Linear Regression with One Regressor: Hypothesis Tests and Confidence Intervals

SW Chapter 5

5.1 Testing hypotheses about one regression coefficient

5.2 Confidence Intervals for β_1

5.3 Regression when X is binary

5.4 Heteroskedasticity and homoskedasticity

5.5 Gauss-Markov Theorem

Learning objectives

- Create hypotheses about slope coefficients and test them using $\hat{\beta}_1$ and its standard error.
- Correctly interpret the results of hypothesis tests
- Calculate confidence intervals for β_1
- Take binary regressors in stride (and interpret them correctly)
- Understand the implications of heteroskedasticity and correct your standard errors
- Know and apply the Gauss-Markov theorem to understand the circumstances under which OLS is BLUE.

Where we are going

We want to learn about the slope of the population regression line. We have data from a sample, so there is sampling uncertainty.

- State the population object of interest
- Provide an estimator of this population object
- Derive the sampling distribution of the estimator (this requires certain assumptions). In large samples, it will be normal by the CLT.
- Find the standard error (SE) of the estimator
- Construct t-statistics (for hypothesis tests) and confidence intervals.

Testing hypotheses about one coefficient

The sampling distribution of $\hat{\beta}_1$

Under the Least Squares Assumptions, for n large, $\hat{\beta}_1$ is approximately distributed,

$$\boxed{\hat{\beta}_1 \sim N(\beta_1, \frac{\sigma_v^2}{n(\sigma_X^2)^2}), \text{ where } v_i = (X_i - \mu_X)u_i)}$$

Note: We won't computer variances by hand, but the intuition is useful!

Hypothesis testing: general setup

- Null hypothesis and **two-sided** alternative:

$$\boxed{H_0 : \beta_1 = \beta_{1,0} \text{ vs } H_1 : \beta_1 \neq \beta_{1,0}}$$

- Null hypothesis and **one-sided** alternative:

$$\boxed{H_0 : \beta_1 = \beta_{1,0} \text{ vs } H_1 : \beta_1 < \beta_{1,0}}$$

where $\beta_{1,0}$ is the hypothesized value of β_1 under the null.

General approach

- In general:

$$\boxed{t = \frac{\text{estimator - hypothesized value}}{\text{SE of estimator}}}$$

  where SE of the estimator is the square root of an estimate of the variance of the estimator.
- For testing $\bar{Y}$, recall that $t = \dfrac{\bar{Y} - \mu_{Y,0}}{s_y/\sqrt{n}}$
- For testing β_1,

$$\boxed{t = \frac{\hat{\beta}_1 - \beta_{1,0}}{SE(\hat{\beta}_1)}}$$

Testing $H_0 \beta_{1,0} = 0$

- Construct your t-statistic:

$$\boxed{t = \frac{\hat{\beta}_1 - \beta_{1,0}}{SE(\hat{\beta}_1)}}$$

- Reject at α significance level if $|t| > c_{\alpha/2}$
- In practice, almost always two-tailed tests.
- This procedure relies on large-n approximately that $\hat{\beta}_1$ is normally distributed, requires at least $n > 30$ for CLT to kick in

Level	α	$c_{\alpha/2}$
1%	0.01	2.58
5%	0.05	1.96
10%	0.10	1.645

Choosing a significance level

- We usually fix α as the **significance level** of our test, or **type I error**, the probability of falsely rejecting the null hypothesis.
- We usually set $\alpha = 0.05$. So 5% of the time, we'll reject the null when it's actually true, a "false positive"
- Is that too high? Why not make α super, super small?

Choosing a significance level

- The smaller is α, the harder it is to reject H_0. So we'll see fewer fase positives, but we'll also see more false negatives!
- **Power** is the probability that we reject the null when the alternate hypothesis is true, equal to $1 - \beta$, where β is probability of **type II error**
- There is a tradeoff between significance, α, and power, β.
- We want to strike a good balance!

Conducting a hypothesis test

Is mother's education associated with birthweight?

```
.
.
.  regress bwght motheduc,robust

Linear regression                               Number of obs     =      1,387
                                                F(1, 1385)        =       7.44
                                                Prob > F          =     0.0065
                                                R-squared         =     0.0048
                                                Root MSE          =     20.318
```

bwght	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
motheduc	.5921371	.2170622	2.73	0.006	.1663308	1.017943
_cons	111.0482	2.849696	38.97	0.000	105.458	116.6384

```
.
```

Testing $H_0 : \beta_{1,0} = a$

- What if we have more general hypotheses?
- Null hypothesis $H_0 : \beta_1 = a$
- Just adjust t-statistic!

$$t = \frac{\text{estimator - hypothesized value}}{\text{SE of estimator}} \frac{\hat{\beta}_1 - a}{SE(\hat{\beta}_1)}$$

Economic vs. statistical significance

- If statistically significant, examine magnitude. Does it actually matter?
  - Statistically significant $\neq$ economically or practically significant!
- If a variable is statistically and economically important but has the "wrong" sign, the regression model might be misspecified
- If a variable is statistically insignificant at the usual levels (10%, 5%, or 1%), may want to exclude it from the regression
  - Not necessarily, though, when samples are small

Confidence Intervals for β_1

Confidence Intervals for β_1

Recall, that a 95% confidence interval is, equivalently:

- The set of points that cannot be rejected at the 5% significance level;
- A set-valued function of the data (an interval that is a function of the data) that contains the true parameter value 95% of the time in repeated samples.

Because the t-statistic for β_1 is distributed $N(0,1)$ in large samples, construction of a 95% confidence for β_1 is just like the case of the sample mean!

95% confidence interval for β_1: $\hat{\beta}_1 \pm 1.96 SE(\hat{\beta}_1)$

Regression when X is binary

Regression when X is binary

Sometimes a regressor is **binary**:

- $X = 1$ if small class size, $X = 0$ if not
- $X = 1$ if female, $X = 0$ if male**
- $X = 1$ if treated (experimental drug),$X = 0$ if not

Binary regressors are sometimes called **dummy variables**.

So far, β_1 has been called a "slope," but that doesn't make sense if X is binary.

Gender is not binary, but it* **is *binary in many, many data sets. Just another example of how data availability shapes our understanding of the world!*

Interpreting a binary X

Recall the population model: $Y_i = \beta_0 + \beta_1 X_i + u_i$

When $X_i = 0$, $Y_i = \beta_0 + u_i$

- The mean of Y_i is β_0
- That is, $E[Y_i|X_i] = 0 = \beta_0$

When $X_i = 1$, $Y_i = \beta_0 + \beta_1 + u_i$

- The mean of Y_i is $\beta_0 + \beta_1$
- That is, $E[Y_i|X_i] = 0 = \beta_0 + \beta_1$

Therefore, $\beta_1 = E(Y_i|X_i = 1) - E(Y_i|X_i = 0)$, which is the population difference in group means.

Interpreting a binary X

Is sex associated with birthweight?

```
.
.
. regress bwght male,robust

Linear regression                               Number of obs     =      1,388
                                                F(1, 1386)        =       7.27
                                                Prob > F          =     0.0071
                                                R-squared         =     0.0052
                                                Root MSE          =     20.308

------------------------------------------------------------------------------
             |               Robust
       bwght |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        male |    2.94235   1.091232     2.70   0.007      .801704    5.082995
       _cons |   117.1669   .7882632   148.64   0.000     115.6206    118.7132
------------------------------------------------------------------------------

.
```

Interpreting a binary X

Is sex associated with birthweight?

$$\widehat{bwght} = 117.17 + 2.94 male$$

Average birthweight of female babies:

$E[bwght|male = 0] = 117.17$ ounces

Average birthweight of male babies:

$E[bwght|male = 1] = 117.17 + 2.94 = 120.11$ ounces

Heteroskedasticity and homoskedasticity

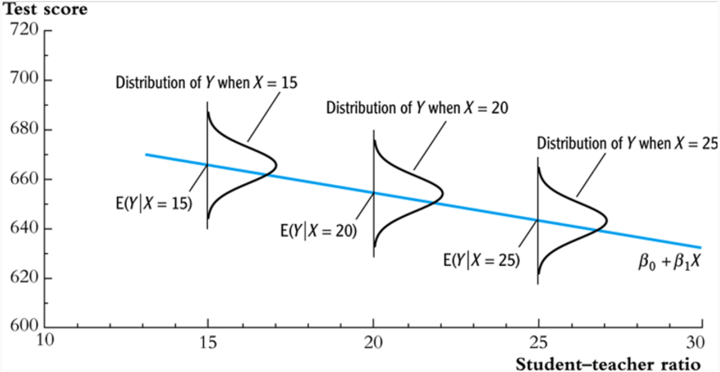
Heteroskedasticity and homoskedasticity

1. WTF?
2. Consequences of heteroskedasticity
3. Implications for computing standard errors

What do these two terms mean?

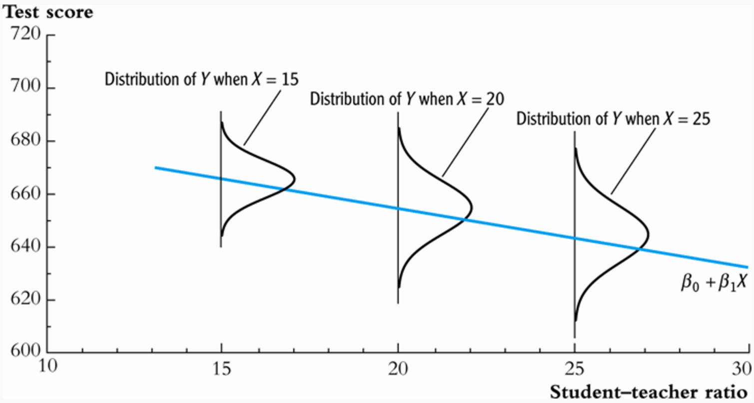
If $var(u|X = x)$ is constant, then u is said to be **homoskedastic**. Otherwise, u is **heteroskedastic**.

Heteroskedasticity in a picture



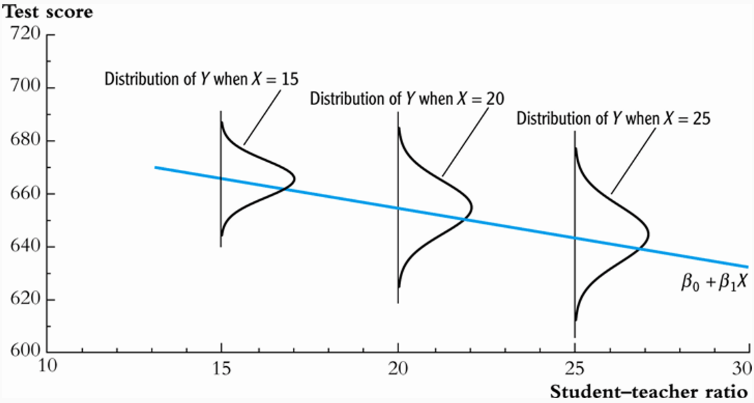
- The variance of u is constant (in fact here, $E(u|X = x) = 0$ (least square assumption 1 satisfied!)
- The variance of u does not depend on X

Homoskedasticity in a picture



- The variance of u not constant
- The variance of u **does** depend on X

Heteroskedasticity in a picture



- The variance of u not constant
- The variance of u **does** depend on X

Does heteroskedasticity affect $\hat{\beta}_1$?

Recall the three least squares assumptions:

1. $E(u|X = x) = 0$
2. $(X_i, Y_i), i = 1, \ldots, n$ are i.i.d.
3. Large outliers are rare

Heteroskedasticity and homoskedasticity concern $var(u|X = x)$.

Because we have not explicitly assumed homoskedastic errors, we have implicitly allowed for heteroskedasticity.

So who cares?

As we just saw, heteroskedasticity does not affect point estimates of β_1. But, as you might expect, it does affect your standard errors!

The previously estimated standard errors are unbiased only under homoskedastic. We will adjust our standard errors to reflect heteroskedasticity, but only in statistical packages. We will call them **heteroskedasticity-robust standard errors**, because they are valid whether or not the errors are heteroskedastic.

Heteroskedasticity-robust standard errors in Stata

```
.
.  regress bwght cigs

      Source |       SS           df       MS      Number of obs   =     1,388
-------------+----------------------------------   F(1, 1386)      =     32.24
       Model |  13060.4194         1  13060.4194   Prob > F        =    0.0000
    Residual |    561551.3     1,386  405.159668   R-squared       =    0.0227
-------------+----------------------------------   Adj R-squared   =    0.0220
       Total |   574611.72     1,387  414.283864   Root MSE        =    20.129

------------------------------------------------------------------------------
       bwght |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cigs |  -.5137721   .0904909    -5.68   0.000    -.6912861   -.3362581
       _cons |   119.7719   .5723407   209.27   0.000     118.6492    120.8946
------------------------------------------------------------------------------

.
```

Heteroskedasticity-robust standard errors in Stata

```
.
.
.  regress bwght cigs,robust

Linear regression                               Number of obs     =      1,388
                                                F(1, 1386)        =      34.29
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0227
                                                Root MSE          =     20.129

------------------------------------------------------------------------------
             |               Robust
       bwght |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cigs |  -.5137721   .0877334    -5.86   0.000    -.6858767   -.3416675
       _cons |   119.7719   .5745494   208.46   0.000     118.6448     120.899
------------------------------------------------------------------------------

.
```

Heteroskedasticity: the bottom line

- If the errors are either homoskedastic or heteroskedastic and you use heteroskedastic-robust standard errors, you are OK
- If the errors are heteroskedastic and you use the homoskedasticity-only formula for standard errors, your standard errors will be wrong
  - Could be too big or too small!
- The two formulas coincide (when n is large) in the special case of homoskedasticity
- So, you should always use heteroskedasticity-robust standard errors.

Gauss-Markov Theorem

The Extended Least Squares Assumptions

Consider our three LS assumptions (needed for unbiasedness):

1. $E(u|X = x) = 0$
2. $(X_i, Y_i), i = 1, \ldots, n$ are i.i.d.
3. Large outliers are rare

Plus, one more!

4. u is homoskedastic

Gauss-Markov Theorem

Under these **four** extended LS assumptions, $\hat{\beta}_1$ has the smallest variance among *all linear estimators* (estimators that are linear functions of $Y_1, ..., Y_n$).

This is the **Gauss-Markov theorem**

Under the GM theory, OLS estimators are **BLUE**:

- Best
- Linear
- Unbiased
- Estimators

OLS limitations

- Homoskedasticity often doesn't hold (homoskedasticity is special)
- The result is only for linear estimators – only a small subset of estimators
  - If we know nature of heteroskedasticity, can model it with weighted least squares, which is more efficient
  - if we have a lot of outliers, then least absolute deviations (LAD) estimators will be more efficient

In most applied regression analysis, we use OLS - so that is what we will do, too!

Conclusion

5.1 Testing hypotheses about one regression coefficient

5.2 Confidence Intervals for β_1

5.3 Regression when X is binary

5.4 Heteroskedasticity and homoskedasticity

5.5 Gauss-Markov Theorem