STAT 650 Assignment-05

Due: October 25, 2022 5:59PM

Instructions :

- This assignment is based on materials coverved in Lectures 11, 12, 13 and 14.
- We highly recommend that you write your solutions in **Jupyter Notebook** and convert them to a **PDF** file. However, you may write the solutions by hand, scan and upload it as **.pdf** file.
- The PDF file should be under 15MB in size. It must be uploaded as a single file and not separate files for separate pages. Do not take a photo of each page and then paste them into a document - this will make your file too big and the results will generally not be very readable anyway.
- Please make sure that the solutions are neat, legible and in order (even if you choose to solve them in different order).
- Include **STAT650--UIN** at the top of the first page.
- Name the file as **UIN_assign5.pdf** (For eg, if someone's UIN is 123456789, then the file should be named 123456789_assign5.pdf). Otherwise, your submission will not be graded.
- You should upload your file through Canvas. You can make multiple submissions within the deadline, but only the latest submission will be considered for grading.
- You may take 6 hours extra after the due time, but 10% of your marks will be deducted.
- It is strictly prohibited to share or distribute the content in this document.

The aim of this assignment is to get familiar with Hypotheses testing.

Problem 1

One microchip manufacturer found that after 5000 hours, 32 out of 200 chips selected at random were defective. Find an approximate 99 percent confidence interval for the proportion of defective chips in the total production.
[5 marks]

Problem 2

Let $X \sim Binomial(n, p)$. When X is measured to be x, we want $x/n \pm 0.0005$ to be an approximate 99 percent confidence interval for p.
(a) If we know p is around 1/8, how should we choose the sample size n?
(b) We do not have any information that p is around 1/8, what sample size should we take?
[$5 + 5 = 10$ marks]

Problem 3

Two rubber compounds were tested for tensile strength. Rectangular materials were prepared and pulled in a longitudinal direction. A sample of 14 specimens: seven from compound A and seven from compound B, was prepared. But it was later found that two specimens from compound B were defective, and they had to be removed from the test. The tensile strength (in units of 100 pounds per square inch) are as follows:

$A = [32, 30, 33, 32, 29, 34, 32]$ and $B = [33, 35, 36, 37, 35]$.

Calculate a 95 percent confidence interval for the difference in the mean tensile strengths of the two rubber compounds. State your assumptions and steps clearly.
$[10$ marks]

Problem 4

(a)

Let the number of successes Y_1 in n independent trials be binomial ($Y_1 \sim Binomial(n, p_1)$). [Hint: $Y_1 \sim Binomial(n, p_1)$ means that the distribution of the number of successes. Then the number of failures, say $Y_2 = n - Y_1$ and $Y_2 \sim Binomial(n, p_2)$, where $p_2 = 1 - p_1$.]

Show that the random variable Q given below has a chi-square distribution with degrees of freedom equal to one.

$$Q = \sum_{i=1}^{2} \frac{(Y_i - np_i)^2}{np_i}$$

(b)

In general, suppose an experiment has k mutually independent and exhaustive outcomes. Let $p_1, p_2, \cdots, p_k$ represent the probabilities of these k outcomes, respectively. Assume that the experiment is performed n independent times, and let $Y_1, Y_2, \cdots, Y_k$ stand respectively for the frequencies of the 1^{st} outcome, the 2^{nd} outcome, ..., and the k^{th} outcome. Jointly, $Y_1, Y_2, \cdots, Y_k$ are said to have a **multinomial distribution** with multinomial probabilities $p_1, p_2, \ldots, p_k$ such that $\sum_{i=1}^{k} p_i = 1$.

If $Y_1, Y_2, \cdots, Y_k$ have a multinormial distribution with parameter $n, p_1, p_2, \cdots, p_k$, then

$$Q_{k-1} = \sum_{i=1}^{k} \frac{(Y_i - np_i)^2}{np_i} \sim \chi^2_{(k-1)}$$

One famous Mendelian theory assumes the crossing of two types of peas states that the respective probabilities are $p_1 = 9/16, p_2 = 3/16, p_3 = 3/16, p_4 = 1/16$ for four mutually exclusive classifications: 1) round and yellow, (2) wrinkled and yellow, (3) round and green, (4) wrinkled and green. To test this assumption, we performed $n = 80$ independent experiments, observing fequencies $y_1 = 42, y_2 = 17, y_3 = 13$, and $y_4 = 8$. Check whether the Mendelian theory assumption is right or wrong.
$[15 + 10 = 25$ marks]

Problem 5

A manufacturer of a new pain relief tablet would like to demonstrate that its product works twice as fast as the competitor's product. Specifically, the manufacturer would like to test

$H_0 : \mu_1 = 2\mu_2$ and $H_a : \mu_1 > 2\mu_2$

where μ_1 is the absorption of the competitive product and μ_2 is the absorption of the new product. Assuming that the variances σ_1^2 and σ_2^2 are known, develop a procedure for testing this hypothesis. Write down the test statistic for testing above hypotheses. Show your steps clearly.
$[15$ marks].

Problem 6

dataFile.csv is needed for this problem.
This dataset consists of two independent variables: x_1 and x_2 and y is the response variable.

(a) Fit a linear regression model and print estimated model parameters, coefficient of determination (R^2), and check the significance of model parameters.
(b) Check the nomality of residuals by providing suitable graphical interpretation(s).
(c) Are there any outliers in the dataset? If yes, remove them and repeat step (a), (b), and (c) to fit a new model for the dataset without outliers.
(d) By using the outliers free data and the corresponding model, find the confidence range for predicted values of Y.
(e) Use your model in (c) to predict Y values when
$(x_1, x_2) = (1, 2), (3, 4), (5, 6), (7, 8), (9, 10), (11, 12), (13, 14)$.

[$5 \times 7 = 35$ marks]