# Problem Statement

A sample of 62 undergraduate students were surveyed and the result of the survey is available in UndergradSurvey.xlsx. In addition to the basic information like gender, age, class and major, it also records whether they would like to go for graduate studies (Grad Intention), current GPA, whether they are currently employed (Employment), their expected annual starting salary (Salary), in how many social networking sites they are registered (Social Networking), their satisfaction level with food and dining services on campus (Satisfaction), spending on textbooks and supplies in current semester (Spending), preferred type of computer (Computer), weekly average number of text messages sent (Text Messages) and how much wealth they would like to accumulate (in millions of dollars) before they would consider themselves rich (Wealth).

# Summary of Data

The data from the survey can be categorized into 2 types, Categorical Data and Numerical Data.

The Categorical Fields are

|  |
| --- |
|  |
| * Gender |
| * Age |
| * Class |
| * Major |
| * Grad Intent |
| * Employment |
| * Social Sites |
| * Satisfaction |
| * Device |

A Graphical Representation of these fields is given in below graph:

The Numerical Fields are

|  |
| --- |
| * Age |
| * GPA |
| * Salary |
| * Spending |
| * Text Messages |
| * Wealth |

\*Age can be both Categorical and Numerical

The Data analysis of these fields is as below.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | *Age* | *GPA* | *Salary* | *Spending* | *Text Messages* | *Wealth* |
|  |  |  |  |  |  |  |
| Mean | 21.12903 | 3.12629 | 48.54839 | 482.0161 | 246.2097 | 7.15121 |
| Standard Error | 0.181777 | 0.048051 | 1.534277 | 28.18816 | 27.2372 | 2.490282 |
| Median | 21 | 3.135 | 50 | 500 | 200 | 1 |
| Mode | 21 | 3.4 | 40 | 500 | 300 | 1 |
| Standard Deviation | 1.431311 | 0.378354 | 12.08091 | 221.9538 | 214.466 | 19.6085 |
| Sample Variance | 2.048652 | 0.143152 | 145.9484 | 49263.49 | 45995.64 | 384.4934 |
| Kurtosis | 1.396513 | -0.57768 | 0.424264 | 4.559914 | 1.135685 | 16.98387 |
| Skewness | 0.736533 | -0.27809 | 0.534701 | 1.585915 | 1.295808 | 4.158268 |
| Range | 8 | 1.56 | 55 | 1300 | 900 | 99.9 |
| Minimum | 18 | 2.34 | 25 | 100 | 0 | 0.1 |
| Maximum | 26 | 3.9 | 80 | 1400 | 900 | 100 |
| Sum | 1310 | 193.83 | 3010 | 29885 | 15265 | 443.375 |
| Count | 62 | 62 | 62 | 62 | 62 | 62 |

|  |  |
| --- | --- |
| **Statistic** | **Size** |
| Whole Class | 62 |
| Female | 33 |
| Male | 29 |

The sample is well distributed based on gender. Variance is more in the data for female students in all aspects compared to that of male students.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Entire Class** | | | | | | |
| **Statistic** | **Age** | **GPA** | **Salary** | **Spend** | **Text Messaging** | **Wealth** |
| **Mean** | 21.1290322 | 3.1 | 48.5 | 482.016129 | 246.2 | 7.2 |
| **Median** | 21 | 3.1 | 50.0 | 500 | 200.0 | 1.0 |
| **Mode** | 21 | 3.4 | 40 | 500 | 300 | 1 |
| **SD** | 1.43131111 | 0.37835378 | 12.0809122 | 221.953805 | 214.4659503 | 19.6085024 |
| **Gender= Female** | | | | | | |
| **Statistic** | **Age** | **GPA** | **Salary** | **Spend** | **Text Messaging** | **Wealth** |
| Mean | 21.0909090 | 3.2 | 48.8 | 452.121212 | 237.4242424 | 2.40 |
| Median | 21 | 3.2 | 50.0 | 450 | 200 | 1.00 |
| Mode | 21 | 3 | 50 | 500 | 300 | 1 |
| SD | 1.35470560 | 0.39434568 | 13.2724045 | 204.584297 | 213.3480183 | 5.36043127 |
| **Gender = Male** | | | | | | |
| **Statistic** | **Age** | **GPA** | **Salary** | **Spend** | **Text Messaging** | **Wealth** |
| Mean | 21.1724137 | 3.1 | 48.3 | 516.034482 | 256.2068966 | 12.55 |
| Median | 21 | 3.1 | 47.0 | 500 | 200 | 2.50 |
| Mode | 21 | 3.4 | 40 | 500 | 300 | 10 |
| SD | 1.5369023 | 0.3598973 | 10.793174 | 239.24501 | 219.0672401 | 27.3499181 |

# 95% Confidence Intervals

* Mean GPA for the entire Class

**95% confidence interval** GPA for entire class: 3.0302-3.2223

|  |  |
| --- | --- |
| *GPA* |  |
|  |  |
| Mean | 3.126290323 |
| Standard Error | 0.048050979 |
| Median | 3.135 |
| Mode | 3.4 |
| Standard Deviation | 0.378353785 |
| Sample Variance | 0.143151586 |
| Sum | 193.83 |
| Count | 62 |
| t-alpha/2 | 1.999623585 |
| Margin of error | 0.09608387 |
| **Confident Interval** | **3.030206452** |
|  | **3.222374193** |

* Mean GPA for males and females separately

GPA Females: 3.03-3.31 GPA *Males*: 2.94–3.21

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *GPA* | *Females* |  | *GPA* | *Males* |
|  |  |  |  |  |
| Mean | 3.169090909 |  | Mean | 3.077586207 |
| Standard Error | 0.068646773 |  | Standard Error | 0.066831262 |
| Median | 3.19 |  | Median | 3.1 |
| Mode | 3 |  | Mode | 3.4 |
| Standard Deviation | 0.394345689 |  | Standard Deviation | 0.359897358 |
| Sample Variance | 0.155508523 |  | Sample Variance | 0.129526108 |
| Sum | 104.58 |  | Sum | 89.25 |
| Count | 33 |  | Count | 29 |
| t-alpha/2 | 2.036933343 |  | t-alpha/2 | 2.048407142 |
| Margin of error | 0.139828901 |  | Margin of error | 0.136897634 |
| **Confident Interval** | **3.029262008** |  | **Confident Interval** | **2.940688573** |
|  | **3.30891981** |  |  | **3.214483841** |

* Mean salary

**95% confidence interval** Salary for entire class: 45.4804-51.6163

|  |  |
| --- | --- |
| *Salary* |  |
|  |  |
| *Mean* | *48.5483871* |
| *Standard Error* | *1.534277386* |
| *Median* | *50* |
| *Mode* | *40* |
| *Standard Deviation* | *12.08091222* |
| *Sample Variance* | *145.94844* |
| *Sum* | *3010* |
| *Count* | *62* |
| *t-alpha/2* | *1.999623585* |
| *Margin of error* | *3.067977246* |
| ***Confident Interval*** | ***45.48040985*** |
|  | ***51.61636434*** |

* Mean salary for males and females separately

Salary *Females*: 44.08-53.49 Salary *Males*: 44.17–52.38

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Salary* | *Females* |  | *Salary* | *Males* |
|  |  |  |  |  |
| Mean | 48.78787879 |  | Mean | 48.27586207 |
| Standard Error | 2.310429079 |  | Standard Error | 2.004242146 |
| Median | 50 |  | Median | 47 |
| Mode | 50 |  | Mode | 40 |
| Standard Deviation | 13.27240459 |  | Standard Deviation | 10.79317427 |
| Sample Variance | 176.1567235 |  | Sample Variance | 116.4926108 |
| Sum | 1610 |  | Sum | 1400 |
| Count | 33 |  | Count | 29 |
| t-alpha/2 | 2.036933343 |  | t-alpha/2 | 2.048407142 |
| Margin of error | 4.706190029 |  | Margin of error | 4.105503926 |
| **Confident Interval** | **44.08168876** |  | **Confident Interval** | **44.17035814** |
|  | **53.49406882** |  |  | **52.381366** |

* Mean spending

**95% confidence interval** spending for entire class: 425.6504-538.3818

|  |  |
| --- | --- |
| Spending |  |
|  |  |
| Mean | 482.016129 |
| Standard Error | 28.18816142 |
| Median | 500 |
| Mode | 500 |
| Standard Deviation | 221.953805 |
| Sample Variance | 49263.49154 |
| Sum | 29885 |
| Count | 62 |
| t-alpha/2 | 1.999623585 |
| Margin of error | 56.36571239 |
| **Confident Interval** | **425.6504166** |
|  | **538.3818414** |

* Mean spending for males and females separately

Spending *Females*: 379.58-524.66 Spending *Males* : 425.03-607.04

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Spending* | *Females* |  | *Spending* | *Males* |
|  |  |  |  |  |
| Mean | 452.1212121 |  | Mean | 516.0344828 |
| Standard Error | 35.61355498 |  | Standard Error | 44.42668402 |
| Median | 450 |  | Median | 500 |
| Mode | 500 |  | Mode | 500 |
| Standard Deviation | 204.5842977 |  | Standard Deviation | 239.2450153 |
| Sample Variance | 41854.73485 |  | Sample Variance | 57238.17734 |
| Sum | 14920 |  | Sum | 14965 |
| Count | 33 |  | Count | 29 |
| t-alpha/2 | 2.036933343 |  | t-alpha/2 | 2.048407142 |
| Margin of error | 72.54243762 |  | Margin of error | 91.00393683 |
| **Confident Interval** | **379.5787745** |  | **Confident Interval** | **425.0305459** |
|  | **524.6636497** |  |  | **607.0384196** |

* Mean wealth

**95% confidence interval** Wealth for entire class: 2.1715-12.1308

|  |  |
| --- | --- |
| Wealth |  |
|  |  |
| Mean | 7.151209677 |
| Standard Error | 2.490282305 |
| Median | 1 |
| Mode | 1 |
| Standard Deviation | 19.60850248 |
| Sample Variance | 384.4933694 |
| Sum | 443.375 |
| Count | 62 |
| t-alpha/2 | 1.999623585 |
| Margin of error | 4.97962723 |
| **Confident Interval** | **2.171582447** |
|  | **12.13083691** |

* Mean wealth for males and females separately

Wealth *Females*: 0.50-4.30 Wealth *Males*: 2.15-22.96

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Wealth* | *Females* |  | *Wealth* | *Males* |
|  |  |  |  |  |
| Mean | 2.403787879 |  | Mean | 12.55344828 |
| Standard Error | 0.933131312 |  | Standard Error | 5.078752302 |
| Median | 1 |  | Median | 2.5 |
| Mode | 1 |  | Mode | 10 |
| Standard Deviation | 5.360431278 |  | Standard Deviation | 27.34991816 |
| Sample Variance | 28.73422348 |  | Sample Variance | 748.0180234 |
| Sum | 79.325 |  | Sum | 364.05 |
| Count | 33 |  | Count | 29 |
| t-alpha/2 | 2.036933343 |  | t-alpha/2 | 2.048407142 |
| Margin of error | 1.900726283 |  | Margin of error | 10.40335249 |
| **Confident Interval** | **0.503061596** |  | **Confident Interval** | **2.150095789** |
|  | **4.304514162** |  |  | **22.95680076** |

# Hypothesis Testing

Salary expectation for male undergraduates is higher than salary expectation for female undergraduates

µsalarymale : Mean Salary Expectation of Male Undergraduates

µsalaryfemale : Mean Salary Expectation of female Undergraduates

Hypothesis:

H0: µsalarymale - µsalaryfemale >= 0

Ha: µsalarymale - µsalaryfemale < 0

|  |  |
| --- | --- |
| t test statistic | -0.16246964 |
| t probability | 0.43574099 |
| t probability > 0.05 | |
| **We do not Reject Null Hypothesis.** | |

There is not enough evidence to reject null hypothesis.

1. Males will consider themselves rich with significantly higher wealth accumulation compared to females.

µwealthmale : Mean of Male wealth accumulation

µwealthfemale : Mean of Female wealth accumulation

Hypothesis:

H0: µwealthmale - µwealthfemale >= 0

Ha: µwealthmale - µwealthfemale < 0

|  |  |
| --- | --- |
| t test statistic | 2.054664 |
| t probability | 0.977863 |
| t probability > 0.05 | |
| **We do not Reject Null Hypothesis** | |

There is not enough evidence to reject null hypothesis.

1. Females spend significantly less on textbooks and supplies compared to males

µspendingmale : Mean of Males spend on textbooks and supplies

µspendingfemale : Mean of Females spend on textbooks and supplies

Hypothesis:

H0: µspendingmale - µspendingfemale >= 0

Ha: µspendingmale - µspendingfemale < 0

|  |  |
| --- | --- |
| t test statistic | 1.115378 |
| t Probability | 0.865432 |
| t probability > 0.05 | |
| **We do not Reject Null Hypothesis** | |

There is not enough evidence to reject null hypothesis.

1. Undergraduate students who have decided to go for graduate studies (Grad Intention = Yes) have a significantly higher GPA than others.

µgpagrad : Mean GPA of Undergraduate students who have decided to go for graduate studies (Grad Intention = Yes)

µgpaothers : Mean GPA of Undergraduate students who have not yet decided and decided not to go for graduate studies (Grad Intention = No, undecided)

Hypothesis:

H0: µgpagrad - µgpaothers >= 0

Ha: µgpagrad - µgpaothers < 0

|  |  |  |
| --- | --- | --- |
| t test statistic |  | -0.89832 |
| t Probability |  | 0.186303 |
| **t probability > 0.05** | | |
| **We do not Reject Null Hypothesis** | | |

There is not enough evidence to reject null hypothesis.

1. Employment status is independent of gender

Hypothesis:

H0: Employment Status and Gender are independent. (There is no relationship between the two variables)

Ha: Employment Status and Gender are dependent. (There is a relationship between the two variables)

|  |  |  |  |
| --- | --- | --- | --- |
| **Observed Frequencies** | | |  |
| Row Labels | Female | Male | Grand Total |
| Full-Time | 3 | 7 | 10 |
| Part-Time | 24 | 19 | 43 |
| Unemployed | 6 | 3 | 9 |
| Grand Total | 33 | 29 | 62 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Expected Frequencies** | | |  |
| Row Labels | Female | Male | Grand Total |
| Full-Time | 5.322580645 | 4.677419355 | 10 |
| Part-Time | 22.88709677 | 20.11290323 | 43 |
| Unemployed | 4.790322581 | 4.209677419 | 9 |
| Grand Total | 33 | 29 | 62 |

|  |  |  |
| --- | --- | --- |
| **Chi Square Contributions** | | |
| Row Labels | Female | Male |
| Full-Time | 1.013489736 | 1.153281424 |
| Part-Time | 0.054115802 | 0.06158005 |
| Unemployed | 0.305474096 | 0.347608454 |

|  |  |
| --- | --- |
| Level of Significance | 0.05 |
| No. of Rows | 3 |
| No. of Columns | 2 |
| Degrees of Freedom | 2 |
| Critical Value | 5.991464547 |
| chi square stat | 2.935549561 |
| p value | 0.230437689 |
| P value > 0.05 |  |
| **We do not Reject Null Hypothesis** | |

There is not enough evidence to reject null hypothesis.

1. Employment status is independent of Grad Intention

H0: Employment Status and Gender are independent. (There is no relationship between the two variables)

Ha: Employment Status and Gender are not independent. (There is a relationship between the two variables)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Observed Frequency** | | | |  |
| Row Labels | No | Undecided | Yes | Grand Total |
| Full-Time | 1 | 1 | 8 | 10 |
| Part-Time | 10 | 16 | 17 | 43 |
| Unemployed | 1 | 5 | 3 | 9 |
| Grand Total | 12 | 22 | 28 | 62 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Expected Frequency** | | | |  |
| Row Labels | No | Undecided | Yes | Grand Total |
| Full-Time | 1.935483871 | 3.548387097 | 4.516129032 | 10 |
| Part-Time | 8.322580645 | 15.25806452 | 19.41935484 | 43 |
| Unemployed | 1.741935484 | 3.193548387 | 4.064516129 | 9 |
| Grand Total | 12 | 22 | 28 | 62 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Chi Square Contributions** | | | |
| Row Labels | No | Undecided | Yes |
| Full-Time | 0.452150538 | 1.830205279 | 2.687557604 |
| Part-Time | 0.338084521 | 0.036077201 | 0.301414639 |
| Unemployed | 0.316009558 | 1.021831215 | 0.278801843 |

|  |  |
| --- | --- |
| Level of Significance | 0.05 |
| No. of Rows | 3 |
| No. of Columns | 3 |
| Degrees of Freedom | 4 |
| Critical Value | 9.487729037 |
| chi square stat | 7.262132398 |
| p value | 0.122667348 |
| p value>0.05 |  |
| **We do not Reject Null Hypothesis** | |

There is not enough evidence to reject null hypothesis.

\*R-Script attached for the complete solution. 