General notes

Midterm will contain 6 problems. All the problems have equal weights. Midterm duration is 120 minutes. No cheatsheet, closed book. All topics before (and including) bootstrap may be included.

Demo variant «Pumpkin»

1. You have estimated the linear regression

$$\hat{y}_i = 3 + 0.2x_i + 0.5b_i - 0.7c_i$$

using big number of observations n. Assume all Gauss — Markov assumptions are satisfied.

The estimate of the covariance matrix of coefficients is

$$\begin{pmatrix} 2.5 & -0.2 & 0.1 & 0 \\ ? & 4.5 & -0.1 & 0.1 \\ ? & ? & 3.2 & -0.1 \\ ? & ? & ? & 3.5 \end{pmatrix}$$

(a) Construct 95% confidence interval for β_x.

(b) Construct 95% confidence interval for conditional expected value of dependent variable if $x = 1$, $b = -1$, $c = 2$.

(c) Test that $\beta_x + \beta_b = 1$ against $\beta_x + \beta_b \neq 1$ at 5% significance level.

2. Consider the simple regression model $y_i = \beta_0 + \beta_1 x_i + u_i$. Assume that Gauss — Markov assumptions are satisfied. Let $z_i = \sqrt{x_i}$.

Consider two estimators of β_1:

$$\hat{\beta}_1' = \frac{\sum_{i=1}^n (z_i - \bar{z}) y_i}{\sum_{i=1}^n (z_i - \bar{z}) x_i}, \quad \hat{\beta}_1'' = \frac{\sum_{i=1}^n (z_i - \bar{z}) y_i}{\sum_{i=1}^n (z_i - \bar{z}) z_i}.$$

(a) Calculate $\mathbb{E}(\hat{\beta}_1' \mid x)$ and $\mathbb{E}(\hat{\beta}_1'' \mid x)$. Are estimators unbiased conditionally on x?

(b) Are estimators $\hat{\beta}_1'$ and $\hat{\beta}_1''$ consistent?

3. You have 300 observations in total and you estimated the regression

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{\beta}_2 v_i + \hat{\beta}_3 w_i$$

using different parts of your dataset.

Observations	Result
$1, \ldots, 300$	$SS^{\text{res}} = 300$, $SST = 500$
$1, \ldots, 200$	$SS^{\text{res}} = 150$
$201, \ldots, 300$	$SS^{\text{res}} = 100$

Assume for each case that Gauss — Markov assumptions for the unrestricted model are satisfied. Use 5% critical level.

(a) Test that $\beta_1 = \beta_2 = \beta_3 = 0$ for the whole dataset against alternative that at least one of the coefficients is non-zero.

(b) Test for the absence of a structural break between part one (observations 1, ..., 200) and part two (observations 201, ..., 300) of your dataset. Here in H_0 you assume that the relation $y_i = \beta_0 + \beta_1 x_i + \beta_2 v_i + \beta_3 w_i + u_i$ holds for the whole dataset. And in H_1 you assume that different linear models may be valid for the two parts of your dataset.

(c) Consider again H_0 that the relation $y_i = \beta_0 + \beta_1 x_i + \beta_2 v_i + \beta_3 w_i + u_i$ holds for the whole dataset. But now in H_1 assume that the linear relation $y_i = \beta_0 + \beta_1 x_i + \beta_2 v_i + \beta_3 w_i + u_i$ holds only for the first part of the dataset. You do not assume linear relation between y_i and predictors in the second part. Test H_0 against H_1.

4. Consider three observations: $x = (1, 1, 0)$ and $y = (1, 3, 5)$.

   (a) Estimate $\hat{\beta}_0$ and $\hat{\beta}_1$ in the regression $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ using OLS.

   (b) Calculate sum of squared residuals, sum of squared explained, total sum squares and R^2.

   (c) Under Gauss — Markov assumptions estimate the variance of random error term.

   (d) Under Gauss — Markov assumptions estimate the variance matrix of $\hat{\beta}$ vector.

5. Consider the model $y = X\beta + u$ where β is non-random, $\mathbb{E}(u \mid X) = 0$, the matrix X of size $n \times k$ has rank $X = k$, but $\mathbb{V}\mathrm{ar}(u \mid X) = \sigma^2 W$ with $W \neq I$. Let $\hat{\beta}$ be the standard OLS estimator of β.

   (a) Find $\mathbb{E}(\hat{\beta} \mid X)$, $\mathbb{E}(\hat{\beta})$.

   (b) Find $\mathbb{V}\mathrm{ar}(\hat{\beta} \mid X)$.

   (c) How do you think, will the standard confidence interval for β be valid in this case?

   (d) Find $\mathbb{C}\mathrm{ov}(y, \hat{\beta} \mid X)$.

6. Consider two multivariate regressions estimated by OLS. Regression A:

$$\hat{y} = X_1 \hat{\beta}_1 + X_2 \hat{\beta}_2.$$

   The matrices X_1 and X_2 are two blocks of regressors, the matrix X_1 has size $n \times k_1$, the matrix X_2 has size $n \times k_2$. Let $X = \begin{pmatrix} X_1 X_2 \end{pmatrix}$ be the full regressors matrix.

   Consider projection matrix H_1 that projects vectors on the span of columns of X_1. Let $M_1 = I - H_1$ where I is an identity matrix, $y' = M_1 y$, $X_2' = M_1 X_2$. Regression B:

$$\hat{y}' = X_2' \hat{\gamma}.$$

   (a) Are estimators $\hat{\beta}_2$ and $\hat{\gamma}$ equal?

   (b) Are residuals in both regressions equal?

Demo variane «Bat»

1. Foma estimated simple regression model using three observations and calculated the forecasts $\hat{y}_i$. Unfortunately he remembers only this table:

y_i	$\hat{y}_i$
0	1
6	?
6	?

He recalls that $\hat{y}_3 > \hat{y}_2$.

Help Foma to reconstruct the missing entries in the table.

2. Consider the partially hidden output from the Stata package:

```
      Source |Sum of Squares   Degrees of Freedom          Number of obs =      229
-------------+-----------------------------------          F(   4,    ???D) =    ???E
   Explained |   13.0563901         ???A                   Prob > F      =   0.0000
    Residual |   ???B               224                    R-squared         =   ???F
-------------+-----------------------------------          Adj R-squared     =   ???G
       Total |   57.617198          ???C                   Standard error    =   ???H

-------------------------------------------------------------------------------
  lnearnings |       Coef.   Std. Err.        t    P>|t|     [95% Conf. Interval]
-------------+-----------------------------------------------------------------
         exp |   ???I        .0166437      3.55   0.000     .0262485     .0918449
           s |   .0962928    .0132887      7.25   ???J      .0701059     .1224797
      female |  -.2069607    .1627401      ???K   NA        ???L          ???M
  exp*female |  -.0086271    ???N          ???O   0.687     NA            NA
       const |   ???P          NA           NA    NA        .6054517      1.67635
```

That's the standard regression table reported by all software packages. Here the letters A, B, ... just play the role of identifiers.

Reconstruct all the yellow entries with question marks.

3. Scared of proofs? That's Halloween, you should be scared!

   (a) Prove that in the simple regression $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ the standard t-statistic t for the test H_0: $\beta_1 = 0$ and the standard F-statistic F for the same test satisfy the relation $t^2 = F$.

   (b) Prove that in the simple regressions $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ and $\hat{x}_i = \hat{\gamma}_0 + \hat{\gamma}_1 y_i$ the determination coefficients are equal.

4. Consider the model $y = X\beta + u$ on the training set where β is non-random, $\mathbb{E}(u \mid X) = 0$, the matrix X of size $n \times k$ has rank $X = k$, but $\mathbb{V}\mathrm{ar}(u \mid X) = \sigma^2 I$, where I is an identity matrix. Observations are independent. Let $\hat{\beta}$ be the standard OLS estimator of β using the training set data. You have a test set with the same type of dependency $y' = X'\beta + u'$, $\mathbb{E}(u' \mid X') = 0$, $\mathbb{V}\mathrm{ar}(u' \mid X') = \sigma^2 I$. You make the forecast $\hat{y}'$ for the test set, $\hat{y}' = X'\hat{\beta}$.

   (a) Find $\mathbb{E}(\hat{y}' \mid X, X')$.

   (b) Find $\mathbb{V}\mathrm{ar}(y' - \hat{y}' \mid X, X')$.

   (c) Find $\mathbb{C}\mathrm{ov}(\hat{\beta}, y' - \hat{y}' \mid X, X')$.

5. Let $u = (u_1, u_2, u_3)$ be the vector of iid standard normal random variables. Denote the linear span of the vector $(1, 1, 0)$ be V. Let $W = V^{\perp}$ be the orthogonal complement of V in $\mathbb{R}^3$.

   (a) Find $\dim V$, $\dim W$.

   (b) Find the projection matrix H that projects vectors from $\mathbb{R}^3$ onto V.

   Let $v = Hu$ and $w = (I - H)u$.

   (c) Find the law of distribution of the random variables $v^T v$ and $\|w\|^2$.

   (d) Find the distribution of $2v^T v/(w^T w)$.

6. Assume that the true model is $y_i = \beta_0 + u_i$ and u_i are iid normally distributed with zero mean and variance σ^2. There are n observations. We estimate regression $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ using OLS.

   (a) Find $\mathbb{E}(\hat{\beta}_1 \mid x)$ and $\mathbb{E}(\hat{\beta}_1)$.

   (b) Find $\mathbb{E}(R^2)$.

   (c) Find $\mathbb{E}(R^2_{adj})$ for the adjusted coefficient

$$R^2_{adj} = 1 - \frac{SS^{\text{res}}/(n-2)}{SST/(n-1)}.$$