'The effects of low-pay and unemployment on psychological well-being; a logistic regression approach '

By Yasmine Calvert

Side note- RMarkdown did not run, despite downloading and installing multiple packages we ran into the same issue again. Also %>% cannot be read by RMarkdown. We have inserted screenshots of our codes and tables, but at the end of this report there will be the codes that you will be able to copy into R.

```
Line 12 Error in contrib.url(repos, "source") :
          trying to use CRAN without setting a mirror
        Calls: <Anonymous> ... withVisible -> eval -> eval -> install.packages -> contrib.url
        Execution halted
```

```
1  ---
2  title: 'The effects of low-pay and unemployment on psychological well-being; a logistic
3    regression approach '
4  author: "Yasmine Calvert, Mina Kamawal"
5  output:
6    word_document: default
7    pdf_document: default
8    html_document: default
9  ---
10
11 ```{r setup, include=FALSE}
12 knitr::opts_chunk$set(echo = TRUE)
13 library("knitr")
14 require("knitr")
15 install.packages("magrittr")
16 library(magrittr)
17 library(dplyr)
18 install.packages('plyr', repos = "http://cran.us.r-project.org")
19 options(repos = list(CRAN="http://cran.rstudio.com/"))
20 install.packages("readxl")
21 library(readxl)
22
23 opts_knit$set(root.dir = "/Users/jasmine/Desktop/Rstudios")
24 getwd()
25 thedataset<- read.csv("bindresp (2).csv", stringsAsFactors=FALSE,
26                       na.strings = ".")
27 ```
```

Introduction

The research paper selected for this project is "The effects of low-pay and unemployment on psychological well-being. The group selected this paper because it appealed to both of us, and the project's goal caught our interest. The project's tables appeared manageable to replicate, and I believed we could replicate the codes because the hypothesis is that there is a correlation between unemployment and psychological well-being. Despite the fact that the paper was issued in 1988, unemployment and psychological well-being are still major issues, especially now during the Covid pandemic. This also prompted me to choose this paper in order to gain a better understanding of the subject and to investigate if there is a substantial correlation between the two variables. The approach taken in this report is to utilise logistic regression to determine the relationship between the variables and whether or not unemployment affects one's psychological well-being.

The British household panel study produced in 1992 is where the data originates from for this report, this amongst the UK data archive data dictionary will be used to select the variables that will be used throughout this report to produce tables. The dependent variable psychological well-being will be measured in six different categories, including one's ability to face issues, loss of confidence, feelings of worthlessness, being less able to enjoy day-to-day activities, and feeling under strain and unhappy. To see whether certain demographics are affected more than others we used variables similar to the original report such as gender (female and male), marital status (single), employment status (unemployed) etc. This would allow us to analysis certain trends with dependent and independent variables. The research's results indicate that unemployment has a significant impact on people's psychological well-being, particularly in the case of men, as shown later in the report. However, it should not be overlooked that both male and female unemployed individuals encounter difficulties, as evidenced by the report.

The report we recreated in RStudio duplicates the data and employs the same logistic regression strategy as the original research paper, employing methods such as coefficient and linear regressions to reach roughly equivalent conclusions. However, our findings differ slightly from those of the original paper. This could be due to a variety of factors, including how we subset the data and how we define certain variables, which differ due to the vagueness of the original report. We also analysed five of the six tables and went into detail from the results.

```
63  mutate(unemployed=bnemst==3,
64         unemployed=if_else(unemployed==TRUE,"1","0"))%>%
65  mutate(councilhousing=btenure==3,
66         councilhousing=if_else(councilhousing==TRUE,"1","0"))%>%
67  mutate(Male =bhgsex==1,
68         Male=if_else(Male==TRUE,"1","0"))%>%
69  mutate(White =brace==1,
70         White=if_else(White==TRUE,"1","0"))%>%
71  mutate(Couple =bmastat==2,
72         Couple=if_else(Couple==TRUE,"1","0"))%>%
73  mutate(Divorced =bmastat==4,
74         Divorced=if_else(Divorced==TRUE,"1","0"))%>%
```

Table One- Means and frequencies of the measures of the psychological well-being

	Mean	1 = Not at all	2 = No more than usual	3 = Rather more	4 = Much more
Constant Strain	2.11	20.63%	46.44%	23.14%	4.04%
Losing Confidence	1.64	46.63%	35.73%	10.46%	1.46%
Belief in Self-Worth	1.39	64.83%	23.05%	5.40%	1.03%

Table One - part two

	Mean	1 = More than usual	2 = Same as usual	3 = Less so	4 = Much less
General Happiness	2.02	12.41%	69.48%	10.57%	1.91%
Ability to Face Problems	2.02	9.16%	75.13%	8.63%	1.37%
Enjoyment of day-to-day	2.13	8.17%	67.89%	15.53%	2.73%

```
#The variable measuring the ability to face problems
summary(newdataset$Problems)
table(newdataset$Problems)

#The Variable measuring the enjoyment of day-to-day activities
summary(newdataset$Activities)
table(newdataset$Activities)
newdataset$Problems[newdataset$Problems== "-9"] <- NA


```

```
        other=if_else(other==TRUE,"1","0"))%>%
 mutate(Olevel=bqfachi==5,
        Olevel=if_else(Olevel==TRUE,"1","0"))%>%
 mutate(Alevel=bqfachi==4,
        Alevel=if_else(Alevel==TRUE,"1","0"))%>%
 mutate(noQu=bqfed==2,
        noQu=if_else(noQu==TRUE,"1","0"))%>%
```

Analysis Table 1:

The results created in table one is similar the findings from the original paper.

The table above shows the dependent variable's means and frequencies (psychological well-being).

The six variables chosen to assess one's psychosocial well-being show that the participants are not in 'psychological distress for the most part,' with the larger percentages under 'not at all,' 'no more than usual,' and the mean of the most part is 2, which connects the option 2 which is the same as usual.

```
cor.test(newdataset$Confidence,newdataset$Strain, method = "kendall")
#0.36
#Selfworth and under strain
cor.test(newdataset$Selfworth,newdataset$Strain, method = "kendall")
#0.28
#Happiness and under strain
cor.test(newdataset$Happiness,newdataset$Strain, method = "kendall")
#0.29

###Column two
#Problems and activities
cor.test(newdataset$Problems,newdataset$Activities, method = "kendall")
#0.36
#Confidence and activities
cor.test(newdataset$Confidence,newdataset$Activities, method = "kendall")
#0.27
#selfworth and activities
cor.test(newdataset$Selfworth,newdataset$Activities, method = "kendall")
#0.23
#Happiness and activities
cor.test(newdataset$Happiness,newdataset$Activities, method = "kendall")
#0.43
###Column three
#Confidence and problems
cor.test(newdataset$Confidence,newdataset$Problems, method = "kendall")
#0.28
#Selfworth and problems
cor.test(newdataset$Selfworth,newdataset$Problems, method = "kendall")
#0.24
#Happiness and problems
cor.test(newdataset$Happiness,newdataset$Problems, method = "kendall")
#0.37

###column four
#Selfworth and confidence
cor.test(newdataset$Selfworth,newdataset$Confidence, method = "kendall")
#0.56
#Happiness and confidence
cor.test(newdataset$Happiness,newdataset$Confidence, method = "kendall")
#0.29
```

```
182
183 ###column five
184 #Happiness and selfworth
185 cor.test(newdataset$Happiness,newdataset$Selfworth, method = "kendall")
186 #0.28
187
188 ```
```

Sidenote: Below are examples of what is shown in the console when you run the code.

```
        Kendall's rank correlation tau

data:  newdataset$Activities and newdataset$Strain
z = 37.134, p-value < 2.2e-16
alternative hypothesis: true tau is not equal to 0
sample estimates:
      tau
0.3486857


        Kendall's rank correlation tau

data:  newdataset$Problems and newdataset$Strain
z = 22.689, p-value < 2.2e-16
alternative hypothesis: true tau is not equal to 0
sample estimates:
      tau
0.2150284


        Kendall's rank correlation tau

data:  newdataset$Confidence and newdataset$Strain
z = 38.695, p-value < 2.2e-16
alternative hypothesis: true tau is not equal to 0
sample estimates:
      tau
0.3616229
```

Table 2 - Correlation coefficients between measures of psychological well-being

	Feeling under strain	Been less able to enjoy day-to-day activities	Been less able to face up to problems	Losing confidence	Thinking of being a worthless person
Been less able to enjoy day-to-day activities	0.34	X	X	X	X
Been less able to face up to problems	0.21	0.36	X	X	X
Losing confidence	0.36	0.27	0.28	X	X
Thinking of being a worthless person	0.28	0.23	0.24	0.56	X
Feeling less happy	0.29	0.43	0.37	0.29	0.28

Analysis Table 2:

When the findings from our tables are compared to the original report, there are some minor differences in correlation results with some of the variables. However, this is not a significant difference; it is due to the fact that we may have handled the data differently and utilised different software to obtain the answers.

The table indicates whether there is a statistical relationship between the two variables, whether positive or negative. This is an attempt to determine whether the measurement of psychological well-being has a relationship. The six measurements used in this research all indicate an overall presence of high correlation, implying that there is a positive linear relationship between the variables measured. This further supports this measurement of psychological wellbeing to be utilised through the report.

All of the measuring variables have a high significant relation with each other because the P-value is 0.002 which is a below 0.05. Since there is no direct hypothesis in this report, the null hypothesis would be rejected if there was a statement suggesting that being unemployed has an influence on one's psychological well-being.

```
#the marital status
table(thedataset$bmastat)
newdataset$bmastat[newdataset$bmastat == "0"] <- NA
newdataset$bmastat[newdataset$bmastat == "1"] <- NA
newdataset$bmastat[newdataset$bmastat == "3"] <- NA
table(newdataset$bmastat)

#The marital Status

newdataset$Single<-as.numeric(as.character(newdataset$Single))
summary(newdataset$Single)

newdataset$Couple<-as.numeric(as.character(newdataset$Couple))
summary(newdataset$Couple)
```

```
newdataset$Divorced<-as.numeric(as.character(newdataset$Divorced))
summary(newdataset$Divorced)

newdataset$Separated<-as.numeric(as.character(newdataset$Separated))
summary(newdataset$Separated+newdataset$Divorced)
```

```
summary(newdatas                                                    hing)

#O-level AND A-l
newdataset$Oleve
summary(newdatas

newdataset$Aleve
summary(newdatas

#No qualificatio
newdataset$noQu<
summary(newdatas
```

Table 3- Means and Definitions of the variables used

Variable	Mean	Definition
Age	50	Year of birth
Male	0.47	Sex - Gender (male or female
White	0.07	Ethnic group
Couple	0.21	Marital status- living as a couple
Divorced	0.18	Marital status- divorced and separated
Single	0.60	Marital status
Degree	0.22	Highest educational qualifications- higher degree, first degree, teaching QF, other QF
A level	0.14	Highest academic qualifications
O level	0.24	Highest academic qualifications
No qualifications	0.04	Highest educational qualifications- No QF
Council	0.16	Housing tenure- local authority rented
Unemployed	0.05	Current labour force status- unemployed
Low pay	0.59	Proxy's personal income banded- value 1-4 (£77-192 per week)
High pay	0.40	Proxy's personal income banded- value 5-14 (£193-674 per week)
Not in Labour Force	0.44	Employment status- value 3-8 - unemployed, retired, family care, FT student, disabled

Analysis Table 3:

The variables in the table above are similar to those in the original study, with the exclusion of No. of children, which was not in the dataset when we inserted the code, we discovered in the data dictionary. We took the upon ourselves to obtain similar variables to replicate the data, however for variables such ‘low pay’ we did not have the same definition of low in our data dictionary, so we group individuals with their weekly pay and categorised low pay as below value 4. This is a reason for our means to be different compared to the original report.

The table introduces independent variables to compare to with the dependent variables to see which is affected by the dependent variable most. The table shows different demographics and who is most likely to answer the questions. For example, the mean age was 50 years old, and respondents were more likely to be single. These are all the assumptions we can make for the mean and the table displayed above. The means allows us to draw assumptions about participants and create our own hypotheses.

```r
#Table 4

#Happiness
m2 <- glm (as.factor(Happiness) ~ unemployed +Single+noQu, data = newdataset,
           family = binomial)
summary(m2)
exp(coef(m2))
tab_model(m2,title = "Happiness")

```

```
Call:
glm(formula = as.factor(Happiness) ~ unemployed + Single + noQu,
    family = binomial, data = newdataset)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0692   0.5001   0.5001   0.5001   0.6765

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.01584    0.03686  54.696  < 2e-16 ***
unemployed  -0.03086    0.12898  -0.239    0.811
Single      -0.51549    0.06954  -7.412 1.24e-13 ***
noQu        -0.11145    0.18944  -0.588    0.556
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 7233.5  on 9290  degrees of freedom
Residual deviance: 7179.8  on 9287  degrees of freedom
  (554 observations deleted due to missingness)
AIC: 7187.8

Number of Fisher Scoring iterations: 4

(Intercept)  unemployed      Single        noQu
  7.5070374   0.9696128   0.5972096   0.8945390
```

Happiness

	as.factor(Happiness)		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	7.51	6.99 – 8.07	**<0.001**
unemployed	0.97	0.76 – 1.26	0.811
Single	0.60	0.52 – 0.69	**<0.001**
noQu	0.89	0.63 – 1.32	0.556
Observations	9291		
R^2 Tjur	0.006		

```
276
277 #Strain
278 m3 <- glm (as.factor(Strain) ~ unemployed +Single+noQu, data = newdataset,
279            family = binomial)
280 summary(m3)
281 exp(coef(m3))
282 tab_model(m3,title = "Strain")
```

```
Call:
glm(formula = as.factor(Happiness) ~ unemployed + Single + noQu,
    family = binomial, data = newdataset)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0692   0.5001   0.5001   0.5001   0.6765

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.01584    0.03686  54.696  < 2e-16 ***
unemployed  -0.03086    0.12898  -0.239    0.811
Single      -0.51549    0.06954  -7.412 1.24e-13 ***
noQu        -0.11145    0.18944  -0.588    0.556
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 7233.5  on 9290  degrees of freedom
Residual deviance: 7179.8  on 9287  degrees of freedom
  (554 observations deleted due to missingness)
AIC: 7187.8

Number of Fisher Scoring iterations: 4

> exp(coef(m2))
(Intercept)  unemployed      Single        noQu
  7.5070374   0.9696128   0.5972096   0.8945390
> tab_model(m2,title = "Happiness")
```

Strain

	as.factor(Strain)		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	3.90	3.68 – 4.13	**<0.001**
unemployed	0.93	0.76 – 1.15	0.489
Single	0.71	0.63 – 0.80	**<0.001**
noQu	0.72	0.54 – 0.97	**0.025**
Observations	9281		
R^2 Tjur	0.005		

```r
#confidence
m4 <- glm (as.factor(Confidence) ~ unemployed +Single+noQu, data = newdataset,
           family = binomial)
summary(m4)
exp(coef(m4))
tab_model(m4,title = "confidence")

```

```
Call:
glm(formula = as.factor(Confidence) ~ unemployed + Single + noQu,
    family = binomial, data = newdataset)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.290  -1.188   1.069   1.167   1.268

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.02376    0.02383   0.997  0.31880
unemployed   0.23710    0.09118   2.600  0.00931 **
Single      -0.05226    0.05157  -1.013  0.31084
noQu        -0.18203    0.13418  -1.357  0.17489
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 12868  on 9282  degrees of freedom
Residual deviance: 12859  on 9279  degrees of freedom
  (562 observations deleted due to missingness)
AIC: 12867

Number of Fisher Scoring iterations: 3

> exp(coef(m4))
(Intercept)  unemployed      Single        noQu
  1.0240414   1.2675651   0.9490797   0.8335762
> tab_model(m4,title = "confidence")
```

confidence

	as.factor(Confidence)		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	1.02	0.98 – 1.07	0.319
unemployed	1.27	1.06 – 1.52	**0.009**
Single	0.95	0.86 – 1.05	0.311
noQu	0.83	0.64 – 1.08	0.175
Observations	9283		
R^2 Tjur	0.000		

```r
#Activities
m5 <- glm (as.factor(Activities) ~ unemployed +Single+noQu, data = newdataset,
           family = binomial)
summary(m5)
exp(coef(m5))
tab_model(m5, title="Activities")
```

```
Call:
glm(formula = as.factor(Activities) ~ unemployed + Single + noQu,
    family = binomial, data = newdataset)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2829   0.3917   0.3917   0.3917   0.6768

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.52905    0.04524  55.907  < 2e-16 ***
unemployed  -0.12741    0.14758  -0.863   0.3880
Single      -0.60795    0.08137  -7.471 7.94e-14 ***
noQu        -0.43638    0.20098  -2.171   0.0299 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 5475.4  on 9286  degrees of freedom
Residual deviance: 5415.3  on 9283  degrees of freedom
  (558 observations deleted due to missingness)
AIC: 5423.3

Number of Fisher Scoring iterations: 5

> exp(coef(m5))
(Intercept)  unemployed      Single        noQu
 12.5415724   0.8803767   0.5444642   0.6463747
> tab_model(m5, title="Activities")
```

Activities

	as.factor(Activities)		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	12.54	11.49 – 13.72	**<0.001**
unemployed	0.88	0.66 – 1.19	0.388
Single	0.54	0.46 – 0.64	**<0.001**
noQu	0.65	0.44 – 0.98	**0.030**
Observations	9287		
R^2 Tjur	0.007		

```r
#self-worth
m6 <- glm (as.factor(Selfworth) ~ unemployed +Single+noQu, data = newdataset,
           family = binomial)
summary(m6)
exp(coef(m6))
tab_model(m6,title="Self-worth")
```

```
Call:
glm(formula = as.factor(Selfworth) ~ unemployed + Single + noQu,
    family = binomial, data = newdataset)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.1093  -0.8541  -0.8541   1.5397   1.5456

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.82062    0.02582 -31.778  < 2e-16 ***
unemployed   0.51724    0.09207   5.618 1.94e-08 ***
Single      -0.01302    0.05563  -0.234    0.815
noQu         0.14092    0.14066   1.002    0.316
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 11534  on 9284  degrees of freedom
Residual deviance: 11502  on 9281  degrees of freedom
  (560 observations deleted due to missingness)
AIC: 11510

Number of Fisher Scoring iterations: 4

> exp(coef(m6))
(Intercept)  unemployed      Single        noQu
  0.4401598   1.6773907   0.9870618   1.1513339
> tab_model(m6,title="Self-worth")
```

Self-worth

	as.factor(Selfworth)		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.44	0.42 – 0.46	**<0.001**
unemployed	1.68	1.40 – 2.01	**<0.001**
Single	0.99	0.88 – 1.10	0.815
noQu	1.15	0.87 – 1.51	0.316
Observations	9285		
R^2 Tjur	0.011		

```r
#Problems

m7 <- glm (as.factor(Problems) ~ unemployed +Single+noQu, data = newdataset,
           family = binomial)
summary(m7)
exp(coef(m7))
tab_model(m7, title = "Problems")
```

```
glm(formula = as.factor(Problems) ~ unemployed + Single + noQu,
    family = binomial, data = newdataset)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2492   0.4076   0.4076   0.4076   0.7060

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.44641    0.04370  55.980   <2e-16 ***
unemployed  -0.01939    0.14443  -0.134   0.8932
Single      -0.77931    0.07598 -10.257   <2e-16 ***
noQu        -0.38556    0.19606  -1.967   0.0492 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 5919.3  on 9283  degrees of freedom
Residual deviance: 5815.5  on 9280  degrees of freedom
  (561 observations deleted due to missingness)
AIC: 5823.5

Number of Fisher Scoring iterations: 5

> exp(coef(m7))
(Intercept)  unemployed      Single        noQu
 11.5468692   0.9807924   0.4587209   0.6800691
> tab_model(m7, title = "Problems")
```

Problems

	as.factor(Problems)		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	11.55	10.61 – 12.59	**<0.001**
unemployed	0.98	0.74 – 1.31	0.893
Single	0.46	0.40 – 0.53	**<0.001**
noQu	0.68	0.47 – 1.02	**0.049**
Observations	9284		
R^2 Tjur	0.012		

Analysis for 4

Table 4 is different from the original report as I focused on selective variables which includes: unemployed, single and no qualifications. I decided to focus on these three because the chosen variables contrast each other as they are not very similar but can all correlate when measuring psychological distress. As the report focuses on unemployment, it would be important to assess its impacts. The variable of no qualifications would show the effect psychological distress would have as participants may view themselves in a negative perspective of being worthless. Additionally, the variable for single would explain how not being in a relationship can differ the experience of psychological distress,

I decided to run a general linear regression model, and I also used the exp(coef()) code to further interpret the regression in detail, and as it would allow me to exponentiate the liner regression. I used the family function (binomial) for the logistic regression as this would show the sampling, density, and distribution function; this allows the glm function to create meaningful findings. Alongside to this, I used the odds ratio function to interpret if the independent variable impacts the dependent variable and how high are the odds of it happening or not happening.

The odds ratio table for self-worth shows that the likeliness is over one for both unemployment (1.68) and for no qualification (1.15), which means that they have more chances of impacting the predictor variables. Correspondingly, the table for confidence shows that unemployment is at 1.27, again revealing that the odds of unemployment affecting an increase in the predictor variable is more significant than one. Whereas the value is below one for all of the other variables, demonstrating that there is lower than one odd in ratio of achieving the specific sum of increase for the predictor variable.

```r
#Table 6
#problems
#Men
menp<-glm(as.factor(Problems*bhgsex) ~ unemployed,
          data = newdataset[newdataset$bhgsex==1,],family = binomial)
summary(menp)
tab_model(menp,title = "Been less able to face up to problems as a men and
          is there a correlation with being unemployed?")
confint(menp)
exp(coef(menp))
#Women
womenp<-glm(as.factor(Problems*bhgsex) ~ unemployed,
          data = newdataset[newdataset$bhgsex==2,],family = binomial)
summary(womenp)
tab_model(womenp,title = "Been less able to face up to problems as a women and
          is there a correlation with being unemployed?")
confint(womenp)
exp(coef(womenp))
#Happiness
#Men
menh<-glm(as.factor(Happiness*bhgsex) ~ unemployed,
          data = newdataset[newdataset$bhgsex==1,],family = binomial)
summary(menh)
tab_model(menh, title = "Feeling less happiness as a man and
          is there a correlation with being unemployed?")
confint(menh)
exp(coef(menh))
#women
womenh<-glm(as.factor(Happiness*bhgsex) ~ unemployed,
          data = newdataset[newdataset$bhgsex==2,],family = binomial)
summary(womenh)
tab_model(menh,title = "Feeling less happiness as a women and
           is there a correlation with being unemployed?")
confint(womenh)
exp(coef(womenh))
#confidence
#Men
menc<-glm(as.factor(Confindence*bhgsex) ~ unemployed,
           data = newdataset[newdataset$bhgsex==1,],family = binomial)
summary(menc)
tab_model(menc, title = "Losing confidence in men and
           is there a correlation with being unemployed? ")
confint(menc)
exp(coef(menc))
#women
womenc<-glm(as.factor(Confidence*bhgsex) ~ unemployed,
           data = newdataset[newdataset$bhgsex==2,],family = binomial)
summary(womenc)
tab_model(womenc,  title = "Losing confidence in women and
           is there a correlation with being unemployed? ")
confint(womenc)
exp(coef(womenc))
```

Feeling less happiness as a man and is there a correlation with being unemployed?

	as.factor(Happiness * bhgsex)		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	6.52	5.95 – 7.15	**<0.001**
unemployed	0.97	0.72 – 1.35	0.871
Observations	4283		
R^2 Tjur	0.000		

Feeling less happiness as a women and is there a correlation with being unemployed?

	as.factor(Happiness * bhgsex)		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	6.52	5.95 – 7.15	**<0.001**
unemployed	0.97	0.72 – 1.35	0.871
Observations	4283		
R^2 Tjur	0.000		

Though of being worthless as a man and is there a correlation with being unemployed?

	as.factor(Selfworth * bhgsex)		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.32	0.30 – 0.35	**<0.001**
unemployed	2.00	1.59 – 2.49	**<0.001**
Observations	4283		
R^2 Tjur	0.005		

Though of being worthless as a woman and is there a correlation with being unemployed?

	as.factor(Selfworth * bhgsex)		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.55	0.52 – 0.59	**<0.001**
unemployed	1.81	1.32 – 2.47	**<0.001**
Observations	5002		
R^2 Tjur	0.008		

Feeling constant strain male participants may feel and is it related to being unemployed?

	as.factor(Strain * bhgsex)		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	2.96	2.76 – 3.18	**<0.001**
unemployed	0.89	0.70 – 1.14	0.365
Observations	4281		
R^2 Tjur	0.000		

Feeling constant strain that female participants may feel and is it related to being unemployed?

	as.factor(Strain * bhgsex)		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	2.96	2.76 – 3.18	**<0.001**
unemployed	0.89	0.70 – 1.14	0.365
Observations	4281		
R^2 Tjur	0.000		

Relationship between men being less able to enjoy day-to- day activities whilst being unemployed

	as.factor(Activities * bhgsex)		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	10.11	9.08 – 11.29	**<0.001**
unemployed	1.02	0.71 – 1.51	0.930
Observations	4282		
R^2 Tjur	0.000		

Relationship between women being less able to enjoy day-to- day activities whilst being unemployed

	as.factor(Activities * bhgsex)		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	11.26	10.17 – 12.50	**<0.001**
unemployed	0.49	0.32 – 0.78	**0.001**
Observations	5005		
R^2 Tjur	0.002		

Losing confidence in men and is there a correlation with being unemployed?

	as.factor(Confidence * bhgsex)		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.73	0.68 – 0.78	**<0.001**
unemployed	1.37	1.11 – 1.71	**0.004**
Observations	4280		
R^2 Tjur	0.001		

Losing confidence in women and is there a correlation with being unemployed?

	as.factor(Confidence * bhgsex)		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	1.32	1.24 – 1.39	**<0.001**
unemployed	1.64	1.18 – 2.30	**0.004**
Observations	5003		
R^2 Tjur	0.001		

Been less able to face up to problems as a men and is there a correlation with being unemployed?

	as.factor(Problems * bhgsex)		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	9.10	8.20 – 10.12	**<0.001**
unemployed	0.93	0.67 – 1.35	0.705
Observations	4279		
R^2 Tjur	0.000		

Been less able to face up to problems as a women and is there a correlation with being unemployed?

	as.factor(Problems * bhgsex)		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	9.67	8.79 – 10.66	**<0.001**
unemployed	0.66	0.43 – 1.08	0.080
Observations	5005		
R^2 Tjur	0.001		

Table 6 :

This is a replication of table six from the original report; however, the difference is that they investigated low paid and not in labour force, whilst I focused on just the unemployment variable as it's one of the main investigations of the report. For this reason, I was not able to analyse the relation the two variables of not being in the labour force and low pay would have on gender. That being said, my results do conclude that male participants have a higher likelihood in being affected by psychological distress than female participants, which is also the case for the original report.

Likewise, for table six, I again utilised both the general linear regression and the odds ratio function. I decided to focus on the variable of gender for table six, by comparing gender to the measuring variables, will allow me to showcase if the participants gender may affect the outcomes. The likelihood of odds ratio is below one for both male and female participates when comparing to the strain and happiness measurements, having lower significance. The self-worth table shows that the likelihood of odds is double for male participants (2), but is also high for female participants (1.81), except it's just below a double chance. This indicates that unemployment impacts self-worth, especially for men.

The table produced to show the measuring variable of being able to do day- to- day activities have odds at more than one for men(1.01), but in the case of women it is only half (0.49), again this show's men can be more impacted. Confidence is the only measuring table that shows female participants (1.64) have a higher odd ratio than male participants (1.37), this indicates that whilst men have higher chances of experiencing all the different measurements of psychological distress when experiencing unemployment, women are mostly impacted on their confidence as a measurement of distress.

Finally, both male(0.93) and female(0.66) participants scored below one for the measurement of dealing with problems, which can show that it's not something that is significantly affected by unemployment. In general, based on how unemployment impacts the different measurements of psychological distress, we can see that male participants are more vulnerable in being impacted from being unemployed, but despite this is less for female participants, unemployment still impacts them somewhat.

Conclusion

"The effects of low-pay and unemployment on psychological well-being: a logistic regression approach" by Theodossiou, has been selected to replicate the findings of my report. As for the struggles that I encountered during the report, I experienced some complications for the coding in terms of creating new datasets which would include the variables I had selected, but I managed to figure and sort this out with the mutate and select function, enabling me to change the data to fit my measurements. I also experienced troubles with the 'family binomial' function when producing the glm model, but after some research and assessing the issue, I managed to use the (as.factor) to get the code working. Another issue that I had was not being able to get Rmarkdown to work, I extensively researched to find a way to resolve this issue, I downloaded Mixtex console and Latexit, as well as installing numerous packages and using cran and trying it on my team members laptop, but unfortunately, none of it worked. I believe this is a common issue as other people were experiencing it too online, possibly because Rmarkdown can't run certain functions like the mutate and select. As I didn't consider the variable of age, I didn't reproduce table 5 for the report, but in the future maybe it would be interesting to look at the age variable too. Lastly, the final problem I encountered was that for the replication of table 3, the variable of a "child is dependent" does not seem to exist in the dataset, the code for this is 'bdepchl'.

As for how I would improve the original report, I would suggest creating an updated version since the report was conducted in 1998, an updated version would include more variables to fit the current

society, for example, using ‘anxiety’ and ‘depression’ as measurements of psychological distress since it is more spoken of in current society compared to when the original report was conducted. Furthermore, the category for race is only defined by one value of ‘white’, consequently this does not show how different ethnic backgrounds may experience unemployment and psychological distress, so for improvement it would be better to include other ethnicities. Similarly, the original report was also vague in defining the variables, so in certain instances, I had to improvise by figuring out values that would be similar to their definitions in order to ensure similar outcomes. Generally, even though we as a pair experienced some challenges, we were able to replicate and evaluate the report sufficiently, we tried utilising our strong points by working as a team and helping each other out, the best outcome for me was how much I have developed in R studio by learning new codes and understanding how I could utilise them.

The raw code

```
---
title: 'The effects of low-pay and unemployment on psychological well-being; a logistic
  regression approach '
author: "Yasmine Calvert, Mina Kamawal"
output:
  word_document: default
  pdf_document: default
  html_document: default
---

```{r setup, include=FALSE}
```
```