7 – Joint Hypothesis Tests

• Recap
• Confidence sets
Exercise

Coefficients:

|        | Estimate | Std. Error | t value | Pr(>|t|) |
|--------|----------|------------|---------|----------|
| (Intercept) | -0.6246  | 0.4660     | -1.340  | 0.182    |
| x1      | 0.2161   | 0.1723     | 1.255   | 0.211    |
| x2      | -0.1092  | 0.1153     | -0.946  | 0.345    |
| x3      | 2.9384   | 0.1092     | 26.914  | <2e-16 ***|

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.575 on 196 degrees of freedom
Multiple R-squared:  0.7921,  Adjusted R-squared:  0.7889
F-statistic: 248.9 on 3 and 196 DF,  p-value: < 2.2e-16

a) Calculate the 95% CI for $b_1$
b) Calculate the 95% CI for $b_2$
a)
Null hypothesis
Confidence set for $b_1$ & $b_2$: reject the null!
• the idea of confidence sets reinforces the idea that individual $t$-tests can’t be used for joint hypotheses
• confidence sets aren’t used in practice (in econometrics)
Aside: the overall F-test

A good idea might be to test if all of the variables are garbage:

\[ H_0: \beta_1 = \beta_1 = \ldots = \beta_k = 0 \]

\[ H_A: \text{at least one } \beta \neq 0 \]

- the intercept is not tested
- this “overall F-test” is usually reported by your econometric software
Coefficients:

|        | Estimate | Std. Error | t value | Pr(>|t|) |
|--------|----------|------------|---------|---------|
| (Intercept) | -0.6246  | 0.4660     | -1.340  | 0.182   |
| x1      | 0.2161   | 0.1723     | 1.255   | 0.211   |
| x2      | -0.1092  | 0.1153     | -0.946  | 0.345   |
| x3      | 2.9384   | 0.1092     | 26.914  | <2e-16  *** |

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.575 on 196 degrees of freedom
Multiple R-squared:  0.7921,  Adjusted R-squared:  0.7889
F-statistic: 248.9 on 3 and 196 DF,  p-value: < 2.2e-16

Residual standard error?

Now we know what all of this R output means.
Model selection/building

- We will typically be interested in studying the marginal effects of a few variables
- Other variables are included to avoid OVB
- So, estimate several “candidate” models – maybe start big
- Use judgement
- Use $t$-tests/F-tests to select among models
- Don’t just try to maximize $R^2$
Presenting results

Now that we have lots of variables in our models, and several different estimated models, we should present our results in tables, and include:

- dependent variable
- estimated regression coefficients
- standard errors
- significance codes (e.g. **)
- measures of fit
- $n$
- relevant F-stats (if any)
Dependent variable: Price. \( n = 1728 \).

<table>
<thead>
<tr>
<th>Regressor</th>
<th>Model (1)</th>
<th>Model (2)</th>
<th>Model (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>20.27</td>
<td>22.46*</td>
<td>17.51*</td>
</tr>
<tr>
<td>Lot.Size</td>
<td>7.60***</td>
<td>7.29***</td>
<td>7.41***</td>
</tr>
<tr>
<td>Waterfront</td>
<td>120.20***</td>
<td>119.20***</td>
<td>120.40***</td>
</tr>
<tr>
<td>Age</td>
<td>-0.13*</td>
<td>-0.14*</td>
<td>-0.14*</td>
</tr>
<tr>
<td>Land.Value</td>
<td>0.00***</td>
<td>0.00***</td>
<td>0.00***</td>
</tr>
<tr>
<td>New.Construct</td>
<td>-45.44***</td>
<td>-45.16***</td>
<td>-44.50***</td>
</tr>
<tr>
<td>Central.Air</td>
<td>9.95**</td>
<td>9.90**</td>
<td>9.65**</td>
</tr>
<tr>
<td>fuel3</td>
<td>-10.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fuel4</td>
<td>-4.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>heat3</td>
<td>-10.45*</td>
<td>-10.53*</td>
<td>-10.55*</td>
</tr>
<tr>
<td>heat4</td>
<td>-0.08</td>
<td>-9.94*</td>
<td>-9.98*</td>
</tr>
<tr>
<td>sewer2</td>
<td>4.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sewer3</td>
<td>3.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Living.Area</td>
<td>0.07***</td>
<td>0.07***</td>
<td>0.07***</td>
</tr>
<tr>
<td>Pct.College</td>
<td>-0.11</td>
<td>-0.10</td>
<td></td>
</tr>
<tr>
<td>Bedrooms</td>
<td>-7.84**</td>
<td>-7.64**</td>
<td>-7.75**</td>
</tr>
<tr>
<td>Fireplaces</td>
<td>1.04</td>
<td>1.06</td>
<td></td>
</tr>
<tr>
<td>Bathrooms</td>
<td>23.11***</td>
<td>23.04***</td>
<td>23.14***</td>
</tr>
<tr>
<td>Rooms</td>
<td>3.02**</td>
<td>3.05**</td>
<td>3.04**</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.65</td>
<td>0.65</td>
<td>0.65</td>
</tr>
<tr>
<td>F-statistic against Model (1)</td>
<td>0.40</td>
<td>0.35</td>
<td></td>
</tr>
</tbody>
</table>

Coefficient is statistically significant at the 5% (*), 1% (**), and 0.1% (***).