Erasmus School of Economics

MOOC Econometrics

Training Exercise S.2

Notes:

- This exercise uses the data file `TrainExerS1` and requires a computer.
- The data set `TrainExerS1` is available on the website.

Questions

1. You continue your investigation of the mean and standard deviations of returns in the stock market from Training Exercise S1. You want to determine how the sample size influences test statistics. You use the same sample of 1000 yearly return observations y_j in the data file `TrainExerS1`.

   (a) First you want to test hypotheses of the form $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$. Construct a series of statistics t_i and corresponding p-values for $\mu_0 = 0\%$ and $\mu_0 = 6\%$, where t_i is the t-statistic based on the first i observations. Use the range $i = 5, 6, \ldots, 30$. Make a table of t-statistics and p-values for both values of μ_0.

   (b) Make a graph with the two series of t-statistics for the range $i = 5, 6, \ldots, 200$. Also graph the series of critical values c_i that correspond with significance levels of 10, 5 and 1%. Note that the critical values depend on the number of observations.

   (c) What do you conclude regarding the hypothesis $H_0 : \mu = 0\%$? And regarding $H_0 : \mu = 6\%$. Based on the graphs, what would you expect for $H_0 : \mu = 6.5\%$. Explain why we never accept null hypotheses.

   (d) Next you analyse tests for hypotheses of the form $H_0 : \sigma^2 = \sigma_0^2$ versus $H_1 : \sigma^2 \neq \sigma_0^2$. Construct a series of statistics χ_i^2 and corresponding p-values for $\sigma_0 = 18\%$ and $\sigma_0 = 15\%$, where χ_i^2 is the χ^2-statistic based on the first i observations. Use the range $i = 5, 6, \ldots, 30$. Make a table of test-statistics and p-values for both values of σ_0.

   (e) Make a graph with the two series of χ_i^2-statistics for the range $i = 5, 6, \ldots, 200$. Also construct series with critical values c_i that correspond with significance levels of 10, 5 and 1%. Note that the critical values depends on the number of observations.

   (f) What do you conclude regarding the hypothesis for $\sigma_0 = 18\%$? And for $\sigma_0 = 15\%$

   (g) Take another look a both graphs. Argue why the significance level should be determined in relation to the sample size.

2. This question is a follow-up on Question 2 of Training Exercise S1. Consider a sample y of n observations of random variables y_i, $i = 1, 2, \ldots, n$. The sample consists of two groups, 1 and 2, with n_1 and n_2 observations per group. The variables are independent, and follow a normal distribution with group dependent mean and group independent variance,

$$y_i \sim \begin{cases} N(\mu_1, \sigma^2), & \text{if } y_i \text{ belongs to group 1} \\ N(\mu_2, \sigma^2), & \text{if } y_i \text{ belongs to group 2.} \end{cases}$$

The sample has been ordered such that the first n_1 observations belong to group 1, and the remaining n_2 observations belong to group 2. Derive a test for the hypothesis $H_0 : \mu_1 = \mu_2$ against the alternative $H_1 : \mu_1 \neq \mu_2$ by the following steps.

(a) Show that the null hypothesis is equivalent to $H_0 : h'\mu = 0$ against the alternative $H_0 : h'\mu \neq 0$ with $h' = (1, -1)$.

(b) In Training Exercise S1 we considered the estimator $m = T^{-1}H'y$ for the mean vector μ, with

$$H = \begin{pmatrix} \iota_{n_1} & 0_{n_1} \\ 0_{n_2} & \iota_{n_2} \end{pmatrix}, \quad T = \begin{pmatrix} n_1 & 0 \\ 0 & n_2 \end{pmatrix}, \quad H'H = T, \quad T^{-1} = \begin{pmatrix} \frac{1}{n_1} & 0 \\ 0 & \frac{1}{n_2}. \end{pmatrix}$$

Derive the distribution of m.

(c) Derive the distribution of $h'm$

(d) Show that $\dfrac{h'(m-\mu)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim N(0, 1)$.

(e) Derive the distribution of

$$t = \frac{h'(m-\mu)}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}},$$

where s^2 is the unbiased estimator for the variance of Training Exercise S1, $s^2 = \dfrac{1}{n-2}(y - Hm)'(y - Hm)$. You need that $(n-2)s^2/\sigma^2 \sim \chi^2(n-2)$ and m and s^2 are independent.