Testing

05 March, 2019

What I did

I wanted to simulate balance comparisons between treatment groups in randomized experiments and in blocked experiments. So I wanted to draw samples of random numbers, repeat those draws many times, randomize treatment for each drawn sample, block treatment for the very same samples, then plot the differences all of the randomized and blocked treatment group distributions.

Simulations Code (repeats for 2, 3, 5, 10 groups and saves .csv files – run on Jeff's machine)

The final simulations code can specify any number of treatment groups and any number of repeats. I can also specify any number of concatenated or sequenced sampled numbers, as long as the total number and sequence is dividable by the number of treatment groups. I can also specify any number of concatenated or sequenced variable levels.

I ran the code below on Jeff's machine. It creates up to 1000 sequenced sampled numbers with 100 repeats for one level (5) in four different loops for 2, 3, 5, and 10 treatment groups.

It takes about 3-4 days to run this code. I looked into ways to speed up the code: Rcpp is an obvious candidate, but I need to read into that more to apply it here.

Plotting Code

I can easily do scatterplots, but they are so dense they're essentially unusable. They're also not very meaningful for the blocked version (and jitter() doesn't help much). Boxplots and transparent/hued histograms are much better here.

The code below reads in the created simulation .csv files and saves a boxplots and a histogram. I adapted this in the actual diss to be produced by markdown, rather than saved.

Simulation for 650 Respondents and 5 Groups (Matt wanted to see this)

Run everything for 650 people and 5 treatment groups, plot it, then test for statistical significance between the diff means of blocked and randomized assignments. I ran the actual simulation on Jeff's machine (faster), then copied the code (adjusting for saving directories) and dragged the results over to my machine.

```
t.test(all_blocked$diff,all_means_variances$diff,alternative="two.sided")
```

```
##
##  Welch Two Sample t-test
##
## data:  all_blocked$diff and all_means_variances$diff
## t = -28.271, df = 101.2, p-value < 2.2e-16
```

```
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.2611598 -0.2269138
## sample estimates:
##  mean of x  mean of y
## 0.02230769 0.26634450
```

```
#       Welch Two Sample t-test

# data:  all_blocked$diff and all_means_variances$diff
# t = -28.271, df = 101.2, p-value < 2.2e-16
# alternative hypothesis: true difference in means is not equal to 0
# 95 percent confidence interval:
#  -0.2611598 -0.2269138
# sample estimates:
#  mean of x  mean of y
# 0.02230769 0.26634450

## Yes, difference in means is statistically significant.
```