

## Sample Test Problems for Test 2

(1). Define the following terms, giving the context, the formula, and the interpretation or use of the concept:

(a) **Sum of Squares for Error** (or Sum of Squares Within) for one-way layout analysis of variance.

(b) **Coefficient of Determination** for a simple linear regression of a variable  $y$  versus a variable  $x$  in a two-column dataset.

(c) **Tukey's studentized range distribution**, in multiple comparisons of group mean differences in one-way layout ANOVA.

(2). For a dataset consisting of the nine observations

5.2 7.1 8.6 12.2 3.8 6.6 10.2 7.9 5.5

(a) Find the sample median and quartiles.

(b) Find the empirical distribution function value at  $x = 10$ .

(3). In a 2009 class of 60 students, it is found that a college professor has given 18 C's, 30 B's and 12 A's. It had been previously reported in an online review of the college's teachers that this professor was well known to give equal numbers of A's, B's and C's. Use the 2009 grade data to test the hypothesis implied by the online review at level  $\alpha = .05$ . Explain in words what assumptions you must make in order that the hypothesis test you present is really applicable.

(4). A set of 60 measurements of operating lifetimes of new DVD players is taken. The specifications given by the manufacturer state that the lifetimes are approximately normally distributed with mean lifetime of 3200 hours and standard deviation of 1000 hours. Test the manufacturer's distributional claim using the following reported data-summary: the sample mean of the 60 observations is 3500 hours and sample standard deviation is 1500 hours, and the observations were tallied into successive intervals as

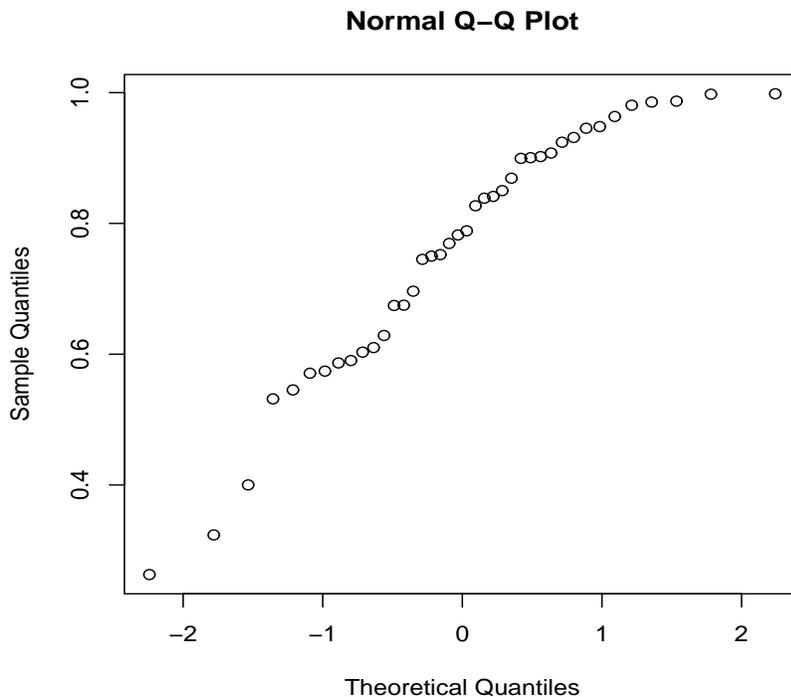
Interval	(0, 1000]	(1000, 2500]	(2500, 4000]	(4000, 10000]
# obs	8	26	20	6

(5). Pairs  $\{(x_i, y_i)\}_{i=1}^{35}$  representing independent measurements of temperature and pressure of a gas within a confined volume are taken, with the following partial results:

$$\bar{x} = 40, \quad \bar{y} = 130, \quad \sum_{i=1}^{35} x_i y_i = 180,000, \quad s_x^2 = 80, \quad s_y^2 = 300$$

Find the least-squares intercept and slope for a line  $y = a + bx$  through these points, and give the residual from the line for an observation  $(36, 150)$ .

(6). A (normal) QQplot is given below for a dataset of size 40. Do you think that these data look approximately normal? If not, what kind of deviation from normality does the plot suggest?



(7). A dataset consisting of diameter measurements in millimeters for 14 randomly sampled examples of a specific type of machined part, is collected from each of 4 different factories. These diameters can be assumed to be all normally distributed with the same variance, and with the same mean within each factory. A summary of the data values within each factory sample of size 14 is as follows:

Factory	1	2	3	4
Samp.Mean	27	25	30	28
Samp.Var	6	8	7	6

(a). Considering only the data from factories 1 and 2, suppose that we started out not quite sure that the variances of the part-diameters were exactly the same in Factories 1 and 2. Give a 95% two-sided confidence interval for the ratio of the variance of part-diameters in Factory 2 over that in Factory 1.

(b). Find SSB, SSW, and the F statistic for the one way layout in this example that would be used to test equality of the **mean** diameters of the machined parts in all of the four factories. Would you accept or reject the null hypothesis of equal means based on these data ?

(c). Give the ANOVA table for this example, and give your best estimate of the unknown common variance  $\sigma^2$  of the diameters in all factories. *Extra: can you find a confidence interval for  $\sigma^2$  in terms of this estimator ?*