Birthday Problem

BIOF2014

Problem

What is the probability that at least two people in this classroom share the same birthday?

Model

Assumptions

- 365 days in a year
- Every birthday is equally likely

Variables

Let N be the number of people in this class.
Let x be the number of people with the same birthday.

Task

1. Derive a mathematical equation for $P(x \geq 2)$.
2. Implement this equation and compute $P(x \geq 2)$.

Hint

$$P(x \geq 2) \;=\; P(x > 1) \;=\; 1 - P(x = 1).$$

Simulation

```
set.seed(1234);

# Simulate birthdays and return the number of people who share the same birthday
# N  number of people
# J  number of possible birthdays
simulate <- function(N, J) {
  birthdays <- sample.int(J, N, replace=TRUE);
  counts <- table(birthdays);
  max(counts)
}

N <- 30;
J <- 365;
S <- 1000;  # number of simulation rounds
x <- vapply(1:S, function(s) simulate(N, J), 0);
event <- as.integer(x > 1);
estimate <- mean(event);
se <- sd(event) / sqrt(S);
z <- qnorm(1 - 0.05 / 2);
```

estimate: 0.712
standard error: 0.0143269
95% confidence interval: (0.6839197, 0.7400803)

ground truth: 0.7063162

```
library(ggplot2)
ggplot(data.frame(x=x), aes(x)) + theme_classic() +
  geom_histogram(binwidth=1) + ylab("count") + ylim(0, S)
```