Lab 2 Demo: Random grid sites

STAT 109: Introductory Biostatistics

Lab 2 Demo: Random grid sites

Goal: Start Lab 2 with a probability question on an **equally likely** sample space, then show how to compute it by **counting** and by **simulation** in R.

Warm-up: a 1D "grid" (5×1 line of sites)

Before the 2D grid, we practice the same ideas on a 1D line of 5 sites.

We label sites as:

$$S = \{1, 2, 3, 4, 5\}$$

We choose 2 distinct sites **without replacement**, so the equally likely sample space is all 2-site sets:

$$|S_2| = \binom{5}{2} = 10$$

1D Question A: at least one end

The "end sites" are 1 and 5.

Use the complement ("no ends" means both picks are from {2,3,4}):

$$P(\text{at least one end}) = 1 - \frac{\binom{3}{2}}{\binom{5}{2}} = 1 - \frac{3}{10} = \frac{7}{10}$$

1D Question B: adjacent

Adjacent pairs are:

- {1,2}, {2,3}, {3,4}, {4,5} (4 pairs total)

So:

$$P(\text{adjacent}) = \frac{4}{\binom{5}{2}} = \frac{4}{10} = 0.4$$

1D simulation (for loop, without replacement)

```
set.seed(109)

sites <- 1:5
B <- 10000

at_least_one_end <- logical(B)
adjacent <- logical(B)

for (b in 1:B) {
  draws <- sample(sites, size = 2, replace = FALSE)
```

```r
  a <- draws[1]
  c <- draws[2]

  # "at least one end" using the Lab 1 sum idea
  at_least_one_end[b] <- sum(c(a == 1, a == 5, c == 1, c == 5)) > 0

  # adjacent on a line if they differ by 1
  adjacent[b] <- (abs(a - c) == 1)
}

mean(at_least_one_end)
mean(adjacent)
```

Setup (2D grid; equally likely sample space)

We have an $N \times N$ grid of possible sampling sites. Each site is equally likely.

For this demo, we will keep things realistic and simple:

- We select **2 distinct sites without replacement** from the grid.
- Every unordered pair of sites is equally likely.

So the sample space size is:

$$|S| = \binom{N^2}{2}$$

- Total sites: N^2
- Corner sites: 4
- Edge sites (including corners): $4N - 4$

Define:

- $p_{\text{corner}} = \frac{4}{N^2}$
- $p_{\text{edge}} = \frac{4N-4}{N^2}$

What Rosanna will demo

- **Types**
  - numeric vs integer vs character vs logical
  - `class(x)` (and optionally `typeof(x)`)
- **Vectors**
  - `c(...)` to create vectors
  - `length(x)` to check size
  - bracketing `x[i]` and bracketing with a logical vector `x[x > 5]`
  - element-by-element operations: `x + 1`, `x * 2`, `x == y`
- **Matrices**
  - `matrix(...)` to create a matrix
  - `dim(m)` to check size
  - bracketing `m[row, col]`, extracting rows/cols with `m[1, ]`, `m[, 2]`
- **Simulation + loops**

- `sample(..., replace = FALSE)` (choose distinct sites)
- `%in%` for "is it in this set?"
- placeholders: `logical(B)`, `numeric(B)`
- `for (b in 1:B) { ... }` storing one result per run
- probability by simulation: `mean(TRUE/FALSE)`

Quick warm-up code for types + structures

```
# Types
x_num <- 3.5
x_int <- 3L
x_chr <- "site"
x_log <- (3 < 5)

class(x_num); class(x_int); class(x_chr); class(x_log)

# Vector + length + scalar operations
v <- c(2, 4, 6, 8, 10)
length(v)
v + 1
v * 2

# Matrix + dim + bracketing
m <- matrix(1:6, nrow = 2, ncol = 3)
dim(m)
m[2, 2]
m[1, ]
m[, 2]
```

---

1) Probability of at least one corner

We are choosing 2 distinct sites from N^2 equally likely sites.

There are 4 corners and $N^2 - 4$ non-corners.

Use the complement ("no corners"):

$$P(\text{at least one corner}) = 1 - P(\text{no corners}) = 1 - \frac{\binom{N^2-4}{2}}{\binom{N^2}{2}}$$

---

2) Probability of at least one edge

Edges (including corners) total $4N - 4$ sites, so non-edges total:

$$N^2 - (4N - 4) = (N - 2)^2$$

Use the complement ("no edges"):

$$P(\text{at least one edge}) = 1 - P(\text{no edges}) = 1 - \frac{\binom{(N-2)^2}{2}}{\binom{N^2}{2}}$$

---

3) Probability the two sites are adjacent (connected)

We will say two sites are adjacent if they are 4-neighbors (up/down/left/right).

Counting (theory)

In an $N \times N$ grid:

- Horizontal adjacent pairs: $N(N-1)$
- Vertical adjacent pairs: $N(N-1)$

So total adjacent (unordered) pairs:

$$2N(N-1)$$

Total possible unordered pairs of sites:

$$\binom{N^2}{2}$$

Therefore:

$$P(\text{adjacent}) = \frac{2N(N-1)}{\binom{N^2}{2}}$$

Plug in numbers (quick numeric example)

Pick values that are easy to talk through in class. For example:

- $N = 5$ (a 5×5 grid, 25 sites)

Then:

- Total pairs: $\binom{25}{2} = 300$
- Adjacent pairs: $2N(N-1) = 2 \cdot 5 \cdot 4 = 40$

We can compute the three probabilities in R.

```r
N <- 5
total_pairs <- choose(N^2, 2)

# 1) at least one corner
p_at_least_one_corner <- 1 - choose(N^2 - 4, 2) / total_pairs

# 2) at least one edge
p_at_least_one_edge <- 1 - choose((N - 2)^2, 2) / total_pairs

# 3) adjacent
p_adjacent <- (2 * N * (N - 1)) / total_pairs

p_at_least_one_corner
p_at_least_one_edge
p_adjacent
```

Simulation in R (for loops, without replacement)

We simulate B random samples of 2 distinct sites, and estimate probabilities by proportions (`mean(TRUE/FALSE)`).

We label each grid cell with a string `"i,j"` so the sampled sites look like coordinates you'd write by hand.

New-to-Lab-1/2 note: this uses `paste()` to create `"i,j"` labels and `strsplit()` + `as.numeric()` to turn `"i,j"` back into numbers for the adjacency check.

```r
set.seed(109)

N <- 5
B <- 10000

# Make a grid whose entries look like "i,j" (row,col)
grid <- matrix("", nrow = N, ncol = N)
for (i in 1:N) {
  for (j in 1:N) {
    grid[i, j] <- paste(i, j, sep = ",")
  }
}

corners <- c(grid[1,1], grid[1,N], grid[N,1], grid[N,N])
edges   <- c(grid[1, ], grid[N, ], grid[2:(N-1), 1], grid[2:(N-1), N])

at_least_one_corner <- logical(B)
at_least_one_edge   <- logical(B)
adjacent            <- logical(B)

for (b in 1:B) {
  picks <- sample(grid, size = 2, replace = FALSE)
  s1 <- picks[1]
  s2 <- picks[2]

  rc1 <- as.numeric(strsplit(s1, ",")[[1]])
  rc2 <- as.numeric(strsplit(s2, ",")[[1]])

  r1 <- rc1[1]; c1 <- rc1[2]
  r2 <- rc2[1]; c2 <- rc2[2]

  # keep the Lab 1 "sum of TRUE/FALSE" idea (avoid new | and & operators)
  at_least_one_corner[b] <- sum(c(s1 %in% corners, s2 %in% corners)) > 0
  at_least_one_edge[b]   <- sum(c(s1 %in% edges,   s2 %in% edges))   > 0

  same_row_adj <- (r1 == r2) * (abs(c1 - c2) == 1)
  same_col_adj <- (c1 == c2) * (abs(r1 - r2) == 1)
  adjacent[b] <- sum(c(same_row_adj, same_col_adj)) > 0
}

mean(at_least_one_corner)
mean(at_least_one_edge)
mean(adjacent)
```

Teaching notes (how this sets up Lab 2)

- **Equally likely sample space**: every 2-site sample (without replacement) is equally likely.
- **Counting**: corners $= 4$; edges $= 4N - 4$; total sites $= N^2$; total pairs $= \binom{N^2}{2}$.
- **Matrices + bracketing**: `grid <- matrix("", ...)`, `corners <- c(grid[1,1], ...)`, etc.
- **For loop**: pre-allocate storage (`logical(B)`) and store one result per iteration.
- **Probability by simulation**: `mean(TRUE/FALSE)` gives a proportion.