



School of Social Sciences and Philosophy Assignment Submission Form

Student Name:	Burak Ozuer
Student ID Number:	18302382
Programme Title:	Applied Social Data Science(P.Grad.Dip)
Module Title:	Applied Statistical Analysis II
Assessment Title:	Replication
Lecturer (s):	Jeffrey Ziegler
Date Submitted	01/04/2024

I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at: <http://www.tcd.ie/calendar>

I have also completed the Online Tutorial on avoiding plagiarism 'Ready, Steady, Write', located at <http://tcd-ie.libguides.com/plagiarism/ready-steady-write>

Signed: **BURAK OZUER**_____

Date: **01/04/2024**_____



Behavioural Functions Associated with Wanting to Reduce Internet Use

The article discusses the development of self-report scales for assessing problematic internet use and the need to study the motivations behind internet usage. It suggests that understanding the environmental consequences that drive internet use can provide valuable insights. The use of behavior rating scales is explained as a way to identify factors that maintain specific behaviors, such as gambling or video gaming. Problem gamblers, for example, tend to endorse items related to negative reinforcement, indicating that they gamble to escape certain aspects of their lives. The Preliminary Internet Consequences Questionnaire (ICQ-P) is introduced as a tool for assessing the consequences that maintain internet use. Eligible participants who chose to take part in the survey were directed to a landing page on PsyToolkit, a web-based service for conducting online surveys. After giving informed consent, participants provided demographic information and answered questions about their desire to spend less time on the internet and their estimated weekly online usage. They then completed the ICQ-P, which consisted of 30 statements related to various behavioral functions. The final part of the survey involved the YDQ, which is an eight-item questionnaire adapted from the APA's criteria for pathological gambling. Answering "yes" to five or more questions on the YDQ indicates problematic internet use.

Finding: Estimated weekly online usage survey answer was not included in the OSF storage data file.

Likert scales

scale: frequency

- never
- almost never
- seldom
- half the time
- usually
- almost always
- always

scale: binary

- Yes
- No



Hypothesis:

Hypothesis 1: The desire to reduce internet use is associated with problematic internet use as measured by the Young Diagnostic Questionnaire (YDQ; Young, 1998). The so-called problematic internet users may be less likely to want to reduce internet use than nonproblematic internet users simply because they value internet use more. Conversely, it is also possible that problematic internet users are more likely to want to reduce internet use than nonproblematic internet users.

Data and analysis can be found: <https://osf.io/kwf4e/>

Data Prep:

One issue is that the variables in question are not clearly defined in the article or supplementary documents. I had to manipulate the data in order to explore different transformation methods.

I've created the following dictionary by analysing the survey questions shared in the of page, which I found very helpful to understand the data.

Attribute Name	Question
IU_W2R	Do you wish you spent less time on the internet?
ICQ_01_t	I go online to make money.
ICQ_02_a	I like the community aspects of using the internet.
ICQ_03_se	I like visual and auditory aspects of using the internet.
ICQ_04_d	I use the internet to get a break from work or other difficult tasks.
ICQ_05_so	I find it is better to confront someone over the internet rather than doing it face to face.
ICQ_06_c	I use the internet when I can't stop thinking about something.



ICQ_07_t	I earn money when I use the internet.
ICQ_08_a	I go online to connect with my friends.
ICQ_09_se	I feel more alive when I am using the internet than when I am doing other types of activities.
ICQ_10_d	I use the internet to procrastinate.
ICQ_11_so	I use the internet after being in an argument or fight.
ICQ_12_c	I use the internet when I am feeling depressed or sad.
ICQ_13_t	I enjoy shopping online.
ICQ_14_a	I like interacting with other people on the internet.
ICQ_15_se	I feel a rush and get excited when I use the internet.
ICQ_16_d	I use the internet to avoid work or other things I need to do.
ICQ_17_so	I use the internet when I am not getting enough support from the people in my life.
ICQ_18_c	I go online when I feel stressed or anxious.
ICQ_19_t	I use coupons and discounts I learn about online.
ICQ_20_a	I enjoy getting attention from people on the internet.
ICQ_21_se	I go online when there is nothing else going on or I have nothing better to do.
ICQ_22_d	I use the internet when I have a work project or class assignment that is due soon.



ICQ_23_s o	I use the internet as a buffer between other people and me.
ICQ_24_ c	I use the internet when I am worried.
ICQ_25_t	When I use the internet, people offer me gifts, money, or other items.
ICQ_26_ a	I enjoy meeting new people online.
ICQ_27_s e	When I use the internet, I get immersed in what's on the screen rather than my physical surroundings.
ICQ_28_ d	If I must do something boring, difficult, or unpleasant, I use the internet first to psych myself up.
ICQ_29_s o	I use the internet when people give me a hard time.
ICQ_30_ c	I use the internet as a way of keeping my mind off my problems.
YDQ.1	Do you feel preoccupied with the internet (think about previous online activity or anticipate next online session)?
YDQ.2	Do you feel the need to use the internet with increasing amounts of time in order to achieve satisfaction?
YDQ.3	Have you repeatedly made unsuccessful efforts to control, cut back, or stop internet use?
YDQ.4	Do you feel restless, moody, depressed or irritable when attempting to cut down or stop internet use?
YDQ.5	Do you stay online longer than originally intended?
YDQ.6	Have you jeopardized or risked the loss of a significant relationship, job, educational, or career opportunity because of the internet?
YDQ.7	Have you lied to family members, a therapist, or others to conceal the extent of involvement with the internet?
YDQ.8	Do you use the internet as a way of escaping from problems or of relieving a dysphoric mood (e.g. feelings



Finding: Please find below the data transformation steps that have been incorporated into the analysis. These steps were not adequately explained in the article. One of the supplementary documents included certain sections of the descriptive analysis; however, without clear explanations regarding data mapping and transformation, replicating the same results from a descriptive analysis perspective proved to be challenging for me.

```
df$YDQ.1[df$YDQ.1 == 2] <- 0
df$YDQ.2[df$YDQ.2 == 2] <- 0
df$YDQ.3[df$YDQ.3 == 2] <- 0
df$YDQ.4[df$YDQ.4 == 2] <- 0
df$YDQ.5[df$YDQ.5 == 2] <- 0
df$YDQ.6[df$YDQ.6 == 2] <- 0
df$YDQ.7[df$YDQ.7 == 2] <- 0
df$YDQ.8[df$YDQ.8 == 2] <- 0

df$ICQ_Ctan <- (df$ICQ_01_t-1)+(df$ICQ_07_t-1)+(df$ICQ_13_t-1)+(df$ICQ_19_t-1)+(df$ICQ_25_t-1)
df$ICQ_Att <- (df$ICQ_02_a-1)+(df$ICQ_08_a-1)+(df$ICQ_14_a-1)+(df$ICQ_20_a-1)+(df$ICQ_26_a-1)
df$ICQ_Sens <- (df$ICQ_03_se-1)+(df$ICQ_09_se-1)+(df$ICQ_15_se-1)+(df$ICQ_21_se-1)+(df$ICQ_27_se-1)
df$ICQ_Demand <- (df$ICQ_04_d-1)+(df$ICQ_10_d-1)+(df$ICQ_16_d-1)+(df$ICQ_22_d-1)+(df$ICQ_28_d-1)
df$ICQ_SocAvoid <- (df$ICQ_05_so-1)+(df$ICQ_11_so-1)+(df$ICQ_17_so-1)+(df$ICQ_23_so-1)+(df$ICQ_29_so-1)
df$ICQ_CogAvoid <- (df$ICQ_06_c-1)+(df$ICQ_12_c-1)+(df$ICQ_18_c-1)+(df$ICQ_24_c-1)+(df$ICQ_30_c-1)
df$ICQ_TOTAL <- df$ICQ_Ctan+df$ICQ_Att+df$ICQ_Sens+df$ICQ_Demand+df$ICQ_SocAvoid+df$ICQ_CogAvoid
df$YDQ <- df$YDQ.1+df$YDQ.2+df$YDQ.3+df$YDQ.4+df$YDQ.5+df$YDQ.6+df$YDQ.7+df$YDQ.8

# Create a new attribute "YDQ_bin" based on the condition df$YDQ >= 5
df$YDQ_bin <- ifelse(df$YDQ >= 5, 1, 0)

> summary(df1)
      IU_W2R      ICQ_TOTAL      ICQ_Ctan      ICQ_Att      ICQ_Sens      ICQ_Demand
Min.   :1      Min.   : 25.00      Min.   : 5.00      Min.   : 0.00      Min.   : 2.00      Min.   : 0.00
1st Qu.:1      1st Qu.: 74.50      1st Qu.:12.00      1st Qu.: 8.00      1st Qu.:13.00      1st Qu.:13.00
Median :1      Median : 90.00      Median :14.00      Median :13.00      Median :16.00      Median :17.00
Mean   :1      Mean   : 89.68      Mean   :14.35      Mean   :13.26      Mean   :15.98      Mean   :16.79
3rd Qu.:1      3rd Qu.:107.00      3rd Qu.:17.00      3rd Qu.:18.00      3rd Qu.:19.00      3rd Qu.:21.00
Max.   :1      Max.   :180.00      Max.   :30.00      Max.   :30.00      Max.   :30.00      Max.   :30.00
      ICQ_SocAvoid      ICQ_CogAvoid      YDQ      YDQ_bin      IU_W2R_F
Min.   : 0.00      Min.   : 0.00      Min.   :0.000      Min.   :0.0000      Length:351
1st Qu.: 8.00      1st Qu.:12.00      1st Qu.:2.000      1st Qu.:0.0000      Class :character
Median :12.00      Median :17.00      Median :3.000      Median :0.0000      Mode  :character
Mean   :12.51      Mean   :16.79      Mean   :3.493      Mean   :0.3048
3rd Qu.:17.00      3rd Qu.:22.00      3rd Qu.:5.000      3rd Qu.:1.0000
Max.   :30.00      Max.   :30.00      Max.   :8.000      Max.   :1.0000
```



Coláiste na Tríonóide, Baile Átha Cliath Trinity College Dublin

Ollscoil Átha Cliath | The University of Dublin

```
> summary(df2)
```

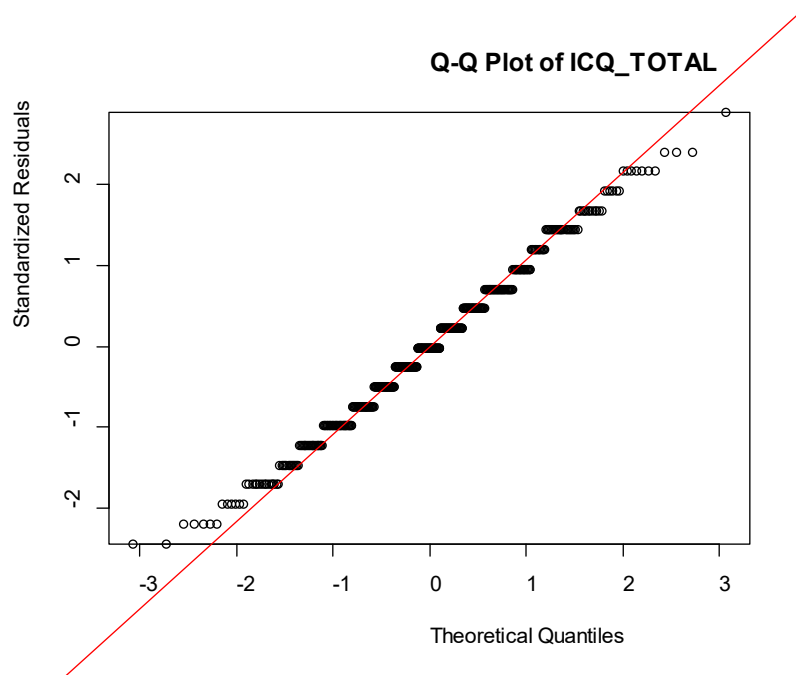
IU_W2R	ICQ_TOTAL	ICQ_Ctan	ICQ_Att	ICQ_Sens	ICQ_Demand
Min. :2	Min. : 14.00	Min. : 5.00	Min. : 0.00	Min. : 4.00	Min. : 0.00
1st Qu.:2	1st Qu.: 65.00	1st Qu.:12.00	1st Qu.: 9.00	1st Qu.:12.00	1st Qu.: 9.00
Median :2	Median : 83.00	Median :15.00	Median :14.00	Median :16.00	Median :14.00
Mean :2	Mean : 82.57	Mean :15.07	Mean :13.61	Mean :15.77	Mean :13.76
3rd Qu.:2	3rd Qu.:101.00	3rd Qu.:18.00	3rd Qu.:18.00	3rd Qu.:19.00	3rd Qu.:18.00
Max. :2	Max. :160.00	Max. :27.00	Max. :30.00	Max. :30.00	Max. :30.00

ICQ_SocAvoid	ICQ_CogAvoid	YDQ	YDQ_bin	IU_W2R_F
Min. : 0.00	Min. : 0.00	Min. :0.000	Min. :0.0000	Length:468
1st Qu.: 5.00	1st Qu.:10.00	1st Qu.:1.000	1st Qu.:0.0000	Class :character
Median :10.00	Median :14.00	Median :2.000	Median :0.0000	Mode :character
Mean :10.29	Mean :14.07	Mean :1.936	Mean :0.1047	
3rd Qu.:15.00	3rd Qu.:19.00	3rd Qu.:3.000	3rd Qu.:0.0000	
Max. :29.00	Max. :30.00	Max. :8.000	Max. :1.0000	



Q-Q Plot

Generated the the following code to be able to visuaise the Q_Q plots as follows, didn't run them for all the other attrributes, just showing the ICQ_TOTAL one as a reference

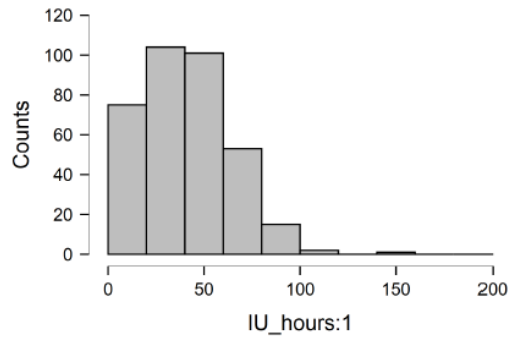




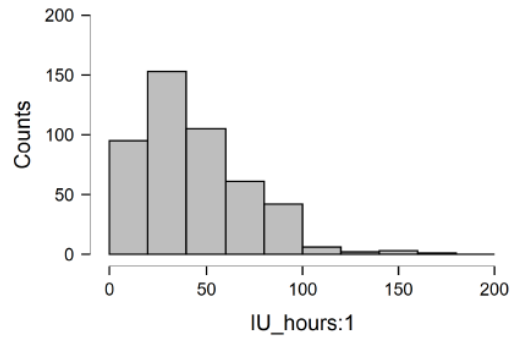
Distribution Plots

IU_hours:1

1

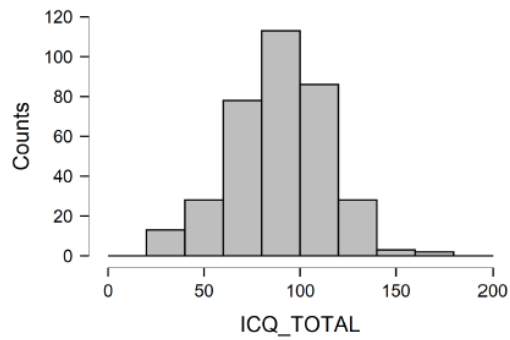


2

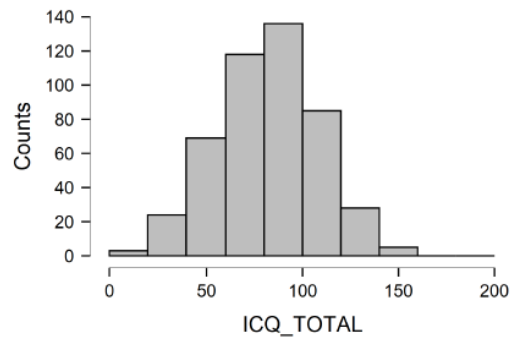


ICQ_TOTAL

1

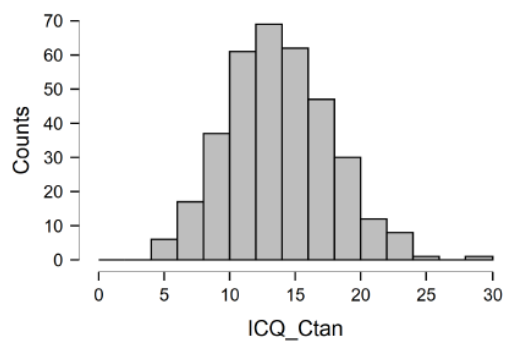


2

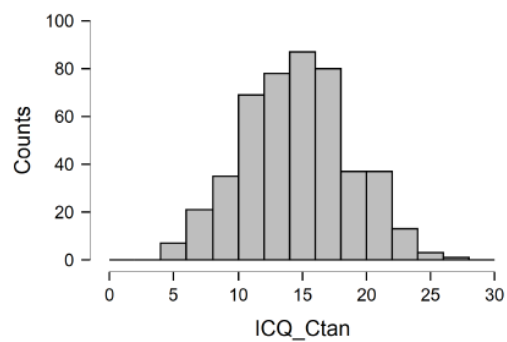


ICQ_Ctan

1



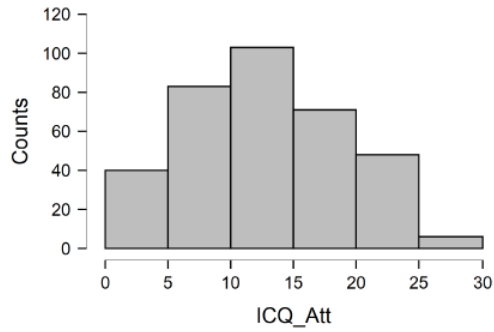
2



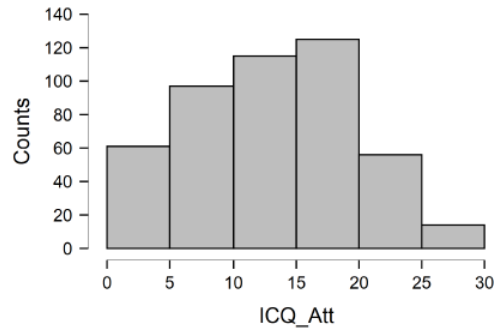


ICQ_Att

1

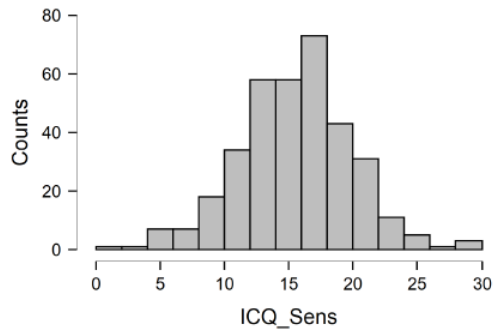


2

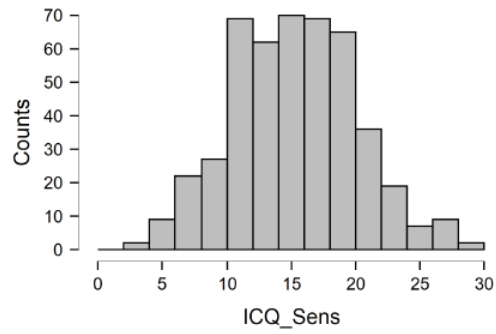


ICQ_Sens

1

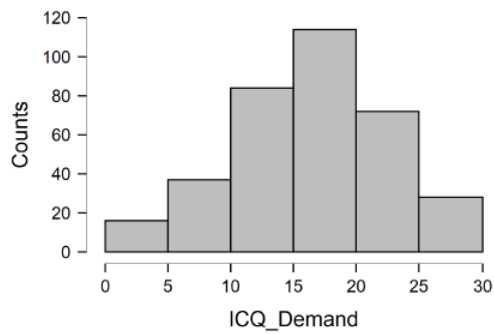


2

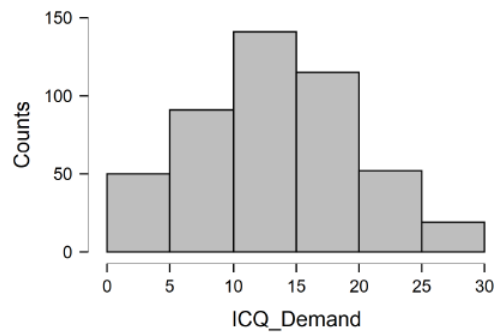


ICQ_Demand

1



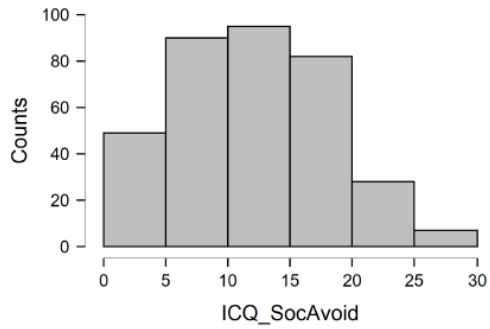
2



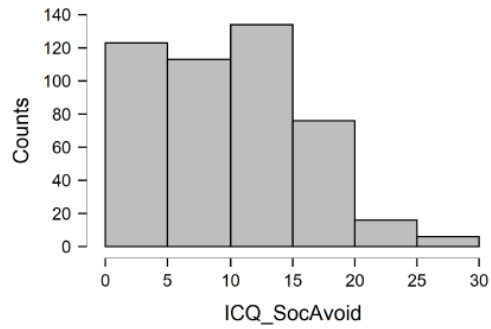


ICQ_SocAvoid

1

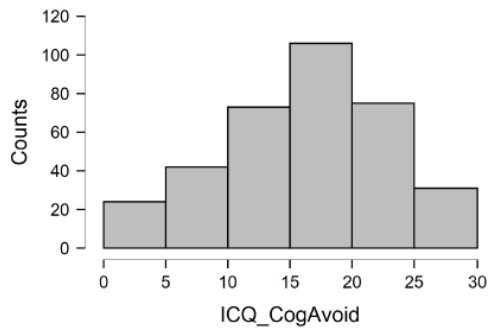


2

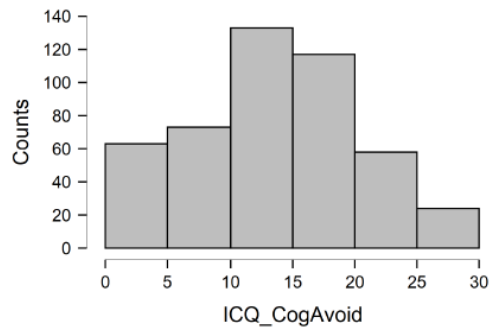


ICQ_CogAvoid

1

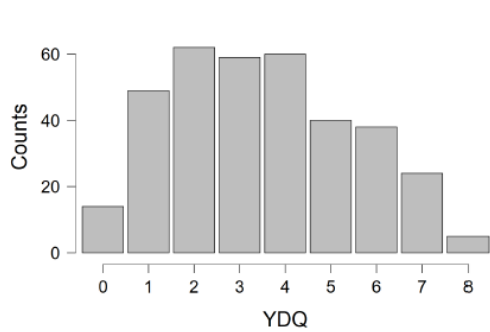


2

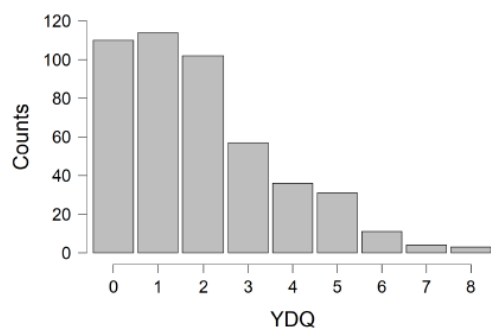


YDQ

1



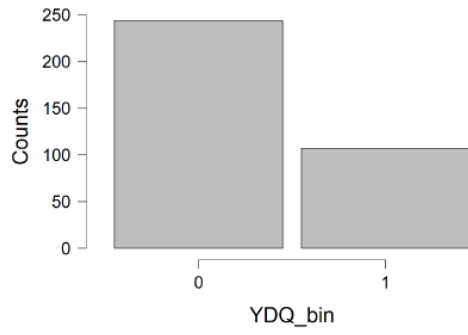
2



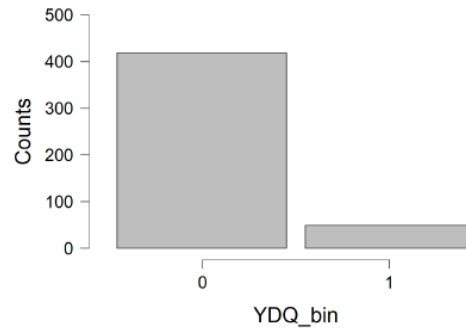


YDQ_bin

1



2





Correlations:

Finding: The article mentions a correlation of 0.39, but there is actually a negative correlation between IU_W2R and YDQ. However, page 12 of the ICQP_W2R_jasp_output.pdf file shows the correct correlations. Adding the calculated correlations below:

```
> # Print the results
> print("Pearson's correlation:")
[1] "Pearson's correlation:"
> print(pearson_r)
      cor
-0.3854862
> print("p-value:")
[1] "p-value:"
> print(pearson_p_value)
[1] 2.071938e-30
> print("Upper 95% CI:")
[1] "Upper 95% CI:"
> print(pearson_ci[1,2])
[1] -0.3854862
> print("Lower 95% CI:")
[1] "Lower 95% CI:"
> print(pearson_ci[1,1])
[1] -0.4423108
>
> print("Spearman's correlation:")
[1] "Spearman's correlation:"
> print(spearman_rho)
      rho
-0.3907634
> print("p-value:")
[1] "p-value:"
> print(spearman_p_value)
[1] 2.85051e-31
> print("Upper 95% CI:")
[1] "Upper 95% CI:"
> print(spearman_ci[1,2])
[1] -0.3907634
> print("Lower 95% CI:")
[1] "Lower 95% CI:"
> print(spearman_ci[1,1])
[1] -0.4472946
> |
```



Chi-square test for the Hypothesis:

According to the results of the Chi-Square test, they found strong evidence of an association between the desire to reduce internet use and problematic internet use as measured by the Young Diagnostic Questionnaire. The p-value obtained ($5.262e^{-13}$) is significantly lower than the threshold they set (0.05), suggesting a strong relationship between the variables.

The odds ratio indicates that there is a negative relationship between the desire to reduce internet use and problematic internet use. Specifically, problematic internet users are less likely to want to reduce their internet use compared to nonproblematic internet users. The confidence intervals further support this finding by indicating that the true odds ratio falls within the specified range.

Overall, these results suggest that there is a significant relationship between the desire to reduce internet use and problematic internet use, with problematic internet users being less motivated to reduce their internet use compared to nonproblematic internet users.

```
# Perform the chi-squared test
chi_squared <- chisq.test(contingency_table)
chi_squared_corrected <- chisq.test(contingency_table, correct = TRUE)
likelihood_ratio <- chisq.test(contingency_table, correct = FALSE, simulate.p.value = TRUE, B = 1000)

# Print the test statistics, degrees of freedom, p-values, and estimated p-values
print("Chi-squared Test:")
print(chi_squared$statistic)
print(chi_squared$parameter)
print(chi_squared$p.value)
print(chi_squared$expected)

print("Chi-squared Test (with continuity correction):")
print(chi_squared_corrected$statistic)
print(chi_squared_corrected$parameter)
print(chi_squared_corrected$p.value) |
print(chi_squared_corrected$expected)

print("Likelihood Ratio Test:")
print(likelihood_ratio$statistic)
print(likelihood_ratio$parameter)
print(likelihood_ratio$p.value)
print(likelihood_ratio$estimate)

# Calculate the probabilities
p1 <- contingency_table[1, 1] / sum(contingency_table[1, ])
p2 <- contingency_table[2, 1] / sum(contingency_table[2, ])

# Calculate the log odds ratio
log_odds_ratio <- log(p1 / (1 - p1)) - log(p2 / (1 - p2))

# Print the log odds ratio
print(log_odds_ratio)

# Fisher's exact test (ICQP_W2R_jasp_output.pdf page 13 replication)
fisher_test <- fisher.test(contingency_table)

# Print the results
print(fisher_test)

# Extract the odds ratio from Fisher's test result
odds_ratio <- fisher_test$estimate

# Calculate the log odds ratio
log_odds_ratio <- log(odds_ratio)
```



Chi-Square Hypothesis 1

Contingency Tables

YDQ_bin		IU_W2R:1		Total
		1	2	
0	Count	244.000	419.000	663.000
	% within column	69.516 %	89.530 %	80.952 %
1	Count	107.000	49.000	156.000
	% within column	30.484 %	10.470 %	19.048 %
Total	Count	351.000	468.000	819.000
	% within column	100.000 %	100.000 %	100.000 %

Chi-Squared Tests

	Value	df	p	VS-MPR*
X ²	52.105	1	5.262e-13	2.473e+10
X ² continuity correction	50.815	1	1.015e-12	1.312e+10
Likelihood ratio	52.062	1	5.377e-13	2.422e+10
N	819			

* Vovk-Sellke Maximum p -Ratio: Based the p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e \cdot p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Log Odds Ratio

	Log Odds Ratio	95% Confidence Intervals		p
		Lower	Upper	
Odds ratio	-1.322	-1.695	-0.949	
Fisher's exact test	-1.320	-1.716	-0.934	8.724e-13

Nominal

	Value
Contingency coefficient	0.245
Phi-coefficient	0.252
Cramer's V	0.252



My Contribution:

As an alternative analysis, I wanted to generate a logistic regression model to examine the impact of all the other variables on the desired reduction in internet use.

Logistic regression model uses the binomial family, indicating that the response variable IU_W2R is binary.

Interpreting the coefficients:

- Intercept: The estimated intercept is -0.09571. However, it is not statistically significant (p-value = 0.8319).
- ICQ variables are not significantly correlated based on the model results. The article was focus on the YDQ score which is also highly correlated based on the model results.
- YDQ: For a one-unit increase in YDQ, the estimated log odds ratio increases by 0.63131, and it is highly statistically significant (p-value < 0.001).
- YDQ_bin: For individuals in the YDQ_bin category (compared to those in the reference category), the estimated log odds ratio decreases by 0.65888, but it is marginally significant.

Confusion Matrix: The confusion matrix shows the predicted labels (0 or 1) compared to the actual labels for the test set. The matrix indicates that there are 115 true negatives (TN), 34 false positives (FP), 39 false negatives (FN), and 58 true positives (TP).

Coefficients: (1 not defined because of singularities)

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.09571	0.45083	-0.212	0.8319
ICQ_TOTAL	-0.02782	0.02580	-1.078	0.2809
ICQ_Ctan	-0.01914	0.03687	-0.519	0.6036
ICQ_Att	0.02260	0.02914	0.776	0.4379
ICQ_Sens	-0.06285	0.04355	-1.443	0.1490
ICQ_Demand	0.08437	0.04052	2.082	0.0373 *
ICQ_SocAvoid	0.03426	0.04366	0.785	0.4327
ICQ_CogAvoid	NA	NA	NA	NA
YDQ	0.63131	0.09556	6.606	3.94e-11 ***
YDQ_bin	-0.65888	0.37785	-1.744	0.0812 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



```
> print(confusion_matrix)
predicted_labels
    0    1
0 115   34
1   39   58
```

