Exercises for Lectures 10-11

1. The Bureau of Business and Economic Research at the University of Montana conducted a poll of opinions of Montana residents in May 1992. Among other things, respondents were asked whether their personal financial status was worse, the same, or better than one year ago. The following table displays the results. Test whether the respondents' answers are uniformly distributed over the three possible responses.

Worse	Same	Better	Total
58	64	67	189

2. At the fifth hockey game of the season at a certain arena, 200 people were selected at random and asked how many of the previous four games they had attended. The results are given in the following table.

Number of games previously attended	Number of people
0	33
1	67
2	66
3	15
4	19

Test the hypothesis that these 200 observed values can be regarded as a random sample from a binomial distribution; that is, there exists a number θ ($0<\theta<1$) such that the probabilities are as follows: $p_0=(1-\theta)^4$, $p_1=4\theta(1-\theta)^3$, $p_2=6\theta^2(1-\theta)^2$, $p_3=4\theta^3(1-\theta)$, $p_4=\theta^4$.

3. Surveys that participants were asked were personal financial status and an income range. The following table gives a cross-tabulation of the answers to both questions.

Income range	Personal financial status			
	Worse	Same	Better	Total
Under \$20,000	20	15	12	47
\$20,000- \$35,000	24	27	32	83
Over \$35,000	14	22	23	59
Total	58	64	67	189

Test the null hypothesis that income is independent of opinion on personal financial status.

4. In an experiment involving 800 subjects, each subject received either treatment I or treatment II, and each subject was classified into one of the following four

categories: older males, younger males, older females, and younger females. At the end of the experiment, it was determined for each subject whether the treatment that the subject had received was helpful or not. The results for each of the four categories of subjects are given in the following table:

Older males	Helpful	Not
Treatment I	120	120
Treatment II	20	10
Younger males		
Treatment I	60	20
Treatment II	40	10
Older females		
Treatment I	10	50
Treatment II	20	50
Younger females		
Treatment I	10	10
Treatment II	160	90

(1) Show that treatment II is more helpful than treatment I within each of the four categories of subjects.

(2) Show that if these four categories are aggregated into only the two categories, older subjects and younger subjects, then treatment I is more helpful than treatment II with each of these categories.

5. Fit a polynomial of the form $y=\beta_0+\beta_1 x+\beta_0+\beta_2 x^2$ (which represents a parabola) to the 10 points given in the following table by the method of least squares.

i	x_i	y_i
1	1.9	0.7
2	0.8	-1.0
3	1.1	-0.2
4	0.1	-1.2
5	-0.1	-0.1
6	4.4	3.4
7	4.6	0.0
8	1.6	0.8
9	5.5	3.7
10	3.4	2.0

6. Suppose we survey 18 students about their two abilities, and got the correlation coefficient 0.5. Are the two abilities related?

7. Suppose we have tested the concentrations of three kinds of heave metals As, Cd and Cr as follows. Conduct centering and scaling on the data.

NO.	As (μg/g)	Cd (ng/g)	Cr (μg/g)
1	7.84	153.80	44.31
2	5.93	146.20	45.05
3	4.90	439.20	29.07
4	6.56	223.90	40.08
5	6.35	525.20	59.35
6	14.08	1092.90	67.96
7	8.94	269.80	95.83
8	9.62	1066.20	285.58
9	7.41	1123.90	88.17
10	8.72	267.10	65.56

8. Suppose the data is as follows. Can we conduct a linear regression $y=\beta_0+\beta_1x_1+\beta_2x_2+\varepsilon$ using the data? Why? Could you use scatter of x_1 and x_2 to show it?

y	-10	-8	-6	-4	-2	0	2	4	6	8	10
x1	1	2	3	4	5	6	7	8	9	10	11
x2	1	3	5	7	9	11	13	15	17	19	21

9. Suppose we are investigating the weights of a mouse. The age in days and its weight are given in the following table:

NO.	Ages in days x_i (d)	Weight (g)
1	6	12
2	9	17
3	12	22
4	15	25
5	18	29

(1) Build the linear regression model ($y=\beta_0+\beta_1x$).

(2) Test if the coefficient of x is significant.

(3) Test if the model is meanful.

10. Suppose we are investigating the yield of rice. Here y is the yield, while x1 and x2 are spike number and number of grain per panicle. Use the data analysis function in Excel to build the multiple linear regression model ($y=\beta_0+\beta_1x_1+\beta_2 x_2+\varepsilon$) for the data.

i	x_{i1}	x_{i2}	y
1	26.7	73.4	1008
2	31.3	59	959
3	30.4	65.9	1051

4	33.9	58.2	1022
5	34.6	64.6	1097
6	33.8	64.6	1103
7	30.4	62.4	992
8	27	74.4	945
9	33.3	64.5	1074
10	30.4	64.1	1029
11	31.5	61.1	1004
12	33.1	56	995
13	34	59.8	1045