

Technologies for Data Analysis Part II

Homework

due 8 June 2014

Post-course survey

Please fill out the post-course satisfaction survey at https://hms.az1.qualtrics.com/SE/?SID=SV_01abcQNXyeFUZ4V This will help us improve the course next year. Thank you!

1 Nobel laureate age

The file `Nobel_laureates_1940s_births.xlsx` contains information about all awardees of the Nobel Prize in Physiology or Medicine who were born in the 1940s. For each of them, the file lists their year of birth and the age (in years) at which they were awarded the Prize.

Summary statistics

Compute and show some summary statistics. Is the mean or median a more appropriate measure?

Is your data normal?

Visualise the distribution of age at award. Does your data look normal?

Are Nobel laureates getting younger?

A similar analysis on Nobel laureates in Physiology or Medicine born in the 1960s has revealed an average age at award of 47 years with a standard deviation of 1.4. Editors at a large daily newspaper plan to run a story about this (“Are Nobel Laureates getting younger?”) and ring you up for your expert opinion. If you were to do a hypothesis test,

- What is your hunch the test will show?
- Would you choose a non-parametric or a parametric test?
- What is your Null Hypothesis?
- Would you choose a one-or two-tailed test?
- What significance level would you choose?
- What more information do you need?
- What might be a problem with this approach and how might you fix it?