Workbook_4

Paul Edwards

17/03/2021

Preliminary data wrangling

```
#Read data
nz_0 <- readr::read_csv2(url("https://raw.githubusercontent.com/go-bayes/psych-447/
main/data/nz/nz.csv"))

#Re-level Kessler 6 variables
f <- c("None Of The Time","A Little Of The Time","Some Of The Time",  "Most Of The
Time", "All Of The Time")

#Transform data
nz_1 <- 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(date = make_date(year = 2009, month = 6, day = 30) + TSCORE)
```

Question 1

```
#Check all variables containing "hours" string
nz_1 %>%
  select(contains("hours")) %>%
    glimpse()
```

```
## Rows: 4,126
## Columns: 6
## $ HLTH.SleepHours <dbl> 6.0, 6.0, 6.0, 4.0, 7.0, 7.0, 6.0, 4.0, 7.0, 8.0, 7.0,~
## $ Hours.Internet  <dbl> 10, 5, 14, 15, 2, 2, 4, 4, 10, 14, 2, 2, 4, 1, 8, 12, ~
## $ Hours.Exercise  <dbl> 14.0, 24.0, 7.0, 10.0, 5.0, 6.0, 1.0, 0.0, 1.0, 7.0, 0~
## $ Hours.Work      <dbl> 14, 0, 35, 65, 60, 50, 41, 42, 0, 0, 45, 40, 40, 42, 3~
## $ Hours.News      <dbl> 4.0, 4.0, 5.0, 4.0, 1.0, 2.0, 0.0, 0.0, 8.0, 7.0, 5.0,~
## $ HoursCharity    <dbl> 2.0, 0.0, 0.0, 2.0, 0.0, 0.0, 0.0, 4.0, 0.0, 0.0, 5.0,~
```

```
#Set above variables as integers
nz_2 <- nz_1 %>%
  dplyr::mutate(HLTH.SleepHours = as.integer(HLTH.SleepHours)) %>%
  dplyr::mutate(Hours.Internet = as.integer(Hours.Internet)) %>%
  dplyr::mutate(Hours.Exercise = as.integer(Hours.Exercise)) %>%
  dplyr::mutate(Hours.Work = as.integer(Hours.Work)) %>%
  dplyr::mutate(Hours.News = as.integer(Hours.News)) %>%
  dplyr::mutate(HoursCharity = as.integer(HoursCharity))
```

Question 2

```
#Standardise and centre Pol.Orient; centre Age within decade brackets
nz_3 <- nz_2 %>%
  dplyr::mutate(Pol.Orient.Std = scale(Pol.Orient, scale = TRUE, center  = TRUE)) %
>%
  dplyr::mutate(Pol.Orient.Cntr = scale(Pol.Orient, scale = FALSE, center  = TRUE))
%>%
  dplyr::mutate(Age.Brackets = cut(
    Age,
    breaks = c(16, 26, 36, 46, 56, 66, 76, Inf),
    labels = c("16-25", "26-35", "36-45", "46-55", "56-65", "66-75", "76+"),
    right = TRUE)) %>%
  group_by(Age.Brackets) %>%
  dplyr::mutate(Age.Cntr = scale(Age, scale = FALSE, center  = TRUE)) %>%
  ungroup()

#Display head values of original and newly created variables
nz_3 %>%
  select(Pol.Orient, Pol.Orient.Std, Pol.Orient.Cntr, Age, Age.Brackets, Age.Cntr)
%>%
  head
```

```
## # A tibble: 6 x 6
##   Pol.Orient Pol.Orient.Std[,~ Pol.Orient.Cntr[~   Age Age.Brackets Age.Cntr[,1]
##        <dbl>             <dbl>             <dbl> <dbl> <fct>               <dbl>
## 1          3            -0.420            -0.582    47 46-55              -4.95
## 2          3            -0.420            -0.582    46 36-45               3.96
## 3          5             1.02              1.42     47 46-55              -4.95
## 4          3            -0.420            -0.582    46 36-45               3.96
## 5          4             0.301             0.418    53 46-55               1.05
## 6          4             0.301             0.418    52 46-55               0.0527
```

```
#Select Hours.Exercise and filter by 2019 survey wave
nz_3 %>%
  filter(Wave=="2019") %>%
  select(Hours.Exercise) %>%
  head
```

```
## # A tibble: 6 x 1
##   Hours.Exercise
##            <int>
## 1             14
## 2              7
## 3              5
## 4              1
## 5              1
## 6              0
```

Question 3

```
#Slice dataframe by maximum daily response count for each survey wave
nz_3 %>%
  group_by(Wave) %>%
  count(Day = floor_date(date, "day")) %>%
  slice(which.max(n))
```

```
## # A tibble: 2 x 3
## # Groups:   Wave [2]
##   Wave  Day            n
##   <fct> <date>     <int>
## 1 2018  2018-06-21   112
## 2 2019  2019-12-03    54
```

Question 4

```
#Find interval between highest and second highest daily responses
nz_3 %>%
  count(Day = floor_date(date, "day")) %>%
  arrange(desc(n)) %>%
  dplyr::mutate(int = lubridate::interval(Day[1], Day[2])) %>%
  dplyr::mutate(int = lubridate::int_standardize(int)) %>%
  dplyr::mutate(tl = lubridate::time_length(int, "day")) %>%
  dplyr::mutate(diff = n[1]-n[2]) %>%
  slice(1:2)
```

```
## # A tibble: 2 x 5
##   Day            n int                                tl  diff
##   <date>     <int> <Interval>                        <dbl> <int>
## 1 2018-06-21   112 2018-06-21 UTC--2018-06-22 UTC     1    19
## 2 2018-06-22    93 2018-06-21 UTC--2018-06-22 UTC     1    19
```

Question 5

```
#Calculate age between two date/times in months.
DOB = make_datetime(year = 1995, month = 12, day = 25, hour = 5, min = 2, sec = 22)
As_at_date = make_datetime(year = 2021, month = 3, day = 20, hour = 13, min = 22, s
ec = 4)
age1 <- time_length(int_standardize(interval(DOB, As_at_date)), "month")
age1
```

```
## [1] 302.8338
```

Question 6

```
#Separate Religion.Church into 3 factors and re-level variable
nz_4 <- nz_3 %>%
  dplyr::mutate(Church.Freq = cut(
    Religion.Church,
    breaks = c(-Inf, 1, 4, Inf),
    labels = c("0", "1-3", "4+"),
    right = TRUE)) %>%
  dplyr::mutate(Church.Freq = forcats::fct_relevel(Church.Freq, c("0", "1-3", "
4+")))
```

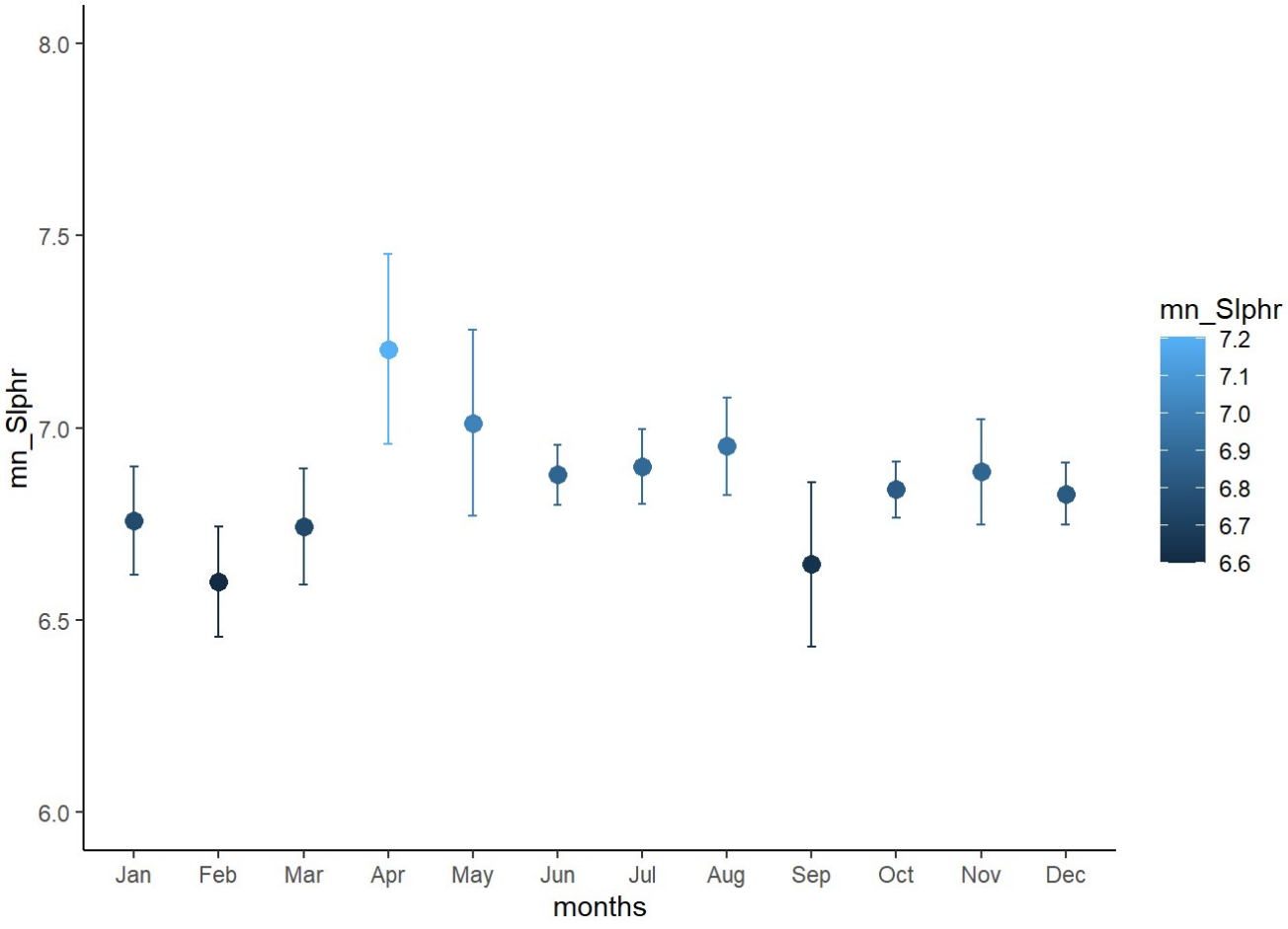
Question 7

```
#Factorise dates by month and label
Mnth <- factor(month(nz_4$date),
               labels=c("Jan",
                        "Feb",
                        "Mar",
                        "Apr",
                        "May",
                        "Jun",
                        "Jul",
                        "Aug",
                        "Sep",
                        "Oct",
                        "Nov",
                        "Dec"))
#Create table of average HLTH.SleepHours by month
table1::table1(~ HLTH.SleepHours | Mnth, data = nz_4, overall = FALSE)
```

	Jan (N=265)	Feb (N=174)	Mar (N=220)	Apr (N=81)	May (N=81)
HLTH.SleepHours					
Mean (SD)	6.76 (1.16)	6.60 (0.961)	6.74 (1.13)	7.21 (1.12)	7.01 (1.0
Median [Min, Max]	7.00 [3.00, 12.0]	7.00 [4.00, 8.00]	7.00 [2.00, 10.0]	7.00 [4.00, 9.00]	7.00 [5.00,

	Jan (N=265)	Feb (N=174)	Mar (N=220)	Apr (N=81)	May (N=81)
Missing	12 (4.5%)	9 (5.2%)	9 (4.1%)	3 (3.7%)	4 (4.9%

```
#Plot average HLTH.SleepHours by month with 95% confidence intervals
nz_4 %>%
  select(Id, date, HLTH.SleepHours) %>%
  mutate(months = month(date, label = TRUE)) %>%
  group_by(months) %>%
  summarise(
    mn_Slphr =  mean(HLTH.SleepHours, na.rm = TRUE),
    sd_Slphr =  sd(HLTH.SleepHours, na.rm = TRUE),
    n_Slphr = n()
  ) %>%
  mutate(
    se_Slphr = sd_Slphr / sqrt(n_Slphr),
    lw_ci = mn_Slphr - qt(1 - (0.05 / 2), n_Slphr - 1) * se_Slphr,
    up_ci = mn_Slphr + qt(1 - (0.05 / 2), n_Slphr - 1) * se_Slphr
  ) %>%
  ggplot(., aes(x = months, y = mn_Slphr, colour = mn_Slphr)) +
  geom_errorbar(aes(ymin = lw_ci, ymax = up_ci), width = .1) +
  geom_point(size = 3)  +
  scale_y_continuous(limits = c(6,8)) +
  theme_classic() + scale_fill_viridis_d()
```



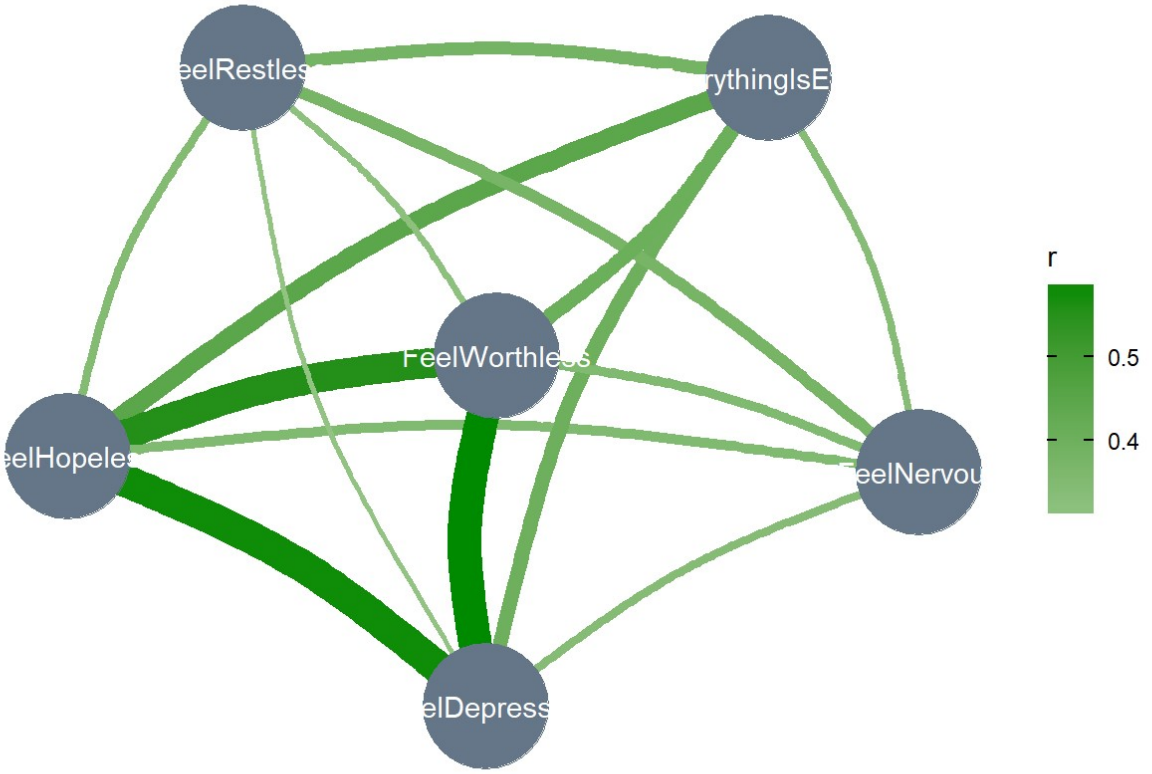
Looking more closely at the sample sizes for each month, there are fewer participants responding in those months which have wider confidence intervals. Smaller sample sizes increase the sampling error and accordingly the precision of our population estimates.

Question 8

```
#Plot correlations between Kessler 6 items, controlling for ID repeated measures
corr1 <- nz_4 %>%
  select(
    FeelHopeless,
    FeelDepressed,
    FeelRestless,
    EverythingIsEffort,
    FeelWorthless,
    FeelNervous,
    Id
  ) %>%
  mutate_all(., as.integer) %>%
  mutate(Id = as.factor(Id)) %>%
  correlation(partial = FALSE, multilevel = TRUE)
summary(corr1)
```

```
## Parameter          | FeelNervous | FeelWorthless | EverythingIsEffort | FeelRest
less | FeelDepressed
##
-------------------------------------------------------------------------------
-----------------
## FeelHopeless       |     0.35*** |       0.56*** |            0.45*** |      0.3
4*** |       0.58***
## FeelDepressed      |     0.34*** |       0.59*** |            0.40*** |      0.3
0*** |
## FeelRestless       |     0.38*** |       0.32*** |            0.38*** |
|
## EverythingIsEffort |     0.33*** |       0.40*** |                    |
|
## FeelWorthless      |     0.35*** |               |                    |
|
```

```
plot(corr1)
```



The correlation plot above displays the strength of correlations between the Kessler 6 items. The strongest correlations are between hopelessness, worthlessness, depression, and to a lesser extent, effortfulness. Some research has identified that a two-factor structure of depression and anxiety underlying the Kessler 6 is a better fitting model than a unidimensional structure (Bessaha, 2017; O'Connor & Parslow, 2010).

Question 9

See attached `Papaja` file.

Question 10

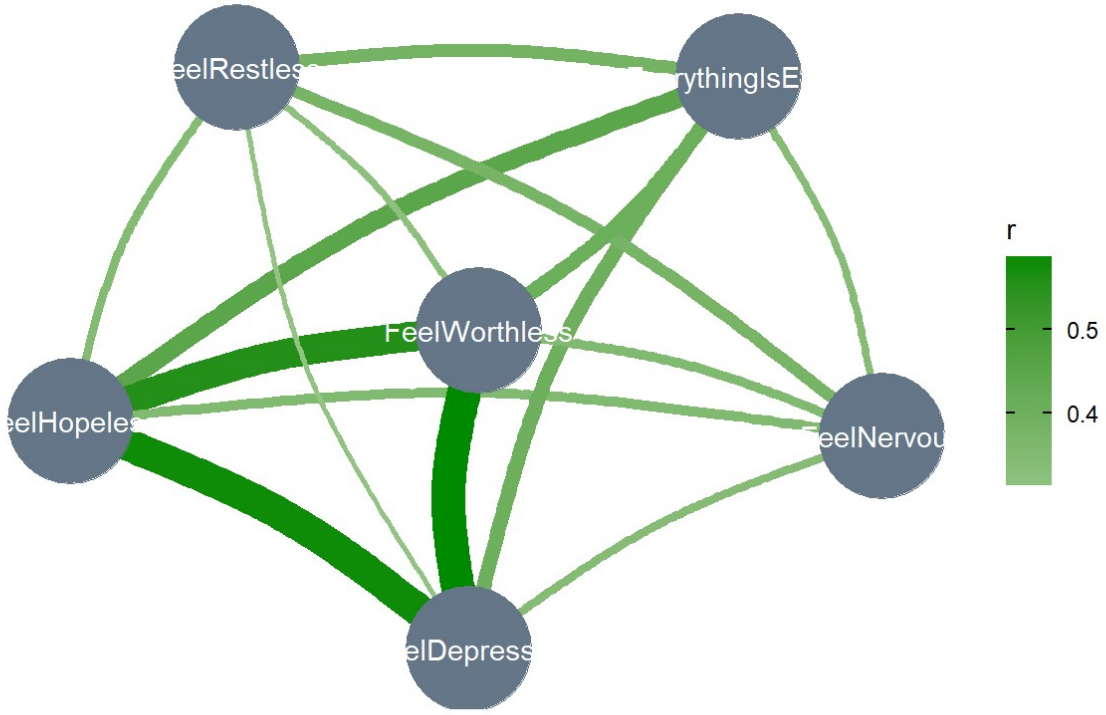
```
#Plot correlations between Kessler 6 items, no controlling for ID repeated measures
corr2 <- nz_4 %>%
  select(
    FeelHopeless,
    FeelDepressed,
    FeelRestless,
    EverythingIsEffort,
    FeelWorthless,
    FeelNervous,
  ) %>%
  mutate_all(., as.integer) %>%
  correlation(partial = FALSE, multilevel = FALSE)
summary(corr2)
```

```
## # Correlation Matrix (pearson-method)
##
## Parameter          | FeelNervous | FeelWorthless | EverythingIsEffort | FeelRest
less | FeelDepressed
##
-----------------------------------------------------------------------------
-----------------
## FeelHopeless       |     0.45*** |       0.65*** |            0.52*** |      0.4
3*** |       0.66***
## FeelDepressed      |     0.41*** |       0.67*** |            0.49*** |      0.3
8*** |
## FeelRestless       |     0.47*** |       0.39*** |            0.46*** |
|
## EverythingIsEffort |     0.43*** |       0.47*** |                    |
|
## FeelWorthless      |     0.43*** |               |                    |
|
##
## p-value adjustment method: Holm (1979)
```

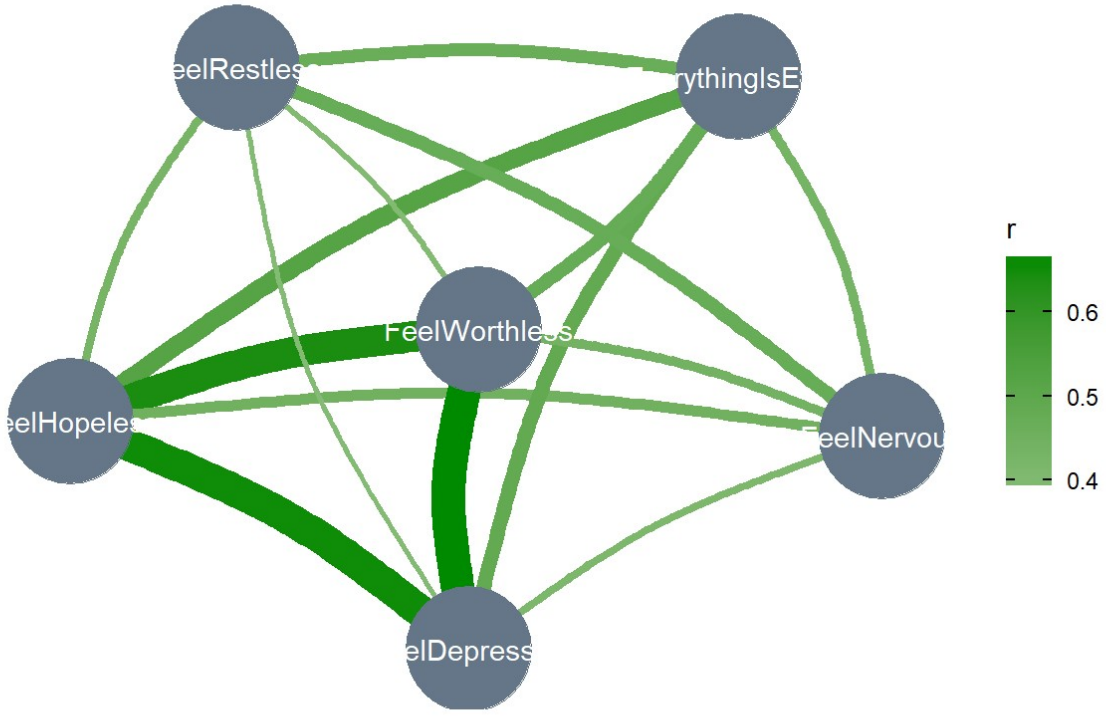
```
#Use patchwork function to merge multilevel and single-level plots
plot(corr1) / plot(corr2) +
  plot_annotation(title = "Plot of multilevel (a) and single-level (b) correlatio
n", tag_levels = 'a')
```

Plot of multilevel (a) and single-level (b) correlation

a



b



References

Bessaha, M. L. (2017). Factor structure of the Kessler Psychological Distress scale (K6) among emerging adults. *Research on Social Work Practice*, *27*(5), 616–624. https://doi.org/10.1177/1049731515594425 (https://doi.org/10.1177/1049731515594425)

O'Connor, D. W., & Parslow, R. A. (2010). Mental health scales and psychiatric diagnoses: Responses to GHQ-12, K-10 and CIDI across the lifespan. *Journal of Affective Disorders*, *121*(3), 263–267. https://doi.org/10.1016/j.jad.2009.06.038 (https://doi.org/10.1016/j.jad.2009.06.038)