

Poster Info and Help

Nicky Wakim

2025-06-04

Some notes from last quarter's posters (1/2)

- Can you all see my feedback from last quarter? Let me know if you cannot, but want to
- In methods, there is a **difference between coding procedure and statistical procedure**
 - Example of something we do NOT report: If we get a dataset where the categorical variables (with worded categories) are represented as integers, then we convert the integers to the categories.
 - This is a **coding procedure** that automatically needs to be done to represent the variable accurately.
 - This is **NOT a statistics procedure** and we don't report it. We're not "converting" a continuous variable to a categorical one. We are performing data management to make sure the data code represent the data.
- In conclusion/discussion, many people mentioned limitations but not how they affect the results
 - Or made a wrong conclusion about the consequences
 - See me to double check!!
- "Data" is a plural word!
 - Use like "Data are represented in Table..."

Some notes from last quarter's posters (2/2)

- Results

- No need to interpret the intercept ~~rules, you have an interaction~~
 - If purposeful model selection, do NOT skip the odds ratios interpretations!!
 - If you are limiting your Regression table to only the main variable, you need to explicitly state that this is from the final model and adjusted for other variables
- There was A LOT of talk about generalizability being an issue because of the spread of our data
 - Sample distribution ALONE does not mean our results are not generalizable
 - The more important question is: Are there enough people (in counts) in each group/demographic/etc to make a decent estimate?
 - Be really careful around language!!
 - Using Why or How in your title is a little misleading – we can only answer “does”
 - Make sure you do not make implications that we are uncovering a causal pathway!!
 - We are not!
 - The study design does NOT support this!

X_1	X_2
No	Yes
No	No
Yes	Yes
Yes	No

Things you can assume and not assume from your reader

- Things people assume you are doing (correctly) and you do not need to share with an audience
 - Following a codebook's categories
 - How you ran a certain model (aka don't need to say we used `glm()` to fit the logistic regression model)
 - You can assume people can read mathematical equations
 - How age is calculated
- Things you cannot assume people know
 - How to read R code: do not include your model equation as an `glm()` function

$$Y = \beta_0 + \dots$$

$$\text{lm}(Y \sim X_1 + X_2)$$

Some notes

- I will NOT be counting exact bullet points
 - If it says 2-3, anything less than 4 is okay
 - If it says 5-6, it shouldn't be 1 and it shouldn't be >10
- Do NOT make bullet points really long!
 - This was a big issue in last quarter's poster
 - Work to make things concise!! You are likely including too much information
- I tried the QMD poster, and it is REALLY hard to troubleshoot
 - I highly suggest making your poster in powerpoint!

Background

- **Length: 5-8 bullets**
- **Purpose:** Introduce the research question and why it is important to study
- This section is non-technical.
 - By reading just the introduction and conclusion, someone without a technical background should have an idea of what they study was about, why it is important, and what the main results are
- You may start with your bullets from Lab 1, but you should edit it and make sure it flows into your report well!
- Should contain some references

Methods

- **Length: 8-10 bullets**
- **Purpose:** Describe the analyses that were conducted and methods used to select variables and check diagnostics
- **Some important methods to discuss** (You may divide these into your sections, not necessarily with these names)
 - General approach to the dataset
 - Variables and variable creation
 - Model building: we performed purposeful selection
 - Final model
 - Model diagnostics

Methods example-ish

- Data were collected from ____ in ____ year
- We performed a complete case analysis on ____
- We generated categorized variables for CO2 emissions and income levels
 - CO2 emissions used quartiles
 - Income levels used the specified groupings by Gapminder
- We used purposeful model selection, a combination of field expertise and statistical methods, to determine the final model
- **Our final model included our outcome, life expectancy, with a main effect for female literacy rate while adjusting for confounders**
 - **We adjusted for: CO2 emissions, income levels, world region, access to improved water, food supply, and intergovernmental groups**
 - **We included an interaction between ____ and ____**
- We investigated model assumptions and diagnostics using ____
- We used R version 4.4.1 to analyze data

Results: What should be in your results?

If you did purposeful model selection:

1. Table 1
2. Table showing odds ratios for your main variable + variable from interaction
3. One additional figure or table to understand your research question
4. Interpretations of odds ratios from Table in #2

If you did LASSO:

1. Table 1
2. Table or plot showing most important variables
3. One additional figure or table to help understand your question
4. Interpretations of important variables: identify top 1-3 that predict outcome

1. Table 1 (table summary in Lesson 2)

2. Regression table or Forest plot (if no interactions)

```
1 reg_table = tbl_regression(  
2   final_model,  
3   label = list(  
4     FemaleLiteracyRate ~ "Female literacy rate (%)",  
5     CO2_q ~ "CO2 emissions quartiles",  
6     income_levels1 ~ "Income levels",  
7     four_regions ~ "World region",  
8     WaterSourcePrct ~ "Access to improved water (%)",  
9     FoodSupplykcalPPD ~ "Food supply (kcal PPD)",  
10    members_oecd_g77 ~ "Intergovernmental group"  
11  ) ) %>%  
12  as_gt() %>%  
13  tab_options(table.font.size = 20) %>%  
14  cols_width(label ~ px(250))  
15 reg_table
```

Characteristic	Beta	95% CI ¹	p-value
Female literacy rate (%)	-0.07	-0.17, 0.02	0.13
CO2 emissions quartiles			
Quartile 1	—	—	
Quartile 2	1.1	-2.7, 4.9	0.6
Quartile 3	-0.29	-5.1, 4.6	>0.9
Quartile 4	-0.60	-5.6, 4.5	0.8
Income levels			
Low income	—	—	
Lower middle income	5.4	0.75, 10	0.024
Upper middle income	6.1	0.20, 12	0.043
High income	8.0	1.4, 15	0.018
World region			
Africa	—	—	
Americas	9.0	4.9, 13	<0.001
Asia	5.3	2.0, 8.5	0.002
Europe	6.9	1.1, 13	0.020
Access to improved water (%)	0.17	0.03, 0.30	0.015
Food supply (kcal PPD)	0.00	0.00, 0.01	0.073
Intergovernmental group			
Group of 77	—	—	
OECD	1.1	-4.2, 6.5	0.7
Others	1.0	-4.0, 6.1	0.7

¹ CI = Confidence Interval

2. Regression table or Forest plot (if no interactions)

```
1 library(broom.helpers)
2 model_tidy = tidy_and_attach(final_model, conf.int=T) %>%
3   tidy_remove_intercept() %>%
4   tidy_add_reference_rows() %>% tidy_add_estimate_to_reference_rows() %>%
5   tidy_add_term_labels() %>%
6   mutate(var_label = case_match(var_label,
7     "Female literacy rate (%)" ~ "",
8     "Food supply (kcal PPD)" ~ " ",
9     "Access to improved water (%)" ~ "  ",
10    "CO2 emissions quartiles" ~ "CO2 emissions \n quartiles",
11    "Income levels" ~ "Income levels",
12    "Intergovernmental group" ~ "Intergovern- \n mental group",
13    .default = var_label
14  ))
15
16 forest_plot = ggplot(data=model_tidy, aes(y=label, x=estimate, xmin=conf.low, xmax=conf.high)) +
17   facet_grid(rows = vars(var_label), scales = "free",
18     space='free_y', switch = "y") +
19   geom_point(size = 3) + geom_errorbarh(height=.2) +
20   geom_vline(xintercept=0, color='#C2352F', linetype='dashed', alpha=1) +
21   theme_classic() +
22   labs(x = "Beta", y = "Variables") +
23   theme(axis.title = element_text(size = 20), axis.text = element_text(size = 20),
24     title = element_text(size = 20), strip.placement = "outside",
25     strip.text.y.left = element_text(size = 18, angle = 0),
26     strip.background = element_blank())
```

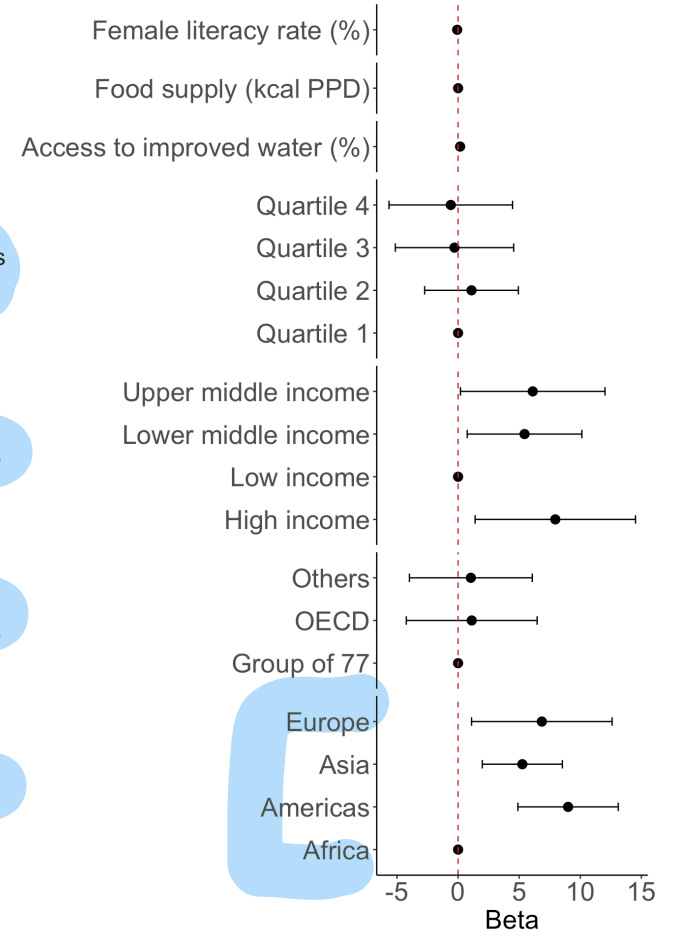
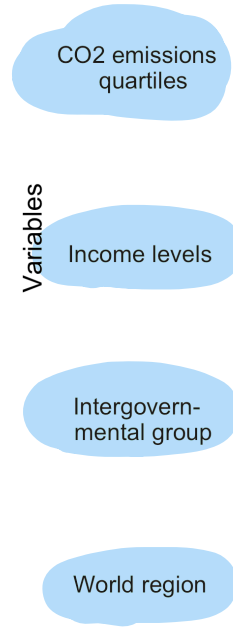
2. Regression table or Forest plot (if no interactions)

- This is a fun one to investigate!
- Stick to the regression table if you are having trouble with this!

```

1 library(broom.helpers)
2
3 model_tidy = tidy_and_attach(final_model, conf.int=
4   tidy_remove_intercept() %>%
5   tidy_add_reference_rows() %>% tidy_add_estimate_t
6   tidy_add_term_labels() %>%
7   mutate(var_label = case_match(var_label,
8     "Female literacy rat
9     "Food supply (kcal P
10    "Access to improved
11    "CO2 emissions quart
12    "Income levels" ~ "I
13    "Intergovernmental g
14    .default = var_label
15  ))
16
17 forest_plot = ggplot(data=model_tidy, aes(y=label,
18   facet_grid(rows = vars(var_label), scales = "free
19     space='free_y', switch = "y") +
20   geom_point(size = 3) + geom_errorbarh(height=.2)
21   geom_vline(xintercept=0, color='#C2352F', linetyp
22   theme_classic() +
23   labs(x = "Beta", y = "Variables") +
24   theme(axis.title = element_text(size = 20), axis.
25     title = element_text(size = 20), strip.plac
26     strip.text.y.left = element_text(size = 16,

```



angle = 0

Important note if you have interactions

- Please check your Quiz 3 feedback!
 - Many of you are not quite interpreting interactions correctly
- There is extra care to reporting results from interactions
- Please make sure you use `estimable()` to get the correct effects!
 - If you have an interaction with a categorical variable, then your main variable's "effect" should be reported for each category
- See HW 3 and Lecture 10 for reporting!
- You will **NOT include** the adjusting variables in the table/plot
 - But you **need to say adjusting for all other variables** in model from methods

Table/plot if you have interactions!!

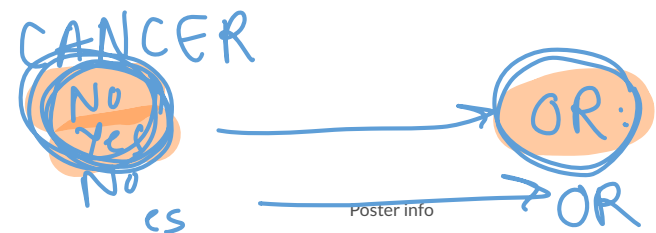
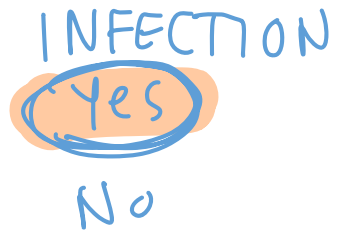
- Let's say my model from HW 3 included other variables and my main research question is about association between death and if infection is probable at ICU intake

Table __: Estimated odds ratios and confidence intervals of death at discharge. Odds ratios are comparing infection probable and not probable at ICU intake depending for two cases when cancer is and is not part of problem. Odds ratios adjust for all other variable in the model presented in methods.

Cancer	Infection	Estimated OR	95% CI
Cancer part of present problem	Infection probable at ICU intake		
	No		
	Yes	FILL HERE	FILL HERE
Cancer not part of present problem	Infection probable at ICU intake		
	No		
	Yes	FILL HERE	FILL HERE

$I(CAN = \text{"Yes"}) = 1$

$I(CAN = \text{Yes}) = 0$



$$\text{logit}(\pi) = \beta_0 + \beta_1 I(\text{CAN} = \text{Yes}) + \beta_2 I(\text{INF} = \text{Yes}) + \beta_3 I(\text{CAN} = \text{Yes}) I(\text{INF} = \text{Yes})$$

CAN N.
INF N.

$$\text{logit}(\pi) = \beta_0$$

CAN No
INF No $\rangle \text{logit}(\pi) = \beta_0$

CAN No
INF Yes

$$\text{logit}(\pi) = \beta_0 + \beta_2$$

CAN Yes
INF No $\rangle \text{logit}(\pi) = \beta_0 + \beta_1$

INF'S MAIN given no cancer: OR of death comparing INF to no INF = $\exp(\beta_2)$

CAN'S MAIN given no inf: OR of death comparing CAN to no CAN = $\exp(\beta_1)$

Saving tables

- For tables: `gtsave` in the `gt` package should work!
- For figures/plots: `ggsave` in the `ggplot2` package should work!

```
1 gtsave(reg_table, filename = here("lessons/19_Poster_help/reg_table.png"))
2 gtsave(table1, filename = here("lessons/19_Poster_help/table1.png"))
3 ggsave(forest_plot, filename = here("lessons/19_Poster_help/forest_plot.png"))
```

- When I first ran the above, I got the following message: `The package "webshot2" is required to save gt tables as images. * Would you like to install it? 1: Yes 2: No`
 - I inputted `1`

Any other questions about the poster?

Office Hours

Katherine

Brent :)

Sophia

Asnira

Jared

Heidi

Fazila

Quiz 3, question 8

```
mod_main = glm(diabetes ~ preg + age, data = pima2, family = binomial)
mod_int = glm(diabetes ~ preg*age, data = pima2, family = binomial)
tbl_regression(mod_int, estimate_fun = ~ style_number(.x, digits = 4))
```

Characteristic	log(OR) ¹	95% CI ¹	p-value
Number of Pregnancies			
0-5	-	-	
6-10	2.3289	0.9926, 3.6712	<0.001
10+	3.4348	-1.7373, 8.7320	0.2
Age (years)	0.0451	0.0278, 0.0631	<0.001
Number of Pregnancies * Age (years)			
6-10 * Age (years)	-0.0471	-0.0802, -0.0142	0.005
10+ * Age (years)	-0.0628	-0.1790, 0.0521	0.3

¹OR = Odds Ratio, CI = Confidence Interval

$$\text{logit}(\hat{\pi}(\text{Age}, \# \text{preg})) =$$

$$\hat{\beta}_0 + \hat{\beta}_1 I(\# \text{preg} = "6-10") + \hat{\beta}_2 I(\# \text{preg} = "10+") + \hat{\beta}_3 \text{Age} + \hat{\beta}_4 I(\# \text{preg} = "6-10") \text{Age} + \hat{\beta}_5 I(\# \text{preg} = "10+") \text{Age}$$

$$\text{logit}(\hat{\pi}) = \hat{\beta}_0 + \hat{\beta}_3 \text{Age}$$

for 0-5 preg,
for every 1 yr
inc in age
the estimated
odds inc
by 4.6%

0-5

6-10

10+

Age OR

$$\exp(\hat{\beta}_3)$$

$$\exp(\hat{\beta}_3 + \hat{\beta}_4)$$

$$\exp(\hat{\beta}_3 + \hat{\beta}_5)$$

$$\text{logit}(\hat{\pi}) = \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_3 \text{Age} + \hat{\beta}_4 \text{Age}$$

$$= (\hat{\beta}_0 + \hat{\beta}_1) + (\hat{\beta}_3 + \hat{\beta}_4) \text{Age}$$

for 6-10 preg,
for every 1 yr
inc in age
the estimated
odds inc

by ~~4.6%~~

$$\rightarrow \frac{\exp(\hat{\beta}_3 + \hat{\beta}_4) - 1}{1} \times 100\%$$

$$\text{logit}(\pi) = \beta_0 + \beta_1 \underbrace{I(\text{unmet})}_{=1} + \beta_2 I(\text{less HS}) + \beta_3 I(\text{HS}) + \beta_4 I(\text{some coll}) + \beta_5 \underbrace{I(\text{unmet})}_{=1} \underbrace{I(\text{less HS})}_{=1} + \beta_6 \underbrace{I(\text{unmet})}_{=1} I(\text{HS}) + \beta_7 \underbrace{I(\text{unmet})}_{=1} I(\text{some coll}) + \text{all other}$$

Yes
(unmet)

educ

OR

$$\beta_0 + \beta_1 + \beta_2 I(\text{less HS})$$

red

HS

$$\exp(\beta_3 + \beta_6)$$

$$+ \beta_3 I(\text{HS}) +$$

less HS

$$\exp(\beta_2 + \beta_5)$$

$$\beta_4 I(\text{some college}) \text{ some college}$$

$$\exp(\beta_4 + \beta_7)$$

$$+ \beta_5 I(\text{less HS})$$

$$+ \beta_6 I(\text{HS}) + \beta_7 I(\text{some coll})$$

$$(\beta_3 + \beta_6) I(\text{HS})$$

$$\exp(\beta_3 + \beta_6)$$

$$\begin{aligned} \text{logit}(\pi) &= \beta_0 + \beta_1 I(\text{MSA} = \text{NonM}) \\ &+ \beta_2 I(\text{gender} = \text{female}) \\ &+ \beta_3 I(\text{MSA} = \text{NonM}) I(\text{gender} = \text{fem}) \\ &+ \text{marital / edue} \quad \downarrow = 1 \end{aligned}$$

male: OR of FI non M to M: $e^{\beta_1} = 1.46$

female: OR of FI non M to M: $e^{\beta_1 + \beta_3}$

females

$$\text{logit}(\pi) = \beta_0 + \beta_1 I(\text{MSA})$$

$$+ \beta_2 + \beta_3 I(\text{MSA})$$

non met
~~MSA~~: $\beta_0 + \beta_1 + \beta_2 + \beta_3$

~~MSA~~
Net: $\beta_0 + \beta_2$