Take Home Exam 4

Due Friday, December 9th, 2016

STAT 4013: Statistical Methods I (FALL 2016)

Directions: This paper is to be the cover page of your exam. All questions are to be typed and attached to the end of this document. You must submit it to D2L Dropbox no later than December 9th, 2016 at 11:59 pm. This portion of the exam is to test your knowledge of both statistical concepts learned thus far and how to effectively use statistical software (in this case R). All work submitted must be your own. You may use anything at your disposal except another human being. Solutions should be written following all rules of English, and points will be deducted as needed. Graphics should appear within the solutions, and an appendix with just your R code should be attached at the end.

You are allowed only one submission to Dropbox, so make sure your answers are as complete as you want them to be before submitting because there is no going back!!!!

Failing to submit your code will result in a reduction of your grade by 20 points.

*Questions:*

1. The shelf life of stored meats is the time a prepackaged cut remains salable, safe, and nutritious. Standard packaging in ambient air atmosphere has a shelf life of about 48 hours after which the meat quality begins to deteriorate from microbial contamination, color degradation, and shrinkage. Vacuum packaging is effective in suppression of microbial growth; however, other quality losses remain a problem.

Recent studies suggested controlled gas atmospheres as possible alternatives to existing packagings. Two atmospheres with promise to combine the capability for suppressing microbial development while maintaining other meat qualities were (1) pure carbon dioxide CO2 and (2) mixtures of carbon monoxide CO, oxygen O2, and nitrogen N. Based on this new information the investigator hypothesized that some form of controlled gas atmosphere would provide a more effective packaging environment for meat storage. The investigators data is below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Psychotropic Bacteria | | | |
| Packaging Condition | log(count/cm2) | | | | |
| 1 Commercial Plastic Wrap | 7.66 | 6.98 | 7.80 | 7.74 | |
| 2 Vacuum packaged | 5.26 | 5.44 | 5.80 | 5.64 | |
| 3 1% CO, 40% O2, 59% N | 7.41 | 7.33 | 7.04 | 7.21 | |
| 4 100% CO2 | 3.51 | 2.91 | 3.66 | 3.48 | |

* 1. What is the factor and what are the factor levels? (5 points)
  2. Provide an appropriate ANOVA table and write an appropriate conclusion. (10 points)
  3. Perform any necessary post-hoc analysis and write an appropriate conclusion. (10 points)

1. A criminologist studying the relationship between level of education and crime rate in medium-sized U.S. counties collected the following data for a random sample of 8 counties; *X* is the percentage of individuals in the county having at least a high-school diploma, and *Y* is the crime rate (crimes reported per 100,000 residents) last year. The data is provided below.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *County:* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| *X:* | 74 | 82 | 81 | 77 | 82 | 88 | 83 | 76 |
| *Y:* | 8487 | 8179 | 8362 | 7815 | 8075 | 8040 | 6981 | 7582 |

* 1. Compute the ordinary least squares line of best fit. (5 points)
  2. Test the significance of the slope. (10 points)
     1. Include the four steps to the hypothesis test
     2. Use the p-value method at a significance level of 0.05
  3. Create and interpret a confidence interval and prediction interval for a county with an 80% high-school completion rate. (10 points)
     1. Use a significance level of 0.01

1.

1. Factors: Packaging Conditions

Levels of Factors:

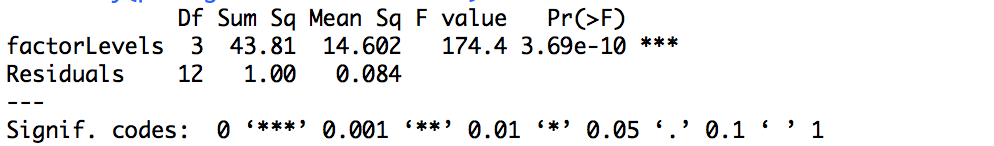
1. Commercial Plastic Wrap

2. Vacuum packaged

3. 1% CO, 40% O2, 59% N

4. 100% CO2

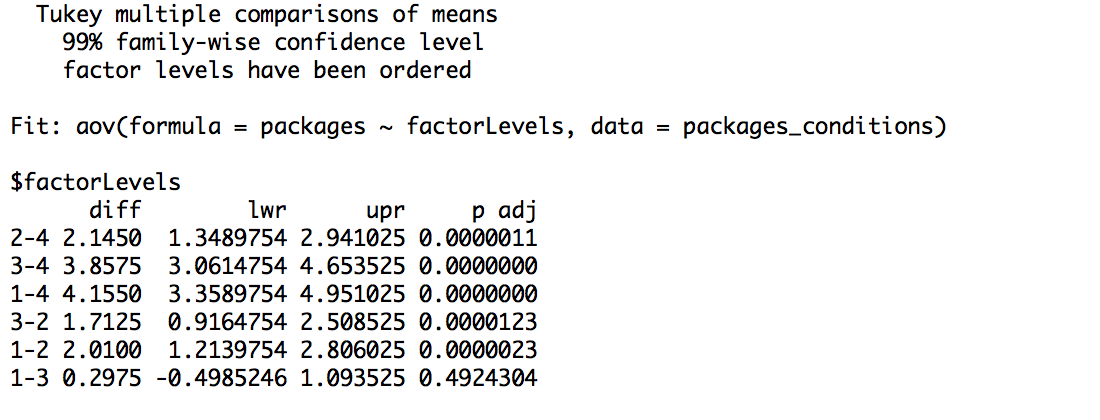
b)



Since, p-value< alpha (0 < 0.05) => true.

There is enough evidence that some form of controlled gas atmosphere would provide a more effective packaging environment for meat storage.

c)



Since zero lies between, 1-3, i.e, Commercial Plastic Wrap and 1% CO, 40% O2, 59% N packing condition’s(factors) means, they only do not significantly differ. Rest all combinations, significantly differ from each other’s mean.

2.

1. y = 9289.79-16.79x

b)

1. Hypothesis: H0 :

Test Statistics: ttest =

Decision: We Fail to Reject H0 because |tteest| >

Conclusion: There is no significant evidence that the slope differs

from Ø.

ii. p-value : 0.7107 < alpha = 0.05 => false. Therefore, with 95% confidence level we fail to reject null hypothesis.

c)

1. We are 99% confident that last year’s crimes report per 100,000 residents of all the counties with an 80% high school completion rate should lie between 7269.82 and 8623.03 crime rate.

We are 99% confident that last year’s crime report per 100,000 residents of an individual county with an 80% high school completion rate should have between 5923.73 and 9969.12 crime rate.

**Appendix**

#1b.

commercial\_plastic\_wrap<-c(7.66,6.98,7.80,7.74)

vacuum\_packaged<-c(5.26,5.44,5.80,5.64)

co\_o2\_n<-c(7.41,7.33,7.04,7.21)

co2<-c(3.51,2.91,3.66,3.48)

factorLevels <- c(1,1,1,1,2,2,2,2,3,3,3,3,4,4,4,4)

factorLevels <- factor(factorLevels)

packages<-c(commercial\_plastic\_wrap, vacuum\_packaged, co\_o2\_n, co2)

packages\_conditions<-data.frame(packages,factorLevels)

packages\_conditions.anova <- aov(packages~factorLevels, data = packages\_conditions)

summary(packages\_conditions.anova)

#1c.

TukeyHSD(x= packages\_conditions.anova, which = "factorLevels", ordered = TRUE, conf.level = .99 )

#2b

x\_axis <- c(74,82,81,77,82,88,83,76)

y\_axis <- c(8487,8179,8362,7815,8075,8040,6981,7582)

model = lm(formula = y\_axis~x\_axis)

summary(model)

#OutPut:

Call:

lm(formula = y\_axis ~ x\_axis)

Residuals:

Min 1Q Median 3Q Max

-915.0 -244.2 195.0 307.7 439.8

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 9289.79 3474.19 2.674 0.0368 \*

x\_axis -16.79 43.17 -0.389 0.7107

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 514.2 on 6 degrees of freedom

Multiple R-squared: 0.0246, Adjusted R-squared: -0.138

F-statistic: 0.1513 on 1 and 6 DF, p-value: 0.7107

#2c

Xnew = list(x\_axis = 80)

predict(model, newdata = Xnew, interval = "confidence", level = .99)

#output:

fit lwr upr

1 7946.422 7269.82 8623.024

predict(model, newdata = Xnew, interval = "prediction", level= .99)

#output:

fit lwr upr

1 7946.422 5923.727 9969.117