STOCKHOLMS UNIVERSITET
Department of Statistics
Raul Cano

VT12
(updated 2012-05-03)

Course description for Statistical Methods (Statistiska metoder), Block II, 7,5 ECTS.

Preliminary plan for teaching

Part I. Bayesian Inference.

F1	Bayesian methods (Daniel Thorburn)
F2	Bayesian methods (Daniel Thorburn)
F3	Bayesian methods (Daniel Thorburn)
F4	Bayesian methods (Daniel Thorburn)

Home assignment for part I. Deadline for the assignment (part I): Thursday 3 may during F8.
(Part I is worth 42 credit points)

1. Read the article Inference for a Bernoulli process (given already by me) and summarize it in your own words in 2 pages. (6 points)

2. Suppose that $X_1, X_2, ..., X_n$ form a random sample from a Bernoulli distribution for which the value of the parameter θ is unknown. Use a prior (conjugate) distribution for θ and find the posterior distribution of θ. (6 points)

3. Suppose that $X_1, X_2, ..., X_n$ form a random sample from an exponential distribution for which the value of the parameter θ is unknown. Use a prior (conjugate) distribution for θ and find the posterior distribution of θ. (6 points)

4. Suppose that the number of defects in a 1200-foot roll of magnetic recording tape has a Poisson distribution for which the value of the mean θ is unknown. Suppose also that the prior distribution of θ is a gamma distribution with parameters $\alpha = 3$ and $\beta = 1$. Suppose finally that when 5 rolls of this tape are selected at random and inspected, the number of defects found on the rolls are 2, 2, 6, 0 and 3. Determine the posterior distribution of θ. (6 points)

5. Suppose that the heights of the individuals in a certain population have a normal distribution for which the value of the mean θ is unknown and the standard deviation is 2 inches. Suppose also that the prior distribution of θ is a normal distribution for which the mean is 68 inches and the standard deviation is 1 inch. Suppose finally that 10 people are selected at random from the population and their average height is found to be 69.5 inches. Determine the posterior distribution of θ. (6 points)
Hint: See the Technical appendix with the details for the posterior calculations of the Normal model in the old teaching material from spring 2011.

6. Read the article Bayes estimates for the linear model (given already by me) and summarize the article in your own words in 3 pages (6 points) and summarize the discussion on the article in your own words in 3 pages (6 points).

Part II. Nonparametric regression and related models.

LITERATURE
The Elements of Statistical Learning by Hastie, Tibshirani and Friedman, 2nd edition, Springer.
(HTF for short) (Se the old teaching material from spring 2011)
Read: HTF 6.1. HTC 9.1. HTF 5.2 (especially 5.2.1) and HTF 3.4.

F5 Nonparametric regression (Saeid Amiri),
(the rest of the home assignment for part I was given at F5. Important: Obligatory attendance at F5)
F6 Nonparametric regression (Saeid Amiri)
F7 Nonparametric regression (Saeid Amiri)

Home assignment for part II. Deadline for the assignment (part II): Thursday 24 may during F10.
(Part II is worth 30 credit points)

1. Consider the cps71 data set that includes a random sample taken from 1971 Canadian Census Public Use Tapes for male individuals having common education (grade 13).

```
library(np)
attach(cps71)
```

1a. Choose the first 175 of 250 observation as a training (estimation) sample and the rest as evaluation sample. Write an R program. (zero points)

Answer to a:

```
library(np)
attach(cps71)
n<-nrow(cps71)
n1<-175
n2<-n-n1
data.train<-cps71[1:n1,]
data.eval<-cps71[(n1+1):n,]
```

1b.Compute the predicted squared error for the evaluation data using the local constant estimator and the parametric model that is linear in age. (Hint1: First fit a model using the training sample, then use command predict to predict the value for the evaluation sample, and then calculate the mean of difference of evaluation sample and predicted value. Hint2: The model you will try to find is f1(wage)~ age. This is a model where the independent variable is age and the dependent variable is a certain function of the variable wage).
Write an R program. Enclose the program, the program output and your answer. (10 points)

2. The data("wage1") consists of a random sample taken from the U.S. Current Population Survey for the year 1976.

```
library(np)
data(wage1)
attach(wage1)
```

2a. Using common regression, explore the relationship of wages to edu, tenur and expr. (Hint1: Look for f1(wage)~educ+tenure+expre+f2(?), find the appropriate f1 and f2 and which one of the independent variables should be used instead of the question mark. Hint2: Remember, in this model f1(wage) means a certain function of the variable wage and f2(?) means a certain function of the variable question mark (?). Hint 3: The Model has educ, tenure, expre and f2(?) as independent variables and f1(wage) as the dependent variable. Hint4: Use R-squared to compare models). Write your best model in an R program. Enclose the program, the program output and your answer. (10 points)

2b. Fit kernel regression using local-constant estimator and local-linear estimator to your best model. Compare the common regression and kernel regressions, which one fits better? (Hint: Use R-squared to compare models). Write an R program. Enclose the program, the program output and your answer. (10 points)

Part III. Introduction to limit theorems in Statistics, the Bootstrap method and an introduction to the foundations of Statistics.

F8 Introduction to limit theorems in Statistics (Raul Cano)
(the home assignment for part II and part III will be given at F8. Important: Obligatory attendance at F8)
F9 The Bootstrap method (Raul Cano)
F10 An introduction to the foundations of Statistics (Raul Cano)

Home assignment for part III. Deadline for the assignment (part III): Thursday 31 may at my office room B742, between 12.00 and 13.00, lunch time. (Part III is worth 28 credit points)

1. Read the material (given by me): Chapter 1, INTRODUCTION. Summarize it in your own words in 2 pages. (7 points)

2. Read the article (given by me): A Leisurely Look at the Bootstrap... . Summarize it in your own words in 2 pages. (7 points)

3. Read the article (given by me): Bayesianism: Its Unifying role for... . Summarize it in your own words in 2 pages. (7 points)

4. Read the article (given by me): Frequentist probability and... . Summarize it in your own words in 2 pages. (7 points)

IMPORTANT: I am expecting to finish grading the last part of the home assignments on Monday 11 june. After the 11 of june, you will get your graded home assignments at the student office B724.

Forms of examination and grading criteria

The course (**Block II)** is examined through individual assignments (home assignments). All assignments are graded as approved or not approved. If they are not approved the assignment can be reworked and handed in once more. In addition all assignments are graded with credit points.

The assignment grades are added together. The maximal credit sum is 100 credit points. To pass the **Block II** (grades A – E) it is required that all assignments are approved.

All the credit points from the assignments need to be achieved at this period of teaching (semester).

Students not passing the course during the present semester, are requested to contact the course coordinator at the beginning of the course (the next time the course will be given) for information about the current rules for examination (they must make all the assignments again at that time).

No credit points from the assignments achieved at this semester can be transferred to the next time the course will be given.

The final grade of the **Block II** is based on a total count of the credit points from the assignments.

The maximal total credit is thus 100 points. Grades are given on a seven-point rating scale:

A.	Excellent	90-100 points
B.	Very good	80-89 points
C.	Good	70-79 points
D.	Satisfactory	60-69 points
E.	Enough	50-59 points
Fx.	Insufficient	40-49 points
F.	Completely insufficient	0-39 points

To pass the **Block II** a minimum grade of E is required.

Final grade for the complete course Statistical Methods (Block I and Block II), 15 ECTS. (Statistiska metoder, AN, 15 hp):

Grade for Block I and Block II (independent of the order)	**Final grade for the complete course**
A + A, A + B	A
A + C, B + B, B + C	B
A + D, A + E, B + D, B + E, C + C, C + D	C
C + E, D + D, D + E	D
E + E	E

Course coordinator and examinator:
Raul Cano, room B742, tel 162977.