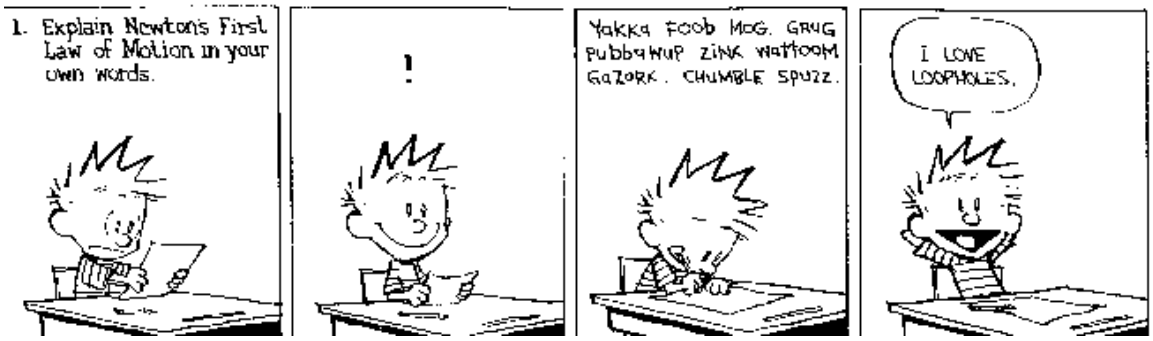
"Final" Homework for MATH 7343

Clearly state your method for analysis and results for each problem. If any computer printouts are included in your answer, you should label clearly each part of the printout which problem is it for. Also summarize your conclusion in plain language that non-statisticians can understand.



1. (40 points)

The file *hospital* on the disk contains data for personal health care expenditures by state for the year 1982. (See page B-11 of the textbook.) The variable *expadm* contains the average expense per admission into a community hospital, the variable *los* contains the average length of stay and the variable *salary* contains the average per employee in 1982. We are interested how the latter two variables affect the average expense per admission.

(a) Draw scatterplots of variables *expadm* vs. *los*, and of variables *expadm* vs. *salary*

(b) Using *expadm* as response variable and *los* as the explanatory variable, write down the least-squares regression equation.

(c) Suppose the average length of stay per admission at Fairyland Community Hospital is 6 days. Construct a 95% prediction interval for the average expense per admission at this hospital.

(d) Is there a significant linear relationship between expense per admission and the length of stay?

(e) Evaluate the fit of the model to data by creating a plot of residuals and a normal probability plot of the residuals. Discuss your findings.

(f) Fit the linear regression model where *expadm* is the response variable and *los* and *salary* are the explanatory variables. Interpret the estimated regression coefficients.

(g) What happens to the estimated coefficient of length of stay when average salary is added to the model?

(h) Does the inclusion of salary in addition to average length of stay improve your ability to predict mean expense per admission? Explain.

2. (30 points)

A student suspects that an e-commerce company is tracking her online shopping history, and not giving her best prices on items that she is likely to buy. She opened another account with the company using the computer at her work place. From her new account on her workplace computer, she looked up the displayed price on 13 items which she already checked the day before from old account on her personal computer. The data is shown below. (The prices are in Dollars.)

Item	Old account	New account
1	27.61	27.55
2	10.29	10.19
3	52.20	52.19
4	33.89	33.87
5	340.49	340.99
6	65.34	65.29
7	54.69	54.64
8	15.23	15.21
9	102.19	102.09
10	4.88	4.89
11	161.39	161.35
12	12.51	12.49
13	96.99	96.96

Find the appropriate statistical method to solve this problem. Report your analysis. What can you conclude about the student's suspicion on the e-commerce company's pricing practice?

3. (30 points) Groundhog Day.

(From https://www.ncdc.noaa.gov/customer-support/education-resources/groundhog-day)

Every February 2, thousands gather at Gobbler's Knob in Punxsutawney, Pennsylvania, to await the spring forecast from a special groundhog. Known as Punxsutawney Phil, this groundhog will emerge from his simulated tree trunk home and look for his shadow, which will help him make his much-anticipated forecast.

According to legend, if Phil sees his shadow the United States is in store for six more weeks of winter weather. But, if Phil doesn't see his shadow, the country should expect warmer temperatures and the arrival of an early spring.

But does Phil really know about the weather? The following table gives the data for the 20 years from 1988 to 2017. (There are more data on the web link. But for this problem, restrict the analysis to the following table only.)

	Phil sees its shadow	
March Temperature	Yes	No
Above normal	16	8
Below normal	5	1

(a) State the appropriate null hypothesis and the alternative hypothesis.
(b) I would like to use the χ^2-test to solve this problem. However, this is a one-sided problem while the χ^2-test is usually used as a two-sided test. What should I do?
(c) Carry out the test at $\alpha = 0.05$ level.
(d) How can we use Groundhog Day's information to predict the March temperature of that year?

4. (30 points)

To demonstrate the effect of nematodes (microscopic worms) on plant growth, a botanist prepares 16 identical planting pots and then introduces different numbers of nematodes into the pots. A tomato seedling is transplanted into each plot. Here are data on the increase in height of the seedlings (in centimeters) 16 days after planting.

Nematodes	Seedling Growth			
0	10.8	9.1	13.5	9.2
1000	11.1	11.1	8.2	11.3
5000	5.4	4.6	7.4	5.0
10000	5.8	5.3	3.2	7.5

(a) I would like to know if the nematodes have any effect on the plant growth at all. Carry out appropriate statistical analysis. What is your conclusion?
(b) Before conducting the experiment, I decided to test whether the introduction of nematodes reduces the plant growth. State the appropriate contrast to test this hypothesis, and carry out the appropriate test.
(c) From the data, it seems that the biggest drop in plant growth is between the group treated with 1000 nematodes per plant and the group treated with 5000 nematodes per plant. If I then decide to test whether these two groups do have different plant growth, what is the appropriate contrast for this test? Carry out the analysis.

5. (40 points).

A study was carried out to see how previous work experience improves the ability of a programmer to complete a complex programming task. 25 randomly chosen programmers were asked to complete a complex programming task within a specified time period. For each programmer, the length of previous programming work experience (in months) was recorded as well as whether the task was successfully completed (1 indicate success). The data was presented in the following table.
(a) Does work experience improve the programmer's ability? Carry out appropriate statistical analysis. What is your conclusion?
(b) With 95% confidence, give an estimation interval of the improvement in the odds of completing the task with specified time period for each extra year of work experience. (Note the data has unit of month.)
(c) The employer values the employee according to the probability of finishing the given task within the time period. A person with 100% probability of finishing the given task within the time period is worth the pay of $90,000 per year; A person with 80% probability of finishing is worth $72,000, etc. Knowing only the work history of an applicant, what

salary (X dollars per year) should the employer pay a programmer with 24 months of previous work experience?
(d) A second programmer with 18 months of work experience is willing to accept a yearly salary of $(X-10,000), where X is the answer you got in part (c) above. According to this analysis, is the second programmer a better deal for the company than the programmer with 24 months of previous work experience?

Experience in months	success
14	0
29	0
6	0
25	1
18	1
4	0
18	0
12	0
22	1
6	0
30	1
11	0
30	1
5	0
20	1
13	0
9	0
32	1
24	0
13	1
19	0
4	0
28	1
22	1
8	1
14	0

6. (20 points) Do 21.5.8 on page 512 of the textbook.

7. (10 points)

I would like to test the null hypothesis that two random variables X and Y are not associated with each other. From a data set, we calculated the sample correlation of X and Y as 0.6 (so that R-square is 0.36). Could I decide on this information only whether to reject the null hypothesis at $\alpha = 0.05$ level? What if I also know that the sample size is 40 (pairs of X and Y)?