MIT 14.32/14.320 **Spring 2023**
J. Angrist

Lecture Note 7

Regression Inference

Start with the bivariate *population* slope and intercept, $\beta = \frac{C(Y_i, X_i)}{V(X_i)}$ and $\alpha = E[Y_i] - \beta E[X_i]$. The corresponding OLS estimates of these parameters can be written:

$$\hat{\beta} = s_{XY}/s_X^2$$
$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X},$$

where s_{XY} and s_X^2 are sample covariance and variance, respectively. Like all sample statistics, OLS estimates vary from sample to sample; the corresponding standard errors quantify the resulting sampling variance.

These notes summarize the theory of statistical inference for regression. We focus on $\hat{\beta}$, first deriving the relevant sampling distribution assuming fixed regressors and Normally distributed residuals. These are *classical regression assumptions* (like the Normal data assumption in our first pass at a sampling distribution for sample means in LN2). We then follow up with *asymptotic theory*, which uses the CLT to derive approximate sampling distributions without the need for unrealistic assumptions.

1 The Wisdom of the Ancients

Classical Regression Assumptions

1. The CEF of Y_i given X_i is linear, in which case we can write:

$$E[Y_i \mid X_i] = \alpha + \beta X_i,$$

   where α and β are as defined above

2. Define the regression residual $\varepsilon_i = Y_i - E[Y_i \mid X_i] = Y_i - [\alpha + \beta X_i]$. Classicists assume:
   (a) $E[\varepsilon_i \varepsilon_j] = 0$; $i \neq j$ — random sampling
   (b) $E[\varepsilon_i^2 \mid X_i] = E[\varepsilon_i^2] = \sigma_\varepsilon^2$ — homoskedasticity
   (c) ε_i is normally distributed — Normality

3. X_i is fixed in repeated samples

Ancient econometricians derived the sampling distribution of $\hat{\beta}$ under a sampling scheme such that, given an overall sample size, the same number with $X_i = x_1$ are always picked, but among these Y_i is drawn randomly; the same number with $X_i = x_2$ is always picked, but among these Y_i is drawn randomly, etc. In this sampling scheme, we get a fixed *distribution* of X_i in every sample. Imagine, say, that you plan to study the difference in wages between MIT and Harvard grads. Let X_i be a dummy indicating MIT grads. A sample design that always selects 100 MIT grads and 100 Harvard grads fixes the distribution of X_i in your sample.

Why should we assume the distribution of regressors is fixed while the dependent variable is treated as random? Expedience: this assumption simplifies sampling theory and doesn't matter much in practice.

- Most econometrics texts also list the following "assumptions":
  (d) $E[\varepsilon_i] = 0$
  (e) $E[X_i \varepsilon_i] = 0$
  - Given our definition of ε_i, however, these are not assumptions. Rather, the regression slope and intercept *are defined* so as to make these things true (a point noted in LN5, and discussed in the appendix to MM Chapter 2 and MHE Section 3.1.3).
- In 14.32, we're much concerned with whether and when regression parameters have a causal interpretation. As discussed in LN6 (and MHE 3.2), causality turns on a conditional independence assumption (CIA) that links regression parameters with potential outcomes. Yet, the conceptual question of whether regression *parameters* are causal is divorced from the technical question of how to quantify the sampling variance of regression *estimates*.

We derive the sampling distribution of OLS estimates using this helpful rewrite of $\hat{\beta}$:

$$\begin{aligned}
\hat{\beta} = s_{XY}/s_X^2 &= \frac{\sum (X_i - \bar{X})Y_i}{\sum (X_i - \bar{X})^2} = \frac{\sum \tilde{X}_i Y_i}{\sum \tilde{X}_i^2},\ \textit{where}\ \tilde{X}_i = X_i - \bar{X} \\
&= \frac{\sum \tilde{X}_i(\alpha + \beta X_i + \varepsilon_i)}{\sum \tilde{X}_i^2} = \beta + \frac{\sum \tilde{X}_i \varepsilon_i}{\sum \tilde{X}_i^2} \qquad (1) \\
\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} &= \alpha + \beta\bar{X} + \sum \frac{\varepsilon_i}{n} - \hat{\beta}\bar{X} = \alpha + (\beta - \hat{\beta})\bar{X} + \sum \frac{\varepsilon_i}{n} \qquad (2)
\end{aligned}$$

Under classical assumptions, randomness in OLS estimates is due to randomness in residuals alone.

<u>Classical Regression Statistical Properties</u>

Result 1 The sample slope and intercept are unbiased:

$$E(\hat{\beta}) = \beta + E\left[\frac{\sum (X_i - \bar{X})\varepsilon_i}{\sum (X_i - \bar{X})^2}\right] = \beta$$

$$E(\hat{\alpha}) = \alpha + E[(\beta - \hat{\beta})\bar{X}] + E\left[\sum \frac{\varepsilon_i}{n}\right] = \alpha$$

Result 2 The sampling variance of $\hat{\beta}$ is $\frac{\sigma_\varepsilon^2}{n s_X^2}$.

Note that:

$$V(\hat{\beta}) = V\left[\frac{\sum \tilde{X}_i \varepsilon_i}{\sum \tilde{X}_i^2}\right] = V[\sum a_i \varepsilon_i],$$

where $a_i = \frac{\tilde{X}_i}{\sum \tilde{X}_i^2}$. Now, use rules for the variance of l.c.'s of r.v.s (review these in LN1) to simplify $V[\sum a_i \varepsilon_i]$

- Larger samples and increased variability in X_i produce a more precisely estimated slope. Variance in regressors is good!
- $\frac{\sigma_\varepsilon}{\sqrt{n} s_X}$ is the *standard error* of the sample slope

- In practice, we work with *estimated standard errors*: $\frac{s_e}{\sqrt{n}s_X}$, where *estimated residuals* are defined as:

$$e_i = Y_i - \hat{\alpha} - \hat{\beta}X_i,$$

and s_e^2 is their sample variance (i.e., an estimate of σ_ε^2)

Result 3 $\hat{\beta}$ is Normally distributed; in particular, $\hat{\beta} \sim N\left[\beta, \frac{\sigma_\varepsilon^2}{ns_X^2}\right]$

Proof. We've already derived the mean and sampling variance of $\hat{\beta}$. Note that $\hat{\beta} = \beta + \sum a_i\varepsilon_i$ and recall that we've assumed ε_i is Normally distributed. Normality of $\hat{\beta}$ follows from the fact that a linear combination of Normal r.v.s is Normal.

Result 4 Under classical assumptions, the OLS estimator, $\hat{\beta}$, is a best linear unbiased estimator (BLUE) of β. This famous result is called the *Gauss-Markov Theorem*.

- $\hat{\beta}$ is said to be a 'linear estimator" because it's a linear combination of the Y_i. In particular, $\hat{\beta} = \sum a_iY_i$
- $\hat{\beta}$ is "best" in the linear class because any other linear unbiased estimator, say $\hat{b} = \sum b_iY_i$, for some other constants b_i not equal to a_i such that $E[\hat{b}] = \beta$, has sampling variance no smaller than that of $\hat{\beta} = \sum a_iY_i$.

Proof that OLS is BLUE (TBD in recitation).

(Wait - you mean there are *other* unbiased estimators of β besides OLS? See Pset 4!)

2 Using the Theory

Hypothesis testing

Test $H_0 : \beta = \beta_0$ using the regression t-statistic:

$$T_n = \frac{\hat{\beta} - \beta_0}{s_e/(\sqrt{n}s_X)} \sim t(n-2)$$

where s_e^2 is an estimate of σ_ε^2 as before (why $n-2$ df?). Recall that:

$$T_n \sim_{approx} N(0,1)$$

We therefore reject H_0 when T_n is a sufficiently surprising draw from a $t(n-2)$ or from a standard Normal distribution.

Confidence intervals

Let c_α be the critical value for a two-sided α-level hypothesis test. Then:

$$P[-c_\alpha \leq T_n \leq c_\alpha] = 1 - \alpha$$

Substituting, we have:

$$P\left[\hat{\beta} - c_\alpha\left(\frac{s_e}{\sqrt{n}s_X}\right) \leq \beta \leq \hat{\beta} + c_\alpha\left(\frac{s_e}{\sqrt{n}s_X}\right)\right] = 1 - \alpha$$

For large n, $T_n \approx N(0,1)$, so for $\alpha = .05$, we have $c_\alpha \approx 2$.

3 Multivariate Regression Standard Errors

Suppose you've got two regressors:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i \qquad E[\varepsilon_i X_{ji}] = 0;\ j = 1, 2 \tag{3}$$

The estimated standard error of $\hat{\beta}_1$ is

$$\frac{s_e}{\sqrt{n} s_{\tilde{x}_1}},$$

where s_e is the sample standard deviation of estimated residuals in (3) and $s^2_{\tilde{x}_1}$ is the variance of the residual from a regression of X_{1i} on X_{2i}.

4 Check the Theory: A Small Sampling Experiment

- The pop parameter here is an immigration coefficient estimated using data from the March 2012 Current Population Survey, with controls for age, age^2, and female
- Don't give me p-vals, yo. It's all about that standard error!

```
.
.         // 14.32 Lecture Note 7: c.i. coverage for regression estimates of immigrant-native wage gap
.
.         // Insheet data from 2016 ACS PUMS
.         import delimited "ss16pusa.csv", clear
(284 vars, 1,623,216 obs)

.
.         // Sample of interest is working men and women aged 40-49
.         keep if age>=40 & age <=49
(1,427,225 observations deleted)

.         keep if wkw==1
(62,348 observations deleted)

.
.         // Generate usual hourly earnings
.         lab var wagp "Raw annual earnings"

.         lab var wkhp "Usual hours per week"

.         gen uhe = wagp/(50*wkhp)

.         quietly sum uhe, d

.         replace uhe=. if uhe>`r(p99)'
(1,331 real changes made, 1,331 to missing)

.         lab var uhe "Usual hourly earnings (truncated at 99pct)"

.         gen loguhe=log(uhe)
(8,324 missing values generated)

.         lab var loguhe "Log hourly earnings"

.
.         // Code regressors
.         gen immig=(nativity==2)

.         lab var immig "Foreign born"

.         gen female=(sex==2)

.         gen age2=agep^2

.
. reg loguhe immig agep age2 female

      Source |       SS           df       MS      Number of obs   =   125,319
-------------+----------------------------------   F(4, 125314)    =   1013.10
       Model |  2067.20528         4   516.80132   Prob > F        =    0.0000
    Residual |  63925.2846   125,314  .510120853   R-squared       =    0.0313
-------------+----------------------------------   Adj R-squared   =    0.0313
       Total |  65992.4899   125,318  .526600248   Root MSE        =    .71423

------------------------------------------------------------------------------
      loguhe |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       immig |  -.1498594   .0049374   -30.35   0.000    -.1595366   -.1401823
        agep |   .0829035   .0246283     3.37   0.001     .0346324    .1311746
        age2 |  -.0008836   .0002765    -3.20   0.001    -.0014254   -.0003417
      female |  -.2275821   .0040461   -56.25   0.000    -.2355124   -.2196519
       _cons |   1.358982   .5467118     2.49   0.013     .2874358    2.430527
------------------------------------------------------------------------------

. gen popbeta=_b[immig]

.
.     // Keep only key variables in a newdata set called "workingextract"
```

```
39 .         drop if missing(loguhe)
   (8,324 observations deleted)

40 .     keep wagp loguhe uhe wkhp immig agep age2 female popbeta

41 .         save workingextract, replace
   file workingextract.dta saved

42 .
43 . summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
agep	125,319	44.61104	2.845942	40	49
wagp	125,319	65587.55	60054.73	4	665000
wkhp	125,319	42.56813	9.995274	1	99
uhe	125,319	30.06187	23.8018	.0013333	183.2727
loguhe	125,319	3.15407	.7256723	-6.620073	5.210975
immig	125,319	.2121466	.4088297	0	1
female	125,319	.4655479	.4988136	0	1
age2	125,319	1998.244	253.5185	1600	2401
popbeta	125,319	-.1498594	0	-.1498594	-.1498594

```
44 . keep if _n==1
   (125,318 observations deleted)

45 . keep popbeta

46 . save expresults, replace
   file expresults.dta saved

47 .
48 . /* draw samples of N=100 500 times compute regression se */
49 .
50 .    forvalues s=1/500 {
  2.         quietly use workingextract, clear
  3.         quietly bsample 100
  4.         quietly reg loguhe immig agep age2 female
  5.         quietly gen betahat=_b[immig]
  6.         quietly gen sebeta=_se[immig]
  7.         quietly gen hi95 = betahat + (sebeta*1.96)
  8.         quietly gen lo95 = betahat - (sebeta*1.96)
  9.         quietly gen cover95=0
 10.         quietly replace cover95 = 1 if lo95<=popbeta &  popbeta<=hi95
 11.         quietly keep if _n==1
 12.         quietly append using expresults
 13.         quietly save expresults, replace
 14.    }

51 .
52 .     keep popbeta betahat sebeta hi95 lo95 cover95

53 .     drop if missing(betahat)
   (1 observation deleted)

54 .
55 .         /* sampling experiment results */
56 .
57 .         list if _n<=5
```

	popbeta	betahat	sebeta	hi95	lo95	cover95
1.	-.1498594	-.3002854	.1793579	.0512562	-.6518269	1
2.	-.1498594	.1582616	.195466	.5413749	-.2248517	1
3.	-.1498594	-.0601462	.1675861	.2683227	-.388615	1
4.	-.1498594	-.2011411	.2040096	.1987176	-.6009999	1
5.	-.1498594	.0736945	.2200027	.5048999	-.3575108	1

```
58 .         summarize betahat lo95 hi95 cover95
```

Variable	Obs	Mean	Std. Dev.	Min	Max
betahat	500	-.1592314	.19725	-.7738841	.5400975
lo95	500	-.511066	.1978349	-1.083843	.1105932
hi95	500	.1926032	.2073171	-.4639261	.9696018
cover95	500	.918	.2746395	0	1

- Coverage is not perfect, but it's not bad. And this is in spite of the fact that the dependent variable isn't Normally distributed, the regressor is as random as the dependent variable, the CEF isn't linear, and the residuals aren't homoskedastic. What's up with that?! As we'll soon see, it's the CLT!
- As we'll also shortly see, *robust* standard errors, which allow for heteroskedasticity, improve on this.

5 Modern Times: Regression in Asymptopia

Life in asymptopia requires large samples but few assumptions. In particular, we no longer need assume linear CEFs, fixed regressors, homoskedastic or Normal residuals. Random sampling (Classical Assumption 2a) is enough for viable statistical inference.

- Asymptopians work with *probability limits*, that is, the LLN and its corollaries, instead of relying on unbiasedness to justify choice of estimator
- Asymptopians use the CLT to argue that, in large random samples, sampling distributions are approximately Normal, without requiring Classical Assumptions 1, 3, 2b, or 2c

The relevant asymptotic theory is detailed in MHE 3.1.3, and sketched at the end of MM Chapter 2 and notes below.

5.1 Consistency of the Sample Slope

1. *Laws of large numbers* (LLN; review definition of *plim* in LN2)

   (a) Let $\bar{X}_n$ denote the sample mean of X_i in a sample of size n. Then: $plim_{n\to\infty}\bar{X}_n = \mu_X$.

   - We say: "The sample mean is a *consistent estimator* of the pop mean."

   (b) LLN for other sample moments. Let $m_n^{r+s} = \sum \frac{X_i^r Y_i^s}{n}$. Then: $plim_{n\to\infty} m_n^{r+s} = E[X_i^r Y_i^s]$.

   - *Any* sample moment is a consistent estimator of the corresponding population moment.

   (c) Plimming functions. Suppose that $plim\, A_n = a$ and $plim\, B_n = b$. Let $h(A_n, B_n)$ be any function continuous at a, b. Then

   $$plim\; h(A_n, B_n) = h(a, b).$$

   - This is called the *Continuous Mapping Theorem*.

Result 1 $\hat{\beta}_n$, the OLS slope estimator computed in a sample of size n, is consistent for β.

Proof. Recall that $\beta = \frac{C(X_i,Y_i)}{V(X_i)}$ and that OLS is the sample analog of this, $\hat{\beta}_n = s_{XY}/s_X^2$. By the Continuous Mapping Theorem,

$$plim\; \hat{\beta}_n = plim(s_{XY}/s_X^2) = plim(s_{XY})/plim(s_X^2).$$

Also,

$$s_{XY} = \frac{1}{n}\sum X_i Y_i - \bar{X}_n \bar{Y}_n,$$

so, plimming inside the covariance formula shows that

$$plim\, s_{XY} = plim\left[\frac{1}{n}\sum X_i Y_i\right] - plim\left[\bar{X}_n\right] plim\left[\bar{Y}_n\right] = E[X_i Y_i] - \mu_X \mu_Y = C(X, Y).$$

Likewise, $plim\, s_X^2 = V(X)$. So, we see that

$$plim(s_{XY}/s_X^2) = C(X,Y)/V(X) = \beta.$$

5.2 Asymptotic Distribution of the Sample Slope

1. *Convergence in Distribution*

   The sequence of statistics S_n has asymptotic or limiting distribution F if:

   $$Lim_{n\to\infty} P(S_n \leq c) = F(c) \qquad \text{for any value } c \text{ where } F \text{ is defined.}$$

   When we treat statistic S_n as if it has distribution F, we are said to be "using an asymptotic approximation." Sometimes, we denote this by $S_n \sim_a F$.

   (a) *Central Limit Theorem*

   The standardized sample mean from a random sample has an asymptotic Normal distribution. This means that:

   $$Lim_{n\to\infty} P\left[\frac{\bar{X} - \mu_X}{s_X/\sqrt{n}} < c\right] = Lim_{n\to\infty} P\left[\frac{\bar{X} - \mu_X}{\sigma_X/\sqrt{n}} < c\right] = \Phi(c), \tag{4}$$

   where Φ is the standard Normal CDF.

   - Note that $\frac{\bar{X}-\mu_X}{s_X/\sqrt{n}}$ is the usual t-statistic for a sample mean
   - The first equals sign in (4) means that replacing σ_X with the estimated std dev, s_X, doesn't (asymptotically) matter. Hence, what we've been calling a "$t - statistic$" is asymptotically standard Normal.
   - The variance of a sample statistic's limiting distribution is referred to as the statistic's *asymptotic variance*. The asymptotic standard error is the square root of this, divided by $\sqrt{n}$. For a sample mean, the asymptotic SE is $\sigma_X/\sqrt{n}$, estimated by $s_X/\sqrt{n}$, the same as the finite-sample standard error.

   (b) CLT for sample moments

   In general, any standardized sample moment has a limiting Normal distribution. In particular:

   $$(Sample\ moment - Population\ moment)/(asymptotic\ SE\ of\ sample\ moment) \sim_a N(0,1)$$

   (c) CLT for functions of moments

   Suppose that A_n and B_n are sample moments and that C_n is a continuous function of these moments:

   $$C_n = h(A_n, B_n).$$

   Then C_n has a limiting Normal distribution:

   $$(C_n - plim\, C_n)/(asymptotic\ SE\ of\ C_n) \sim_a N(0,1)$$

This is a remarkably powerful result: it encapsulates most of the asymptotic theory an applied econometrician needs to know.

- Formulas for the asymptotic standard error of a function of sample moments are typically more complicated than formulas for the asymptotic standard errors of the underlying sample moments themselves. But we can look these up when needed (and Stata knows them).

Result 2
Consider the regression model:

$$Y_i = \alpha + \beta X_i + \varepsilon_i.$$

The OLS estimator $\hat{\beta}_n$ computed in a random sample of size n is approximately Normally distributed with probability limit β and asymptotic variance equal to:

$$AV(\hat{\beta}_n) = \frac{E[(X_i - \mu_X)^2 \varepsilon_i^2]}{(\sigma_X^2)^2}.$$

This in turn means that $\frac{\sqrt{n}(\hat{\beta}_n - \beta)}{\sqrt{AV(\hat{\beta}_n)}} \sim_a N(0,1)$

Proof. $plim\ \hat{\beta}_n = \beta$ is Result 1, above. Asymptotic normality is a consequence of the CLT for functions of sample moments. The formula for the asymptotic variance is derived in MHE 3.1.3.

$AV(\hat{\beta})$ becomes an *asymptotic standard error* when we take the square root and divide by $\sqrt{n}$. Stata *estimates* the asymptotic standard error of a regression coefficient using the *robust standard error* formula:

$$\widehat{RSE}(\hat{\beta}_n) = \frac{1}{\sqrt{n}} \left(\frac{\frac{1}{n}\sum_i [(X_i - \bar{X})^2 e_i^2]}{(s_X^2)^2} \right)^{\frac{1}{2}} \tag{5}$$

This is $\frac{1}{\sqrt{n}}\sqrt{AV(\hat{\beta}_n)}$ after replacing expectations with sums, ε_i with the estimated resids, e_i, and replacing σ_X^2 with s_X^2. Stata reports this standard error when you use option "robust." MM notation differs slightly, using $RSE(\hat{\beta}_n)$ to denote population robust standard errors, without worrying about whether resids and σ_X^2 are estimated or known.

RSE for multivariate regression

When the slope coefficient of interest is $\hat{\beta}_j$, that is, the jth coefficient in a multivariate model, we get the relevant robust standard error by modifying (5) to be ... (see MM p. 97 for the answer)

Homoskedasticity in large samples

Result 3 When residuals are homoskedastic, i.e., $E[\varepsilon_i^2 \mid X_i] = \sigma_\varepsilon^2$, then $AV(\hat{\beta}_n) = \frac{\sigma_\varepsilon^2}{\sigma_X^2}$.

Proof: The general AV formula is $\frac{E[(X_i - \mu_X)^2 \varepsilon_i^2]}{(\sigma_X^2)^2}$

$$
\begin{aligned}
&= \frac{E[E((X_i - \mu_X)^2 \varepsilon_i^2 \mid X_i)]}{(\sigma_X^2)^2} = \frac{E[(X_i - \mu_X)^2 E(\varepsilon_i^2 \mid X_i)]}{(\sigma_X^2)^2} \\
&= \frac{\sigma_\varepsilon^2 E[(X_i - \mu_X)^2]}{(\sigma_X^2)^2} \qquad \text{(since } E[\varepsilon_i^2 \mid X] = \sigma_\varepsilon^2\text{)} \\
&= \frac{\sigma_\varepsilon^2}{\sigma_X^2}.
\end{aligned}
$$

- Assuming homoskedastic resids, the RSE std error formula simplifies to the old-fashioned formula (to complete the path from asymptotic to old-fashioned, divide by $\sqrt{n}$)
- Ancient econometricians fretted much about heteroskedasticity. Modern masters have only one thing to say in the face of the heteroskedastic challenge: “robust”

5.3 Empirical SE Formulas Compared

```
    opened on:   5 Apr 2018, 10:00:06

  1 . 
  2 .         // Merge housing and person record from 2016 ACS PUMS
  3 .         use pfile, clear

  4 .         merge m:1 serialno using hfile

       Result                           # of obs.
       -----------------------------------------
       not matched                       124,853
           from master                         0  (_merge==1)
           from using                    124,853  (_merge==2)

       matched                         3,156,487  (_merge==3)
       -----------------------------------------

  5 .         keep if _m==3
(124,853 observations deleted)

  6 .         drop _m

  7 . 
  8 .         // Sample of interest is all women aged 25-49
  9 .         keep if sex==2 & age>=25 & age <=49
(2,684,757 observations deleted)

 10 . 
 11 .         // Children
 12 .         lab var noc "Number of own children"

 13 . 
 14 .         // Employed variable
 15 .         gen employed=(cow>=1 & cow<=7)

 16 .         lab var employed  "Individual is employed"

 17 . 
 18 .         // Weekly hours
 19 .         replace wkhp=0 if employed==0
(65,870 real changes made)

 20 .         lab var wkhp "Usual hours per week"
```

```
      sum employed wkhp noc yearsEd age white

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
    employed |    471,730     .860365    .3466083          0          1
        wkhp |    435,760    32.05255    17.07636          0         99
         noc |    464,543    1.082944    1.219274          0         16
     yearsEd |    471,730    13.98169    3.078403          0         21
        agep |    471,730    37.15064    7.276517         25         49
-------------+---------------------------------------------------------
       white |    471,730    .7336273    .4420619          0          1

46 .         // Hours worked on number of kids, old-fashioned vs. robust SEs
47 .         reg wkhp noc yearsEd age white

      Source |       SS           df       MS      Number of obs   =   429,843
-------------+----------------------------------   F(4, 429838)    =  10946.29
       Model |  11395392.4         4   2848848.1   Prob > F        =    0.0000
    Residual |   111868309   429,838  260.256907   R-squared       =    0.0924
-------------+----------------------------------   Adj R-squared   =    0.0924
       Total |   123263701   429,842  286.765139   Root MSE        =    16.132

------------------------------------------------------------------------------
        wkhp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         noc |  -2.274926   .0203607  -111.73   0.000    -2.314832    -2.23502
     yearsEd |   1.353838   .0080881   167.39   0.000     1.337986    1.369691
        agep |  -.0496699   .0033886   -14.66   0.000    -.0563114   -.0430284
       white |   .4837675   .0560483     8.63   0.000     .3739146    .5936205
       _cons |   17.14356   .1782207    96.19   0.000     16.79425    17.49286
------------------------------------------------------------------------------

48 .         reg wkhp noc yearsEd age white, r

Linear regression                               Number of obs     =    429,843
                                                F(4, 429838)      =    9620.16
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0924
                                                Root MSE          =     16.132

------------------------------------------------------------------------------
             |               Robust
        wkhp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         noc |  -2.274926   .0218837  -103.96   0.000    -2.317817   -2.232035
     yearsEd |   1.353838   .0090033   150.37   0.000     1.336192    1.371484
        agep |  -.0496699   .0033541   -14.81   0.000    -.0562439   -.0430959
       white |   .4837675   .0578343     8.36   0.000      .370414    .5971211
       _cons |   17.14356   .1879141    91.23   0.000     16.77525    17.51186
------------------------------------------------------------------------------
```

- Robustness matters little in this case, but the fact that robust SEs are a little bigger than old-fashioned SEs improves confidence interval coverage