UTM UNIVERSITI TEKNOLOGI MALAYSIA

SCHOOL OF COMPUTING
Faculty of Engineering

SESSION 2020/2021 SEMESTER 2

SECI2143 – PROBABILITY AND STATISTICAL DATA ANALYSIS
DR NOR AZIZAH ALI

PROJECT 2

SECTION: 05

GROUP: F

NO	NAME	MATRIC NUMBER
1.	NURIN FARZANAH BINTI MOHD HILMI	A20EC0122
2.	NURUL HUDA BINTI NOR DIN	A20EC0126
3.	NAFIS AHMED	A20EC4044
4.	SHAH SAJID	A20EC4050

TABLE OF CONTENTS

INTRODUCTION

In this project 2, we will practise and use all of the knowledge about inferential statistic that we learn throughout the semester. There are four inferential statistics that will be used in this project which are hypothesis testing on two sample, correlation, analysis of variance (ANOVA) and chi square test (one way contingency). Other than that, all test will be calculated by using R Studio since the objective of this project is to apply the uses of R Studio in calculation of four inferential statistic.

DATASET

The secondary data set that we choose is from the American Community Survey 2010-2012. This data set is about how many people who got employed based on their major and major category. This data set includes the major, major category, total of college graduates per major, the total of the employed, employed full time and unemployed college graduates per major. Firstly, for hypothesis testing on two sample, we will study about the differences proportion of college graduates that got employed between Architecture major and Computer Science major. Secondly, for chi square test (one way contingency), the differences proportion of employed college graduates between five majors (General Agriculture, Agriculture Production and Management, Agricultural Economics, Animal Sciences, and Food Science). Thirdly, we will find the correlation coefficient between employed full time and unemployed under Computer and Mathematics category. Lastly, for ANOVA, the differences mean of employed college graduates between three major categories (Education, Health and Business) will be tested.

DATA ANALYSIS

HYPOTHESIS TESTING ON TWO SAMPLE

The claim is that "students of Architecture major are less likely to be employed than college graduates of Computer Science major".

Sample data:

Table 1: Sample data hypothesis testing

	Architecture	Computer Science
Employed, x	$x_1 = 216770$	$x_2 = 656372$
Sample, n	$n_1 = 294692$	$n_2 = 783292$

- Significance level, $\alpha = 0.05$
- Hypothesis Statement:

$$H_0 : p_1 = p_2$$

$$H_1 : p_1 < p_2$$

- Pooled Sample Proportion:

$$\bar{p} = \frac{x_1+x_2}{n_1+n_2} = 0.80998 \qquad \bar{q} = 1 - \bar{p} = 0.19002$$

- Test Statistic for Two Proportions:

$$z = \frac{(\widehat{p_1} - \widehat{p_2}) - (p_1 - p_2)}{\sqrt{\frac{\bar{p}\bar{q}}{n_1} + \frac{\bar{p}\bar{q}}{n_2}}}$$

z	-120.762990889292
z.alpha	-1.95996398454005

Figure 1:Test statistic & critical value

Conclusion:

Test Statistic, $z \leq$ Critical Value

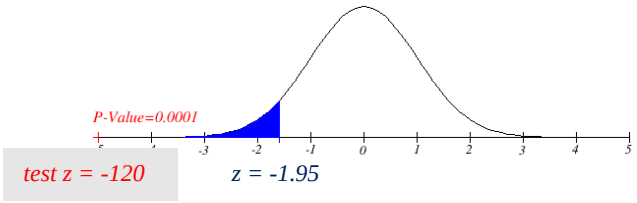

Figure 2: Normal distribution graph

We reject H_0 because the *P*-value of 0.0001 is less than $\alpha = 0.05$. There is sufficient evidence to support the claim of H_1 , where students that took Architecture major are less likely to be employed than students that took Computer Science major.

CORRELATION

What is the relationship between employed full time, x and unemployed, y for 8 major codes under Computers and Mathematics major category?

Sample data:

Table 2 Correlation(x,y)

Major Code	**Employed Full Time, x**	**Unemployed, y**
MATHEMATICS AND COMPUTER SCIENCE	5039	150
APPLIED MATHEMATICS	12109	892
STATISTICS AND DECISION SCIENCE	14468	1138
COMPUTER PROGRAMMING AND DATA PROCESSING	18747	2265
COMPUTER ADMINISTRATION MANAGEMENT AND SECURITY	28156	2626
COMPUTER NETWORKING AND TELECOMMUNICATIONS	35954	2748
COMMUNICATION TECHNOLOGIES	37261	4609
INFORMATION SCIENCES	57604	3704

- Compute the value of the correlation coefficient, r between x and y.
- Construct a scatter plot and label it.

Solution:

Hypothesis:

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

$$test\ statistic, t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} \qquad \alpha = 0.05$$

correlation coefficient, r:

```
> cor(x,y)
[1] 0.8621754
>
```

Figure 3: r=0.862

r = 0.862

- test statistic, t and critical value, t(α/2, n-2):

t	4.16536927761342
t.alpha	2.44691185114497

Figure 4: test statistic, t and critical value

Conclusion:

$$\therefore \ test\ statistic, t\ =\ 4.165\ >\ 2.447\ , we\ reject\ null\ hypothesis, H_0.$$

Discussion:

Because there is sufficient evidence that linear correlation exists between x and y.

- Construct a scatter plot and label it:

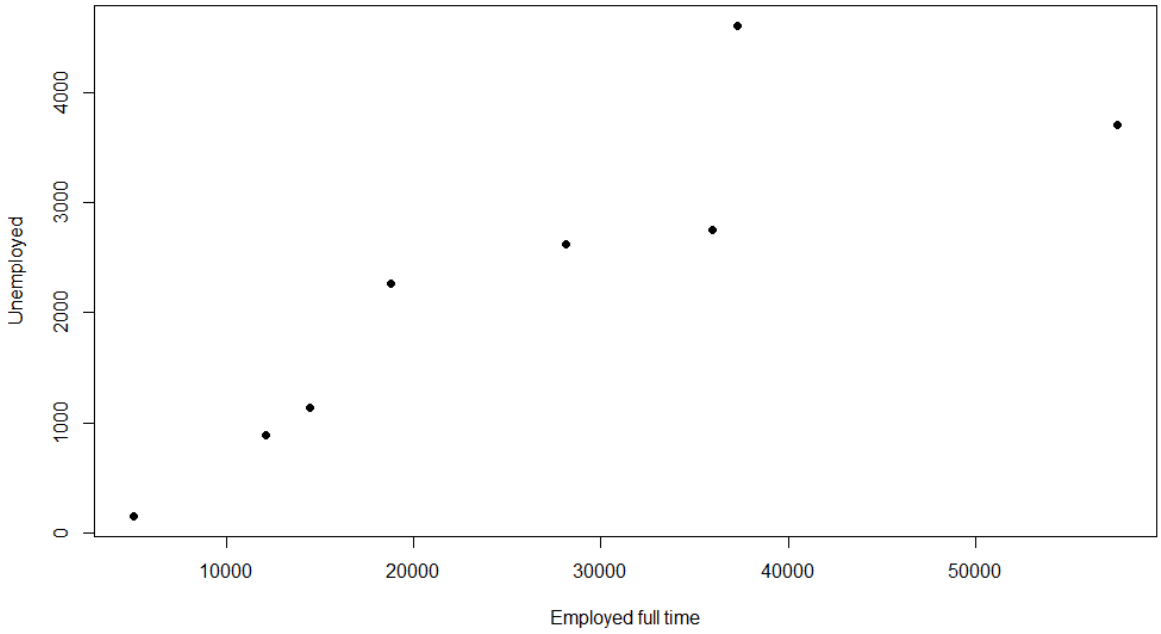

Figure 5: Scatter graph

Conclusion:

∴ Since r > 0.8, the relationship between x and y is strong positive linear relationship. It means that the when employed full time in major Computer and Mathematics increases, the unemployed also increase. We can conclude that if a lot of employees offer a worker as full-time worker, it causes a lot of other people are unemployed.

CHI SQUARE TEST (ONE WAY CONTINGENCY):

A study has been conducted on 287330 employed people who are from major category Agriculture & Natural Resources department and they have 5 different majors.

Test the claim that the proportion of employed people is equal for the 5 majors under the major category.

Sample data:

Table 3: Sample Data Chi Square

Major	GENERAL AGRICULTURE	AGRICULTURE PRODUCTION AND MANAGEMENT	AGRICULTURAL ECONOMICS	ANIMAL SCIENCES	FOOD SCIENCE
Employed	90245	76865	26321	81177	12722

Information:

- Significance level, $\boldsymbol{\alpha = 0.05}$
- Hypothesis Statement:

$$\boldsymbol{H_0}: \text{p1} = \text{p2} = \text{p3} = \text{p4} = \text{p5}$$

$\boldsymbol{H_1}$: At least 1 of the 5 proportions is different from others

Observed Value = O

Expected Value = E = $\frac{N}{K}$, where N= 287330, K= 5

Test Statistic: $X^2 = \frac{\sum(O-E)^2}{E}$

R-Program Output :

```
        Chi-squared test for given probabilities

data:  Employed
X-squared = 86748, df = 4, p-value < 2.2e-16
```

Conclusion:

The value of the test statistic that is $\boldsymbol{X^2}$, we get 86748. From the table with significance level, α = 0.05 we get the **critical value** $\boldsymbol{X^2_{4,0.05} = 9.48}$ Thus**,** as the test statistic is greater than the critical value, that is $\boldsymbol{X^2 > X^2_{4,0.05}}$ we **reject the null hypothesis**. That is, we reject the claim that the proportion of employed people is equal for the 5 categories under the major.

ANALYSIS OF VARIANCE (ANOVA)

The purpose of this test is to determine whether or not the mean number of employed from three random and independent major categories are similar or different to one another.

We will use a 0.05 significance level to test either they have same mean or at least one mean is different.

Table 4:Employed of three Departments

Employed		
Education	Health	Business
843693	78198	1580978
3113	49393	1335825
1492	85360	7846
819393	51279	2354398
47203	121479	47341
193542	19009	57983
113460	1325711	890125
36224	124058	670681
129486	199174	142879
108272	42543	66453

Solution:

$$H_0: \quad \mu_1=\mu_2=\mu_3$$
$$H_1: at\ least\ one\ mean, \mu\ is\ different$$

Find n, $\bar{x}$, s for each test subject:

We have n=10 and k=3.

Table 5:n, mean and s

	Employed		
	Education,A	Health,B	Business,C
n	10	10	10
$\bar{x}$	217934	209620.4	715450.9
s	328979.03	395609.52	815205.53

Find $\bar{\bar{x}}$, $s_{\bar{x}}$, $ns_{\bar{x}}^2$ between samples:

Table 6:value of mean, s.d, variance

$\bar{\bar{x}}$	381001.76
$S_{\bar{x}}$	289671.27
$ns_{\bar{x}}^2$	839094460768.034

Find variance within sample, S

Table 7:value of variance within samples

S_p^2	**309764722629.603**

Calculate test statistic, *F*:

Test statistic, F is calculated using R Studio and the formula is as below:

$$F = \frac{variance\ between\ sample\ (nS_{\bar{x}}^2)}{variance\ within\ sample\ (S_p^2)}$$

H	2.70881220316158

Figure 6:value of test statistic

Calculate Critical Value:

- Find numerator and denominator

  Numerator = k-1 = 2 Denominator = k*(n-1) = 27

  Find F critical value:

```
> #Count F Critical Value
> qf(.95,df1=num,df2=Den)
[1] 3.354131
```

Figure 7: F critical value = 3.354

Conclusion:

Since $F_{\text{test statistic}} < F_{\text{critical value}}$ (2.7088 < 3.354), the test statistic does not fall within the critical region, therefore we fail to reject hypothesis null. There is sufficient evidence to claim that the employed from three different department have the same mean.

CONCLUSION

There are many valuable experiences that we got from this project that can never be learn in the theoretically class only. For example, our skill in using R Studio to do inferential analysis improve massively while working on this project. Our understanding about inferential analysis become more excellent because we need to apply it to the data set that we choose by our own self. Any deficiency in this project mainly came from us, because our knowledge about R Studio is limited as this program rarely used by us in daily life. Therefore, we hope in the future, we can use R Studio program more skilfully.

REFERENCE

FiveThirtyEight. (2019). *FiveThirtyEight College Majors Dataset*. Kaggle.com. https://www.kaggle.com/fivethirtyeight/fivethirtyeight-college-majors-dataset

Hypothesis Test Graph Generator. (2021). Imathas.com. http://www.imathas.com/stattools/norm.html

Video Link : https://youtu.be/DymkscsWAKw

APPENDIX

Original/raw dataset:

Major_code	Major	Major_category	Total	Employed	Employed_f	Unemploye	Unemployn	Median	P25th	P75th
1100	GENERAL AGRICULTURE	Agriculture & Natural Resources	128148	90245	74078	2423	0.0261471	50000	34000	80000
1101	AGRICULTURE PRODUCTION AND MANAGEMENT	Agriculture & Natural Resources	95326	76865	64240	2266	0.0286361	54000	36000	80000
1102	AGRICULTURAL ECONOMICS	Agriculture & Natural Resources	33955	26321	22810	821	0.0302483	63000	40000	98000
1103	ANIMAL SCIENCES	Agriculture & Natural Resources	103549	81177	64937	3619	0.0426789	46000	30000	72000
1104	FOOD SCIENCE	Agriculture & Natural Resources	24280	17281	12722	894	0.0491884	62000	38500	90000
1105	PLANT SCIENCE AND AGRONOMY	Agriculture & Natural Resources	79409	63043	51077	2070	0.0317909	50000	35000	75000
1106	SOIL SCIENCE	Agriculture & Natural Resources	6586	4926	4042	264	0.0508671	63000	39400	88000
1199	MISCELLANEOUS AGRICULTURE	Agriculture & Natural Resources	8549	6392	5074	261	0.0392304	52000	35000	75000
1301	ENVIRONMENTAL SCIENCE	Biology & Life Science	106106	87602	65238	4736	0.0512898	52000	38000	75000
1302	FORESTRY	Agriculture & Natural Resources	69447	48228	39613	2144	0.0425633	58000	40500	80000
1303	NATURAL RESOURCES MANAGEMENT	Agriculture & Natural Resources	83188	65937	50595	3789	0.0543413	52000	37100	75000
1401	ARCHITECTURE	Engineering	294692	216770	163020	20394	0.0859911	63000	40400	93500
1501	AREA ETHNIC AND CIVILIZATION STUDIES	Humanities & Liberal Arts	103740	75798	50530	5525	0.067939	46000	32000	71000
1901	COMMUNICATIONS	Communications & Journalism	987676	790696	595739	54390	0.0643603	50000	35000	80000
1902	JOURNALISM	Communications & Journalism	418104	314438	235407	20754	0.0619168	50000	35000	80000
1903	MASS MEDIA	Communications & Journalism	211213	170474	125489	15431	0.0830048	48000	32000	70000
1904	ADVERTISING AND PUBLIC RELATIONS	Communications & Journalism	186829	147433	111552	10624	0.0672163	50000	34000	75000
2001	COMMUNICATION TECHNOLOGIES	Computers & Mathematics	62141	49609	37261	4609	0.0850087	50000	34500	75000
2100	COMPUTER AND INFORMATION SYSTEMS	Computers & Mathematics	253782	218248	189950	11945	0.0518912	65000	45000	90000
2101	COMPUTER PROGRAMMING AND DATA PROCESSING	Computers & Mathematics	29317	22828	18747	2265	0.0902642	60000	40000	85000
2102	COMPUTER SCIENCE	Computers & Mathematics	783292	656372	561052	34196	0.0495187	78000	51000	105000
2105	INFORMATION SCIENCES	Computers & Mathematics	77805	66393	57604	3704	0.0528411	68000	46200	95000
2106	COMPUTER ADMINISTRATION MANAGEMENT AND SECURITY	Computers & Mathematics	39362	32366	28156	2626	0.0750457	55000	40000	80000
2107	COMPUTER NETWORKING AND TELECOMMUNICATIONS	Computers & Mathematics	51771	44071	35954	2748	0.0586941	55000	36000	80000
2201	COSMETOLOGY SERVICES AND CULINARY ARTS	Industrial Arts & Consumer Services	42325	33388	25780	1941	0.0549407	40000	26200	60000
2300	GENERAL EDUCATION	Education	1438867	843693	591863	38742	0.0439035	43000	32000	59000
2301	EDUCATIONAL ADMINISTRATION AND SUPERVISION	Education	4037	3113	2468	0	0	58000	44750	79000
2303	SCHOOL STUDENT COUNSELING	Education	2396	1492	1093	169	0.1017459	41000	33200	50000
2304	ELEMENTARY EDUCATION	Education	1446701	819393	501786	32685	0.0383592	40000	31000	50000
2305	MATHEMATICS TEACHER EDUCATION	Education	68808	47203	29494	1610	0.032983	43000	34000	60000
2306	PHYSICAL AND HEALTH EDUCATION TEACHING	Education	281661	193542	136343	9389	0.046267	48400	34000	66500
2307	EARLY CHILDHOOD EDUCATION	Education	157079	113460	71133	5890	0.0493506	35300	27000	45800
2308	SCIENCE AND COMPUTER TEACHER EDUCATION	Education	56477	36224	24817	1596	0.0421999	46000	35000	61000
2309	SECONDARY TEACHER EDUCATION	Education	224262	129486	88917	5925	0.0437557	45000	34000	60000
2310	SPECIAL NEEDS EDUCATION	Education	149689	108272	71615	5357	0.0471447	42000	34000	53000
2311	SOCIAL SCIENCE OR HISTORY TEACHER EDUCATION	Education	127022	78785	51632	3800	0.0460132	45000	33000	64000
2312	TEACHER EDUCATION: MULTIPLE LEVELS	Education	88067	58885	37892	2032	0.0333569	40000	30000	51000
2313	LANGUAGE AND DRAMA EDUCATION	Education	181445	111347	67651	5624	0.0480803	42000	32000	54000
2314	ART AND MUSIC EDUCATION	Education	231861	155159	94756	6629	0.0409734	42600	32000	56000
2399	MISCELLANEOUS EDUCATION	Education	225553	126054	91322	5145	0.0392152	50000	35600	71000
2400	GENERAL ENGINEERING	Engineering	503080	359172	312023	17986	0.0476882	75000	50000	1.00E+05
2401	AEROSPACE ENGINEERING	Engineering	65734	44944	38491	1969	0.0419713	80000	58000	110000
2402	BIOLOGICAL ENGINEERING	Engineering	32748	24270	18621	1521	0.0589741	62000	40000	91000
2403	ARCHITECTURAL ENGINEERING	Engineering	19587	13713	11180	1017	0.0690428	78000	50000	102000
2404	BIOMEDICAL ENGINEERING	Engineering	18347	12876	9202	1105	0.0790358	65000	40000	96000
2405	CHEMICAL ENGINEERING	Engineering	188046	131697	109406	6388	0.0462614	86000	60000	120000
2406	CIVIL ENGINEERING	Engineering	358593	262831	220528	14823	0.0533866	78000	55000	105000
2407	COMPUTER ENGINEERING	Engineering	154160	128742	111025	7456	0.0547438	80000	60000	107000
2408	ELECTRICAL ENGINEERING	Engineering	671647	489965	422317	26064	0.0505088	88000	60000	116000
2409	ENGINEERING MECHANICS PHYSICS AND SCIENCE	Engineering	20582	14909	12257	683	0.0438045	65000	45000	1.00E+05
2410	ENVIRONMENTAL ENGINEERING	Engineering	13016	9849	8104	472	0.045732	70000	50000	95000
2411	GEOLOGICAL AND GEOPHYSICAL ENGINEERING	Engineering	6264	4120	3350	0	0	85000	55000	125000
2412	INDUSTRIAL AND MANUFACTURING ENGINEERING	Engineering	138366	101273	85014	5498	0.0514934	75000	50000	101000
2413	MATERIALS ENGINEERING AND MATERIALS SCIENCE	Engineering	21430	14687	11871	933	0.0597311	78000	55000	105000
2414	MECHANICAL ENGINEERING	Engineering	581529	422207	362053	19360	0.0438439	80000	59000	110000
2415	METALLURGICAL ENGINEERING	Engineering	12818	6939	5462	326	0.0448727	96000	65000	123000
2416	MINING AND MINERAL ENGINEERING	Engineering	10746	7416	6419	366	0.0470316	92000	52000	124000
2417	NAVAL ARCHITECTURE AND MARINE ENGINEERING	Engineering	16094	10690	9226	449	0.0403088	97000	60000	125000
2418	NUCLEAR ENGINEERING	Engineering	9826	7320	6474	527	0.0671594	95000	65000	128000
2419	PETROLEUM ENGINEERING	Engineering	19631	14002	11636	617	0.0422053	125000	75000	210000
2499	MISCELLANEOUS ENGINEERING	Engineering	57006	43906	37194	2744	0.058821	70000	50000	1.00E+05
2500	ENGINEERING TECHNOLOGIES	Engineering	37382	30102	25651	1475	0.0467112	63000	40000	93000
2501	ENGINEERING AND INDUSTRIAL MANAGEMENT	Engineering	47098	27275	22104	1577	0.0546583	74000	50000	107000
2502	ELECTRICAL ENGINEERING TECHNOLOGY	Engineering	94697	73737	64157	4572	0.0583841	67000	46900	91000
2503	INDUSTRIAL PRODUCTION TECHNOLOGIES	Engineering	82142	65401	57266	3431	0.049846	70000	48000	98000
2504	MECHANICAL ENGINEERING RELATED TECHNOLOGIES	Engineering	29348	24190	21273	1101	0.0435333	60000	40000	82000
2599	MISCELLANEOUS ENGINEERING TECHNOLOGIES	Engineering	64196	53097	46183	3401	0.0601968	63000	42000	90000
2601	LINGUISTICS AND COMPARATIVE LANGUAGE AND LITERATURE	Humanities & Liberal Arts	75791	45657	30428	4043	0.0813481	48000	30000	70000
2602	FRENCH GERMAN LATIN AND OTHER COMMON FOREIGN LANGU	Humanities & Liberal Arts	236342	153654	99459	9598	0.0587925	48000	33500	69000
2603	OTHER FOREIGN LANGUAGES	Humanities & Liberal Arts	57793	34696	23046	2707	0.0723739	45000	30000	75000
2901	FAMILY AND CONSUMER SCIENCES	Industrial Arts & Consumer Services	402038	241585	154167	14273	0.0557848	40500	30000	60000
3201	COURT REPORTING	Law & Public Policy	9330	7270	6143	518	0.0665126	50000	34000	75000
3202	PRE-LAW AND LEGAL STUDIES	Law & Public Policy	67037	49259	37254	3699	0.0698478	48000	34000	70000
3301	ENGLISH LANGUAGE AND LITERATURE	Humanities & Liberal Arts	1098647	708882	482229	52248	0.0686453	50000	32900	75000
3302	COMPOSITION AND RHETORIC	Humanities & Liberal Arts	59211	44913	29628	3569	0.073615	40000	28800	65000
3401	LIBERAL ARTS	Humanities & Liberal Arts	601221	404932	296792	29348	0.0675785	50000	33000	75000
3402	HUMANITIES	Humanities & Liberal Arts	46188	29971	19460	2530	0.0778438	46700	30000	70000
3501	LIBRARY SCIENCE	Education	16193	7091	4330	743	0.094843	40000	30000	55000
3600	BIOLOGY	Biology & Life Science	839454	583079	422788	36757	0.0593012	51000	35000	80000
3601	BIOCHEMICAL SCIENCES	Biology & Life Science	75322	52594	37103	4056	0.0715975	53000	33000	82000
3602	BOTANY	Biology & Life Science	14135	9284	6333	327	0.0340235	50000	32000	75000
3603	MOLECULAR BIOLOGY	Biology & Life Science	28197	20221	13366	1303	0.0605371	45000	30000	70000
3604	ECOLOGY	Biology & Life Science	45368	36708	25677	1888	0.048917	47500	32000	73000
3605	GENETICS	Biology & Life Science	6362	4747	3498	206	0.041591	48000	33000	80000
3606	MICROBIOLOGY	Biology & Life Science	68885	45422	33990	2435	0.0508807	60000	40000	85000
3607	PHARMACOLOGY	Biology & Life Science	5015	3481	2579	57	0.0161108	60000	35000	105000
3608	PHYSIOLOGY	Biology & Life Science	43984	31394	20207	1692	0.0511395	50000	30000	75000
3609	ZOOLOGY	Biology & Life Science	55395	35714	26152	1815	0.0483626	55000	34000	85000
3611	NEUROSCIENCE	Biology & Life Science	13676	8987	5446	665	0.0688976	35000	28000	52000
3699	MISCELLANEOUS BIOLOGY	Biology & Life Science	29389	22298	16508	1114	0.0475824	52000	33500	72800
3700	MATHEMATICS	Computers & Mathematics	432806	280902	209838	15701	0.0529361	66000	42000	1.00E+05
3701	APPLIED MATHEMATICS	Computers & Mathematics	19112	15136	12109	892	0.0556526	70000	47000	106000
3702	STATISTICS AND DECISION SCIENCE	Computers & Mathematics	24806	18808	14468	1138	0.057054	70000	43000	102000
3801	MILITARY TECHNOLOGIES	Industrial Arts & Consumer Services	4315	1650	1708	187	0.1017964	64000	39750	90000
4000	MULTI/INTERDISCIPLINARY STUDIES	Interdisciplinary	45199	35706	26038	2990	0.077269	43000	32000	55000
4001	INTERCULTURAL AND INTERNATIONAL STUDIES	Humanities & Liberal Arts	56580	43114	27499	3462	0.0743301	45000	30000	69000
4002	NUTRITION SCIENCES	Health	64534	43878	28112	2961	0.0632166	49500	34000	69000
4005	MATHEMATICS AND COMPUTER SCIENCE	Computers & Mathematics	7184	5874	5039	150	0.0249004	92000	53000	136000
4006	COGNITIVE SCIENCE AND BIOPSYCHOLOGY	Biology & Life Science	6898	5527	3639	284	0.0488728	53000	31500	93000
4007	INTERDISCIPLINARY SOCIAL SCIENCES	Social Science	61871	43312	31078	3030	0.0653835	45000	33000	67000
4101	PHYSICAL FITNESS PARKS RECREATION AND LEISURE	Industrial Arts & Consumer Services	350409	286683	204242	14108	0.046903	44000	30000	60000
4801	PHILOSOPHY AND RELIGIOUS STUDIES	Humanities & Liberal Arts	205763	138734	95429	11252	0.0750203	45000	30000	70000
4901	THEOLOGY AND RELIGIOUS VOCATIONS	Humanities & Liberal Arts	232865	164827	121606	7317	0.0425051	40000	27000	56000
5000	PHYSICAL SCIENCES	Physical Sciences	8856	5872	4961	238	0.0389525	60000	38000	89000
5001	ASTRONOMY AND ASTROPHYSICS	Physical Sciences	4700	3400	2447	320	0.0860215	80000	40000	106000
5002	ATMOSPHERIC SCIENCES AND METEOROLOGY	Physical Sciences	14051	11252	9098	429	0.0367263	60000	39000	90000
5003	CHEMISTRY	Physical Sciences	308062	198075	153308	11468	0.0547286	59000	38700	90000
5004	GEOLOGY AND EARTH SCIENCE	Physical Sciences	107902	75698	59262	5272	0.0651105	65000	42000	1.00E+05
5005	GEOSCIENCES	Physical Sciences	8267	6129	5020	274	0.0427924	57000	36000	80000
5006	OCEANOGRAPHY	Physical Sciences	10741	7882	6508	449	0.0538951	55000	40000	90000
5007	PHYSICS	Physical Sciences	122620	80797	60812	4440	0.0520901	70000	40000	110000

118	5202	CLINICAL PSYCHOLOGY	Psychology & Social Work	7638	5128	3297	587	0.1027122	45000	26100	62000
119	5203	COUNSELING PSYCHOLOGY	Psychology & Social Work	17633	13071	8888	954	0.0680214	39000	25000	50000
120	5205	INDUSTRIAL AND ORGANIZATIONAL PSYCHOLOGY	Psychology & Social Work	17969	11878	8631	1084	0.0836291	62000	40000	84000
121	5206	SOCIAL PSYCHOLOGY	Psychology & Social Work	10871	6897	5226	660	0.0873362	47000	35000	70000
122	5299	MISCELLANEOUS PSYCHOLOGY	Psychology & Social Work	34102	23921	15688	2137	0.0820094	45000	30000	70000
123	5301	CRIMINAL JUSTICE AND FIRE PROTECTION	Law & Public Policy	757141	613369	517599	35037	0.0540356	50000	35000	74000
124	5401	PUBLIC ADMINISTRATION	Law & Public Policy	54636	37879	31262	2836	0.0696549	56000	40000	85000
125	5402	PUBLIC POLICY	Law & Public Policy	14782	11147	8196	959	0.0792169	60000	38000	90000
126	5403	HUMAN SERVICES AND COMMUNITY ORGANIZATION	Psychology & Social Work	81786	61402	44959	4794	0.0724213	38000	29000	52000
127	5404	SOCIAL WORK	Psychology & Social Work	319163	225081	164020	14208	0.0593759	40000	30000	53000
128	5500	GENERAL SOCIAL SCIENCES	Social Science	127363	80165	57036	6132	0.0710569	50000	35000	75000
129	5501	ECONOMICS	Social Science	757616	535446	430580	34974	0.0613127	69000	42000	110000
130	5502	ANTHROPOLOGY AND ARCHEOLOGY	Humanities & Liberal Arts	143087	102399	66046	8684	0.0781758	43000	30000	68000
131	5503	CRIMINOLOGY	Social Science	75085	59534	48763	4106	0.0645192	49000	35000	72000
132	5504	GEOGRAPHY	Social Science	115423	83671	63070	6202	0.0690085	54000	38300	78000
133	5505	INTERNATIONAL RELATIONS	Social Science	77371	56564	42091	4278	0.0703133	55000	40000	89000
134	5506	POLITICAL SCIENCE AND GOVERNMENT	Social Science	748956	541630	421761	40376	0.0693739	58000	38000	90000
135	5507	SOCIOLOGY	Social Science	674558	459174	336515	32344	0.0658043	47000	33000	70000
136	5599	MISCELLANEOUS SOCIAL SCIENCES	Social Science	15882	12307	9444	708	0.0543988	52000	40000	80000
137	5601	CONSTRUCTION SERVICES	Industrial Arts & Consumer Services	92346	79055	65916	4257	0.0510971	65000	45000	95000
138	5701	ELECTRICAL, MECHANICAL, AND PRECISION TECHNOLOGIES AND	Industrial Arts & Consumer Services	15726	12607	10453	692	0.052034	48000	32000	75000
139	5901	TRANSPORTATION SCIENCES AND TECHNOLOGIES	Industrial Arts & Consumer Services	126639	98814	83519	4902	0.0472637	67000	42500	98000
140	6000	FINE ARTS	Arts	571961	386961	256747	29912	0.0717533	45000	30000	70000
141	6001	DRAMA AND THEATER ARTS	Arts	174817	135071	81519	11789	0.0802737	42000	29000	62000
142	6002	MUSIC	Arts	276262	192704	116142	11155	0.0547192	45000	30000	67000
143	6003	VISUAL AND PERFORMING ARTS	Arts	55141	41098	23479	4297	0.094658	40000	27000	59000
144	6004	COMMERCIAL ART AND GRAPHIC DESIGN	Arts	504657	379980	266671	30330	0.0739197	46600	32000	70000
145	6005	FILM VIDEO AND PHOTOGRAPHIC ARTS	Arts	133508	107651	69303	10080	0.0856189	47000	30000	70000
146	6006	ART HISTORY AND CRITICISM	Humanities & Liberal Arts	90852	61295	38989	4185	0.0639126	44500	30000	70000
147	6007	STUDIO ARTS	Arts	81008	58799	36943	5372	0.0837138	37600	24900	58000
148	6099	MISCELLANEOUS FINE ARTS	Arts	8511	6431	3802	1190	0.1561475	45000	30000	60000
149	6100	GENERAL MEDICAL AND HEALTH SERVICES	Health	104516	78198	53746	4525	0.0547006	50000	35000	73000
150	6102	COMMUNICATION DISORDERS SCIENCES AND SERVICES	Health	74977	49393	26085	2407	0.0464672	42000	30000	60000
151	6103	HEALTH AND MEDICAL ADMINISTRATIVE SERVICES	Health	108510	85360	67294	5160	0.057004	50000	35000	75000
152	6104	MEDICAL ASSISTING SERVICES	Health	64316	51279	25118	1660	0.0313568	55000	37000	75000
153	6105	MEDICAL TECHNOLOGIES TECHNICIANS	Health	164990	121479	92128	4564	0.0362099	60000	45000	76000
154	6106	HEALTH AND MEDICAL PREPARATORY PROGRAMS	Health	32514	19009	13147	1431	0.0700098	50000	34000	85000
155	6107	NURSING	Health	1769892	1325711	947546	36503	0.0267968	62000	48000	80000
156	6108	PHARMACY PHARMACEUTICAL SCIENCES AND ADMINISTRATION	Health	180084	124058	89234	4414	0.0343577	106000	78000	125000
157	6109	TREATMENT THERAPY PROFESSIONS	Health	252138	199174	128115	5378	0.0262916	61000	40000	80000
158	6110	COMMUNITY AND PUBLIC HEALTH	Health	56741	42543	28912	3032	0.0665277	47000	32800	70000
159	6199	MISCELLANEOUS HEALTH MEDICAL PROFESSIONS	Health	77647	52610	35676	2978	0.0535727	45000	32000	62000
160	6200	GENERAL BUSINESS	Business	2148712	1580978	1304646	85626	0.0513775	60000	40000	95000
161	6201	ACCOUNTING	Business	1779219	1335825	1095027	75379	0.0534147	65000	42500	1.00E+05
162	6202	ACTUARIAL SCIENCE	Business	9763	7846	6880	466	0.0560635	72000	53000	115000
163	6203	BUSINESS MANAGEMENT AND ADMINISTRATION	Business	3123510	2354398	1939384	147261	0.0588653	58000	39500	86000
164	6204	OPERATIONS LOGISTICS AND E-COMMERCE	Business	57200	47341	41104	2141	0.0432683	65000	45000	90000
165	6205	BUSINESS ECONOMICS	Business	75547	57983	48471	3816	0.0617486	65000	45000	1.00E+05
166	6206	MARKETING AND MARKETING RESEARCH	Business	1114624	890125	704912	51839	0.0550329	56000	38500	90000
167	6207	FINANCE	Business	816548	670681	561073	34166	0.0484729	65000	45000	1.00E+05
168	6209	HUMAN RESOURCES AND PERSONNEL MANAGEMENT	Business	187274	142879	116466	9241	0.0607481	54000	38000	80000
169	6210	INTERNATIONAL BUSINESS	Business	86064	66453	51012	5106	0.0713537	54000	38600	80000
170	6211	HOSPITALITY MANAGEMENT	Business	200854	163393	122499	8862	0.051447	49000	33000	70000
171	6212	MANAGEMENT INFORMATION SYSTEMS AND STATISTICS	Business	156673	134478	118249	6186	0.0439771	72000	50000	1.00E+05
172	6299	MISCELLANEOUS BUSINESS & MEDICAL ADMINISTRATION	Business	102753	77471	61603	4308	0.0526786	53000	36000	83000
173	6402	HISTORY	Humanities & Liberal Arts	712509	478416	354163	33725	0.065851	50000	35000	80000
174	6403	UNITED STATES HISTORY	Humanities & Liberal Arts	17746	11887	8204	943	0.0734996	50000	39000	81000

Processed dataset:

- Hypothesis Testing on 2 Sample

	Architecture	Computer Science
Employed, x	$x_1 = 216770$	$x_2 = 656372$
Sample, n	$n_1 = 294692$	$n_2 = 783292$

- Correlation

Major Code	**Employed Full Time, x**	**Unemployed, y**
MATHEMATICS AND COMPUTER SCIENCE	5039	150
APPLIED MATHEMATICS	12109	892
STATISTICS AND DECISION SCIENCE	14468	1138
COMPUTER PROGRAMMING AND DATA PROCESSING	18747	2265
COMPUTER ADMINISTRATION MANAGEMENT AND SECURITY	28156	2626
COMPUTER NETWORKING AND TELECOMMUNICATIONS	35954	2748
COMMUNICATION TECHNOLOGIES	37261	4609
INFORMATION SCIENCES	57604	3704

- Chi Square Test (One way contingency)

Major	GENERAL AGRICULTURE	AGRICULTURE PRODUCTION AND MANAGEMENT	AGRICULTURAL ECONOMICS	ANIMAL SCIENCES	FOOD SCIENCE
Employed	90245	76865	26321	81177	12722

-

- ANOVA

Employed		
Education	Health	Business
843693	78198	1580978
3113	49393	1335825
1492	85360	7846
819393	51279	2354398
47203	121479	47341
193542	19009	57983
113460	1325711	890125
36224	124058	670681
129486	199174	142879
108272	42543	66453