SAMPLE RStudio lab assignment: dummy variables

[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 be creating dummy variables and adding them to the model you generated during the RStudio lab assignment linear regression basics assignment.

First, review any feedback given on “RStudio lab assignment linear regression basics” and revise your model and/or interpretations accordingly. If there were any issues with your codes and/or interpretation, I provided the correct codes and/or interpretation as feedback.

After completing this step, you can begin creating dummy variables to use in your model.

Questions

1) Paste the code, table, and interpretation of the fourth model (the last model) from the RStudio lab assignment linear regression basics (revise them first, if you received a feedback) **[10 points]**:

Code (-5):	*model4 <- lm(sei10 ~ childs + educ + prestg10 + lifenow, data = gss)* *tab_model(model4, show.std = T, show.ci = F, collapse.se = T, p.style = "stars")*
Table (5 points):	

	r's socioeconomic index(2010)	
Predictors	*Estimates*	*std. Beta*
(Intercept)	-37.74 *** (2.07)	0.00 (0.01)
number of children	-0.11 (0.20)	-0.01 (0.01)
highest year of school completed	1.93 *** (0.12)	0.23 (0.01)
r's occupational prestige score(2010)	1.21 *** (0.02)	0.71 (0.01)
r's rating of life overall now from 0-10	0.65 ** (0.20)	0.04 (0.01)
Observations	1732	
R^2 / R^2 adjusted	0.719 / 0.718	

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

Interpretation (5 points):

First paragraph:

Respondents' education in years, occupational prestige score, and quality of life rating are statistically significant predictors of socio-economic index score since the p values are less than 0.05. The number of children respondents have is not a statistically significant predictor of socio-economic index score since the p value is greater than 0.05.

Second paragraph:

A year increase in respondents' education increases respondents' socio-economic index score by 1.93 points. One unit increase in

	occupational prestige score increases respondents' socio-economic index score by 1.21 points. One unit increase in the quality of life rating increases socio-economic index score by 0.65 points. **Third paragraph**: The strongest predictor of respondents' socio-economic index score is occupational prestige score (std.Beta=0.71), followed by respondents' education in years (std.Beta=0.23), and quality of life rating (std.Beta=0.04). **Fourth paragraph**: The adjusted R squared value indicates that 71.8% of the variation in respondents' socio-economic index score can be explained by the occupational prestige score, respondents' education in years, and quality of life rating.

2a) Create a frequency table of **compuse.** Paste the frequency table **(5 points)** and interpret the table **(5 points)**.

Code (-5):	*frq(gss$compuse, out = "v")*
Table:	**r use computer (x)** <*numeric*> *val* | *label* | *frq* | *raw.prc* | *valid.prc* | *cum.prc* 1 | yes | 1932 | 54.51 | 81.80 | 81.80 2 | no | 430 | 12.13 | 18.20 | 100.00 NA | NA | 1182 | 33.35 | NA | NA *total N=3544 · valid N=2362 · x̄=1.18 · σ=0.39*
Interpretation:	Tthe use of computer variable shows that 81.80% of the respondents use computer and 18.20% of the respondents do not use computer.

2b) Using the original variable of **compuse,** create two dummy variables.

- The name of the first dummy variable: **usecomputeratwork**
- The name of the second dummy variable: **nousecomputeratwork**

Code (-5):	*gss$usecomputeratwork <- ifelse(gss$compuse == 1, 1, 0)*

Code (-5):	*gss$nousecomputeratwork <- ifelse(gss$compuse == 2, 1, 0)*

3a) Create a frequency table of **satfin**. Paste the frequency table **(5 points)** and interpret the table **(5 points)**.

Code (-5):	*frq(gss$satfin, out = "v")*
Table:	**satisfaction with financial situation (x)** *<numeric>* *val* | *label* | *frq* | *raw.prc* | *valid.prc* | *cum.prc* 1 | pretty well satisfied | 877 | 24.75 | 24.87 | 24.87 2 | more or less satisfied | 1568 | 44.24 | 44.47 | 69.34 3 | not satisfied at all | 1081 | 30.50 | 30.66 | 100.00 NA | NA | 18 | 0.51 | NA | NA *total N=3544 · valid N=3526 · x̄=2.06 · σ=0.74*
Interpretation:	The level of financial satisfaction variable shows that 24.87% of the respondents are pretty well satisfied; 44.47% of the respondents are more or less satisfied; 30.66% of the respondents are not satisfied at all.

3b) Using the original variable of **satfin,** create three dummy variables.

- The name of the first dummy variable: **prettywellsatisfied**
- The name of the second dummy variable: **moreorlesssatisfied**
- The name of the third dummy variable: **notsatisfiedatall**

Code (-5):	*gss$prettywellsatisfied <- ifelse(gss$satfin == 1, 1, 0)*
Code (-5):	*gss$moreorlesssatisfied <- ifelse(gss$satfin == 2, 1, 0)*
Code (-5):	*gss$notsatisfiedatall <- ifelse(gss$satfin == 3, 1, 0)*

4) Run the fifth model. The fifth model should include all the variables used in the fourth model of RStudio lab assignment linear regression basics (the code section of the first question of this assignment) + the dummy variables.

dependent variable: **sei10**
independent variable 1: **childs**
independent variable 2: **educ**
independent variable 3: **prestg10**
independent variable 4: **lifenow**
dummy variable 1: **usecomputeratwork** (reference category: nousecomputeratwork)
dummy variable 2: **prettywellsatisfied** (reference category: notsatisfiedatall)
dummy variable 3: **moreorlesssatisfied** (reference category: notsatisfiedatall)

Paste the fifth model linear regression model table **[20 points]**.

Finally interpret the model 5 **[50 points]**.

Linear regression analysis interpretation breakdown:

Make sure to use "Reporting of dummy variable estimates (coefficients)" information when interpreting dummy variables.

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].** (use linear regression interpretation breakdown template: first paragraph)

Second paragraph: [The explanation of coefficients (Estimates column)] Mention how independent variables increase or decrease the value of the dependent variable, using the "Estimates" column. When reporting the estimates (coefficients), ensure that the sentence includes the units (one unit, score, year, dollars, etc.) of both the independent and the dependent variable **[20 points].** (use linear regression interpretation breakdown template: second paragraph **AND** reporting of estimates (coefficients))

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. Only mention the statistically significant ones. "std.Beta" is an absolute number,

which means, for example, -.56 is stronger than .45. **[15 points].** (use linear regression interpretation breakdown template: third paragraph)

Fourth paragraph: The explanation of Adjusted R-squared value (R2 adjusted) **[5 points]** (use linear regression interpretation breakdown template: fourth paragraph) **AND** (reporting of adjusted R-squared)

Code (-5):	*model5 <- lm(sei10 ~ childs + educ + prestg10 + lifenow + usecomputeratwork + prettywellsatisfied + moreorlesssatisfied, data = gss)* *tab_model(model5, show.std = T, show.ci = F, collapse.se = T, p.style = "stars")*
Table:	

	r's socioeconomic index(2010)	
Predictors	*Estimates*	*std. Beta*
(Intercept)	-38.38 *** (2.61)	-0.00 (0.02)
number of children	0.02 (0.24)	0.00 (0.02)
highest year of school completed	1.89 *** (0.15)	0.22 (0.02)
r's occupational prestige score(2010)	1.17 *** (0.03)	0.69 (0.02)
r's rating of life overall now from 0-10	0.30 (0.25)	0.02 (0.02)
usecomputeratwork	4.82 *** (1.19)	0.06 (0.02)
prettywellsatisfied	3.23 ** (1.06)	0.06 (0.02)
moreorlesssatisfied	0.66 (0.85)	0.01 (0.02)
Observations	1203	
R^2 / R^2 adjusted	0.725 / 0.723	

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

Interpretation:	**First paragraph**: Respondents' education in years, occupational prestige score, using computer at work, and being pretty well satisfied with finances are statistically significant predictors of socio-economic index score since the p values are less than 0.05. The number of children respondents have, the quality of life rating, and being more or less satisfied with finances are not statistically significant predictors of socio-economic index score since the p values are greater than 0.05. **Second paragraph**: A year increase in respondents' education increases respondents' socio-economic index score by 1.89 points. One unit increase in occupational prestige score increases respondents' socio-economic index score by 1.17 points. Using a computer at work increases respondents' socio-economic index score by 4.82 points compared to respondents who do not use computer at work. Being pretty well satisfied with finances increases respondents' socio-economic index score by 3.23 points compared to respondents who are not satisfied at all with their finances. **Third paragraph**: The strongest predictor of respondents' socio-economic index score is occupational prestige score (std.Beta=0.69), followed by respondents' education in years (std.Beta=0.22), using a computer at work (std.Beta=0.06) and being pretty well satisfied with finances (std.Beta=0.06). **Fourth paragraph**: The adjusted R squared value indicates that 72.3% of the variation in respondents' socio-economic index score can be explained by the occupational prestige score, respondents' education in years, using a computer at work, and being pretty well satisfied with finances.