Exam Statistics

Bachelor Econometrics and Operations Research
Bachelor Econometrics and Data Science
School of Business and Economics
Tuesday, March 27, 2019

Exam:	Statistics
Code:	E_EOR1_STAT
Coordinator:	M.H.C. Nientker
Co-reader:	J.M. Sneek
Date:	March 27, 2019
Time:	08:45
Duration:	2 hours
Calculator:	Not allowed
Graphical calculator:	Not allowed
Number of questions:	3
Type of questions:	Open
Answer in:	English
Credit score:	100 credits counts for a 10, you get 1 free credit
Grades:	Made public within 10 working days
Number of pages:	3, including front page

- Read the entire exam carefully before you start answering the questions.
- Be clear and concise in your statements, but justify every step in your derivations.
- The questions should be handed back at the end of the exam. Do not take it home.

Good luck!

Question 1 (12 points). During class we assumed that the heights of men can be modelled accurately by a normal distribution and thus we used the statistical model $\{g_{(\mu,\sigma^2)} \mid \mu \in \mathbb{R}, \sigma^2 > 0\}$, where

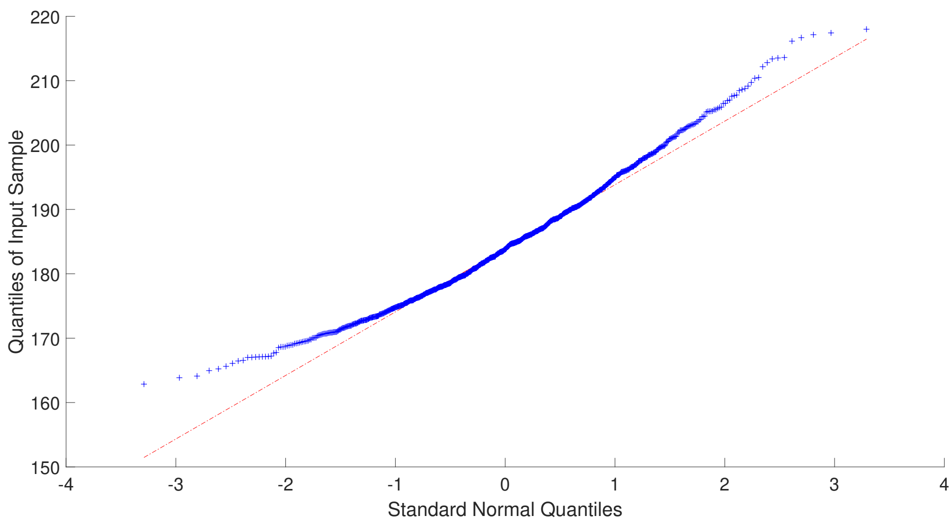
$$g_{(\mu,\sigma^2)}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}, \qquad x \in \mathbb{R}.$$

In practice we would like to check this assumption to make sure that our analysis is valid.

(6 points) a. Let $Z \sim \mathrm{N}(0,1)$. Show that the set of normal distributions $\{g_{(\mu,\sigma^2)} \mid \mu \in \mathbb{R}, \sigma^2 > 0\}$ is equal to the location-scale family of Z.

(6 points) b. We have measured one thousand random men and constructed a QQ-plot. Based on Figure 1, what do you conclude concerning the normality assumption we made? Why?

Figure 1: QQ-plot of data against the standard normal distribution



Question 2 (42 points). Let $X_1, \ldots, X_n$ be an independent and identically distributed sequence of random variables from a population in $\{g_\theta \mid \theta > 0\}$, where

$$g_\theta(x) = \frac{1}{\theta}, \qquad 0 \leq x \leq \theta.$$

(6 points) a. Derive the first moment $\mathbb{E}_\theta X_1$ and show that the moment estimator $\hat{\theta}_{MOM}$ of θ_0 is equal to $2\overline{X}$.

(9 points) b. Describe a weakness of $\hat{\theta}_{MOM}$ in this case and give a suggestion on how to improve it.

(12 points) c. Show that the ML estimator $\hat{\theta}_{ML}$ of θ_0 is equal to $X_{(n)} = \max\{X_1, \ldots, X_n\}$.

(6 points) d. Describe a weakness of $\hat{\theta}_{ML}$ in this case. Is this estimator biased for θ_0? Do not use any calculations to answer this question.

(9 points) e. Suppose that we know that $\theta_0 = 6$ and that we have simulated $x = (1, 1, 4, 4, 3, 4)$. Show that $\hat{\theta}_{MOM}$ is closer to θ_0 than $\hat{\theta}_{ML}$ and conclude that the moment estimator is closer for this realization. Why can't we conclude that $\hat{\theta}_{MOM}$ is better than $\hat{\theta}_{ML}$ in general? Propose a method to compare estimators and explain why it is better than comparing for a given realization.

Question 3 (45 points). Let $X = (X_1, \ldots, X_n)$ be an independent and identically distributed sequence of random variables from a population in $\{g_\theta \mid \theta > 0\}$, where

$$g_\theta(x) = \theta x^{\theta - 1} \qquad 0 \leq x \leq 1.$$

(12 points) a. Show that the maximum likelihood estimator of θ_0 is given by

$$\hat{\theta}_{ML} = -1/\overline{\log X} = -\left(\frac{1}{n}\sum_{i=1}^{n} \log X_i\right)^{-1}.$$

(9 points) b. Show that $T(X) = \sum_{i=1}^{n} \log X_i$ is a sufficient statistic for θ_0 using the factorization theorem.

(12 points) c. Calculate the Cramér-Rao lower bound for θ_0. Hint: use the theory on exponential families.

(12 points) d. Find the UMVU estimator for θ_0. Show that part c won't help here and explain why. You are allowed to use that

$$\mathbb{E}_\theta(1/T(X)) = -\frac{1}{n-1}\theta \qquad \text{and} \qquad \mathbb{V}\mathrm{ar}_\theta(1/T(X)) = \frac{1}{(n-1)^2(n-2)}\theta^2.$$