2.469   0.014   0.1139   0.2405   0.1170    0.4866
         art |   0.02280     0.789   0.430   0.0514   0.0234   0.0528    2.2561
         cit |   0.00448     2.275   0.023   0.1481   0.0046   0.1521   33.0599
       _cons |   1.06718     6.424   0.000
-------------------------------------------------------------------------------
       b = raw coefficient
       t = t-score for test of b=0
   P>|t| = p-value for t-test
   bStdX = x-standardized coefficient
   bStdY = y-standardized coefficient
  bStdXY = fully standardized coefficient
   SDofX = standard deviation of X
```


Interpreting standardized coefficients

job	b	t	P>|t|	bStdX	bStdY	bStdXY	SDofX
fem	-0.13919	-1.543	0.124	-0.0680	*-0.1430*	-0.0698	0.4883
phd	0.27268	5.529	0.000	0.2601	0.2801	0.2671	0.9538
ment	0.00119	1.692	0.091	0.0778	0.0012	0.0799	65.5299
fel	0.23414	2.469	0.014	0.1139	0.2405	0.1170	0.4866
art	0.02280	0.789	0.430	0.0514	0.0234	0.0528	2.2561
cit	0.00448	2.275	0.023	*0.1481*	0.0046	*0.1521*	33.0599
_cons	1.06718	6.424	0.000				

Your turn