DTU Library

Diversity in statistics projects – do students choose according to their gender?

Olsen, Niels Aske Lundtorp

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Olsen, N. A. L. (2022). Diversity in statistics projects – do students choose according to their gender? Technical University of Denmark.

Diversity in statistics projects – do students choose according to their gender?

Niels Lundtorp Olsen
Assistant Professor
DTU Compute
nalo@dtu.dk

Background

Diversity in STEM and how to increase diversity is a hot topic in today's discussions about research and education policy.

Some of the recent discussion in this regard points towards projects/examples that are used in education: are the projects more "appealing" to men than women? There are some natural constraints and focuses that STEM education must have, but even so there can still be a wide scope without changes to the actual content. A recent project at IT University of Denmark found a gender difference when high school students were to choose between two projects that differed in scope but were otherwise similar[1].

This report analyses student preferences at the course *Introduction to statistics* at Technical University of Denmark (DTU). The very large number of people in the study (>4000) makes it ideal for studying gender effects in relation to project preferences.

Introduction to Statistics (Introstat).

Introduction to statistics is a semi-mandatory course for all bachelor students at DTU worth 5 ECTS. It exists in three versions 02323 for Bachelor of Engineering (Diplomingeniør), 02402 for Bachelor of Science (Civilingeniør) and 02403 for BSc students specialized in mathematics and data analysis. 02403 is excluded from this study since the students do not get a choice of projects.

The curriculum and course material is around 95% similar for 02323 and 02402. Both courses have two mandatory projects that must be passed in order to qualify for the exam:

(1) Each student shall submit his or her project individually.
(2) The student can choose between four subjects, i.e. each project exists in four versions. These are thematically different and use different data sets, but the methods/analysis and "statistical content" is similar and also communicated as such – thus the individual student's selection is considered to be only due to "preference for subject".
(3) The subjects are similar in projects 1 and 2. Thus the student most likely selects the same subject for both projects, and our analysis therefore only covers project 1.

---

[1] I.M. Christensen et al.: *Computing Educational Activities Involving People Rather Than Things Appeal More to Women* (2021)*.* Proceedings of the 17th ACM Conference on International Computing Education Research.

This setup has been running since 2016 with minor changes in 2018. In total there have been around 4500 students on the two courses 2017-2021. Each course runs twice a year in spring and fall.

Figure 1: number of students who handed in project 1 in the different semesters. Black is 02323, red is 02402:
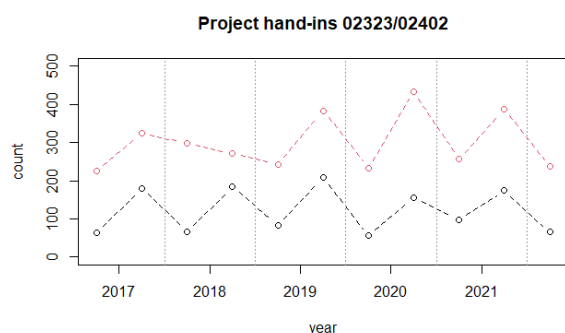
About the projects

BMI: This project studies the relation of body mass index (BMI) to other variables in the Danish population

Finans: This project analyses a data set of exchange-traded funds (ETF), a particular financial asset.

Skive Fjord: This project studies changes in nitrate and phosphate emissions in Skive Fjord.

Sønderborg: This project analyses heat consumption of four houses in Sønderborg.

This report does not analyse the projects in detail. A detailed analysis of the projects is indeed relevant but left for future studies.

Analysis

Data: Students who handed in a project since 2017, and who took the exam in 2021 or before.

Choice of variables

Gender: Students' gender (male/female).

Course: Did the student sign up for 02323 or 02402?

Study line: study line of the student. Study line is an obvious confounder as this is associated with both gender and choice of project. All study lines with less than 100 students were grouped into "other". This includes foreign students, master students and single-course students. See the appendix for an overview of abbreviations and study line names.

Project: Which of the four projects did the student select?

Methods

Aggregated tables of various parameters and subsets of students. All statistical tests were χ^2 tests for testing independence.

Results

Overall results

Table 2: Number of men and women selecting each project:

	female	male	percent	total
bmi	1015	1470	55%	**2485**
finans	131	604	16%	**735**
skivefjord	356	471	18%	**827**
sønderborg	122	346	10%	**468**
percent	36%	64%		
total	**1624**	**2891**		**4515**

Table 3: Frequencies of men and women selecting each project:

	female	male	both
bmi	63%	51%	55%
finans	8%	21%	16%
skivefjord	22%	16%	18%
sønderborg	8%	12%	10%

As can be seen from the Table 2 and 3, "bmi" has the highest preference both genders, with more than half of the students selecting "bmi", and "sønderborg" is the least popular choice. However women are more likely than men to choose "bmi": the odds-ratio for this event is 1.61. The biggest difference is seen in "finans", which men are more than twice as likely to select as women are (OR = 3.01).

Table 4: Frequencies for the two courses:

2323				2402		
	female	male			female	male
bmi	65%	54%		**bmi**	62%	50%
finans	5%	14%		**finans**	9%	24%
skivefjord	20%	16%		**skivefjord**	22%	17%
sønderborg	10%	17%		**sønderborg**	7%	10%

The only major difference between 2323 and 2402 is the higher preference for "finans" and lower preference for "sønderborg" among 2402 students compared to 2323 students.

All associations between gender and projects were extremely significant (p < 0.0001).

Results on different study lines

In this section, we report the project preferences among the different study lines.

Table 5: Frequencies of student selecting each project grouped onto study lines:

Study line	bmi	finans	skivefjord	sønderborg
miljø	38%	4%	53%	6%
fødevare	74%	2%	23%	1%
other	47%	26%	17%	11%
energi	48%	13%	9%	31%
biotek	58%	11%	28%	3%
byggetek	68%	9%	16%	7%
bygdesign	68%	4%	13%	15%
design og innov	41%	31%	18%	9%
elektro diplom	49%	9%	16%	25%
elektro civil	42%	25%	17%	16%
gen.eng	47%	27%	13%	12%
itelek	57%	13%	9%	22%
kemi og tek	48%	18%	31%	3%
kemi og biotek	54%	9%	24%	13%
kvantbio	76%	12%	10%	2%
medtek	84%	8%	6%	2%
software diplom	59%	12%	14%	16%
software civil	55%	21%	18%	6%
biomedicin	65%	15%	19%	0%
itøk	55%	23%	13%	10%

There is a clear evidence that study lines influence the choice of project. BMI was most popular across all study lines except miljø, but the three top scorers have a direct or indirect focus on human physiology. More patterns that can be observed from the table, we will not go through these here.

Study lines and gender

Having demonstrated that study line influences the choice of project, we investigated if there remain significant gender differences after adjusting for this confounder. This was done by tables of counts and χ^2 tests of independence, stratified on study lines.

Only categories with 20+ of each gender are included. Results of the χ^2 tests are reported in Table 6, see Table 7 in the appendix for frequency tables.

Table 6: p-values from chi^2 tests of independence, grouped on study lines

Study line	p value
miljø	0.039
fødevare	0.54
other	0
energi	0.53
biotek	0.015
byggetek	0.28
bygdesign	0.25
design og innov	0.018
gen.eng	0.6
kemi og tek	0.13
kemi og biotek	0
kvantbiologi	0.89
medtek	0.0001
software diplom	0.23
software civil	0.18
biomedicin	0.032
itøk	0.72

Discussion and conclusion

Even though a large part of the variation in preferences could be explained by the students' study lines, there still remained significant gender effects with a clear pattern: "bmi" was more preferred among women than men (yet with a high preference among men, too), and "finans" was more preferred among men than women.

We observed that study lines could explain much of the variation in project choices. However, the projects should still reflect diversity at the level of student population: study lines more popular among women seem to have less "choice" in their selection of projects.

The outline of projects in Introstat is a tiny detail in the studies of DTU students and of the diversity discussions at the university. However, the results of this study likely reflects many other projects and activities at DTU and other engineering universities. This study provides statistics that could serve as an indicator and guideline in these discussions.

Table 7: Project preferences according to gender and study line (left column = female, right column = male)

	miljø		fødevare		other	
bmi	34%	42%	72%	80%	52%	46%
finans	2%	8%	3%	0%	9%	31%
skivefjord	59%	42%	25%	20%	25%	13%
sønderborg	6%	7%	1%	0%	14%	10%
	energidesign		**biotek**		**byggetek**	
bmi	59%	44%	64%	51%	76%	65%
finans	7%	15%	6%	17%	4%	11%
skivefjord	7%	9%	29%	27%	12%	18%
sønderborg	28%	32%	1%	6%	8%	6%
	bygdesign		**design og innov**		**gen.eng**	
bmi	70%	65%	55%	33%	41%	49%
finans	2%	7%	19%	39%	29%	27%
skivefjord	12%	14%	18%	18%	13%	13%
sønderborg	16%	14%	7%	10%	16%	11%
	kemi og tek		**kemi og biotek**		**kvantbiologi**	
bmi	55%	43%	65%	42%	76%	76%
finans	6%	29%	5%	13%	12%	12%
skivefjord	36%	27%	22%	27%	10%	12%
sønderborg	4%	2%	8%	18%	2%	0%
	medtek		**software diplom**		**software civil**	
bmi	89%	77%	70%	58%	69%	53%
finans	3%	16%	5%	13%	8%	23%
skivefjord	8%	3%	15%	13%	17%	19%
sønderborg	1%	4%	10%	16%	6%	6%
	biomedicin		**itøk**			
bmi	68%	57%	65%	53%		
finans	10%	26%	17%	24%		
skivefjord	21%	16%	9%	14%		
sønderborg	1%	0%	9%	9%		

Study lines and abbreviations

Table 8: Study lines: names, abbreviations and programme.

Abbreviation	Danish title	English title	Programme
biomedicin	teknisk biomedicin	Human Life Science Engineering	BSc
biotek	bioteknologi	Biotechnology	BSc
bygdesign	bygningsdesign	Architectural engineering	BSc
byggetek	byggeteknologi	Civil engineering	BSc
design og innov	Design og innovation	Design and innovation	BSc
elektro civil	Elektroteknologi	Electrical engineering	BSc
elektro diplom	Elektroteknologi	Electrical engineering	BEng
energi	Bæredygtigt energidesign	Design of Sustainable Energy Systems	BSc
fødevare	Fødevaresikkerhed og -kvalitet	Food Safety and Quality	Beng
gen.eng	General engineering	General engineering	BSc
Itelek	It-elektronik	Computer Engineering	Beng
itøk	it og økonomi	IT and Economics	BEng
kemi og biotek	Kemi- og Bioteknik	Chemical and Bio Engineering	Beng
kemi og tek	kemi og teknologi	Chemistry and Technology	BSc
kvantbio	Kvantitativ Biologi og Sygdomsmodellering	Quantitative Biology and Disease Modelling	BSc
medtek	medicin og teknologi	Biomedical engineering	BSc
miljø	Miljøteknologi	Environmental Engineering	BSc
software civil	Softwareteknologi	Software technology	BSc
software diplom	Softwareteknologi	Software technology	BEng