CFAspace

Provided by APF

Academy of Professional Finance 专业金融学院

CFA Level II

Multiple Regression Analysis

Part I

CFA Lecturer: Jiahao Gu

Content

Multiple Regression

Coefficients

(Los a, b, c, d, e)

Interpret the Multiple Regression Results

T-test for the regression coefficients

Hypothesis Test of Regression Coefficients

Other Hypothesis Test

Confidence interval and predicted value

Multiple Regression Coefficients

Multiple Regression Model

Multiple Linear Regression Model:

$$Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \cdots b_k X_{ki} + e_i$$

b_0: regression intercept term
b_i : regression slope coefficient
e_i : residual for the ith observation
k : number of independent variables

Multiple Regression Equation:

Minimize $\sum_{i=1}^{n} e_i^2$

$$\hat{Y}_i = \hat{b}_0 + \hat{b}_1 X_{1i} + \hat{b}_2 X_{2i} + \cdots \hat{b}_k X_{ki}$$

No.	Age(X_1)	Working hour(X_2)	Salary(Y)
1	42	232	356
2	23	217	70
3	37	248	418
4	24	220	12
5	29	225	22
6	51	205	455
7	38	237	339
8	27	229	252
9	41	260	278
10	39	201	233

	Coefficient	*Standard Error*	*t Stat*	*p-level*
Intercept	-665.71733	407.95987	-1.63182	0.14674
Age(X1)	14.68151	3.46332	4.23914	0.00384
Working hour(X2)	1.73217	1.72266	1.00552	0.34813

Salary(Y) = - 665.72 + 14.68 * Age(X_1) + 1.73 * Working hour(X_2)

Multiple Regression Coefficients

Interpret Regression Results

	Coefficient	*Standard Error*	*t Stat*	*p-level*
Intercept	-665.71733	407.95987	-1.63182	0.14674
Age(X1)	14.68151	3.46332	4.23914	0.00384
Working hour(X2)	1.73217	1.72266	1.00552	0.34813

Salary(Y) = - 665.72 + 14.68 * Age(X_1) + 1.73 * Working hour(X_2)

-- The intercept term is -665.72, and this is the value of Y when all independent variables are equal to zero.

-- The slope coefficient of X_1 is 14.68 and the slope coefficient of X_2 is 1.73. The slope coefficient of each independent variable is the estimated change of Y given one-unit change in that independent variable, holding the other independent variables constant.

-- The slope coefficient of X_1 changes when we add X_2 to the model. This means X_1 is correlated with X_2 to some extent. Because when X_1 changes, X_2 will also change. If X_1 and X_2 are uncorrelated, the slope coefficient of X_1 will not change when we add more independent variables.

Multiple Regression Coefficients

Hypothesis Test of Regression Coefficients

	Coefficient	Standard Error	t Stat	p-level
Intercept	-665.71733	407.95987	-1.63182	0.14674
Age(X1)	14.68151	3.46332	4.23914	0.00384
Working hour(X2)	1.73217	1.72266	1.00552	0.34813

$$t = \frac{\hat{b}_j - b_j}{s_{\hat{b}_j}}$$

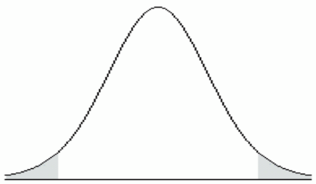
The t-statistic has n – k - 1 degrees of freedom.

To test statistical significance, we set $b_j = 0$. So H_0: $b_j = 0$; H_a: $b_j \neq 0$

The t-statistic in the table above is computed when $b_j = 0$.

The p-value is the smallest level of significance for which the null hypothesis can be rejected. An alternative method of doing hypothesis testing of the coefficients is to compare the p-value to the significance level:

- If the p-value is less than significance level, the null hypothesis can be rejected.
- If the p-value is greater than the significance level, the null hypothesis cannot be rejected.
- p-value in the table is applied to a two-tail test.



Multiple Regression Coefficients

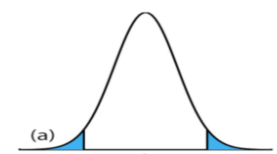
Other Hypothesis Test

	Coefficient	*Standard Error*	*t Stat*	*p-level*
Intercept	-665.71733	407.95987	-1.63182	0.14674
Age(X1)	14.68151	3.46332	4.23914	0.00384
Working hour(X2)	1.73217	1.72266	1.00552	0.34813

We can also formulate one- and two-tailed tests in which the null hypothesis is that the coefficient is equal to some value other than zero, or that it is greater than or less than some value.

- If the null hypothesis is that the coefficient is equal to some value other than zero, we still use a two-tail test. ($\alpha = 0.025$)
- If the null hypothesis is that the coefficient is greater than or less than some value, we will use a one-tail test. ($\alpha = 0.05$)

Example: (1) Test the null hypothesis that the slope coefficient of Age is equal to 10 using a 5% significance level. (2) Test the null hypothesis that the slope coefficient of Age is greater than 20 using a 5% significance level.



(1) $t = (14.68 – 10) / 3.463 = 1.35$ $t_c = 2.365$ ($df = 7, \alpha = 0.025$)

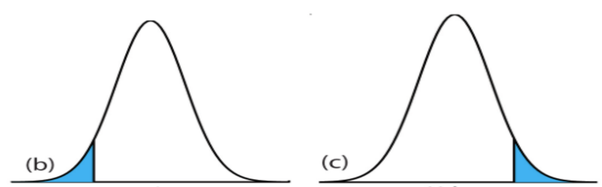
Because $t < t_c$, we cannot reject H_0

(2) H_0: $b_{age} \geq 20$; H_a: $b_{age} < 20$

$t = (14.68 – 20) / 3.463 = -1.54$

$t_c = 1.895$ ($df = 7, \alpha = 0.05$)

Because $t > -t_c$, cannot reject H_0



Multiple Regression Coefficients

Confidence interval and predicted value

The confidence interval of b_k is $\hat{b}_k - t_c \times s_{\hat{b}_k} < b_k < \hat{b}_k + t_c \times s_{\hat{b}_k}$

t_c is a two-tailed value with n - k - 1 degrees of freedom.

	Coefficient	*Standard Error*	*t Stat*	*p-level*
Intercept	-665.71733	407.95987	-1.63182	0.14674
Age(X1)	14.68151	3.46332	4.23914	0.00384
Working hour(X2)	1.73217	1.72266	1.00552	0.34813

t_c (df = 7, α = 0.025) is 2.365, so the confidence interval of b_{age} is

14.68 – 2.365 × 3.463 < b_{age} < 14.68 + 2.365 × 3.463, 6.49 ~ 22.87

From the confidence interval, we can also see the bage is significantly different from zero at the 5% level.

-- Predict the value of dependent variable

Salary(Y) = - 665.72 + 14.68 * Age(X_1) + 1.73 * Working hour(X_2)

If X_1 = 30 and X_2 = 240, we can predict the salary is

-665.72 + 14.68 × 30 + 1.73 × 240 = 189.88

CFAspace

Thank You!