Exercises Sheet 6

Classical Inference: Many-sample tests

Note: Some of these questions may be easier using SPSS (or SAS) or equivalent

1. The following is a record of bacterial counts on old samples, collected over an extended period. No. of positive results /day and frequency of obtaining this no. over a given period are recorded. On the basis of these data, would it be reasonable to assume that the No. of positive counts is a Poisson random variable?
   (a) Test at the 0.05 level of significance
   (b) Test at the 0.01 level of significance
   [**Note:** If you are told that test requirements are usually that each expected frequency should be no less than 5 for chi-squared viable, re-write the distribution below before performing any test]

No./day	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
No. of days	151	118	77	19	0	0	0	0

2. In order to assess measurement accuracy on new laboratory equipment, 100 measurements in g. made on a standard item were treated as a random sample and the following distribution was obtained. The view was that sensitivity was poor and measurements within the given range were equally likely. Would you support that view on the evidence available?
   (a) at the 0.01 level of significance?
   (b) at the 0.90 level of significance?
   (c) What can you say about the nature of statistical testing from your results on parts (a) and (b)?

Weight in g.	**No. in range**
15.5-15.6	5
15.6-15.7	9
15.7-15.8	7
15.8-15.9	10
15.9-16.0	12
16.0-16.1	8
16.1-16.2	13
16.2-16.3	15
16.3-16.4	13
16.4-16.5	8

3. Two laboratories were inspected for quality performance and rated over a random sample of records on 1000 tasks. Ratings were as follows.

		Rating		
Lab	**Poor**	**Average**	**Good**	**Total**
A	98	256	188	542
B	43	202	213	458
Total	141	458	401	1000

(a) Formulate and test a suitable hypothesis at the 0.025 level of Significance
(b) Given that, for the test to be valid, ideally none but certainly no more than 20% of expected frequencies must be less than 5, is this condition a problem here?
(c) If expected frequencies were lower than 5 in column 1, row 2 and column 2, row 1, how might you sensibly *collapse* the table and still test the principal hypothesis of interest? How would your test be affected? You can illustrate with the frequencies given, if helpful.

4. Medical researchers experiment to determine if four brands of headache remedies provide different levels of relief for patients with chronic headaches. Twelve subjects are randomly assigned to each of the four factor levels and controlled also for use of blood pressure medication, drug B. The data are as follows.

Drug B		**Head. Tablet**	**Brand**	
Usage	**1**	**2**	**3**	**4**
	8.68	6.23	4.92	7.43
None	8.14	6.73	5.21	6.76
	11.57	4.31	6.34	4.63
	8.98	5.80	5.00	5.87
In last 30 days	8.62	8.25	6.35	7.38
	3.35	7.88	8.41	5.58
	4.95	5.65	6.42	7.37
	5.16	6.47	4.70	7.28
Currently	7.89	3.78	7.72	5.46
	5.62	1.27	5.16	6.44
	7.08	0.83	2.60	6.02
	6.91	5.85	5.82	8.64

Use SPSS (or equivalent) to perform a basic analysis of these data. Formulate, test and report on appropriate hypotheses and results.

5. To assess whether students' exam. performances vary with the advance description of an exam., a lecturer uses different descriptors in different classes and no descriptor for one class. He also feels that interpretation of descriptors may be gender-related. Students take a common exam. Simultaneously. The data on exam. marks are as follows.

	Descriptor Used		
Gender	**Easy**	**Difficult**	**None**
Female	90	92	90
	85	76	85
	87	98	95
	85	65	75
Male	70	78	76
	72	70	70
	78	55	60
	65	50	62
	58	40	68

(a) Initially ignoring the gender effect, use SPSS to analyse these data. Formulate and test suitable hypotheses. What assumptions would you expect to apply?
(b) Omit the last row of the table (for the sake of illustration only) and analyse these data as a two-factor ANOVA, again formulating and testing suitable hypotheses.
(c) Indicate how you would analyse the complete data, using a regression (linear model) type approach. What are the independent variables here?

6. Yields of a particular substance are dependent on four properties of the original material, labelled X_1, X_2, X_3, X_4 here for convenience. It is required to use these data to provide a model equation for predicting yield (Y). Assume the observations Y_i to be independent and Normally distributed with constant variance. Use SPSS to perform the analysis and report briefly on your results.

X_1	X_2	X_3	X_4	Y
38.4	6.1	220	235	6.9
40.3	4.8	231	307	14.4
40.0	6.1	217	212	7.4
31.8	0.2	316	365	8.5
40.8	3.5	210	218	8.0
41.3	1.8	267	235	2.8
38.1	1.2	274	285	5.0
50.8	8.6	190	205	12.2
32.2	5.2	236	267	10.0
38.4	6.1	220	300	15.2
40.3	4.8	231	367	26.8
32.2	2.4	284	351	14.0
31.8	0.2	316	379	14.7
41.3	1.8	267	275	6.4
38.1	1.2	274	365	17.6
50.8	8.6	190	275	22.3
32.2	5.2	236	360	24.8
38.4	6.1	220	365	26.0
40.3	4.8	231	395	34.9
40.0	6.1	217	272	18.2

32.2	2.4	284	424	23.2
31.8	0.2	316	428	18.0
40.8	3.5	210	273	13.1
41.3	1.8	267	358	16.1
38.1	1.2	274	444	32.1
50.8	8.6	190	345	34.7
32.2	5.2	236	402	31.7
38.4	6.1	220	410	33.6
40.0	6.1	217	340	30.4
40.8	3.5	210	347	26.6
41.3	1.8	267	416	27.8
50.8	8.6	190	407	45.7