Fairness homework

Kayané Robach

December 2021

1 Simpson's paradox

1.1

30.33% of the women who apply are admitted ($\frac{556.62}{1835}$) whereas 44.52% of the men who apply are admitted ($\frac{1198.02}{2691}$). A higher percentage of men is admitted but also we have to take into account that more men are applying. We cannot tell if it is biased without doing a statistical test to correctly answer the question.

1.2

In department A 82% of the women who applied get admited whereas only 62% of the men were admitted. However there is much less women applying in this department than men, indeed in reality around 88 women and 511 men have been admitted. Again it is hard to say that this result is biased. We have a similar result in department B, the percentage of admission between men and women are closer but there is much less women applying. In department C it is the opposit ; 37% of men are admitted over 325 applications whereas 34% of women are admitted over 593 applications. Except for department A, there is a small difference in the percentage of women or men admitted but there is a large difference in the number of applications that are form women or from men. (... similar results for other departments).

1.3

With treatment A, 78.9% of the kidney stones are properly treated. With the B, 82.6% are properly treated. Indeed over the 350 kidney stones considered for each of both treatments, the B properly treats 289 whereas the A properly treats only 276 cases.

1.4

There is 96.6% of success of treatment A on small stones, 73.3% on large ones. There is 86.7% of success of treatment B on small stones and 68.8% on large ones. Apparently the treatment A could seem to be the best one, nonetheless

both treatments have been tested on 350 kidney stones but not on the same proportions for the two kind of stones considered.

To conclude the treatment B seems to be the best. The error we can do by looking at the table is to not consider that, for small kidney stones as for large ones, the number of cases treated with A and with B are not the same.

1.5

$$P(C|T=1) = \frac{84+192}{87+263} = \frac{276}{350} = 78.9\%$$

$$P(C|T=0) = \frac{234+55}{270+80} = \frac{289}{350} = 82.6\%\%$$

$$\implies P(C|T=0) > P(C|T=1)$$

$$P(C|T=1, S=small) = \frac{84}{87} = 96.6\%$$

$$P(C|T=1, S=large) = \frac{192}{263} = 73.3\%$$

$$P(C|T=0, S=small) = \frac{234}{270} = 86.7\%$$

$$P(C|T=0, S=large) = \frac{55}{80} = 68.8\%$$

$$\implies P(C|T=1, S=small) > P(C|T=0, S=small)$$

$$\implies P(C|T=1, S=large) > P(C|T=0, S=large)$$

Selection bias could be introduced by the selection of individuals in a program. In order to avoid selection bias it is necessary to compare population that would have behave likely under the same treatment (but the challenge is that we can never answer this question). In our case it would have been better if treatment A was tested on the same proportion of small kidney than large ones, same for treatment B. So that we would have been able to correctly understand the behavior of both treatments.

2 Some fairness criteria

2.1

$$R \perp A|Y \iff P(R|Y, A=a) = P(R|Y, A=b)$$

Remind that the false positive rate expresses as $P(R=1|Y=0)$, whereas false negative rate expresses as $P(R=0|Y=1)$. In other words, separation critera means that both groups have the same false positive and false negative rates (as well as same true positive and true negative rates).

2.2

The classifier satisfying the separation criteria should be the one achieving, in both groups, the same true positive and false positive.

Considering the ROC curve presenting true positive against false positive, we have to consider the points of the intersection of both group-level curves, that is to say, the region below both curves.

2.3

We want to show that $l(R)$ satisfies calibration by group

$$\iff P(Y = 1|l(R) = r, A = a) = P(Y = 1|l(R) = r, A = b) = r$$

$$P(Y = 1|l(R) = r, A = a) = P(Y = 1|R \in l^{-1}(r), A = a)$$

Since we know that R satisfies sufficiency we get

$$P(Y = 1|l(R) = r, A = a) = P(Y = 1|R \in l^{-1}(r), A = a) = P(Y = 1|R \in l^{-1}(r), A = b)$$

$$\iff P(Y = 1|l(R) = r, A = a) = P(Y = 1|l(R) = r, A = b)$$

Then by choosing $l(r) = P(Y = 1|R = r, A = a)$ we obtain $r = P(Y = 1|R \in l^{-1}(r), A = a)$, we finally get the result :

$$r = P(Y = 1|l(R) = r, A = a) = P(Y = 1|l(R) = r, A = b)$$

2.4

We want to show that :

$$A \not\perp Y \; and \; R \not\perp Y \implies$$

independence and separation are incompatible.

In other words, the only possibility to have separation and independence is to have either $A \perp Y$ or $R \perp Y$.

Let us show that : $A \perp R$ and $A \perp R|Y \implies A \perp Y$ or $R \perp Y$.

Assume independence, then

$$P(R = r) = P(R = r|A = a)$$

$$\iff \sum_y P(R = r|Y = y)P(Y = y) = \sum_y P(R = r|A = a, Y = y)P(Y = y|A = a)$$

Assume separation, then

$$\iff \sum_y P(R = r|Y = y, A = a)P(Y = y) = \sum_y P(R = r|A = a, Y = y)P(Y = y|A = a)$$

and also :

$$\iff \sum_y P(R=r|Y=y)P(Y=y) = \sum_y P(R=r|Y=y)P(Y=y|A=a)$$

Using the second sentence, we develop :

$$P(R=r|Y=0)P(Y=0) + P(R=r|Y=1)P(Y=1)$$
$$= P(R=r|Y=0)P(Y=0|A=a) + P(R=r|Y=1)P(Y=1|A=a)$$

$$P(R=r|Y=0)P(Y=0) + P(R=r|Y=1)(1-P(Y=0))$$
$$= P(R=r|Y=0)P(Y=0|A=a) + P(R=r|Y=1)(1-P(Y=0|A=a))$$

$$P(Y=0)[P(R=r|Y=0) - P(R=r|Y=1)] + P(R=r|Y=1)$$
$$= P(Y=0|A=a)[P(R=r|Y=0) - P(R=r|Y=1)] + P(R=r|Y=1)$$

$$P(Y=0)[P(R=r|Y=0) - P(R=r|Y=1)]$$
$$= P(Y=0|A=a)[P(R=r|Y=0) - P(R=r|Y=1)]$$

$$\implies P(Y=0) = P(Y=0|A=a) \text{ or } P(R=r|Y=0) = P(R=r|Y=1)$$

Which implies $Y \perp A$ or $R \perp Y$.