Lab 3 (Hypothesis testing)

This file uses the same data on bid-ask spreads (tradecost) and candidate determinants used in lab2 HW2.

. Import the data set and set up variable names

```
% import dataset lab2.xls (using Import Dataset)


% rename variables
tradecost = lab2{ : ,1};
logvolatility= lab2{ : ,2};
logsize = lab2{ : ,3};
logtrades = lab2{ : ,4};
logturn = lab2{ : ,5};
numberanalysts = lab2{ : ,6};
logtradecost = log(tradecost);
```

We run a regression of the log of the bid-ask spread on the 5 explanatory variables, reproduced below.

We interpreted the signs of the estimated coefficients of the 5 explanatory variables in light of the theories outlined in the background information of lab2 HW2..

```
x1 = table(logtradecost,logvolatility,logsize,logtrades,logturn,numberanalysts);
lm1 =fitlm(x1,'logtradecost ~ logvolatility + logsize + logtrades + logturn + numberanalysts')
```

```
lm1 =
Linear regression model:
    logtradecost ~ 1 + logvolatility + logsize + logtrades + logturn + numberanalysts

Estimated Coefficients:
                      Estimate        SE         tStat        pValue  
                      _________    _________    ________    __________

    (Intercept)        -0.82932      0.44277      -1.873      0.064173
    logvolatility         1.023     0.053245      19.213     2.574e-34
    logsize            -0.14029      0.02391     -5.8673    6.5918e-08
    logtrades          -0.16936     0.035596      -4.758    7.0564e-06
    logturn           -0.098714     0.032502     -3.0371     0.0030911
    numberanalysts    5.044e-05    0.0029135    0.017313       0.98622


Number of observations: 100, Error degrees of freedom: 94
Root Mean Squared Error: 0.142
R-squared: 0.882,  Adjusted R-Squared: 0.876
F-statistic vs. constant model: 141, p-value = 4.16e-42
```

coefTest

Linear hypothesis test on linear regression model coefficients

Description

`p = coefTest(mdl)` computes the *p*-value for an *F*-test that all coefficient estimates in `mdl`, except for the intercept term, are zero.

`p = coefTest(mdl,H)` performs an *F*-test that $H \times B = 0$, where *B* represents the coefficient vector. Use `H` to specify the coefficients to include in the *F*-test.

`p = coefTest(mdl,H,C)` performs an *F*-test that $H \times B = C$.

Note: in Lecture 4, H = R, and the vector of beta coefficient is B = Beta, and C is the r vector

`Examples:`

1. Test the null that all coefficients are 0.

```
[p1,F1] = coefTest(lm1)
```

```
p1 = 4.1630e-42
F1 = 141.1555
```

These match with the output

2. Test the assumption that the coefficients associated with Logsize and Logturn are equal to each other.

```
H = [0 0 1 0 -1 0];
%p= coefTest(lm1,H)

[p2,F2] = coefTest(lm1,H)
```

```
p2 = 0.2863
F2 = 1.1499
```

H is a 1x5 vector, B is the 5x1 vector of coefficients, threfore $H \times B = 0$ results in an equation.

$H \times B = 0$ means in this case: beta3 - beta5 = 0

3. Test the assumption that the coefficients associated associated with all variables except the constant are 0.

```
H = [0 1 1 1 1 1];
[p3,F3] = coefTest(lm1, H)
```

```
p3 = 1.2639e-17
F3 = 111.2644
```

4. Test the assumption that the coefficients associated associated with number analysts is 0.

```
H = [0 0 0 0 0 1];
[p4,F4] = coefTest(lm1, H)
```

```
p4 = 0.9862
F4 = 2.9973e-04
```

`mdl` — Linear regression model object

`LinearModel` object | `CompactLinearModel` object

Linear regression model object, specified as a `LinearModel` object created by using `fitlm` or `stepwiselm`, or a `CompactLinearModel` object created by using `compact`.

`H` — Hypothesis matrix

numeric index matrix

Hypothesis matrix, specified as an `r`-by-`s` numeric index matrix, where `r` is the number of coefficients to include in an *F*-test, and `s` is the total number of coefficients.

- If you specify `H`, then the output `p` is the *p*-value for an *F*-test that $H \times B = 0$, where *B* represents the coefficient vector.
- If you specify `H` and `C`, then the output `p` is the *p*-value for an *F*-test that $H \times B = C$.

Example: `[1 0 0 0 0]` tests the first coefficient among five coefficients

Data Types: `single` | `double`

`C` — Hypothesized value

numeric vector

Hypothesized value for testing the null hypothesis, specified as a numeric vector with the same number of rows as `H`.

If you specify `H` and `C`, then the output `p` is the *p*-value for an *F*-test that $H \times B = C$, where *B* represents the coefficient vector.

Data Types: `single` | `double`

Output Arguments

collapse all

p — *p*-value for *F*-test

numeric value in the range [0,1]

p-value for the *F*-test, returned as a numeric value in the range [0,1].

`F` — Value of test statistic for *F*-test

numeric value

Value of the test statistic for the *F*-test, returned as a numeric value.

`r` — Numerator degrees of freedom for *F*-test

positive integer

Numerator degrees of freedom for the *F*-test, returned as a positive integer. The *F*-statistic has `r` degrees of freedom in the numerator and `mdl.DFE` degrees of freedom in the denominator.

Algorithms

The *p*-value, *F*-statistic, and numerator degrees of freedom are valid under these assumptions:

- The data comes from a model represented by the formula in the `Formula` property of the fitted model.
- The observations are independent, conditional on the predictor values.

Under these assumptions hold, let β represent the (unknown) coefficient vector of the linear regression. Suppose H is a full-rank matrix of size r-by-s, where `r` is the number of coefficients to include in an *F*-test, and `s` is the total number of coefficients. Let c be a vector the same size as β. The following is a test statistic for the hypothesis that $H\beta = c$:

$F=(H\hat{}\beta-v)'(HCH')-1(H\hat{}\beta-v)$.

Here $\hat{}\beta$ is the estimate of the coefficient vector β, stored in the `Coefficients` property, and V is the estimated covariance of the coefficient estimates, stored in the`CoefficientCovariance` property. When the hypothesis is true, the test statistic F has an F Distribution with r and u degrees of freedom, where u is the degrees of freedom for error, stored in the `DFE` property.

3. Test the assumption that the coefficients associated associated with all variables except the constant are 0 (general formulation).

```
H = [0 0 0 0 0 0; 0 1 0 0 0 0;0 0 1 0 0 0;0 0 0 1 0 0;0 0 0 0 1 0;0 0 0 0 0 1 ]
```

```
H = 6×6
     0     0     0     0     0     0
     0     1     0     0     0     0
     0     0     1     0     0     0
     0     0     0     1     0     0
     0     0     0     0     1     0
     0     0     0     0     0     1
```

```
C = [0; 0; 0; 0; 0; 0]
```

```
C = 6×1
```

```
     0
     0
     0
     0
     0
     0
```

```
[p5,F5] = coefTest(lm1, H, C)
```

```
p5 = 4.1630e-42
F5 = 141.1555
```

5. Test the assumption that the coefficient associated with log size equals -0.5.

```
H = [0 0 0 0 0 0; 0 0 0 0 0 0;0 0 1 0 0 0;0 0 0 0 0 0;0 0 0 0 0 0;0 0 0 0 0 0 ]
```

```
H = 6×6
     0     0     0     0     0     0
     0     0     0     0     0     0
     0     0     1     0     0     0
     0     0     0     0     0     0
     0     0     0     0     0     0
     0     0     0     0     0     0
```

```
C = [0; 0; -0.5; 0; 0; 0]
```

```
C = 6×1
         0
         0
   -0.5000
         0
         0
         0
```

```
[p6,F6] = coefTest(lm1,H,C)
```

```
p6 = 9.1569e-27
F6 = 226.3294
```