Write-up for the Beat the Blues Data

Eric Reed, Sara Nunez, Yiding Zhang, Kostis Gourgoulias
University of Massachusetts, Amherst

February 4, 2014

1 Background

This is a dataset containing information from a clinical trial with "Beat the Blues", an interactive multimedia program of cognitive-behavioural techniques. Since there are way more patients suffering from anxiety and depression than therapists, the study of the dataset makes the case that this program can indeed help with the treatment. One of the metrics used to gauge depression levels was BDI or Beck Depression Inventory.

The dataset contains one hundred observations of one hundred patients and eight variables. BDI was tracked before treatment, after two months and then after one, three and six month follow-ups. Two groups are studied, one that uses the BtB program and one that has the usual anti-depression treatment. Those can be further splitted to sub-groups of patients that were taking anti-depressant drugs.

```
library(ggplot2)
library(MASS)
library(HSAUR2)

## Loading required package:  tools

BtheB <- BtheB
attach(BtheB)
summary(BtheB)

##   drug     length   treatment    bdi.pre         bdi.2m          bdi.3m
##  No :56   <6m:49    TAU  :48   Min.   : 2.0   Min.   : 0.0   Min.   : 0.0
##  Yes:44   >6m:51    BtheB:52   1st Qu.:15.0   1st Qu.: 8.0   1st Qu.: 6.0
##                                Median :22.0   Median :15.0   Median :13.0
##                                Mean   :23.3   Mean   :16.9   Mean   :14.8
##                                3rd Qu.:30.2   3rd Qu.:23.0   3rd Qu.:20.0
##                                Max.   :49.0   Max.   :48.0   Max.   :53.0
##                                               NA's   :3      NA's   :27
##      bdi.5m          bdi.8m
##  Min.   : 0.0   Min.   : 0.0
##  1st Qu.: 3.0   1st Qu.: 3.0
##  Median :10.0   Median :10.5
##  Mean   :12.8   Mean   :11.1
##  3rd Qu.:20.0   3rd Qu.:15.2
##  Max.   :47.0   Max.   :40.0
##  NA's   :42     NA's   :48
```

2 Variables and Hypothesis

The variables we chose to work with were bdi.2m and bdi.3m, which were the Beck Depression Inventory II test results of the 100 patients after two months

and after three months, respectively.

The purpose of this analysis was to intperpret the results from a linear regression for the Beck Depression Inventory II test at week 3 as, a function of week 2. The slope of the regression line for this data, represents the proportion between the BDI at week 3 and week 2. We therefore hypothesize, that the slope of the regression line for the subset of the patients who recieved the "BtheB" treatment, will be smaller than that of the patients who recieved the "TAU" treatment.

The following summarizes the two variables of interest for the usual treatment group and the beat the blues group:

```
h1 = qplot(bdi.2m, fill = treatment, main = "BDI 2 month Follow-up Histogram")
h2 = qplot(bdi.3m, fill = treatment, main = "BDI 3 month Follow-up Histogram")
multiplot(h1, h2)

## Loading required package:  grid
## stat_bin:  binwidth defaulted to range/30.  Use 'binwidth = x' to
adjust this.
## stat_bin:  binwidth defaulted to range/30.  Use 'binwidth = x' to
adjust this.
```
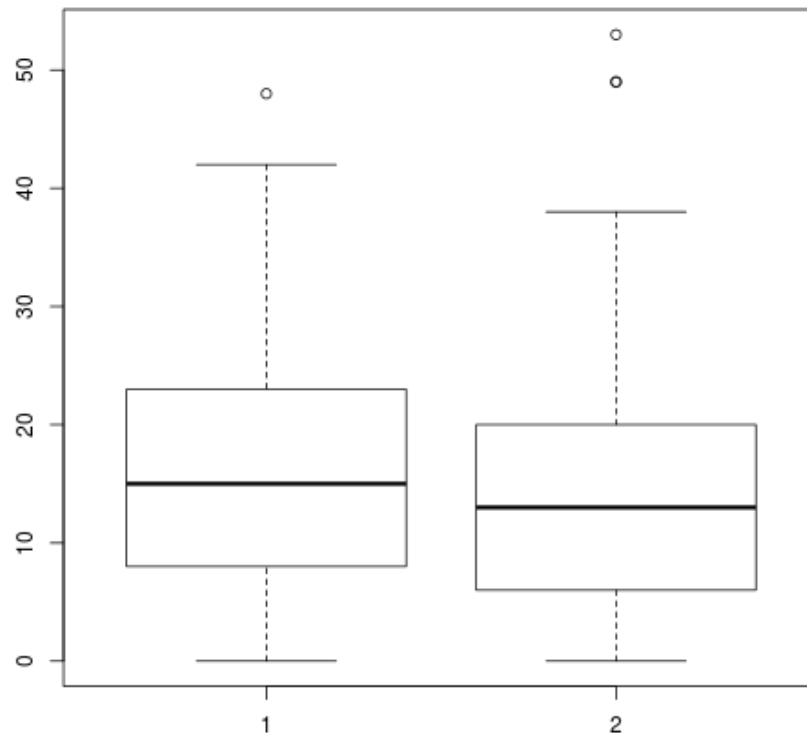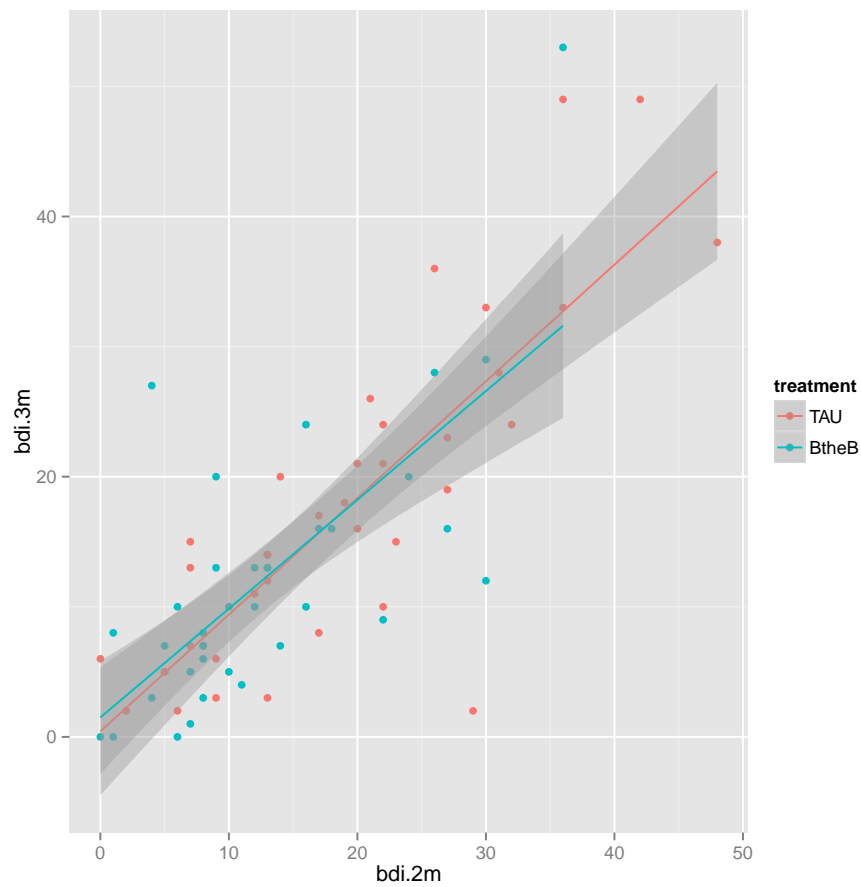
```
qplot(bdi.2m, bdi.3m, data = BtheB, geom = c("point", "smooth"), method = "lm",
    color = treatment)
```

```
## Warning:  Removed 12 rows containing missing values (stat_smooth).
## Warning:  Removed 15 rows containing missing values (stat_smooth).
## Warning:  Removed 27 rows containing missing values (geom_point).
```

```
m2m3m <- lm(bdi.3m ~ bdi.2m, data = BtheB)
summary(m2m3m)

##
## Call:
## lm(formula = bdi.3m ~ bdi.2m, data = BtheB)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -24.261  -4.697  -0.338   2.585  22.534
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.9784     1.5321    0.64     0.53
## bdi.2m        0.8718     0.0802   10.87   <2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.29 on 71 degrees of freedom
##   (27 observations deleted due to missingness)
## Multiple R-squared:  0.625,Adjusted R-squared:  0.619
## F-statistic:  118 on 1 and 71 DF,  p-value: <2e-16

tau <- lm(bdi.3m & treatment == "TAU" ~ bdi.2m & treatment == "TAU", data = BtheB)
btb <- lm(bdi.3m & treatment == "BtheB" ~ bdi.2m & treatment == "BtheB", data = BtheB)
summary(tau)

##
## Call:
## lm(formula = bdi.3m & treatment == "TAU" ~ bdi.2m & treatment ==
##     "TAU", data = BtheB)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.0189 -0.0189 -0.0189  0.0000  0.9811
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     0.0189     0.0147    1.29      0.2
## bdi.2m & treatment == "TAU"TRUE 0.9811     0.0233   42.17   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.107 on 86 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  0.954,Adjusted R-squared:  0.953
## F-statistic: 1.78e+03 on 1 and 86 DF,  p-value: <2e-16

summary(btb)

##
## Call:
## lm(formula = bdi.3m & treatment == "BtheB" ~ bdi.2m & treatment ==
##     "BtheB", data = BtheB)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.9444  0.0000  0.0000  0.0556  0.0556
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   7.71e-16   2.16e-02     0.0        1
```

```
## bdi.2m & treatment == "BtheB"TRUE 9.44e-01  3.31e-02    28.5   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.151 on 83 degrees of freedom
##   (15 observations deleted due to missingness)
## Multiple R-squared:  0.907,Adjusted R-squared:  0.906
## F-statistic:  813 on 1 and 83 DF,  p-value: <2e-16
```

3 Missing Data

```
##        bdi.2m bdi.3m Both
## TAU        3     12    3
## BtheB      0     15    0
## Total      3     27    3
## [1] 73
## [1] 36
## [1] 37
```

There was missing data from 3 of the participants from week two and 27 from week three. All 3 of the participants who had missing data from week two recieved the "TAU" treatment. Of the 27 participants who had missing data from week three, 12 recieved the "TAU" treatment and 15 recieved the "BtheB" treatment. All 3 of the missing participants from week two were also missing from week three, therefore a total of 27 patients are not included in our analysis, 12 from "TAU" and 15 from "BtheB", giving us 36 and 37 respective participants from each group. A sample size of only 73, is certainly very small, and will be detrimental to the confidence we have in the outcome. However, we have considerably more data on studay participants than if we had used data from week 5, or week 8, as each had missing data for 42 and 48 participants respectively.

4 Results and Interpretation

The regression line for the whole dataset had a β value of 0.8718, and a correlational coefficient of 0.625. The regression line for the subset of the data for paricipants in the "TAU" treatment group had a β value of 0.9811, and had a correlational coeficcient of 0.954. The regression line for the subset of the data for paricipants in the "BtheB" treatment group had a β value of 0.944, and had a correlational coeficcient of 0.907. We can interpret these results to say that for the whole datset, given a one unit increase in bdi.2w we can expect a 0.8718 unit increase in bdi.3w. For the subset containing partipants who underwent the "TAU" treatment, given a 1 unit increase in bdi.2w we can expect

a 0.9811 unit increase in bdi.3w. And, For the subset containing partipants who underwent the "BtheB" treatment, given a 1 unit increase in bdi.2w we can expect a 0.944 unit increase in bdi.3w. This supports our hypothesis since the proportion, β value, between our before, "bdi,2w", and after, "bdi.3w", is smaller for the "BtheB" group.

It should also be noted that our regression lines proved to be a much better predictor when we subsetted the data, as the correlational coefficients are much close to 1 for the each subset than the whole dataset.

References

[1] J. Proudfoot, D. Goldberg, A. Mann, B. S. Everitt, I. Marks and J. A. Gray, (2003). Computerized, interactive, multimedia cognitive-behavioural program for anxiety and depression in general practice. Psychological Medicine, 33(2), 217227.*