

STAT 3202: Homework 09

Spring 2019, OSU

Due: Monday, April 15

Please see the Homework Policy on Carmen for information about homework formatting, submission, and grading.

Exercise 1

In the previous homework, we considered data from four chemical plants that discharged effluents into streams in the vicinity of their locations. To monitor the extent of pollution created by the effluents and to determine whether this differs from plant to plant, the company collected random samples of liquid waste, five specimens from each plant. The data are given in the accompanying table.

Plant	Polluting Effluents (lb/gal of waste)				
A	1.65	1.72	1.50	1.37	1.60
B	1.70	1.85	1.46	2.05	1.80
C	1.40	1.75	1.38	1.65	1.55
D	2.10	1.95	1.65	1.88	2.00

We found that there were differences among these plants. In R, please use the **TukeyHSD** function to determine what differences there are among the plants. What plants are different at family-wise error rate (FWER) 0.05? Summarize in a sentence or two what this tells us about the plants.

Exercise 2

In the previous homework, we considered a study in which maneuver times were measured (in seconds) for vehicles of various sizes that were involved in making a left turn at an intersection. A summary of the data is reproduced below, with numbers assigned to the three categories.

Vehicle Type	Sample Size	Mean	Standard Deviation
1: Small Car	45	4.59	0.70
2: Large Car	102	4.88	0.64
3: Truck or bus	18	6.24	0.90

Investigate how small cars compare to the other groups by calculating confidence intervals for two comparisons: $\mu_1 - \mu_2$ and $\mu_1 - \mu_3$. Note: A confidence interval for $\mu_i - \mu_j$ is given by:

$$\bar{Y}_i - \bar{Y}_j \pm t_{\frac{\alpha}{2}, n-K} \sqrt{\text{MSE} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$$

where:

- K is the number of groups,

- n_i is the number of units in group i ,
- \bar{Y}_i is the sample mean in group i ,
- $100(1 - \alpha)\%$ is the desired confidence level for the interval.

Construct the two confidence intervals using a *Bonferroni correction* to ensure that the *family-wise* confidence level is 98%.

Exercise 3

Because hockey programs for kids are often separated by age groups, there is a theory that the month in which you are born can determine your success in hockey. Specifically, if you are on the “older” side of your age range (by some number of months) you might be slightly larger and stronger than the other kids in your age range, and thus be more successful and receive more attention and training early on. This initial difference can then be propagated forward, and result in your being more likely to play professional hockey. This is called “The Relative Age Effect.”

One data source for this argument is a 1985 article by Barnsley, Thompson, and Barnsley, called “Hockey Success and Birthdate: The Relative Age Effect.” In it, the authors investigated the birth months of National Hockey League players in the 1982/83 season and compared it with the baseline birthrate. Separated into quarters, they found:

Months	Number of National Hockey League Players
Jan / Feb / Mar	229
Apr / May / Jun	213
Jul / Aug / Sep	157
Oct / Nov / Dec	116

The proportions of (male) births by month was:

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
0.090	0.074	0.080	0.082	0.092	0.082	0.087	0.087	0.087	0.083	0.079	0.077

Assuming the baseline birthrates are fairly consistent across years, does the data provide evidence that National Hockey League players are more likely to be born at certain times of the year than others? Please perform this calculation by hand (you can use R or a calculator, but not just the `chisq.test` function in R). Test at $\alpha = 0.05$. Please report:

- The value of the χ^2 test statistic.
- The degrees of freedom.
- Your conclusion (accept / reject).

Exercise 4

A manufacturer of buttons wished to determine whether the fraction of defective buttons produced by three machines varied from machine to machine. Samples of 400 buttons were selected from each of the three machines, and the number of defectives were counted for each sample. The results are shown in the accompanying table. Do these data present sufficient evidence to indicate that the fraction of defective

buttons varied from machine to machine? Please perform this calculation by hand (you can use R or a calculator, but not just the `chisq.test` function in R). Please report:

- H_0 and H_A
- The value of the χ^2 test statistic.
- The degrees of freedom.
- The p -value.
- Your conclusion (accept / reject) at level $\alpha = 0.05$.

Machine Number	Good Buttons	Defective Buttons	Total Buttons
1	384	16	400
2	376	24	400
3	391	9	400
Total	1151	49	1200

Exercise 5

Returning to the button manufacturer, calculate three confidence intervals for $p_1 - p_2$, $p_1 - p_3$, $p_2 - p_3$, where p_i is the defective rate in Machine i . The confidence intervals can be calculated using the standard method for finding the confidence interval for comparing two proportions. Use a Bonferroni correction to calculate these confidence intervals so as to maintain a *family-wise confidence level* of 95%.