Final project STAT5060

2023-11-23

Data Introduction

The dataset "Nurses.csv" is from the third edition of Multilevel analysis: Techniques and applications (Hox, J. J., Moerbeek, M., & Schoot, R, 2018), which contains data from a cross-sectional hypothetical study on stress experienced by nurses in hospitals. In each of 25 hospitals, four wards were selected and randomly assigned to an experimental or control condition, where the intervention included a training program to cope with job-related stress. After the intervention was implemented, a random sample of 10 nurses from each ward completed a survey to test job-related stress level.

Variables in the dataset:

hospital: ID number for a hospital $(1, \ldots, 25)$, a total of 25 hospitals

ward: ward numbering within a hospital $(1, 2, 3, 4)$, a total of 4 wards from each hospital

wardid: ID number of ward (not used in programming of analyses)

nurse: ID number for a nurse $(1, \ldots, 1000)$, a total of 1000 nurses (not used in programming)

age: nurse's age in years gender: nurse's gender $(0 = \text{male}, 1 = \text{female})$

experience: nurse's years of experience working stress: outcome variable, a scale from 1 to 7

wardtype: type of hospital ward ('general care', 'special care')

hospize: hospital size ('small', 'medium', 'large')

expcon: intervention indicator variable ('control', 'experiment') administered at Ward

Basic data manipulations

```
library(ggplot2)
library(nlme)
library(lme4)
```

```
## Loading required package: Matrix
```

```
##
## Attaching package: 'lme4'
```

```
## The following object is masked from 'package:nlme':
##
##     lmList
```

```
library(mgcv)
```

```
## This is mgcv 1.8-40. For overview type 'help("mgcv-package")'.
```

```
library(readr)
library(furniture)      # nice table1() descriptives
library(stargazer)      # display nice tables: summary & regression
```

```
##
## Please cite as:

##  Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.

##  R package version 5.2.3. https://CRAN.R-project.org/package=stargazer
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --

## v tibble  3.2.1     v dplyr   1.1.2
## v tidyr   1.2.1     v stringr 1.4.1
## v purrr   0.3.4     v forcats 0.5.2
## -- Conflicts ------------------------------------------- tidyverse_conflicts() --
## x dplyr::collapse() masks nlme::collapse()
## x tidyr::expand()   masks Matrix::expand()
## x dplyr::filter()   masks stats::filter()
## x dplyr::lag()      masks stats::lag()
## x tidyr::pack()     masks Matrix::pack()
## x tidyr::unpack()   masks Matrix::unpack()
```

```
library(dplyr)
nurses <- read_csv("./Nurses.csv")
```

```
## Rows: 1000 Columns: 11
## -- Column specification ------------------------------------------------------
## Delimiter: ","
## chr (3): wardtype, hospsize, expcon
## dbl (8): hospital, ward, wardid, nurse, age, gender, experience, stress
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
nurses$hospsize.n<-recode(nurses$hospsize, 'small'=0, 'medium'=1, 'large'=2)
nurses$wardtype.n<-recode(nurses$wardtype, 'general care'=0, 'special care'=1)
nurses$expcon.n<-recode(nurses$expcon, 'control'=0, 'experiment'=1)
nurses$cgender<-ifelse(nurses$gender==0, 'Male', 'Female')
nurses$age.C <- nurses$age-mean(nurses$age)#mean of 43.0005 years
nurses$experience.C <- nurses$experience-mean(nurses$experience)#mean of 17.057 years
nurses$Chospsize <-(nurses$hospsize.n-1)
nurses$Cexpcon <- (nurses$expcon.n-0.5)
nurses = data.frame(nurses)
head(nurses)
```

```
##   hospital ward wardid nurse age gender experience stress    wardtype hospsize
## 1        1    1     11     1  36      0         11     7 general care    large
## 2        1    1     11     2  45      0         20     7 general care    large
## 3        1    1     11     3  32      0          7     7 general care    large
## 4        1    1     11     4  57      1         25     6 general care    large
## 5        1    1     11     5  46      1         22     6 general care    large
## 6        1    1     11     6  60      1         22     6 general care    large
##         expcon hospsize.n wardtype.n expcon.n cgender   age.C experience.C
## 1 experiment          2          0        1    Male  -7.005       -6.057
## 2 experiment          2          0        1    Male   1.995        2.943
## 3 experiment          2          0        1    Male -11.005      -10.057
## 4 experiment          2          0        1  Female  13.995        7.943
## 5 experiment          2          0        1  Female   2.995        4.943
## 6 experiment          2          0        1  Female  16.995        4.943
##    Chospsize Cexpcon
## 1         1     0.5
## 2         1     0.5
## 3         1     0.5
## 4         1     0.5
## 5         1     0.5
## 6         1     0.5
```

# Exploratory Data Analysis

```
stargazer(data.frame(nurses), title  = "Descriptive statistics, aggregate over entire sample",
                      header = FALSE,
                      type = "text")
```

```
##
## Descriptive statistics, aggregate over entire sample
## ====================================================
## Statistic       N    Mean    St. Dev.   Min    Max
## ----------------------------------------------------
## hospital     1,000 13.016    6.947       1     25
## ward         1,000  2.493    1.119       1      4
## wardid       1,000 132.653  69.476      11     254
## nurse        1,000 500.500 288.819       1    1,000
## age          1,000 43.005   12.042      23     64
## gender       1,000  0.735    0.442       0      1
## experience   1,000 17.057    6.042       1     38
## stress       1,000  4.977    0.980       1      7
## hospsize.n   1,000  0.776    0.689       0      2
## wardtype.n   1,000  0.501    0.500       0      1
## expcon.n     1,000  0.504    0.500       0      1
## age.C        1,000 -0.000   12.042    -20.005 20.995
## experience.C 1,000  0.000    6.042    -16.057 20.943
## Chospsize    1,000 -0.224    0.689      -1      1
## Cexpcon      1,000  0.004    0.500    -0.500  0.500
## ----------------------------------------------------
```
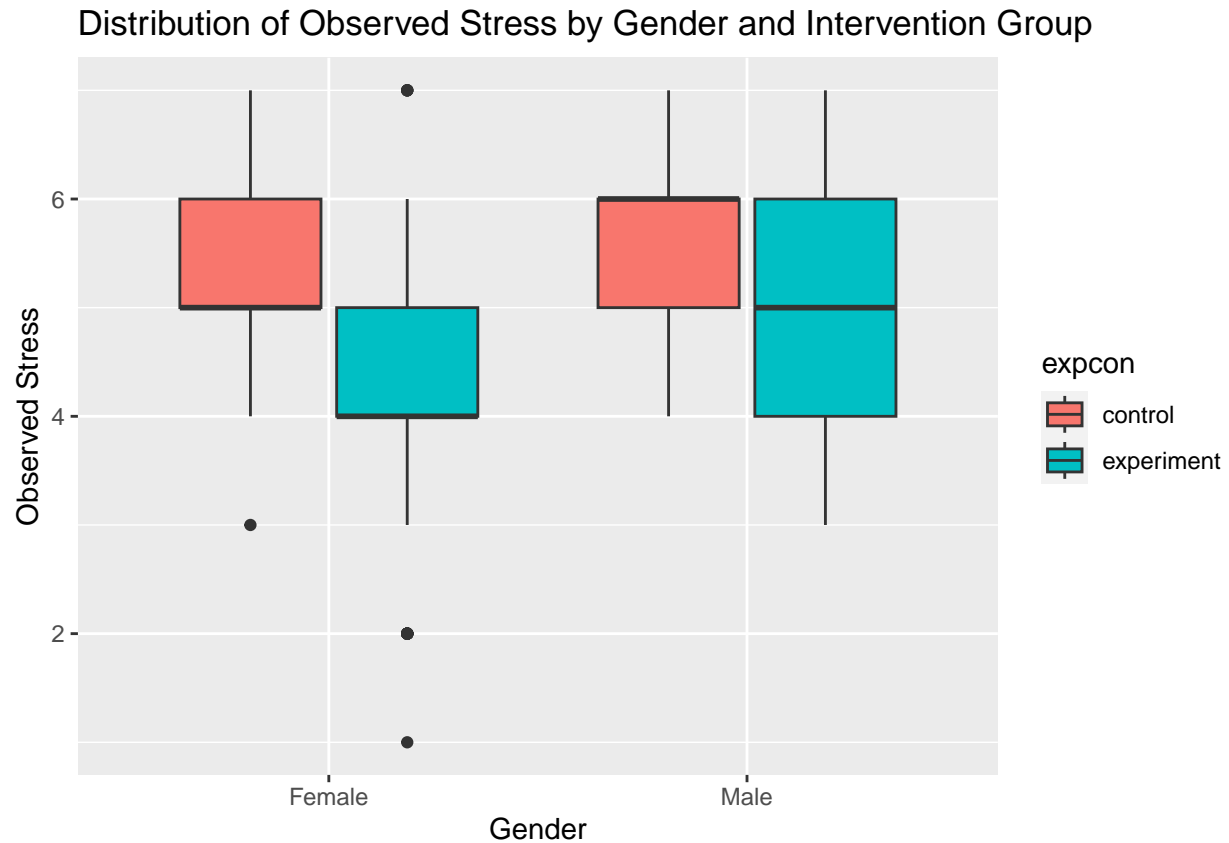
3

```
library(ggplot2)
p1 <- ggplot(nurses, aes(x=age, y=stress))
p1 + geom_smooth(aes(group=expcon, color=expcon), method="loess", se=FALSE) +
  labs(title="Observed Stress versus Nurse's Age by Intervention Group",
       x="Nurse's Age (yrs)", y="Observed Stress")
```

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

Observed Stress versus Nurse's Age by Intervention Group
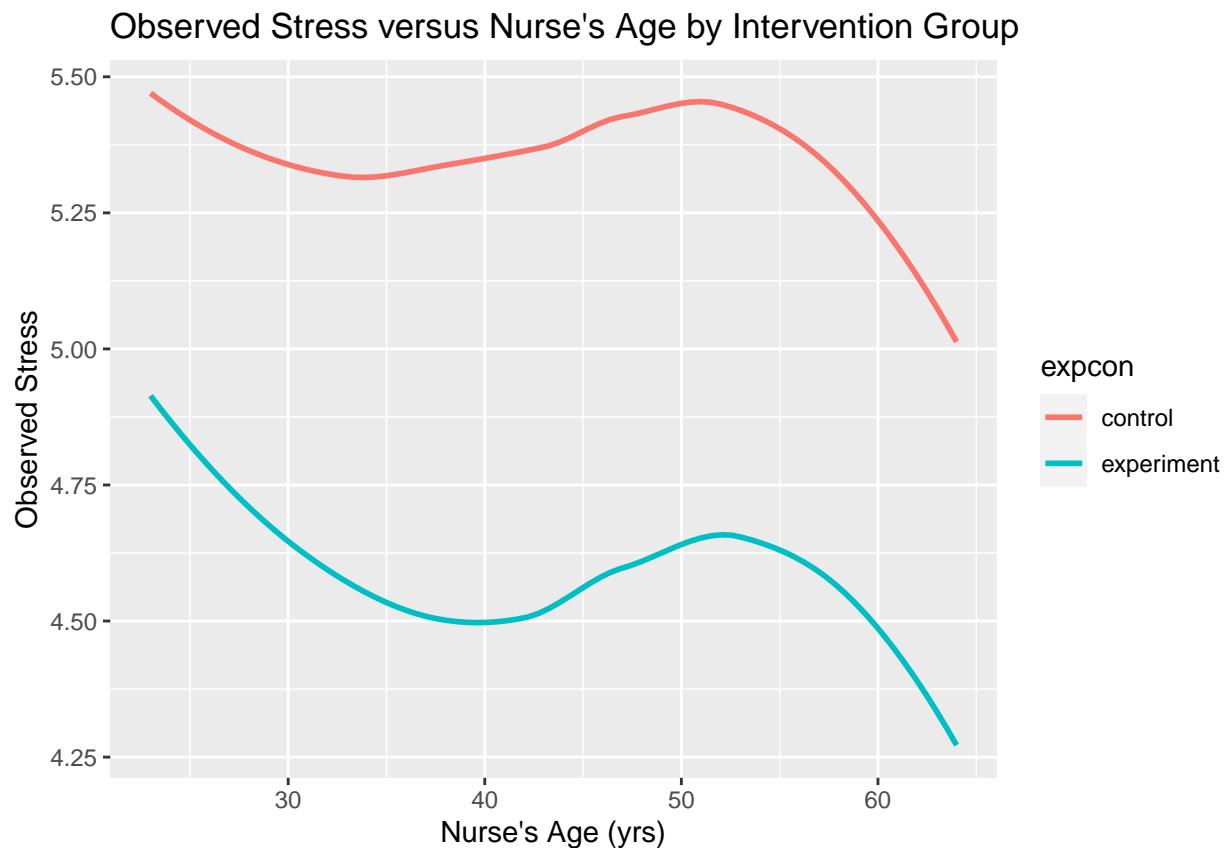
- The relationship between Nurse's age and Stress can be described as either linear if we assume that the smoothing was too smooth or as a bimodal (decreasing through age 40, increasing between age 40 and 52 years, and decreasing after that). I will use a linear relationship between because there are fewer nurses in the older ages, and that could just be an artifact of the data.
- Given that both lines are simply parallel to each other, there is no moderating effect of intervention on this association.
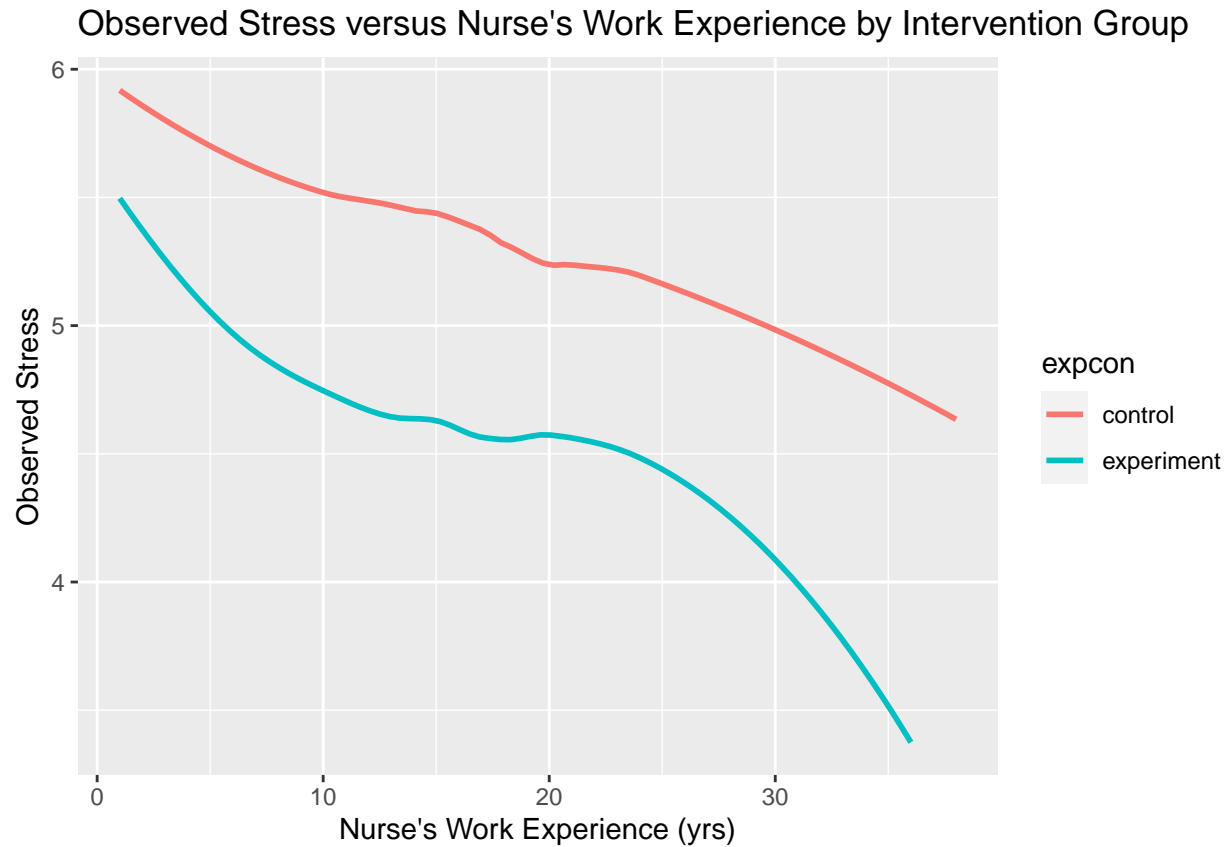
```
p2 <- ggplot(data=nurses, aes(y=stress, x=cgender, fill=expcon))
p2 + geom_boxplot() +
  labs(title="Distribution of Observed Stress by Gender and Intervention Group",
       x="Gender", y="Observed Stress")
```

Distribution of Observed Stress by Gender and Intervention Group

- There may or may not be an association between gender and stress, as on average, stress seems to be very similar between intervention and control groups.
- Because among males, there is a large overlaps in stress, and it is unclear whether there is an intervention effect among females, the visual inspection of the plot does not suggest a moderating effect of intervention on the association between stress and gender.

```r
p3 <- ggplot(nurses, aes(x=experience, y=stress))
p3 + geom_smooth(aes(group=expcon, color=expcon), method="loess", se=FALSE) +
  labs(title="Observed Stress versus Nurse's Work Experience by Intervention Group",
       x="Nurse's Work Experience (yrs)", y="Observed Stress")
```

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

Observed Stress versus Nurse's Work Experience by Intervention Group

- There appears to be a liner relationship between observed stress and nurse's experience in years.

- Given that the two trajectories behave in parallel, this does not suggest a moderating effect of intervention on this association.

```
p4 <- ggplot(data=nurses, aes(y=stress, x=wardtype, fill=expcon))
p4 + geom_boxplot() +
  labs(title="Distribution of Observed Stress by Hospital Ward Type and Intervention Group",
       x="Hospital Ward Type", y="Observed Stress")
```

Distribution of Observed Stress by Hospital Ward Type and Intervention Group
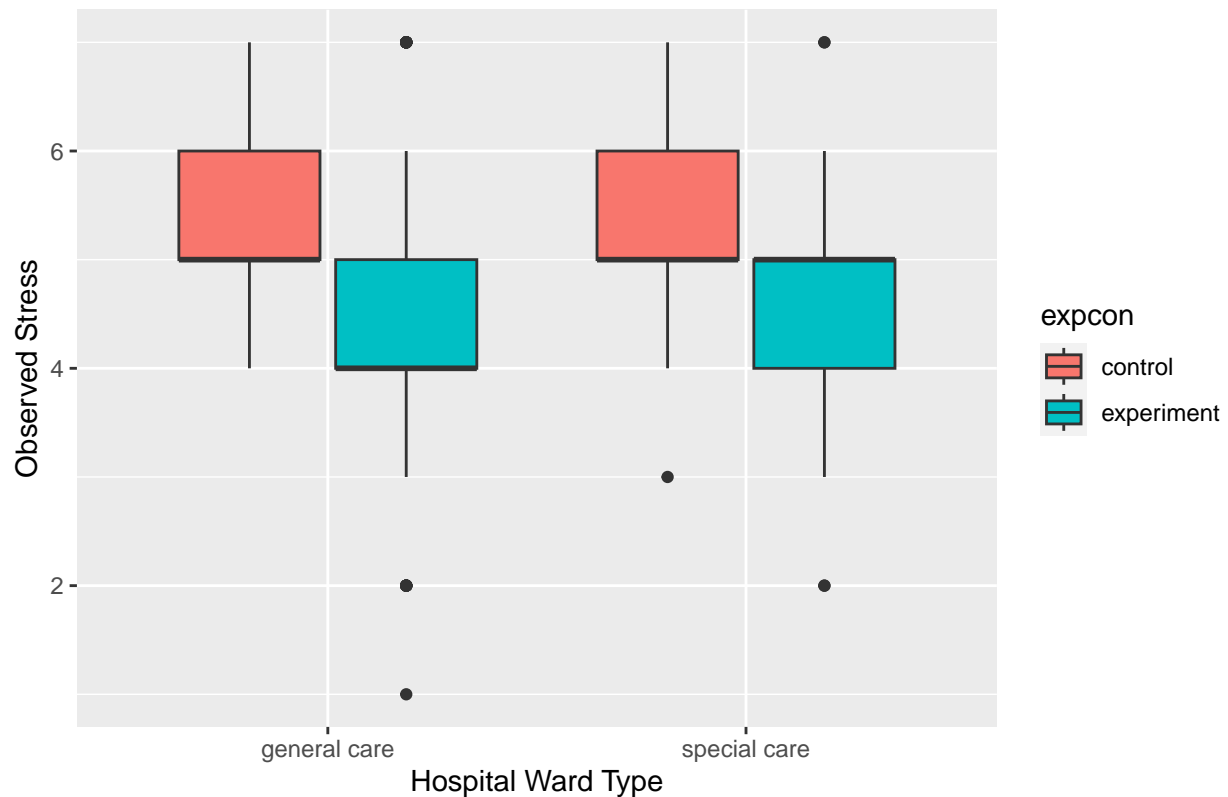
- Stress does not seem to depend on ward type, as if we were to average the control and intervention boxplots within each ward type, they would be approximately the same;

- Since the intervention effect size seems to be the same within each ward type, therefore, there does not seem to be a moderating effect.

```r
p5 <- ggplot(data=nurses, aes(y=stress, x=hospsize, fill=expcon))
p5 + geom_boxplot() +
  labs(title="Distribution of Observed Stress by Hospital Size and Intervention Group",
       x="Hospital Size", y="Observed Stress")
```

Distribution of Observed Stress by Hospital Size and Intervention Group
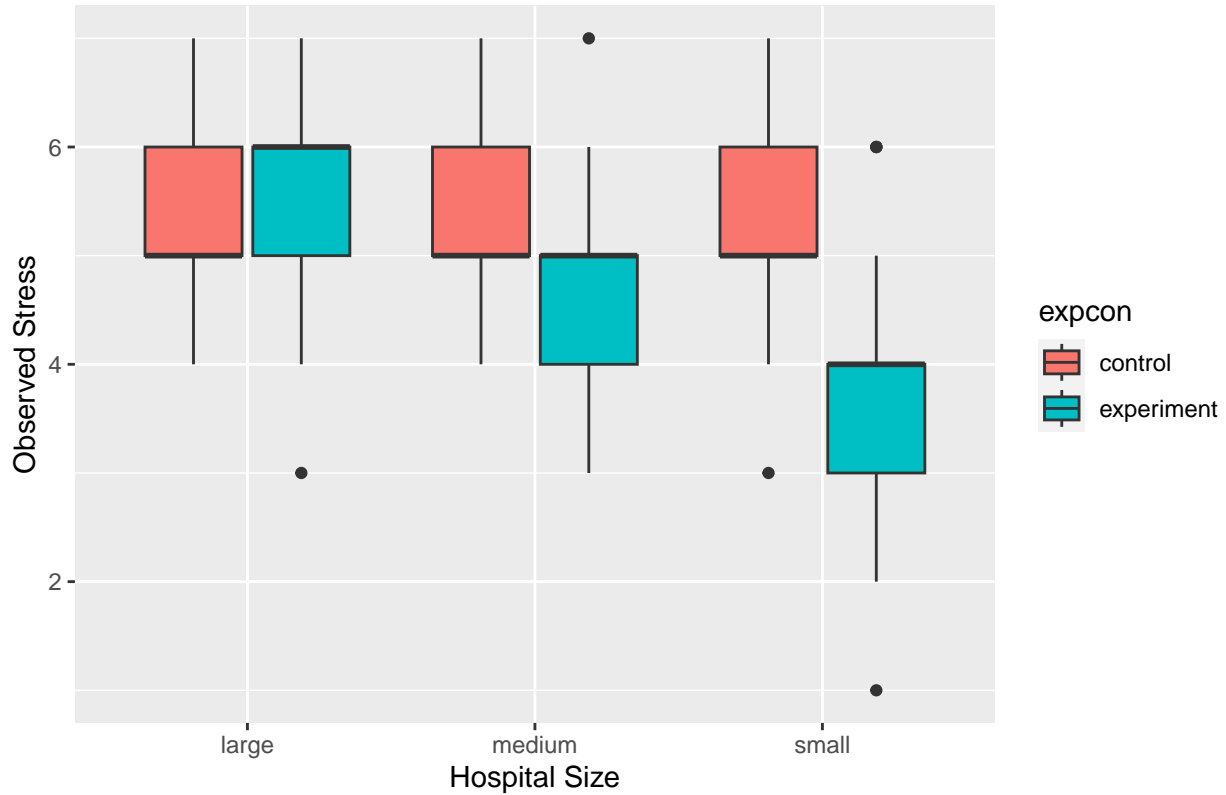
- If you average within each hospital size the effect size of intervention, it seems that there is a linear decline in stress across the hospital size;
- Because the effect size of intervention is different between the hospital sizes, there is a potential moderating effect.

Multi-level models

Three-level model using only a random intercept at each appropriate level of hierarchy

Level 1 (nurse): $Y_{ijk} = \gamma_{0jk} + \gamma_{1jk} \, \mathrm{Sex}_{ijk} + \gamma_{2jk} \, \mathrm{Cexp}_{ijk} + \gamma_{3jk} \, \mathrm{Cage}_{ijk} + \epsilon_{ijk}$

Level 2 (ward): $\gamma_{0jk} = \gamma_{00k} + b_{0jk} + \gamma_{01k} \mathrm{Nwardtype}_{jk} + \gamma_{02k} \mathrm{Cexpcon}_{ik}$

Level 3 (hospital): $\gamma_{00k} = \beta_{000} + b_{0k} + \beta_{001} \mathrm{Chospsize}_k;$

$\gamma_{02k} = \beta_{020} + \beta_{021} \mathrm{Chospsize}_k; \gamma_{01k} = \beta_{010}; \gamma_{1jk} = \beta_{100}; \gamma_{2jk} = \beta_{200}; \gamma_{3jk} = \beta_{300}$

Combined Model: $Y_{ijk} = \beta_{000} + b_{0k} + b_{0jk} + \beta_{001} \mathrm{Chospsize}_k + \beta_{010} \mathrm{Nwardtype}_{ik}$
$+ (\beta_{020} + \beta_{021} \mathrm{Chospsize}_k) \, \mathrm{Cexpcon}_{jk} + \beta_{100} \mathrm{Sex}_{ijk} + \beta_{200} \, \mathrm{Cexp}_{ijk} + \beta_{300} \, \mathrm{Cage}_{ijk} + \epsilon_{ijk}$

```
m1 <- lme(stress ~   age.C + gender + experience.C + wardtype.n + Chospsize*Cexpcon, random = ~ 1+ward|
              data=nurses, na.action=na.omit, method="ML")
summary(m1)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: nurses
##       AIC      BIC    logLik
##   1701.85 1760.743 -838.9249
##
## Random effects:
##  Formula: ~1 + ward | hospital
##  Structure: General positive-definite, Log-Cholesky parametrization
##            StdDev    Corr
## (Intercept) 0.8884047 (Intr)
## ward       0.2721405 -0.887
## Residual   0.5198284
##
## Fixed effects:  stress ~ age.C + gender + experience.C + wardtype.n + Chospsize *    Cexpcon
##                    Value Std.Error  DF   t-value p-value
## (Intercept)     5.382961 0.10546040 969  51.04248  0.0000
## age.C           0.021435 0.00241943 969   8.85954  0.0000
## gender         -0.457715 0.03823967 969 -11.96963  0.0000
## experience.C   -0.059380 0.00485302 969 -12.23571  0.0000
## wardtype.n      0.058101 0.06388064 969   0.90952  0.3633
## Chospsize       0.377360 0.12256341  23   3.07890  0.0053
## Cexpcon        -0.580067 0.11589863 969  -5.00495  0.0000
## Chospsize:Cexpcon  0.596068 0.09508083 969   6.26906  0.0000
##  Correlation:
##                (Intr) age.C  gender expr.C wrdty. Chspsz Cexpcn
## age.C           0.007
## gender         -0.264 -0.010
## experience.C   -0.013 -0.818  0.033
## wardtype.n     -0.476 -0.014 -0.004  0.016
## Chospsize       0.231  0.001  0.005  0.002  0.002
## Cexpcon         0.441  0.012 -0.003 -0.004 -0.805  0.018
## Chospsize:Cexpcon  0.021  0.041  0.006 -0.021  0.013  0.112  0.169
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -3.557085991 -0.672914337  0.003559349  0.678777749  3.170685039
##
## Number of Observations: 1000
## Number of Groups: 25
```

```
VarCorr(m1)
```

```
## hospital = pdLogChol(1 + ward)
##            Variance   StdDev    Corr
## (Intercept) 0.78926291 0.8884047 (Intr)
## ward       0.07406044 0.2721405 -0.887
## Residual   0.27022153 0.5198284
```

```r
icc1 <- (as.numeric(VarCorr(m1)[2])+as.numeric(VarCorr(m1)[4]))/(as.numeric(VarCorr(m1) [2])+as.numeric
icc1
```

```
## [1] 0.7795729
```

```r
icc2<-as.numeric(VarCorr(m1)[2])/(as.numeric(VarCorr(m1)[2])+as.numeric(VarCorr(m1)[4
])+ as.numeric(VarCorr(m1)[5]))
icc2
```

```
## [1] 0.05998712
```

- The estimated correlation among responses (stress level) within the same ward and the same hospital is 58.42%.

- The estimated correlation among responses (stress level) from the same hospital regardless of ward type is 26.20%.

Three-level model adding a random slope for centered intervention indicator variable due to hospital

$$\text{Level 1 (nurse): } Y_{ijk} = \gamma_{0jk} + \gamma_{1jk}\text{Sex}_{ijk} + \gamma_{2jk}\text{Cexp}_{ijk} + \gamma_{3jk}\text{Cage}_{ijk} + \epsilon_{ijk}$$
$$\text{Level 2 (ward): } \gamma_{0jk} = \gamma_{00k} + b_{0jk} + \gamma_{01k}\text{Nwardtype}_{ik} + \gamma_{02k}\text{ Cexpcon }_{ik}$$
$$\text{Level 3 (hospital): } \gamma_{00k} = \beta_{000} + b_{0k} + \beta_{001}\text{Chospsize}_{k};$$

$$\gamma_{02k} = \beta_{020} + \beta_{021}\text{Chospsize}_{k} + b_{02k}; \gamma_{01k} = \beta_{010}; \gamma_{1jk} = \beta_{100}; \gamma_{2jk} = \beta_{200}; \gamma_{3jk} = \beta_{300}$$

$$\text{Combined Model: } Y_{ijk} = \beta_{000} + b_{0k} + b_{0jk} + \beta_{001}\text{Chospsize}_{k} + \beta_{010}\text{Nwardtype}_{jk} +$$
$$d\left(\beta_{020} + \beta_{021}\text{Chospsize}_{k} + b_{02k}\right)\text{Cexpcon}_{jk} + \beta_{100}\text{Sex}_{ijk} + \beta_{200}\text{Cexp}_{ijk} + \beta_{300}\text{Cage}_{ijk} + \epsilon_{ijk}$$

```r
m2 <- lme(stress ~age.C + gender + experience.C + wardtype.n + Chospsize*Cexpcon, random = list(hospital
          data=nurses, na.action=na.omit, method="ML")
summary(m2)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: nurses
##        AIC      BIC    logLik
##   1576.761 1640.562 -775.3804
##
## Random effects:
##  Formula: ~1 + Cexpcon | hospital
##  Structure: General positive-definite, Log-Cholesky parametrization
##            StdDev    Corr
## (Intercept) 0.3893819 (Intr)
## Cexpcon     0.4223545 0.324
##
##  Formula: ~1 | ward %in% hospital
##         (Intercept)  Residual
## StdDev:   0.3297284 0.4654225
##
## Fixed effects:  stress ~ age.C + gender + experience.C + wardtype.n + Chospsize *      Cexpcon
```

```
##                        Value  Std.Error  DF   t-value  p-value
## (Intercept)         5.399837  0.10010103  897   53.94386  0.0000
## age.C               0.022297  0.00220087  897   10.13113  0.0000
## gender             -0.454752  0.03497276  897  -13.00303  0.0000
## experience.C       -0.061808  0.00447027  897  -13.82642  0.0000
## wardtype.n          0.053134  0.07258371   72    0.73203  0.4665
## Chospsize           0.458217  0.12444194   23    3.68218  0.0012
## Cexpcon            -0.499094  0.11626939   72   -4.29257  0.0001
## Chospsize:Cexpcon   0.998446  0.16120292   72    6.19372  0.0000
##  Correlation:
##                    (Intr) age.C  gender expr.C wrdty. Chspsz Cexpcn
## age.C               0.007
## gender             -0.254 -0.014
## experience.C       -0.011 -0.817  0.028
## wardtype.n         -0.362 -0.011 -0.004  0.013
## Chospsize           0.249 -0.001  0.003  0.002  0.000
## Cexpcon             0.201  0.009 -0.004 -0.003  0.000  0.062
## Chospsize:Cexpcon   0.056  0.015 -0.003 -0.010  0.000  0.223  0.280
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -2.77549510 -0.69778033  0.01178434  0.64256902  2.81372299
##
## Number of Observations: 1000
## Number of Groups:
##          hospital ward %in% hospital
##                25               100
```

```
VarCorr(m2)
```

```
##             Variance            StdDev    Corr
## hospital =  pdLogChol(1 + Cexpcon)
## (Intercept) 0.1516182           0.3893819 (Intr)
## Cexpcon     0.1783833           0.4223545 0.324
## ward =      pdLogChol(1)
## (Intercept) 0.1087208           0.3297284
## Residual    0.2166181           0.4654225
```

```
icc2.1 = (0.166 + 0.203 + 0.111) / (0.166 + 0.203 + 0.111 + 0.217)
icc2.2 = (0.166 + 0.111) / (0.166 + 0.203 + 0.111 + 0.217)
icc2.3 = (0.166) / (0.166 + 0.203 + 0.111 + 0.217)
icc2.1
```

```
## [1] 0.6886657
```

```
icc2.2
```

```
## [1] 0.3974175
```

```
icc2.3
```

```
## [1] 0.2381636
```

```
icc2<-as.numeric(VarCorr(m2)[2])/(as.numeric(VarCorr(m2)[2])+as.numeric(VarCorr(m2)[4
])+ as.numeric(VarCorr(m2)[5]))
```

```
## Warning: NAs introduced by coercion
```

```
icc2
```

```
## [1] NA
```

```
intervals(m2)
```

```
## Approximate 95% confidence intervals
##
##   Fixed effects:
##                        lower        est.       upper
## (Intercept)        5.20416450  5.39983659  5.59550868
## age.C              0.01799518  0.02229732  0.02659947
## gender            -0.52311482 -0.45475196 -0.38638910
## experience.C      -0.07054606 -0.06180782 -0.05306958
## wardtype.n        -0.09097931  0.05313373  0.19724676
## Chospsize          0.20182104  0.45821703  0.71461301
## Cexpcon           -0.72994385 -0.49909404 -0.26824423
## Chospsize:Cexpcon  0.67838226  0.99844641  1.31851056
##
##   Random Effects:
##    Level: hospital
##                              lower       est.      upper
## sd((Intercept))          0.2771991 0.3893819 0.5469652
## sd(Cexpcon)              0.2554489 0.4223545 0.6983133
## cor((Intercept),Cexpcon) -0.2634529 0.3238953 0.7360486
##    Level: ward
##                      lower      est.      upper
## sd((Intercept)) 0.2605456 0.3297284 0.4172812
##
##   Within-group standard error:
##      lower       est.      upper
## 0.4444111 0.4654225 0.4874272
```

- Estimated variance of b_{0k} is 0.1516 and it describes the variability in stress across different hospitals (Level 3 variance component);

- Estimated variance of b_{0jk} is 0.1087 and it describes the variability in stress across different wards within the same hospital (Level 2 variance component);

- Estimated variance of b_{02k} is 0.1784 describes variability in intervention effects across hospitals (Level 3 variance component);

- Estimated correlation between b_{0k} and b_{02k} is 0.34, which means the larger the deviation of a hospital's mean for nurses' stress from the overall mean of stress averaged across all hospitals, the larger the deviation for that hospital's intervention effect from the overall intervention effect size averaged across all hospitals.

- Estimated residual variance of 0.2166 is the estimated variance of epsilon and it captures the variability of the individual nurse's stress levels around the mean stress level for a specific ward within a hospital.

Similar to M1, the majority of heterogeneity in stress (response variable) is explained by the correlation among responses from the same ward within a hospital, but it is now composed of an additional source of heterogeneity due to the varying effect sizes of intervention across different hospitals: the variance of b_{0k} increased from 0.1361 (M1) to 0.1516 (M2), the variance of b_{0jk} decreased from 0.1676 (M1) to 0.1087 (M2) but this decrease is compensated by the additional variance of 0.1783 from the random effect of the heterogeneity of intervention effect on stress (which is a ward-level covariate) across different hospitals.

Comparison between two models

To compare the two models, the only difference between them is in the covariance matrix – therefore, we need to conduct a LRT using REML:

The hypotheses are:

$$H_0 : \sigma^2_{b_{02k}} = 0$$

$$H_1 : \sigma^2_{b_{02k}} \neq 0$$

```
m1 <- lme(stress ~   age.C + gender + experience.C + wardtype.n + Chospsize*Cexpcon, random = ~ 1|hospi
          data=nurses, na.action=na.omit)
m2 <- lme(stress ~age.C + gender + experience.C + wardtype.n + Chospsize*Cexpcon, random = list(hospita
              data=nurses, na.action=na.omit)
anova(m1, m2)
```

```
##     Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## m1      1 11 1619.668 1673.565 -798.8343
## m2      2 13 1615.262 1678.959 -794.6311 1 vs 2 8.406378  0.0149
```

This is the likelihood ratio test (LRT), implemented via REML, with df=2, G2=8.406 and p=0.0149 (smaller than 0.10), indicating that the addition of the random effect of intervention across different hospitals is significantly improving the fit of the model to the data.

Hence, we decide to choose model2.

```
summary(m2)
```

```
## Linear mixed-effects model fit by REML
##   Data: nurses
##        AIC      BIC    logLik
##   1615.262 1678.959 -794.6311
##
## Random effects:
##  Formula: ~Cexpcon | hospital
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev    Corr
## (Intercept) 0.4086432 (Intr)
## Cexpcon     0.4501755 0.315
##
##  Formula: ~1 | ward %in% hospital
##         (Intercept)  Residual
## StdDev:   0.3337242 0.4661913
##
## Fixed effects:  stress ~ age.C + gender + experience.C + wardtype.n + Chospsize *      Cexpcon
```

```
##                     Value  Std.Error  DF   t-value p-value
## (Intercept)       5.399858 0.10327271 897  52.28737  0.0000
## age.C             0.022300 0.00219598 897  10.15512  0.0000
## gender           -0.454716 0.03489625 897 -13.03050  0.0000
## experience.C     -0.061804 0.00446073 897 -13.85515  0.0000
## wardtype.n        0.053069 0.07304260  72   0.72655  0.4699
## Chospsize         0.458206 0.12922692  23   3.54575  0.0017
## Cexpcon          -0.498967 0.12074955  72  -4.13225  0.0001
## Chospsize:Cexpcon 0.998592 0.16741718  72   5.96469  0.0000
##  Correlation:
##                   (Intr) age.C  gender expr.C wrdty. Chspsz Cexpcn
## age.C              0.007
## gender            -0.246 -0.014
## experience.C      -0.011 -0.817  0.028
## wardtype.n        -0.353 -0.011 -0.004  0.013
## Chospsize          0.250 -0.001  0.003  0.002  0.000
## Cexpcon            0.202  0.009 -0.004 -0.003  0.000  0.062
## Chospsize:Cexpcon  0.056  0.015 -0.003 -0.010  0.000  0.223  0.279
##
## Standardized Within-Group Residuals:
##        Min          Q1         Med          Q3         Max
## -2.76185798 -0.69352704  0.01071273  0.64130667  2.80413622
##
## Number of Observations: 1000
## Number of Groups:
##         hospital ward %in% hospital
##               25                100
```

Interpretation of the final model

- *Intercept*: estimated mean stress for a male nurse working on a general ward of a medium sized hospital, with average age and average years of experience, is 5.40 points;
- *Age.c*: **for every one year of increase in age, mean stress level significantly increases by 0.02 points;**
- *Gender*: **female nurses reported on average significantly lower stress than male nurses, and that difference is estimated to be 0.45 points;**
- *Experience.C*: **for every one year of increase in work experience, mean stress level significantly decreases by 0.06 points;**
- *Wardtype.n*: while the estimated mean stress in a specialty care ward was on average higher than in the general ward by 0.05 points, **this difference was neither clinically nor statistically significant or meaningful;**
- *Chospsize, Cexpcon and Chospsize*×Cexpcon should be interpreted together: as hospital size increased from small to medium to large, the significant reduction in mean stress level due to intervention was significantly diminished/diluted: there was **a significant reduction in mean stress level due to intervention at small hospitals** (an estimated reduction of Beta6-Beta7=-0.499-0.998=-1.5), **a significant reduction in mean stress level due to intervention in medium-sized hospitals** (an estimated reduction of Beta6=-0.499 points), and a non-significant increase in large-sized hospitals (an estimated increase of Beta6+Beta7=-0.499+.998=0.499), meaning that **intervention did not seem to significantly impact mean stress level among nurses in large hospitals.**

Residual and random effects visual diagnostic check

```
m.finalb = m2
library(mgcv)
est.cov.mfinalb<-extract.lme.cov(m.finalb)
pred.m.finalb = fitted(m.finalb, level=0:2)
pred.m.finalb0 = fitted(m.finalb, level=0)
pred.m.finalb1=fitted(m.finalb, level=1)
pred.m.finalb2=fitted(m.finalb, level=2)
resid.m.finalb=resid(m.finalb, level=0:2)

cr.final <- solve(t(chol(est.cov.mfinalb))) %*% residuals(m.finalb, level=0)

pred.mfinalb = data.frame(id=nurses$nurse,ctrt=nurses$expcon,trt=nurses$expcon.n,   age=nurses$age,expe:
```

Histogram and a Q-Qplot for the transformed residuals

```
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##       combine
```

```
#Histogram of transformed residuals for final model
p1 <- ggplot(pred.mfinalb, aes(x=resid))
p1<- p1 + geom_histogram(aes(y=..density..), color="black", fill="lightblue") +
  geom_density() + labs(title="Distribution of Transformed Residuals-LME Final Model", x="Residual")

#Q-Q plot of transformed residuals for m3
p2 <- ggplot(pred.mfinalb, aes(sample=resid))
p2<- p2 + geom_qq_line(color="blue",size=1) + geom_qq() +
  labs(title="Q-Q plot of Transformed Residuals-LME Final Model", x="Quantile", y="Residual")
```

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

```
#install.packages("gridExtra")
#library("gridExtra")
plots.Q3di<-grid.arrange(p1,p2, nrow=1)
```

```
## Warning: The dot-dot notation ('..density..') was deprecated in ggplot2 3.4.0.
## i Please use 'after_stat(density)' instead.
```

```
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```
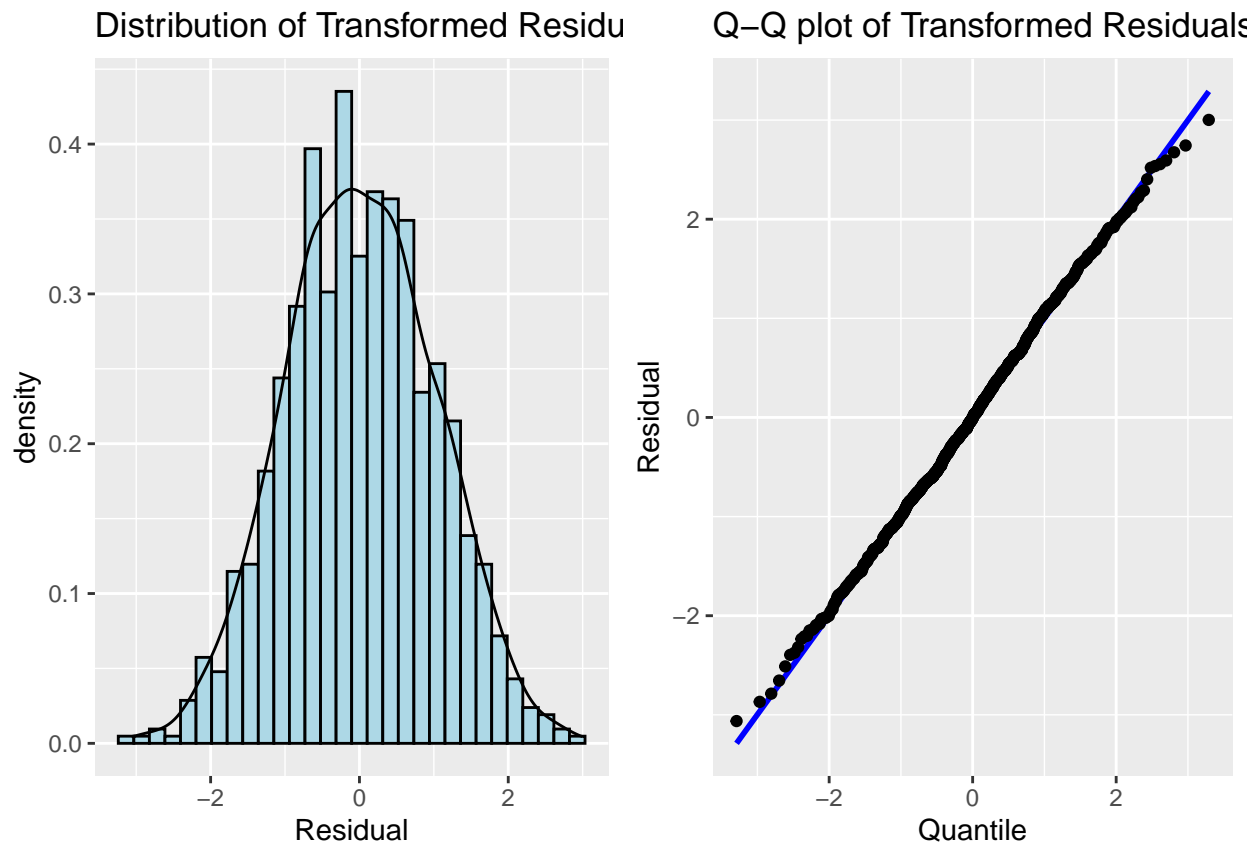


```
plots.Q3di
```

```
## TableGrob (1 x 2) "arrange": 2 grobs
##   z     cells    name              grob
## 1 1 (1-1,1-1) arrange gtable[layout]
## 2 2 (1-1,2-2) arrange gtable[layout]
```

The histogram of residuals looks pretty normal, and the Q-Q plot indicates no significant departures from the Normality assumption.

Transformed residuals against the marginal mean response and each covariate in the final model

```
p3 <- ggplot(pred.mfinalb, aes(x=pred, y=resid))
p3 <- p3 + geom_point() + geom_hline(yintercept=0,colour="red") +
  geom_smooth(method="loess", se=F) +
```

```r
  labs(x="Predicted Mean", y="Residual")

#Scatter plot of transformed residual against age with LOESS smoothed curve for final model
p4 <- ggplot(pred.mfinalb, aes(x=age, y=resid))
p4 <- p4 + geom_point() + geom_hline(yintercept=0,colour="red") +
  geom_smooth(method="loess", se=F) +
  labs(x="Age (yrs)", y="Residual")

#Scatter plot of transformed residual against Experience with LOESS smoothed curve for final model
p5 <- ggplot(pred.mfinalb, aes(x=experience, y=resid))
p5 <- p5 + geom_point() + geom_hline(yintercept=0,colour="red") +
  geom_smooth(method="loess", se=F) +
  labs(x="Experience (yrs)", y="Residual")

#Scatter plot of transformed residual against Hospital Size with LOESS smoothed curve for final model
p6 <- ggplot(pred.mfinalb, aes(x=hospsize, y=resid))
p6 <-p6 + geom_point() + geom_hline(yintercept=0,colour="red") +
  geom_smooth(method="loess", se=F) +
  labs(x="Hospital Size, 0=Small, 1=Medium, 2=Large", y="Residual")

#Scatter plot of transformed residual against Gender with LOESS smoothed curve for final model
p7 <- ggplot(pred.mfinalb, aes(x=gender, y=resid))
p7 <- p7 + geom_point() + geom_hline(yintercept=0,colour="red") +
  geom_smooth(method="loess", se=F) +
  labs(x="Gender, 0=Male, 1=Female", y="Residual")

#Scatter plot of transformed residual against Treatment with LOESS smoothed curve for final model
p8 <- ggplot(pred.mfinalb, aes(x=trt, y=resid))
p8 <- p8 + geom_point() + geom_hline(yintercept=0,colour="red") +
  geom_smooth(method="loess", se=F) +
  labs(x="Treatment", y="Residual")

#Scatter plot of transformed residual against Ward Type with LOESS smoothed curve for final model
p9 <- ggplot(pred.mfinalb, aes(x=type, y=resid))
p9 <- p9 + geom_point() + geom_hline(yintercept=0,colour="red") +
  geom_smooth(method="loess", se=F) +
  labs(x="Ward Type, 0=General Care, 1=Special Care", y="Residual")
plots.Q3dii<-grid.arrange(p3,p4,p5,p6,p7,p8,p9, nrow=3, top="Transformed Residuals by Covariates")
```

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

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : pseudoinverse used at -0.01
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : neighborhood radius 1.01
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : reciprocal condition number 1.5387e-30
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : There are other near singularities as well. 4.0401

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

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : pseudoinverse used at -0.005

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : neighborhood radius 1.005

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : reciprocal condition number 0

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : There are other near singularities as well. 1.01

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

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : pseudoinverse used at -0.005

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : neighborhood radius 1.005

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : reciprocal condition number 1.2049e-29

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : There are other near singularities as well. 1.01

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

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : pseudoinverse used at -0.005

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : neighborhood radius 1.005

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : reciprocal condition number 6.0711e-30

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : There are other near singularities as well. 1.01
```
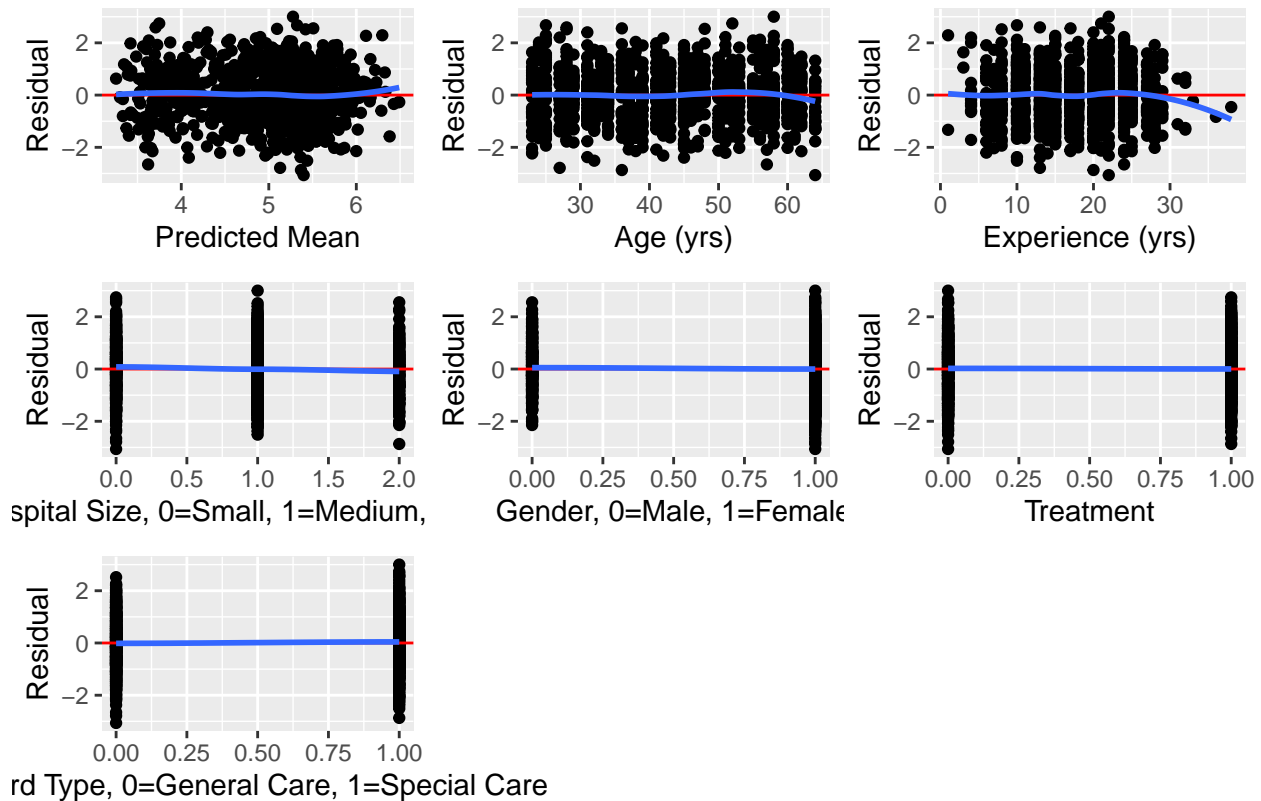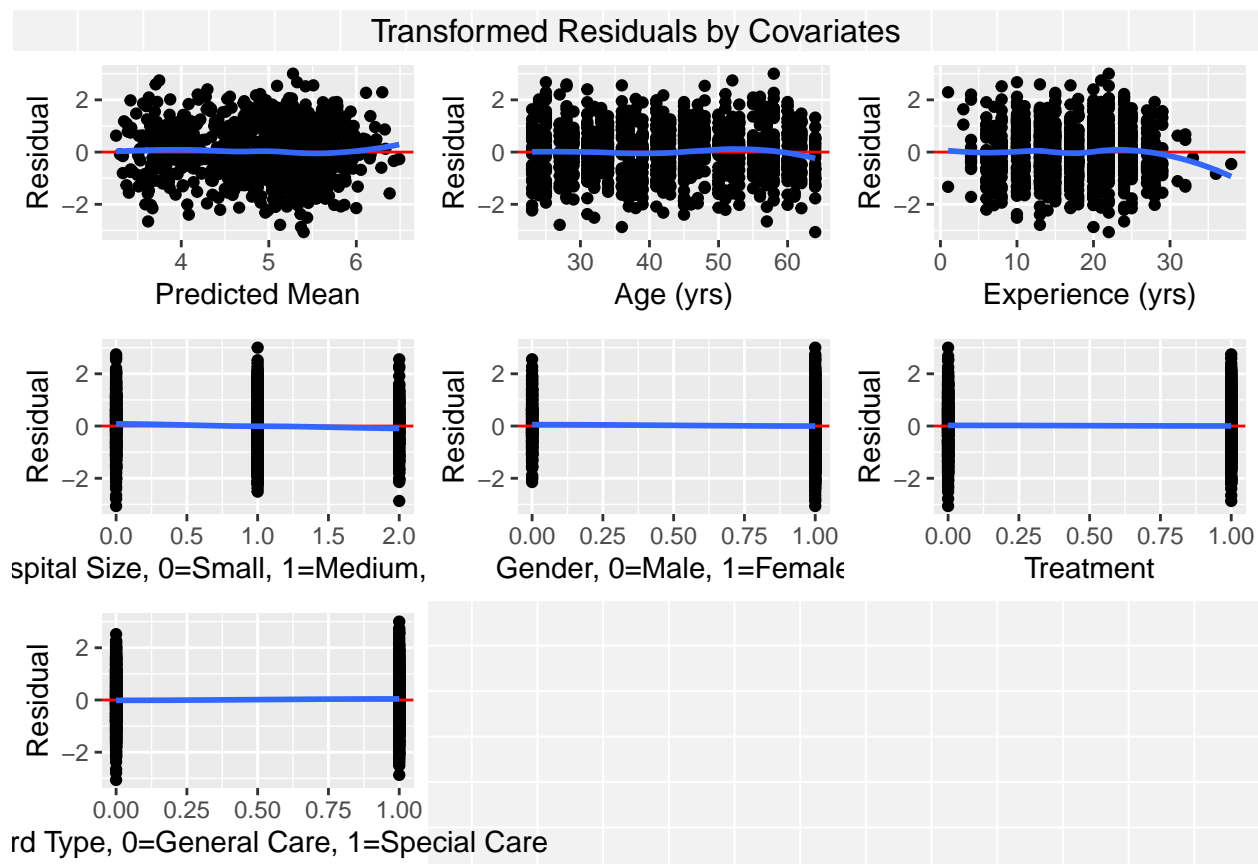
Transformed Residuals by Covariates


```
plots.Q3dii
```

```
## TableGrob (4 x 3) "arrange": 8 grobs
##   z     cells      name                grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (2-2,3-3) arrange      gtable[layout]
## 4 4 (3-3,1-1) arrange      gtable[layout]
## 5 5 (3-3,2-2) arrange      gtable[layout]
## 6 6 (3-3,3-3) arrange      gtable[layout]
## 7 7 (4-4,1-1) arrange      gtable[layout]
## 8 8 (1-1,1-3) arrange text[GRID.text.837]
```

```
plot(plots.Q3dii)
```

Transformed Residuals by Covariates

The distribution of transformed residuals relative to each covariate indicates that the model for the mean response fits the data well and that the functional form for each covariate seems to be appropriate, as there are no systematic trends identified visually: the residuals are approximately 0 on average and their variability is approximately half above and half below the average of 0.

Histograms and Q-Q plots of the random effects (BLUPS)

```
###Assessing normality of random effects for final model
mfinalb.re <- random.effects(m.finalb)
str(mfinalb.re)
```

```
## List of 2
##  $ hospital:'data.frame':    25 obs. of  2 variables:
##   ..$ (Intercept): num [1:25] -0.168 -0.192 0.212 -0.209 0.36 ...
##   ..$ Cexpcon    : num [1:25] 0.6758 0.1834 -0.0166 -0.6745 0.4038 ...
##   ..- attr(*, "effectNames")= chr [1:2] "(Intercept)" "Cexpcon"
##  $ ward    :'data.frame':    100 obs. of  1 variable:
##   ..$ (Intercept): num [1:100] 0.4078 -0.0936 -0.3578 -0.2234 0.2993 ...
##   ..- attr(*, "effectNames")= chr "(Intercept)"
##  - attr(*, "class")= chr [1:2] "ranef.lme" "list"
##  - attr(*, "label")= chr "Random effects"
##  - attr(*, "level")= int 2
##  - attr(*, "standardized")= logi FALSE
##  - attr(*, "grpNames")= chr [1:2] "hospital" "ward %in% hospital"
```

```
hosp.level.re<-as.data.frame(mfinalb.re$hospital)
head(hosp.level.re)
```

```
##   (Intercept)      Cexpcon
## 1  -0.1676849   0.67575112
## 2  -0.1921443   0.18336053
## 3   0.2122419  -0.01663303
## 4  -0.2088782  -0.67445845
## 5   0.3603665   0.40375277
## 6  -0.4767881  -0.51860237
```

```
length(hosp.level.re$Cexpcon)#25 hospitals
```

```
## [1] 25
```

```
ward.level.re<-as.data.frame(mfinalb.re$ward)
head(ward.level.re)
```

```
##       (Intercept)
## 1/1    0.4077696
## 1/2   -0.0936426
## 1/3   -0.3577565
## 1/4   -0.2233794
## 2/1    0.2992588
## 2/2   -0.2372915
```

```
length(ward.level.re$`(Intercept)`)#100 wards
```

```
## [1] 100
```

```
#Calculate the variance of the empirical BLUPs and compare them to the estimated variances for final mo
apply(hosp.level.re, 2, var)
```

```
## (Intercept)      Cexpcon
##   0.1336653    0.1197785
```

```
#(Intercept)      Cexpcon
#0.130260      0.110292
apply(ward.level.re, 2, var)
```

```
## (Intercept)
##   0.05881797
```

```
#(Intercept)
#0.06012455

#Histogram of transformed Random Effects for final model
p1 <- ggplot(hosp.level.re, aes(x=`(Intercept)`))
p1 <- p1 + geom_histogram(aes(y=..density..), color="black", fill="lightblue") +
```

```r
    geom_density() + labs(title="Random Int for Hospital", x="Hospital RE")

#Q-Q plot
p2 <- ggplot(hosp.level.re, aes(sample=`(Intercept)`))
p2 <- p2 + geom_qq_line(color="blue",size=1) + geom_qq() +
  labs(title="Random Int for Hospital", x="Quantile", y="Hospital RE")

#Histogram
p3 <- ggplot(hosp.level.re, aes(x=Cexpcon))
p3 <- p3 + geom_histogram(aes(y=..density..), color="black", fill="lightblue") +
  geom_density() + labs(title="Random Trt across Hospitals", x="Hospital*Treat RE")

#Q-Q plot
p4 <- ggplot(hosp.level.re, aes(sample=Cexpcon))
p4 <- p4 + geom_qq_line(color="blue",size=1) + geom_qq() +
  labs(title="Random Trt across Hospitals", x="Quantile", y="Hospital*Treat RE")

#Histogram
p5 <- ggplot(ward.level.re, aes(x=`(Intercept)`))
p5 <- p5 + geom_histogram(aes(y=..density..), color="black", fill="lightblue") +
  geom_density() + labs(title="Random Int for Ward", x="Ward RE")

#Q-Q plot
p6 <- ggplot(ward.level.re, aes(sample=`(Intercept)`))
p6 <- p6 + geom_qq_line(color="blue",size=1) + geom_qq() +
  labs(title="Q-Q plot of Random Int for Ward", x="Quantile", y="Ward RE")

plots.Q3diii<-grid.arrange(p1,p2,p3,p4,p5,p6, nrow=2, top="Distribution of Random Effects: Histogram an
```
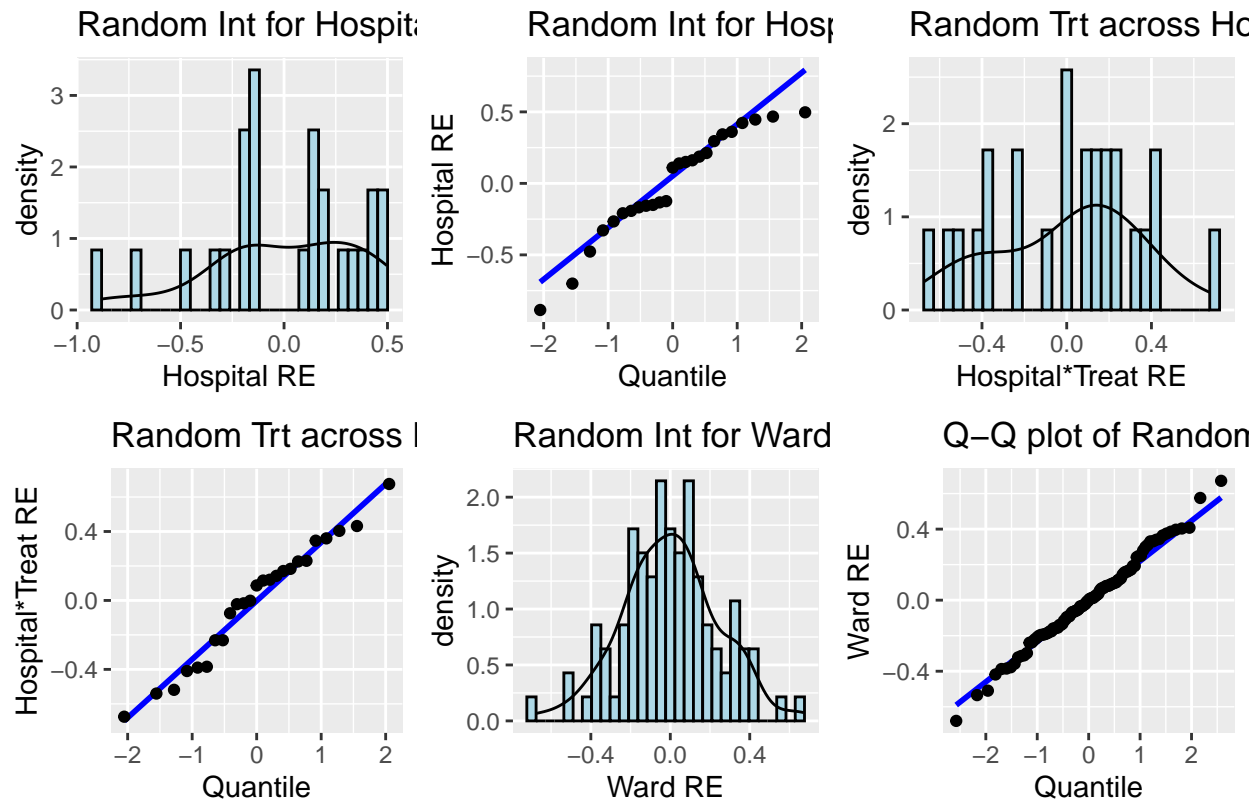
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Distribution of Random Effects: Histogram and Q–Q Plots
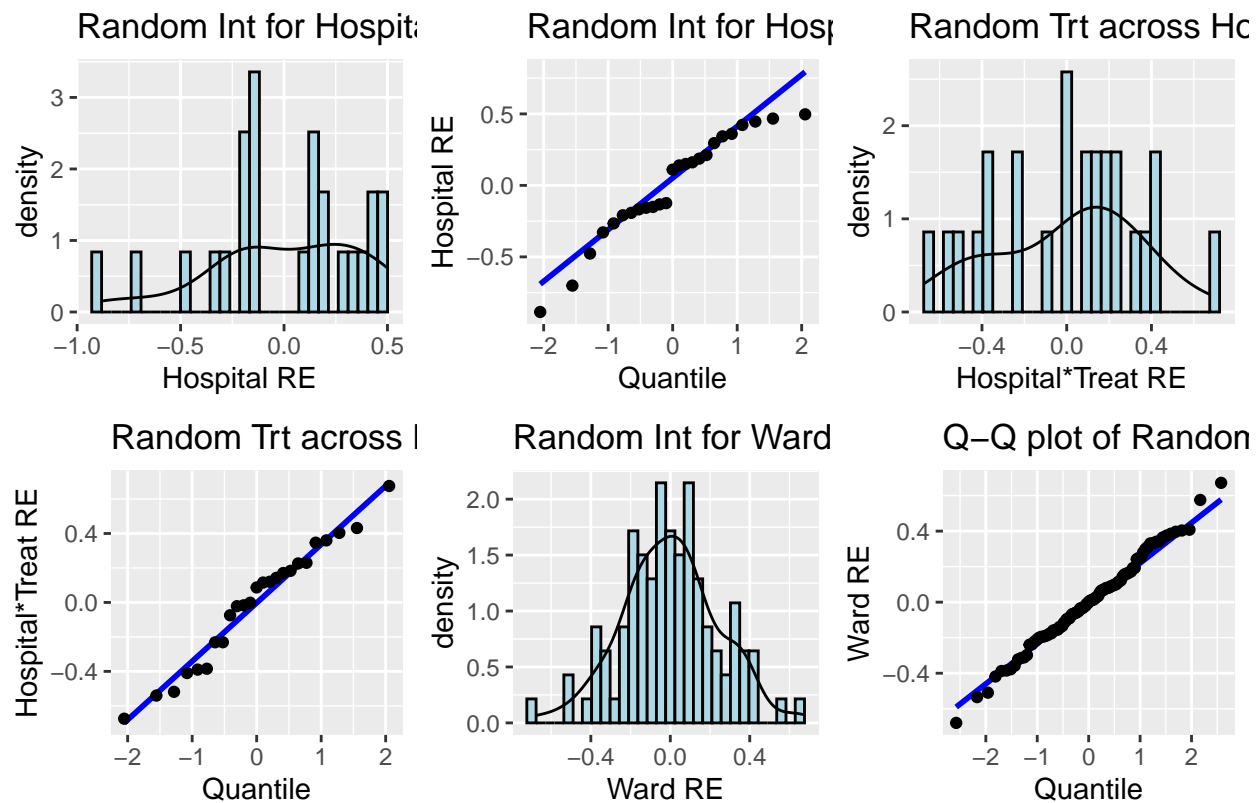
```
plots.Q3diii
```

```
## TableGrob (3 x 3) "arrange": 7 grobs
##   z     cells    name                grob
## 1 1 (2-2,1-1) arrange      gtable[layout]
## 2 2 (2-2,2-2) arrange      gtable[layout]
## 3 3 (2-2,3-3) arrange      gtable[layout]
## 4 4 (3-3,1-1) arrange      gtable[layout]
## 5 5 (3-3,2-2) arrange      gtable[layout]
## 6 6 (3-3,3-3) arrange      gtable[layout]
## 7 7 (1-1,1-3) arrange text[GRID.text.1096]
```

```
plot(plots.Q3diii)
```

Distribution of Random Effects: Histogram and Q–Q Plots

- Random Intercept for Hospital: there do not appear to be fatal departures from normality for the random intercepts for Hospital, and their empirical distribution simply reflects the relatively small number of hospitals in the study (only 25 hospitals).

- Random Treatment Effect across Hospitals: similarly to the distribution of RE for Hospital, this RE suffers from the small sample size of hospitals, and does not appear to have fatal departures from normality (Q-Q plot).

- Random Intercept for Ward: there do not appear to be fatal departures from normality for the RE for ward; because there were more wards than hospitals, the histogram and the Q-Q plot for this RE looks better for the other two Res, but it's only due to the increased sample size (number of wards is 100).