Jane Maguire
SC321-Project03
5/13/24

Data Analysis: Final Project

Introduction:

This data comes from an edited subset of the Second Longitudinal Study of Aging(LSOA II). This dataset is “representative of the civilian noninstitutionalized U.S. population 70 years of age and over at mid-year 1995”. Six different variables were recorded for each person in the dataset. The variables recorded are if a person is frequently depressed/anxious, if they have trouble seeing even with glasses, if present social activities feel like enough, if they are male, their age in years, and the number of days leaving the house in the past 2 weeks. The objective of this research is to determine if the proportion of elderly people with depression/anxiety is higher when subsetting based on certain variables. Specifically focusing on how vision problems and how socially active a person is affects the proportion of depressed/anxious elderly people.

Methods:

This dataset consists of six variables: DeprAnx, VisionBad, SocialEnough, Male, Age, and DaysOut; defined in table 1. We perform separate simple logistic regressions for each explanatory variable to find the "unadjusted" ORs and their CIs. We then perform a single multiple logistic regression using all of the main effects in order to find the "adjusted" ORs and their CIs.

Table 1: Shows the variable name and definition for each variable in the dataset

Variable Name:	**Definition:**
DeprAnx	1 = is frequently depressed and/or anxious, 0 = is not
VisionBad	1 = has trouble seeing even with glasses, 0 = does not
SocialEnough	1 = present social activities feel like enough, 0 = too little
Male	1 = male, 0 = female
Age	In years
DaysOut	Number of days leaving the house in past 2 weeks (from 0 if did not leave house, to 14 if left every day)

Results:

Table 2 shows the descriptive statistics for the four binary variables; DeprAnx, VisionBad, SocialEnough and Male. This table provides a count and percentage of people in each group (0 or 1) for each variable. Table 3 shows the summary statistics for the 2 quantitative variables; Age and DaysOut. This table provides the mean, median, standard deviation, and interquartile range for these two variables.

Table 2: Descriptive Statistics of Binary Variables

	DeprAnx		**VisionBad**		**SocialEnough**		**Male**	
	Count	%	Count	%	Count	%	Count	%
0	6829	92.4	6296	85.19	1751	23.69	4539	61.42
1	561	7.59	1094	14.80	5639	76.31	2851	38.58

Table 3: Summary Statistics of Quantitative Variables

Variable Name:	**Mean**	**Median**	**SD**	**IQR**
Age	76.06	75.00	5.510276	8.00
DaysOut	10.01	14.00	4.982683	9.00

After looking at the descriptive statistics for each variable, we then move on to creating linear models for predicting depression/anxiety based on each of the predictor variables individually. By doing so we were then able to extract the coefficients of the model and calculate the odds ratios for each individual predictor as well as calculating the confidence intervals for each odds ratio. The unadjusted odds ratios and confidence intervals for each predictor variable is shown in table 4.

Table 4: Unadjusted OR's and CI's that estimate relationships separately

	VisionBad	SocialEnough	Male	Age	DaysOut
Odds Ratio	2.228535	0.3769467	0.563215	1.012572	0.9105441
Confidence Interval	(1.824643, 2.721829)	(0.3158954, 0.4497969)	(0.4644519, 0.6829795)	(0.997235, 1.028145)	(0.8957989, 0.9255320)

After looking at the Unadjusted OR's and CI's that estimate relationships separately, we then move on to look at the adjusted OR's and CI's for a model that controls for all 5 variables jointly. We first create the model and check the LIBS conditions. These conditions are all met, so we then move on to extract the coefficients of the model and calculate the odds ratios for each predictor as well as calculating the confidence intervals for each odds ratio. The adjusted odds ratios and confidence intervals for each predictor variable is shown in table 5.

Table 5: Adjusted OR's and CI's based on model that controls for all 5 variables jointly

	VisionBad	SocialEnough	Male	Age	DaysOut
Odds Ratio	1.8089328	0.4582022	0.6849279	0.9860326	0.9322844
Confidence Interval	(1.4683802, 2.2284677)	(0.3817496, 0.5499658)	(0.5607487, 0.8366069)	(0.9702704, 1.0020508)	(0.9158418, 0.9490223)

We can see that the VisionBad and SociaEnough variables have the most deviation from 1 which means it is not just a 50/50 shot and there is more likely a correlation. Because of this it makes sense to focus efforts on these two variables. Based on these results we can then use the odds ratios to interpret if depression/anxiety is higher when subsetting based on certain variables.

Discussion:

We found the proportion of elderly people with depression/anxiety is higher when subsetting based on certain variables, specifically, those who are vision impaired and are socially isolated. In order to determine this, we needed to interpret the odds ratios provided to understand how each variable affects the odds of experiencing depression and/or anxiety. The VisionBad variable has an odds ratio of 1.8089. In context this means that if a person has trouble seeing even with glasses (VisionBad = 1), the odds of experiencing depression and/or anxiety are 1.8089 times higher, while holding other variables constant. Individuals with vision impairment are at a higher risk of depression and/or anxiety compared to those without vision impairment. The SocialEnough variable has an odds ratio of 0.4582. In context this means that if a person's present social activities feel like enough(SocialEnough =1), the odds of experiencing depression and/or anxiety are 0.4582 times lower, while holding other variables constant. Individuals who feel social enough are at a lower risk of depression and/or anxiety compared to those who do not feel socially enough. The male variable has an odds ratio of 0.6849. This means that males have odds of experiencing depression and/or anxiety that are 0.6849 times those of females, while holding other variables constant. Therefore, females are at a higher risk of depression and/or anxiety compared to males, as the odds ratio is less than 1. The age variable has an odds ratio of 0.9860. This means that for every one-year increase in age, the odds of experiencing depression and/or anxiety decrease by a factor of 0.9860, while holding other variables constant. Therefore, younger individuals are at a slightly higher risk of depression and/or anxiety compared to older individuals, although the effect size is small. The DaysOut variable has an odds ratio of 0.9323. This means that for every one-unit increase in the number of days spent outside, the odds of experiencing depression and/or anxiety decrease by a factor of 0.9323, while holding other variables constant. Therefore, individuals who spend more days outside are at a slightly lower risk of depression and/or anxiety compared to those who spend fewer days outside. Based on these odds ratios, the combination of explanatory variables associated with the highest risk of depression and/or anxiety in the sampled population includes individuals with vision impairment, individuals who feel socially isolated, females, younger individuals, individuals who spend fewer days outside.

A recommendation I would give based on these results is to focus on enhancing strategies to create social support networks for individuals with vision problems in order to help mitigate the risk of depression and anxiety. Encouraging participation in social activities, fostering connections with peers, and providing emotional support could be beneficial. Also, prioritizing vision care and accessibility measures could help reduce vision-related distress and improve overall well-being. Access to regular vision screenings, assistive devices, and vision rehabilitation programs can increase quality of life for individuals with vision problems. A limitation to this dataset is that binary variables are used to measure vision and socialness and because of this we are not able to track depression and anxiety on a continuous scale. I think in future research it would be interesting to measure with quantitative continuous variables. In future research I would also like to see how other variables may play a factor. For example, the elderly population is not as well versed in technology and technology is an easy way to communicate with friends and family and can also provide entertainment. Because of this I would be curious to see if the use of technology would affect levels of depression/anxiety. In conclusion, targeting interventions towards older females with vision impairment, social isolation, and limited outdoor activities may be a priority to address the highest risk of depression and/or anxiety in the elderly population.

Appendix

Jane Maguire

2024-05-13

Load Libraries and Data

```
library(Stat2Data)
load("soa2_SC321.RData")
```

Descriptive Statistics for Binary Variables

First, we look at the descriptive statistics for the binary variables by pulling the counts and percentages for each factor (0 or 1) for each variable.

```
# Counts/percentages of binary statistics for Depr/Anx
table(soa2$DeprAnx)
```

```
##
##    0    1
## 6829  561
```

```
prop.table(table(soa2$DeprAnx)) * 100
```

```
##
##        0        1
## 92.40866  7.59134
```

```
# Counts/percentages of binary statistics for VisionBad
table(soa2$VisionBad)
```

```
##
##    0    1
## 6296 1094
```

```
prop.table(table(soa2$VisionBad)) * 100
```

```
##
##        0        1
## 85.19621 14.80379
```

```
# Counts/percentages of binary statistics for SocialEnough
table(soa2$SocialEnough)
```

```
##
##    0    1
## 1751 5639
```

```
prop.table(table(soa2$SocialEnough)) * 100
```

```
##
##        0        1
```

```
## 23.69418 76.30582
```

```r
# Counts/percentages of binary statistics for Male
table(soa2$Male)
```

```
##
##    0    1
## 4539 2851
```

```r
prop.table(table(soa2$Male)) * 100
```

```
##
##        0        1
## 61.42084 38.57916
```

Summary Statistics for Qunatitative Variables

Then, we look at the summary statistics for the two quantitative variables in order to look at Mean, median, and IQR. We also pull that standard deviation for each variable.

```r
# Summary statistics for Age
summary(soa2$Age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   69.00   72.00   75.00   76.06   80.00   99.00
```

```r
sd(soa2$Age)
```

```
## [1] 5.510276
```

```r
# Summary statistics for DaysOut
summary(soa2$DaysOut)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00    5.00   14.00   10.01   14.00   14.00
```

```r
sd(soa2$DaysOut)
```

```
## [1] 4.982683
```

Unadjusted OR's and CI's that estimate relationships seperatley

To begin our analysis, we fit models for predicting depression and anxiety with each of the predictor variables separately. We then pull the coefficients from each of these models in order to calculate the odds ratios. We then also calculate the confidence interval for each of these odds ratios.

```r
#DeprAnx and VisionBad
model_vision <- glm(DeprAnx ~ VisionBad, data = soa2, family = binomial)
vision_odds <- exp(coef(model_vision))
vision_odds[2] #Odds Ratio
```

```
## VisionBad
##  2.228535
```

```r
exp(confint.lm(model_vision)[2,]) # Confidence Interval
```

```
##    2.5 %   97.5 %
## 1.824643 2.721829
```

```r
#DeprAnx and SocialEnough
model_social <- glm(DeprAnx ~ SocialEnough, data = soa2, family = binomial)
```

```r
social_odds <- exp(coef(model_social))
social_odds[2] #Odds Ratio
```

```
## SocialEnough
##    0.3769467
```

```r
exp(confint.lm(model_social)[2,]) # Confidence Interval
```

```
##     2.5 %    97.5 %
## 0.3158954 0.4497969
```

```r
#DeprAnx and Male
model_male <- glm(DeprAnx ~ Male, data = soa2, family = binomial)
male_odds <- exp(coef(model_male))
male_odds[2] #Odds Ratio
```

```
##     Male
## 0.563215
```

```r
exp(confint.lm(model_male)[2,]) # Confidence Interval
```

```
##     2.5 %    97.5 %
## 0.4644519 0.6829795
```

```r
#DeprAnx and Age
model_age <- glm(DeprAnx ~ Age, data = soa2, family = binomial)
age_odds <- exp(coef(model_age))
age_odds[2] #Odds Ratio
```

```
##      Age
## 1.012572
```

```r
exp(confint.lm(model_age)[2,]) # Confidence Interval
```

```
##    2.5 %   97.5 %
## 0.997235 1.028145
```

```r
#DeprAnx and DaysOut
model_days <- glm(DeprAnx ~ DaysOut, data = soa2, family = binomial)
days_odds <- exp(coef(model_days))
days_odds[2] #Odds Ratio
```

```
##   DaysOut
## 0.9105441
```

```r
exp(confint.lm(model_days)[2,]) # Confidence Interval
```

```
##     2.5 %    97.5 %
## 0.8957989 0.9255320
```

Adjusted OR's and CI's based on model that controls for all 5 variables jointly

After fitting the models for each of the predictor variables separately, we now fit a model that controls for all 5 variables jointly.

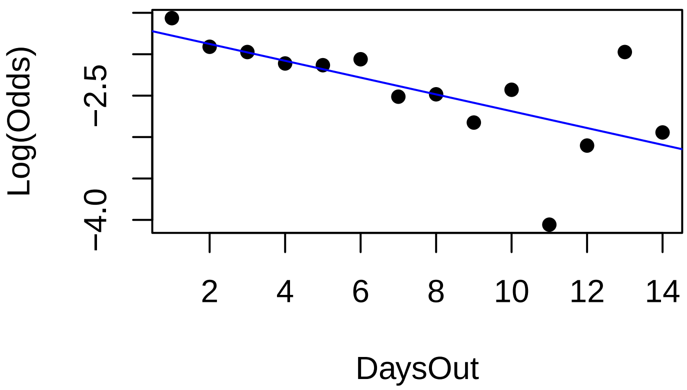
```r
# Fit logistic regression model with all predictors
model_all <- glm(DeprAnx ~ VisionBad + SocialEnough + Male + Age + DaysOut, data = soa2, family = binomi
```

We then check the LIBS conditions to confirm that conditions are met. The LIBS conditions are as follows.

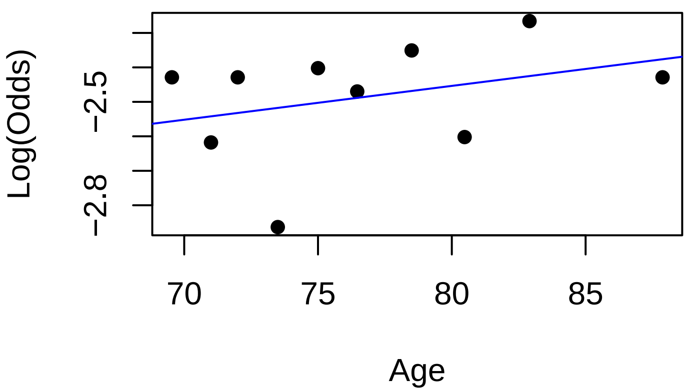
L: the log odds at a given x, i.e. $logit(\pi(x_i))$, is a linear function $\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots$ I: the y_i values are independent of each other B: the responses y_i each have a Bernoulli distribution that depends only on $\pi(x_i)$ S: the data come from a simple random sample

We can check the linearity assumption by creating a empirical logit plot. Because there are multiple binary variables that are also predictors we can not have more than 2 groups for those so linearity is guaranteed. We can check the Age and days out predictors with the graphs below and see than linearity is met.

```
attach(soa2)
emplogitplot1(model_days, breaks = 0:14)
```




```
emplogitplot1(model_age, ngroups = 10)
```




```
detach(soa2)
```

We know from the data given to us that the y_i values are independent of each other so the I assumption is met.

The B assumption is met because the responses y_i each have a Bernoulli distribution meaning the response(DeprAnx) is a binary variable.

The S assumption is met because, again, we know from the information given that the data comes from a simple random sample.

After checking all conditions, we now can use the model to obtain the odds ratios and confidence intervals just as we did in the first step of our analysis.

```
# Get adjusted odds ratios and confidence intervals
adjusted_odds <- exp(coef(model_all))
adjusted_ci <- exp(confint.lm(model_all))
adjusted_odds
```

```
##  (Intercept)    VisionBad SocialEnough         Male          Age      DaysOut
```

```
##    0.7744133    1.8089328    0.4582022    0.6849279    0.9860326    0.9322844
```

```
adjusted_ci
```

```
##                  2.5 %    97.5 %
## (Intercept)  0.2174186 2.7583473
## VisionBad    1.4683802 2.2284677
## SocialEnough 0.3817496 0.5499658
## Male         0.5607487 0.8366069
## Age          0.9702704 1.0020508
## DaysOut      0.9158418 0.9490223
```

After obtaining the odds ratios and confidence intervals we have enough information to make inferences about our data. As mentioned in the discussion, based off the odds ratios for the full model, targeting interventions towards older females with vision impairment, social isolation, and limited outdoor activities may be a priority to address the highest risk of depression and/or anxiety in the elderly population.