Basic Information and Instructions for Test - 3

You need to use your favorite software to compute the basic output and then use it to answer the questions. You need to submit the program and output in the drop box and submit the papers as well.

Exam Materials:

Test will be based on the materials covered in class.

Module-7 - Design of Experiments - (ANOVA) - Note that the lecture slides contains more materials than the book

Module-8 - Generalized Linear Model (Logistic Regression)

Module-9 - Regression with Time Series data

Name: ____________________________ **Instructor: D. Kushary**

1. A farmer wants to determine the effect of five different concentrations of lime on the pH of the soil on a firm. Fifteen soil samples are to be used in the experiment, five from each of three different locations. The five soil samples from each location are then randomly assigned to the five concentrations of lime and 1 week after the lime is applied the pH of the soil is measured.

   (a) Write down the model with factors and levels (3 points).

   It is a two factor model with no interaction:

   Model: $Y_{ij} = \mu + \alpha_j + \beta_i + e, \quad j = 1, 2, 3, 4, 5 \ \& \ i = 1, 2, 3.$

   Concentration = Factor A (column) (k=5) and Location = Factor B (row) (b=3)

   (b) Complete the ANOVA table below and perform a appropriate testing for concentrations ? Use $\alpha = 0.05$. (7 points)

Source	Sum of Squares	Deg of Freedom	Mean Squares	F - Stat	P -value
Concentration					
Location	0.170			5.812	
Error					
Total	1.429				

   Solution:

   Step - 1 We know the degrees of freedoms are (k-1) = 5 -1 =4, b -1 = 3 -1 =2 and for error it is (k-1)(b-1) = 8.

   Step - 2 - MSB = SSB/(b-1) = 0.170/2 = 0.085

   Step - 3 - $\frac{MSB}{MSE} = F_B - Stat \quad \Rightarrow \quad MSE = \frac{MSB}{F_B - Stat} = \frac{0.085}{5.812} = 0.0146$

   Step - 4 - $SSE = 0.0146 \times 8 = 0.117 \quad \Rightarrow \quad SSA = SST - SSB - SSE = 1.429 - 0.170 - 0.117 = 1.142$

   Step - 5 $MSA = 1.142/4 = 0.286 \quad \Rightarrow \quad F_A - stat = MSA/MSE = 0.286/0.015 = 19.52$

   To test the hypothesis about the concentration we set
   $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0 \ vs \ H_1$: Not all of them 0.

   As $F_A - Stat = 19.52 > F_{0.05,4,8} = 3.84$ we reject the null hypothesis.

   (c) Is concentration-1 different from concentration-5 at 5% level? $T_1 = 15.5$ & $T_5 = 16.5$ (5 points)

   Now to do the multiple comparison between 1st and the 5th concentration, we use Fisher's method. Concentration-1 is different from concentration-5 if and only if

$$
\begin{aligned}
|(\bar{X}_1 - \bar{X}_5)| &> t_{0.05/2,8} \sqrt{MSE\left(\frac{1}{3} + \frac{1}{3}\right)} \\
\left|\frac{16.5}{3} - \frac{17.5}{3}\right| &> 2.306 \sqrt{0.0146\left(\frac{1}{3} + \frac{1}{3}\right)} \\
0.33 &> 0.2277 \quad \Rightarrow \quad True
\end{aligned}
$$

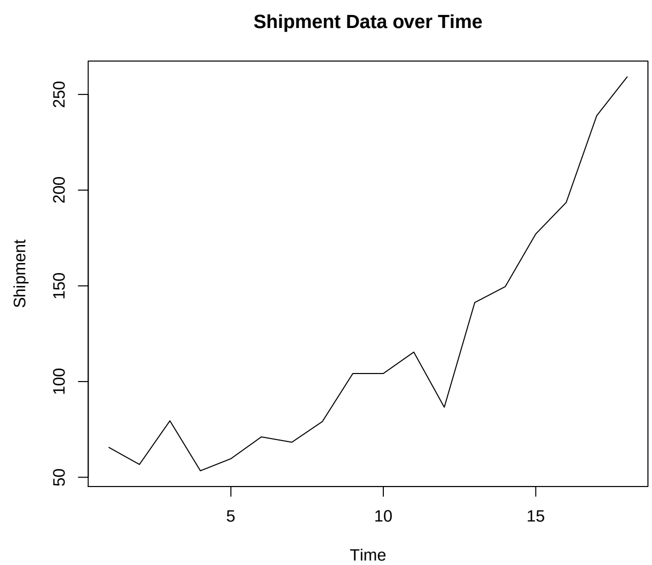
   Hence concentration-1 is different from concentration-5 at 5% level.

2. The following tables provides the shipment data over the time.

t	Yt	t	Yt	t	Yt
1	65.6	7	68.3	13	141.3
2	56.7	8	79.1	14	149.6
3	79.5	9	104.2	15	177.1
4	53.4	10	104.2	16	193.6
5	59.7	11	115.4	17	238.9
6	71.1	12	86.6	18	259.2

(1) Plot the data and fit a simple linear regression using time as the independent variable and test for positive autocorrelation at 5% level (must use tables). (7.5 points)

Ans: The SLR model is significant and the estimated regression line is $\hat{Y}_t = 15.203 + 10.701\, t$.



To test for positive correlation we set $H_0 : \phi = 0 \; vs \; H_1 : \phi > 0$

As $d = 0.5819 < 1.08 = d_L$ (at n=15) $< d_L$(at n=18) , $H_0 : \phi = 0$ is rejected and hence we conclude that auto correlation is present. (you can use p-value)

(2) Based on your conclusion, comment on the appropriateness of using this model and if not then try the quadratic model and test for the autocorrelation. (7.5 points)

As one of the reasons for auto correlation is "missing regressors" and the original plot shows a possibility of second degree relationship we fit a quadratic relationship. Hence the new model is $Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \epsilon$.

The estimated model is $\hat{Y}_t = 78.22 - 8.204\, t + 0.995\, t^2$ and all the coefficients are significant at 5% level.

To test for positive correlation for this model we set $H_0 : \phi = 0 \; vs \; H_1 : \phi > 0$

As $d = 2.1877 > 1.54 = d_U$ (at n=20) $> d_L$(at n=18) , hence $H_0 : \phi = 0$ is not rejected.

3. On 28 January 1986 the space shuttle Challenger was destroyed in an explosion shortly after launch from Cape Kennedy. The cause of the explosion was eventually identified as catastrophic failure of the O-rings on the solid rocket booster. The failure likely occurred because the O-ring material was subjected to a lower temperature at launch (31°F) than was appropriate. The material and the solid rocket joints had never been tested at temperatures this low. Some O-ring failures had occurred during other shuttle launches (or engine static tests). The failure data observed prior to the Challenger launch is given in the data:

a. Fit a logistic regression model to the response variable Fail. Use a simple linear regression model as the structure for the linear predictor. Interpret the coefficient of temperature.(5 points)

The estimated model is

$$\hat{\pi} = \frac{1}{1 + e^{-(10.875 - 0.1713\,x)}} \quad \text{where x = Temperature}$$

b. Find a 97% confidence interval for the probability of failure when temperature is 31.(5 points)

Answer: At temperature 31 the estimated probability of failure is

$$\hat{\pi} = \frac{1}{1 + e^{-(10.875 - 0.1713 \times 31)}} = 0.9962 \quad \text{and} \quad 97\% \text{ CI is } (0.2243, 0.9999)$$

c. Estimate and interpret the odds ratio for temperature.(5 points)

Odds Ration for temperature $= e^{\hat{\beta}_1} = e^{-0.1713} = 0.8426$

It means that if temperature goes up by 1 degree then odds of failure goes down by (100 - 84.26) = 15.74%.

d. Expand the linear predictor to include a quadratic term in temperature. Is there any evidence that this quadratic term is required in the model?(5 points)

After adding the quadratic term the parameter estimate table is as follows:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	39.1593	44.8255	0.87	0.3823
Temp	-1.0192	1.3164	-0.77	0.4388
Temp2	0.0063	0.0096	0.65	0.5130

To test the addition of quadratic term: we set up the test as

$H_0 : \beta_2 = 0 \quad vs \quad H_1 : \beta_2 \neq 0$

The p-value for the above test is 0.5130 which is much bigger than any reasonable α and hence we do not reject the null hypothesis.