Assignment #3

Stat 332: Sampling and Experimental Design (Spring 2024)

Professor: Christian Boudreau　　　　　　　　Due at 11:59 pm on Saturday Jul 13, 2024

Instructions:

- Teamwork is allowed and encouraged; however, everyone must hand in their own assignment showing that they understood what they wrote. You are not allowed to collaborate with anyone who is not a Stat 332 classmate.
- Justify all your answers.
- One solution per question; if you provide multiple solutions, you will receive 1/2 the points of the inferior solution.
- Write legibly; if the TA or I have to guess at your answers, it's unlikely to be in your favour.
- Make sure to include your `R` code and relevant output for Q#3 and Q#4.
- Crowdmark:
  - You should soon received an automated email from Crowdmark saying that you can now submit/upload your solution. If you have not received the email by noon on Wednesday, please start by checking your spam or junk mail folder. If the email is not there, email me. Note that, even if you did not get the email, you should still be able to submit/upload your assignment by simply going to Crowdmark, and signing into your account.
  - To ensure smooth and efficient grading, please make sure to upload your solution separately for each question.
  - Crowdmark allows you to resubmit anytime before the due date; once the due date has passed, your submission is locked in and you will thus not be able to resubmit.
  - Late submissions will be accepted, but are subject to a 5% per hour penalty*; click here for more information on Crowdmark and late submissions.
  - Technical difficulties with Crowdmark:
    - Consult Crowdmark Help
    - Watch this short video about submitting an assignment on Crowdmark
    - E-mail me
    - If tragedy strikes and you cannot upload your assignment to Crowdmark, email it to me, so I have proof of when you completed your assignment (otherwise, the above mentioned lateness penalty will apply).
- Recall that each of you is allowed a one-time extension with a reduced lateness penalty. You have 3 option: 1) a 5% reduced penalty for an additional 24 hours, 2) a 10% reduced penalty for an additional 48 hours, and 3) a 0% reduced lateness penalty if you submit your assignment within 3 hours past the deadline. To take advantage of the first two options (i.e., 5% and 10%), you have to email me (at least) 12 hours before the original due date. Since this is meant to be a last minute thing, you do not need to email me in advance to take advantage of the 0% (3 hours) lateness penalty. However, since it counts towards your one-time reduced lateness penalty, you still need to email me me to inform me that you want to take advantage of this option.
- See course outline for more information on assignments, and how much each of them contributes to your overall grade.

*This is the default penalty, and is applied when there are no extenuating circumstances.

Question #1 (5 points)**:**

In Tutorial # 6, we proved that $\hat{\mu}$ and $\hat{\alpha}_i$ $(i = 1, \ldots, I)$ are maximum likelihood estimators (MLEs). Building on this, derive $\hat{\sigma}^2$, the MLE of σ^2, in a one-way ANOVA. Is $\hat{\sigma}^2$ the same as MSE?

Question #2 (10 points)**:**

This question is concerned with expressing ANOVA models as regression models.

(a) Let's consider a one-way ANOVA where $I = 2$ (i.e., where the factor being studied has 2 levels; e.g., gender, where the 2 levels are males and females), $n_1 = 5$ and $n_2 = 6$. Write the regression model corresponding to this one-way ANOVA model when the first parametrization is used. That is, write the corresponding $\boldsymbol{X}$ matrix, and Y and β vectors. Make sure to specify the dimensions of $\boldsymbol{X}$ and the lengths of Y and β.

(b) Same as (a), but when the second parametrization is used.

(c) Same as (a), but with $I = 4$, $n_1 = 3$, $n_2 = 4$, $n_3 = 5$ and $n_4 = 2$.

(d) Same as (c), but when the second parametrization is used.

$I=6 \quad n_i=20 \quad n=20\times 6=120$

Question #3 (15 points)**:**

A company uses six filling machines to place detergent into cartons that show a label weight of 32 ounces. The production manager has complained that the six machines do not place the same amount of fill into the cartons. A consultant requested that 20 filled cartons be selected randomly from each of the six machines and the content of each carton carefully weighed. The observations (stated for convenience as deviations from targeted 32.00 ounces) are given in the table below, and are also available on the STAT 332 LEARN website; see file `machines.txt`.

(a) Produce the ANOVA table.

(b) Test, at the 5% level, whether or not the mean fill differs between the 6 machines.

(c) Compute all pairwise comparisons and, using Tukey's method, obtain their CI's with a family level of 95%.

(d) Repeat par (c), but using the Bonferroni method for multiple comparisons. How do the CI's computed in part (d) compare with those obtained in part (c)? Which of the two multiple comparison methods is preferable here?

use 好

(e) Machines #1 and #2 were purchased new 5 years ago. Machines #3 and #4 were also purchased 5 years ago, but were purchased reconditioned instead of new. Lastly, Machines #5 and #6 were purchased new last year. Using contrast, and Scheffé method for multiple comparisons, test:

i) If machines purchased 5 years ago differ from those purchased last year.
ii) If machines purchase new differ from those purchased reconditioned.
iii) If the 2 machines purchased new 5 years ago differ than those purchased reconditioned.

(f) Check if the assumptions of the ANOVA models are satisfied. $1-\alpha/(2n)$ 120 因为 n_i 都一样，所以只有一组上下线

(e) i) A − B

$L = (\mu_1+\mu_2+\mu_3+\mu_4)/4 - (\mu_5+\mu_6)/2$

$= \sum_{i=1}^{I} c_i\mu_i$ where $c_1=c_2=c_3=c_4=\frac{1}{4}$, $c_5=c_6=-\frac{1}{2}$

95% CI for L: $\hat{L} \pm S_{1-\alpha}\sqrt{\boxed{\hat{V}(\hat{L})}}$, $\hat{V}(\hat{L}) = MSE\sum_{i=1}^{I}\frac{c_i^2}{n_i}$

$\hat{L} = \frac{\hat{\mu}_1+\hat{\mu}_2+\hat{\mu}_3+\hat{\mu}_4}{4} - \frac{\hat{\mu}_5+\hat{\mu}_6}{2}$

$S_{1-\alpha} = \sqrt{(I-1)F_{I-1,\,n-I,\,1-\alpha}}$, $I=6$, $I-1=5$, $n-I=114$, $1-\alpha=0.95$

如果CI包含0，那么A和B没差别.
不包含0，若且都大于0，那么A大于B
都小于0，那么A小于B } B和A有差别

ii) $L = \frac{\mu_1+\mu_2+\mu_5+\mu_6}{4} - \frac{\mu_3+\mu_4}{2}$

⋮

$L = \frac{\mu_1+\mu_2}{2} - \frac{\mu_3+\mu_4}{2}$

Machine					
#1	#2	#3	#4	#4	#6
-0.14	0.46	0.21	0.49	-0.19	0.05
0.20	0.11	0.78	0.58	0.27	-0.05
0.07	0.12	0.32	0.52	0.06	0.28
0.18	0.47	0.45	0.29	0.11	0.47
0.38	0.24	0.22	0.27	0.23	0.12
0.10	0.06	0.35	0.55	0.15	0.27
-0.04	-0.12	0.54	0.40	0.01	0.08
-0.27	0.33	0.24	0.14	0.22	0.17
0.27	0.06	0.47	0.48	0.29	0.43
-0.21	-0.03	0.62	0.34	0.14	-0.07
0.39	0.05	0.47	0.01	0.20	0.20
-0.07	0.53	0.55	0.33	0.30	0.01
-0.02	0.42	0.59	0.18	-0.11	0.10
0.28	0.29	0.71	0.13	0.27	0.16
0.09	0.36	0.45	0.48	-0.20	-0.06
0.13	0.04	0.48	0.54	0.24	0.13
0.26	0.17	0.44	0.51	0.20	0.43
0.07	0.02	0.50	0.42	0.14	0.35
-0.01	0.11	0.20	0.45	0.35	-0.09
-0.19	0.12	0.61	0.20	-0.18	0.05

Question #4 (15 points)**:** $n_0=7$ $n_1=10$ $n_2=10$ $n_3=5$

If you're an impatient pedestrian like myself, you push the "walk button" (i.e., button to activate the crossing signal) many times in a hope that the pedestrian crossing light/signal will quickly turn on. However, some reports suggest that many of these walk buttons are actually placebo buttons designed to give pedestrians the illusion of control whereas the crossing signal continues its operation as programmed. To investigate this, someone measured the waiting time (in seconds) after pushing the button 0, 1, 2 and 3 times. The 32 observations are given in the table below.

(a) Produce the ANOVA table.

(b) Test, at the 5% level, the null hypothesis that the number of pushes of the pedestrian button has no effect on the waiting time for the walk signal.

(c) Estimate the mean waiting time after pushing the button once and after pushing the button three times; give the corresponding 95% CIs for those two estimates.

(d) Compute the pairwise comparison between the two mean waiting times of part (c), and give the corresponding 95% CI using Tukey's method. $D_{42} = \mu_4 - \mu_2$ $\hat{D}_{42} = \hat{\mu}_4 - \hat{\mu}_2 = \bar{y}_{4\cdot} - \bar{y}_{2\cdot}$

$\bar{y}_{4\cdot} - \bar{y}_{2\cdot} \pm \frac{1}{\sqrt{2}} q_{I,n-I,1-\alpha} \sqrt{MSE(\frac{1}{n_4}+\frac{1}{n_2})}$

(e) Give a contrast to compare the effect of no pushes vs. pushing the button once or more. Estimate that constrast, and give the corresponding 95% CI. No need to make multiple comparison adjustment, as this is the only constrast of interest.

用9t.

$L = \mu_1 - \frac{\mu_2+\mu_3+\mu_4}{3}$

(Q3 用F 因为 Scheffe)

by using 7.5.3.

# of pushes			
0	1	2	3
38.14	38.28	38.17	38.14
38.20	38.17	38.13	38.30
38.31	38.08	38.16	38.21
38.14	38.25	38.30	38.04
38.29	38.18	38.34	38.37
38.17	38.03	38.34	
38.20	37.95	38.17	
	38.26	38.18	
	38.30	38.09	
	38.21	38.06	