Quantitative Analysis and Empirical Methods

Multiple Regression

Jan Rovny

Sciences Po, Paris, CEE / LIEPP

Overview

- Control variables
- Comparing coefficients
- Testing hypotheses
- Model comparison

Introduction

- We now know how to model a relationship between two variables X and Y using simple regression
- In reality, of course, we believe that there are multiple predictors (or Xs) which cause Y
- We thus want to model Y as a function of *multiple* predictors
- Furthermore, we want to introduce certain *control variables* into our model
- To do this, we must run **multiple regression** analysis

Multiple Regression Model

- $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_K x_{iK} + \epsilon_i$
- R:
  ```
  model<-lm(depvar ~ indepvar + indepvar + indepvar...)
  ```
- β_K, where $K \geq 1$, is a partial regression coefficient, which gives the effect of the predictor x_K purged of all the influence of the other predictors.
- This amounts to a *ceteris paribus* clause (or holding all else constant)
- A unit increase in x_K thus leads to β_K change in y.

Control Variables

- Our theory generally implicates one (or a small number) of predictors as the key determinants of the dependent variable.
- We, however, often agree that there may be other causes of our dependent variable
- In such a scenario, we want to establish that our predictor of interest (X_1) causes Y independently of other predictors X_k.
- That is to say that our hypothesized X_1 has a substantively and statistically significant effect on Y even when other predictors are present.

Control Variables Eg.

- Example with religiosity and happiness.
- Surely, income is an important determinant of happiness too...
- What if income trumps religiosity in explaining happiness?

		Estimate	Std. Error	t value	Pr(>|t|)
Model 1	(Intercept)	6.8792	0.0813	84.65	0.0000
	Religiosity	0.1455	0.0463	3.15	0.0017
		R^2:	0.01254	Adj R^2:	0.01127

		Estimate	Std. Error	t value	Pr(>|t|)
Model 2	(Intercept)	6.1846	0.1771	34.91	0.0000
	Religiosity	0.1573	0.0458	3.43	0.0006
	Income	0.1629	0.0370	4.40	0.0000
		R^2:	0.03647	Adj R^2:	0.034

Clearly, both income and religiosity cause happiness

The Dilemma

Parsimony	Validity
Preference given to models with fewer parameters	Preference given to models that fit data better
To improve parsimony, drop predictors	To improve validity, add predictors

- Since R^2 is a non-decreasing function as we add additional predictors, irresponsible analysts may 'overfit' their models, atheoretically adding variables.
- To guard ourselves from non-parsimonious models, we prefer the adjusted R^2 (denoted as $\bar{R}^2$) as a measure of model fit, since it adjusts for the number of predictors included in the model.
- $\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k-1}$ Model summary in R reports $\bar{R}^2$

Comparing Regression Coefficients

- Since in most cases the scales of predictors are different, it is impossible to directly compare regression coefficients.
- In order to be able to compare the coefficients, they must be converted to comparable scales.
- The most common solution is to *standardize all predictors*
  - $x_{std} = \frac{x - mean_x}{sd_x}$
  - The problem is that we create variables with identical means and standard deviations, which makes other important comparisons impossible.
  - Easy solution in R:
    `library(lm.beta) [Return] lm.beta(modelname)`

Hypothesis Testing Procedure

1. State a null and alternative hypothesis: $H_0 : \mu = \mu_0$, $H_a : \mu \neq \mu_0$
2. Select a level of significance of interest: $\alpha = .05$ (we want to be 95% sure.)
3. Determine the sampling distribution of the test statistic. (If we are dealing with a means test and we know σ, we use the standard normal distribution and its Z statistic, if we are dealing with a means test and we don't know σ we use Student's t distribution and the T statistic.)
4. Calculate the test statistic (for t: $t = \frac{\hat{\beta}_k - b}{\hat{\sigma}_{\hat{\beta}_k}}$)
5. Find the critical value in the appropriate statistical table
6. Make a conclusion about the null hypothesis (reject or fail to reject)

T-test

- The T-test provides a method for testing whether our β coefficients are equal to a certain number.
- The most common **null hypothesis** is that $\beta_x = 0$ (x has no effect on y).
- Every null hypothesis indicates an **alternative hypothesis** or H_a, which is the logical opposite of H_0. In our example $H_a : \beta_x \neq 0$ (x has an effect on y).
- To test our H_0 we solve: $t = \frac{\hat{\beta}_x - 0}{\hat{\sigma}_{\hat{\beta}_x}}$.
- The resulting t-value is then applied to the t-table with the relevant degrees of freedom.
- To avoid going to the tables, we can use the short-hand "2-t rule of thumb:" (assuming sufficient d.f.)

T-test Example

	Coef.	Std. Err.	t	$P > \|t\|$	[95% Conf. Interval]
x1	.9302514	.2155127	4.32	0.000	.5068189, 1.353684
x2	-1.146999	.2490548	-4.61	0.000	-1.636334, -.6576643
x3	.6640873	.3977576	1.67	0.096	-.1174141, 1.445589
x4	.495263	.4368451	1.13	0.257	-.3630363, 1.353562
cons	.3553831	.1233497	2.88	0.004	.1130296 .5977366
N=500	F(4, 495)=30.33	$P > F$=.000	$R^2 = 0.1969$	AdR^2=0.1904	Root MSE=2.7512

Test the possibility that the true value of $\beta_{x1} = 0.7$

Is $\hat{\beta}_{x1}$ significantly different from 0.7?

$t = \frac{\hat{\beta}_k - b}{\hat{\sigma}_{\hat{\beta}_k}} = \frac{.930 - .7}{.216} = \frac{.23}{.216} = 1.06$

Check this t-value in the t-table with 495 degrees of freedom at the 0.05 confidence level.

The critical value here is 1.96, which is greater than 1.06

We thus fail to reject $H_0 : \beta_{x1} = 0.7$.

$\hat{\beta}_{x1} = .93$ is not significantly different from 0.7.

Two-Tailed v. One-Tailed Tests

- Until now, we have been doing our t-tests as if we had no expectation about the direction of the relationship between x and y.
- As a result, when we were testing whether our β_x is significantly different from 0, we looked at both ends (or tails) of the distribution of β_x. This was a **two-tailed test**.
- In reality, we often have theoretical expectations about the direction of the relationship between x and y (increased income causes increased happiness etc.)
- Consequently, when testing whether β_x is significantly different from 0, we would expect 0 to be on one particular side of the distribution of β_x. Here we can do a **one-tailed test**.
- In practice, this means that we can divide our p-values by two.

F-test

- In the most classical F-test we assess the null hypothesis that all our predictors are equal to zero:
- $H_0 : \beta_1 = \beta_2 = ... = \beta_k = 0$,
- This tests the overall significance of our regression model. R automatically reports this test.
- The logic of the F-test is that if all predictors are 0, then the sole source of variation in y is the error term e.
- This effectively means that we need to check the relative size of the Regression Sum of Squares (RSS) to Error Sum of Squares (ESS), while taking into account the relevant df.
- The test thus is: $F = \frac{RSS/df}{ESS/df} = \frac{RSS/(k)}{ESS/(n-k-1)} = \frac{R^2/k}{(1-R^2)/(n-k-1)}$
- Can test other restrictions...

Combined T-test

- Testing the null hypothesis that: $H_0 : \beta_1 = \beta_2$ or equivalently $\beta_1 - \beta_2 = 0$, we use the Combined t-test:
- $Combined - T = \frac{\hat{\beta}_1 - \hat{\beta}_2}{\sqrt{\sigma^2_{\hat{\beta}_1} + \sigma^2_{\hat{\beta}_2} - 2\sigma^2_{\hat{\beta}_1, \hat{\beta}_2}}}$
- Here we need to know the variance and covariance of our estimators.
- We can get this information from the so-called **variance-covariance matrix of estimators (VCE)**
- In **R**: 'vcov(model)'

Model Comparison

When we compare different regression models we differentiate between Nested and Non-Nested models:

1. **Nested Models**
   - A nested model M2 is nested in another model M1, if M2 is a special case of M1:
   - $M2: y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \epsilon_i$ and
   - $M1: y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 w_i + \beta_4 r_i + \epsilon_i$
   - Here it is clear that M2 is a special case of M1, since M1 arises when $\beta_3 = \beta_4 = 0$ Thus M2 is nested in M1.
2. **Non-Nested Models**
   - $M2: y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 z_i + \epsilon_i$
   - $M1: y_i = \gamma_0 + \gamma_1 q_{i1} + \gamma_2 r_i + \epsilon_i$
   - Here y_i is seen as a function of different predictors. These models are thus non-nested.

Comparing Nested Models

- *Restricted* : $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \epsilon_i$
- *Full* : $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 w_i + \beta_4 r_i + \epsilon_i$
  - We want to know whether adding w and r improves our model significantly
- To test this, we perform the Joint F-test:
  - $F - test = \frac{RSS_R - RSS_{UR}/m}{RSS_{UR}/n-k-1} \sim F_{m,n-k-1}$
  - Here m is the number of restrictions (number of excluded variables - in our case 2)
  - k is the number of predictors in the full model.
  - We apply the result to the F distribution where the 1st d.f. is m and the second d.f. is $n-k-1$.
  - If the associated p-value is smaller than our selected significance level, we conclude that the full model is a significant improvement.

Comparing Non-Nested Models

- Since models are not linear restrictions of each other, we cannot use the F-test
- We must rely on another approach:

Akaike's Information Criterion (AIC):

- $AIC_i = -2l_i + 2K_i$,

Baysian Information Criterion (BIC):

- $BIC_i = -2l_i + K_i ln(n)$ (penalizes more for predictor inclusion)
- where i denotes a particular model,
- l_i is the log-likelihood function of a particular model
- K_i is the number of estimated parameters of a given model.
- AIC / BIC is a combined measure of *fit* and *parsimony*.
- The model with the **smallest** AIC / BIC is preferred.
- R: `AIC(modelname)` or `BIC(modelname)`