Bayesian Comparison of Conditions

Nele Albers

04 January, 2023

Contents

Introduction

This file is to reproduce our Bayesian comparison of the two conditions with regards to participant characteristics, reported in the "Participants"-section of the paper. The participant characteristics we consider are the age, sex, Godin leisure time activity based on (Godin (2011)), the pre-measurement of running or walking self-efficacy based on an adaptation of the exercise self-efficacy scale by (McAuley (1993)), number of hours spent sitting on weekend days (Craig et al. (2003)), smoking frequency, Transtheoretical Model (TTM)-stage for becoming physically active based on adapting the question by (Norman et al. (1998)) to physical activity, and the weekly exercise amount.

Authored by Nele Albers, Beyza Hizli, Bouke L. Scheltinga, Eline Meijer, and Willem-Paul Brinkman.

Required Files

The following files are required: - Data/pre_C.csv

Setup

First, we load the packages that we need.

```
library(formatR)  # To wrap lines
library(rethinking)  # For Bayesian models
```

Also, we set the number of iterations and chains used for fitting the models.

```
NUM_ITERATIONS = 3000  # our value: 3000
NUM_CHAINS = 4  # our value: 4
```

Data Files

```
df = read.csv(file = 'Data/pre_C.csv')

# Turn smoking frequency into binary variable based on whether or not people smoke at least one per day
df$smoking_freq = as.integer(df$smoking_freq >= 3)

# Create separate variables for male and female
df$Sex_Male = as.integer(df$sex == 2)
df$Sex_Female = as.integer(df$sex == 1)

# Make condition numeric
df$group = as.integer(df$group == "personalized") + 1

dat_list <- list(
  age = df$age,
  s_f = df$Sex_Female,
  s_m = df$Sex_Male,
  godin = df$godin_activity,
  se = df$pre,
  sit = df$sitting_weekend,
  sq = df$smoking_freq,
  ttm = df$ttm_pa,
  weekex = df$weekly_exercise,
  cond = df$group
)
```

Analysis

We will now analyze the impact of the condition on the different participant characteristics. First, we fit a model with the respective participant characteristic as dependent variable and only an intercept as predictor. Then, we fit a model that additionally contains the conditions as predictors. We compare these two models based on the WAIC. A lower WAIC suggest a better approximate out-of-sample deviance and thus a better model. Based on the WAIC result table, we conclude that the condition is not an important predictor of the dependent variable if the WAIC is lower for the baseline model.

Age

Let's first get the lowest and highest observed age.

```
age_min = min(df$age)
age_max = max(df$age)

paste("Minimum observed age:", age_min)
```

```
## [1] "Minimum observed age: 19"
```

```r
paste("Maximum observed age:", age_max)
```

```
## [1] "Maximum observed age: 72"
```

Let's first create a model with only the intercept. The age ranges from 19 to 72, so the midpoint is 45.5. So we set the mean of the normal prior to 45.5.

```r
set.seed(18)  # For reproducibility
ml.age.basis <- ulam(alist(age ~ dnorm(mu, sigma), mu <- a_bar, a_bar ~
    dnorm(45.5, 10), sigma ~ dexp(1)), data = dat_list, chains = NUM_CHAINS,
    log_lik = TRUE, cores = NUM_CHAINS, iter = NUM_ITERATIONS)
```

```r
# Print estimators
precis(ml.age.basis, prob = 0.95)
```

```
##           mean       sd     2.5%    97.5%    n_eff    Rhat4
## a_bar 41.77439 2.296730 37.15299 46.21445 3563.149 1.000037
## sigma 14.27096 1.364781 11.80412 17.24416 3621.218 1.000248
```

Next, let's add the conditions as predictors.

```r
set.seed(18)  # For reproducibility
ml.age.cond <- ulam(alist(age ~ dnorm(mu, sigma), mu <- a_bar + b * cond,
    a_bar ~ dnorm(45.5, 10), b ~ dnorm(0, 10), sigma ~ dexp(1)), data = dat_list,
    chains = NUM_CHAINS, log_lik = TRUE, cores = NUM_CHAINS, iter = NUM_ITERATIONS)
```

```r
# Print estimators
precis(ml.age.cond, prob = 0.95, depth = 2)
```

```
##            mean       sd      2.5%     97.5%    n_eff    Rhat4
## a_bar 47.705438 5.592923  36.73875 58.457855 1651.035 1.001474
## b     -4.186833 3.636305 -11.29857  2.998101 1680.554 1.001600
## sigma 14.122158 1.354318  11.73823 17.118185 2358.381 1.000422
```

And now let's compare the two models based on WAIC. A lower WAIC suggest a better approximate out-of-sample deviance and thus a better model. Based on the WAIC result table, we conclude that the condition is not an important predictor of the dependent variable if the WAIC is lower for the baseline model.

```r
compare(ml.age.cond, ml.age.basis)
```

```
##                  WAIC       SE     dWAIC      dSE    pWAIC    weight
## ml.age.basis 332.5308 6.911691 0.0000000       NA 1.752005 0.5355017
## ml.age.cond  332.8153 7.311640 0.2844925 2.290321 2.563093 0.4644983
```

Sex: Female

Let's first create a model with only the intercept. The dependent variable now is a binary variable.

```r
set.seed(18)  # For reproducibility
ml.s_f.basis <- ulam(alist(s_f ~ dbinom(1, p), logit(p) <- a_bar, a_bar ~
    dnorm(0, 1.5)), data = dat_list, chains = NUM_CHAINS, log_lik = TRUE,
    cores = NUM_CHAINS, iter = NUM_ITERATIONS)
```

```r
# Print estimators
precis(ml.s_f.basis, prob = 0.95)
```

```
##             mean        sd       2.5%     97.5%    n_eff    Rhat4
## a_bar 0.04296208 0.3247741 -0.6168248 0.6491108 1797.241 1.001754
```

Next, let's add the condition as predictor.

```
set.seed(18)  # For reproducibility
ml.s_f.cond <- ulam(alist(s_f ~ dbinom(1, p), logit(p) <- a_bar + b * cond,
    a_bar ~ dnorm(0, 1.5), b ~ dnorm(0, 1.5)), data = dat_list, chains = NUM_CHAINS,
    log_lik = TRUE, cores = NUM_CHAINS, iter = NUM_ITERATIONS)
```

```
# Print estimators
precis(ml.s_f.cond, prob = 0.95, depth = 2)
```

```
##              mean        sd       2.5%    97.5%    n_eff    Rhat4
## a_bar -0.05804037 0.8051700 -1.6888830 1.461491 1398.371 1.002892
## b      0.07684279 0.5209729 -0.9000322 1.119207 1392.659 1.003108
```

And now let's compare the two models based on the WAIC. A lower WAIC suggest a better approximate out-of-sample deviance and thus a better model. Based on the WAIC result table, we conclude that the condition is not an important predictor of the dependent variable if the WAIC is lower for the baseline model.

```
compare(ml.s_f.cond, ml.s_f.basis)
```

```
##                  WAIC        SE    dWAIC       dSE    pWAIC    weight
## ml.s_f.basis 56.12323 0.2725332 0.000000        NA 1.041283 0.6584222
## ml.s_f.cond  57.43577 0.4522511 1.312542 0.2835141 1.710183 0.3415778
```

Sex: Male

Let's first create a model with only the intercept. The dependent variable now is a binary variable.

```
set.seed(18)  # For reproducibility
ml.s_m.basis <- ulam(alist(s_m ~ dbinom(1, p), logit(p) <- a_bar, a_bar ~
    dnorm(0, 1.5)), data = dat_list, chains = NUM_CHAINS, log_lik = TRUE,
    cores = NUM_CHAINS, iter = NUM_ITERATIONS)
```

```
# Print estimators
precis(ml.s_m.basis, prob = 0.95)
```

```
##              mean        sd       2.5%     97.5%    n_eff   Rhat4
## a_bar -0.06779758 0.3206477 -0.7126552 0.5681148 2011.881 1.00116
```

Next, let's add the condition as predictor.

```
set.seed(18)  # For reproducibility
ml.s_m.cond <- ulam(alist(s_m ~ dbinom(1, p), logit(p) <- a_bar + b * cond,
    a_bar ~ dnorm(0, 1.5), b ~ dnorm(0, 1.5)), data = dat_list, chains = NUM_CHAINS,
    log_lik = TRUE, cores = NUM_CHAINS, iter = NUM_ITERATIONS)
```

```
# Print estimators
precis(ml.s_m.cond, prob = 0.95, depth = 2)
```

```
##              mean        sd      2.5%     97.5%    n_eff    Rhat4
## a_bar  0.03642257 0.8163832 -1.599480 1.6222812 1314.139 1.000123
## b     -0.06664703 0.5260895 -1.112771 0.9885673 1226.269 1.000221
```

And now let's compare the two models based on the WAIC. A lower WAIC suggest a better approximate out-of-sample deviance and thus a better model. Based on the WAIC result table, we conclude that the condition is not an important predictor of the dependent variable if the WAIC is lower for the baseline model.

```
compare(ml.s_m.cond, ml.s_m.basis)
```

```
##                  WAIC        SE    dWAIC       dSE    pWAIC    weight
```

```
## ml.s_m.basis 56.07229 0.4346687 0.000000        NA 1.015160 0.6678365
## ml.s_m.cond  57.46912 0.4707884 1.396832 0.2356739 1.725622 0.3321635
```

Godin Leisure Time Activity

Let's get the largest observed value.

```
max_val_godin = max(df$godin_activity)
max_val_godin
```

```
## [1] 101
```

Let's first create a model with only the intercept. The largest observed value is 101, so we set the normal prior to a mean of 50.5.

```
set.seed(18)  # For reproducibility
ml.godin.basis <- ulam(alist(godin ~ dnorm(mu, sigma), mu <- a_bar, a_bar ~
    dnorm(50.5, 10), sigma ~ dexp(1)), data = dat_list, chains = NUM_CHAINS,
    log_lik = TRUE, cores = NUM_CHAINS, iter = NUM_ITERATIONS)
```

```
# Print estimators
precis(ml.godin.basis, prob = 0.95)
```

```
##           mean       sd     2.5%    97.5%    n_eff     Rhat4
## a_bar 35.65016 3.385557 29.24673 42.54760 3928.072 0.9999495
## sigma 22.14998 1.897541 18.76414 26.27783 3659.514 1.0007594
```

Next, let's add the conditions as predictors.

```
set.seed(18)  # For reproducibility
ml.godin.cond <- ulam(alist(godin ~ dnorm(mu, sigma), mu <- a_bar + b *
    cond, a_bar ~ dnorm(50.5, 10), b ~ dnorm(0, 10), sigma ~ dexp(1)),
    data = dat_list, chains = NUM_CHAINS, log_lik = TRUE, cores = NUM_CHAINS,
    iter = NUM_ITERATIONS)
```

```
# Print estimators
precis(ml.godin.cond, prob = 0.95, depth = 2)
```

```
##            mean       sd      2.5%     97.5%    n_eff     Rhat4
## a_bar 44.481061 6.676608  31.63953 57.444597 2104.877 1.0004500
## b     -6.706183 4.437163 -15.55804  2.111151 2130.970 1.0004718
## sigma 22.060633 1.911105  18.59269 26.092808 3131.602 0.9996664
```

And now let's compare the two models based on WAIC. A lower WAIC suggest a better approximate out-of-sample deviance and thus a better model. Based on the WAIC result table, we conclude that the condition is not an important predictor of the dependent variable if the WAIC is lower for the baseline model.

```
compare(ml.godin.cond, ml.godin.basis)
```

```
##                    WAIC       SE     dWAIC      dSE    pWAIC    weight
## ml.godin.basis 376.3470 14.49204 0.0000000       NA 2.882197 0.5341732
## ml.godin.cond  376.6208 14.09030 0.2738122 2.367791 3.274098 0.4658268
```

Running/Walking Self-Efficacy

Let's first create a model with only the intercept. The self-efficacy scale ranges from 0 to 100, so we set the normal prior to a mean of 50.

```
set.seed(18)  # For reproducibility
ml.se.basis <- ulam(alist(se ~ dnorm(mu, sigma), mu <- a_bar, a_bar ~ dnorm(50,
```

```
    10), sigma ~ dexp(1)), data = dat_list, chains = NUM_CHAINS, log_lik = TRUE,
    cores = NUM_CHAINS, iter = NUM_ITERATIONS)
```

```
# Print estimators
precis(ml.se.basis, prob = 0.95)
```

```
##           mean       sd     2.5%    97.5%    n_eff     Rhat4
## a_bar 67.60091 3.396502 60.76580 74.27294 3903.489 0.9998533
## sigma 22.32651 1.918748 18.95292 26.52592 3696.083 0.9999780
```

Next, let's add the conditions as predictors.

```
set.seed(18)  # For reproducibility
ml.se.cond <- ulam(alist(se ~ dnorm(mu, sigma), mu <- a_bar + b * cond,
    a_bar ~ dnorm(50, 10), b ~ dnorm(0, 10), sigma ~ dexp(1)), data = dat_list,
    chains = NUM_CHAINS, log_lik = TRUE, cores = NUM_CHAINS, iter = NUM_ITERATIONS)
```

```
# Print estimators
precis(ml.se.cond, prob = 0.95, depth = 2)
```

```
##            mean       sd      2.5%    97.5%    n_eff    Rhat4
## a_bar 65.169867 6.800005 51.027610 77.82143 2282.762 1.000259
## b      1.830408 4.488141 -6.624642 10.82482 2289.148 1.000736
## sigma 22.471781 1.925695 19.056663 26.66356 3036.404 1.000278
```

And now let's compare the two models based on WAIC. A lower WAIC suggest a better approximate out-of-sample deviance and thus a better model. Based on the WAIC result table, we conclude that the condition is not an important predictor of the dependent variable if the WAIC is lower for the baseline model.

```
compare(ml.se.cond, ml.se.basis)
```

```
##                 WAIC       SE   dWAIC       dSE    pWAIC   weight
## ml.se.basis 376.4154 10.89365 0.00000        NA 2.335232 0.657024
## ml.se.cond  377.7155 10.75443 1.30012 0.6602526 2.831735 0.342976
```

Sitting Hours Weekend Day

Let's first create a model with only the intercept. The possible values range from 0 to 24, so we set the mean of the normal prior to 12.

```
set.seed(18)  # For reproducibility
ml.sit.basis <- ulam(alist(sit ~ dnorm(mu, sigma), mu <- a_bar, a_bar ~
    dnorm(12, 10), sigma ~ dexp(1)), data = dat_list, chains = NUM_CHAINS,
    log_lik = TRUE, cores = NUM_CHAINS, iter = NUM_ITERATIONS)
```

```
# Print estimators
precis(ml.sit.basis, prob = 0.95)
```

```
##           mean        sd     2.5%    97.5%    n_eff    Rhat4
## a_bar 7.791622 0.6441331 6.515785 9.069427 3440.423 1.000174
## sigma 3.985021 0.4454787 3.244343 5.000734 3595.515 1.000423
```

Next, let's add the conditions as predictors.

```
set.seed(18)  # For reproducibility
ml.sit.cond <- ulam(alist(sit ~ dnorm(mu, sigma), mu <- a_bar + b * cond,
    a_bar ~ dnorm(12, 10), b ~ dnorm(0, 10), sigma ~ dexp(1)), data = dat_list,
    chains = NUM_CHAINS, log_lik = TRUE, cores = NUM_CHAINS, iter = NUM_ITERATIONS)
```

```r
# Print estimators
precis(ml.sit.cond, prob = 0.95, depth = 2)
```

```
##             mean        sd      2.5%      97.5%    n_eff    Rhat4
## a_bar 10.490774 1.9442845  6.688613 14.3682925 1996.110 1.000542
## b     -1.826769 1.2407776 -4.267188  0.6569373 1997.310 1.000287
## sigma  3.944749 0.4312982  3.228057  4.9209282 2554.179 1.000286
```

And now let's compare the two models based on WAIC. A lower WAIC suggest a better approximate out-of-sample deviance and thus a better model. Based on the WAIC result table, we conclude that the condition is not an important predictor of the dependent variable if the WAIC is lower for the baseline model.

```r
compare(ml.sit.cond, ml.sit.basis)
```

```
##                   WAIC       SE    dWAIC      dSE    pWAIC    weight
## ml.sit.basis 223.4329 11.39858 0.000000       NA 2.280206 0.5367041
## ml.sit.cond  223.7271 13.68060 0.294162 3.765262 3.563686 0.4632959
```

Smoking Frequency

Let's first create a model with only the intercept. Note that the dependent variable now is a binary variable.

```r
set.seed(18)  # For reproducibility
ml.sq.basis <- ulam(alist(sq ~ dbinom(1, p), logit(p) <- a_bar, a_bar ~
    dnorm(0, 1.5)), data = dat_list, chains = NUM_CHAINS, log_lik = TRUE,
    cores = NUM_CHAINS, iter = NUM_ITERATIONS)
```

```r
# Print estimators
precis(ml.sq.basis, prob = 0.95)
```

```
##              mean        sd      2.5%     97.5%    n_eff     Rhat4
## a_bar -0.4628361 0.3197273 -1.104532 0.1539674 2057.866 0.9998417
```

Next, let's add the condition as predictor.

```r
set.seed(18)  # For reproducibility
ml.sq.cond <- ulam(alist(sq ~ dbinom(1, p), logit(p) <- a_bar + b * cond,
    a_bar ~ dnorm(0, 1.5), b ~ dnorm(0, 1.5)), data = dat_list, chains = NUM_CHAINS,
    log_lik = TRUE, cores = NUM_CHAINS, iter = NUM_ITERATIONS)
```

```r
# Print estimators
precis(ml.sq.cond, prob = 0.95, depth = 2)
```

```
##              mean        sd      2.5%     97.5%    n_eff    Rhat4
## a_bar  0.1777112 0.8084039 -1.422562 1.7611620 1251.645 1.000571
## b     -0.4628809 0.5357714 -1.577205 0.5648121 1225.069 1.001128
```

And now let's compare the two models based on the WAIC. A lower WAIC suggest a better approximate out-of-sample deviance and thus a better model. Based on the WAIC result table, we conclude that the condition is not an important predictor of the dependent variable if the WAIC is lower for the baseline model.

```r
compare(ml.sq.cond, ml.sq.basis)
```

```
##                 WAIC       SE    dWAIC      dSE     pWAIC   weight
## ml.sq.basis 53.88136 2.885696 0.000000       NA 0.9542468 0.596108
## ml.sq.cond  54.65991 3.479820 0.778549 1.526095 1.6961311 0.403892
```

TTM-Stage for Becoming Physically Active

Let's first create a model with only the intercept. Note that the dependent variable now is an ordered categorical variable.

```
set.seed(18)  # For reproducibility
ml.ttm.basis <- ulam(alist(ttm ~ dordlogit(0, cutpoints), cutpoints ~ dnorm(0,
    1.5)), data = dat_list, chains = NUM_CHAINS, log_lik = TRUE, cores = NUM_CHAINS,
    iter = NUM_ITERATIONS)
```

```
# Print estimators
precis(ml.ttm.basis, prob = 0.95, depth = 2)
```

```
##                    mean        sd       2.5%      97.5%    n_eff     Rhat4
## cutpoints[1] -1.0232948 0.3464294 -1.7210970 -0.3636606 3641.015 1.0007806
## cutpoints[2] -0.3941068 0.3150827 -1.0125987  0.2187715 5823.544 1.0004492
## cutpoints[3]  0.5545425 0.3162734 -0.0446307  1.1870455 7352.900 1.0002631
## cutpoints[4]  2.3450459 0.5232169  1.3948668  3.4530812 6342.075 0.9997613
```

Next, let's add the conditions as predictors.

```
set.seed(18)  # For reproducibility
ml.ttm.cond <- ulam(alist(ttm ~ dordlogit(phi, cutpoints), phi <- b[cond],
    cutpoints ~ dnorm(0, 1.5), b[cond] ~ dnorm(0, 0.5)), data = dat_list,
    chains = NUM_CHAINS, log_lik = TRUE, cores = NUM_CHAINS, iter = NUM_ITERATIONS)
```

```
# Print estimators
precis(ml.ttm.cond, prob = 0.95, depth = 2)
```

```
##                     mean        sd       2.5%      97.5%    n_eff     Rhat4
## cutpoints[1] -1.09210770 0.4421055 -1.9573467 -0.2413296 3853.895 1.0005032
## cutpoints[2] -0.46542858 0.4200466 -1.2901507  0.3728089 4157.259 1.0001940
## cutpoints[3]  0.49009995 0.4227904 -0.3480943  1.3239255 4661.050 1.0002166
## cutpoints[4]  2.28089271 0.5720354  1.2017720  3.4297275 6102.029 0.9998908
## b[1]         -0.09079321 0.3983316 -0.8727943  0.6962745 4551.822 1.0006316
## b[2]         -0.06187009 0.3951267 -0.8400281  0.7104832 4296.401 0.9997545
```

And now let's compare the two models based on the WAIC. A lower WAIC suggest a better approximate out-of-sample deviance and thus a better model. Based on the WAIC result table, we conclude that the condition is not an important predictor of the dependent variable if the WAIC is lower for the baseline model.

```
compare(ml.ttm.cond, ml.ttm.basis)
```

```
##                  WAIC       SE    dWAIC      dSE    pWAIC    weight
## ml.ttm.basis 125.4406 4.699210 0.000000       NA 3.605273 0.6435535
## ml.ttm.cond  126.6223 4.783278 1.181642 0.161478 4.205498 0.3564465
```

Weekly Exercise Amount

Let's first create a model with only the intercept. The possible values range from 1 to 3, so we set the mean of the normal prior to 2.

```
set.seed(18)  # For reproducibility
ml.weekex.basis <- ulam(alist(weekex ~ dnorm(mu, sigma), mu <- a_bar, a_bar ~
    dnorm(2, 10), sigma ~ dexp(1)), data = dat_list, chains = NUM_CHAINS,
    log_lik = TRUE, cores = NUM_CHAINS, iter = NUM_ITERATIONS)
```

```
# Print estimators
precis(ml.weekex.basis, prob = 0.95)
```

```
##            mean        sd      2.5%    97.5%    n_eff    Rhat4
## a_bar 1.9749417 0.1373302 1.7047577 2.248409 3686.747 1.000479
## sigma 0.8599879 0.1025061 0.6870954 1.094513 3918.271 1.000746
```

Next, let's add the conditions as predictors.

```
set.seed(18)  # For reproducibility
ml.weekex.cond <- ulam(alist(weekex ~ dnorm(mu, sigma), mu <- a_bar + b *
    cond, a_bar ~ dnorm(2, 10), b ~ dnorm(0, 10), sigma ~ dexp(1)), data = dat_list,
    chains = NUM_CHAINS, log_lik = TRUE, cores = NUM_CHAINS, iter = NUM_ITERATIONS)
```

```
# Print estimators
precis(ml.weekex.cond, prob = 0.95, depth = 2)
```

```
##           mean        sd       2.5%     97.5%    n_eff    Rhat4
## a_bar 1.739579 0.4268862  0.9034902 2.5759123 1821.092 1.002598
## b     0.158164 0.2731078 -0.3918901 0.6961071 1811.625 1.003074
## sigma 0.868688 0.1024159  0.6970190 1.0997330 2103.556 1.000441
```

And now let's compare the two models based on WAIC. A lower WAIC suggest a better approximate out-of-sample deviance and thus a better model. Based on the WAIC result table, we conclude that the condition is not an important predictor of the dependent variable if the WAIC is lower for the baseline model.

```
compare(ml.weekex.cond, ml.weekex.basis)
```

```
##                      WAIC       SE    dWAIC      dSE    pWAIC   weight
## ml.weekex.basis  99.57285 4.141785 0.000000       NA 1.196200 0.689453
## ml.weekex.cond  101.16798 4.256858 1.595126 1.199328 2.087316 0.310547
```

References

Craig, Cora L, Alison L Marshall, Michael Sjöström, Adrian E Bauman, Michael L Booth, Barbara E Ainsworth, Michael Pratt, et al. 2003. "International Physical Activity Questionnaire: 12-Country Reliability and Validity." *Medicine & Science in Sports & Exercise* 35 (8): 1381–95.

Godin, Gaston. 2011. "The Godin-Shephard Leisure-Time Physical Activity Questionnaire." *The Health & Fitness Journal of Canada* 4 (1): 18–22.

McAuley, Edward. 1993. "Self-Efficacy and the Maintenance of Exercise Participation in Older Adults." *Journal of Behavioral Medicine* 16 (1): 103–13.

Norman, GJ, SV Benisovich, CR Nigg, and JS Rossi. 1998. "Examining Three Exercise Staging Algorithms in Two Samples." In *19th Annual Meeting of the Society of Behavioral Medicine*.