Ovsanna

Hsueh-Pin Liu, Nuo Chen, Tao Guo, Huifei Xu

2023-05-04

Previous steps

Before making the EDAs, I first changed the variable's names of the .xlsx file to make it easier to import, so the new .xlsx file is different from the original one.

EDAs

According to the data, there are 2 "two-week" patients, 4"three-week" patients and 2 "four-week" patients. I make plots for each patient's each question and the black line seperates the pre-treatment and treatment time.

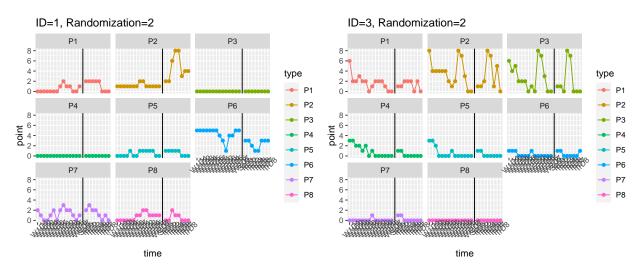


Figure 1: 2-week pre-treatment patients

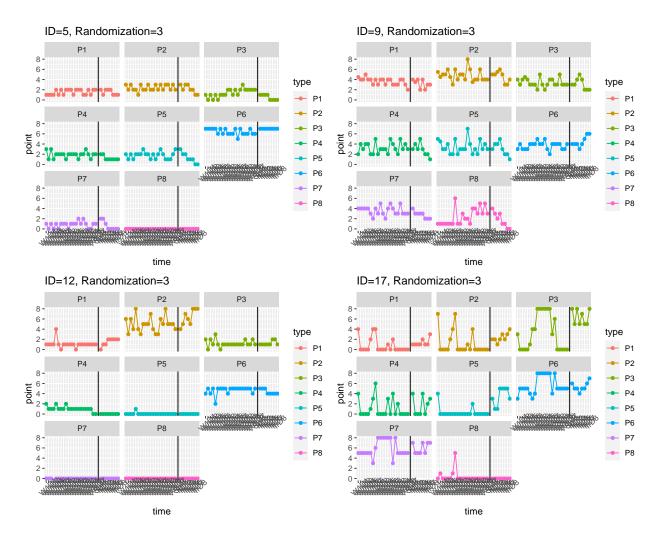


Figure 2: 3-week pre-treatment patients

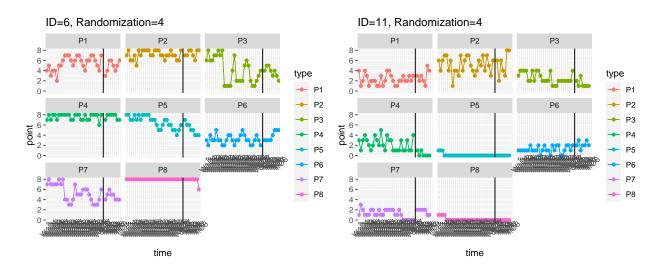


Figure 3: 4-week pre-treatment patients

Comment on these EDAs

After analyzing the plots, several noteworthy observations can be made:

- 1. The score of Q6 differs from the other questions. While the scores of other questions increase, the score of Q6 decreases.
- 2. The treatment appears to be ineffective for many questions. For instance, Q8 shows no change in trend for six patients. Moreover, the treatment may even worsen the condition, as seen in the first patient's Q2 score, which increased after treatment.
- 3. Due to the limited dataset consisting of only eight patients with a scoring scale ranging from 0-8, it is challenging to create effective models and it may not be possible to develop highly accurate models. Nevertheless, efforts can be made to create models despite these limitations.

Methods

By the previous EDA part, we can see the data has some traits.

First, as different patients may have different feeling about the anxiety level for pre-treatment and posttreatment. To analyze the fixed effect and random effect, the multilevel regression model is the best way to solve it.

Second, the data is actually collected over time. So for such a time series data, we can check the acf plot to see if there is a trend or seasonality.

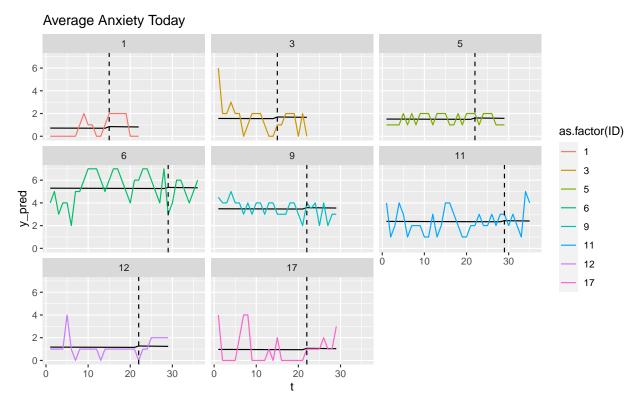
Third, both the question score y and the time t are ordinal, and the indicator variable Treatment T is binary. So we also need to check the data by fitting a ordinal regression model and seeing the performance. However the ordinal regression model has the limitation that it is hard for the model to fit the data between different groups. And the model fit is bad indeed. So this model will not be shown in next part.

We expect that, after fitting these models, there are patterns of increasing scores in Question "Max Anxiety Today", "Confident in My Ability to Handle Situations", "Push Away Thoughts and Feelings I Do Not Like", and decreasing scores in Question "Average Anxiety Today", "Scares Me when I am Nervous", "Jump to Negative Conclusions", "Avg Avoidance of Thoughts, Situations, Sensations". For Question "Physical Feelings in My Body Scare Me", the daily scores are all stable among all patients so we consider not to analyze this question.

Result

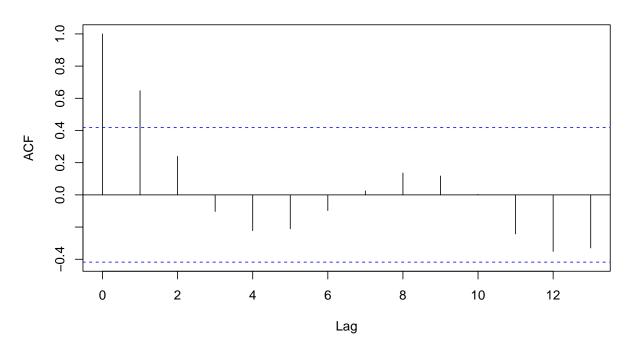
We fit separated model for each problem because different problem may have different effect. Although the data is actually ordinal but 9 points is enough to do linear model, and it is more easier to diagnose linear model. We The result of most model is same with what we found in EDA, the score of patient is same after making treatment. For instance, in most problem, only the intercept is significant.

	Estimate	Std. Error	df	t value	$\Pr(> t)$
(Intercept)	2.136820	0.575868	7.9303	-28.2	0.006 **
${f T}$	0.197230	0.747232	222.20	0.264	0.79206
t	-0.001098	0.012335	221.08	-0.089	0.92916
T:t	-0.003483	0.029450	221.41	-0.118	0.90596



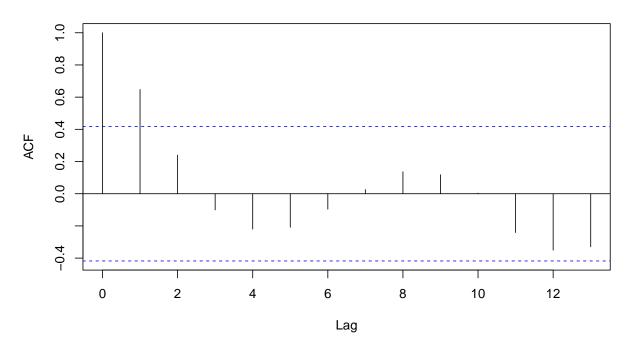
After fitting model, we do the auto-correlation analysis. From acf plot, there could be temporal effect. Therefore we try to use time series model (AR model). After fitting with AR model, the acf plot almost did not change for each problem. temporal effect may be too small.

Series m[m\$ID == 1,]\$resid



After fitting with AR model, the acf plot almost did not change for each problem. temporal effect may be too small.

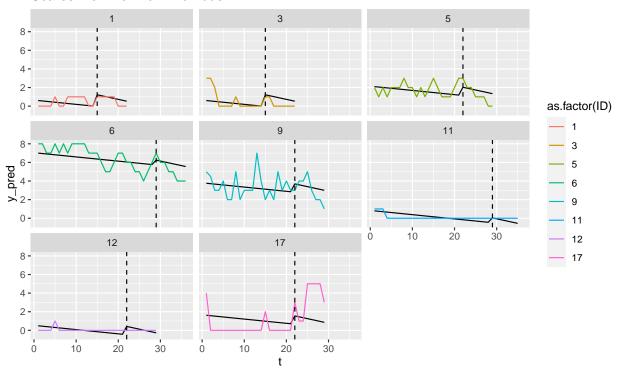
Series m[m\$ID == 1,]\$resid



There are only two question the treatment seem to effects to patients, Scares Me when I am Nervous and Confident in My Ability to Handle Situations. However, for Scares Me when I am Nervous and Confident, after using AR models, this effect disappeared. From plot of this question, we could not find significant decreasing of score after taking treatment, which are correspond the result of AR model

	Estimate	Std. Error	df	t value	$\Pr(> t)$
(Intercept)	2.16590	0.81736	7.42557	2.650	0.031243 *
${ m T}$	2.04479	0.72973	221.06774	2.802	0.005527**
t	-0.04548	0.01204	220.51531	-3.778	0.000203***
T:t	-0.05264	0.02874	220.67366	-1.831	0.068379

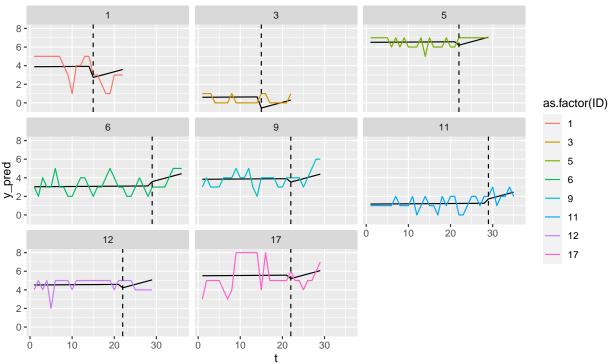
Scares Me when I am Nervous



The result of Confident in My Ability to Handle Situations is much better than result of Scares Me when I am Nervous. The plot also correspond to result of model. Both T and T:t are significant. Form plot and fitted line, we could find some patient increase after treatment. Finally, we also check with mixed ordinal model because data is ordinal, but compare with mixed ordinal model is hard to visualize. We put the all result in appendix.

	Estimate	Std. Error	df	t value	$\Pr(> t)$
(Intercept)	3.631392	0.728410	7.451728	4.985	0.00132**
${ m T}$	-2.978075	0.669260	221.129847	-4.450	1.36e-05***
\mathbf{t}	0.003036	0.011040	220.545842	0.275	0.78354
T:t	0.118754	0.026362	220.713363	4.505	1.08e-05***

Confident in My Ability to Handle Situations



#Discussion

The results of the ordinal regression model developed to assess the effectiveness of the treatment in improving anxiety among adolescents based on the survey questions scored on an ordinal scale from 0 to 8 should be interpreted with caution due to several limitations.

First and foremost, the sample size of the dataset is extremely small, consisting of only 8 people's data. This small sample size may limit the statistical power of the model and affect the accuracy of its predictions. Additionally, the small sample size may not be representative of the entire population of adolescents experiencing anxiety issues, which may limit the generalizability of the model's findings.

Second, the assumption of equal intervals between levels of the ordinal scale may not hold true in reality. The difference between each level of the ordinal scale may not be equal, which could affect the accuracy of the model's predictions.

Third, the model assumes a linear relationship between the independent and dependent variables. While ordinal regression models can capture nonlinear relationships between variables, the linear model may not accurately represent the relationship between the survey questions and anxiety symptoms. Other models, such as polynomial regression or spline regression, may be more appropriate for capturing nonlinear relationships.

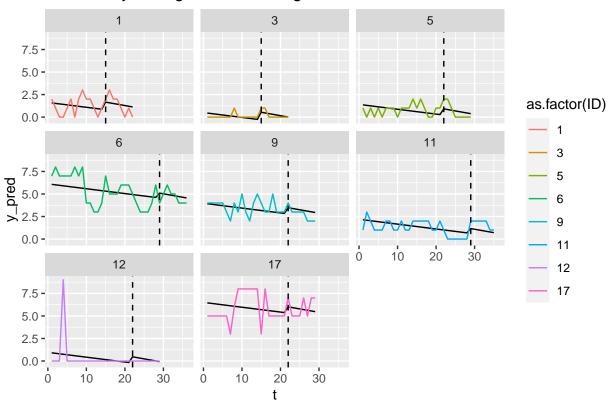
Additionally, the survey questions are self-reported and may not accurately reflect the severity or frequency of anxiety symptoms. Furthermore, there may be other factors that contribute to anxiety symptoms that are not captured by the survey questions, such as genetic or environmental factors.

Overall, while the ordinal regression model developed in this project provides some initial insights into the effectiveness of the treatment in improving anxiety among adolescents based on the survey questions scored on an ordinal scale from 0 to 8, the above limitations should be taken into consideration when interpreting the results. Future research with a larger sample size and more comprehensive measures of anxiety symptoms may be necessary to provide more reliable and generalizable findings.

Appendix

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: y ~ T + t * T + (1 | as.factor(ID))
     Data: m6
## REML criterion at convergence: 760.2
##
## Scaled residuals:
    Min 1Q Median
##
                            3Q
                                  Max
## -2.7610 -0.5608 -0.1045 0.4846 7.2786
## Random effects:
## Groups
               Name
                        Variance Std.Dev.
## as.factor(ID) (Intercept) 5.532
                                  2.352
## Residual
                          1.285
                                  1.134
## Number of obs: 231, groups: as.factor(ID), 8
##
## Fixed effects:
             Estimate Std. Error
                                      df t value Pr(>|t|)
## (Intercept) 2.91240
                        0.84920 7.47426 3.430 0.00994 **
## T
              1.17029 0.80633 221.19792 1.451 0.14809
## t
              -0.02202 0.03176 220.75374 -0.693 0.48878
## T:t
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
## Correlation of Fixed Effects:
##
      (Intr) T
                  t
## T -0.027
## t -0.173 0.021
## T:t 0.056 -0.950 -0.258
```

Push Away Thoughts and Feelings I Do Not Like

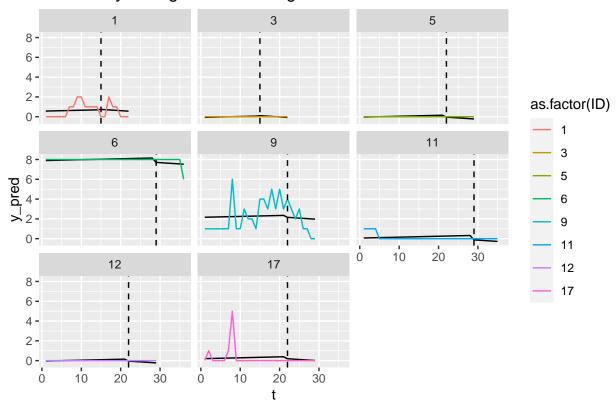


```
## Linear mixed-effects model fit by REML
     Data: m6
##
##
          AIC
                   BIC
                          logLik
##
     750.9272 774.9019 -368.4636
##
## Random effects:
    Formula: ~1 | as.factor(ID)
##
           (Intercept) Residual
## StdDev:
              2.351128 1.155872
##
## Correlation Structure: AR(1)
    Formula: ~1 | as.factor(ID)
##
    Parameter estimate(s):
##
         Phi
## 0.3271351
## Fixed effects: y \sim T + t * T
##
                    Value Std.Error DF
                                          t-value p-value
## (Intercept) 2.8771326 0.8633003 220 3.332713 0.0010
## T
                1.3259463 1.0232815 220
                                         1.295779
## t
               -0.0516173 0.0175745 220 -2.937062
                                                   0.0037
               -0.0288226 0.0410666 220 -0.701851 0.4835
   Correlation:
##
##
       (Intr) T
       -0.044
## T
       -0.228 0.083
## T:t 0.079 -0.951 -0.308
```

```
##
## Standardized Within-Group Residuals:
                    Q1
                             Med
## -2.6735718 -0.5284025 -0.1054118 0.4539406 7.1520062
## Number of Observations: 231
## Number of Groups: 8
## Cumulative Link Mixed Model fitted with the Laplace approximation
## formula: y \sim T + t * T + (1 \mid as.factor(ID))
## data:
          m6
##
## link threshold nobs logLik AIC
                                               max.grad cond.H
                                    niter
## logit flexible 231 -289.61 605.22 1085(7898) 1.29e-03 2.9e+05
##
## Random effects:
## Groups
                Name
                            Variance Std.Dev.
## as.factor(ID) (Intercept) 21.71
                                    4.66
## Number of groups: as.factor(ID) 8
##
## Coefficients:
      Estimate Std. Error z value Pr(>|z|)
       ## t -0.08166 0.02327 -3.509 0.00045 ***
## T:t -0.06905 0.05422 -1.273 0.20285
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
##
## Threshold coefficients:
##
      Estimate Std. Error z value
## 0|1 -3.6579
                 1.6718 -2.188
## 1|2 -1.8403
                  1.6657 -1.105
## 213
       0.7256
                  1.6768 0.433
## 3|4
       2.8651
                 1.6913
                          1.694
## 4|5
       4.1530
                 1.6984 2.445
## 5|6
       5.7648
                           3.371
                  1.7102
## 6|7
       6.1799
                  1.7132
                           3.607
## 718
       7.1735
                  1.7265 4.155
## 819
       9.7671
                  1.9727 4.951
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: y ~ T + t * T + (1 | as.factor(ID))
##
     Data: m7
##
## REML criterion at convergence: 556.6
##
## Scaled residuals:
##
      Min
             1Q Median
                              3Q
                                    Max
## -2.8165 -0.2910 -0.0668 0.1268 6.6611
##
## Random effects:
## Groups
                Name
                           Variance Std.Dev.
```

```
## as.factor(ID) (Intercept) 7.5792
   Residual
                              0.5043
                                      0.7102
## Number of obs: 231, groups: as.factor(ID), 8
##
## Fixed effects:
##
                Estimate Std. Error
                                            df t value Pr(>|t|)
## (Intercept)
                 1.336496
                           0.979308
                                      7.140944
                                                  1.365
                                                          0.2138
                           0.505673 220.362513
## T
                                                          0.2683
                 0.561208
                                                  1.110
## t
                 0.009391
                            0.008337 220.174137
                                                  1.126
                                                          0.2613
## T:t
                -0.034975
                           0.019912 220.227741 -1.757
                                                          0.0804 .
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
## Correlation of Fixed Effects:
##
       (Intr) T
## T
       -0.015
## t
      -0.094 0.020
## T:t 0.030 -0.950 -0.257
```

Push Away Thoughts and Feelings I Do Not Like

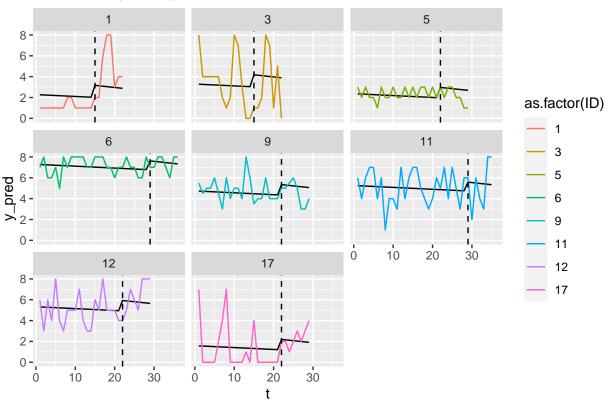


```
## Linear mixed-effects model fit by REML
## Data: m7
## AIC BIC logLik
## 543.8212 567.7959 -264.9106
##
## Random effects:
## Formula: ~1 | as.factor(ID)
```

```
(Intercept) Residual
## StdDev:
             2.745514 0.7291276
##
## Correlation Structure: AR(1)
## Formula: ~1 | as.factor(ID)
## Parameter estimate(s):
        Phi
## 0.3603219
## Fixed effects: y \sim T + t * T
                  Value Std.Error DF t-value p-value
## (Intercept) 1.3543270 0.9824060 220 1.378582 0.1694
              0.6377633 0.6572827 220 0.970303 0.3330
## T
              0.0068061 0.0113633 220 0.598956 0.5498
## t
## T:t
              -0.0364790 0.0264524 220 -1.379041 0.1693
## Correlation:
##
      (Intr) T
## T
     -0.026
## t -0.130 0.092
## T:t 0.045 -0.951 -0.313
## Standardized Within-Group Residuals:
                      Q1
        {	t Min}
                                 Med
## -2.64643410 -0.30198290 -0.05895287 0.11544185 6.47916091
## Number of Observations: 231
## Number of Groups: 8
## Cumulative Link Mixed Model fitted with the Laplace approximation
## formula: y \sim T + t * T + (1 \mid as.factor(ID))
## data:
         m7
##
## link threshold nobs logLik AIC niter max.grad cond.H
## logit flexible 231 -117.22 256.43 1128(12026) 1.12e+02 1.2e+07
##
## Random effects:
## Groups
                            Variance Std.Dev.
                 Name
## as.factor(ID) (Intercept) 65.07
## Number of groups: as.factor(ID) 8
## Coefficients:
       Estimate Std. Error z value Pr(>|z|)
       5.3052169  0.0006685  7935.49  <2e-16 ***
       0.0080511 0.0006694 12.03 <2e-16 ***
## T:t -0.2772134  0.0006686 -414.61  <2e-16 ***
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
## Threshold coefficients:
       Estimate Std. Error z value
## 0|1 2.863e+00 6.686e-04 4282.21
## 1|2 5.151e+00 4.280e-01
                           12.04
## 2|3 5.811e+00 4.794e-01
                           12.12
## 3|4 6.704e+00 5.624e-01
                           11.92
```

```
## 4|5 7.363e+00 6.512e-01 11.31
## 5|6 8.637e+00 9.682e-01 8.92
## 6|8 1.113e+01 2.213e+00
                         5.03
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: y ~ T + t * T + (1 | as.factor(ID))
     Data: m1
##
## REML criterion at convergence: 925.7
## Scaled residuals:
                        3Q
##
      Min 1Q Median
## -2.4789 -0.6876 -0.0800 0.5115 3.3440
##
## Random effects:
## Groups
                          Variance Std.Dev.
               Name
## as.factor(ID) (Intercept) 3.912
                                  1.978
                          2.756
## Residual
                                  1.660
## Number of obs: 231, groups: as.factor(ID), 8
##
## Fixed effects:
             Estimate Std. Error df t value Pr(>|t|)
## (Intercept) 4.02130 0.74339 8.44541 5.409 0.000532 ***
              1.48380 1.17753 223.22366 1.260 0.208951
## T
## t
             -0.01821 0.01945 221.62260 -0.936 0.350308
## T:t
             ## ---
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
## Correlation of Fixed Effects:
##
     (Intr) T
## T -0.046
## t -0.289 0.025
## T:t 0.095 -0.950 -0.262
```

Max Anxiety Today

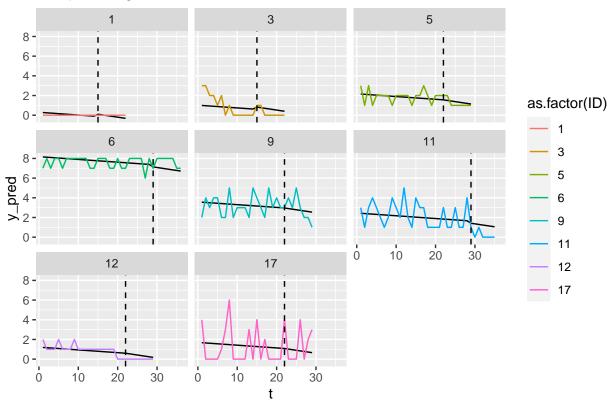


```
## Linear mixed-effects model fit by REML
     Data: m1
##
##
         AIC
                  BIC
                        logLik
##
     920.3059 944.2806 -453.153
##
## Random effects:
    Formula: ~1 | as.factor(ID)
##
           (Intercept) Residual
## StdDev:
               1.9288 1.705292
## Correlation Structure: AR(1)
  Formula: ~1 | as.factor(ID)
   Parameter estimate(s):
##
        Phi
## 0.3228426
## Fixed effects: y \sim T + t * T
                   Value Std.Error DF
                                       t-value p-value
## (Intercept) 4.096115 0.7630646 220 5.367980 0.0000
## T
               0.931428 1.4995962 220 0.621119 0.5352
## t
              -0.020236 0.0257955 220 -0.784460 0.4336
               -0.001454 0.0602599 220 -0.024126 0.9808
   Correlation:
##
       (Intr) T
      -0.075
## T
## t -0.379 0.088
## T:t 0.133 -0.951 -0.313
```

```
##
## Standardized Within-Group Residuals:
                       Q1
## -2.42294175 -0.69885558 -0.09197737 0.49122749 3.13779198
## Number of Observations: 231
## Number of Groups: 8
## Warning in update.uC(rho): Non finite negative log-likelihood
    at iteration 350
## Warning in update.uC(rho): Non finite negative log-likelihood
    at iteration 618
## Warning in update.uC(rho): Non finite negative log-likelihood
    at iteration 758
## Warning in update.uC(rho): Non finite negative log-likelihood
    at iteration 1471
## Cumulative Link Mixed Model fitted with the Laplace approximation
## formula: ordered(y) \sim T + t * T + (1 | as.factor(ID))
## data:
           m1
##
## link threshold nobs logLik AIC
                                       niter
## logit flexible 231 -439.51 909.03 1588(9505) 1.79e-04 9.1e+04
##
## Random effects:
## Groups
                 Name
                             Variance Std.Dev.
## as.factor(ID) (Intercept) 4.534
## Number of groups: as.factor(ID) 8
##
## Coefficients:
##
       Estimate Std. Error z value Pr(>|z|)
       1.230838 1.457687
                            0.844
                                     0.398
      -0.017650
                 0.020088 -0.879
                                      0.380
## T:t -0.004956
                 0.057092 -0.087
                                      0.931
##
## Threshold coefficients:
##
        Estimate Std. Error z value
## 0|1
         -3.6952
                    0.8575 -4.309
## 1|2
         -2.3857
                     0.8302 -2.874
## 213
         -1.2853
                     0.8216 -1.564
## 3|3.5 -0.2761
                     0.8175 -0.338
## 3.5|4 -0.2409
                     0.8174 -0.295
## 4|4.5 0.8387
                     0.8180
                             1.025
## 4.5|5 0.9052
                     0.8182 1.106
## 5|5.5 1.5909
                    0.8213 1.937
## 5.5|6 1.6572
                     0.8216 2.017
## 6|7
          2.4975
                     0.8293 3.011
## 7|8
          3.6631
                    0.8485
                             4.317
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: y \sim T + t * T + (1 \mid as.factor(ID))
##
     Data: m3
## REML criterion at convergence: 709.9
## Scaled residuals:
                            3Q
      Min 1Q Median
                                     Max
## -1.6271 -0.6634 -0.0498 0.3780 4.4834
## Random effects:
## Groups
                 Name
                            Variance Std.Dev.
## as.factor(ID) (Intercept) 6.153
                                     2.481
## Residual
                            1.019
                                     1.009
## Number of obs: 231, groups: as.factor(ID), 8
##
## Fixed effects:
              Estimate Std. Error
                                        df t value Pr(>|t|)
                          0.89034 7.33575 2.897 0.0219 *
## (Intercept) 2.57940
## T
               0.70662
                          0.71833 220.86647 0.984
                                                    0.3263
## t
               -0.02803
                          0.01185 220.41221 -2.366
## T:t
              -0.03221
                          0.02829 220.54215 -1.139 0.2561
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
## Correlation of Fixed Effects:
      (Intr) T
## T -0.023
## t -0.147 0.020
## T:t 0.048 -0.950 -0.258
```

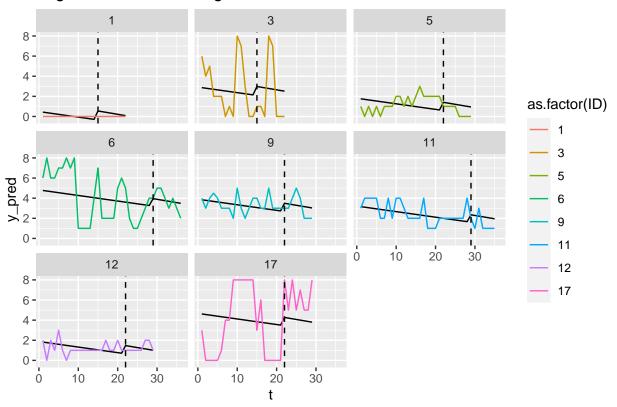
Jump to Negative Conclusions



```
Value Std.Error DF
                                             t-value
                                                         p-value
## (Intercept) 2.58344986 0.89225260 220
                                           2.8954243 0.004168142
## T
               0.76273737 0.77339309 220
                                          0.9862221 0.325107520
               -0.02825279 0.01284489 220 -2.1995349 0.028881183
## t
## T:t
              -0.03430151 0.03055475 220 -1.1226242 0.262820803
## Cumulative Link Mixed Model fitted with the Laplace approximation
##
## formula: ordered(y) \sim T + t * T + (1 | as.factor(ID))
## data:
           mЗ
##
   link threshold nobs logLik AIC
                                                   max.grad cond.H
                                       niter
   logit flexible 231 -288.62 601.23 1089(11331) 1.06e-04 3.0e+05
##
##
## Random effects:
   Groups
                              Variance Std.Dev.
                 Name
   as.factor(ID) (Intercept) 25.07
## Number of groups: as.factor(ID) 8
##
  Coefficients:
      Estimate Std. Error z value Pr(>|z|)
##
## T
       0.98291
                  1.63590 0.601
                                    0.5479
                  0.02322 -2.357
                                     0.0184 *
      -0.05472
## T:t -0.05465
                  0.06253 -0.874
                                     0.3821
## ---
```

```
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
##
## Threshold coefficients:
     Estimate Std. Error z value
## 0|1 -2.7378 1.8524 -1.478
## 1|2 -1.0018 1.8465 -0.543
## 2|3 0.3107 1.8461 0.168
      1.6044
## 3|4
              1.8525 0.866
              1.8769 1.609
## 4|5 3.0195
## 5|6 4.5991 1.9919 2.309
## 6|7 6.3342
                 2.3704 2.672
## 7|8 9.1337
                 2.5087 3.641
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: y \sim T + t * T + (1 \mid as.factor(ID))
##
     Data: m2
##
## REML criterion at convergence: 969.5
##
## Scaled residuals:
      Min
##
           1Q
                  Median
                               3Q
## -2.46862 -0.46695 -0.08721 0.50280 3.04935
##
## Random effects:
## Groups
          Name
                         Variance Std.Dev.
## as.factor(ID) (Intercept) 2.403 1.550
## Residual
                          3.412
                                  1.847
## Number of obs: 231, groups: as.factor(ID), 8
##
## Fixed effects:
            Estimate Std. Error df t value Pr(>|t|)
## (Intercept) 2.96285 0.61566 9.86430 4.812 0.000738 ***
## T
             1.05237 1.30502 225.35824 0.806 0.420863
## t
             ## T:t
             -0.01090 0.05154 223.69761 -0.211 0.832764
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
## Correlation of Fixed Effects:
##
     (Intr) T
## T -0.063
## t -0.389 0.030
## T:t 0.128 -0.950 -0.267
```

Avg Avoidance of Thoughts, Situatoins, Sensations



```
## Warning in update.uC(rho): Non finite negative log-likelihood
    at iteration 822
## Cumulative Link Mixed Model fitted with the Laplace approximation
## formula: ordered(y) \sim T + t * T + (1 | as.factor(ID))
## data:
##
  link threshold nobs logLik AIC
                                       niter
                                                  max.grad cond.H
  logit flexible 231 -404.26 834.53 1189(5966) 7.16e-05 7.4e+04
##
## Random effects:
  Groups
                 Name
                             Variance Std.Dev.
   as.factor(ID) (Intercept) 5.559
## Number of groups: as.factor(ID) 8
##
## Coefficients:
       Estimate Std. Error z value Pr(>|z|)
##
## T
                 1.488199
                             0.393
                                     0.6940
       0.585447
      -0.043191
                  0.020933 -2.063
                                     0.0391 *
## T:t -0.007855
                  0.055678 -0.141
                                     0.8878
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
## Threshold coefficients:
```

##		${\tt Estimate}$	Std. Error	z value
##	0 1	-2.1710	0.7514	-2.889
##	1 2	-0.5730	0.7539	-0.760
##	2 3	0.6361	0.7566	0.841
##	3 4	1.3162	0.7586	1.735
##	4 4.5	2.0933	0.7650	2.736
##	4.5 5	2.1310	0.7655	2.784
##	5 6	2.6462	0.7735	3.421
##	6 7	2.9689	0.7804	3.804
##	7 8	3.4512	0.7962	4.334