

Sampling Report

Khang Thai

2025-03-14

Introduction

Research Question

The purpose of this study is to estimate the proportion of male pedestrians on Bruin Walk. The purpose of this research is to find the demographics of all students, faculty, and visitors that walk on Bruin Walk everyday.

Target Population, Sampling Frame, and Sampling Units

Target Population

The target population for this study includes all individuals who walk through Bruin Walk. This includes students, faculty, staff, and visitors at UCLA.

Sampling Frame

The sampling frame would be the group of individuals observed during a specific time period on Bruin Walk.

Sampling Unit

The Sampling Unit would be the individual pedestrian who pass through the observation point.

Enumerating the Sampling Frame

The sample population (n) was estimated using a 95% confidence interval with a 60% proportion and a margin of error of 0.07:

$$n_0 = \frac{(1.96)^2(0.6)(1-0.6)}{(0.07)^2} = \frac{0.921984}{0.0049} = 188.16$$

$$n_0 = 188$$

After applying the Finite Population Correction:

$$n = \frac{188}{1 + \frac{188-1}{4000}} = 180$$

The total population (N) of pedestrians on Bruin Walk is estimated to be 4000 individuals per day. Using a systematic sampling approach, individuals selected at a fixed interval (k) was calculated by:

$$k = \frac{N}{n} = \frac{4000}{180} = 22$$

Every 22nd person walking through Bruin Walk was observed and recorded.

Sampling Plan and Methodology

Systematic random sampling was used to estimate the proportion of male pedestrians on Bruin Walk. Collecting data was taken at fixed location on Bruin Walk observing every 22nd pedestrian passing through the observation point. 180 individuals will be observed and will then be classified as either male or female. One strength about this sampling plan is that since systematic sampling is used, taking sample is feasible because it does not require continuous observation. It is also able to collect a reasonable estimate without needing to full enumeration. One weakness is that there could be potential bias depending on the time the sample was taken. Another issue is that there could be potential pattern in pedestrian flow, so the sample might not be entirely random. Some possible improvements would be considering collecting data over a longer period of time that way it would improve the representation of the sample.

Method of Analysis

Proportion estimation will be used to find the proportion of males in the sample:

$$\hat{p} = \frac{\text{number of males observed}}{\text{total pedestrians observed}}$$

The sampling variance will then be estimated: $Var(\hat{p}) = \frac{\hat{p}(1-\hat{p})}{n} * \frac{N-n}{N-1}$

where: $\hat{p} = 0.51$, $n = 180$, $N = 4000$

Population Parameter

```
observed_male <- 92
tot_sample_size <- 180
p_hat <- observed_male / tot_sample_size
p_hat
```

```
[1] 0.5111111
```

Sampling Variance

```
N <- 4000
n <- 180

var_p <- (p_hat * (1 - p_hat) / n) * ((N - n) / (N - 1))

var_p
```

```
[1] 0.001326065
```

Confidence Interval

```
# Standard Error
SE <- sqrt(var_p)
SE
```

```
[1] 0.03641518
```

```
# Lower-Upper Bound
lower_CI <- p_hat - 1.96 * SE
upper_CI <- p_hat + 1.96 * SE

lower_CI
```

```
[1] 0.4397374
```

```
upper_CI
```

```
[1] 0.5824849
```

Results

Gender	Count	Proportion
Male	92	0.51
Female	88	0.49
Total	180	1.00

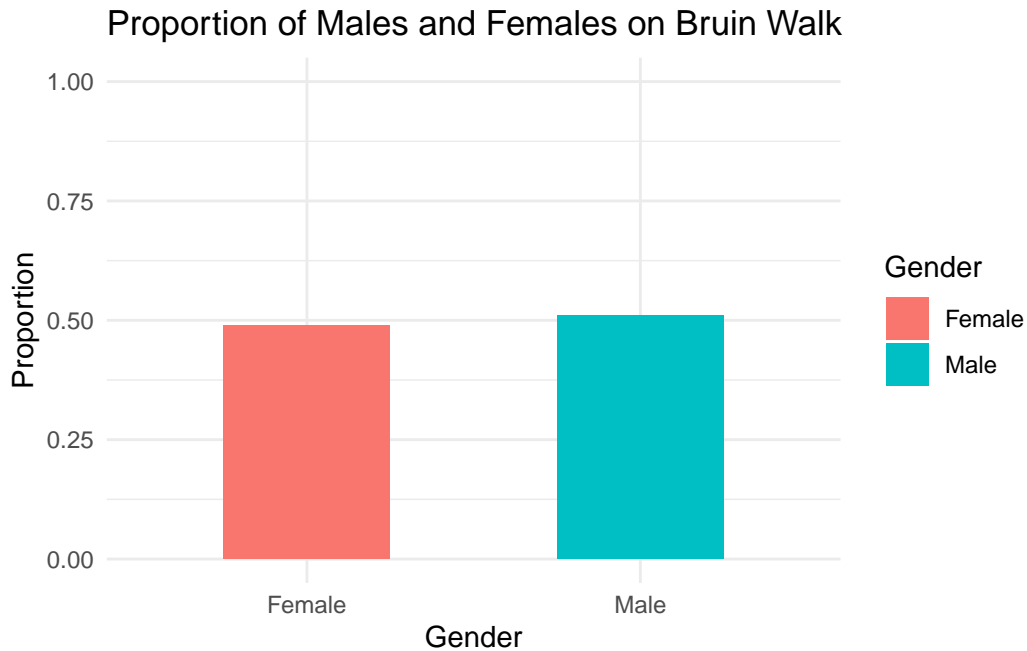
Table 1: Gender Distribution of Pedestrians on Bruin Walk

```
library(ggplot2)
```

Warning: package 'ggplot2' was built under R version 4.3.3

```
data <- data.frame(
  Gender = c("Male", "Female"),
  Proportion = c(0.51, 0.49)
)

ggplot(data, aes(x = Gender, y = Proportion, fill = Gender)) +
  geom_bar(stat = "identity", width = 0.5) +
  labs(title = "Proportion of Males and Females on Bruin Walk",
       x = "Gender",
       y = "Proportion") +
  theme_minimal() +
  scale_y_continuous(limits = c(0, 1))
```



Findings

Based on the collected data, the estimated proportion of male pedestrians on Bruin is 0.51, with a 95% confidence interval ranging from 43.97% to 58.25%. This suggests that the proportion of male pedestrians on Bruin Walk is likely in this range. The finite population correction was applied to improve the accuracy of the variance estimation with a total estimated population of $N = 4000$ and sample size of $n = 180$.

Strengths

Using systematic sampling ensured that pedestrians were selected at fixed intervals, thus reducing the risk of over-sampling individuals from specific demographic groups. Additionally, including finite population correction helps reduce the estimated variance, leading to a more precise confidence interval.

Limitations

Some limitations that were encountered during the sampling process were that the study only sampled individuals at a specific time, which may not capture the full demographic diversity of Bruin Walk. Another limitation is that the flow of pedestrian follows a pattern where there

are more students on Bruin Walk near the hour because classes typically starts and end on the hour. This means that using systematic sampling might not be truly random.

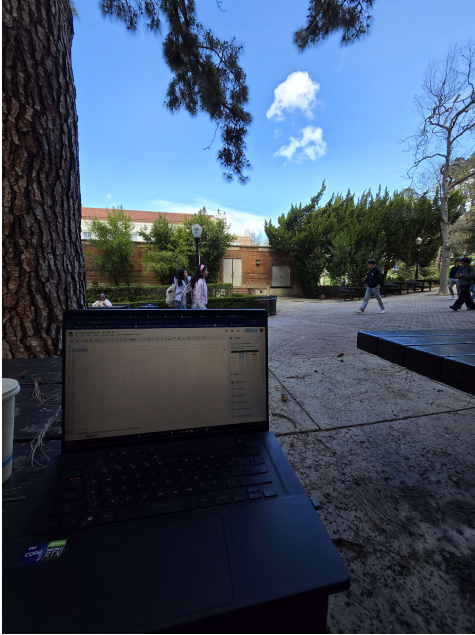


Figure 1: 1:05 pm



Figure 2: 2:14 pm

The raw data collected for this study can be downloaded here:

[Download CSV File](#)