HW 05: Multinomial logistic regression

Karen Dong

2023-11-27

Set up

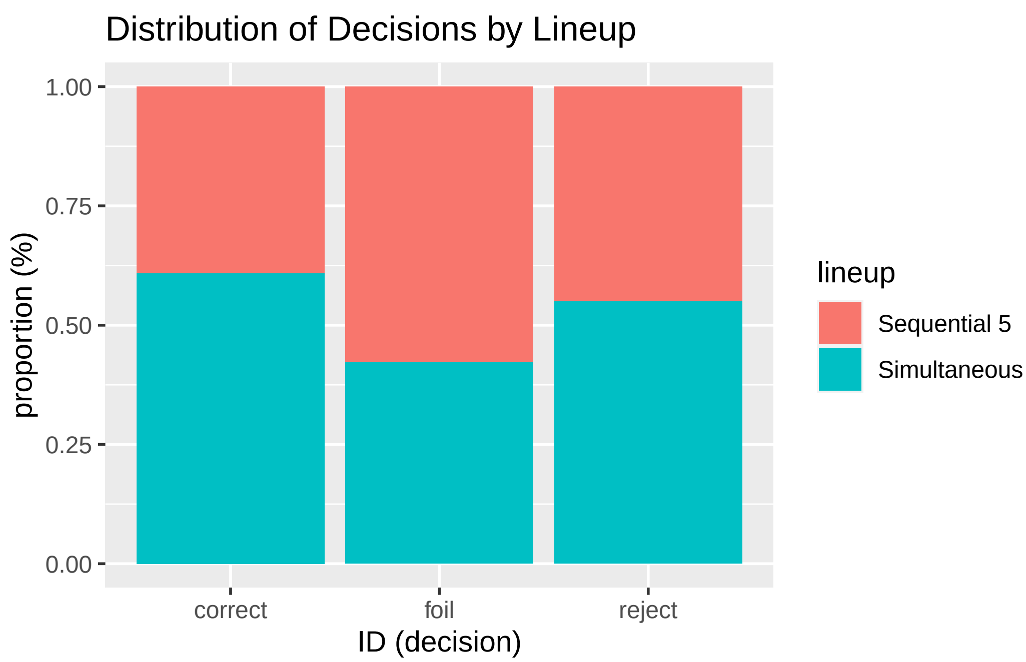
```r
library(tidyverse)
library(tidymodels)
library(knitr)
library(patchwork)
```

```r
ew <- read_csv("data/eyewitness.csv")
ew <- ew |>
  mutate(id = as_factor(id))
```

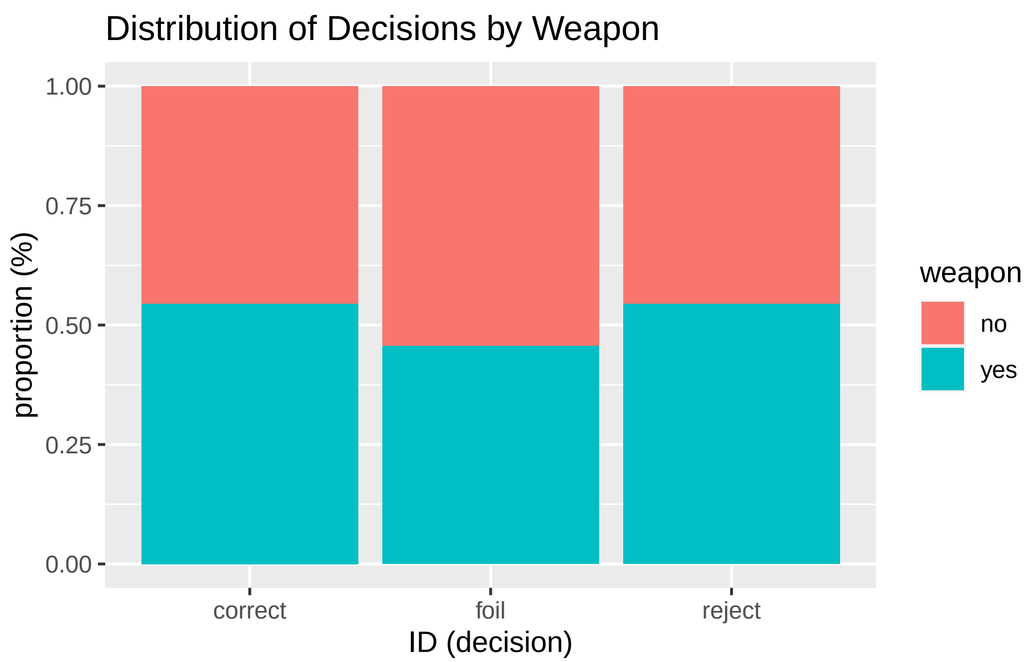
Exercises

Exercise 1

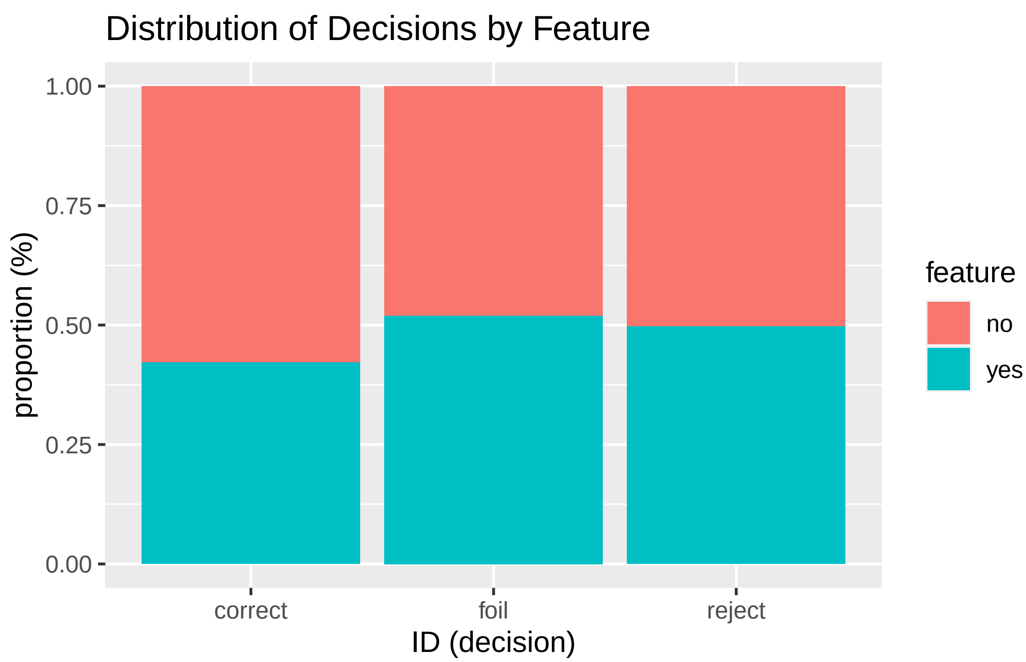
```r
ew |>
  ggplot(aes(x = id, fill = lineup)) +
  geom_bar(position = "fill") +
  labs(x = "ID (decision)",
       y = "proportion (%)",
       title = "Distribution of Decisions by Lineup")
```




```r
ew |>
  ggplot(aes(x = id, fill = weapon)) +
  geom_bar(position = "fill") +
  labs(x = "ID (decision)",
       y = "proportion (%)",
       title = "Distribution of Decisions by Weapon")
```




```
ew |>
  ggplot(aes(x = id, fill = feature)) +
  geom_bar(position = "fill") +
  labs(x = "ID (decision)",
       y = "proportion (%)",
       title = "Distribution of Decisions by Feature")
```



Distribution of Decisions by Feature
1.00
0.75
0.50
0.25
0.00
proportion (%)
correct
foil
reject
ID (decision)
feature
no
yes

Exercise 2

- One thing I learned about the data from the univariate plots is that the most frequent decision is foil, which means on average, people incorrectly identified the "foil" (**foil**) more times than correctly identifying the true perpetrator (**correct**), and more times than incorrectly concluding the true perpetrator is not in the lineup (**reject)**.
- `lineup` appears to have significant effect on the `id` since there seems to be the greatest difference between the decision being foil vs. correct when we observe the lineup being Sequential 5 vs. Simultaneous.

Exercise 3

Since response variable `id` is a categorical variable that have multiple potential outcomes, we should use a multinomial logistic regression model to predict id from `lineup`, `weapon` and `feature`.

Exercise 4

```r
ew_fit <- multinom_reg() |>
  set_engine("nnet") |>
  fit(id ~ lineup + weapon + feature, data = ew)

ew_fit |>
  tidy() |>
  kable(digits = 3)
```

y.level	term	estimate	std.error	statistic	p.value
foil	(Intercept)	1.063	0.165	6.442	0.000
foil	lineupSimultaneous	-0.803	0.166	-4.849	0.000
foil	weaponyes	-0.377	0.164	-2.300	0.021
foil	featureyes	0.450	0.165	2.725	0.006
reject	(Intercept)	-0.199	0.205	-0.974	0.330
reject	lineupSimultaneous	-0.266	0.199	-1.340	0.180
reject	weaponyes	-0.007	0.196	-0.037	0.971
reject	featureyes	0.327	0.197	1.660	0.097

- The baseline category for the response variable is the participant making the **correct** decision to identify the true perpetrator.
- The odds of a participant incorrectly identifying the foil (**foil**) vs. correctly identifying the true perpetrator (**correct**) when lineup is sequential 5, weapon is no, and feature is no, is approximately **2.895{exp(1.063)}**.
- The odds of a participant incorrectly concluding the true perpetrator is not in the lineup decision (**reject**) vs. correctly identifying the true perpetrator (**correct**) when lineup is sequential 5, weapon is no, and feature is no, is approximately **0.820(exp{-0.199)}**.
- When lineup is **simultaneous**, odds of a participant making incorrectly identifying the foil (**foil**) vs. correctly identifying the true perpetrator (**correct**) is expected to multiply by approximately **0.450{exp(-0.803)}** compared to when the lineup is sequential 5, holding weapon and feature constant.
- When lineup is **simultaneous**, odds of a participant incorrectly concluding the true perpetrator is not in the lineup decision (**reject**) vs. correctly identifying the true perpetrator (**correct**) is expected to multiply by approximately **0.766{exp(-0.266)}** compared to when the lineup is sequential 5, holding weapon and feature constant.

Exercise 5

```
ew_fit1 <- multinom_reg() |>
     set_engine("nnet") |>
     fit(id ~ (lineup + weapon + feature)^2, data = ew)

ew_fit1 |>
  tidy() |>
  kable(digits = 3)
```

y.level	term	estimate	std.error	statistic	p.value
foil	(Intercept)	1.087	0.225	4.829	0.000
foil	lineupSimultaneous	-1.134	0.282	-4.019	0.000
foil	weaponyes	-0.031	0.290	-0.107	0.914
foil	featureyes	0.354	0.304	1.166	0.244
foil	lineupSimultaneous:weaponyes	-0.042	0.333	-0.126	0.900
foil	lineupSimultaneous:featureyes	0.810	0.334	2.429	0.015
foil	weaponyes:featureyes	-0.714	0.333	-2.146	0.032
reject	(Intercept)	0.000	0.275	-0.001	1.000
reject	lineupSimultaneous	-0.761	0.345	-2.205	0.027
reject	weaponyes	0.148	0.348	0.426	0.670
reject	featureyes	-0.125	0.379	-0.329	0.742
reject	lineupSimultaneous:weaponyes	-0.014	0.403	-0.034	0.972
reject	lineupSimultaneous:featureyes	1.127	0.402	2.803	0.005
reject	weaponyes:featureyes	-0.376	0.400	-0.940	0.347

```
anova(ew_fit$fit, ew_fit1$fit, test = "Chisq") |>
  kable(digits = 3)
```

Model	Resid. df	Resid. Dev	Test	Df	LR stat.	Pr(Chi)
lineup + weapon + feature	1764	1773.458		NA	NA	NA
(lineup + weapon + feature)^2	1758	1759.317	1 vs 2	6	14.14	0.028

- Based on my test, the p-value is approximately 0.028. This is lower than the significance threshold of 0.05, so there is evidence of at least one significant interaction effect since there is a difference between the model **with** interaction effects and the model **without** interaction effects.

Exercise 6

- If there was no weapon but the perpetrator had a distinctive feature in the mock crime, the log-odds of reject vs. correct ID increases by **0.366** when there is a simultaneous vs. a sequential lineup.
- If there was no weapon but the perpetrator had a distinctive feature in the mock crime, the odds of reject vs. correct ID multiplies by **1.442{exp(0.366)}** when there is a simultaneous lineup vs. a sequential lineup.
- The intercept describes the group of participants under the condition of **sequential lineup**, **no weapons** present in the mock crime video, and **no distinctive features** on the perpetrator as part of the baseline category of making the **correct** decision to identify the true perpetrator.

Exercise 7

- **Linearity:** Our model only consists of categorical variables, so we **can** assume linearity is satisfied since only quantitative predictors need to test for linearity.
- **Randomness:** The article mentions the researchers collected data from "720 undergraduates, either online (n = 220) or in a laboratory setting (n = 500) across three Midwestern universities, and 2269 additional adult online participants from across the U.S. with SurveyMonkey." Undergrad students at Midwestern colleges and SurveyMonkey users are most likely volunteers who chose to opt into the study, so random selection cannot be established. It is likely the conditions are randomly assigned to the volunteers, but we cannot fully confirm the article's methodologies without reading the full paper. Randomness **cannot** be assumed to be satisfied.
- **Independent:** It is reasonable to assume any given participant's decision when identifying perpetrators do not influence another participant's decision. Independence **can** be assumed to be satisfied.

Exercise 8

```r
ew_aug <- augment(ew_fit1, new_data = ew)
  ew_aug |> select(contains("pred"))
```

```
# A tibble: 886 x 4
   .pred_class .pred_correct .pred_foil .pred_reject
   <fct>               <dbl>      <dbl>        <dbl>
 1 correct             0.413      0.394        0.193
 2 correct             0.413      0.394        0.193
 3 correct             0.413      0.394        0.193
 4 correct             0.413      0.394        0.193
 5 correct             0.413      0.394        0.193
 6 correct             0.413      0.394        0.193
 7 correct             0.413      0.394        0.193
 8 correct             0.413      0.394        0.193
 9 correct             0.413      0.394        0.193
10 correct             0.413      0.394        0.193
# i 876 more rows
```

```r
ew_conf <- ew_aug |>
  count(id, .pred_class, .drop = FALSE) |>
  pivot_wider(names_from = .pred_class, values_from = n)

ew_conf |>
  kable(digits = 3)
```

id	correct	foil	reject
correct	88	147	0
foil	81	379	0
reject	44	147	0

- Using `lineup`, `weapon` and `feature`, the model calculates log-odds for each outcome and converts into probabilities. The most probable `id` choice was chosen for each of the levels.
- The misclassifaction rate is 47.291%
  - (147 + 81 + 44 + 147) / (147 + 81 + 44 + 147 + 88 + 379)