# Week 8 Workbook

Odelia

4/28/2021

libraries

```
library("tidyverse")
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.1     v dplyr   1.0.5
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library("patchwork")
library("lubridate")
```

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library("kableExtra")
```

```
##
## Attaching package: 'kableExtra'

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

```
library("gtsummary")
library("lubridate")
library("tidyLPA")
```

```
## You can use the function citation('tidyLPA') to create a citation for the use of {tidyLPA}.
## Mplus is not installed. Use only package = 'mclust' when calling estimate_profiles().
```

```r
library("ggplot2")
library("performance")
library("qqplotr")
```

```
##
## Attaching package: ’qqplotr’

## The following objects are masked from ’package:ggplot2’:
##
##     stat_qq_line, StatQqLine
```

read data

```r
nz_0 <- as.data.frame(readr::read_csv2(
  url(
    "https://raw.githubusercontent.com/go-bayes/psych-447/main/data/nzj.csv"
  )))
```

```
## i Using ’,’ as decimal and ’.’ as grouping mark. Use ‘read_delim()‘ for more control.

##
## -- Column specification --------------------------------------------------------
## cols(
##   .default = col_double(),
##   Male = col_character(),
##   BigDoms = col_character(),
##   GenCohort = col_character(),
##   Religious = col_character(),
##   Believe.God = col_character(),
##   Believe.Spirit = col_character(),
##   FeelHopeless = col_character(),
##   FeelDepressed = col_character(),
##   FeelRestless = col_character(),
##   EverythingIsEffort = col_character(),
##   FeelWorthless = col_character(),
##   FeelNervous = col_character()
## )
## i Use ‘spec()‘ for the full column specifications.
```

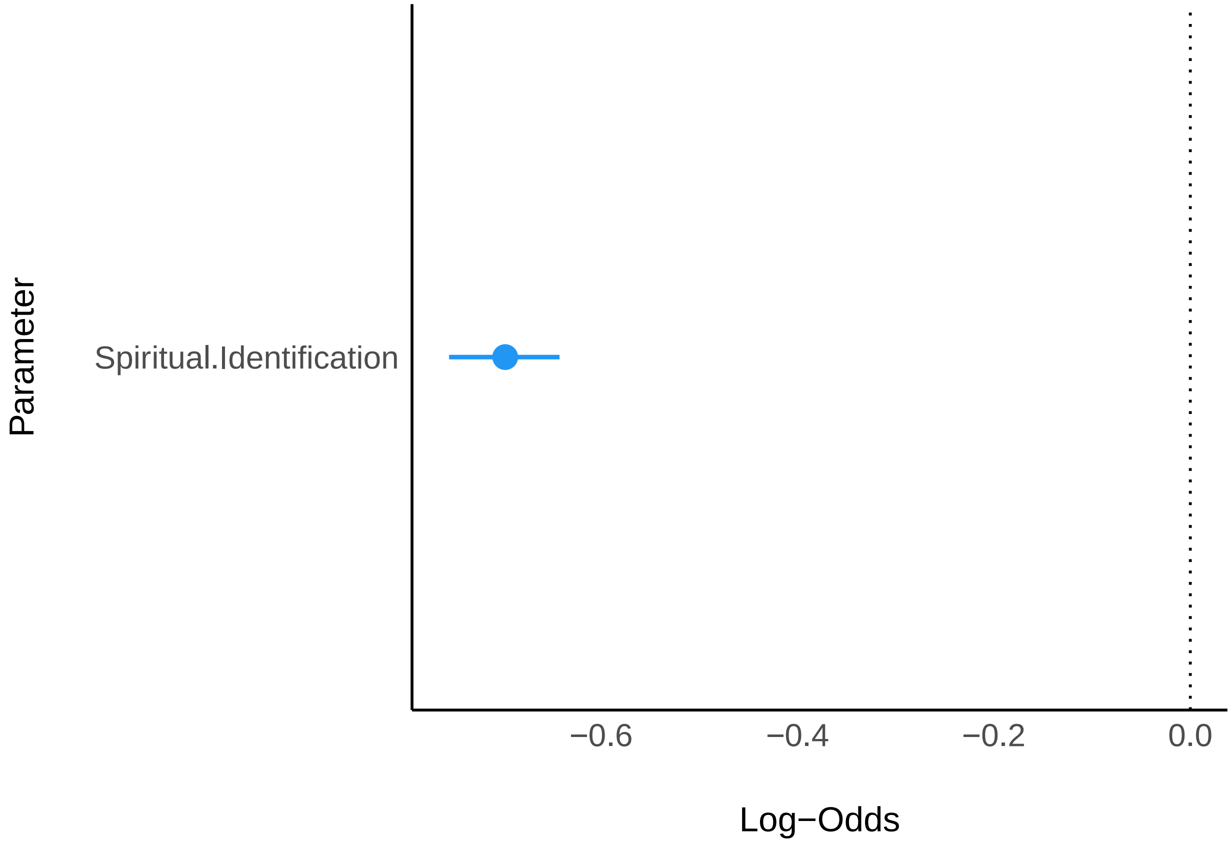
```r
f <-
  c(
    "None Of The Time",
    "A Little Of The Time",
    "Some Of The Time",
    "Most Of The Time",
    "All Of The Time"
  )
nz <- nz_0 %>%
  dplyr::mutate_if(is.character, factor) %>%
  select(
    -c(
      SWB.Kessler01,
      SWB.Kessler02,
      SWB.Kessler03,
      SWB.Kessler04,
      SWB.Kessler05,
      SWB.Kessler06
    )
  ) %>%
  dplyr::mutate(Wave = as.factor(Wave)) %>%
  mutate(FeelHopeless = forcats::fct_relevel(FeelHopeless, f)) %>%
  mutate(FeelDepressed = forcats::fct_relevel(FeelDepressed, f)) %>%
  mutate(FeelRestless = forcats::fct_relevel(FeelRestless, f)) %>%
  mutate(EverythingIsEffort = forcats::fct_relevel(EverythingIsEffort, f)) %>%
  mutate(FeelWorthless = forcats::fct_relevel(FeelWorthless, f)) %>%
  mutate(FeelNervous = forcats::fct_relevel(FeelNervous, f)) %>%
  dplyr::mutate(Wave = as.factor(Wave)) %>%
  dplyr::mutate(male_id = as.factor(Male)) %>%
  dplyr::mutate(date = make_date(year = 2009, month = 6, day = 30) + TSCORE)%>%
  dplyr::filter(Wave == 2018)
```

**Write brief report that predicts belief in spirit or a life-force (Believe.Spirit) from no more than five covariates. Explain your model and results.**

```
belief <- glm(Believe.Spirit ~ Spiritual.Identification, data = nz, family = "binomial")
parameters::model_parameters(belief)
```

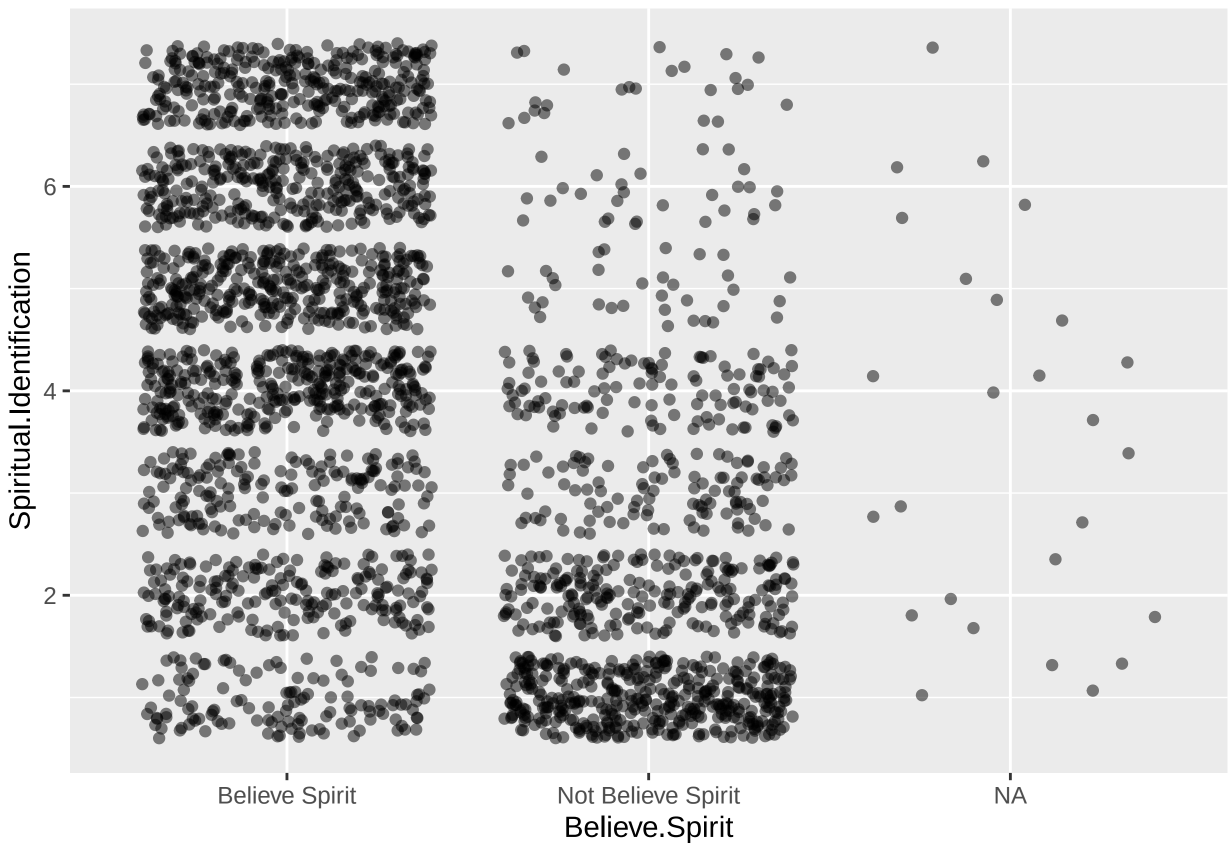
```
## Parameter                | Coefficient |   SE |         95% CI |      z |      p
## ------------------------------------------------------------------------------------
## (Intercept)              |        1.65 | 0.10 | [ 1.46,  1.84] |  16.98 | < .001
## Spiritual.Identification |       -0.70 | 0.03 | [-0.75, -0.64] | -24.32 | < .001
```

```
plot(parameters::model_parameters(belief))
```



```
ggplot(nz, (aes(Believe.Spirit, Spiritual.Identification))) + geom_jitter(alpha = .5)
```

```
## Warning: Removed 109 rows containing missing values (geom_point).
```



```
performance_accuracy(belief)
```

```
## # Accuracy of Model Predictions
##
## Accuracy: 81.52%
##       SE: 2.44%-points
##   Method: Area under Curve
```

```
equatiomatic::extract_eq(belief, use_coefs = TRUE)
```

$$\log\left[\frac{P(\operatorname{Believe.Spirit} \widehat{=} \operatorname{Not\ Believe\ Spirit})}{1 - P(\operatorname{Believe.Spirit} \widehat{=} \operatorname{Not\ Believe\ Spirit})}\right] = 1.65 - 0.7(\operatorname{Spiritual.Identification})$$

A binary logistic regression was fitted to predict Believe.Spirit from Spiritual.Identification. The effect of Spiritual.Identification is significantly negative (beta = -0.70, 95% CI [-0.75, -0.64], p < .001; Std. beta = -1.42, 95% CI [-1.54, -1.31]). The chance of a type 1 error is 0.1%. As further clarified with the plots, plot 1 tells us that the the CI is narrow and thus a high accuracy of Spiritual.Identification being a predictor. Plot 2 tells us that Believe.Spirit can be predicted by increasing identification of spirituality. The plot also tells us that whilst the opposite is true, the rate of every increasing unit of Spiritual.Identification predicting for believing in spirit is visibly higher than the rate of decreasing unit of Spiritual.Identification predicting for not believing in spirit. The significant effect of Spiritual.Identification as a predictor for Believe.Spirit is also evident with the high accuracy of the model at over 80%.

**Write report that predicts charitable donations (CharityDonate) from no more than five covariates. Explain your model and results.**

```
library(MASS)
```

```
##
## Attaching package: ’MASS’

## The following object is masked from ’package:gtsummary’:
##
##     select

## The following object is masked from ’package:patchwork’:
##
##     area

## The following object is masked from ’package:dplyr’:
##
##     select
```

```
pois <- glm(CharityDonate ~ Household.INC, data = nz, family = "poisson")
parameters::model_parameters(pois)
```

```
## Parameter     | Coefficient |       SE |       95% CI |        z |      p
## -------------------------------------------------------------------------
## (Intercept)   |        6.70 | 6.18e-04 | [6.70, 6.71] | 10851.41 | < .001
## Household.INC |    2.16e-06 | 1.00e-09 | [0.00, 0.00] |  2161.20 | < .001
```

```
pois2 <- glm(CharityDonate ~ Household.INC + Standard.Living, data = nz, family = "poisson")
parameters::model_parameters(pois2)
```

```
## Parameter       | Coefficient |       SE |       95% CI |       z |      p
## --------------------------------------------------------------------------
## (Intercept)     |        4.31 | 3.55e-03 | [4.31, 4.32] | 1214.96 | < .001
## Household.INC   |    1.92e-06 | 1.07e-09 | [0.00, 0.00] | 1796.31 | < .001
## Standard.Living |        0.29 | 4.11e-04 | [0.29, 0.30] |  716.94 | < .001
```

```
pois3 <- glm(CharityDonate ~ Household.INC + Standard.Living + Religious, data = nz, family = "poisson"
parameters::model_parameters(pois3)
```

```
## Parameter             | Coefficient |       SE |       95% CI |       z |      p
## --------------------------------------------------------------------------------
## (Intercept)           |        3.93 | 3.60e-03 | [3.92, 3.93] | 1091.71 | < .001
## Household.INC         |    2.07e-06 | 1.09e-09 | [0.00, 0.00] | 1898.42 | < .001
## Standard.Living       |        0.30 | 4.09e-04 | [0.29, 0.30] |  721.48 | < .001
## Religious [Religious] |        0.82 | 1.15e-03 | [0.82, 0.82] |  712.60 | < .001
```

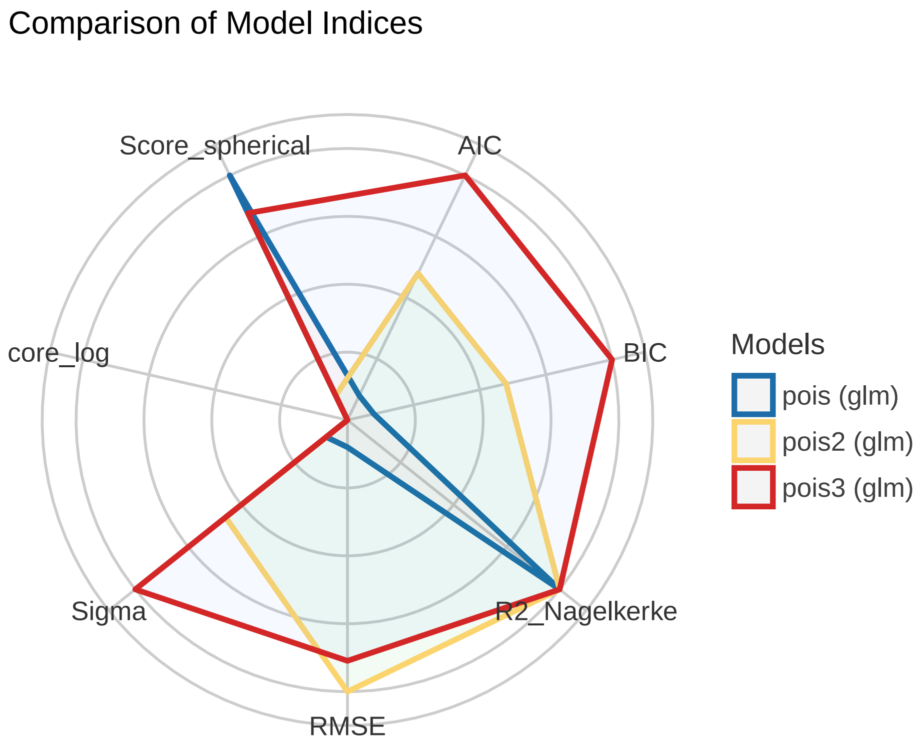
```
perf <- performance::compare_performance(pois,pois2,pois3)
```

```
## Warning: When comparing models, please note that probably not all models were
## fit from same data.
```

```
plot(perf)
```

```
## Warning in change_scale.numeric(X[[i]], ...): A ‘range‘ must be provided for
## data with only one observation.

## Warning in change_scale.numeric(X[[i]], ...): A ‘range‘ must be provided for
## data with only one observation.
```



```
check_overdispersion(pois3)
```

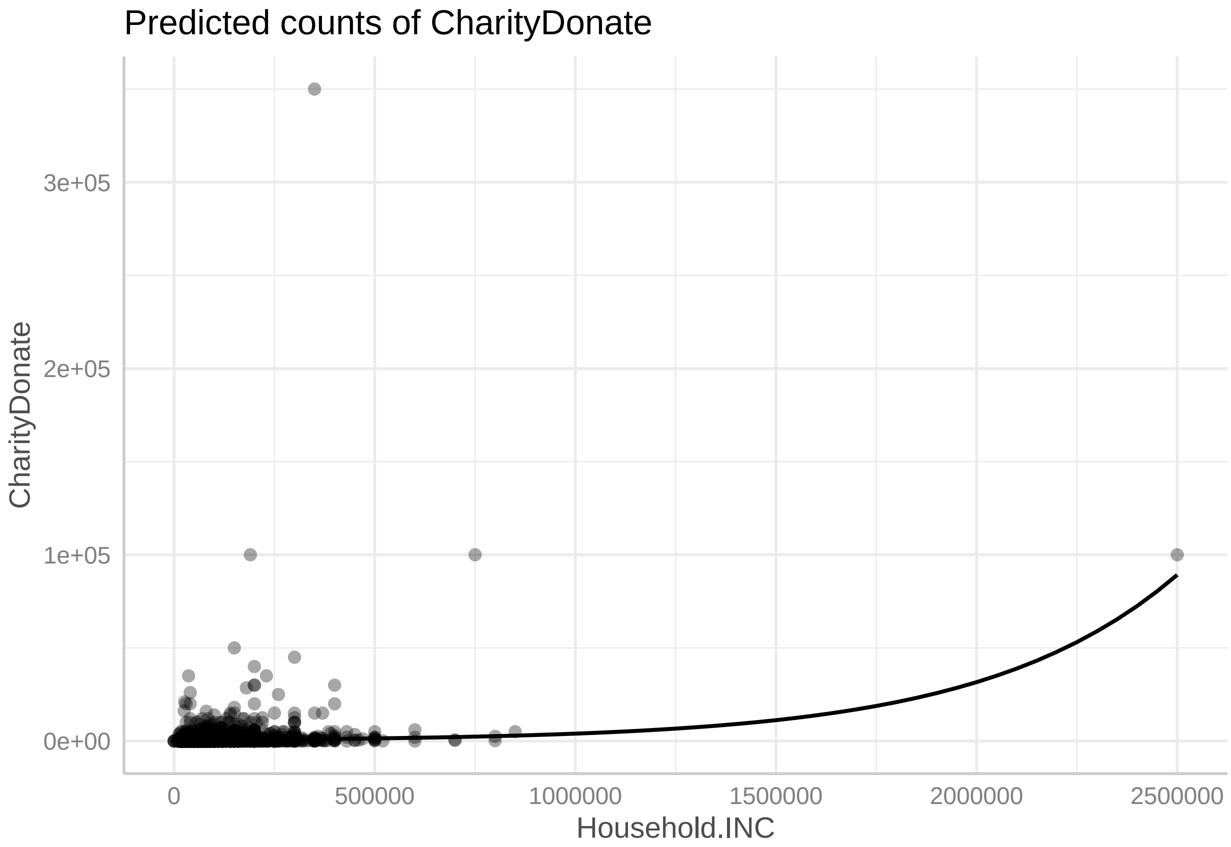
```
## # Overdispersion test
##
##        dispersion ratio =     36377.437
##   Pearson's Chi-Squared = 103421053.909
##                 p-value =       < 0.001

## Overdispersion detected.
```

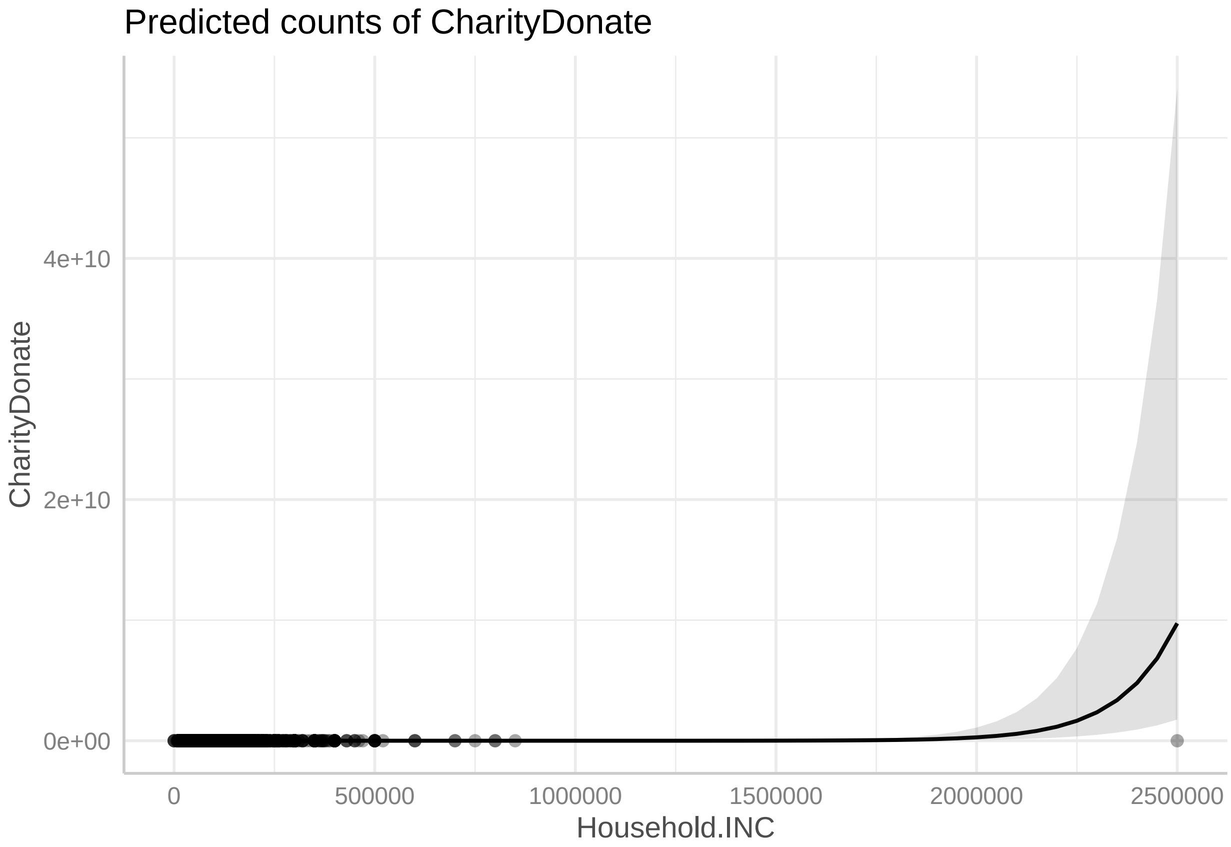
```
nb <- glm.nb(CharityDonate ~ Household.INC + Standard.Living + Religious, data = nz,)
parameters::model_parameters(nb)
```

```
## Parameter                | Coefficient |       SE |       95% CI |      z |      p
## ---------------------------------------------------------------------------
## (Intercept)           |        3.77 |     0.15 | [3.50, 4.04] | 25.89 | < .001
## Household.INC         |    7.09e-06 | 3.67e-07 | [0.00, 0.00] | 19.33 | < .001
## Standard.Living       |        0.20 |     0.02 | [0.16, 0.23] | 10.66 | < .001
## Religious [Religious] |        1.28 |     0.07 | [1.14, 1.43] | 17.90 | < .001
```

```
plot(ggeffects::ggpredict(pois3, terms = "Household.INC"), add.data = TRUE)
```



```
plot(ggeffects::ggpredict(nb, terms = "Household.INC"), add.data = TRUE)
```

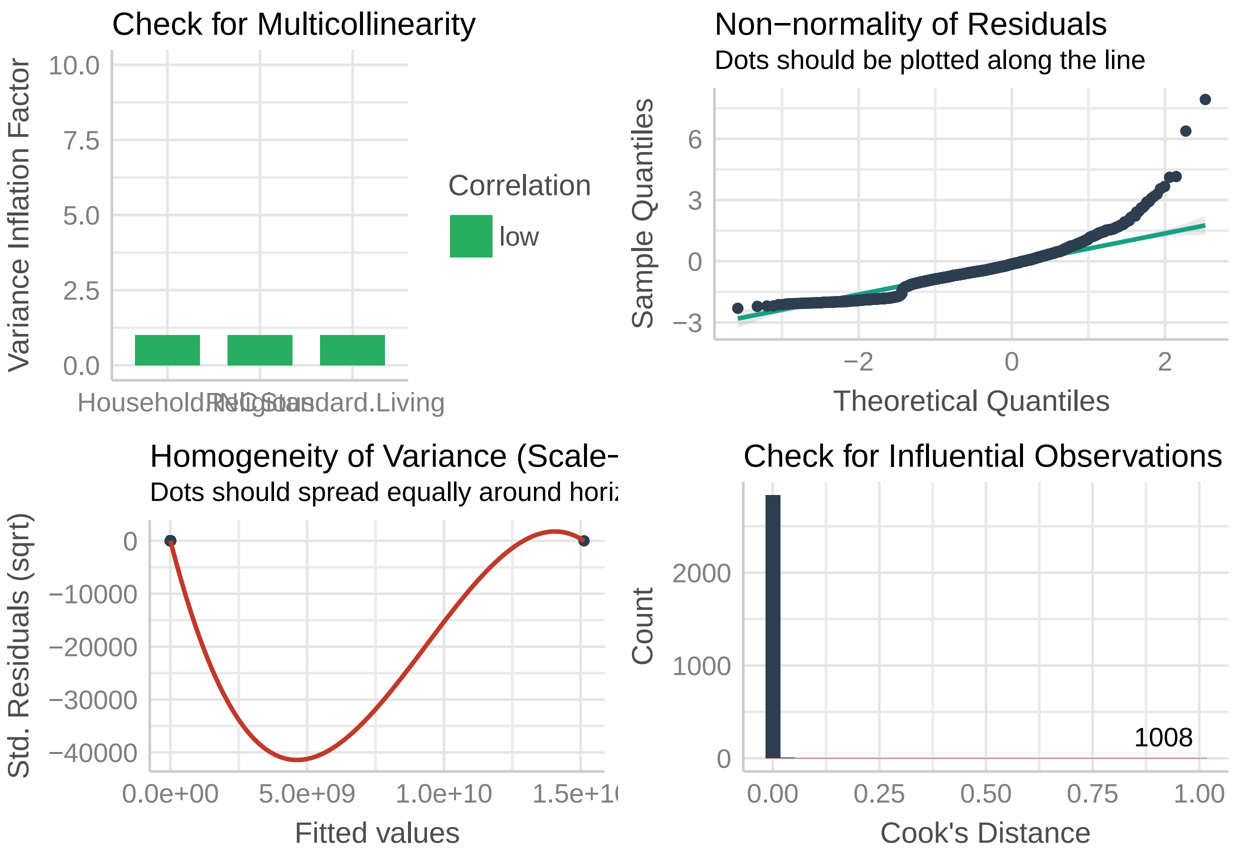


```
performance::check_model(nb)
```

```
## ‘geom_smooth()‘ using formula ’y ~ x’
```

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

```
## Warning: Removed 2846 rows containing missing values (geom_text_repel).
```



```
equatiomatic::extract_eq(nb, use_coefs = TRUE)
```

$$\log(E(\widehat{\text{CharityDonate}})) = 3.77 + 0(\text{Household. INC}) + 0.2(\text{Standard. Living}) + 1.28(\text{Religious})$$

A poisson regression model was first fitted to predict CharityDonate from Household.INC, Standard.Living and Religious. 3 models were fitted with increasing covariate respectively; based on the tables and "perf" plot, pois3 (all 3 covariates) is observed to be the best model in predicting for CharityDonate. However, overdispersion was detected and the test was done. Dispersion ratio was above 1 and p-value was above .05. Hence, a negative binomial model was fitted instead. The effect of Household.INC is significantly positive (beta = 7.09e-06, 95% CI [6.29e-06, 7.91e-06], p < .001; Std. beta = 0.68, 95% CI [0.60, 0.75]). The effect of Standard.Living is significantly positive (beta = 0.20, 95% CI [0.16, 0.23], p < .001; Std. beta = 0.37, 95% CI [0.31, 0.44]). The effect of Religious [Religious] is significantly positive (beta = 1.28, 95% CI [1.14, 1.43], p < .001; Std. beta = 1.28, 95% CI [1.14, 1.43]). Although all 3 covariates are statistically significant in predicting for CharityDonate, Religious correlates the most out of the 3 covariates, followed by Standard.Living and Household.INC. The plots (pois3 against nb) tells us that the negative binomial regression model is the appropriate model to be used instead of a poisson regression or linear model. From the performance plots, we can tell that the covariates do not interact with one another which means that the assumption of multicollinearity is not violated. However, the assumptions for normality and homogeneity of variance appear to be violated. This tells us that the data is likely to be falsely skewed probably due to the outliers and that these results are likely not adequately representative of the population. Therefore, this model is informative but not as good as it typically needs to be as there is a higher likelihood of type I error.