Does the experimental 'Intervention' have a significant impact on nurse stress (post-test)?

Weronika Cyprian

April 2nd, 2024

Introduction

Multisite trials are pivotal in clinical research, as they are conducted at multiple locations simultaneously. They are essential in assessing the effectiveness of intervention due to multiple reasons. One of the crucial advantages of replicating interventions in different settings is the consideration of diverse population demographics, which enhances the generalizability of the results. Moreover, these trials tend to have bigger sample sizes which also impacts the validity of conclusions and results themselves.

It is also worth noting that multisite trials reduce the impact of bias. By doing so, alongside accounting for confounding variables, these studies improve the accuracy of the results and strengthen their significance. Moreover, multisite trials enable more efficient recruitment of participants by using a larger pool of participants.

Research question Does the experimental intervention have a significant impact on nurse stress levels (post-test)?

Data The dataset considered in this report is a result of a longitudinal multisite trial on the stress of nurses working in Accident and Emergency (A&E) departments in hospitals. For each of the 20 hospitals, nurses working in the A&E departments were randomly assigned to either an experimental or a control condition. In the experimental condition, all nurses receive a training program to cope with job-induced anxiety. After the program is completed, a sample of 10 nurses from each hospital's A&E department are given a test that measures job-related stress, at intervals of 1 month, 2 months and 3 months after the date of the training program.

```r
MST <- read.csv("https://andygolightly.github.io/teaching/MATH43515/summative/weronika.csv",
                header=TRUE)
head(MST)
```

```
##   ID Hospital Responset1 Responset2 Responset3 Trt Experience Gender Size
## 1  1        1         38         39         40   0        9.6      1    1
## 2  2        1         37         37         44   1       11.8      1    1
## 3  3        1         37         38         37   1        8.8      0    1
## 4  4        1         43         38         41   0        5.1      0    1
## 5  5        1         38         39         39   1       10.5      1    1
## 6  6        1         36         38         41   1        6.3      0    1
```

```
dim(MST)
```

```
## [1] 200    9
```

It is therefore visible that this data set includes the records of 200 nurses and there are 9 columns in this dataset. These columns represent the following variables:

- `ID`: anonymized nurse identifier;
- `Hospital`: hospital identifier $\{1, 2, \ldots, 20\}$;
- `Trt`: treatment / experiment indicator (0=control; 1=training program given (treatment));
- `Experience`: nurse experience (continuous scale, units of years);
- `Gender`: gender indicator (0=male; 1=female);
- `Size`: indicator of A&E department size (0=small, 1=large);
- `Responset1`: A post-test stress score for each nurse at time 1;
- `Responset2`: A post-test stress score for each nurse at follow-up time 2;
- `Responset3`: A post-test stress score for each nurse at follow-up time 3;

All test scores were recorded on an integer-valued scale on $[0, 100]$ ranging from "no stress" (0) to "maximum stress" (100).

Descriptive data exploration

Before commencing with the multilevel modelling, it is essential to consider exploratory data analysis, which enables us to understand the structure of the dataset.

```
sum(is.na(MST))
```

```
## [1] 0
```

There are no missing values in this dataset.

```
summary(MST)
```

```
##        ID            Hospital        Responset1      Responset2
##  Min.   :  1.00   Min.   : 1.00   Min.   :30.00   Min.   :33.00
##  1st Qu.: 50.75   1st Qu.: 5.75   1st Qu.:38.00   1st Qu.:38.75
##  Median :100.50   Median :10.50   Median :40.00   Median :40.00
##  Mean   :100.50   Mean   :10.50   Mean   :39.99   Mean   :40.65
##  3rd Qu.:150.25   3rd Qu.:15.25   3rd Qu.:43.00   3rd Qu.:43.00
##  Max.   :200.00   Max.   :20.00   Max.   :48.00   Max.   :50.00
##    Responset3        Trt          Experience        Gender           Size
##  Min.   :34.0   Min.   :0.00   Min.   : 2.100   Min.   :0.000   Min.   :0.0
##  1st Qu.:39.0   1st Qu.:0.00   1st Qu.: 5.100   1st Qu.:0.000   1st Qu.:0.0
##  Median :41.0   Median :0.00   Median : 6.950   Median :1.000   Median :0.5
##  Mean   :41.3   Mean   :0.49   Mean   : 8.056   Mean   :0.615   Mean   :0.5
##  3rd Qu.:44.0   3rd Qu.:1.00   3rd Qu.: 9.825   3rd Qu.:1.000   3rd Qu.:1.0
##  Max.   :49.0   Max.   :1.00   Max.   :23.600   Max.   :1.000   Max.   :1.0
```

The mean value of the post-test stress score at time 1 is 39.99, and it is increasing a little with time, considering that the mean value at follow-up time 3 is 41.3.

```
boxplot(MST$Responset1 ~ MST$Trt,
        main="Response Scores 1 by Treatment group",
        xlab="Treatment Group", ylab="Response 1 Score",
        col=c("purple","pink"))
```

Response Scores 1 by Treatment group
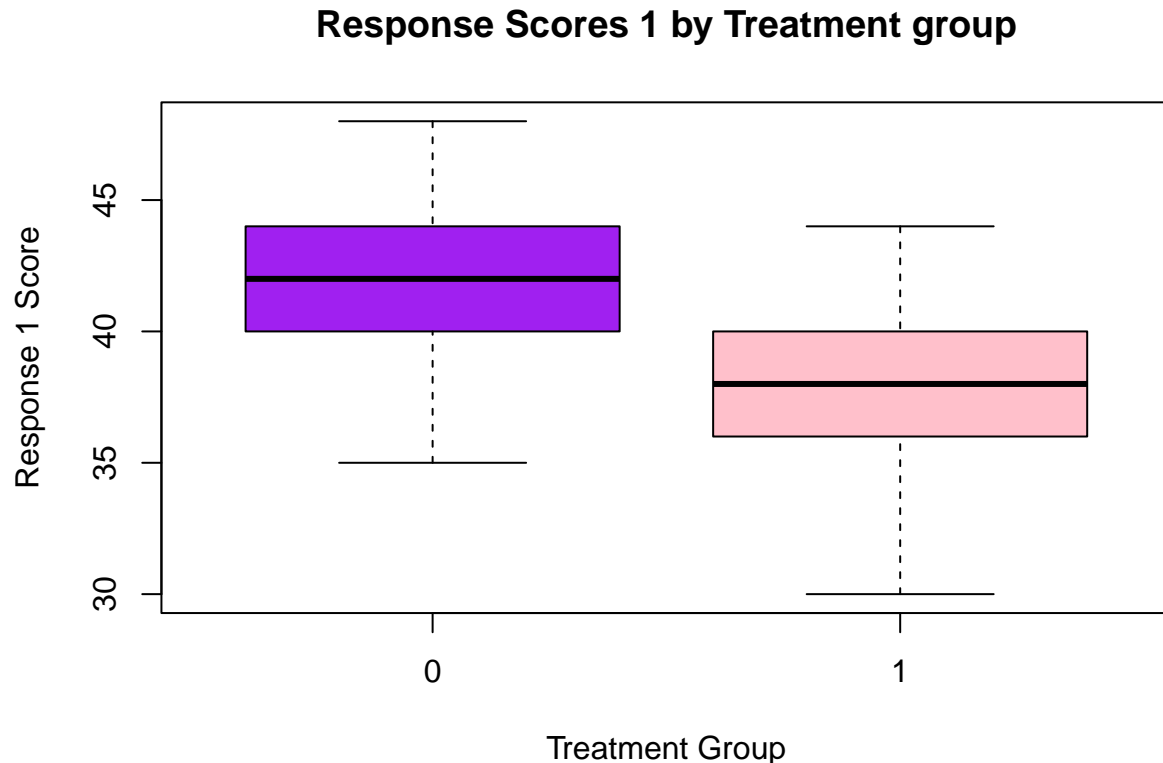
As mentioned in the previous section, nurses were divided into two groups - the experimental one which received a training program to cope with job-related stress (referred to as "1" in a variable "Trt"), and the control one, without intervention ("referred to as"0" in the same variable). The boxplot of the post-test stress scores at time 1 shows differences between treatment and control groups. The median for the experimental group is much lower, which suggests the positive impact of a training program on reducing the level of work-induced stress for nurses.

In order to notice any relationship between the response against time, I've created some additional boxplots.

```
# Combine Responset1, Responset2, and Responset3 into a single column
responset_data <- data.frame(Response = c(MST$Responset1, MST$Responset2, MST$Responset3),
                             Time = factor(rep(c("Responset1", "Responset2", "Responset3"),
                             each = nrow(MST))),
                             stringsAsFactors = FALSE)

# Create boxplots of the response against time, separated by treatment
par(mfrow=c(1,3))

# Boxplot for Response at Time 1
boxplot(Responset1 ~ Trt, data = MST,
        main="Response Scores at Time 1",
        xlab="Treatment Group", ylab="Response Score",
```
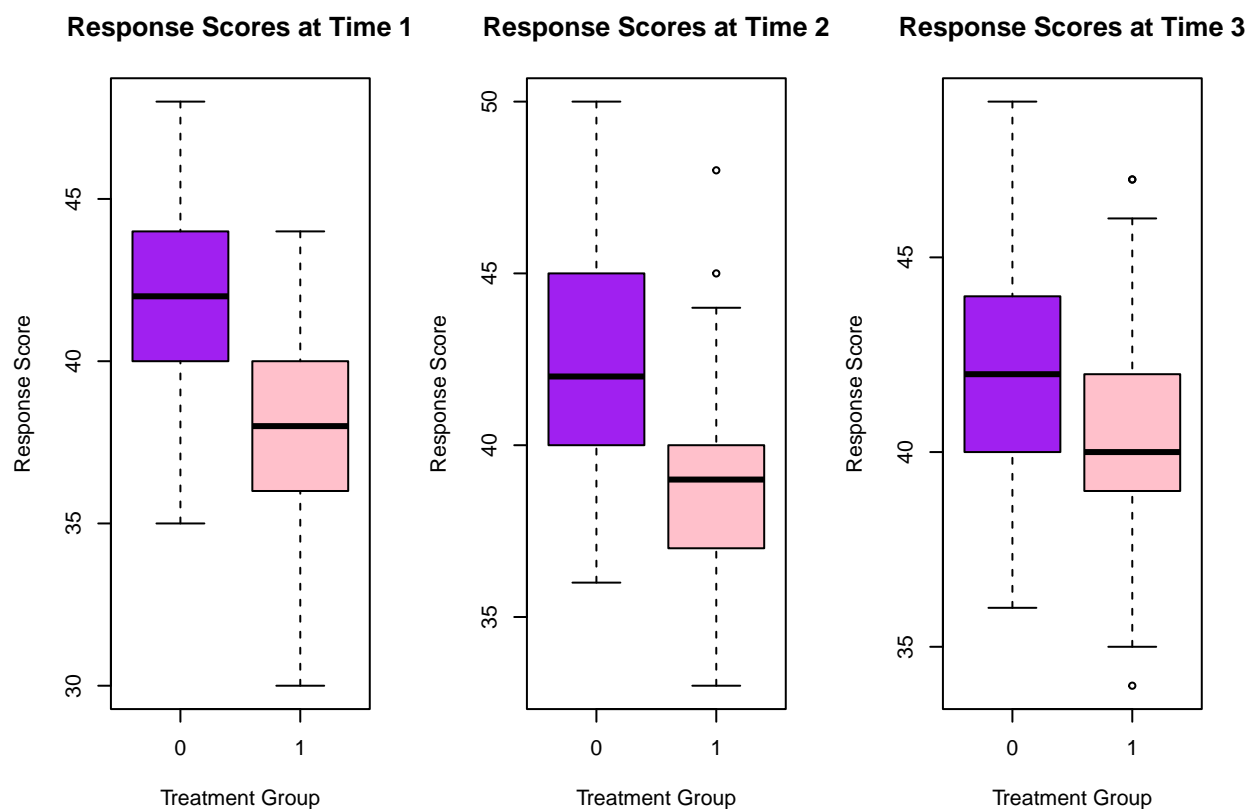
```
        col=c("purple","pink"))

# Boxplot for Response at Time 2
boxplot(Responset2 ~ Trt, data = MST,
        main="Response Scores at Time 2",
        xlab="Treatment Group", ylab="Response Score",
        col=c("purple","pink"))

# Boxplot for Response at Time 3
boxplot(Responset3 ~ Trt, data = MST,
        main="Response Scores at Time 3",
        xlab="Treatment Group", ylab="Response Score",
        col=c("purple","pink"))
```


It suggests that the median of the group that did not receive the treatment is getting closer and closer to the value of the median of the group that received treatment.

Then, I calculated the correlation coefficient between `Responset1` and `Responset3`.

```
cor(MST$Responset1, MST$Responset3)
```

```
## [1] 0.6330064
```

The calculated correlation coefficient between `Responset1` and `Responset3` is approximately 0.63, which indicates a significant positive correlation between the stress scores right after obtaining the training and at

the follow-up time 3. It means that no matter what the level of stress for each nurse at the beginning, it tended to decrease after the training program.

```r
par(mfrow=c(1,3))
boxplot(MST$Responset1 ~ MST$Hospital,
        main="Response Scores 1 by Hospital",
        xlab="Hospital", ylab="Response 1 Score",
        col= "pink")
boxplot(MST$Responset2 ~ MST$Hospital,
        main="Response Scores 2 by Hospital",
        xlab="Hospital",
        ylab="Response 2 Score",
        col= "purple")
boxplot(MST$Responset3 ~ MST$Hospital,
        main="Response Scores 3 by Hospital",
        xlab="Hospital",
        ylab="Response 3 Score",
        col= "lightblue")
```
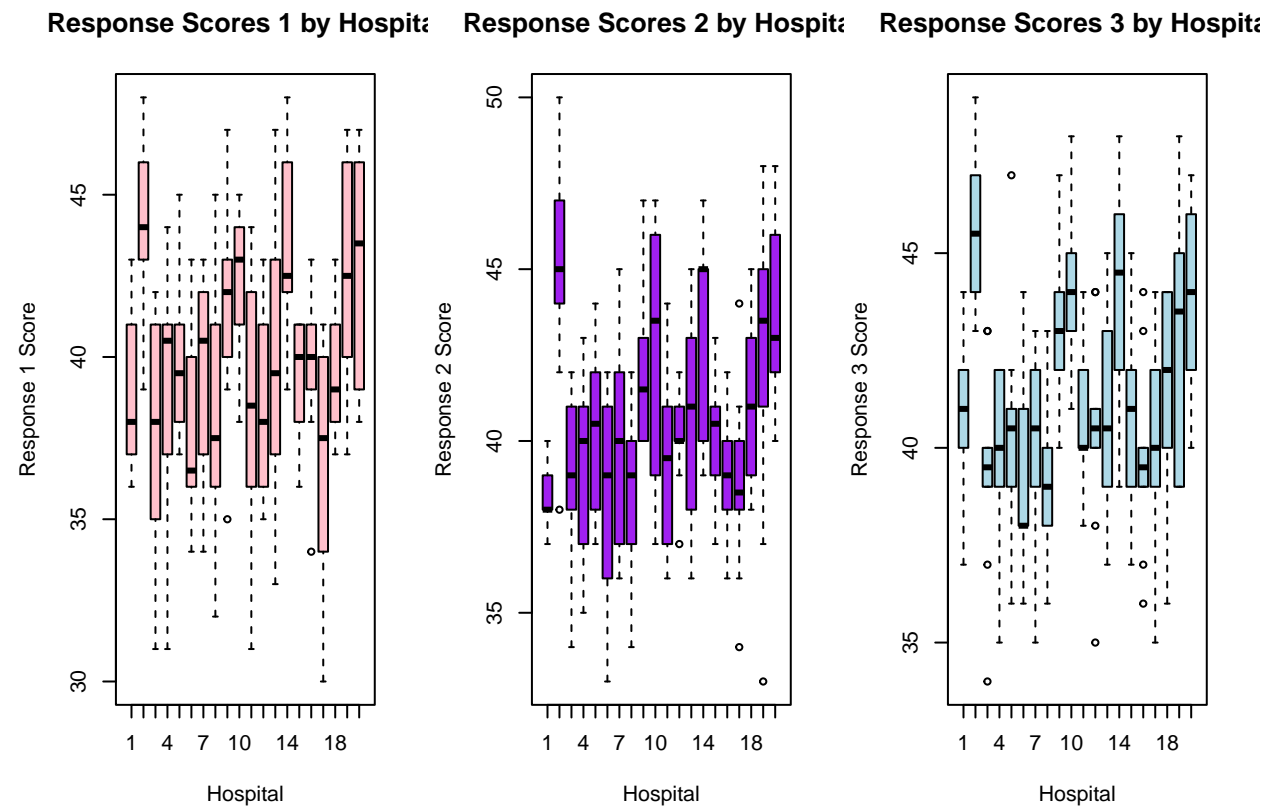

The first figure shows the distribution of post-test stress scores (Response 1 Score) for nurses in 20 hospitals. The x-axis represents the hospital identifier, ranging from 1 to 20. The y-axis shows the response score 1, which represents the nurses' self-reported stress levels.

Given that this is a small sample (approximately 10 nurses per hospital), it is important to keep in mind that in this boxplot I am looking at trends rather than statistically significant results. With that caveat in mind, it appears that there is some variation in the post-test stress scores across the hospitals. Hospitals 2, 14, 19 and 20 appear to have nurses with the highest post-test stress scores. However, it is important to consider

5

that the stress scores could be due to factors other than the training program itself. Therefore, I decided to see if there are any significant changes to this boxplot if I consider `Responset3` instead of `Responset1`.

Although the third figure displays more outliers compared to the first two, it is clear that the stress scores are generally reduced.

```
ggplot(MST, aes(x = Experience)) +
  geom_histogram(binwidth = 1, fill = "skyblue", color = "black") +
  labs(x = "Experience (years)", y = "Frequency", title = "Distribution of Experience") +
  theme(axis.text.y = element_text(vjust = 0.5))
```
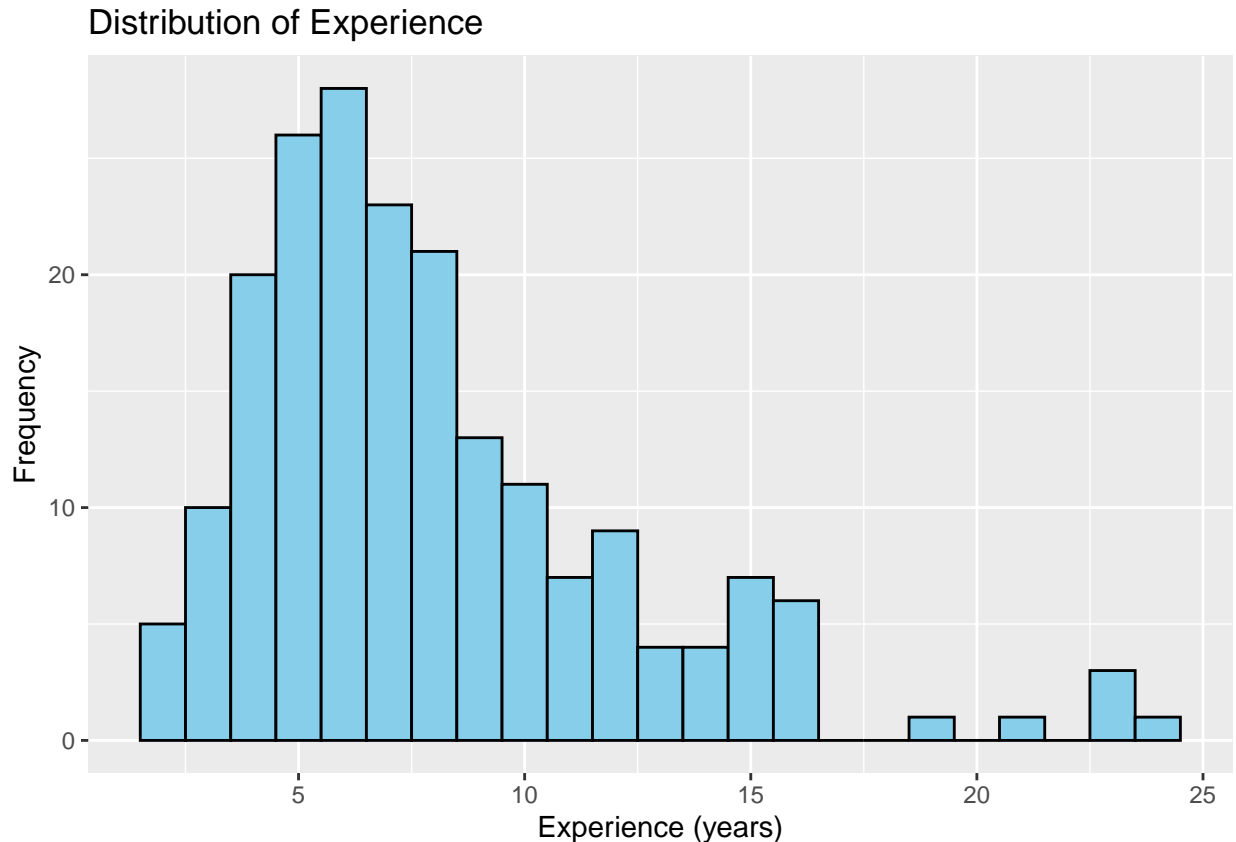

The distribution of experience is skewed to the left, showing that the vast majority of nurses have betwen 0 to 16 years of experience.

```
ggplot(MST, aes(x = Experience, y = Responset1)) +
  geom_point(alpha = 0.6, color = "blue") +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(x = "Experience (years)", y = "Responset1",
       title = "Relationship between Experience and Responset1")
```

Relationships between variables

```
## `geom_smooth()` using formula = 'y ~ x'
```

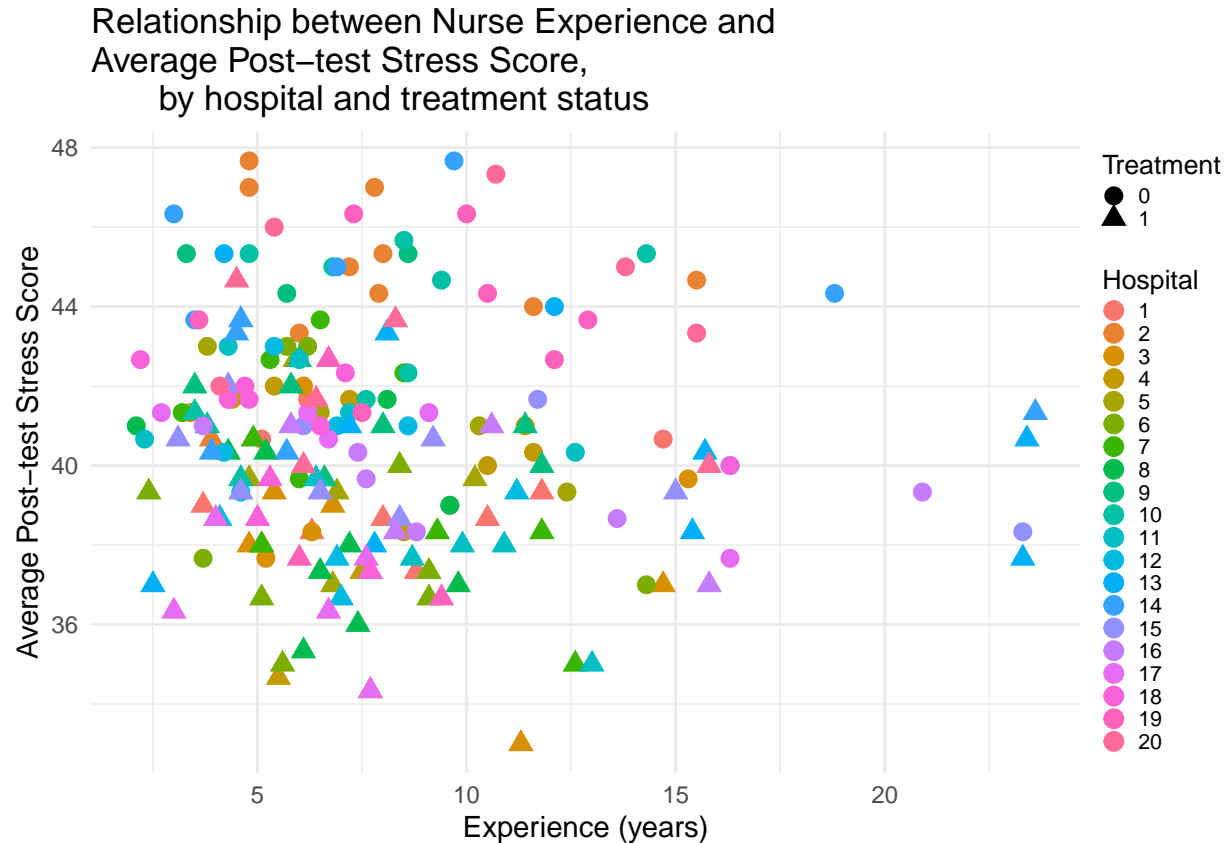## Relationship between Experience and Responset1
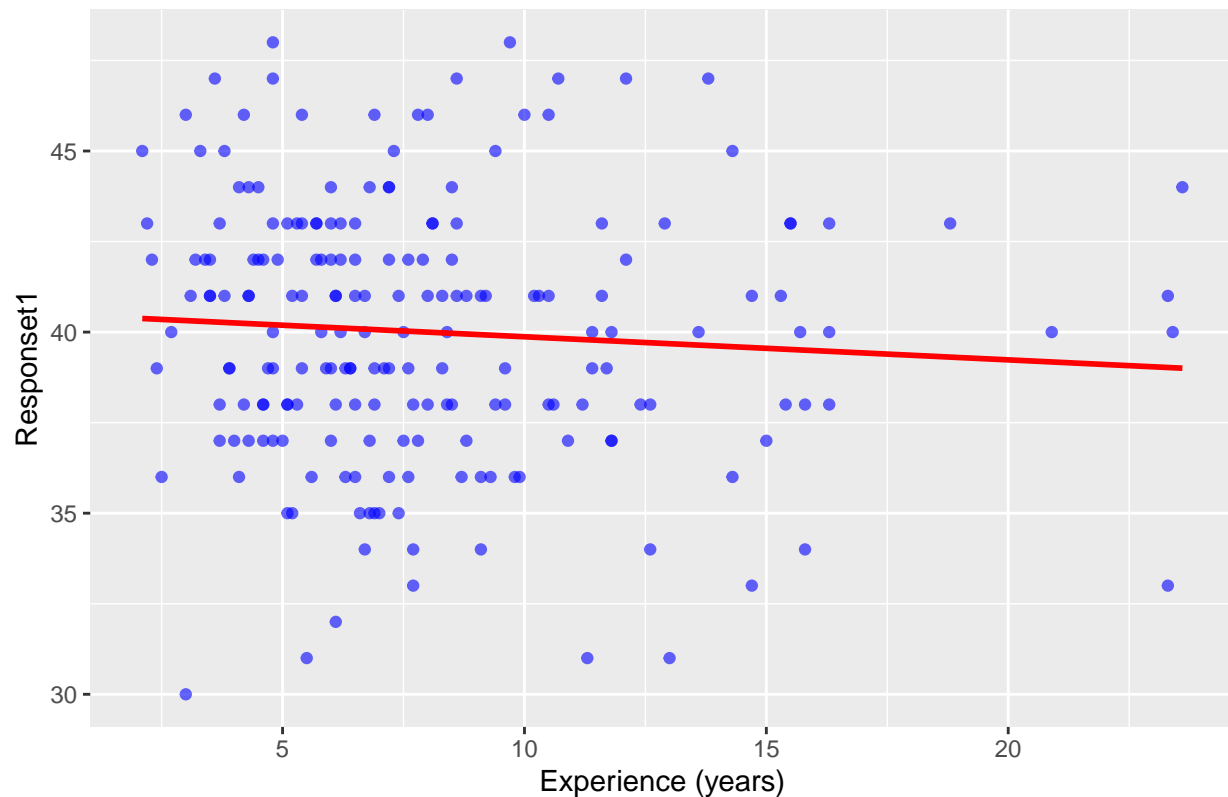
I decided it would be useful to also create a scatterplot displaying the results, for instance, coloured by `Hospital` and separated by `Trt`. To do so, I had to create a new variable for the combined post-test stress scores.

```
MST <- MST %>%
  mutate(Responset_Avg = (Responset1 + Responset2 + Responset3) / 3)
```

Now, with the newly created variable, the scatterplot can be created.

```
ggplot(MST, aes(x=Experience, y=Responset_Avg, color=as.factor(Hospital), shape=as.factor(Trt))) +
  geom_point(size=3) +
  labs(title="Relationship between Nurse Experience and\nAverage Post-test Stress Score,
       by hospital and treatment status",
       x="Experience (years)",
       y="Average Post-test Stress Score",
       color="Hospital",
       shape="Treatment") +
  theme_minimal() +
  theme(legend.key.height=unit(0.1,"cm"),
        legend.key.width=unit(0.1, "cm"),
        legend.spacing.y=unit(0.1, "cm"),
        legend.box.spacing=unit(0.1, "cm"),
        legend.title=element_text(size = 10),
        legend.text=element_text(size = 8),
        legend.position="right")
```

Relationship between Nurse Experience and Average Post−test Stress Score, by hospital and treatment status

Based on this scatterplot, as well as previous figure `Relationship between Experience and Responset1` it appears that there is no clear relationship between years of experience and average positive stress score.

Methods

This section explains why multilevel models are crucial for analysing this dataset and the specific model choices considered for the report will be discussed.

Multilevel Models and Nested Data   Multilevel models are statistical models that are used to analyse data with either hierarchical or nested structures. They are relevant for analysing data where observations are grouped within higher-level units, such as individuals within hospitals, students within schools, or even repeated measurements within subjects over time. What they do is recognise the existence of such data hierarchies, by enabling residual components at each level in the hierarchy.

Multilevel models are implemented, as there should be some similarities of lower-level units belonging to the same upper-level items which cause a correlation between the units. For example, using simple regression ignores group structure and therefore can lead to smaller standard errors of test statistics which eventually might lead to insignificant effects labelled as significant.

In this particular case, multilevel models are relevant because nurses are nested within hospitals. This hierarchical structure implies that there may be some variability in outcomes not only at the individual nurse level but also at the hospital level. By choosing to use multilevel models, it is possible to account for this nested structure and appropriately model the dependence among observations within the given hospital.

It can be said, that this dataset has a longitudinal nature, as there are repeated measurements of stress scores over a given time and therefore it is beneficial to use longitudinal multilevel models. Such models

can handle both the within-subject correlation (due to repeating measurement) and the between-subject variability (due to individual differences).

To decompose the variance in the data, we can calculate the Intraclass Correlation Coefficient (ICC), which quantifies the proportion of total variance that is attributable to between-hospital variation. This helps us understand the extent to which outcomes are similar or dissimilar within hospitals compared to between hospitals.

Decomposing variance In multilevel modelling, variance decomposition refers to the process of partitioning the total variance in the outcome variable into components attributable to different levels of the hierarchy. This decomposition helps us to understand the proportion of variability in the outcome that exists between higher-level units (i.e. hospitals) and within lower-level units (i.e. nurses within hospitals).

The intraclass correlation coefficient (ICC) is calculated as a ratio of:

$$\frac{\text{variance of interest}}{\text{total variance}} = \frac{\text{variance of interest}}{\text{variance of interest} + \text{unwanted variance}}$$

In terms of this dataset, the ICC would help us understand the proportion of variance in post-test stress scores that is attributable to differences between hospitals compared to differences within said hospitals. Values of ICC range from 0 to 1, where a high value indicates a stronger clustering effect at the higher level.

Another measure, similar to ICC, is the variance partition coefficient (VPC). It is commonly used in logistic or generalised linear mixed methods. It is an estimation of the proportion of variance that exists at each level of the model hierarchy relative to the total variance. The VPC provides insight into the relative importance of different levels in explaining the variability in the outcome.

Modeling Strategy

For model building, variables are attributed to specific levels. In this case, `Gender` and `Experience` will be measured at the nurse level, while `Size` will be measured at the hospital level. We adopt a bottom-up modeling strategy, beginning with a null model for each time stamp and gradually incorporating lower-level and upper-level variables. We will use appropriate hypothesis tests such as t-tests and likelihood ratio tests (LRT) to evaluate model fit and variable significance.

Following the modeling strategy, we will construct a two-level model with random intercepts and slopes. We will then compare the nested models using LRT to assess the significance of additional predictors and interactions. Finally, we will select the model with the best fit based on statistical criteria and theoretical considerations.

Analysis

First time stamp - `Responset1` First, I am starting with null model for `Responset1`, which will be then essential for calculating ICC value.

```
model1 <- lmer(formula = Responset1 ~ 1+ Trt + (1+ Trt|Hospital), data = MST)
summary(model1)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Responset1 ~ 1 + Trt + (1 + Trt | Hospital)
##    Data: MST
##
## REML criterion at convergence: 938.3
```

```
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.51923 -0.59564  0.05227  0.64995  1.99424
##
## Random effects:
##  Groups    Name         Variance Std.Dev. Corr
##  Hospital (Intercept) 3.0762   1.7539
##           Trt          0.1437   0.3791   0.08
##  Residual              5.2729   2.2963
## Number of obs: 200, groups:  Hospital, 20
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  42.1747     0.4565 18.7293   92.39  < 2e-16 ***
## Trt          -4.4512     0.3528 16.7037  -12.62 5.82e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##     (Intr)
## Trt -0.339
```

According to the summary of this model, the post-test stress score lowered by 4.4512 points when training was implemented. It is visible that there is a significant variation between hospitals, as random effect variances are 3.0762 and 0.1437 for the `Hospital` and `Trt` groups respectively. There is also a really low value of the correlation of random effects, which is 0.08 to be precise. This indicates a weak positive correlation. I also noticed a high residual variance (5.2729), which may imply that there are other factors influencing the responses except for `Trt` and `Hospital`. It will be further evaluated later in this section.

Next, I decided it would be useful to have a look at the ICC value for this model:

```
icc(model1)
```

```
## # Intraclass Correlation Coefficient
##
##     Adjusted ICC: 0.378
##   Unadjusted ICC: 0.238
```

This value for adjusted ICC suggests a moderate clustering effect of post-test stress scores within hospitals, highlighting how important it is to account for the hierarchal structure of the dataset.

Now, I have decided to move on to the random intercept model, by appropriate fitting:

```
model0 <- lmer(formula = Responset1 ~ 1+ Trt+(1|Hospital), data=MST)
summary(model0)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Responset1 ~ 1 + Trt + (1 | Hospital)
##    Data: MST
##
## REML criterion at convergence: 938.3
```

```
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.52815 -0.61155  0.05405  0.65237  1.98542
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Hospital (Intercept) 3.153    1.776
##  Residual             5.309    2.304
## Number of obs: 200, groups:  Hospital, 20
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  42.1755     0.4609  25.0836   91.50   <2e-16 ***
## Trt          -4.4499     0.3433 185.2218  -12.96   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##     (Intr)
## Trt -0.365
```

The next step is producing a plot of regression lines for each hospital.

```
MST$pred0 <- predict(model0)

ggplot(MST,
       aes(x= Trt, y = Responset1, col = Hospital,
           group = Hospital))+
       geom_line(aes(y=pred0, group=Hospital, col=Hospital)) +
       scale_color_gradientn(colours = rainbow(20)) +
       labs(title="The plot of regression lines for each hospital",
            x="Treatment",
            y="Post-test stress scores at time 1")
```

The plot of regression lines for each hospital
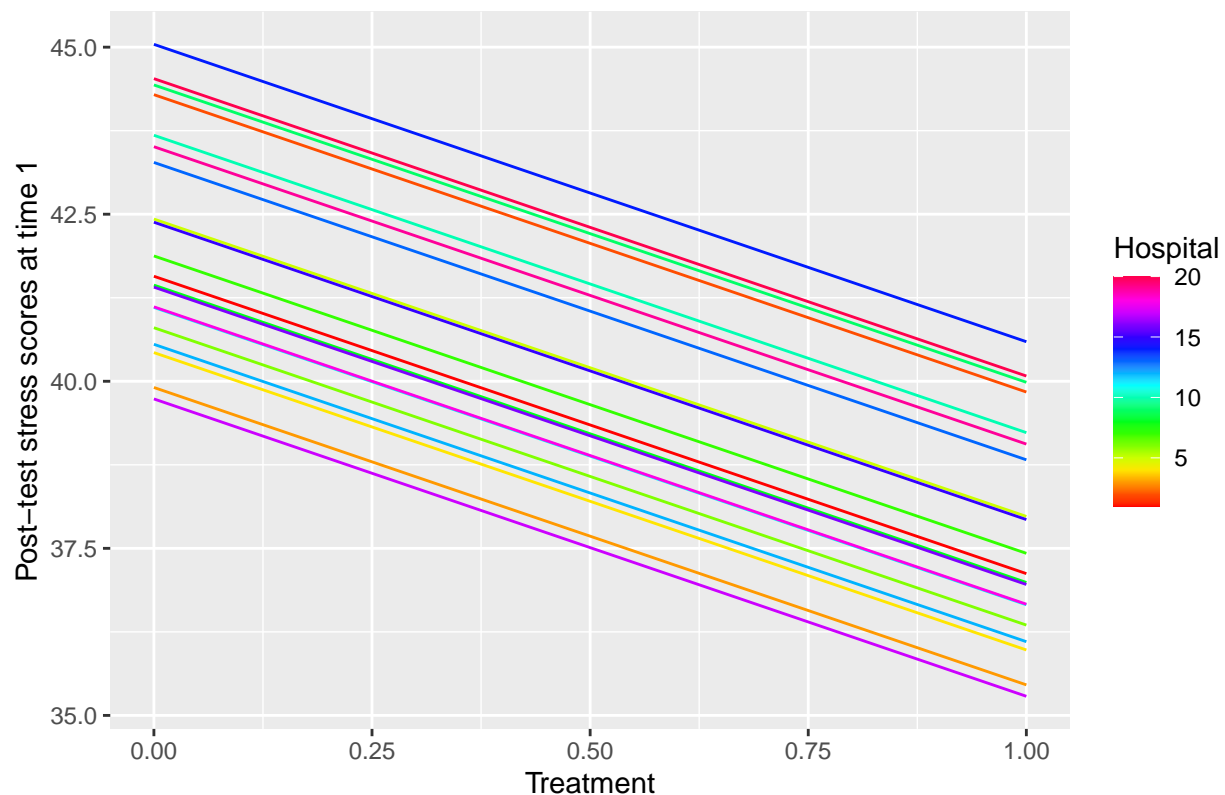
As it was suspected, all the slopes are the same.

Now, I fitted a model with a random slope for `Trt` and performed a likelihood ratio test:

```
model0_slope <- lmer(formula= Responset1 ~ Trt + (1 + Trt|Hospital), data=MST)
lrt <- anova(model0, model0_slope)
```

```
## refitting model(s) with ML (instead of REML)
```

```
lrt
```

```
## Data: MST
## Models:
## model0: Responset1 ~ 1 + Trt + (1 | Hospital)
## model0_slope: Responset1 ~ Trt + (1 + Trt | Hospital)
##              npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## model0          4 946.15 959.34 -469.08   938.15
## model0_slope    6 950.11 969.90 -469.06   938.11 0.0388  2     0.9808
```

Performing this test allowed me to test whether there is evidence in heterogeneity of the intervention effect between hospitals. Since the p-value associated with test statistics is 0.9808, so much greater than the significance level, it means that there is no significant evidence. It means, that the treatment program seems to have a consistent effect in decreasing stress scores across different hospitals and variation in it does not seem to be strongly influenced by which hospital they belong to.

```
model2 <- lmer(Responset2 ~ Trt + (1 | Hospital), data=MST)
summary(model2)
```

Second time stamp = Responset2

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Responset2 ~ Trt + (1 | Hospital)
##    Data: MST
##
## REML criterion at convergence: 924.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1623 -0.6495  0.0645  0.6300  2.8344
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Hospital (Intercept) 2.832    1.683
##  Residual             4.967    2.229
## Number of obs: 200, groups:  Hospital, 20
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  42.1759     0.4392  25.2844   96.03   <2e-16 ***
## Trt          -3.1140     0.3320 185.4114   -9.38   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##     (Intr)
## Trt -0.370
```

It is interesting to observe that although the post-test stress score still lowers, it is happening at a decreased rate (3.1140 scores compared to 4.4512 of `model1`). The residual variance is a bit lower compared to `Responset1`, therefore it indicates that other variables than `Trt` have less impact on post-test stress scores later in time.

```
model3 <- lmer(Responset3 ~ Trt +  (1 | Hospital), data=MST)
summary(model3)
```

Third time stamp = Responset3

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Responset3 ~ Trt + (1 | Hospital)
##    Data: MST
##
```

```
## REML criterion at convergence: 935
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.1670 -0.6351 -0.0321  0.6701  3.1691
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Hospital (Intercept) 2.733    1.653
##  Residual             5.278    2.297
## Number of obs: 200, groups:  Hospital, 20
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  42.1646     0.4372  25.7593   96.45  < 2e-16 ***
## Trt          -1.7747     0.3420 185.8646   -5.19 5.48e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##     (Intr)
## Trt -0.383
```

Once again, it is visible here that even if the post-test stress score keeps lowering, it is continuously happening at a decreased rate (now it is -1.7747 which compared to the initial -4.4512 is a significant decline).

```r
MST$responset_data <- rowSums(MST[, c("Responset1", "Responset2", "Responset3")])

model_ALL <- lmer(responset_data ~ Trt + (1 | Hospital), data=MST)
summary(model_ALL)
```

Combined stress - final fitted model

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: responset_data ~ Trt + (1 | Hospital)
##    Data: MST
##
## REML criterion at convergence: 1266.4
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.42568 -0.67886  0.04836  0.61645  2.20179
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Hospital (Intercept) 26.59    5.156
##  Residual             26.64    5.161
## Number of obs: 200, groups:  Hospital, 20
##
## Fixed effects:
```

```
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept) 126.4968     1.2671  22.8096   99.83   <2e-16 ***
## Trt          -9.2995     0.7717 183.0572  -12.05   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##     (Intr)
## Trt -0.298
```

This model gives us an estimate of `Trt` equal to -9.2995, which means that overall nurses in the treatment group tend to get a 9.2995 lower stress score, compared to nurses from the control group. This is a proof that the intervention effect is significant.

Next, I decided to fit a model that includes a nurse level covariates - `Experience` and `Gender`, as well as a hospital level covariate - `Size`. By adding all these covariates in one comprehensive model, I can control their effects simultaneously and better understand the combined impact on stress.

```
model_FINAL <- lmer(responset_data ~ Trt + Experience + Gender + Size + (1|Hospital),
                    data=MST)
summary(model_FINAL)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: responset_data ~ Trt + Experience + Gender + Size + (1 | Hospital)
##    Data: MST
##
## REML criterion at convergence: 1187.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.8592 -0.6783  0.1413  0.6801  2.2812
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Hospital (Intercept) 12.59    3.548
##  Residual             18.87    4.344
## Number of obs: 200, groups:  Hospital, 20
##
## Fixed effects:
##              Estimate Std. Error        df t value Pr(>|t|)
## (Intercept) 121.89690    1.48224  39.56303  82.238  < 2e-16 ***
## Trt          -9.34112    0.64942 182.51054 -14.384  < 2e-16 ***
## Experience   -0.26171    0.07733 182.01665  -3.384 0.000874 ***
## Gender        4.95430    0.66750 181.75737   7.422 4.29e-12 ***
## Size          7.36340    1.70294  18.02602   4.324 0.000408 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) Trt    Exprnc Gender
## Trt        -0.261
## Experience -0.467  0.060
## Gender     -0.335  0.015  0.160
## Size       -0.567  0.029 -0.018 -0.022
```

It looks like each year of nurse experience results in a decrease in stress level by 0.26171 score points. Also, female nurses tend to obtain approximately 4.95430 units higher stress scores compared to male nurses, and the p-value associated with it is low, making the results significant. Nurses working in bigger A&E departments seem to obtain approximately 7.36340 higher stress scores compared to nurses working in smaller A&E departments. All of these data suggest that all these covariates are important in explaining stress and therefore I decided to use them all in my final model.

Since my final fitted model is done, I would like to add confidence interval based on it:

```
confint(model_FINAL)
```

```
## Computing profile confidence intervals ...
```

```
##                    2.5 %       97.5 %
## .sig01         2.3510070    4.9085724
## .sigma         3.8989203    4.7949568
## (Intercept) 119.0414617  124.7594891
## Trt          -10.6049232   -8.0685337
## Experience    -0.4118867   -0.1096939
## Gender         3.6554026    6.2628663
## Size           4.0431581   10.6840302
```

These confidence intervals give us information about the uncertainty connected with the estimated coefficients in the final fitted model.

Now, I will move on to assessing the assumptions and diagnostics for this final model, starting with residual and normal Q-Q plots. Then, I will conduct the Shapiro-Wilk test for normality, and finally, I will assess VIF.

```
par(mfrow=c(2,2))
plot(model_FINAL, main= "Residual Plot", xlab="Fitted values", ylab="Residuals")
```

Residual Plot
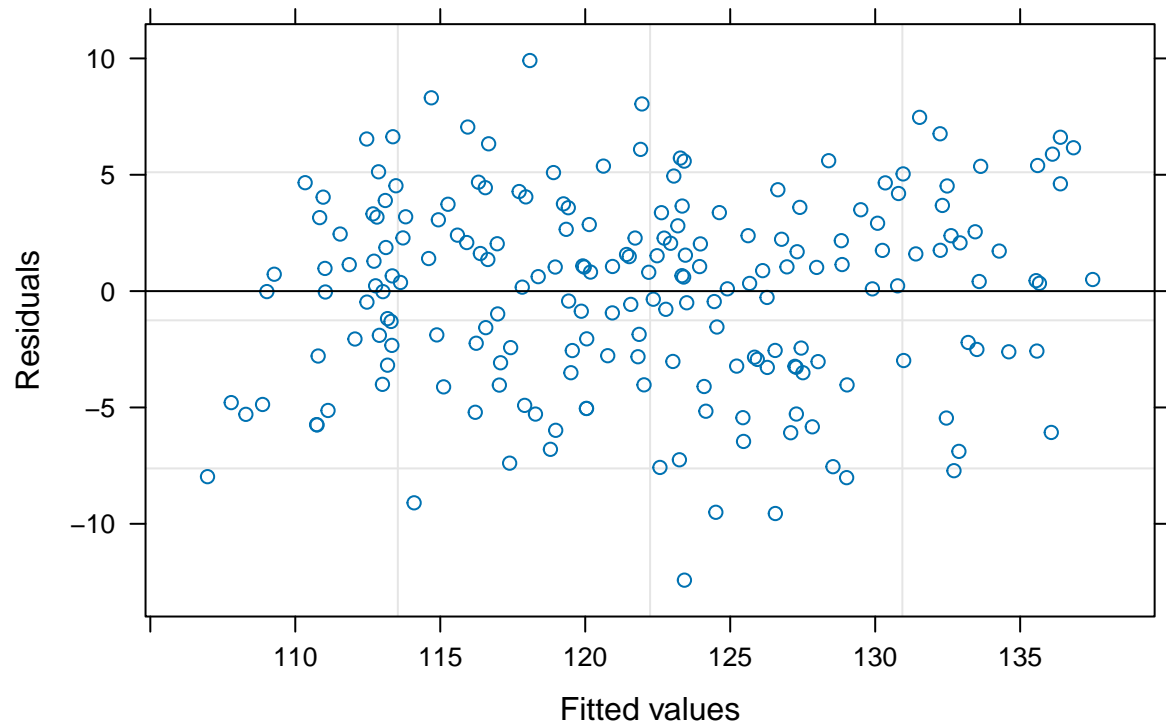

```r
qqnorm(resid(model_FINAL), main="Normal Q-Q Plot", xlab="Theoretical Quantiles",
       ylab="Sample Quantiles")
qqline(resid(model_FINAL))
```
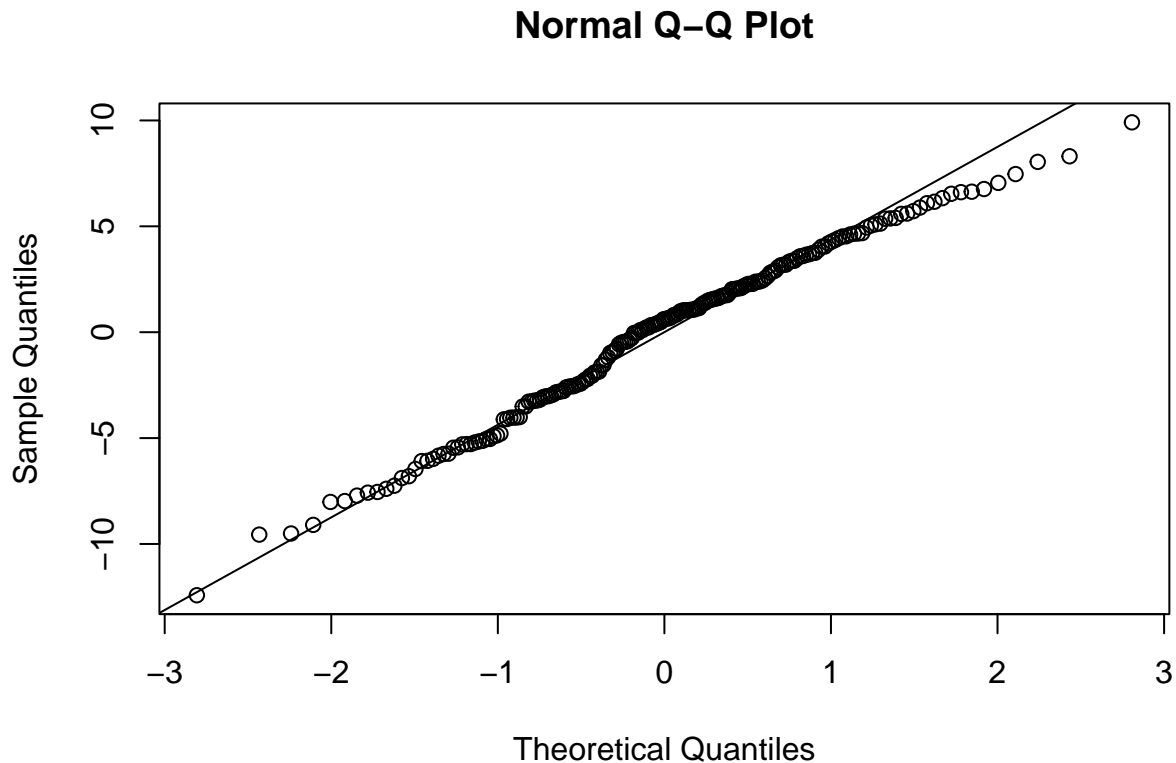
Normal Q–Q Plot

It is visible that the residuals more or less follow a straight line on a Q-Q plot. There are some departures from normality at both ends.

```
shapiro.test(resid(model_FINAL))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resid(model_FINAL)
## W = 0.98789, p-value = 0.08675
```

The p-value is greater than 0.05, therefore there is not enough evidence of departure from normality. The visual inspection of the Q-Q plot supports the assumption of normality.

```
vif_values <- car::vif(model_FINAL)
print(vif_values)
```

```
##       Trt Experience     Gender       Size
##  1.004593   1.029965   1.026549   1.001636
```

All the VIF values are close to 1, indicating scant multicollinearity. It is a good thing, as it suggests that all of these predictor variables provide us with a unique set of information and therefore are not redundant with the others.

When fitting the final model, some assumptions were considered to ensure the validity of the statistical inference. To be more precise, I assessed the linearity, independence, homoscedasticity and normality of

residuals. These assumptions were deemed reasonable based on diagnostic checks, including residual plots and normal Q-Q plots. No significant deviations were observed, indicating that the model assumptions were met adequately.

Discussion of results

The analysis was meant to critically assess the effectiveness of the intervention program in reducing stress levels for nurses, taking into consideration various variables and also examining what changes evolve over time. The intervention can be deemed effective, as it significantly reduced stress scores for nurses (with a p-value of less than 0.001) and they experienced a reduction of approximately 9.34 stress points more compared to the control group.

Other variables than `Trt`, such as years of nurses' experience (`Experience`), their gender (`Gender`) and the size of the A&E department (`Size`) also had an impact on the stress scores.

The results achieved for `model1`, `model2`, `model3`, as well as `model_FINAL` show that the effectiveness of the intervention was visible in each of the time stamps, which implies the persistent benefit of time. Overall, the conducted modelling shows that the intervention effectively reduced the level of stress among nurses, which is what was the ultimate goal of the implemented treatment. This is essential information for any medical-related organisation that would like to lead to better well-being of nurses and ultimately better quality of their work.

Limitations and recommendations Regardless of decent results, there are certain limitations of this modelling method, as well as the dataset itself. This research consisted of qualitative data, which could be interrupted by some distortionary variables such as affecting hormones or other personal issues. Also, these results could be more precise if more nurses and therefore more hospitals would be involved in the collection of data. It is worth highlighting that the scores of stress were only collected in three-time stamps, all of them happening after implementing the training process. Including the stress scores before implementing it, as well as a year after, would be beneficial. First of all, it would be more clear to see the correlation and long-term effect of the intervention. Also, my decision of aggregating the data in the final model removed the ability to assess the treatment effect through time.

Future scholars would benefit from examining this data as longitudinal and expanding the range of analysis.