

Homework 3

Winter 2024

Question 1

Part a)

Using dummy codes, run a regression to assess whether *depress2001* differs by marriage status (*married01*). Use “Married” as the reference group. Create a table to display the results of the analysis (Hint: try using the `bind_rows()`, `tidy()`, and `kable`, or `stargazer`’ functions. You could also use any other functions that you would like, as long as the table is clean and clear). Then, in the space below, write the regression equation and interpret the overall model (was there a significant effect of *married01* on *depress2001*? Cite the appropriate statistics) and the intercept.

```
arh_q1 <- arh |>
  dplyr::mutate(divorced_v_married = ifelse(married01 == "Divorced", 1, 0),
               nevmarried_v_married = ifelse(married01 == "Never Married", 1, 0),
               separated_v_married = ifelse(married01 == "Separated", 1, 0),
               widowed_v_married = ifelse(married01 == "Widowed", 1, 0),
               )

mod1a <- lm(depress2001 ~ divorced_v_married + nevmarried_v_married + separated_v_married + widowed_v_married)
stargazer(mod1a, type = "latex", header = FALSE, ci=TRUE, digits=2, title = "Depression predicted by Mar")
```

$$\widehat{\text{depress2001}}_i = 11.39 + 1\text{Divorced}_i + 1.68\text{NevMarr}_i + 0.012\text{Separated}_i + 1.38\text{Widowed}_i$$

Interpretations:

- **Overall model:** The overall model is statistically significant ($F(4, 817) = 4.0, p = 0.003$), meaning that marriage status has an effect on Depression. However, the model only explains a small amount of the variance in the dependent variable ($R^2_{adj}=0.14$).
- **Intercept:** The coefficient for the intercept is 11.39 and statistically significant ($p<.001$), which means that the predicted value of Depression for a person who is Married is 11.39.

Part b)

Based on the regression performed above in Part A, which groups can you conclude are significantly different (at $p < .05$)? List each specific pair of groups that differed.

Answer:

- Widowed vs. Married: $b = 1.38, p < 0.001$.

Table 1: Depression predicted by Marriage Status.

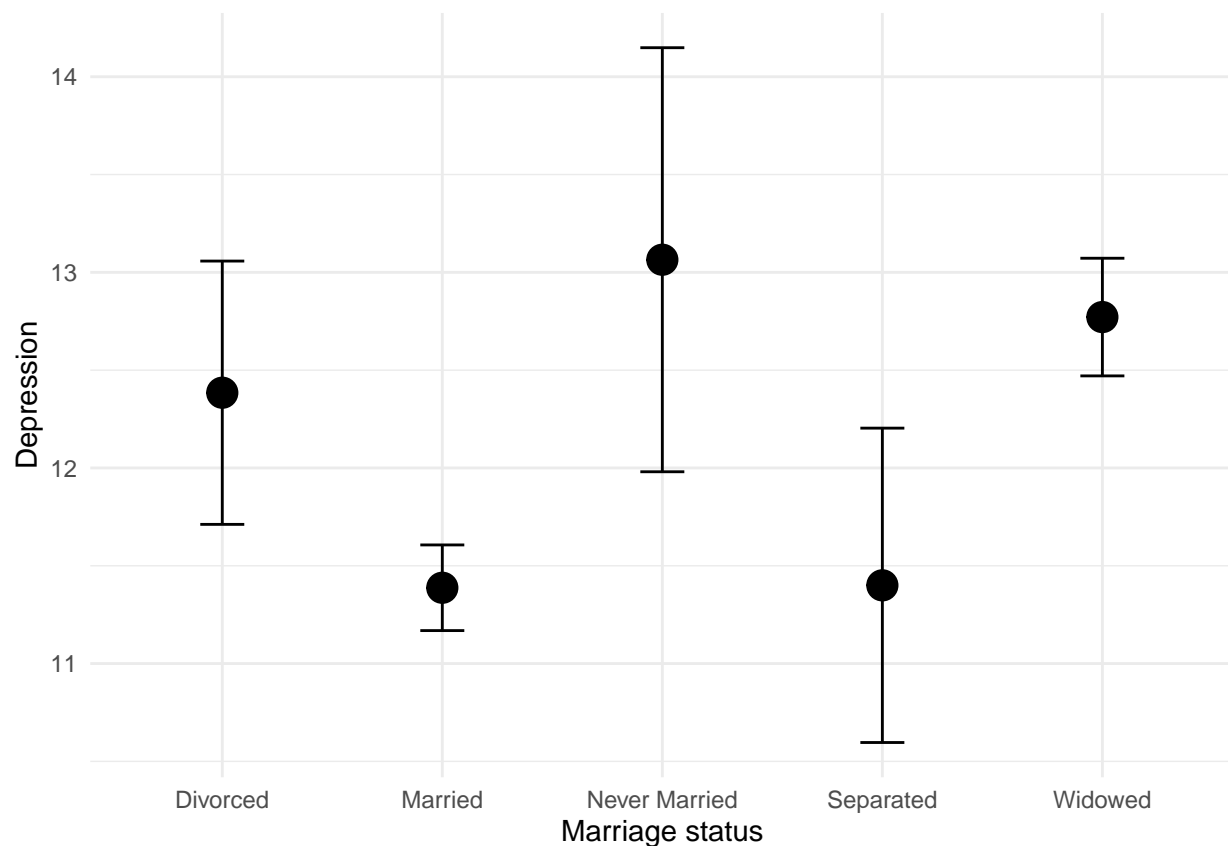
	<i>Dependent variable:</i>
	depress2001
divorced_v_married	1.00 (−0.28, 2.27)
nevmarried_v_married	1.68* (−0.10, 3.46)
separated_v_married	0.01 (−2.50, 2.53)
widowed_v_married	1.38*** (0.65, 2.11)
Constant	11.39*** (10.92, 11.86)
Observations	822
R ²	0.02
Adjusted R ²	0.01
Residual Std. Error	4.88 (df = 817)
F Statistic	3.99*** (df = 4; 817)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Part c)

Create a plot with *married01* on the x-axis and average levels of *depress2001* on the y-axis (i.e., using `geom_point()`). Include standard error bars for plus and minus 1 SE.

```
arh_mean <- arh |>
  dplyr::with_groups(
    married01,
    summarise,
    mean = mean(depress2001, na.rm = TRUE),
    n = length(depress2001),
    sd = sd(depress2001, na.rm = TRUE),
    se = sd / sqrt(n),
    upper = mean + se,
    lower = mean - se
  ) |>
  dplyr::filter(!is.na(married01))

arh_mean |>
  ggplot(aes(x = married01, y = mean)) +
  geom_point(size = 5) +
  geom_errorbar(aes(ymin = lower, ymax = upper), width = 0.2) +
  theme_minimal() +
  labs(x = "Marriage status", y = "Depression")
```



Question 2

Part a)

Repeat the regression from Question 1, Part A (assessing differences in *depress01* based on *married01* grouping), but this time use “Divorced” as the reference group. Create a table to display the results of the analysis (Hint: try using the `bind_rows()`, `tidy()`, and `kable` functions. You could also use any other functions that you would like, as long as the table is clean and clear).

```
arh_q2 <- arh |>
  dplyr::mutate(married_v_divorced = ifelse(married01 == "Married", 1, 0),
    nevmarried_v_divorced = ifelse(married01 == "Never Married", 1, 0),
    separated_v_divorced = ifelse(married01 == "Separated", 1, 0),
    widowed_v_divorced = ifelse(married01 == "Widowed", 1, 0),
  )

mod2a <- lm(depress2001 ~ married_v_divorced + nevmarried_v_divorced + separated_v_divorced + widowed_v_divorced)
stargazer(mod2a, type = "latex", header = FALSE, ci=TRUE, digits=2, title = "Depression predicted by Ma
```

Table 2: Depression predicted by Marriage Status.

	<i>Dependent variable:</i>
	depress2001
married_v_divorced	-1.00 (-2.27, 0.28)
nevmarried_v_divorced	0.68 (-1.41, 2.77)
separated_v_divorced	-0.98 (-3.73, 1.76)
widowed_v_divorced	0.39 (-0.93, 1.70)
Constant	12.38*** (11.20, 13.57)
Observations	822
R ²	0.02
Adjusted R ²	0.01
Residual Std. Error	4.88 (df = 817)
F Statistic	3.99*** (df = 4; 817)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

Part b)

Interpret the slope for the dummy code for “Never Married” in both regressions (the one with “Married” as the reference group, and the one with “Divorced” as the reference group).

Answer:

- **“Married” Reference Group:** In this mode, the slope for “Never Married” is statistically non-significant and positive ($b = 1.68$, 95% CI [-0.11, 3.46], $p = 0.065$). It represents the difference between the mean level of depression of the “Married” group and the mean depression of the “Never Married” group.
- **“Divorced” Reference Group:** In this mode, the slope for “Never Married” is statistically non-significant and positive ($b = 0.68$, 95% CI [-1.41, 2.77], $p = 0.524$). It represents the difference between the mean level of depression of the “Divorced” group and the mean depression of the “Never Married” group.

Part c)

Based on the analysis performed with “Divorced” as the reference group, which groups can you conclude are significantly different at $p < .05$?

Answer:

- WRITE ANSWER HERE

Question 3

Part a)

Using dummy codes, run a two-way factorial ANOVA model with *self_worth2001* as the outcome and *married01*, *smoke01*, and the interaction between these variables as predictors. For the *married01* variable, use “Married” as the reference group; for the *smoke01* variable, use “Non-Smoker” as the reference group. Display the results of the analysis in a table in the same way as above (i.e., with `tidy()` and `kabl()`).

Part b)

Interpret the intercept.

- **Answer:** WRITE ANSWER HERE

Part c)

Among the non-smoker group, which group(s) are significantly different from the reference group, at $p < .05$?

- **Answer:** WRITE ANSWER HERE

Part d)

Create a graph showing the means of *self_worth2001* (on the y axis) broken down by *married01* and *smoke01*, with ± 1 SE bars. Try different ways of graphing the data (e.g., try putting *married01* on the x axis and coloring by *smoke01*, then try it the other way around; you could also try faceting the graph by *married01* and *smoke01*), and use the combination that you think emphasizes the most interesting effects in the data. Include a brief explanation of why you chose the graphed the data the way you did. There is not necessarily one right way to do this, as long as you can justify your decision.

- **Explanation:** WRITE ANSWER HERE