

Statistics 5021

Professor: Glen Meeden, 313A Ford Hall, 612-625-8321, glen@stat.umn.edu

Home Page: <http://www.stat.umn.edu/~glen/classes/5021/>

Homework assignments and Exam schedule
Text: McClave and Sincich, Statistics, 9th edition

Assignment 1: Chapter 2 #'s 37 (also find the 65th percentile) and 57.

Generate a random sample of size 200 from a normal population with mean 50 and standard deviation 10. Using the computer construct a histogram for this sample. Also find its mean, variance, standard deviation, range and the 15th, 25th, 40th, 50th, 65th, 75th and 90th percentiles.

Using the data on page 99 compare the values of current and retired beanie babies by computing summary statistics and graphs of the data. If you are using Rweb you can access the data by entering the URL's

<http://www.stat.umn.edu/~glen/HW5021/beanieC.txt>

<http://www.stat.umn.edu/~glen/HW5021/beanieR.txt>

In the following when the URL's are given for other data sets only the last bit will be given since the rest of the address is the same.

Assignment 2: Chapter 3 #'s 102, 103, 104, 114 a) and b), 116, 120, 121, 129, 131 and 133.

Assignment 3: Chapter 4 #'s 22, 27 a) and b), 45, 93, 94, 96, 99, 107, 108 and 110 a), b) and c).

Let the joint distribution of X and Y be given in the table below.

		Y			
		1	2	3	4
	1	.12	.21	.24	.03
X	2	.06	.06	.12	.06
	3	.02	.03	.04	.01

Find the marginal distributions of X and Y . Are X and Y independent? Find $E(X)$, $E(Y)$, $E(4X - 2Y)$, $E(XY)$, $COV(XY)$ and $V(X - 3Y)$

Assignment 4: Chapter 5 #'s 89, 90, 91, 93, 95, 106 and 114. Chapter 6 #'s 37, 41, 42 and 44.

Assignment 5: Chapter 7 #'s 6, 10, 16, 22 a), b) and c), 23, 24, 27, 42 a) and b) and 86.

Generate ten random samples of size ten from a standard normal distribution, i.e. mean = 0 and variance = 1. For each sample compute the usual 90% t -confidence interval for the mean and check how many of your intervals contain 0. Suppose 60 students in a class do this exercise. For how many would you expect that their computed intervals contained 0 at most 7 out of the 10 times.

Assignment 6: Chapter 8 #'s 18, 19, 33, 34, 35, 49, 51, 58, 73, 77, 81 and 111

1st Exam

Assignment 7: Chapter 9 #'s 85 a) and b), 87 a) and b), 88 a) and b), 89, 93 and 100. Chapter 13 #'s 33, 39, 49 and 50.

A survey was conducted to find out whether a teenager's family status has any relationship to the amount of alcohol he or she consumes. A random sample of 200 teenagers were questioned about their alcohol use. The results are given in the table below.

	Alcohol		
	None	Occasional	Frequent
Upper Class	4	16	10
Upper Middle Class	11	40	24
Lower Middle Class	9	47	9
Lower Class	6	17	7

a. Do the data provide sufficient evidence to indicate a relationship between family status and the use of alcohol? Test using $\alpha = 0.05$.

b. What assumptions are necessary to justify the test of part a? Do you think they are satisfied in this application.

c. Calculate the percentage of the sample that fell into each cell. Graph these percentages on the vertical axis against family status on the horizontal axis. Does the graph support the results of the test in part a? Explain.

Assignment 8: Chapter 11 #'s 77 (all but e), 78, 83 and 88. In problem 78 find a 95% confidence interval for the life span of a horse with a gestation period of 349 days. For problem 88 in Rweb use the command 'lsfit(number,weight,intercept=F)' to find the least squares line through the origin.

Assignment 9: Chapter 12 #'s 13 and 47, 9 and 140, and 135.

In problem 13 find a 95% confidence interval for the expected value of Y when $x_1 = 60, x_2 = 1, x_3 = 23, x_4 = .25, x_5 = 3, x_6 = .5$ and $x_7 = 1.5$.

The URL for 13 is wateroil.txt, for 9 is mnsales.txt and for 135 is phosphor.txt.

A supermarket chain is interested in exploring the relationship between sales of its store-brand canned vegetables, y , and the amount spent on advertising, x_1 , and the amount of allotted shelf space, x_2 . One of the chain's supermarkets was randomly selected, and over a 20-week period y, x_1 and x_2 were observed. The URL for the data is canveg.txt.

a. Fit the following model to the data;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + Z$$

b. Conduct an F -test to investigate the overall usefulness of this model at $\alpha = 0.05$.

c. Test $\beta_3 = 0$ at $\alpha = 0.05$.

d. Explain what it means to say that advertising expenditures and shelf space interact.

e. Explain how you could be misled by using a first-order model instead of an interaction model to explain how advertising expenditures and shelf space influence sales.

A firm has developed a new type of light bulb and is interested in investigating the drop in light output as a % of original output, y . It is know that y depends on the cleanliness of the surface, x_1 , and the length of time the bulb has been in operation, x_2 . There are two possible values for x_1 are 0 and 1. Zero denotes a clean bulb and one denotes a dirty bulb. The URL for the data is lightbulb.txt. Use the data to build a regression model that relates y to some function of x_1 and x_2 . Include a residual analysis.

Assignment 10: Chapter 10 #'s 82, 65 and 85 a), b), c) and d).

2nd Exam

Final Exam