Monte Carlo Analysis of Ant Nest Counts in Field and Forest Habitats

EA030 Reflection

October 8, 2025

Introduction

In ecological research, we often want to understand whether observed patterns, such as the number of ant nests in different habitats, are **statistically unusual** or could have occurred by chance. Traditional parametric tests may not always be appropriate, especially with small sample sizes or non-normal distributions.

Monte Carlo simulations provide a flexible, non-parametric approach to test hypotheses by *randomly reshuffling data* many times to create a null distribution. By comparing observed statistics (like the mean number of ant nests in fields and forests) to this null distribution, we can assess the probability that the observed pattern arose by chance.

In this exercise, we analyze ant nest counts across two habitat types, **Field** and **Forest**, using Monte Carlo simulations to estimate expected mean and standard deviation under random redistribution of nests.

Learning Goals

By the end of this activity, learners will be able to:

1. Understand Monte Carlo simulations as a non-parametric method for hypothesis testing.
2. Compute summary statistics (mean and standard deviation) for ecological count data.
3. Perform Monte Carlo randomizations to generate null distributions of ecological data.
4. Compare observed statistics to null distributions to estimate tail probabilities (p-values).
5. Visualize simulation results using histograms with observed values highlighted.
6. Interpret the ecological significance of observed patterns in the context of habitat differences.

Method & Analysis Workflow

Step 1: Input Data

- Collect ant nest counts per quadrat.
- Record the habitat type (Field or Forest) for each quadrat.

We can enter the data into R as follows:[1].

```
# --------------------------
# Step 1: Input ant nest counts
# --------------------------
nest_counts <- data.frame(
  Habitat = c("Forest", "Forest", "Forest", "Forest", "Forest",
              "Forest", "Field", "Field", "Field", "Field"),
  Nests = c(9,6,4,6,7,10,12,9,12,10)
)
```

[1] Alternatively, you can use the an-combe data that is built into R by typing `data("anscombe")`. Note that the dataframe name will differ from what I have used in the handout!

Step 2: Compute Observed Statistics

- Calculate the mean and standard deviation of ant nests for Field and Forest habitats.

We can use a data wrangling function... Google how to use dplyr() and see if you can replace the question marks with a productive result as noted below!

```
library(dplyr)
observed_stats <- ?? %>%
  group_by(??) %>%
  summarise(
    Mean = mean(??),
    SD = sd(??)
```

```
## # A tibble: 2 x 4
##   Habitat     N  Mean    SD
##   <chr>   <int> <dbl> <dbl>
## 1 Field       4  10.8  1.5
## 2 Forest      6   7    2.19
```

Let's order the data.frame by habitat

```
nest_counts <- nest_counts[order(nest_counts$Habitat), ]
nest_counts

##    Habitat Nests
## 7    Field    12
## 8    Field     9
## 9    Field    12
## 10   Field    10
```

```
## 1   Forest     9
## 2   Forest     6
## 3   Forest     4
## 4   Forest     6
## 5   Forest     7
## 6   Forest    10
```

What's the difference between forest mean and field mean for ant colonies?[2]

[2] Let's think about that magnigitude of the difference as an absolute value.

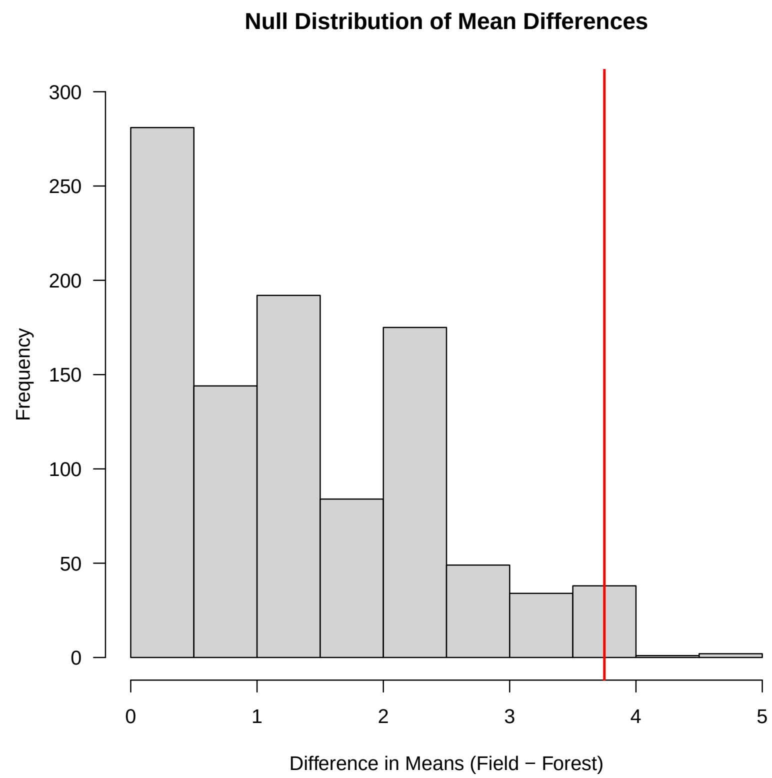
```
## [1] 3.75
```

Step 3: Monte Carlo Randomization

- Randomly shuffle nest counts across quadrats 1000 times (or more).
- For each randomization, compute the mean (and SD) for Field and Forest habitats.
- Store the simulated means (and SDs) to create null distributions.

Step 4: Visualize the Null Distribution

```
# Make a histogram of the simulated differences
par(las=1)
hist(dif_sim, main="Null Distribution of Mean Differences",
     xlab="Difference in Means (Field - Forest)", ylim=c(0, 300))
abline(v=observed_diff, col="red", lwd=2)
```



Step 5: Count the Number of Simulated Difference Above the Observed Difference

- Count how many times the simulated difference in means is greater than or equal to the observed difference.

- Divide this count by the total number of simulations to get the p-value.

```
# Count the number of simulated difference are above the observed difference
p_value_diff <- sum(dif_sim >= observed_diff) / nsimul
noise_diff <- sum(dif_sim < observed_diff) / nsimul
p_value_diff

## [1] 0.041
```

Step 6: Interpretation

- Compare observed means and SDs to the null distributions.
- Assess whether ant nest densities in each habitat are significantly higher or lower than expected by random chance.

Reflection Questions

1. What does a Monte Carlo simulation tell us about the likelihood of observed patterns compared to random expectation?
2. How do the observed mean nest counts in Field and Forest compare to their respective simulated null distributions? What does this imply?
3. Why might a Monte Carlo approach be preferred over traditional parametric tests in ecological datasets with small sample sizes or skewed distributions?
4. How might the interpretation of results change if we examined standard deviation rather than the mean?
5. Can you think of other ecological questions or datasets where Monte Carlo simulations could provide insight?