Section 3. Simple Regression
Confidence Intervals, $R^2$

1. Who invented regression?
2. $R^2$
3. Review – 2 formulas and the story so far
1. Who invented regression?

• Francis Galton,
  - climatologist,
  - gentleman explorer
  - social scientist
Heredity and Height

“regression” to the mean

### Table 1

<table>
<thead>
<tr>
<th>Mid-Parents</th>
<th>Adult Children</th>
<th>Heights in inches</th>
<th>Deviations in inches</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>64</td>
<td>-4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>65</td>
<td>-3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>66</td>
<td>-2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>67</td>
<td>-1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>68</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>69</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>70</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>71</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>72</td>
<td>4</td>
</tr>
</tbody>
</table>

![Diagram](image-url)
2. \( R^2 \)

- How much of the variation in \( Y \) did we explain with the regression line?

\[
R^2 = \frac{\sum_{i=1}^{N} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2}
\]

\[
= 1 - \frac{\sum_{i=1}^{N} e_i^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2}
\]
Coffee Demand – High $R^2$
E.g. Coffee Demand – high $R^2$

- $p$ is the price of coffee,
- $q$ is the quantity (in cups)

```
. reg q p, robust

Regression with robust standard errors

|                 | Coef.    | Std. Err. | t     | P>|t|    | [95% Conf. Interval] |
|-----------------|----------|-----------|-------|--------|----------------------|
| $q$             | 17.5064  | 1.822797  | 9.60  | 0.000  | 13.71568 - 21.29711  |
| $p$             | -6.246766| 1.176301  | -5.31 | 0.000  | -8.693018 - -3.800513|
| _cons           | 17.5064  | 1.822797  | 9.60  | 0.000  | 13.71568 - 21.29711  |
```

Number of obs = 23
$F( 1, 21) = 28.20$
Prob > F = 0.0000
R-squared = 0.7349
Root MSE = 4.0549
Eg. Wage Regression - Low $R^2$

* Lhwage is log(hourly wage), ed is years of education

```
regress lhwage ed, robust
```

Regression with robust standard errors

|            | Coef.       | Std. Err. | t    | P>|t|  | [95% Conf. Interval] |
|------------|-------------|-----------|------|------|---------------------|
| lhwage     | ed          | 0.0704563 | 0.0016628 | 42.37 | 0.000 | 0.0671969 - 0.0737156 |
|            | _cons       | 0.9852746 | 0.0238393 | 41.33 | 0.000 | 0.9385464 - 1.032003 |
```

Number of obs = 13743
F( 1, 13741) = 1795.40
Prob > F = 0.0000
R-squared = 0.1185
Root MSE = 0.5083
3. Review

- Confidence Intervals
- Omitted Variable Bias
How to Review:
Notes, Problem Sets, Reading, OH

1. Introduction: Why Study Econometrics?

2. Probability and Statistics: A quick review
Probability, random variables, the normal distribution and the central limit theorem, inference, confidence intervals and hypothesis testing. Asymptotics of the sample mean. Using Stata. Reading: Chapters #2 and #3.

3. Simple Regression (one regressor)
Fitting a line through a cloud of points. Least squares, unbiased estimates, consistent estimates, confidence intervals, hypothesis testing, omitted variable bias, $R^2$. Reading: Chapter #4. And a little Ch #5 on multivariate regression and omitted variable bias.

{Review and midterm about here}
Confidence Intervals

Confidence Intervals for $\beta_1$

A 95% two-sided confidence interval for $\beta_1$ is an interval that contains the true value of $\beta_1$ with a 95% probability, that is, it contains the true value of $\beta_1$ in 95% of all possible randomly drawn samples. Equivalently, it is also the set of values of $\beta_1$ that cannot be rejected by a 5% two-sided hypothesis test. When the sample size is large, it is constructed as

$$95\% \text{ confidence interval for } \beta_1 = (\hat{\beta}_1 - 1.96SE(\hat{\beta}_1), \hat{\beta}_1 + 1.96SE(\hat{\beta}_1)). \quad (4.27)$$

..and the same for $\beta_0$
Experiments eliminate OVB

\[ b_1^s = b_1^L + b_2^L b_{21}, \]

- So there’s no OVB if \( b_{21} = 0 \)
  i.e., \( b_{21} = 0 \) implies \( b_1^s = b_1^L \)
  .. Which you can guarantee if you design an experiment in which \( X_1 \) is uncorrelated with other \( X \)’s (omitted variables).
    Random assignment of \( X_1 \) is sure to do that.
  .. Back to examples to demonstrate