SAMPLE RStudio lab assignment: logistic regression

[How to submit an assignment]

[Code templates in R]

[Interpretation templates]

[Variables in GSS]

Note: Do not run or interpret the analyses without opening [Variables in GSS] file, using "What it measures" columns, reading variable type, how the questions were asked, and the response sets.

Assignment instructions

In this assignment, you will work on two questions.

- In the first question, you will create the dummy variables that you will be using for this assignment.
- In the second question, you will run a logistic regression model.

Reminder: Ensure the highlighted parts are the same within the model.

model1 ➜ model1

```
model1 <- lm(conrinc ~ god, data = gs
tab_model(model1, show.std = T, show.
```

Questions

1) Create all the dummy variables that you will be using for this assignment.

1a) Dependent variable: Using the original variable of **cohabok,** create two dummy variables.

- The name of the first dummy variable: **opposingcohabitation**
    - Merge “disagree” and “strongly disagree” categories
    - What it measures: “opposing cohabitation**”**
    - This dummy variable is your dependent variable.
- The name of the second dummy variable: **favoringcohabitation**
    - Merge “neither agree nor disagree”, “agree”, “strongly agree categories

Code (-5):	*gss$opposingcohabitation <- ifelse(gss$cohabok == 4 | gss$cohabok == 5, 1, 0)*
Code (-5):	*gss$favoringcohabitation <- ifelse(gss$cohabok == 1 | gss$cohabok == 2 \| gss$cohabok == 3, 1, 0)*

1b) Independent dummy variable 1: Using the original variable of **marital,** create two dummy variables.

- The name of the first dummy variable: **married**
    - Use the “married” category
    - What it measures: “being married”
    - Use this dummy variable as an independent variable.
- The name of the second dummy variable: **othermarital**
    - Merge “widowed”, “divorced”, “separated”, and “never married” categories
    - What it measures: “being non-married”
    - This is a comparison category. You will omit this from the model.

Code (-5):	*gss$married <- ifelse(gss$marital == 1, 1, 0)*
Code (-5):	*gss$othermarital <- ifelse(gss$marital == 2 | gss$marital == 3 \| gss$marital == 4 | gss$marital == 5, 1, 0)*

1c) Independent dummy variable 2: Using the original variable of **sex,** create two dummy variables.

- The name of the first dummy variable: **male**
    - Use the “male” category
    - What it measures: “being male”
    - Use this dummy variable as an independent variable.
- The name of the second dummy variable: **female**
    - Use the “female” category
    - What it measures: “being female”
    - This is a comparison category. You will omit this from the model.

Code (-5):	*gss$male <- ifelse(gss$sex == 1, 1, 0)*
Code (-5):	*gss$female <- ifelse(gss$sex == 2, 1, 0)*

2) Run a logistic regression model using:

dependent variable: **opposingcohabitation**
independent variable 1: **married**
independent variable 2: **male**
independent variable 3: **educ**
independent variable 4: **age**

Finally interpret the model 1 **[interpretation: 80 points]**.

Logistic regression analysis interpretation breakdown:

First paragraph: [The significance levels] Mention which variables (“what it measures”) are statistically significant, and which variables are statistically insignificant. Variables with at least one asterisk (*) are statistically significant **[10 points].** (logistic regression interpretation breakdown template: first paragraph)

Second paragraph: [The explanation of odd ratios] Mention how independent variables increase or decrease the probability of the dependent variable happening, using the "Odd ratios" column. When reporting the Odd ratios, ensure that the sentence includes the units (one unit, score, year, dollars, etc.) of the continuous independent variables. When reporting the Odd ratios of the dummy independent variables, ensure that the sentence mentions the comparison category. When reporting the negative Odd ratios of the dummy independent variables, make sure to divide 1 by the Odd ratios **[40 points].** (logistic regression interpretation breakdown template: second paragraph**) AND** (reporting the odd ratios of continuous and dummy independent variables) **AND** (reporting the negative odd ratios and Std.Betas)

Third paragraph: [The explanation of standardized betas (std.Beta column)] Mention the strongest predictors (variables) of the dependent variable using the "std.Beta" (standardized beta) column in order. When reporting the negative standardized betas of the dummy independent variables, make sure to divide 1 by the standardized betas **[20 points].** (logistic regression interpretation breakdown template: third paragraph) **AND** (reporting the negative odd ratios and Std.Betas)

Fourth paragraph: [The explanation of Tjur R-Squared value] Mention the Tjur R-Squared as a percentage with the statistically significant variables **[10 points]** (logistic regression interpretation breakdown template: fourth paragraph**) AND** (reporting of Tjur R-squared)

Summary of the variables to be used in the model and interpretation

Variable	Type	Function	What it measures	Comparison category
opposingcohabitation	dummy	dependent	opposing cohabitation	No comparison category for dependent variables
married	dummy	independent	being married	being non-married
male	dummy	independent	being male	being female
educ	continuous	independent	respondents' education in years	No comparison category for continuous
age	continuous	independent	respondents' age in years	No comparison category for continuous

Code (-5):

```
model4 <- glm(opposingcohabitation ~ married + male + educ + age, data = gss, family = binomial(link="logit"))
tab_model(model4, show.std = TRUE, show.ci = FALSE, collapse.se = TRUE, p.style = "stars")
```

Table (20 points):

	opposingcohabitation	
Predictors	*Odds Ratios*	*std. Beta*
(Intercept)	0.14 *** (0.07)	0.19 (0.02)
married	2.00 *** (0.34)	1.41 (0.12)
male	0.69 * (0.12)	0.83 (0.07)
highest year of school completed	0.97 (0.03)	0.92 (0.08)
age of respondent	1.01 ** (0.00)	1.28 (0.11)
Observations	1055	
R^2 Tjur	0.032	

* *p*<0.05 ** *p*<0.01 *** *p*<0.001

Interpretation (80 points):

***First paragraph**:*

Being married, being male, and respondents' age in years are significant predictors of opposing cohabitation since the p values are less than 0.05. respondents' education in years is not a significant predictor of opposing cohabitation since the p value is less than 0.05.

***Second paragraph**:*

Being married increases the probability of opposing cohabitation by 2 times compared to being non-married. Being male decreases the probability of opposing cohabitation by 1.44 times compared to being female. A year increase in respondents' age in years increases the probability of opposing cohabitation by 1.01 times.

***Third paragraph**:*

The strongest predictor of opposing cohabitation is Being married (std.Beta=1.41), followed by respondents' age in years (std.Beta=1.28), and being male (std.Beta=1.20).

	***Fourth paragraph**:* The Tjur R-squared value indicates that 3.2% of the variation in opposing cohabitation can be explained by being married, being male, and respondents' age in years.