STAT 24400 Statistics Theory and Methods I

Homework 8: Due 10:00AM Thurs, March 10, 2016.
Put it in the clearly marked box outside 111 Zoology.

---

1. Call "This, it, thus, and" Class I words; Class II is "everything else." For each of 215 groups of 5 of James Mill's sentences, the number of Class I words was counted.

# Class I words	0	1	2	3	4	5
# groups	87	11	51	42	20	4

   Test whether a Binomial distribution (n = 5, θ) fits these data.

2. The members of a community are classified by Blood type:

O	A	B	AB	Total
121	120	79	33	353

   Theory has it that the probabilities of those types depends on gene frequency parameters r, p, q, where $r+p+q = 1$ and $P(\text{"O"}) = r^2$, $P(\text{"A"}) = p^2+2pr$, $P(\text{"B"}) = q^2+2qr$, and $P(\text{"AB"}) = 2pq$. Using numerical methods (that is, a method such as that described in Chapter 5 of Stigler's notes) we can find the MLEs of r, p, q; they are 0.580, 0.246, and 0.173 [you may use these values as the MLEs without verifying that they are]. Test if the community fits the theory.

3. Are fingerprint patterns genetic, or are they developmental? In 1892 Francis Galton compiled the following table on the relationship between the patterns on the same finger of 105 sibling pairs. Test the hypothesis that the patterns are independent for example, that knowing one sibling (A) has a Whorl on the finger does not help in predicting the pattern of the other (B).

	A children			
B children	Arches	Loops	Whorls	Totals
Arches	5	12	2	19
Loops	4	42	15	61
Whorls	1	14	10	25
Totals	10	68	27	105

4. For the Bortkiewicz Death by Horsekick Data, test the hypothesis that the data follow a Poisson distribution. You should group the counts for "4 or more" as one category.

Number of deaths	Frequency count
0	144
1	91
2	32
3	11
4	2
More	0
Total	280

5. An American roulette wheel is spun $n = 3880$ times in order to test if it is fair (i.e. to test if each slot has probability 1/38). Suppose that each of the 36 numbered slots (1, 2,.., 36) comes up exactly 100 times and each of "0" and "00" comes up 140 times.

   (a) Test at the 5% level using the χ^2 test if the wheel is fair.

   (b) Now suppose that before you had looked at the data you had suspected that the numbered slots were less likely than the "0" and "00", and you had decided to test the binomial hypothesis H_0: P("0" or "00") = 2/38 *versus* H_1: P("0" or "00") ¿ 2/38. We know that the UMP test of these hypotheses rejects $H_0$0 if Z (= total number of "0" and "00"s) is greater than C, where C is chosen for a level 0.05 test. Use the fact that under H_0 Z has approximately a Normal N(n*(2/38), n*(2/38)*(36/38)) distribution [this follows from the Central Limit Theorem, for example] to find C and perform the test.

   (c) Compare the result in (b) with that in (a). (This is intended to illustrate that with exactly the same data different conclusions may be reached depending upon how focused the test is upon a narrow hypothesis. The test in (a) tests against all alternatives to a fair wheel and the test of (b) is focused narrowly upon deviations of the probabilities of "0" and "00" slots from the others. As this example should suggest, it would be statistically impermissible to focus the test after seeing the data without compensating for this choice of a narrow hypothesis in some way.)