HW13-Solutions

2024-04-20

1. In section 8.2, the author presented an example in which treatment allocation was confounded with sex. Create the contingency table for treatment and recovery that would have resulted had treatment allocation been balanced among gender groups. Your contingency table need not report counts; you may simply report proportions/probabilities. As the conditional row probabilities, $P(R|T)$ and $P(R|T^c)$, are the primary quantities of interest, add the conditional row probabilities to the table.

From the provided table in the textbook, we have:

$$P(R|T,M) = 0.6$$
$$P(R|T,F) = 0.2$$
$$P(R|T^c,M) = 0.7$$
$$P(R|T^c,F) = 0.3$$

Where $P(R|T,M)$ means the probability of recovery given Treatment AND Male.

we can see calculate the rest of the table. We will have equal proportions in each of the categories: Male & Treatment, Female & Treatment, Male & No Treatment, and Female & No Treatment.

To find the $P(T \cap R)$, we can use the law of total probability.

$$\begin{aligned} P(T \cap R) &= P(R|T,M)P(T,M) + P(R|T)P(T,F) \\ &= 0.6 * 0.25 + 0.2 * 0.25 = 0.2 \end{aligned}$$

The same idea can be used for $P(T^c \cap R)$.

$$\begin{aligned} P(T^c \cap R) &= P(R|T^c,M)P(T^c,M) + P(R|T^c)P(T^c,F) \\ &= 0.7 * 0.25 + 0.3 * 0.25 = 0.25 \end{aligned}$$

We know from our equal assignment that the $P(T) = P(T^c) = 0.5$

Plugging these values into the table:

	R	R^c	Total
T	0.2		0.5
row			

	R	R^c	Total
T^c	0.25		0.5
row			
Total	0.45		1

Using the complement and calculating the row conditionals we finish with:

	R	R^c	Total
T	0.2	0.3	0.5
row	0.4	0.6	
T^c	0.25	0.25	0.5
row	0.5	0.5	
Total	0.45	0.55	1

2. In the trial example, men and women were equally likely to participate. How would the table change if women were twice as likely than men to join the trial? Create the resulting contingency table.

When there are twice as many women as men in the study, then:

$$P(M) = \frac{1}{3}, \quad P(F) = \frac{2}{3}$$

We still have $P(T) = P(T^c) = 0.5$. Since treatment and sex are independent, we can calculate:

$$P(T, M) = P(T \cap M) = P(T)P(M) = \frac{1}{2} * \frac{1}{3} = \frac{1}{6}$$

Same can be done for $P(T, F)$:

$$\begin{gathered} P(T, F) = P(T \cap F) = P(T)P(F) \\ = \frac{1}{2} * \frac{2}{3} = \frac{2}{6} = \frac{1}{3} \end{gathered}$$

Solving for $P(R|T)$ using the law of total probability again:

$$\begin{gathered} P(T \cap R) = P(R|T, M)P(T, M) + P(R|T)P(T, F) \\ = 0.6 * \frac{1}{6} + 0.2 * \frac{1}{3} = \frac{1}{6} \end{gathered}$$

Same for $P(R|T^c)$:

$$\begin{gathered} P(T^c \cap R) = P(R|T^c, M)P(T^c, M) + P(R|T^c)P(T^c, F) \\ = 0.7 * \frac{1}{6} + 0.3 * \frac{1}{3} = \frac{13}{60} = 0.21\bar{6} \end{gathered}$$

	R	R^c	Total
T	$\frac{1}{6}$	$\frac{1}{3}$	0.5
row			
T^c	$\frac{13}{60}$	$\frac{17}{60}$	0.5
row			
Total	$\frac{23}{60}$	$\frac{37}{60}$	1

Filling in the row conditional values:

	R	R^c	Total
T	$\frac{1}{6}$	$\frac{1}{3}$	0.5
row	$\frac{1}{3}$	$\frac{2}{3}$	
T^c	$\frac{13}{60} = 0.21\bar{6}$	$\frac{17}{60} = 0.28\bar{3}$	0.5
row	$\frac{13}{30} = 0.4\bar{3}$	$\frac{17}{30} = 0.5\bar{6}$	
Total	$\frac{23}{60} = 0.38\bar{3}$	$\frac{37}{60} = 0.61\bar{6}$	1

3. The treatment effect may be calculated on the relative scale or the absolute scale. Which of these quantities is absolute and which is relative? Why?

$$\Delta = P(R|T) - P(R|T^c), \quad \rho = \frac{P(R|T)}{P(R|T^c)}$$

Δ is the absolute scale because it is finding the difference between the two values. On the other hand, ρ is the relative scale because it is a ratio of the probabilities.

4. Calculate both Δ and ρ for the tables created in 1 and 2. Which summary measure is the same and which differs as the ratio of men to women changes? Can you show this generally? (Let α be the proportion of women. Derive an expression for Δ and ρ.)

Subscripts indicate which table it came from: the first (1) or second (2)

$$\Delta_1 = 0.4 - 0.5 = -0.1, \quad \rho_1 = \frac{0.4}{0.5} = \frac{4}{5}$$

$$\Delta_2 = \frac{1}{3} - \frac{13}{30} = -0.1, \quad \rho_2 = \frac{\frac{1}{3}}{\frac{13}{30}} = \frac{10}{13} = 0.769$$

From these findings, the value for Δ (absolute) does not change from the first table to the second, but ρ (relative) does change.

We can show this in a general sense by expanding the formula for Δ and ρ:

$$
\begin{gathered}
\Delta = P(R|T) - P(R|T^c) \\
= [P(R|T,M)(1-\alpha) + P(R|T,F)(\alpha)] - [P(R|T^c,M)(1-\alpha) + P(R|T^c,F)(\alpha)] \\
\text{Plug in the values from the previous questions:} \\
= [0.6(1-\alpha) + 0.2\alpha] - [0.7(1-\alpha) + 0.3\alpha] \\
= 0.6 - 0.6\alpha + 0.2\alpha - 0.7 - 0.7\alpha - 0.3\alpha \\
\text{Combining terms:} \\
= 0.6 - 0.7 = -0.1
\end{gathered}
$$

Doing the same for ρ:

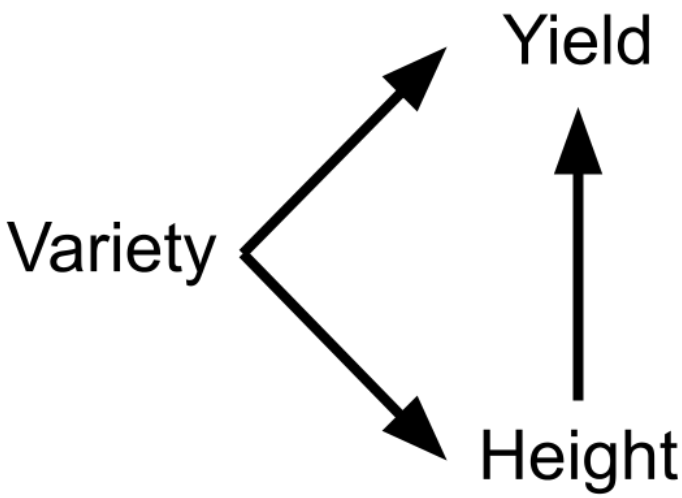
$$
\begin{gathered}
\frac{P(R|T)}{P(R|T^c)} \\
= \frac{[P(R|T,M)(1-\alpha) + P(R|T,F)(\alpha)]}{[P(R|T^c,M)(1-\alpha) + P(R|T^c,F)(\alpha)]} \\
= \frac{0.6(1-\alpha) + 0.2\alpha}{0.7(1-\alpha) + 0.3\alpha} \\
= \frac{0.6 - 0.4\alpha}{0.7 - 0.4\alpha} \\
= \frac{6 - 4\alpha}{7 - 4\alpha}
\end{gathered}
$$

From these results, we see that the absolute scale is relative while the relative scale depends on α.

5. In section 8.6, the author compares a medical trial to an agricultural trial. Summarize why the conclusions of the trials were different, even though the data were exactly the same.

The difference is due to the exchangability of the data. We don't know if it'll be tall or short beforehand. We know $P(R|T)$ but not $P(R|T,S)$. We can't know the height beforehand, so it doesn't make sense to stratify on it. Height is on the causal pathway.

6. Consider the figure below, which describes the medical trial. Create a figure that represents the agricultural trial, using the terms variety, yield, and height.



7. Suppose that you were tasked with creating a randomization table for a trial. The investigators are concerned that outcomes differ by diabetes status. Create randomization tables using R which accounts for diabetes status. Suppose the total number of trial participants is 100, of which 25 will have diabetes.

```r
### create random labels for the treatment

# first for the diabetes group
# 0 means no treatment, 1 means treatment

# we can have 12 not have the treatment and 13 have it since 25 is not even

diabetes_treatment_options <- c(rep(0, 12),
                                rep(1, 13))

diabetes_treatment <- sample(diabetes_treatment_options,
                             size = 25,
                             replace = FALSE)
```

```r
# next for the non diabetes group

# have 37 with no treatment, 38 with treatment since 75 isn't even

non_diabetes_treatment_options <- c(rep(0, 37),
                                    rep(1, 38))

non_diabetes_treatment <- sample(non_diabetes_treatment_options,
                                 size = 75,
                                 replace = FALSE)

# now create labels for Diabetes (25) and No Diabetes

labels <- c(rep('Diabetes', 25), rep('No Diabetes', 75))

# combine the treatment and labels

randomization_data <- data.frame(diabetes_Status = labels,
                                 treatment = c(diabetes_treatment,
                                              non_diabetes_treatment))

# combine into a table to view

table(randomization_data)
```

```
##                treatment
## diabetes_Status  0  1
##     Diabetes    12 13
##     No Diabetes 37 38
```