Aashish Panta, Mairo Yamano, Myadaggarav Chuluundorj, Rebecca Mintah

Professor Bronchetti

F23 - ECON031-01

Dec 23, 2023

Statistical Analysis of Factors Affecting Career Concerns Relating to Major Choice of Swarthmore Students

Introduction

Around the United States, international students represent a significant portion of the student body, notably for Swarthmore it's16%. International students' academic paths are informed not only by their interests but also by their backgrounds and experiences. An in-depth understanding of their academic choices can only be synthesized in reference to the myriad of factors that affect decision-making. Since all of our group members were international students from different countries we agreed that our experiences at Swarthmore regarding family and career expectations were different from domestic students. Thus, our goal with this project is to investigate whether international and domestic students differ in their academic journeys at Swarthmore when choosing their major, specifically what other factors could contribute to the influence of career in major choices. The following questions summarize our areas of inquiry:

1. Does career consideration and family influence for major choice vary between domestic and international students?
2. How do international status, class year, gender, and personality affect the career consideration of students?

Methods

We surveyed 143 Swarthmore students of Swarthmore College using Google Forms (see Appendix 1) and compiled and analyzed the resulting data in Stata. To find Swarthmore student

respondents, we used convenience sampling as we asked people we knew, shared survey links in group chats and social media handles as well as approached students in public spaces such as Sci Commons to take our survey. Alongside this, we also sent out the survey to both sections of the ECON031 class. However, this method of data collection may have skewed the data as it is influenced by our connections on campus, which likely depend on our class year and major. On the other hand, due to practicality concerns, we did not implement other sampling methods.

Using multiple choice questions (see Appendix A), we collected data on class year, number of major changes and international status. Subsequently, studentswere asked to specify their gender and primary major using text boxes, offering flexibility in their answers. Moreover, they were asked to self-identify their personality on a numerical scale from 1, very introverted, to 5, very extroverted. In addition, the respondents rated the importance of career prospects and family expectations on majors using numerical scales from 1 to 5. Finally, we used Stata to conduct Ordinary Least Squares Multiple Linear Regression Analysis and t-tests to find statistically significant relationships between variables and test our hypotheses.

Preliminary Findings

Among the 143 responses across all grades, approximately 17% were seniors, 26% were juniors, 37% were sophomores, and 20% were freshmen (see Appendix 4). From these responses, we divided students into four groups of major divisions based on their primary major responses, including Natural Science, Social Science, Humanities, and Undecided. The majority of participants fell into the Natural Science and Social Science categories, accounting for an impressive 94% of the total respondents (see Appendix 5). This reveals the prevalent academic interest of our student population, although it is clear that the Humanities division is undersampled as it was reported to be 18% of students' majors (Swarthmore, 2022).

Students were asked to characterize their personalities on a scale from 1 (introverted) to 5 (extroverted). The observation had a roughly normal distribution, and the mean rating settled at 2.78 (see Appendix 6). This suggests a nuanced blend of introverted and extroverted tendencies in the sample, indicating that many students identify in between.

Furthermore, the survey asked students to report the number of times they have changed their majors. Initially, we anticipated a progressive increase in major changes with each passing year. We found a growth in the number of changes up until junior year, which means that current seniors had fewer changes than sophomores and juniors (see Appendix 7). This unexpected trend could be attributed to various factors, including external influences such as the COVID-19 pandemic, which might have discouraged current senior students from changing their majors during their crucial period of time.

In addition, we wanted to check if family expectations influence students' major choices. Around 75% of students acknowledged some level of family influence, rating between 2 and 4 on our scale (see Appendix 18). Remarkably, these ratings were evenly distributed, emphasizing that the family's voice might be a substantial part of the major decision-making process.

A significant portion of the survey investigates the impact of career considerations on students' decisions regarding their majors. Preliminary findings indicate that nearly half of the participants (61 respondents) acknowledge a substantial influence (rating of 5 on a scale of 1-5) on their major choices based on career considerations.

Approximately 30% of the surveyed participants identified themselves as international students, representing a demographic that is double the original size of Swarthmore's international student body, meaning we have oversampled the international student population at Swarthmore.

t-tests

We started our statistical analysis by conducting t-tests to better understand the differences in the means between the datasets and enable ourselves to draw meaningful conclusions. The t-tests helped us to determine if there were statistically significant differences between the means of the two groups.

From the beginning of the project, we assumed that international students might be more worried about career prospects when choosing their major, as the difficulty of obtaining work permits after graduation is a significant concern for many international students (Loo et al.). So, for the first t-test, we started by creating a dummy variable "international_num" where 1 represented international students and 0 represented non international students. Then, we compared whether the two groups were different in terms of "career_score" that asked the extent to which their consideration of their career or job search influenced their choice of major.

The results of the t-test (see Appendix 8) show that the mean "career_score" for international students is 3.84, whereas the mean "career_score" for non-internationals is slightly higher at 4.1. Despite the .25 difference in mean, the t-statistic of 1.33 and p-value of 0.1854 suggest that we fail to reject the null hypothesis using the significance level of 0.05. It implies that there is no significant difference in "career_score" between international and non-international students, which is counterintuitive to our initial assumptions.

The result of the first t-test (see Appendix 8) prompted us to inquire whether—despite no difference in the impact of career choice—international students and non international students have differences in the influence of family in their major choice. So, using a t-test, we compared the two groups in terms of "family_score" (out of five, to what extent their family expectations affect their major choice). The results of the t-test (see Appendix 8) show that the mean "family_score" for international students is 2.68 whereas the mean "career_score" for non-internationals is slightly

higher at 2.9. The t-statistic of 1.36 and p-value of 0.1748 suggest that we fail to reject the null hypothesis using the significance level of 0.05. It implies that there is no significant difference in "family_score" between international and non international students.

After observing a roughly normal distribution of career_score and family_score, we were curious if the students highly influenced by family expectations also had high career considerations in their major choice. So, for the third t-test, we created a dummy variable high_family_inf, which equaled 1 for students with family_score 3 or higher, and compared the two groups' career_score.

The results of the t-test (see Appendix 11) show that the mean career_score for high_family_inf is 4.25, whereas the mean career_score for non high_family_inf is much lower at 3.67. The t-statistic of -3.30 and p-value of 0.0012 suggest that we successfully reject the null hypothesis using the significance level of 0.05. It implies that there is a significant difference in career_score between students with high family influence and students with lower family influence. It allowed us to understand that students,who experience the strong influence of family expectations, prioritize career concerns in their major choice.

We further conducted multiple tests for career_score by categorizing groups by gender (using a dummy variable of female_num, see Appendix 10), by class year (using a dummy variable of senior, see Appendix 12), and by academic division (using a dummy variable of natural science, see Appendix 13). All three of these t-tests had a p-value of more than 0.05; thus, we failed to reject our null hypothesis using the significance level of 0.05.

However, we got an unexpected result when conducting a t-test for "career_score" by "very_introvert"—a dummy variable that equals one if a student marked their personality 1 (very introvert) in the personality score. The results of the t-test (see Appendix 5) show that the mean career_score for very introverted students is 3.46 whereas the mean career_score for more extroverted students is significantly higher at 4.076. The t-statistic of 1.9831 and p-value of 0.0493

suggest that we successfully reject the null hypothesis using the significance level of 0.05. Expect a naive hypothesis that very introverted students might go to graduate school and are not really concerned about their careers; we do not have any reasoning behind the result.

Regression Analysis

We conducted simple and multiple regressions to estimate the effect of known variables on career_score (dependent variable). We decided on the independent variables based on our results of the t-test, i.e., we intentionally focused on estimating the variation in career_score caused by statistically significant variables like very_introvert and high_family_inf.

First, we regressed "career_score" on the variable "very_introvert" to estimate the strength and direction of the relationship between the two variables. The results of the regression (see Appendix 15) show that very_introvert is a statistically significant variable at a 0.05 significance level. The -.615 negative coefficient of very_introvert implies that, on average, students who are "very introverts" tend to have a career_score lower by .615 compared to students who are not "very_introverts". The difference is nearly 15% of the mean, which is quite significant. However, adjusted R-squared is only 2%, which means a weak overall fit.

To strengthen the regression further, we added family_score as a variable and changed our linear regression to regression. The result of the regression (see Appendix 16) shows that both of the variables are statistically significant at a 0.05 significance level. The positive coefficient of family_score suggests that controlling other variables, on average, for every one-unit increase in family_score, the career_score increases by approximately 0.29 units. The coefficient of very_introvert remains more or less the same. At the same time, we observe a significant improvement in the adjusted R-squared value from 2% to 11%. This means adding the family_score as an independent variable in the regression provides a better prediction of career_score.

Finally, we conducted a regression with multiple variables by adding three more variables: international_num, natural_science, and social_science. The results (see Appendix 17) show that except family_score and very_introvert, none of the added variables are statistically significant. And the adjusted R-squared value does not improve from 10%.

Conclusion

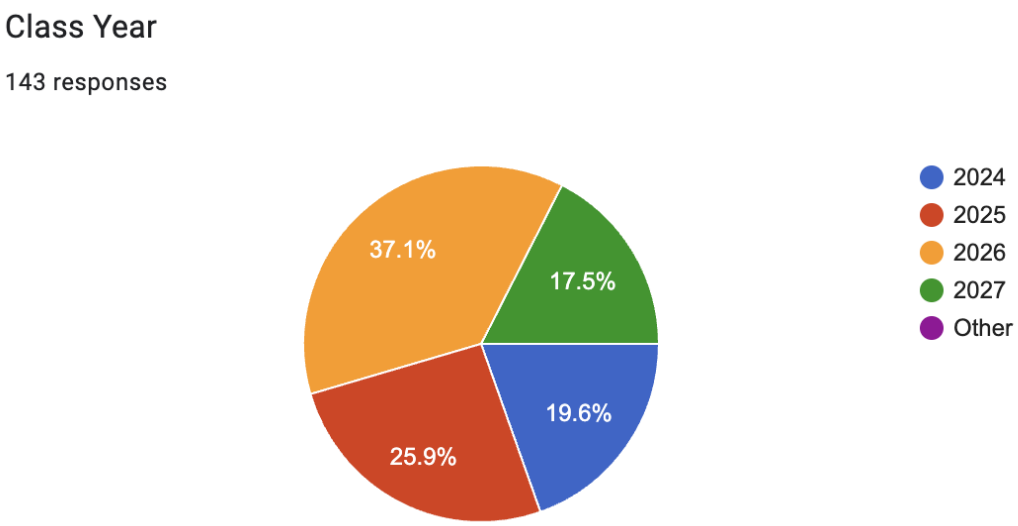
The absence of significant difference in career_score between international and non international students is counterintuitive to our initial assumptions. One potential explanation for the results might be the different economic backgrounds between international students at Swarthmore that most often tend to fall into very affluent or economically disadvantaged categories. Therefore, we couldn’t find a straightforward relationship between international status and career scores.

Besides, we are able to conclude that only the influence of family expectations in major choice and personality quality of being very introverted are strong predictors of the influence of career concerns in the major choice of a student based on the data we collected and the statistical analysis we conducted. However, we were only able to explain only 11% variability of career_score from its average using the independent variables. We believe that there are several variables we haven't accounted for, including family income, geographic location, education and profession of parents, number of dependents on family, and more that explain the strength of career concerns in major choice of students.

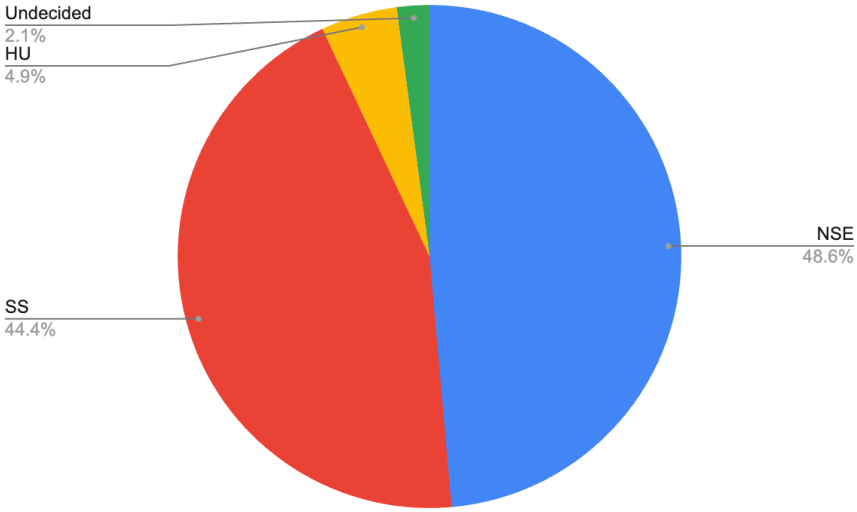
Overall, the analytical paper, even in a miniature way, provides a glimpse of numerous intricate factors that affect students' major choices. We believe that a nuanced understanding of how students' backgrounds and expectations from themselves enable higher education institutes to adequately support the academic success of students coming from diverse backgrounds.

Appendix:

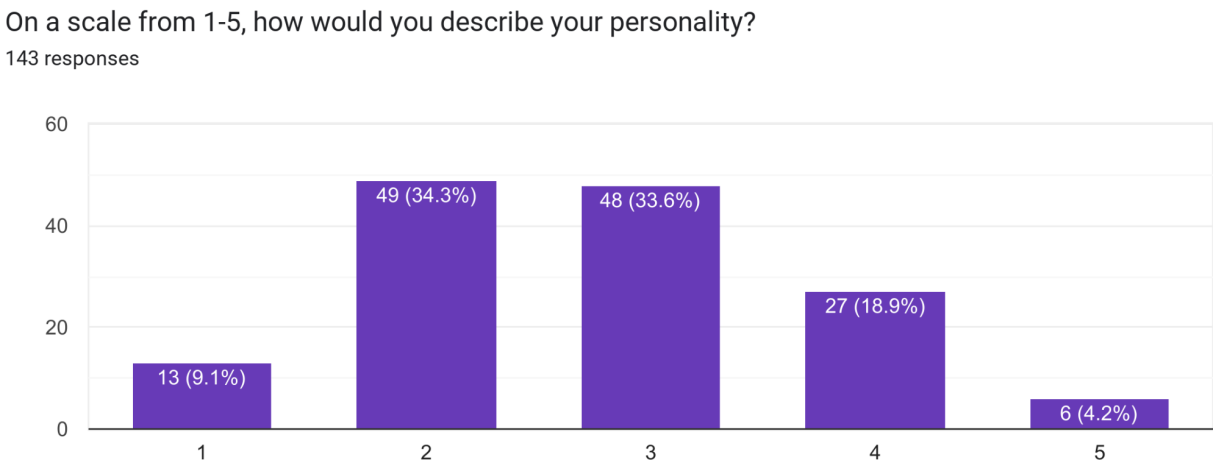
1. Link to the Survey Instrument
2. Project Data file
3. Project Do file
4. Pie chart of class distribution of response



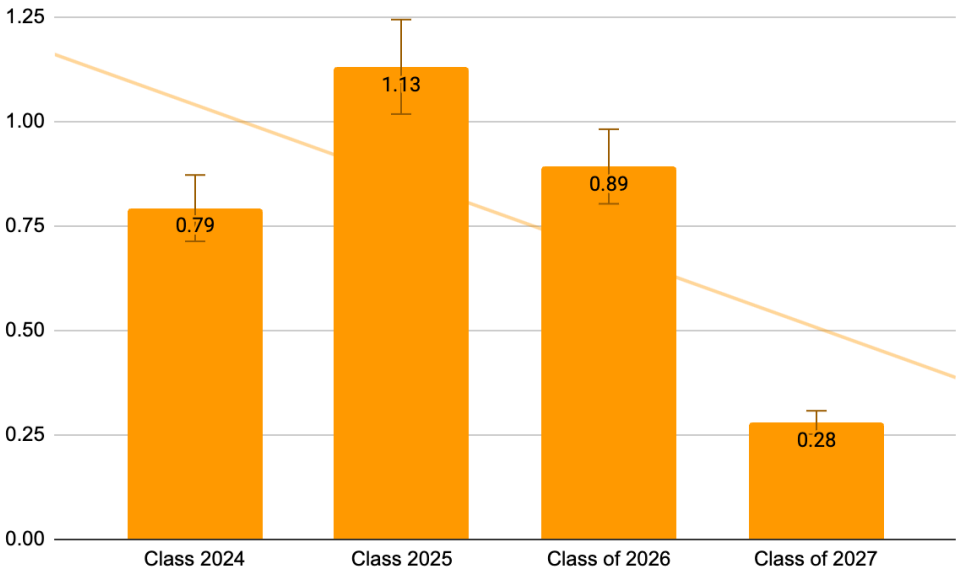
5. Pie chart of Division Distribution



6. Distribution of Personality Score



7. Average number of times students have changed their majors for class years



8. ttest result of career_score by international_num

```
. ttest career_score, by (international_num)

Two-sample t test with equal variances

   Group |     Obs        Mean    Std. err.   Std. dev.   [95% conf. interval]
---------+--------------------------------------------------------------------
       0 |      98    4.102041    .0988997    .9790569    3.905752    4.298329
       1 |      45    3.844444    .1879059    1.260511    3.465745    4.223144
---------+--------------------------------------------------------------------
Combined |     143    4.020979    .0901253    1.077742    3.842818     4.19914
---------+--------------------------------------------------------------------
    diff |            .2575964    .1935474               -.1250336    .6402263
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =   1.3309
H0: diff = 0                     Degrees of freedom =      141

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9073         Pr(|T| > |t|) = 0.1854          Pr(T > t) = 0.0927

.
```

9. ttest result of family_score by international_num

```
. ttest family_score, by (international_num)

Two-sample t test with equal variances

   Group |     Obs        Mean    Std. err.   Std. dev.   [95% conf. interval]
---------+--------------------------------------------------------------------
       0 |      98    2.979592    .1178205    1.166364    2.745751    3.213433
       1 |      45    2.688889    .1820817    1.221441    2.321927    3.055851
---------+--------------------------------------------------------------------
Combined |     143    2.888112    .0992957    1.187404    2.691823    3.084401
---------+--------------------------------------------------------------------
    diff |            .2907029    .2131752               -.1307298    .7121357
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =   1.3637
H0: diff = 0                     Degrees of freedom =      141

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9126         Pr(|T| > |t|) = 0.1748          Pr(T > t) = 0.0874
```

10. ttest career_score, by (female_num)

```
. ttest career_score, by (female_num)

Two-sample t test with equal variances

   Group |     Obs        Mean    Std. err.   Std. dev.   [95% conf. interval]
---------+--------------------------------------------------------------------
       0 |      87    4.034483    .1126202    1.050451    3.810601    4.258364
       1 |      56           4    .1507557    1.128152    3.697879    4.302121
---------+--------------------------------------------------------------------
Combined |     143    4.020979    .0901253    1.077742    3.842818     4.19914
---------+--------------------------------------------------------------------
    diff |            .0344828    .1852723               -.331788    .4007535
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =   0.1861
H0: diff = 0                     Degrees of freedom =      141

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.5737         Pr(|T| > |t|) = 0.8526          Pr(T > t) = 0.4263
```

11. ttest result of career_score by high_family_inf

```
. ttest career_score, by (high_family_inf)

Two-sample t test with equal variances

   Group |     Obs        Mean    Std. err.   Std. dev.   [95% conf. interval]
---------+--------------------------------------------------------------------
       0 |      58    3.672414    .1780243    1.355792    3.315926    4.028901
       1 |      85    4.258824    .0822528    .7583337    4.095255    4.422392
---------+--------------------------------------------------------------------
Combined |     143    4.020979    .0901253    1.077742    3.842818     4.19914
---------+--------------------------------------------------------------------
    diff |           -.5864097    .1774585               -.937233   -.2355865
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =  -3.3045
H0: diff = 0                     Degrees of freedom =      141

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0006         Pr(|T| > |t|) = 0.0012          Pr(T > t) = 0.9994
```

12. ttest career_score, by (senior)

```
. ttest career_score, by (senior)

Two-sample t test with equal variances
```

Group	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
0	115	4.052174	.0979138	1.050008	3.858207	4.24614
1	28	3.892857	.2261905	1.196887	3.428753	4.356962
Combined	143	4.020979	.0901253	1.077742	3.842818	4.19914
diff		.1593168	.2275285		-.2904916	.6091251

```
    diff = mean(0) - mean(1)                                      t =   0.7002
H0: diff = 0                                     Degrees of freedom =      141

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.7575         Pr(|T| > |t|) = 0.4850          Pr(T > t) = 0.2425
```

13. ttest career_score, by (natural_science)

```
. ttest career_score, by (natural_science)

Two-sample t test with equal variances
```

Group	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
0	75	3.906667	.1344637	1.16449	3.638742	4.174591
1	68	4.147059	.1171522	.9660615	3.913222	4.380896
Combined	143	4.020979	.0901253	1.077742	3.842818	4.19914
diff		-.2403922	.1799708		-.5961821	.1153978

```
    diff = mean(0) - mean(1)                                      t =  -1.3357
H0: diff = 0                                     Degrees of freedom =      141

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0919         Pr(|T| > |t|) = 0.1838          Pr(T > t) = 0.9081
```

14. ttest career_score, by (very_introvert)

```
. ttest career_score, by (very_introvert)

Two-sample t test with equal variances
```

Group	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
0	130	4.076923	.0904605	1.031409	3.897945	4.255901
1	13	3.461538	.3858953	1.391365	2.620745	4.302332
Combined	143	4.020979	.0901253	1.077742	3.842818	4.19914
diff		.6153846	.3103134		.0019162	1.228853

```
    diff = mean(0) - mean(1)                                  t =   1.9831
H0: diff = 0                     Degrees of freedom =      141

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9754         Pr(|T| > |t|) = 0.0493          Pr(T > t) = 0.0246
```

15. Regress career_score very introvert

```
. regress career_score very_introvert
```

Source	SS	df	MS	Number of obs	=	143
				F(1, 141)	=	3.93
Model	4.47552448	1	4.47552448	Prob > F	=	0.0493
Residual	160.461538	141	1.1380251	R-squared	=	0.0271
				Adj R-squared	=	0.0202
Total	164.937063	142	1.16152861	Root MSE	=	1.0668

career_score	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
very_introvert	-.6153846	.3103134	-1.98	0.049	-1.228853	-.0019162
_cons	4.076923	.093563	43.57	0.000	3.891955	4.261891

16. Regress career_score family_score very_introvert

```
. regress career_score family_score very_introvert
```

Source	SS	df	MS	Number of obs	=	143
				F(2, 140)	=	10.34
Model	21.2315461	2	10.615773	Prob > F	=	0.0001
Residual	143.705517	140	1.02646798	R-squared	=	0.1287
				Adj R-squared	=	0.1163
Total	164.937063	142	1.16152861	Root MSE	=	1.0131

career_score	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
family_score	.2896653	.0716941	4.04	0.000	.1479221	.4314084
very_introvert	-.6755459	.2950876	-2.29	0.024	-1.25895	-.0921418
_cons	3.245807	.2240788	14.49	0.000	2.802791	3.688822

17. Multiple regression

```
. regress career_score family_score international_num natural_science social_science female_n
> um very_introvert
```

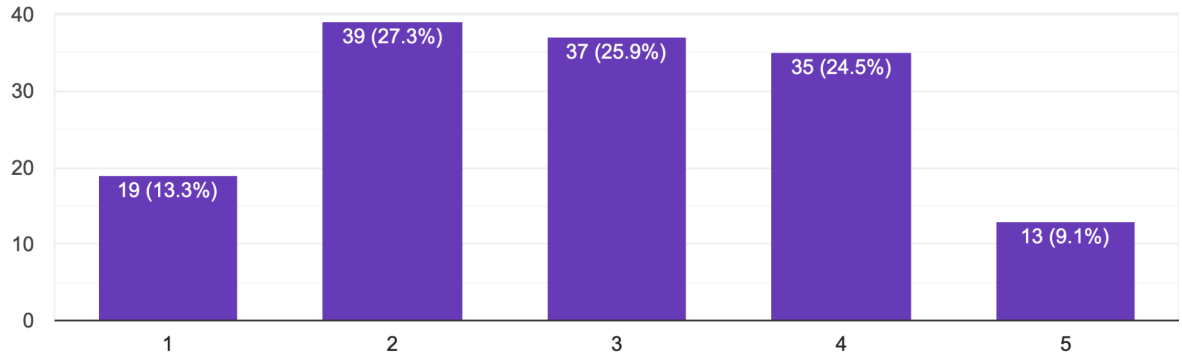
Source	SS	df	MS	Number of obs	=	143
				F(6, 136)	=	4.20
Model	25.7690199	6	4.29483665	Prob > F	=	0.0007
Residual	139.168043	136	1.02329443	R-squared	=	0.1562
				Adj R-squared	=	0.1190
Total	164.937063	142	1.16152861	Root MSE	=	1.0116

career_score	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
family_score	.2991033	.0737891	4.05	0.000	.1531809	.4450256
international_num	-.1291921	.1852735	-0.70	0.487	-.4955818	.2371976
natural_science	.3714232	.3207791	1.16	0.249	-.262937	1.005783
social_science	.1111264	.3244952	0.34	0.733	-.5305827	.7528354
female_num	-.1422454	.1798295	-0.79	0.430	-.4978692	.2133784
very_introvert	-.7362683	.3015457	-2.44	0.016	-1.332593	-.1399433
_cons	3.09485	.400436	7.73	0.000	2.302963	3.886736

18.

On a scale from 1-5, how much do your family or their expectations affect your major choice?

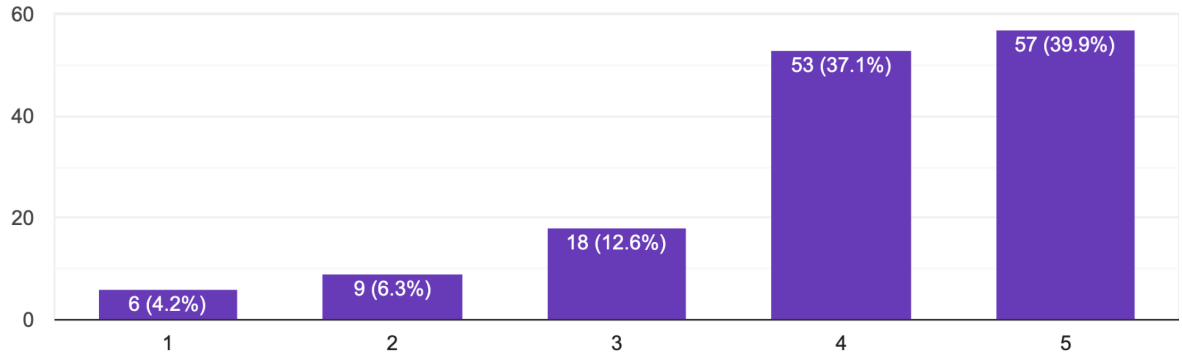
143 responses



19.

On a scale from 1-5, to what extent has the consideration of your career or job search influenced your major choice?

143 responses



Bibliography:

Swarthmore College, Division of Graduate, 2022 https://www.swarthmore.edu/sites/default/files/assets/documents/institutional-research/degsmajorchart.pdf

Loo, Bryce, et al. "Career Prospects and Outcomes of U.S.-Educated International Students: Improving Services, Bolstering Success. Report 09." World Education Services, Oct. 2017, pp. 1-64. https://eric.ed.gov/?id=ED586191