

# Assignment 04

Out: 04/26

Due: 05/10

## Instructions

### Collaboration:

Collaboration on solving the assignment is allowed, after you have thought about the problem sets on your own. It is also OK to get clarification (but not solutions) from online resources, again after you have thought about the problem sets on your own. There are two requirements:

- Cite your collaborators fully and completely (e.g., “XXX explained to me what is asked in problem set 3”). Or cite online resources (e.g., “I got inspired by reading XXX”) that helped you.
- Write your scripts and report independently - the scripts and report must come from you only.

### Late Submission:

Late submissions will not receive full credit. Half credit will be awarded to correct solutions submitted within 24 hours of the original deadline. Otherwise, no credit will be given.

### Submitting your assignment:

Submit your scripts and report via mail [pudc2020@mail.sustech.edu.cn](mailto:pudc2020@mail.sustech.edu.cn) .

1. [15 points] Toss a sieve 300 times repeatedly, the results are as follows:

The number of dice	1	2	3	4	5	6
Frequency	40	70	48	60	52	30

Is the structure of the dice uniform?? (  $\alpha = 0.05$  )

2. [20 points] Injecting antibiotics into the human body will cause the phenomenon of antibiotics and plasma protein binding, which will reduce the efficacy of the drug. The following is a list of the percentage of antibiotics bound to plasma proteins when five antibiotics are injected into the body of cattle:

antibiotic	penicillin	tetracycline	streptomycin	erythromycin	chloramphenicol
%	29.6	27.3	5.8	19	29.2
	24.3	32.6	6.2	18.3	32.8
	28.5	30.8	11	17.4	25
	32	34.8	8.3	21.6	24.2

(1) Are there any significant differences in the percentage of the five antibiotics bound to plasma proteins?

(2) Calculate the confidence interval of the average of the percentage of 5 antibiotics bound to plasma protein (95% confidence level)

(3) Find the confidence interval of the mean difference between penicillin and streptomycin, and the confidence interval of the mean difference between erythromycin and chloramphenicol (95% confidence level)

3. [15 points] The lapse rate is the rate at which temperature drops as you increase elevation.

Some hardy students were interested in checking empirically if the lapse rate of 9.8 degrees C km<sup>-1</sup> was accurate for their hiking. To investigate, they grabbed their thermometers and their Huawei wrist altimeters and found the following data on their hike.

	1	2	3	4	5	6	7	8
Elevation (m)	180	305	381	488	549	640	762	883
Temperature (degrees C)	13.3	12.2	13.3	10.0	8.3	9.4	8.3	7.2

Draw a scatter plot with regression line, and investigate if the lapse rate is 9.8 degrees C km<sup>-1</sup>

4. Edwin Hubble used the power of the Mount Wilson Observatory telescopes to measure features of the nebulae outside the Milky Way. He was surprised to find a relationship between a nebula's distance from the Earth and the velocity with which it was going away from the Earth. Hubble's initial data on 24 nebulae are listed below.

	Nebula	Velocity	Distance
1	S. Mag.	170	0.032
2	L. Mag.	290	0.034
3	NGC 6822	-130	0.214
4	NGC 598	-70	0.263
5	NGC 221	-185	0.275
6	NGC 224	-220	0.275
7	NGC 5457	200	0.450
8	NGC 4736	290	0.500
9	NGC 5194	270	0.500
10	NGC 4449	200	0.630
11	NGC 4214	300	0.800
12	NGC 3031	-30	0.900
13	NGC 3627	650	0.900
14	NGC 4626	150	0.900
15	NGC 5236	500	0.900
16	NGC 1068	920	1.000
17	NGC 5055	450	1.100
18	NGC 7331	500	1.100
19	NGC 4258	500	1.400
20	NGC 4151	960	1.700
21	NGC 4382	500	2.000
22	NGC 4472	850	2.000
23	NGC 4486	800	2.000
24	NGC 4649	1090	2.000

The recession velocity is in km per second, which was determined with considerable accuracy by the red shift in the spectrum of light from a nebula. The distance measures the distance from the Earth, in megaparsecs (1 megaparsec is 1 million parsecs, and 1 parsec is about 30.9 trillion km). The distance was measured by comparing mean luminosities of the nebulae to those of certain star types, a method that is NOT particularly accurate.

(1) [5 points] Make a scatter plot with distance as the Y-axis and recession velocity as the X-axis.

Describe what you see.

(2) [5 points] Add a simple linear regression line to the above scatter plot.

(3) [15 points] If Hubble's Big Bang Theory is correct, explain why the following two assumptions about the regression line you made in 5.2 need to be true:

- The intercept should be zero
- And the slope is the age of the universe

Address the first assumption with your regression results; and estimate the age of the universe.

(4) [5 points] Explain why improved measurement of distance would lead to more precise estimates of the regression coefficients.

Pregnant nonvegetarians	Pregnant vegetarians	Nonpregnant vegetarians
185	171	210
189	174	139
187	202	172
181	171	198
150	207	177
176	125	
	189	
	179	
	163	
	174	
	184	
	186	

5. [20 points] A researcher used 3 different accelerators and 4 different amounts of zinc oxide to synthesize rubber. The measured strength of rubber is as follows:

accelerators A zinc oxide B	B1	B2	B3	B4
A1	32	35	35.5	38.5
A2	33.5	36.5	38	39.5
A3	36	37.5	39.5	43

Do different accelerators and different amounts of zinc oxide have an important effect on the strength of rubber? ( $\alpha = 0.05$ )