# Exercise 2: Sampling

For this assignment, complete the dataset, and then answer the questions to the best of your ability.

(adapted from CH4 ex.12)  
A college has N=150 faculty members. The dean wants to conduct a faculty salary survey selecting   
systematic sample of n=20 faculty. The list of faculty members appears on page 147. I have begun an Excel file, included with this problem set, with the first 100 entries filled in.

**Dataset**

First, complete the dataset by adding the final 50 entries from page 147. The following coding was used:

Division: 1= Eng&Prof 2= Medicine 3= Lit&SocSci 4= Biol&Sci  
Sex: 1= Male 2=Female

**Questions**

Then, answer the following questions:  
1) What should be the sampling interval for this design? (1 pt)

The ratio of population to sample size for this dataset is 7.5. The sampling interval will be rounded down to 7 and I will continue counting and selecting each kth element from the random starting point until 20 elements are chosen.   
2) What are the possibilities for the first random start? (1 pt)

The probabilities for the first random start are 1, 2, 3, 4, 5, 6, and 7.   
3) Draw a sample of 20 faculty members (2pts)

Using the random start (RS) of 1, the elements chosen for the n=20 sample through systematic selection are elements 1, 8 , 15, 22, 29, 36, 43, 50, 57, 64 , 71, 78, 85, 92, 99, 106, 113, 120, 127, and 134. The R code used to obtain this sample is as follows. Tables for each of the three samples produced for this exercise are included at the end of the document.

```{r}

library(readxl)

```

```{r}

salary <- read\_xlsx('Faculty\_Salaries.xlsx')

```

```{r}

set.seed(42)

start <- sample(1:7, 1)

sample\_indices = seq(start, by = 7, length.out = 20)

sample <- salary[salary$no %in% sample\_indices,]

```

4) Estimate the mean salary for the faculty using your n=20 sample. (3 pts)

The mean salary for the faculty in the sample was $78,050 (SD = $32,500, n=20). The confidence intervals for this estimate are $63,800 and $92,300.   
5) What is the variance for that mean? (3 pts)  
The variance for this mean estimate is approximately $1,058,050.

Next, the dean of the school of medicine suggested that their professors were not paid as well as the other divisions’ professors and wondered if you could draw a sample to see if that is true.  
6) Draw a stratified sample from the dataset, using a sample size n=10, for each of the four  
divisions. (2 pts)

```{r}

set.seed(42)

library(tidyverse)

strat\_sample <- salary %>% group\_by(Division) %>%

sample\_n(size = 10)

```

Results are shown in Table 2 in the end of the document.

7) Show your estimated mean salary for each division. (4 pts)  
Faculty in the “Eng&Prof” division earn $76,600 on average while those in the medicine division have the highest mean earnings at $86,300. Faculty in the Literary and Social Sciences division have the lowest average earnings at $69,300 and Biology/Science faculty had an average salary of $77,300. However, the results of an one-way ANOVA test show that differences in salary among divisions are not significant, p = 0.646. This suggests that faculty division is not a major factor of salary earnings, or that the sample size of n=10 from each division was insufficient to detect any effects. The R code used to obtain these means is provided below.

```{r}

strat\_sample %>% group\_by(Division) %>%

summarize(mean = mean(salary))

anova <- aov(salary ~ as.factor(Division),

data = strat\_sample)

summary(anova)

```

Finally, some of the female professors said the real issue wasn’t divisional but rather that women were not paid as much as their male counterparts. They asked you to check the data and see if they are correct.  
8) Draw a stratified sample size n=20 for both sexes (2 pts)

```{r}

set.seed(42)

sex\_sample <- salary %>% group\_by(sex) %>%

sample\_n(size = 20)

```

Results are shown in Table 3 at the end of the document.

9) Show your estimated mean salary by sex. (2 pts)

An independent samples t-test is used to compare salary earnings of male and female faculty. The results show a significant difference in earnings between male faculty (M = $81,200, SD = $24,500) and female faculty (M = $51,850, SD = $18,450) at the 0.05 level. The confidence intervals show that there is a 95% chance that the true difference in mean earnings is between $15,500 and $43,200. Thus, there seems to exist a systematic earnings disparity among college faculty based on the gender of staff.

```{r}

sex\_sample %>% group\_by(sex) %>%

summarize(mean = mean(salary), sd = sd(salary))

sex\_t <- t.test(salary ~ sex, data = sex\_sample, var.equal = T)

print(sex\_t)

```

**Figures**

## Table 1: Systematic Sample Table

| no | Division | sex | rank | salary |
| --- | --- | --- | --- | --- |
| 1 | 1 | 1 | 3 | 88 |
| 8 | 4 | 1 | 3 | 47 |
| 15 | 2 | 1 | 1 | 181 |
| 22 | 2 | 2 | 1 | 106 |
| 29 | 2 | 1 | 2 | 79 |
| 36 | 2 | 2 | 2 | 53 |
| 43 | 2 | 2 | 3 | 32 |
| 50 | 2 | 1 | 2 | 114 |
| 57 | 2 | 1 | 3 | 64 |
| 64 | 1 | 2 | 1 | 86 |
| 71 | 3 | 1 | 1 | 91 |
| 78 | 1 | 1 | 2 | 49 |
| 85 | 2 | 1 | 1 | 47 |
| 92 | 1 | 1 | 1 | 85 |
| 99 | 3 | 1 | 2 | 50 |
| 106 | 2 | 1 | 1 | 93 |
| 113 | 1 | 1 | 1 | 72 |
| 120 | 2 | 1 | 2 | 87 |
| 127 | 1 | 1 | 2 | 65 |
| 134 | 1 | 1 | 1 | 72 |

## Table 2: Stratified Sample By Division

| no | Division | sex | rank | salary |
| --- | --- | --- | --- | --- |
| 1 | 1 | 1 | 3 | 88 |
| 109 | 1 | 1 | 2 | 121 |
| 45 | 1 | 1 | 3 | 43 |
| 18 | 1 | 1 | 1 | 91 |
| 75 | 1 | 1 | 2 | 71 |
| 113 | 1 | 1 | 1 | 72 |
| 67 | 1 | 2 | 3 | 44 |
| 60 | 1 | 1 | 3 | 58 |
| 103 | 1 | 1 | 1 | 114 |
| 34 | 1 | 1 | 2 | 64 |
| 81 | 2 | 2 | 3 | 42 |
| 43 | 2 | 2 | 3 | 32 |
| 57 | 2 | 1 | 3 | 64 |
| 115 | 2 | 1 | 2 | 167 |
| 105 | 2 | 1 | 1 | 112 |
| 4 | 2 | 1 | 1 | 133 |
| 87 | 2 | 1 | 3 | 83 |
| 55 | 2 | 1 | 3 | 84 |
| 59 | 2 | 2 | 1 | 87 |
| 77 | 2 | 2 | 2 | 59 |
| 142 | 3 | 2 | 3 | 46 |
| 71 | 3 | 1 | 1 | 91 |
| 25 | 3 | 1 | 1 | 61 |
| 99 | 3 | 1 | 2 | 50 |
| 9 | 3 | 2 | 3 | 39 |
| 122 | 3 | 1 | 1 | 79 |
| 40 | 3 | 1 | 2 | 60 |
| 12 | 3 | 1 | 1 | 62 |
| 6 | 3 | 1 | 1 | 113 |
| 47 | 3 | 1 | 1 | 92 |
| 89 | 4 | 1 | 1 | 78 |
| 118 | 4 | 1 | 1 | 107 |
| 41 | 4 | 2 | 1 | 73 |
| 79 | 4 | 1 | 1 | 83 |
| 52 | 4 | 1 | 1 | 49 |
| 32 | 4 | 1 | 1 | 103 |
| 98 | 4 | 1 | 1 | 85 |
| 135 | 4 | 2 | 3 | 36 |
| 10 | 4 | 1 | 1 | 74 |
| 37 | 4 | 1 | 1 | 85 |

## Table 3: Stratified Sample By Sex

| no | Division | sex | rank | salary |
| --- | --- | --- | --- | --- |
| 57 | 2 | 1 | 3 | 64 |
| 123 | 2 | 1 | 2 | 113 |
| 82 | 2 | 1 | 2 | 97 |
| 30 | 2 | 1 | 1 | 61 |
| 93 | 3 | 1 | 1 | 53 |
| 122 | 3 | 1 | 1 | 79 |
| 21 | 4 | 1 | 2 | 55 |
| 150 | 1 | 1 | 1 | 96 |
| 55 | 2 | 1 | 3 | 84 |
| 29 | 2 | 1 | 2 | 79 |
| 89 | 4 | 1 | 1 | 78 |
| 145 | 3 | 1 | 1 | 90 |
| 109 | 1 | 1 | 2 | 121 |
| 45 | 1 | 1 | 3 | 43 |
| 24 | 2 | 1 | 3 | 79 |
| 31 | 2 | 1 | 1 | 86 |
| 4 | 2 | 1 | 1 | 133 |
| 49 | 4 | 1 | 2 | 57 |
| 131 | 1 | 1 | 3 | 53 |
| 32 | 4 | 1 | 1 | 103 |
| 22 | 2 | 2 | 1 | 106 |
| 26 | 3 | 2 | 3 | 37 |
| 67 | 1 | 2 | 3 | 44 |
| 143 | 3 | 2 | 1 | 57 |
| 94 | 2 | 2 | 3 | 40 |
| 5 | 1 | 2 | 2 | 71 |
| 43 | 2 | 2 | 3 | 32 |
| 9 | 3 | 2 | 3 | 39 |
| 2 | 2 | 2 | 3 | 45 |
| 61 | 2 | 2 | 3 | 39 |
| 63 | 2 | 2 | 2 | 46 |
| 76 | 2 | 2 | 3 | 42 |
| 135 | 4 | 2 | 3 | 36 |
| 144 | 2 | 2 | 2 | 50 |
| 81 | 2 | 2 | 3 | 42 |
| 36 | 2 | 2 | 2 | 53 |
| 136 | 3 | 2 | 1 | 66 |
| 142 | 3 | 2 | 3 | 46 |
| 64 | 1 | 2 | 1 | 86 |
| 141 | 3 | 2 | 1 | 60 |