Tutorial 8:

Introduction to Data Science

Hypothesis Testing

Ernst C. Wit

November 11, 2024

Exercises

1. **The Lady Tasting Tea**. The following example is given by the inventor of modern statistics, Ronald A. Fisher, as the motivation for hypothesis testing. It is known as the "Lady tasting tea" example. Fisher described an acquaintance, an *English lady*, who claimed to have the ability to determine whether the tea was poured in the cup before or after the milk. Fisher thought that this was impossible and proposed her to do an experiment (the Lady was fully informed of the experimental method):

   - The experiment provided the Lady with 8 cups of tea at one time, 4 prepared with each method, presented in random order.
   - She was asked to select the 4 cups prepared by each method (i.e. milk before tea, MBT, or milk after the tea, MAT).
   - The test statistic was the number of successes in selecting the 4 cups.

   (a) What is the null and alternative hypothesis in this case?

   (b) How many configurations exist of putting 4 MBT and 4 MAT assignments on the 8 trials? (HINT: think permutations!)

   (c) Of these configurations, how many involve 0, 1, 2, 3 and 4 correct assignments. Fill in the following table:

Success count	Number of permutations
0	
1	
2	
3	
4	
Total	

   HINT: It may help to make a drawing of how many times you can get 0, 1 ... 4 correct assignments:

	number of successes	True MAT 1	2	3	4	True MBT 5	6	7	8
Our claim (A=MAT, B=MBT)	0	B	B	B	B	A	A	A	A
	1	B	A	B	B	A	A	A	B
				⋮				⋮	
	4	A	A	A	A	B	B	B	B

(d) The Lady tasted the eight cups and for all of them she correctly determined whether the milk was put in before or after the tea. How unusual is that if she was just guessing?
(HINT: If the null hypothesis is true, i.e. she was just guessing, then the probability of x successes is just equal to the number of ways you can get x successes divided by the total number of permutations.)

(e) If we use the conventional probability criterion, i.e. we reject H_0 is the p-value is less than 5%, what would be Fisher's conclusion in this case?

2. **Elections in Red-Blue Land.** Elections are held in Red-Blue Land, where the winner of more than 50% of the vote will become president. A newspaper wants to predict the elections by means of a sample of voters on election day (so-called *exit polls*).

(a) What are the null and alternative hypotheses that the newspaper formulates? Put the hypotheses both in words and in population parameters.

(b) The newspaper takes a sample of 100 individuals. What test statistic would you choose to test the null hypothesis?

(c) If the null hypothesis is true, then what is the approximate distribution of the test statistic?

(d) In the random sample 61 individuals selected Mr T over Ms H. Calculate the p-value associated with the hypothesis test.

(e) What do you conclude on the basis of the p-value if you were to use a significance level of $\alpha = 0.05$?

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.000	0.004	0.008	0.012	0.016	0.020	0.024	0.028	0.032	0.036
0.1	0.040	0.044	0.048	0.052	0.056	0.060	0.064	0.067	0.071	0.075
0.2	0.079	0.083	0.087	0.091	0.095	0.099	0.103	0.106	0.110	0.114
0.3	0.118	0.122	0.126	0.129	0.133	0.137	0.141	0.144	0.148	0.152
0.4	0.155	0.159	0.163	0.166	0.170	0.174	0.177	0.181	0.184	0.188
0.5	0.191	0.195	0.198	0.202	0.205	0.209	0.212	0.216	0.219	0.222
0.6	0.226	0.229	0.232	0.236	0.239	0.242	0.245	0.249	0.252	0.255
0.7	0.258	0.261	0.264	0.267	0.270	0.273	0.276	0.279	0.282	0.285
0.8	0.288	0.291	0.294	0.297	0.300	0.302	0.305	0.308	0.311	0.313
0.9	0.316	0.319	0.321	0.324	0.326	0.329	0.331	0.334	0.336	0.339
1.0	0.341	0.344	0.346	0.348	0.351	0.353	0.355	0.358	0.360	0.362
1.1	0.364	0.367	0.369	0.371	0.373	0.375	0.377	0.379	0.381	0.383
1.2	0.385	0.387	0.389	0.391	0.393	0.394	0.396	0.398	0.400	0.401
1.3	0.403	0.405	0.407	0.408	0.410	0.411	0.413	0.415	0.416	0.418
1.4	0.419	0.421	0.422	0.424	0.425	0.426	0.428	0.429	0.431	0.432
1.5	0.433	0.434	0.436	0.437	0.438	0.439	0.441	0.442	0.443	0.444
1.6	0.445	0.446	0.447	0.448	0.449	0.451	0.452	0.453	0.454	0.454
1.7	0.455	0.456	0.457	0.458	0.459	0.460	0.461	0.462	0.462	0.463
1.8	0.464	0.465	0.466	0.466	0.467	0.468	0.469	0.469	0.470	0.471
1.9	0.471	0.472	0.473	0.473	0.474	0.474	0.475	0.476	0.476	0.477
2.0	0.477	0.478	0.478	0.479	0.479	0.480	0.480	0.481	0.481	0.482
2.1	0.482	0.483	0.483	0.483	0.484	0.484	0.485	0.485	0.485	0.486
2.2	0.486	0.486	0.487	0.487	0.487	0.488	0.488	0.488	0.489	0.489
2.3	0.489	0.490	0.490	0.490	0.490	0.491	0.491	0.491	0.491	0.492

Table 1: Standard Normal Distribution. This means that values in the table correspond to probabilities $P(0 < Z \leq z)$, where Z is a standard normal distributed variable.