

Exercises on Analysis of Categorical Data

- 1) To determine the possible effect of a chemical treatment on the status of seed germination, a random sample of 100 chemically treated seeds, and a random sample of 150 untreated seeds of the same type are sown. The number of seeds that germinate and do not germinate are recorded below:

	Germinated	Not Germinated	Total
Treated	84	16	100
Untreated	132	18	150
Total	216	18	250

- Using the relevant tables and graphs comment on the similarities or differences between the germination status of the chemically treated and untreated seeds.
 - Using the chi-square test, determine if the data provide sufficient evidence that the germination status is different for the chemically treated and untreated seeds. Use the 5% level of significance
 - Comment if any of the cells contribute more to the outcome of the test in Part (b) than others.
 - Explain whether or not the requirement for the use of the chi-square test is met. If this requirement is not met, which alternative procedure should be used?
- 2) A survey on household technology was conducted in five countries in the Asia Pacific region and it was based on a representative sample of 500 respondents from each country. The number of respondents whose household members own more than device, viz., desktop computers, laptops, notebooks or tablets are as follows:

	Australia	New Zealand	China	Japan	South Korea	Total
Own more than one device	265	240	190	270	245	1210
Do not own more than one device	235	260	310	230	255	1290
Total	500	500	500	500	500	2500

- Using the relevant tables and graphs comment on the similarities or differences between countries regarding ownership of devices.
- Explain how the expected frequency and standardised residual of Japanese respondents who do not own more than one device is obtained.
- At the 5% level of significance, determine if there is a difference in the proportion of households in the five countries who own more than one device.
- Comment if any of the cells contribute more to the outcome of the test in Part (c) than others.

- 3) Managed funds provide the individual investor with a convenient medium for diversification. A managed fund spreads its investment over numerous securities, so that individuals who purchase a share in a managed fund are investing in a highly diversified portfolio, something they could not achieve on their own with limited funds. Managed funds tailor their diversification strategies to different investor groups, their capital-growth funds and so on. The table below shows the levels of return over five years for a random sample of investors for various categories of managed funds.

Managed fund type	Level of return			Total
	High	Medium	Low	
Maximum capital gain	25	41	52	118
Long-term growth	22	31	42	95
Growth and current income	33	41	53	127
Balanced income	35	39	42	116
Common stock	28	15	10	53
Total	143	167	199	509

- Using the relevant tables and graphs comment on level of returns from the different type of funds
- Test at the 5% level of significance if there is evidence of a relationship between level of return and type of managed fund.
- Comment if any of the cells contribute more to the outcome of the test in Part (c) than others.
- What does this fund say about managed fund investments?

- 4) Suppose that the sales records of a random sample 50 salespersons in a large retail business were scrutinised to determine over a 6-month period last year as to whether or not their targets were achieved. At the end of last year, they were subjected to a week-long train exercise and their sales targets for the first three months of this year scrutinised. The results are given in the table below:

		Without Training		
		Targets not reached	Targets reached	Total
With Training	Targets not reached	10	5	15
	Targets reached	15	20	35
Total		25	25	50

- a) Using the relevant tables and graphs comment on the data.
 - b) Using the McNemar procedure, test at the 5% level of significance that the proportion of those salespersons whose targets were not reached is the same before and after the training program.
 - c) Using the Binomial exact test procedure, test at the 5% level of significance that the proportion of those salespersons whose targets were not reached is the same before and after the training program.
 - d) Comment on the outcome of the test in Part (b) and (c) explain which of the two tests is possibly more appropriate.
- 5) A manufacturer wishes to compare the effectiveness of four methods of treating raw fabric to render it water-repellent. Six types of fabric were used in the operation. A specimen of each type was divided into quarters and each quarter was randomly assigned to one of the four methods. After treatment, each piece was tested whether or not it was water repellent and scored 0 if unsatisfactory and 1 if satisfactory. The results are shown below and, in the file, *water.csv*.

	Method			
Fabric	A	B	C	D
1	1	1	0	0
2	1	0	0	1
3	1	1	0	0
4	1	1	1	0
5	1	1	0	1
6	1	1	0	1

- a) Using the relevant tables and graphs comment on the data.
- b) Using the Cochran Q test, can it be concluded that these data provide sufficient evidence to indicate a difference among the methods. Use the 5% level of significance.