ENGI 4421 Probability and Statistics
Faculty of Engineering and Applied Science

Problem Set 10

Chi-Square Tests; Simple Linear Regression

1. Is the following set of observations of objects in eight different directions consistent with a uniform distribution?

Direction	Number of objects observed
N	4
NE	6
E	13
SE	11
S	12
SW	18
W	11
NW	5

2. A model of a chemical process predicts the numbers of runs of a process that should fall in each of five time intervals. Are the values shown in the table below consistent with the values expected from the model at a 5% level of significance?

Time interval	observed	expected
0 - 5	7	10
5 - 10	19	25
10 - 15	27	30
15 - 20	31	25
> 20	16	10

3. Random samples of cables are taken from shipments from four manufacturers. Each cable in each sample is graded according to the load applied when that cable breaks. The results are summarized in this contingency table.

		Quality of cable			
		A	B	C	D
Manufacturer	X	8	22	14	6
	Y	14	61	19	6
	Z	10	21	10	9

Are these data consistent with independence between manufacturer and quality of cable?

4. A particular type of motor is known to have an output torque whose range in normal operation follows a normal distribution. Seven motors are chosen at random and are tested with the old and new methods of controlling the range of torque values. The results of the tests are as follows:

Motor:	1	2	3	4	5	6	7
New method:	5.25	3.16	4.43	6.12	5.75	2.21	6.01
Old method:	7.83	6.22	7.46	8.83	8.19	5.64	8.88

(a) Justify your choice of method in (b) below.
(b) Conduct an appropriate hypothesis test to determine whether there is sufficient evidence to conclude that the range of torques with the new method is at least 2 units less than with the old method.
(c) Use the simple linear regression model on these data to find the equation of the line of best fit to these data.
(d) Find the coefficient of determination R^2 and use it to comment on your answer to part (a) above.

---

5. A study was conducted to analyze the relationship between advertising expenditure and sales. The following data were recorded:

X	***Y***
Advertising ($)	**Sales ($)**
20	310
24	340
30	400
32	420
35	490

Assume a simple linear regression between sales Y and advertising X. Calculate the coefficients β_0 and β_1 of the line of best fit to these data and estimate the sales when \$28 are spent on advertising. Is there a significant linear association between Y and X?

---

6. [This is an extension of Example 12.06 from the lecture notes.]

A sample of 10 diesel trucks were run both hot and cold to estimate the difference in fuel economy. The results, in miles per gallon, are presented in the following table. (from "In-use Emissions from Heavy-Duty Diesel Vehicles", J. Yanowitz, Ph.D. thesis, Colorado School of Mines, 2001.)

Truck	Hot	Cold
1	4.56	4.26
2	4.46	4.08
3	6.49	5.83
4	5.37	4.96
5	6.25	5.87
6	5.90	5.32
7	4.12	3.92
8	3.85	3.69
9	4.15	3.74
10	4.69	4.19

(a) Use the simple linear regression model on these data to find the equation of the line of best fit (for "Hot" as the response and "Cold" as the regressor) to these data manually.
(b) Calculate the 95% prediction interval for a single future observation of fuel efficiency when run hot, for a truck whose fuel efficiency when run cold is 4.00 miles per gallon.
(c) Use Minitab (or another software package) to check that the distribution of the residuals is consistent with a normal distribution.
(d) Use Minitab (or another software package) to show the ANOVA table, the equation of the line of best fit and to display the line of best fit, the 95% confidence intervals and the 95% prediction intervals on a single diagram.
(e) Find the sample correlation coefficient r between hot and cold fuel efficiencies, correct to three significant figures.

☜ Back to the index of questions On to the solutions to this problem set ☞