1 Hypothesis Testing in Multiple Regression

Let us start with a concrete example:

$$\texttt{testscr} = \beta_0 + \beta_1 \texttt{str} + \beta_2 \texttt{expn} + \beta_3 \texttt{pct\_el} + \epsilon$$

where `expn` measures expenditure per pupil and `pct_el` measures percentage of english learners in the classroom.

The null hypothesis that "school resources don't matter" is

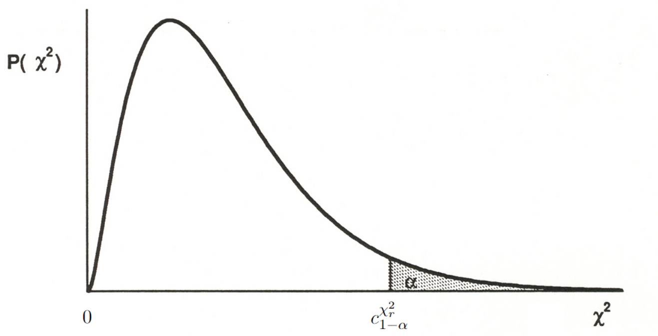
$$H : \beta_1 = 0, \beta_2 = 0$$
$$K : \text{ either } \beta_1 \neq 0 \text{ or } \beta_2 \neq 0 \text{ or neither is equal to } 0$$

We wish to write this as $H : R\beta = c$ against the two-sided alternative $K : R\beta \neq c$ for a known $r \times (k+1)$ dimensional restriction matrix R and a known $r \times 1$ vector c.

In our example,

$$R_{2\times 4} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$
$$\beta_{4\times 1} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = (\beta_0, ..., \beta_3)'$$

Strategy: Reject H if the ("squared") distance between $\left(R\hat{\beta} - c\right)$ is large enough and look up critical values in a χ^2_r distribution.[1]



Speficially, recall that, under Assumptions 1-5, it is reasonable to assume that $\sqrt{n}\left(\hat{\beta} - \beta\right)$ was distributed as a $(k+1)$ dimensional normal random variable with mean vector $\mu_{(k+1)\times 1} = (0, 0..., 0)'$ and variance matrix $\hat{V}_{(k+1)\times(k+1)}$ where

$$\hat{V} = \left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i'\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i e_i^2 \mathbf{X}_i'\right) \left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i'\right)^{-1}$$

for the residual $e_i = Y_i - \mathbf{X}_i'\hat{\beta}$.

Therefore,

$$\sqrt{n}\left(R\hat{\beta} - R\beta\right) \xrightarrow[n\to\infty]{d} \mathcal{N}_r\left(0, RVR'\right)$$

where r is the number of (linearly independent) rows of R.

[1]See Appendix table attached to these notes.

Now, under H, $R\beta = c$ so

$$\sqrt{n}\left(R\hat{\beta} - c\right) \xrightarrow[n\to\infty]{d} \mathcal{N}_r\left(0, RVR'\right)$$

Also, we have a consistent estimator $\hat{V} \underset{n\to\infty}{\overset{p}{\to}} V$.

So under H, the Wald Statistic

$$W_n = \sqrt{n}\left(R\hat{\beta} - c\right)'\left(R\hat{V}R'\right)^{-1}\sqrt{n}\left(R\hat{\beta} - c\right)$$

converges to a chi-squared distribution. The degrees of freedom corresonds to the rank of R, or the number of restrictions in H. This gives us the distributional results we need. Under the null H, the random variable W_n is distributed as a χ^2_r.

We then ask if it is likely that our observed value of W_n comes from a χ^2_r distribution. If our observed value of W_n is much larger than we would expect if the distribution were χ^2_r, then we reject the null H.

As before, to find a precise definition of large we find a critical value c_1 such that the probability of making a Type I error is equal to some small α,

$$\mathbb{P}_H\left(W_n > c_1\right) = \alpha$$

and this critical value is the $(1-\alpha)^{th}$ quantile of a chi-squared distribution with r degrees of freedom and which we denote by $c^{\chi^2_r}_{1-\alpha}$. We can use a critical value table to find this critical value.

Exercise 1:

1. Explain how you would test the null hypothesis that $\beta_1 = 0$ in the multiple regression model $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$.

2. Explain how you would test the null hypothesis that $\beta_2 = 0$.

3. Explain how you would test the joint hypothesis that $\beta_1 = 0$ and $\beta_2 = 0$.

4. Why isn't the result of the joint test implied by the results of the first two tests?

Exercise 2:

The following regression was computed using data on employees in a developing country. The data set consists of information on over 10,000 full-time, full-year workers. The highest educational achievement for each worker is either a high school diploma or a bachelor's degree. The workers' ages range from 25 to 40 years. The data set also contains information on the region of the country where the person lives, gender, and age. For the purposes of these exercises, let

$$\begin{aligned}
AHE &= \text{average hourly earnings} \\
College &= \text{binary variable (1 if college, 0 if high school)} \\
Female &= \text{binary variable (1 if female, 0 if male)} \\
Age &= \text{(in years)} \\
Northeast &= \text{binary variable (1 if Region = Northeast, 0 otherwise)} \\
Midwest &= \text{binary variable (1 if Region = Midwest, 0 otherwise)} \\
South &= \text{binary variable (1 if Region = South, 0 otherwise)} \\
West &= \text{binary variable (1 if Region = West, 0 otherwise)}
\end{aligned}$$

The regression results are as below:

Results of Regressions of Average Hourly Earnings on Sex and Education Binary Variables and Other Characteristics Using 2015 Data from the Current Population Survey

Dependent variable: average hourly earnings (*AHE*).

Regressor	(1)	(2)	(3)
College (X_1)	10.47 (0.29)	10.44 (0.29)	10.42 (0.29)
Female (X_2)	−4.69 (0.29)	−4.56 (0.29)	−4.57 (0.29)
Age (X_3)		0.61 (0.05)	0.61 (0.05)
Northeast (X_4)			0.74 (0.47)
Midwest (X_5)			−1.54 (0.40)
South (X_6)			−0.44 (0.37)
Intercept	18.15 (0.19)	0.11 (1.46)	0.33 (1.47)
Summary Statistics and Joint Tests			
F-statistic testing regional effects $= 0$			9.32
SER	12.15	12.03	12.01
R^2	0.165	0.182	0.185
n	7178	7178	7178

Using the regression results in column (3):

1. Do there appear to be important regional differences? Use an appropriate hypothesis test to explain your answer.

2. How can we test if the regional effects for the Northeast, Midwest, and South are equivalent?

3. Juanita is a 28-year-old female college graduate from the South. Molly is a 28-year-old female college graduate from the West. Construct a 95% confidence interval for the difference in expected earnings between Juanita and Molly.

2 Nonlinear Regression Functions

The population regression function has been assumed to be linear; that is, it has a constant slope. In the context of causal inference, this constant slope corresponds to the effect on Y of a unit change in X being the same for all values of the regressors. But what if the population regression function is nonlinear? Here we lay out a general strategy for modeling nonlinear population regression functions that are extensions of the multiple regression model and therefore can be estimated and tested using the tools we have developed so far.

The general non-linear regression model

$$Y = g\left(X_1, ..., X_k\right) + \epsilon$$

where

1. $\mathbb{E}\left(\epsilon|X_1, ..., X_k\right) = 0$ This could be either because the errors are defined directly as $y - \mathbb{E}\left(y|\mathbf{X}\right)$ and $\mathbb{E}\left(y|\mathbf{X}\right) = g\left(\mathbf{X}\right)$. Alternatively, this could be a ceterus paribus model and the ϵ are unobservable determinants of y satisfying the conditional mean zero assumption.
2. We observe an i.i.d. sample from the (joint) distribution of $\left(Y, X_1, ..., X_k\right)$
3. No perfect multicollinearity (precise statement depends upon form of $g\left(\cdot\right)$)
4. Enough moments exist (precise statement depends upon precise form of $g\left(\cdot\right)$)

In this course, we restrict attention to functions

$$g\left(\mathbf{X}\right) = \alpha' h\left(\mathbf{X}\right)$$

for vector α and $h\left(\cdot\right)$ known (same dimension as α).

We can use estimated $\hat{\alpha}$ to compute derivatives of interest (for continuously distributed regressors) at prespecified value $\mathbf{x}_0$

$$\delta_j\left(\mathbf{x}_0\right) \equiv \frac{\partial \alpha' h\left(\mathbf{x}_0\right)}{\partial x_j} = \sum_{s=1}^{k} \alpha_s \frac{\partial h_s\left(\mathbf{x}_0\right)}{\partial x_j}$$

consistently estimated by

$$\hat{\delta}_j\left(\mathbf{x}_0\right) = \sum_{s=1}^{k} \hat{\alpha}_s \frac{\partial h_s\left(\mathbf{x}_0\right)}{\partial x_j} = \hat{\alpha}' \frac{\partial h\left(\mathbf{x}_0\right)}{\partial x_j}$$

2.1 Polynomials

One way to specify a nonlinear regression function is to use a polynomial in X. In general, let r denote the highest power of X that is included in the regression. The **polynomial regression model** of degree r is

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \cdots + \beta_r X^r + \epsilon \tag{1}$$

The polynomial regression model is similar to the multiple regression model as stated in handout 4 except that the regressors were distinct independent variables, whereas here the regressors are powers of the same dependent variable, X; that is, the regressors are X, X^2, X^3, and so on. Thus the techniques for estimation and inference developed for multiple regression can be applied here. In particular, the unknown coefficients $\beta_0, \beta_1, \ldots, \beta_r$ in Equation 1 can be estimated by OLS regression of Y against $X, X^2, \ldots X^r$. In the polynomial model, a 1 unit change in X is associated with a change in Y of $\beta_1 + 2\beta_2 X + \cdots + r\beta_r X^{r-1}$.

2.2 Logarithms

There are three different cases in which logarithms might be used: when X is transformed by taking its logarithm but Y is not; when Y is transformed to its logarithm but X is not; and when both Y and X are transformed to their logarithms. The interpretation of the regression coefficients is different in each case. We discuss these three cases in turn.

Case I: X is in logarithms, Y is not. In this case, the regression model is

$$Y = \beta_0 + \beta_1 \log(X) + \epsilon, i = 1, \ldots, n \tag{3}$$

In the **linear-log** model, a 1% change in X is associated with a change in Y of $0.01\beta_1$.

Case II: Y is in logarithms, X is not. In this case, the regression model is

$$\log(Y) = \beta_0 + \beta_1 X + \epsilon, i = 1, \ldots, n \tag{4}$$

In the **log-linear** model, a one-unit change in X ($\Delta X = 1$) is associated with a $(100 \times \beta_1)\%$ change in Y.

Case III: Both X and Y are in logarithms. In this case, the regression model is

$$\log(Y) = \beta_0 + \beta_1 \log(X) + \epsilon, i = 1, \ldots, n \tag{5}$$

In the **log-log** model, a 1% change in X is associated with a $\beta_1\%$ change in Y.

The only difference between the regression model in Equations 3, 4, and 5, and the regression model with a single regressor is that the right-hand variable is now the logarithm of X rather than X itself or the left-hand variable is now the logarithm of Y rather than Y itself. To estimate the coefficients β_0 and β_1 in the equations above, first compute new variables, $ln(Y)$ or $ln(X)$, which are readily done using a spreadsheet or statistical software. Then $\boldsymbol{\beta}$ can be estimated by the OLS regression of the transformation of Y on the transformation of X, hypotheses about β_1 can be tested using the t-statistic, and a 95% confidence interval for β_1 can be constructed as $\hat{\beta}_1 \pm 1.96 \times SE(\hat{\beta}_1)$.

2.3 Interaction Terms

2.3.1 Interacting two dummy variables

Consider the following of wages on marital status, gender and their interactions.

$$wage = \beta_0 + \beta_1 married + \beta_2 female + \beta_3 female \times married + u$$

The conditional expectations of wages are as below:

$$\begin{aligned}
\mathbb{E}[wage|female = 0, married = 0] &= \beta_0 \\
\mathbb{E}[wage|female = 0, married = 1] &= \beta_0 + \beta_1 \\
\mathbb{E}[wage|female = 1, married = 0] &= \beta_0 + \beta_2 \\
\mathbb{E}[wage|female = 1, married = 1] &= \beta_0 + \beta_1 + \beta_2 + \beta_3
\end{aligned}$$

Now let's think about what each coefficient in the regression:

β_0 : average wage for single males.
β_1 : effect of being married for a male.
β_2 : effect of being female for single individuals
β_3 : differential effect of being married for a woman relative to what it is for a man

2.3.2 Interacting a continuous variable and a dummy

$$wage = \beta_0 + \beta_1 female + \beta_2 educ + \beta_3 female \times educ + u$$

What is the marginal effect of education?

We can express the marginal effect of educ as follows:

$$\frac{\partial \mathbb{E}[wage|educ, female]}{\partial educ} = \beta_2 + \beta_3 female$$

Now, you may be asked to substitute a value for female. There are only two possible values for female: 0 and 1. Substituting $female = 0$ (i.e., you are a male) gives:

$$\frac{\partial \mathbb{E}[wage|educ, female = 0]}{\partial educ} = \beta_2$$

Therefore we say that the marginal effect of education on expected wage for **males** is β_2.

Substituting $female = 1$ (i.e., you are a female) gives:

$$\frac{\partial \mathbb{E}[wage|educ, female = 1]}{\partial educ} = \beta_2 + \beta_3$$

Therefore we say that the marginal effect of education on expected wage for **females** is $\beta_2 + \beta_3$.

What is the effect of female?

We can express the additional effect of being female

$$\mathbb{E}[wage|educ, female = 1] - \mathbb{E}[wage|educ, female = 0] = \beta_1 + \beta_3 educ$$

Then you might be asked to substitute a particular value of education (usually we select the median). Substituting $educ = 10$ for example gives:

$$\mathbb{E}[wage|educ = 10, female = 1] - \mathbb{E}[wage|educ = 10, female = 0] = \beta_1 + \beta_3 * 10$$

Therefore we say that the effect of being female on expected wage for people with 10 years of education is $\beta_1 + \beta_3 * 10$.

2.3.3 Interacting two continuous variables

Let's start with the following model:

$$wage = \beta_0 + \beta_1 age + \beta_2 educ + \beta_3 age \times educ + u$$

What is the marginal effect of education?

We can express the marginal effect of education is:

$$\frac{\partial \mathbb{E}[wage|educ, age]}{\partial educ} = \beta_2 + \beta_3 age$$

Then you might be asked to substitute a particular value of educ (usually we select the median). Substituting $age = 20$ for example gives:

$$\frac{\partial \mathbb{E}[wage|educ, age = 20]}{\partial educ} = \beta_2 + \beta_3 * 20$$

Therefore we say that the marginal effect of education on expected wage for people with 20 years of age is $\beta_2 + \beta_3 * 20$.

Exercise 3:

Suppose a researcher collects data on houses that have sold in a particular neighborhood over the past year and obtains the regression results in the table shown below.

Regression Results for Exercise 8.2

Dependent variable: ln(*Price*)

Regressor	(1)	(2)	(3)	(4)	(5)
Size	0.00042 (0.000038)				
ln(*Size*)		0.69 (0.054)	0.68 (0.087)	0.57 (2.03)	0.69 (0.055)
$[\ln(Size)]^2$				0.0078 (0.14)	
Bedrooms			0.0036 (0.037)		
Pool	0.082 (0.032)	0.071 (0.034)	0.071 (0.034)	0.071 (0.036)	0.071 (0.035)
View	0.037 (0.029)	0.027 (0.028)	0.026 (0.026)	0.027 (0.029)	0.027 (0.030)
$Pool \times View$					0.0022 (0.10)
Condition	0.13 (0.045)	0.12 (0.035)	0.12 (0.035)	0.12 (0.036)	0.12 (0.035)
Intercept	10.97 (0.069)	6.60 (0.39)	6.63 (0.53)	7.02 (7.50)	6.60 (0.40)
Summary Statistics					
SER	0.102	0.098	0.099	0.099	0.099
$\bar{R}^2$	0.72	0.74	0.73	0.73	0.73

Variable definitions: *Price* = sale price ($); *Size* = house size (in square feet); *Bedrooms* = number of bedrooms; *Pool* = binary variable (1 if house has a swimming pool, 0 otherwise); *View* = binary variable (1 if house has a nice view, 0 otherwise); *Condition* = binary variable (1 if real estate agent reports house is in excellent condition, 0 otherwise).

1. Using the results in column (1), what is the expected change in price of building a 500-square-foot addition to a house? Construct a 95% confidence interval for the percentage change in price.

2. Comparing columns (1) and (2), is it better to use *Size* or $ln(Size)$ to explain house prices?

3. The regression in column (3) adds the number of bedrooms to the regression. How large is the estimated effect of an additional bedroom? Is the effect statistically significant? Why do you think the estimated effect is so small? (Hint: Which other variables are being held constant?)

4. Is the quadratic term $ln(Size)^2$ important?

.

5. Use the regression in column (5) to compute the expected change in price when a pool is added to a house that doesn't have a view. Repeat the exercise for a house that has a view. Is there a large difference? Is the difference statistically significant?

.

3 Appendix

TABLE 3 Critical Values for the χ^2 Distribution

	Significance Level		
Degrees of Freedom	**10%**	**5%**	**1%**
1	2.71	3.84	6.63
2	4.61	5.99	9.21
3	6.25	7.81	11.34
4	7.78	9.49	13.28
5	9.24	11.07	15.09
6	10.64	12.59	16.81
7	12.02	14.07	18.48
8	13.36	15.51	20.09
9	14.68	16.92	21.67
10	15.99	18.31	23.21
11	17.28	19.68	24.72
12	18.55	21.03	26.22
13	19.81	22.36	27.69
14	21.06	23.68	29.14
15	22.31	25.00	30.58
16	23.54	26.30	32.00
17	24.77	27.59	33.41
18	25.99	28.87	34.81
19	27.20	30.14	36.19
20	28.41	31.41	37.57
21	29.62	32.67	38.93
22	30.81	33.92	40.29
23	32.01	35.17	41.64
24	33.20	36.41	42.98
25	34.38	37.65	44.31
26	35.56	38.89	45.64
27	36.74	40.11	46.96
28	37.92	41.34	48.28
29	39.09	42.56	49.59
30	40.26	43.77	50.89

This table contains the 90^{th}, 95^{th}, and 99^{th} percentiles of the χ^2 distribution. These serve as critical values for tests with significance levels of 10%, 5%, and 1%.