Inferences on Linear Combinations of Coefficients

R Tutorials for Applied Statistics

Note on required packages: The following code required the package `multcomp` to test hypotheses on linear combinations of regression coefficients. If you have not already done so, download and install the libraries (needed only once per computer), and load the libraries (need to do every time you start R) with the following code:

```
# This only needs to be executed once for your machine
install.packages("multcomp")

# This needs to be executed every time you load R
library("multcomp")
```

---

1 Motivation

With interactions effects, it becomes interesting to construct hypothesis tests and confidence intervals on linear combinations of regression coefficients. Consider the following estimated regression:

$$y_i = b_0 + b_1 x_{1,i} + b_2 x_{2,i} + b_3 x_{1,i} x_{2,i} + e_i$$

yi=b0+b1x1,i+b2x2,i+b3x1,ix2,i+ei

The estimated *marginal effect* of x_1x1 on y_1y1 is not only the coefficient on x_1x1, but is also a function of the coefficient on the interaction term and the value of the variable it is interacted with (x_2x2),

$$me_1 = \frac{\Delta \hat{y}}{\Delta x_1} = b_1 + b_3 x_{2,i}$$

me1=Δy^Δx1=b1+b3x2,i

The *average marginal effect* is equal to the marginal effect given above, where $x_{2,i}$x2,i is set equal to its mean, $\bar{x}_2$x¯2:

$$ame_1 = \frac{\Delta\hat{y}}{\Delta x_1} = b_1 + b_3\bar{x}_2$$

ame1=Δy^Δx1=b1+b3x‾2

The hypothesis test on b_1b1 which comes out of typical summary statistics from a regression run is not relevant in terms of inferring what the impact of x_1x1 is on y_1y1. We need a hypothesis test for the *linear combination* of $b_1 + b_3\bar{x}_2$b1+b3x‾2

2 Example: Factors Affecting Monthly Earnings

Let us examine a data set that explores the relationship between total monthly earnings (`MonthlyEarnings`) and a number of factors that may influence monthly earnings including including each person's IQ (`IQ`), a measure of knowledge of their job (`Knowledge`), years of education (`YearsEdu`), years experience (`YearsExperience`), and years at current job (`Tenure`).

The code below downloads data on the above variables from 1980 for 663 individuals and assigns it to a dataframe called `df` .

```
load(url("http://murraylax.org/datasets/wage2.RData"))
```

The following call to `lm()` estimates a multiple regression predicting monthly earnings based on the five explanatory variables given above and includes an interaction term between education and experience. Summary regression statistics are displayed with the call to `summary()`

```
lmwages <- lm(MonthlyEarnings
              ~ IQ + Knowledge + YearsEdu + YearsExperience + Tenure
              + YearsEdu*YearsExperience,
              data = df)
summary(lmwages)
```

```
##
## Call:
## lm(formula = MonthlyEarnings ~ IQ + Knowledge + YearsEdu + YearsExperience +
##     Tenure + YearsEdu * YearsExperience, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -863.30 -241.41  -41.33  190.62 2213.41
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -1.020    284.112  -0.004 0.997135
## IQ                          3.800      1.202   3.162 0.001636 **
## Knowledge                   7.859      2.243   3.504 0.000489 ***
## YearsEdu                    9.403     19.937   0.472 0.637325
## YearsExperience           -33.477     23.097  -1.449 0.147699
## Tenure                      4.235      3.037   1.395 0.163607
## YearsEdu:YearsExperience    3.547      1.719   2.064 0.039450 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 368.7 on 656 degrees of freedom
## Multiple R-squared:  0.1849, Adjusted R-squared:  0.1774
## F-statistic:  24.8 on 6 and 656 DF,  p-value: < 2.2e-16
```

Suppose that we want to make inferences about the effect that experience has on monthly earnings. The coefficient on `YearsExperience` is negative, but not statistically significant. The coefficient on the interaction term of `YearsExperience` with `YearsEdu` is positive, and statistically significant. Alone, these statistics say almost nothing of the average effect experience has on monthly earnings. Is the average effect positive or negative? Is the average effect statistically significant?

The **average marginal effect** of experience on monthly earnings is given by:

$$ame_{exp} = b_{exp} + b_{exp*edu}\bar{x}_{edu}$$

ameexp=bexp+bexp∗edux¯edu

The expression for ame_{exp} ameexp is a **linear combination** of coefficients b_{exp} bexp and $b_{exp*edu}$ bexp∗edu.

3 Specifying linear combination expressions in R

We can construct this linear expression in R by creating a matrix with one row and number of columns equal to the number of coefficients we have.

```
coefeq <- matrix(data=0, nrow=1, ncol=length(lmwages$coefficients))
```

This call to `matrix()` created a matrix that we named `coefeq` with 1 row and 7 columns and filled it in with all 0s. The call within the function `ncol=length(lmwages$coefficients)` set the parameter for the number of columns equal to the number of coefficients in the `lmwages` object.

We can look at the matrix we created by typing its name at the command prompt,

```
coefeq
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7]
## [1,]    0    0    0    0    0    0    0
```

Next, we need to give column names in the matrix that are exactly equal to the coefficient names in `lmwages`. We set the column names of a matrix with a call to `colnames()`. We get the names of the coefficients in lmwages with the function `names()`.

```
colnames(coefeq) <- names(lmwages$coefficients)
```

Let us again look at the matrix with our coefficient names labeling each column:

```
coefeq
```

```
##      (Intercept) IQ Knowledge YearsEdu YearsExperience Tenure
## [1,]           0  0         0        0               0      0
##      YearsEdu:YearsExperience
## [1,]                        0
```

Now we want to set the values for some of the matrix elements equal to the numbers multiplying the regression coefficients in our linear combination. Looking at the equation for the average marginal effect of experience above, we can see the number multiplying b_{exp} bexp is equal to 1 and the number multiplying $b_{exp*edu}$ bexp*edu is $\bar{x}_{edu}$ x̄edu. The code below sets these values and again displays the matrix.

```
coefeq[1,"YearsExperience"] <- 1
coefeq[1,"YearsEdu:YearsExperience"] <- mean(df$YearsEdu, na.rm=TRUE)
coefeq
```

```
##      (Intercept) IQ Knowledge YearsEdu YearsExperience Tenure
## [1,]          0  0         0        0               1      0
##      YearsEdu:YearsExperience
## [1,]                 13.68024
```

Inside the square brackets that follow `coefeq` is the row number and column number (separated by a comma) for the element of the matrix to set. Since there is only one row, both lines above specify row 1. The first line specifies the column associated with "YearsExperience" coefficient. The second line specifies the interaction coefficient with "YearsEdu:YearsExperience". The first line sets the number multiplying the years experience coefficient equal to 1. The second line sets the number multiplying the interaction coefficient equal to the mean of the years of education variable.

The matrix `coefeq` multiplied by vector of coefficients is equal to the average marginal effect of experience on monthly earnings. We can compute this with the code,

```
coefeq %*% lmwages$coefficients
```

```
##            [,1]
## [1,] 15.05299
```

The operator `%*%` is the multiplication operator for matrices. We can see that the estimated average marginal effect of experience on monthly earnings is equal to $15.05.

4 Hypothesis Testing for Linear Combination

Let us test the hypothesis that experience has a positive marginal effect on monthly earnings at the sample mean for education. The null and alternative hypotheses are given by,

$$H_0 : \beta_{exp} + \beta_{exp*edu} \bar{x}_{edu} = 0$$
H0:βexp+βexp∗edux̄edu=0

$$H_1 : \beta_{exp} + \beta_{exp*edu} \bar{x}_{edu} > 0$$
H1:βexp+βexp∗edux̄edu>0

Note: seeing a sample estimate like $\bar{x}$x̄ in null and alternative hypotheses is *very unusual*. We are not conducting a hypothesis test on this mean. We are conditioning our hypothesis test for the marginal effect *conditioning on a value for education* that we happen to set equal to the mean in our sample.

The library `multmod` (we loaded this into memory at the beginning of the tutorial) includes a function called `glht()` which stands for *general linear hypothesis test*. We can use this function to compute hypothesis tests and confidence for any given linear combination of coefficients. The call for our linear combination is given by,

```
ametest <- glht(model=lmwages, linfct=coefeq, rhs=0, alternative="greater")
```

The parameter `model=lmwages` specifies the regression model for which we are conducting hypotheses, `linfct=coefeq` specifies the linear combination using our matrix that we constructed for the average marginal effect of experience, `rhs=0` specifies that the right-hand side of the null and alternative hypotheses are equal to 0, and `alternative="greater"` specifies that our alternative hypothesis has a greater-than sign, i.e. this is a one-sided test.

We can view summary statistics with a call to `summary()`:

```
summary(ametest)
```

```
##
##    Simultaneous Tests for General Linear Hypotheses
##
## Fit: lm(formula = MonthlyEarnings ~ IQ + Knowledge + YearsEdu + YearsExperience +
##      Tenure + YearsEdu * YearsExperience, data = df)
##
## Linear Hypotheses:
##          Estimate Std. Error t value   Pr(>t)
## 1 <= 0    15.053      4.128    3.647 0.000143 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Again we see that the average marginal effect is equal to $15.05. The one-sided p-value is equal to $0.000143$0.000143 which is less than 0.05. We reject the null hypothesis and conclude that there is sufficient statistical evidence that there is a positive marginal effect of experience on monthly earnings, *at the mean level of education*.