BDP2-389 Hiller: Power and sample size simulation

Benjamin Chan (chanb@ohsu.edu)

2018-11-12 08:06:50

source("simulatePower.R")

# Summary

The outcome is the post-intervention PSS score, assumed to be normally distributed. The model is a random intercept model with subjects nested within wave. The model adjusts for baseline PSS score.

where is the index for wave. The random intercept accounts for the potential for correlation between subjects within wave. For example, if the overall economy is good or bad at a particular calendar time, then all subjects recruited before that wave might have correlated stress levels. The amount of correlation between subjects within wave is the intraclass correlation set by iccWithinWave, 5%.

Correlation is also assumed between a subject’s baseline and post-intervention stress level. For example, a person who doesn’t get stressed is likely to not be under stress at both time points. The amount of correlation between time points within subject is the intraclass correlation set by iccWithinSubject, 40%.

Power simulations assume a study design with 6 waves and 4 subjects in randomized to each intervention arm in each wave. Simulations results are based on 5000 simulated data sets.

power %>% kable(digits = c(0, rep(3, 6), rep(0, 3), 3))

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| term | nominalEffSize | sigma | meanBeta | sdBeta | iccWithinSubject | iccWithinWave | groupSize | nReject | nSim | power |
| MBSR | 0.78 | 1 | 0.783 | 0.288 | 0.4 | 0.05 | 24 | 4044 | 5000 | 0.809 |
| yBaseline | NA | NA | 0.398 | 0.140 | 0.4 | 0.05 | 24 | 4080 | 5000 | 0.816 |

With a total of 24 subjects in each intervention arm (48 subjects in total), and a within subject correlation of 40% and a within wave correlation of 5%, there is

* 80.9% power to detect an effect size of 0.78

# Details

V1 is the covariance matrix for within subject covariance (dimension 2, baseline to post-intervention). V2 is the covariance matrix for within wave covariance (dimension 4, number of subjects in each wave in each intervention arm).

V1

## [,1] [,2]  
## [1,] 1.0 0.4  
## [2,] 0.4 1.0

V2

## [,1] [,2] [,3] [,4]  
## [1,] 1.00 0.05 0.05 0.05  
## [2,] 0.05 1.00 0.05 0.05  
## [3,] 0.05 0.05 1.00 0.05  
## [4,] 0.05 0.05 0.05 1.00

V

## [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]  
## [1,] 1.00 0.05 0.05 0.05 0.40 0.02 0.02 0.02  
## [2,] 0.05 1.00 0.05 0.05 0.02 0.40 0.02 0.02  
## [3,] 0.05 0.05 1.00 0.05 0.02 0.02 0.40 0.02  
## [4,] 0.05 0.05 0.05 1.00 0.02 0.02 0.02 0.40  
## [5,] 0.40 0.02 0.02 0.02 1.00 0.05 0.05 0.05  
## [6,] 0.02 0.40 0.02 0.02 0.05 1.00 0.05 0.05  
## [7,] 0.02 0.02 0.40 0.02 0.05 0.05 1.00 0.05  
## [8,] 0.02 0.02 0.02 0.40 0.05 0.05 0.05 1.00

Check simulation data. Make sure it meets expectation.

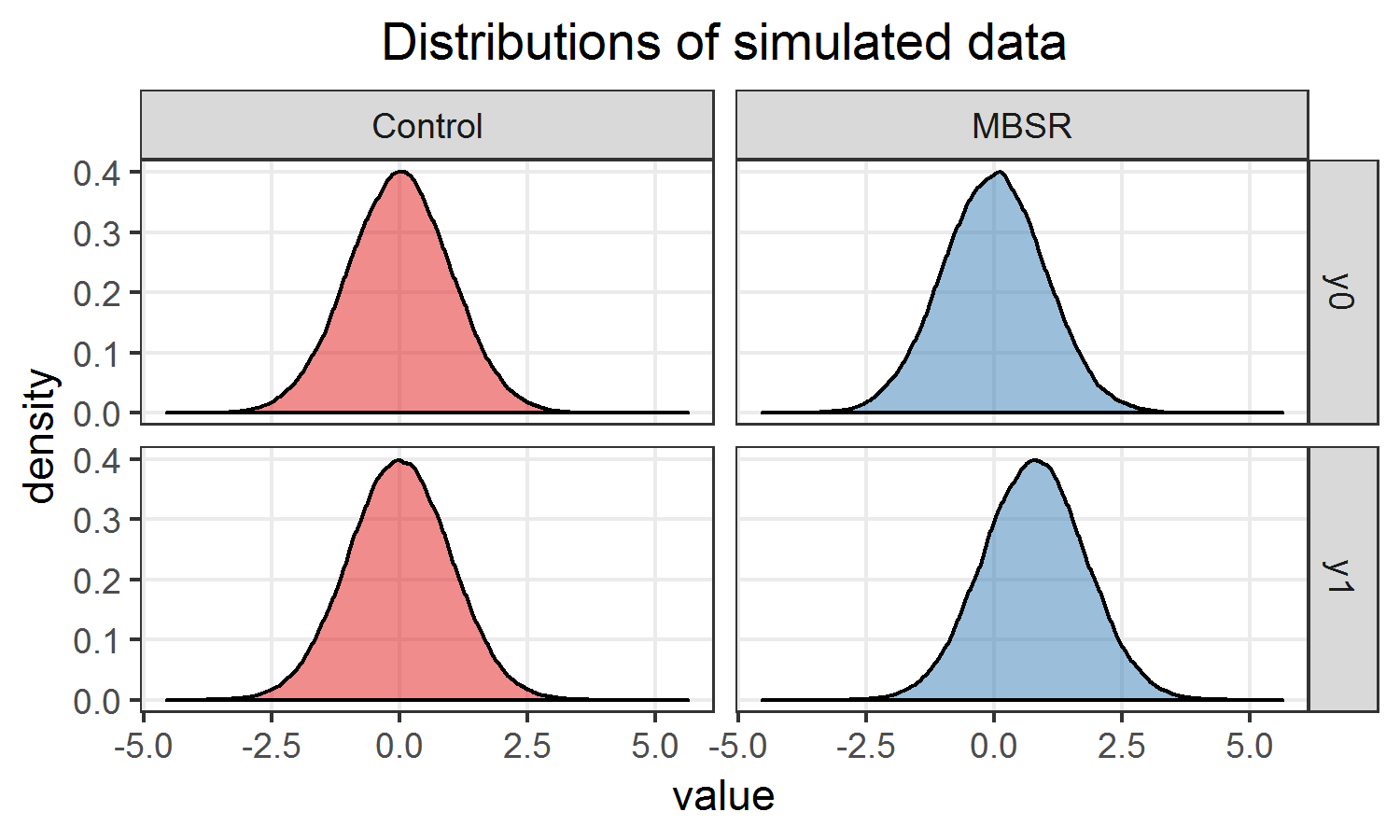
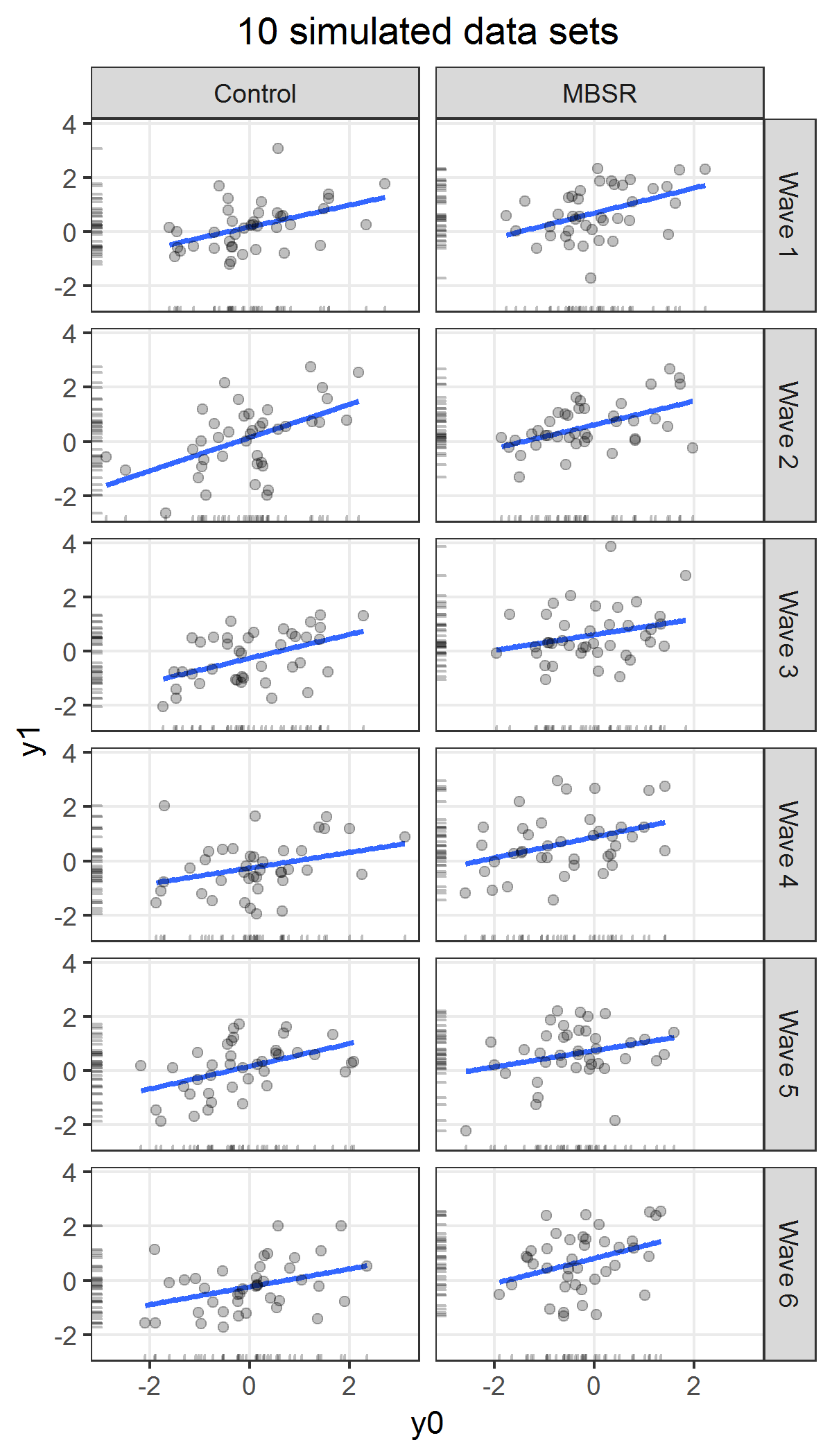
df %>%  
 group\_by(intervention) %>%  
 summarize(meanBaseline = mean(y0),  
 sdBaseline = sd(y0),  
 medianBaseline = median(y0),  
 minBaseline = min(y0),  
 maxBaseline = max(y0)) %>%  
 kable(digits = 2)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| intervention | meanBaseline | sdBaseline | medianBaseline | minBaseline | maxBaseline |
| Control | 0.00 | 1 | 0 | -4.55 | 4.43 |
| MBSR | -0.01 | 1 | 0 | -4.29 | 4.42 |

cor(df %>% select(y0, y1)) %>% kable(digits = 4)

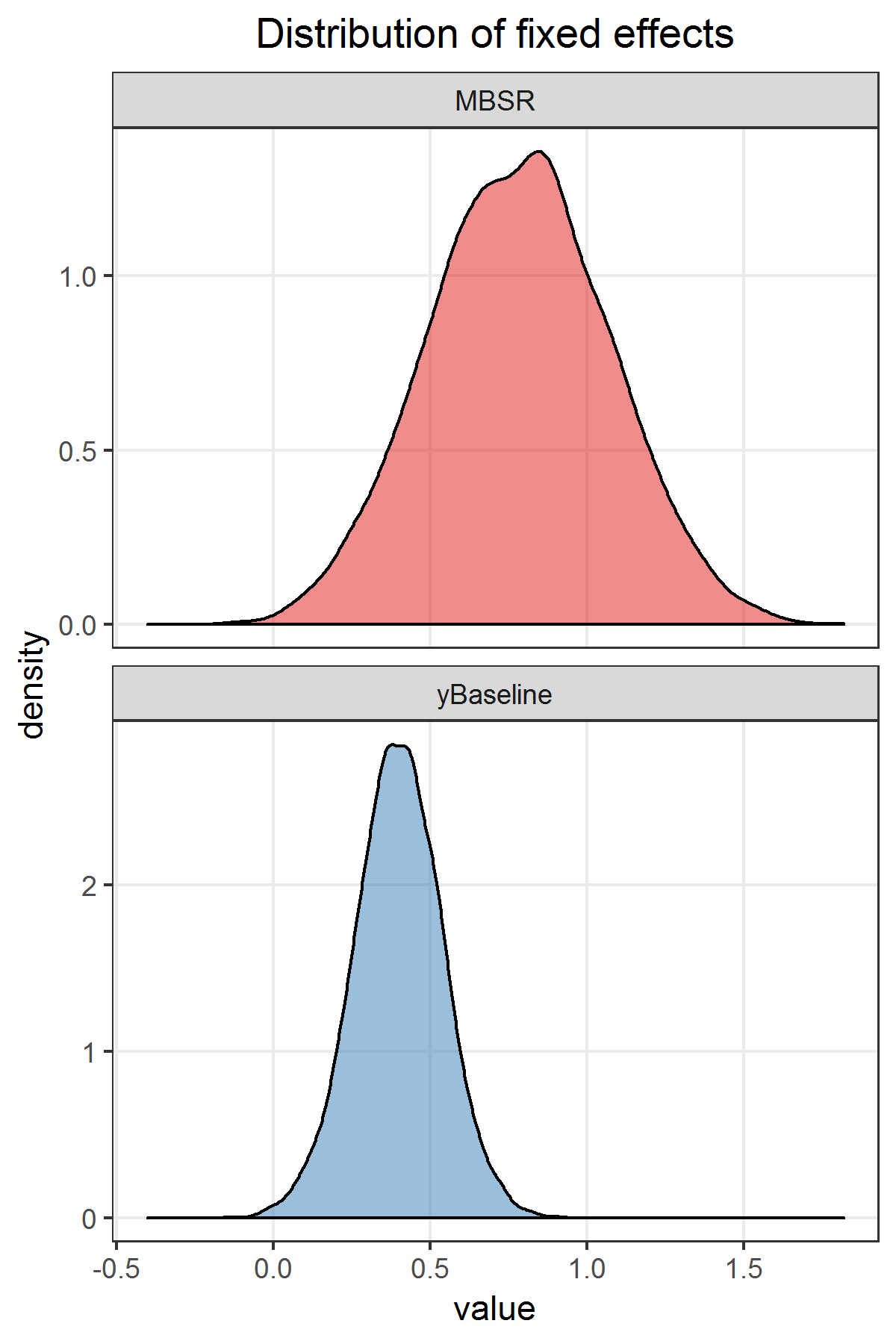
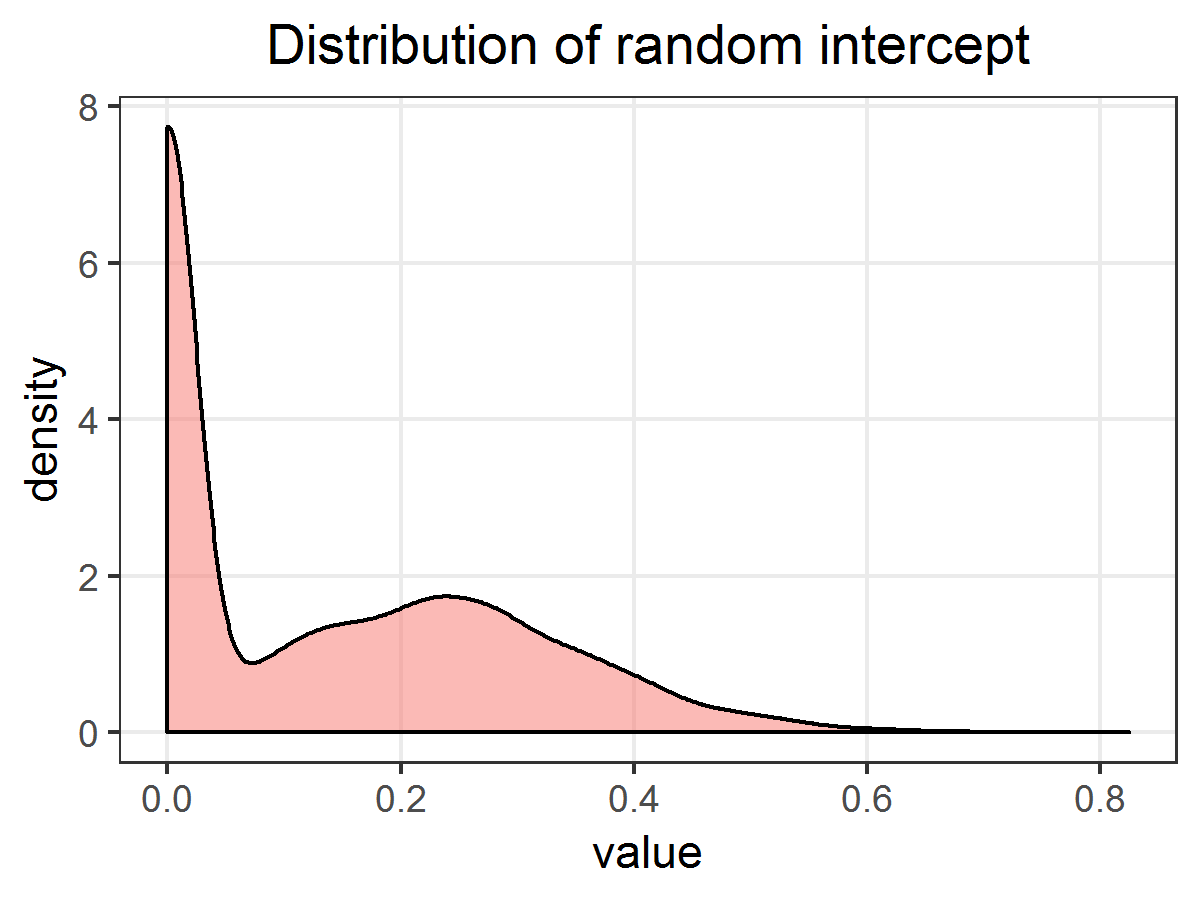
|  |  |  |
| --- | --- | --- |
|  | y0 | y1 |
| y0 | 1.0000 | 0.3715 |
| y1 | 0.3715 | 1.0000 |

G <-   
 df %>%  
 gather(rep, value, y0, y1) %>%  
 ggplot(aes(x = value)) +  
 ggtitle("Distributions of simulated data") +  
 geom\_density(aes(fill = intervention), alpha = 1/2) +  
 scale\_fill\_brewer("Intervention group", palette = "Set1") +  
 facet\_grid(rep ~ intervention) +  
 theme\_bw() +  
 theme(legend.position = "none",  
 panel.grid.minor = element\_blank(),  
 plot.title = element\_text(hjust = 0.5))  
ggsave("../figures/simDataDensities.png", width = 5, height = 3)  
G <-  
 df %>%  
 filter(sim <= 10) %>%  
 mutate(wave = sprintf("Wave %d", wave)) %>%  
 ggplot(aes(x = y0, y = y1)) +  
 ggtitle("10 simulated data sets") +  
 geom\_smooth(method = "lm", se = FALSE) +  
 geom\_point(alpha = 1/4) +  
 geom\_rug(alpha = 1/4) +  
 facet\_grid(wave ~ intervention) +  
 theme\_bw() +  
 theme(panel.grid.minor = element\_blank(),  
 plot.title = element\_text(hjust = 0.5))  
ggsave("../figures/simCorrData.png", width = 4)

Check model simulation output.

G <-   
 simResults %>%  
 filter(group == "fixed") %>%  
 gather(variable, value, estimate) %>%  
 ggplot(aes(x = value)) +  
 ggtitle("Distribution of fixed effects") +  
 geom\_density(aes(fill = term), alpha = 1/2) +  
 facet\_wrap(~ term, scale = "free\_y", ncol = 1) +  
 scale\_fill\_brewer("Model term", palette = "Set1") +  
 theme\_bw() +  
 theme(legend.position = "none",  
 panel.grid.minor = element\_blank(),  
 plot.title = element\_text(hjust = 0.5))  
ggsave("../figures/simModelEstimateDensity.png", width = 4, height = 6)  
G <-   
 simResults %>%  
 filter(group == "wave") %>%  
 gather(variable, value, estimate) %>%  
 ggplot(aes(x = value)) +  
 ggtitle("Distribution of random intercept") +  
 geom\_density(aes(fill = term), alpha = 1/2) +  
 theme\_bw() +  
 theme(legend.position = "none",  
 panel.grid.minor = element\_blank(),  
 plot.title = element\_text(hjust = 0.5))  
ggsave("../figures/simRandomEffectsDensity.png", width = 4, height = 3)

R session information.

sessionInfo()

## R version 3.5.1 (2018-07-02)  
## Platform: x86\_64-w64-mingw32/x64 (64-bit)  
## Running under: Windows 7 x64 (build 7601) Service Pack 1  
##   
## Matrix products: default  
##   
## locale:  
## [1] LC\_COLLATE=English\_United States.1252   
## [2] LC\_CTYPE=English\_United States.1252   
## [3] LC\_MONETARY=English\_United States.1252  
## [4] LC\_NUMERIC=C   
## [5] LC\_TIME=English\_United States.1252   
##   
## attached base packages:  
## [1] parallel stats graphics grDevices utils datasets methods   
## [8] base   
##   
## other attached packages:  
## [1] bindrcpp\_0.2.2 doParallel\_1.0.14 iterators\_1.0.10   
## [4] foreach\_1.4.4 ggplot2\_3.0.0 broom\_0.5.0   
## [7] lme4\_1.1-18-1 Matrix\_1.2-14 simstudy\_0.1.10   
## [10] data.table\_1.11.8 dplyr\_0.7.6 tidyr\_0.8.1   
## [13] magrittr\_1.5 knitr\_1.20 checkpoint\_0.4.3   
##   
## loaded via a namespace (and not attached):  
## [1] Rcpp\_0.12.18 RColorBrewer\_1.1-2 highr\_0.7   
## [4] plyr\_1.8.4 pillar\_1.3.0 compiler\_3.5.1   
## [7] nloptr\_1.2.1 bindr\_0.1.1 tools\_3.5.1   
## [10] mvnfast\_0.2.5 evaluate\_0.11 tibble\_1.4.2   
## [13] nlme\_3.1-137 gtable\_0.2.0 lattice\_0.20-35   
## [16] pkgconfig\_2.0.2 rlang\_0.2.2 withr\_2.1.2   
## [19] stringr\_1.3.1 grid\_3.5.1 tidyselect\_0.2.4   
## [22] glue\_1.3.0 R6\_2.2.2 minqa\_1.2.4   
## [25] reshape2\_1.4.3 purrr\_0.2.5 codetools\_0.2-15   
## [28] backports\_1.1.2 scales\_1.0.0 splines\_3.5.1   
## [31] MASS\_7.3-50 assertthat\_0.2.0 colorspace\_1.3-2   
## [34] labeling\_0.3 stringi\_1.2.4 lazyeval\_0.2.1   
## [37] munsell\_0.5.0 crayon\_1.3.4