

STAT 332 - Assignment 3

Due 9:00am Friday, November 9th - to be submitted through Crowdmark

Student Name: Your name here

Student ID: 123456789

General instructions: You may show your work using one or a combination of the following ways:

- Present the R commands you used to calculate the quantities indicated. We strongly recommend that you do so using ‘R Markdown’. An example of an ‘R Markdown’ document is provided in Learn.
- Type out work in ‘Latex’ or ‘R Markdown’.
- Present scan/photographs of handwritten work. If you choose this option ensure your work is clear and easy to read.

When you are using R (which we strongly encourage), you must present all R commands you used. Your answer must include **all** necessary commands to calculate the quantities indicated. Someone reading your work must be able to take your commands, copy them into R, and obtain identical results. Note that this includes the `set.seed()` and `sample()` commands that select your sample.

Note that for data analysis questions you must also clearly present your final answers in addition to the R commands you used.

START OF ASSIGNMENT

Question 1 [5 marks]

Suppose a finite population has four strata with sizes $N_1 = 3374$, $N_2 = 1960$, $N_3 = 52$ and $N_4 = 5$. A stratified simple random sample is drawn from the population and the sample data are summarized in the following table:

Stratum h	n_h	$\hat{\mu}_h$	$\hat{\sigma}_h^2$
1	100	491	921
2	50	1050	1278
3	45	2568	2343
4	5	18054	154054

- Compute the stratified mean estimate $\hat{\mu}_s$ and its estimated variance.
- Construct an approximate 95% confidence interval for the **population total** τ .

Question 2 [10 marks]

For the population of question 1, suppose that you wish to take another stratified sample of the population one year later. This time, you wish to take a sample of size 300 and estimate μ . Assume that unknown stratum quantities μ_h and σ_h^2 can be roughly approximated using the previous year’s survey.

- Calculate the stratum sample sizes under proportional allocation.
- Calculate the stratum sample sizes under optimal allocation.
- What are the values (approximately) of $Var(\tilde{\mu}_s)$ with sample size allocations of (a) and (b), respectively? Compare these two variances and comment.

Question 3 [15 marks]

An entirely fictional university is home to several cafes named “Feridonuts”. The owner of the cafes (known as Mr. Goose) would like to introduce a new promotion to boost profits. First, they decide to compare two new promotional schemes:

1. A range of juice drinks under the brand name “Goose Bar”.
2. A “\$9 Dinner” special, where for \$9 customers can buy a meal, drink, side, and dessert.

Mr. Goose is interested in whether the two promotions result in the same average profit per cafe. There are 30 Feridonuts cafes located around the university, and 20 of these are chosen via simple random sampling without replacement to be used in trials of the new promotions. Of the 20 stores, 10 were chosen to spend 1 week trialling the “Goose Bar” promotion, while the other 10 spent 1 week trialling the “\$9 Dinner” promotion. The total profits for each store are provided in the table below. The units are the fictional currency the Feridollar.

```
dataQ1 = matrix(NA, nrow=2, ncol=10)
dataQ1[1,] = c(100, 95, 125, 105, 100, 90, 135, 120, 85, 101)
dataQ1[2,] = c(90, 110, 85, 90, 95, 110, 115, 110, 105, 120)
rownames(dataQ1) = c("Goose Bar", "$9 Dinner")
dataQ1
```

##		[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]
##	Goose Bar	100	95	125	105	100	90	135	120	85	101
##	\$9 Dinner	90	110	85	90	95	110	115	110	105	120

- (a) Define the response, factor, and the factor levels in this study.
- (b) Write down the number of: (i) treatments (ii) replicates for each treatment.
- (c) Let y_{ij} equal the profit recorded at the j^{th} store that trialled promotion i . (“Goose Bar” for promotion 1, and “\$9 Dinner” for promotion 2.) We shall model these data as

$$Y_{ij} = \mu + \tau_i + R_{ij} \quad i = 1, 2 \quad j = 1, 2, \dots, 10 \quad R_{ij} \sim N(0, \sigma^2) \quad (1)$$

independently, subject to $\tau_1 + \tau_2 = 0$. Based on these data, calculate the following point estimates.

- (i) $\hat{\mu}$
 - (ii) $\hat{\tau}_1$
 - (iii) $\hat{\tau}_2$
 - (iv) The **pooled** variance estimate $\hat{\sigma}^2$
- (d) Construct a 95% confidence interval for $\tau_1 - \tau_2$: the difference in average weekly profits per store between the two promotions. You must show your working, but note that if you use R’s built-in t -test command you must **clearly** present your final confidence interval. Assume equal variances in both groups, and write your answer to 3 decimal places.
 - (e) Based on your answer to part (d), what conclusion would you reach concerning the original question Mr. Goose is interested in? Your answer must be phrased in formal statistical language, in particular making reference to a specific null hypothesis. You may assume all necessary model assumptions are satisfied.

- (f) The model specified above assumes that our R_{ij} terms are mutually independent. *Based on the information in this question*, discuss whether you think this is a reasonable assumption.

END OF ASSIGNMENT