

Department of Computer Science, University of Karachi  
**BSCS: 306: Probability and Statistical Methods**

Course Tutor: Dr Tahseen A. Jilani

**Assignment-10: Hypothesis Testing - III**

**Test for Independence, Test for homogeneity and test for multiple proportions**

- Q.1. A criminologist conducted a survey to determine whether the incidence of certain types of crime varied from one part of a large city to another. The particular crimes of interest were assault, burglary, larceny, and homicide. The following table shows the numbers of crimes committed in four areas of the city during the past year. Type

District	Type of Crime			
	Assault	Burglary	Larceny	Homicide
1	162	118	451	18
2	310	196	996	25
3	258	193	458	10
4	280	175	390	19

Can we conclude from these data at the 0.01 level of significance that the occurrence of these types of crime is dependent on the city district?

- Q.2. A college infirmary conducted an experiment to determine the degree of relief provided by three cough remedies. Each cough remedy was tried on 50 students and the following data recorded:

	Cough Remedy		
	NyQuil	Robitussin	Triaminic
No relief	11	13	9
Some relief	32	28	27
Total relief	7	9	14

Test the hypothesis that the three cough remedies are equally effective. Use 1% level of significance.

- Q.3. The following responses concerning the standard of living at the time of an independent opinion poll of 1000 households versus one year earlier seems to be in agreement with the results of a study published in Across the Board (June 1981):

Period	Standard of Living			Total
	Somewhat	Not as		
	Better	Same	Good	
1980: Jan.	72	144	84	300
May	63	135	102	300
Sept.	47	100	53	200
1981: Jan.	40	105	55	200

Test, the hypothesis that the proportions of households within each standard of living category are the same for each of the four time periods. Use 5% level of significance.

- Q.4. The following responses concerning the standard of living at the time of an independent opinion poll of 1000 households versus one year earlier seems to be in agreement with the results of a study published in Across the Board (June 1981):

Period	Standard of Living			Total
	Somewhat	Same	Not as	
	Better		Good	
1980: Jan.	72	144	84	300
May	63	135	102	300
Sept.	47	100	53	200
1981: Jan.	40	105	55	200

Test the hypothesis that the proportions of households within each standard of living category are the same for each of the four time periods. Use 5% level of significance.

- Q.5. A random sample of 175 Cal Poly State University students was selected, and both the email service provider and cell phone provider were determined for each one, resulting in the accompanying data. State and test the appropriate hypotheses.

		Cell Phone Provider		
		ATT	Verizon	Other
Email Provider	gmail	28	17	7
	Yahoo	31	26	10
	Other	26	19	11

- Q.6. Each individual in a random sample of high school and college students was cross-classified with respect to both political views and marijuana usage, resulting in the data displayed in the accompanying two-way table (“Attitudes About Marijuana and Political Views,” Psychological Reports, 1973: 1051–1054). Does the data support the hypothesis that political views and marijuana usage level are independent within the population? Test the appropriate hypotheses using level of significance .01.

		Usage Level		
		Never	Rarely	Frequently
Political Views	Liberal	479	173	119
	Conservative	214	47	15
	Other	172	45	85

- Q.7. A random sample of smokers was obtained and each individual was classifier both with respect to gender and with respect to the age at which he/she first started smoking. The data in the accompanying table is consistent with summary results reported in the article “Cigarette Tar Yields in Relation to Mortality in the Cancer Prevention Study II Prospective Cohort” (British Med. J. 2004: 72-79). Calculate the proportion of males in each age category and then do the same for females. (a). Based on the proportions, does it appear that there might be an association between gender and the age at which an individual first smokes? (b) Carry out a test of hypotheses to decide whether there is an association between the two factors.

		Gender	
		Male	Female
Age	<16	25	10
	16 – 17	24	32
	18 – 20	28	17
	>20	19	34

## Goodness of fit test

- Q.1. A die is tossed 180 times with the following results:

x	1	2	3	4	5	6
f	28	36	36	30	27	23

Is this a balance die? Use a 0.01 level of significance.

- Q.2. A coin is thrown until a head occurs and the number X of tosses recorded. After repeating the experiment 256 times, we obtained the following results:

<b>x</b>	1	2	3	4	5	6	7	8
<b>f</b>	136	60	34	12	9	1	3	1

Test the hypothesis at the 0.05 level of significance that the observed distribution of X may be fitted by the geometric distribution  $g(x; 1/2)$ ,  $x=1,2,3,\dots$

**Q.3.** In 100 tosses of a coin, 63 heads and 37 tails are observed. Is this a balanced coin? Use a 0.05 level of significance?

**Q.4.** A machine is supposed to mix peanuts, hazelnuts, cashews and pecans in the ratio 5:2:2:1. A can containing 500 of these mixed nuts was found to have 269 peanuts, 112 hazelnuts, 74 cashews and 45 pecans. At the 0.05 level of significance, test the hypothesis that the machine is mixing the nuts in the ratio 5:2:2:1.

**Q.5.** The grades in a statistics course for a particular semester were as follows:

<b>Category</b>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>F</b>
<b>Grade</b>	14	18	32	20	16

Test the hypotheses, at 5% level of significance, that the distribution of the grades is uniform.

**Q.6. (Test for normal distribution)** Apply the chi-square goodness of fit test between the observed class frequencies and the corresponding expected frequencies of a normal distribution with mean = 65 and S.D. = 21, using 5% level of significance.

23	60	79	32	57	74	52	70	82
36	80	77	81	95	41	65	92	85
55	76	52	10	64	75	78	25	80
98	81	67	41	71	83	54	64	72
88	62	74	43	60	78	89	76	84
48	84	90	15	79	34	67	17	82
69	74	63	80	85	61			

**Q.7. (Test for normal distribution)** Apply the chi-square goodness of fit test between the observed class frequencies and the corresponding expected frequencies of a normal distribution with mean = 65 and S.D. = 0.4, using 1% level of significance.

2.0	3.0	0.3	3.3	1.3	0.4
0.2	6.0	5.5	6.5	0.2	2.3
1.5	4.0	5.9	1.8	4.7	0.7
4.5	0.3	1.5	0.5	2.5	5.0
1.0	6.0	5.6	6.0	1.2	0.2