Multiple Regression Analysis

\[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u \]
Assumptions of the Classical Linear Model (CLM)

- So far, we know that given the Gauss-Markov assumptions, OLS is BLUE
- In order to do classical hypothesis testing, we need to add another assumption (beyond the Gauss-Markov assumptions)
- The Normality Assumption
  - Assume that \( u \) is independent of \( x_1, x_2, \ldots, x_k \) and \( u \) is normally distributed with zero mean and variance \( \sigma^2 \): \( u \sim \text{Normal}(0, \sigma^2) \)
CLM Assumptions (cont)

- Under CLM, OLS is not only BLUE, but is the minimum variance unbiased estimator.
- We can summarize the population assumptions of CLM as follows:
  \[ y \mid x \sim \text{Normal}(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k, \sigma^2) \]
- While for now we just assume normality, clear that sometimes not the case.
- Large samples will let us drop normality.
The homoskedastic normal distribution with a single explanatory variable

\[ E(y|x) = \beta_0 + \beta_1 x \]

Normal distributions
Normal Sampling Distributions

Under the CLM assumptions, conditional on the sample values of the independent variables

\[ \hat{\beta}_j \sim \text{Normal}[\beta_j, \text{Var}(\hat{\beta}_j)] \]

so that

\[ \frac{(\hat{\beta}_j - \beta_j)}{\text{sd}(\hat{\beta}_j)} \sim \text{Normal}(0, 1) \]

\( \hat{\beta}_j \) is distributed normally because it is a linear combination of the errors
The $t$ Test

Under the CLM assumptions

\[
\frac{(\hat{\beta}_j - \beta_j)}{se(\hat{\beta}_j)} \sim t_{n-k-1}
\]

Note this is a $t$ distribution (vs normal) because we have to estimate $\sigma^2$ by $\hat{\sigma}^2$

Note the degrees of freedom: $n - k - 1$
The $t$ Test (cont)

- Knowing the sampling distribution for the standardized estimator allows us to carry out hypothesis tests
- Start with a null hypothesis
- For example, $H_0: \beta_j = 0$
- If accept null, then accept that $x_j$ has no effect on $y$, controlling for other $x$'s
The \( t \) Test (cont)

To perform our test we first need to form

"the" \( t \) statistic for \( \hat{\beta}_j : t_{\hat{\beta}_j} = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \)

We will then use our \( t \) statistic along with a rejection rule to determine whether to accept the null hypothesis, \( H_0 \).
\( t \) Test: One-Sided Alternatives

Besides our null, \( H_0 \), we need an alternative hypothesis, \( H_1 \), and a significance level.

- \( H_1 \) may be one-sided, or two-sided.
- \( H_1: \beta_j > 0 \) and \( H_1: \beta_j < 0 \) are one-sided.
- \( H_1: \beta_j \neq 0 \) is a two-sided alternative.

If we want to have only a 5\% probability of rejecting \( H_0 \) if it is really true, then we say our significance level is 5\%.
One-Sided Alternatives (cont)

- Having picked a significance level, $\alpha$, we look up the $(1 - \alpha)^{th}$ percentile in a $t$ distribution with $n - k - 1$ df and call this $c$, the critical value.
- We can **reject** the null hypothesis if the $t$ statistic is greater than the critical value.
- If the $t$ statistic is less than the critical value then we **fail to reject** the null.
One-Sided Alternatives (cont)

\[ y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + u_i \]

\[ H_0: \beta_j = 0 \]
\[ H_1: \beta_j > 0 \]

Fail to reject

\[ (1 - \alpha) \]

\[ \alpha \]

0, c
One-sided vs Two-sided

Because the $t$ distribution is symmetric, testing $H_1: \beta_j < 0$ is straightforward. The critical value is just the negative of before.

We can reject the null if the $t$ statistic < $-c$, and if the $t$ statistic > than $-c$ then we fail to reject the null.

For a two-sided test, we set the critical value based on $\alpha/2$ and reject $H_1: \beta_j \neq 0$ if the absolute value of the $t$ statistic > $c$. 
Two-Sided Alternatives

\[ y_i = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_k X_{ik} + u_i \]

\( H_0: \beta_j = 0 \)
\( H_1: \beta_j \neq 0 \)

 rejects \[ \alpha/2 \] and \[ (1 - \alpha) \]

 fail to reject \[ \alpha/2 \] and \[ (1 - \alpha) \]
Summary for $H_0: \beta_j = 0$

- Unless otherwise stated, the alternative is assumed to be two-sided.
- If we reject the null, we typically say “$x_j$ is statistically significant at the $\alpha \%$ level.”
- If we fail to reject the null, we typically say “$x_j$ is statistically insignificant at the $\alpha \%$ level.”
Testing other hypotheses

- A more general form of the $t$ statistic recognizes that we may want to test something like $H_0: \beta_j = a_j$

- In this case, the appropriate $t$ statistic is

$$t = \frac{\hat{\beta}_j - a_j}{se(\hat{\beta}_j)}$$

where $a_j = 0$ for the standard test
Confidence Intervals

Another way to use classical statistical testing is to construct a confidence interval using the same critical value as was used for a two-sided test.

A \((1 - \alpha)\) % confidence interval is defined as

\[
\hat{\beta}_j \pm c \cdot se(\hat{\beta}_j), \text{ where } c \text{ is the } \left(1 - \frac{\alpha}{2}\right) \text{ percentile in a } t_{n-k-1} \text{ distribution}
\]
Computing \( p \)-values for \( t \) tests

- An alternative to the classical approach is to ask, “what is the smallest significance level at which the null would be rejected?”
- So, compute the \( t \) statistic, and then look up what percentile it is in the appropriate \( t \) distribution – this is the \( p \)-value
- \( p \)-value is the probability we would observe the \( t \) statistic we did, if the null were true
Stata and $p$-values, $t$ tests, etc.

- Most computer packages will compute the $p$-value for you, assuming a two-sided test.
- If you really want a one-sided alternative, just divide the two-sided $p$-value by 2.
- Stata provides the $t$ statistic, $p$-value, and 95% confidence interval for $H_0: \beta_j = 0$ for you, in columns labeled “t”, “P > |t|” and “[95% Conf. Interval]”, respectively.

**IMPORTANT:** small $p$-values are evidence against the null; large values provide little evidence against the null.
Statistical VS Economic Significance

- The statistical significance of a variable $x_j$ is determined entirely by the size of $t_\beta$ which depends on:
  - The size of $\beta$
  - Standard error of $\beta$

- The economic (or practical) significance depends not just on the size but the sign of $\beta$
Testing a Linear Combination

Suppose instead of testing whether $\beta_1$ is equal to a constant, you want to test if it is equal to another parameter, that is $H_0 : \beta_1 = \beta_2$

Use same basic procedure for forming a $t$ statistic

$$t = \frac{\hat{\beta}_1 - \hat{\beta}_2}{se(\hat{\beta}_1 - \hat{\beta}_2)}$$
Testing Linear Combo (cont)

Since

\[ se(\hat{\beta}_1 - \hat{\beta}_2) = \sqrt{Var(\hat{\beta}_1 - \hat{\beta}_2)} \]

then

\[ Var(\hat{\beta}_1 - \hat{\beta}_2) = Var(\hat{\beta}_1) + Var(\hat{\beta}_2) - 2Cov(\hat{\beta}_1, \hat{\beta}_2) \]

\[ se(\hat{\beta}_1 - \hat{\beta}_2) = \left\{ se(\hat{\beta}_1)^2 + se(\hat{\beta}_2)^2 - 2s_{12} \right\}^{\frac{1}{2}} \]

where \( s_{12} \) is an estimate of \( Cov(\hat{\beta}_1, \hat{\beta}_2) \)
Testing a Linear Combo (cont)

- So, to use formula, need $s_{12}$, which standard output does not have.
- Many packages will have an option to get it, or will just perform the test for you.
- In Stata, after `reg y x1 x2 ... xk` you would type `test x1 = x2` to get a $p$-value for the test.
- More generally, you can always restate the problem to get the test you want.
Multiple Linear Restrictions

Everything we’ve done so far has involved testing a single linear restriction, (e.g. $\beta_1 = 0$ or $\beta_1 = \beta_2$)

However, we may want to jointly test multiple hypotheses about our parameters

A typical example is testing “exclusion restrictions” – we want to know if a group of parameters are all equal to zero
Testing Exclusion Restrictions

- Now the null hypothesis might be something like \( H_0: \beta_{k-q+1} = 0, \ldots, \beta_k = 0 \)
- The alternative is just \( H_1: H_0 \) is not true
- Can’t just check each \( t \) statistic separately, because we want to know if the \( q \) parameters are jointly significant at a given level – it is possible for none to be individually significant at that level
Exclusion Restrictions (cont)

To do the test we need to estimate the “restricted model” without $x_{k-q+1}, \ldots, x_k$ included, as well as the “unrestricted model” with all $x$’s included.

Intuitively, we want to know if the change in SSR is big enough to warrant inclusion of $x_{k-q+1}, \ldots, x_k$.

$$F \equiv \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k-1)}, \text{ where}$$

$r$ is restricted and $ur$ is unrestricted.
The $F$ statistic

- The $F$ statistic is always positive, since the SSR from the restricted model can’t be less than the SSR from the unrestricted.
- Essentially the $F$ statistic is measuring the relative increase in SSR when moving from the unrestricted to restricted model.
- $q = \text{number of restrictions, or } df_r - df_{ur}$
- $n - k - 1 = df_{ur}$
The $F$ statistic (cont)

To decide if the increase in SSR when we move to a restricted model is “big enough” to reject the exclusions, we need to know about the sampling distribution of our $F$ stat.

Not surprisingly, $F \sim F_{q,n-k-1}$, where $q$ is referred to as the numerator degrees of freedom and $n - k - 1$ as the denominator degrees of freedom.
The $F$ statistic (cont)

Reject $H_0$ at $\alpha$ significance level if $F > c$
The $R^2$ form of the $F$ statistic

Because the SSR’s may be large and unwieldy, an alternative form of the formula is useful.

We use the fact that $SSR = SST(1 - R^2)$ for any regression, so can substitute in for $SSR_u$ and $SSR_{ur}$

$$F \equiv \frac{(R_{ur}^2 - R_r^2)/q}{(1 - R_{ur}^2)/(n - k - 1)}$$

where again $r$ is restricted and $ur$ is unrestricted
Overall Significance

- A special case of exclusion restrictions is to test $H_0$: $\beta_1 = \beta_2 = \ldots = \beta_k = 0$
- Since the $R^2$ from a model with only an intercept will be zero, the $F$ statistic is simply

$$F = \frac{R^2 / k}{\left(1 - R^2\right) / (n - k - 1)}$$
General Linear Restrictions

- The basic form of the $F$ statistic will work for any set of linear restrictions.
- First estimate the unrestricted model and then estimate the restricted model.
- In each case, make note of the SSR.
- Imposing the restrictions can be tricky – will likely have to redefine variables again.
**F Statistic Summary**

- Just as with $t$ statistics, $p$-values can be calculated by looking up the percentile in the appropriate $F$ distribution.

- Stata will do this by entering: `display fprob(q, n - k - 1, F)`, where the appropriate values of $F$, $q$, and $n - k - 1$ are used.

- If only one exclusion is being tested, then $F = t^2$, and the $p$-values will be the same.