ENGI 4421 Final Examination

2020 August

---

1. How many distinct rearrangements are there of the nine letters in the word "`INFERENCE`"? Express your answer as an integer. [9]

---

2. 100 beams in a random sample are each assigned to one of seven categories, based on their strength and weight. A model has been developed that indicates how many beams we would expect to see in these seven categories. The table below shows the numbers observed and the numbers expected in each category. [14]

Category	**observed**	**expected**
A	4	5
B	7	10
C	23	20
D	24	30
E	17	20
F	13	10
G	12	5
Total	**100**	**100**

Use the χ^2 goodness-of-fit test to determine, at a level of significance of .05, whether or not these data are consistent with the model.

---

3. The mass of a cable is measured using two sets of scales. [10]

One set reports a mass of (577.4 ± 3.0) g.

The other set reports a mass of (574.9 ± 4.0) g.

Find the estimate of the mass that has the minimum uncertainty ***and*** find that uncertainty.

☞

4. The time T from installation to failure of a high-load pipe follows an exponential distribution and is independent of the times to failure of all other pipes. Upon failure a pipe is replaced immediately with a new pipe whose lifetime follows the same probability distribution. The mean time to failure μ is known to be 5 years.

(a) Find $\mathrm{P}[T > 12]$. [3]

(b) Find the time $\tilde{\mu}$ (correct to two decimal places) before which half of all pipes fail. [3]

(c) The number of pipes N that fail in 20 years follows a Poisson distribution. Find the probability that exactly four pipes fail during those 20 years. [4]

(d) Suppose that these pipes are installed in fifty (50) locations and that the number N of failures is recorded for each location over the period of 20 years. Find the approximate probability that the mean number $\overline{N}$ of pipe failures in 20 years among these 50 locations is less than 3.5. (Linear interpolation is **not** required). [6]

---

5. The efficiencies of a random sample of 15 machines were measured (in the appropriate units) before and after the application of a new process to those machines. You may assume that these efficiencies are normally distributed. [14]

The summary statistics are:

For y = efficiencies after the application of the new process

$n_Y = 15, \quad \sum y = 1631.31, \quad \sum y^2 = 177\,687.28$

For x = efficiencies before the application of the new process

$n_X = 15, \quad \sum x = 1449.23, \quad \sum x^2 = 140\,284.76$

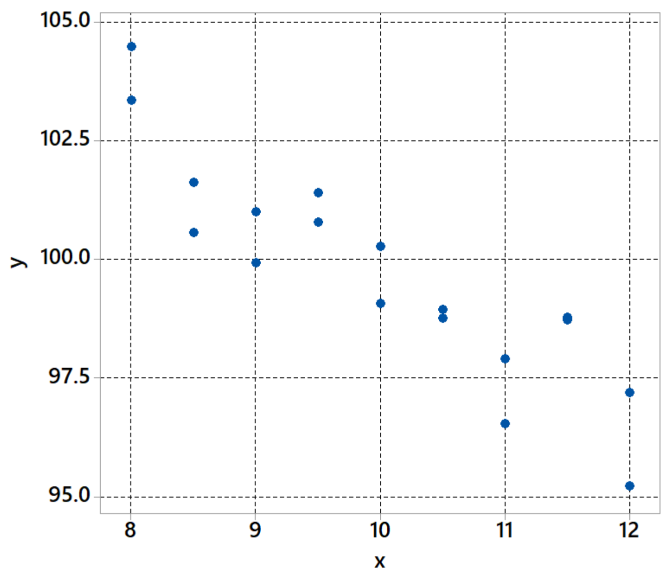
For $d = y - x$:

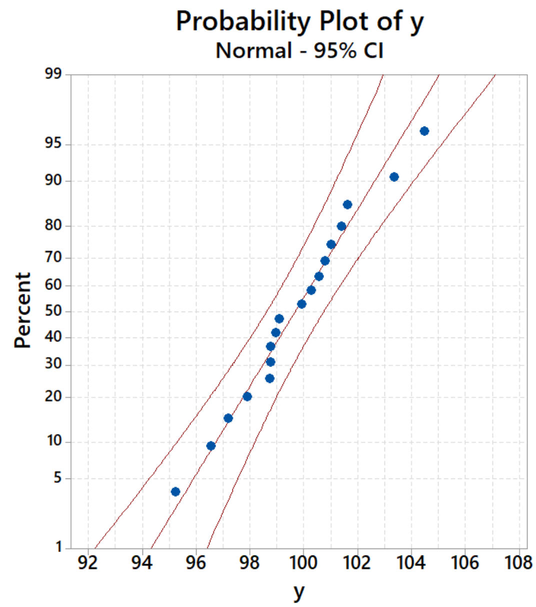
$n_D = 15, \quad \sum d = 182.08, \quad \sum d^2 = 2\,307.58$

Conduct an appropriate hypothesis test, at a level of significance of .01, to determine whether or not there is sufficient evidence to conclude that the process has improved the efficiency ***by more than* 10 *units***. Justify your choice of test. Use whichever set of summary statistics above is appropriate.

---

6. For 18 observations of a pollution index (x), the lifetime (y) of a pipeline section is measured. A scatterplot and a normal probability plot of the observations are presented here:





(a) State ***two*** reasons why it is appropriate to use the simple linear regression model. [2]

(b) Given the summary statistics [3]

$$n = 18 \qquad \sum x = 180.0 \qquad \sum y = 1794.4$$
$$\sum x^2 = 1830.0 \qquad \sum xy = 17896.9 \qquad \sum y^2 = 178961.6$$
$$n\,S_{xx} = 540.00 \qquad n\,S_{xy} = -847.80 \qquad n\,S_{yy} = 1437.44$$

find the equation of the regression line $y = \hat{\beta}_0 + \hat{\beta}_1 x$.

(c) Construct the ANOVA table from these data. [4]

(d) To the nearest one per cent, what proportion of the variation in Y is explained by the linear regression? [2]

(e) Conduct an hypothesis test to determine whether or not there is a useful linear relationship between Y and x. [3]

(f) Find the 95% prediction interval for a future observation of Y at x = 10.0 . [5]

BONUS QUESTION

(g) Is there sufficient evidence to conclude that the slope β_1 is steeper than –1? [+5]

Find the boundary of the one-sided 99% confidence interval for β_1.

7. A manager knows that 60% of all boxes of faucet gaskets are "good" (that is, contain no defective gaskets), while the other 40% of boxes are "bad" (that is, contain at least one defective gasket).

A test is available, costing \$10 per box, to determine whether a box contains any defective gaskets. However, the test is not perfect. A "bad" box has a probability of 0.9 to fail the test. A "good" box has a probability of 0.2 to fail the test.

Accepting a good box earns \$50.
Accepting a bad box causes a loss of \$65.
Rejecting a box causes a loss of \$15 (regardless of whether it is good or bad).

(a) Show that the probability that a box is good, given that it passed the test, is $\frac{12}{13}$. [4]

(b) Draw a decision tree to illustrate this situation and determine the optimum strategy that maximizes the expected gain per box for the contractor and find the exact value of that maximum expected gain. [14]

☜ Back to the index of questions On to the solutions ☞