Integer Sequences - Statistical Analysis

E. Ajao, M. C. Gîrjău, A. Lee

We are investigating whether there is any significant difference in the performance/accuracy of UMAD with single-point crossover and UMAD with no crossover.

```r
# read in data
data <- readRDS("data/clean.Rds")



data %>%
  group_by(sequence, has_crossover) %>%
  summarize(total_runs = n(),
            success_rate = mean(found_solution),
            avg_generations = mean(generations_taken),
            avg_runtime = mean(time_taken)) %>%
  kable(booktabs = TRUE, digits = 2, linesep = "", align = "c") %>%
  kable_styling(latex_options = "hold_position")
```

sequence	has_crossover	total_runs	success_rate	avg_generations	avg_runtime
:A000292	FALSE	100	0.00	300.00	4272747.04
:A000292	TRUE	100	0.00	300.00	5253421.36
:A037270	FALSE	100	0.49	228.76	5049518.51
:A037270	TRUE	100	0.48	229.60	3670713.52
:A114241	FALSE	100	0.00	300.00	2649490.87
:A114241	TRUE	100	0.00	300.00	2870666.38
:A168392	FALSE	100	0.00	300.00	3834728.97
:A168392	TRUE	100	0.06	291.34	4807402.77
:simple	FALSE	100	1.00	14.45	61252.36
:simple	TRUE	100	1.00	9.65	33147.59

```r
prop.test(c(48 + 6 + 100, 49 + 100), c(1000, 1000), alternative = "g") # overall
```

```
    2-sample test for equality of proportions with continuity correction

data:  c(48 + 6 + 100, 49 + 100) out of c(1000, 1000)
X-squared = 0.062234, df = 1, p-value = 0.4015
alternative hypothesis: greater
95 percent confidence interval:
 -0.02237327  1.00000000
sample estimates:
prop 1 prop 2
 0.154  0.149
```

```r
prop.test(c(6, 0), c(100, 100), alternative = "g") # just for :A16839
```

```
    2-sample test for equality of proportions with continuity correction

data:  c(6, 0) out of c(100, 100)
X-squared = 4.2955, df = 1, p-value = 0.01911
alternative hypothesis: greater
95 percent confidence interval:
 0.01093689 1.00000000
sample estimates:
prop 1 prop 2
  0.06   0.00
```