Assignment 1

Exercise 1

Headline 1	**Just ONE pint a day 'poisons your brain and increases your risk of dementia'**
causal wording	poisons your brain, increases your risk
source	https://www.thesun.co.uk/news/5341512/alcohol-one-pint-a-day-poisons-brain-increases-risk-of-dementia/ The Sun article of January 2018 study of more than 13,000 boozers, led by Oxford academics, published in the Journal of Public Health last week Prof. Simon Moore source of original study not indicated

Headline 2	**Psychose vom Kiffen?**
causal wording	vom Kiffen
source	https://www.drugcom.de/newsuebersicht/topthemen/psychose-vom-kiffen/ article from April 2010 they cite several sources: ● D'Souza DC, Sewell RA, Ranganathan M. Cannabis and psychosis/schizophrenia: human studies. Eur Arch Psychiatry Clin Neurosci. 2009 Oct;259(7):413-31. doi: 10.1007/s00406-009-0024-2. Epub 2009 Jul 16. PMID: 19609589; PMCID: PMC2864503. ○ meta-analysis ○ subjects of included studies: people ○ peer-reviewed ● Moore, T., Zammit, S., Lingfort-Huges, A., Barnes, T., Jones, P., Burke, M. & Lewis, G. (2007). Cannabis use and risk of psychotic or affective mental health outcomes: a systematic review. The Lancet, 370, 319-328 ○ meta-analysis ○ included longitudinal and population-based studies ○ data extraction in duplicate ○ subjects of the included studies: people ○ peer-reviewed reporting is in agreement with those 2 original papers

Headline 3	**Why caffeine may limit weight gain**

causal wording	may limit
source	https://www.medicalnewstoday.com/articles/327391 Medical News Today Article from January 2020 source is indicated in the article: Fatima J. Zapata, Miguel Rebollo-Hernanz, Jan E. Novakofski, Manabu T. Nakamura, Elvira Gonzalez de Mejia, Caffeine, but not other phytochemicals, in mate tea (Ilex paraguariensis St. Hilaire) attenuates high-fat-high-sucrose-diet-driven lipogenesis and body fat accumulation, Journal of Functional Foods, Volume 64, 2020 ● peer-reviewed ● study subjects: rats ● experimental study ● results: caffeine from natural and synthetic sources promoted reduction of body fat accumulation in animals fed with a high-fat-high-sucrose diet ○ caffeine can be considered as anti-obesity agents hence the rather careful reporting is supported by the findings in the original study

Headline 4	**No, 5G radiation doesn't cause or spread the coronavirus. Saying it does is destructive**
causal wording	doesn't cause
source	https://theconversation.com/no-5g-radiation-doesnt-cause-or-spread-the-coronavirus-saying-it-does-is-destructive-135695 The Conversaton article from April 2020 cite WHO: https://www.who.int/health-topics/coronavirus#tab=tab_1 ● no scientific article ● but probably bases on scientific articles ● therefore no hit for an original study reporting agrees with WHO information

Exercise 2

Claim 1: *Data show that income and marriage have a high positive correlation. Therefore, your earnings will increase if you get married.*

Correlation does not imply causation. In claim 1, this implication is assumed.

Claim 2: *Data show that as the number of fires increase, so does the number of fire fighters. Therefore, to cut down on fires, you should reduce the number of fire fighters.*

The increase of the number of fire fighters is a reaction of society an the increase of the number of fires. This reaction is not necessary. Therefore, concluding that reducing the number of fire fighters reduces the number of fires is not valid.

Claim 3: *Data show that people who hurry tend to be late to their meetings. Don't hurry, or you'll be late.*

Here we observe correlation between hurrying and being late. The conclusion *Don't hurry, or you'll be late* implies that hurrying causes being late. This conclusion is not valid because the correlation could arise from other things like for example being very busy.

Assignment 1

3

screening test for doping: 90% sensitivity, 95% specificity

1 in 50 athletes is truly doping at any time

d: doping
¬d: not doping

p: positive test
n: negative test

If an athlete

- is doping, what is the probability that test is positive?

corresponds to sensitivity: $P(p \mid d) = 90\%$

- is not doping, what is the prob. that test positive?

$$P(p \mid \neg d) = 1 - \underbrace{P(n \mid \neg d)}_{\text{specificity}} = 1 - 0.95 = 0.05$$

- gets a pos. res., what is the probability that they doped?

this is the positive pred. value

$$P(d \mid p) = \frac{\text{sensitivity} \times \text{prevalence}}{\text{sensitivity} \times \text{prevalence} + (1-\text{specificity}) \times (1-\text{prevalence})}$$

$$= \frac{0.9 \cdot \frac{1}{50}}{0.9 \cdot \frac{1}{50} + (1-0.95) \cdot \left(1 - \frac{1}{50}\right)}$$

$$= \frac{\frac{9}{500}}{\frac{9}{500} + 0.05 \cdot \frac{49}{50}}$$

$$= \frac{\frac{9}{500}}{\frac{9}{500} + \frac{49}{1000}}$$

$$= \frac{\frac{9}{500}}{\frac{57}{1000}}$$ Algebra error

$$= \frac{9}{500} \cdot \frac{1000}{57}$$

$$= \frac{3 \cdot 2}{19}$$

$$= \underline{\underline{\frac{6}{19}}}$$

- gets a neg. test res., what is the probability they did not dope

corresponds to negative pred. value

$$P(\neg d \mid n) = \frac{\text{specificity} \times (1-\text{prevalence})}{\text{specificity} \times (1-\text{prevalence}) + (1-\text{sensitivity}) \times \text{pre}}$$

$$= \frac{0.95 \cdot \frac{49}{50}}{0.95 \cdot \frac{49}{50} + 0.1 \cdot \frac{49}{50}}$$ no

$$= \frac{\frac{931}{1000}}{\frac{931}{1000} + \frac{98}{1000}}$$

$$= \frac{931}{1029}$$

$$\approx 0.9$$

4

$$P(A|B) = \frac{P(AB)}{P(B)}$$

claim: A and B independent $\Rightarrow P(A|B) = P(A)$

proof: if A and B are independent, we have $P(AB) = P(A) \cdot P(B)$

thus $P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(A) \cdot \cancel{P(B)}}{\cancel{P(B)}} = P(A)$

- relationship between $P(A|B)$ and $P(B|A)$

$P(A|B) = \frac{P(AB)}{P(B)}$ and $P(B|A) = \frac{P(BA)}{P(B)}$

$\Rightarrow P(A|B) \cdot P(B) = P(AB) = P(BA) = P(B|A) \cdot P(B)$

$\Rightarrow P(A|B) \cdot P(B) = P(B|A) \cdot P(B)$

5

- $P(Y=1) = \frac{391}{646}$ $\quad P(Y=0) = \frac{255}{646}$

  $P(Y=1 \mid X=1) = \frac{307}{472}$ $\quad P(Y=1 \mid X=0) = \frac{84}{174}$

- odds exposed: $\frac{P(Y=1 \mid X=1)}{1-P(Y=1 \mid X=1)} = \frac{\frac{307}{472}}{1-\frac{307}{472}} \approx 1.861$

  odds unexposed $\frac{P(Y=1 \mid X=0)}{1-P(Y=1 \mid X=0)} = \frac{\frac{84}{174}}{1-\frac{84}{174}} = 0.9\overline{3}$

  hence odds ratio $= \frac{1.861}{0.9\overline{3}} \approx 1.99$

-

claim: the odds ratio for a 2x2 contingency table is symmetric

proof:

$$\frac{odds(Y=1 \mid X=1)}{odds(Y=1 \mid X=0)} = \frac{\frac{P(Y=1 \mid X=1)}{1-P(Y=1 \mid X=1)}}{\frac{P(Y=1 \mid X=0)}{1-P(Y=1 \mid X=0)}}$$

$$\frac{P(Y=1, X=1)}{P(X=1)}$$

$$\frac{P(Y=1, X=0)}{P(X=0)}$$

$$= \frac{\frac{1-P(Y=0 \mid X=1)}{P(Y=0 \mid X=1)}}{\frac{1-P(Y=0 \mid X=0)}{P(Y=0 \mid X=0)}}$$

You need to apply Bayes rule

relative risk

$$\frac{P(Y=1 \mid X=1)}{P(Y=1 \mid X=0)} = \frac{\frac{307}{472}}{\frac{84}{179}} \approx 1.35$$

claim: under the rare disease assumption ($P(Y=1) \to 0$), the RR under an exposure X can be approx. by the OR

proof:

$$\lim_{P(Y=1)\to 0} \frac{odds(Y=1 \mid X=1)}{odds(Y=1 \mid X=0)} = \lim_{P(Y=1)\to 0} \frac{\frac{P(Y=1 \mid X=1)}{1-P(Y=1 \mid X=1)}}{\frac{P(Y=1 \mid X=0)}{1-P(Y=1 \mid X=0)}}$$

$$= \frac{\frac{\lim_{P(Y=1)\to 0} P(Y=1 \mid X=1)}{\lim_{P(Y=1)\to 0} 1-P(Y=1 \mid X=1)}}{\frac{\lim_{P(Y=1)\to 0} P(Y=1 \mid X=0)}{\lim_{P(Y=1)\to 0} 1-P(Y=1 \mid X=0)}}$$

$$= \frac{\frac{\lim_{P(Y=1)\to 0} P(Y=1 \mid X=1)}{1}}{\frac{\lim_{P(Y=1)\to 0} P(Y=1 \mid X=0)}{1}}$$

$$= \lim_{P(Y=1)\to 0} RR$$

Are X and Y independent?

x

Assignment 2

Exercise 1

it is easier to change the treatments instead of the outcomes

we can modify the table on slide 28 by changing the outcome of Ceres:

But then the causal effect is not zero anymore. The first column changes

	Potential outcomes		Treatment	Outcome
Unit	$Y^{a=1}$	$Y^{a=0}$	A	Y
Juno	0	0	0	0
Ceres	0	0	1	1 ~~0~~
Vulcan	0	0	1	0
Jupiter	0	1	1	0
Minerva	0	1	1	0
Mercury	0	1	1	0
Neptune	1	0	0	0
Mars	1	0	0	0
Venus	1	0	1	1
Diana	1	1	0	1
Apollo	1	1	0	1
Vesta	1	1	1	1

for the **potential outcomes**, we still get:

no, Ceres now has Y^{a=1} = 1

$$P(Y^{a=1} = 1) = P(Y^{a=0} = 1) = 6/12 = .5$$

hence there is no average causal effect

but the conditional probabilites of getting flu we get something different:

$$P(Y = 1 \mid A = 1) = 3/7 \quad > \quad 2/5 = P(Y = 1 \mid A = 0)$$

Hence we observe an association between getting the treatment and getting the flu

Exercise 4

	Player A			Player B		
	Times at bat	Hits	Average	Times at bat	Hits	Average
Against right-handed pitchers (C_1)	202	45	$.223(=r_1)$	250	58	$.232(=R_1)$
Against left-handed pitchers (C_2)	250	71	$.284(=r_2)$	108	32	$.296(=R_2)$
Overall	452	116	$.257(=r)$	358	90	$.251(=R)$

Source: https://www.maa.org/sites/default/files/0746834219623.di020717.02p0042n.pdf (consulted 04.03.2022)

Causal Inference

Homework 2

Benedikt Schmidt

04 March 2022

Exercise 2

I chose a presentation of M. Lorez of the Foundation National Institute for Cancer Epidemiology and Registration (NICER). The numbers in the presentation are drawn from

Six et al. (2017). Age-dependent risk and lifetime risk of developing cancer in Switzerland. SCB 37(3), 284-291

and

Bruder et al. (2018). Estimating lifetime and 10-year risk of lung cancer. Preventive Medicine Reports 11, 125-13

I accessed the on March 4th, 2022, via https://www.nicer.org/assets/files/publications/presentations/spgpathlecture-lung-cancer_epi_2018-web.pdf

life time risk of lung cancer in female (heavy) smokers: 11 % life time lung cancer risk in female never smokers: 1 %

~~In this article, the life time risk of lung cancer~~

```
library(ggplot2)
library(broom)

# Exercise 2

# life time risk of lung cancer in female (heavy) smokers: 11 %
# life time lung cancer risk in female never smokers: 1 %

p_LC_s = 0.11
p_LC_ns = 0.01

# risk ratio
RR <- p_LC_s / p_LC_ns
RR
```

```
## [1] 11
```

```
# odds
odds_p_LC_s <- p_LC_s / (1 - p_LC_s)
odds_p_LC_ns <- p_LC_ns / (1 - p_LC_ns)

# odds ratio
OR <- odds_p_LC_s / odds_p_LC_ns
OR
```

```
## [1] 12.23596
```

```r
# ratio gsi
xi <- RR / OR
xi
```

```
## [1] 0.8989899
```

```r
# gsi is not good approximation for RR
RR - xi
```

```
## [1] 10.10101
```

```r
# estimated smoking prevalence in swiss female population in 2012 (from the same
# presentation)
p_s = 0.22

# The rare disease assumption is obviously not satisfied

# ratio chi

chi <- p_LC_ns / p_s
chi
```

```
## [1] 0.04545455
```

```r
seq_p_s <- seq(0.22, 0.01, -0.01)
seq_p_s
```

```
##  [1] 0.22 0.21 0.20 0.19 0.18 0.17 0.16 0.15 0.14 0.13 0.12 0.11 0.10 0.09 0.08
## [16] 0.07 0.06 0.05 0.04 0.03 0.02 0.01
```

```r
seq_p_LC_ns <- 1:22
for (i in 1:22) {
  seq_p_LC_ns[i] <- chi * seq_p_s[i]
}
seq_p_LC_ns
```

You can vectorize this (faster):
chi * seq_p_LC_ns

```
##  [1] 0.0100000000 0.0095454545 0.0090909091 0.0086363636 0.0081818182
##  [6] 0.0077272727 0.0072727273 0.0068181818 0.0063636364 0.0059090909
## [11] 0.0054545455 0.0050000000 0.0045454545 0.0040909091 0.0036363636
## [16] 0.0031818182 0.0027272727 0.0022727273 0.0018181818 0.0013636364
## [21] 0.0009090909 0.0004545455
```

```r
# pairwise RR, OR and p_lc
seq_RR <- 1:22
seq_OR <- 1:22
seq_p_LC <- 1:22

for (i in 1:22) {
  seq_RR[i] <- p_LC_s / seq_p_LC_ns[i]
  seq_OR[i] <- odds_p_LC_s / (seq_p_LC_ns[i] / (1 - seq_p_LC_ns[i]) )
  seq_p_LC[i] <- seq_p_s[i] * p_LC_s + (1 - seq_p_s[i]) * seq_p_LC_ns[i]
}

seq_RR
```

```
##  [1]  11.00000  11.52381  12.10000  12.73684  13.44444  14.23529  15.12500
##  [8]  16.13333  17.28571  18.61538  20.16667  22.00000  24.20000  26.88889
```

```
## [15]  30.25000  34.57143  40.33333  48.40000  60.50000  80.66667 121.00000
## [22] 242.00000
```

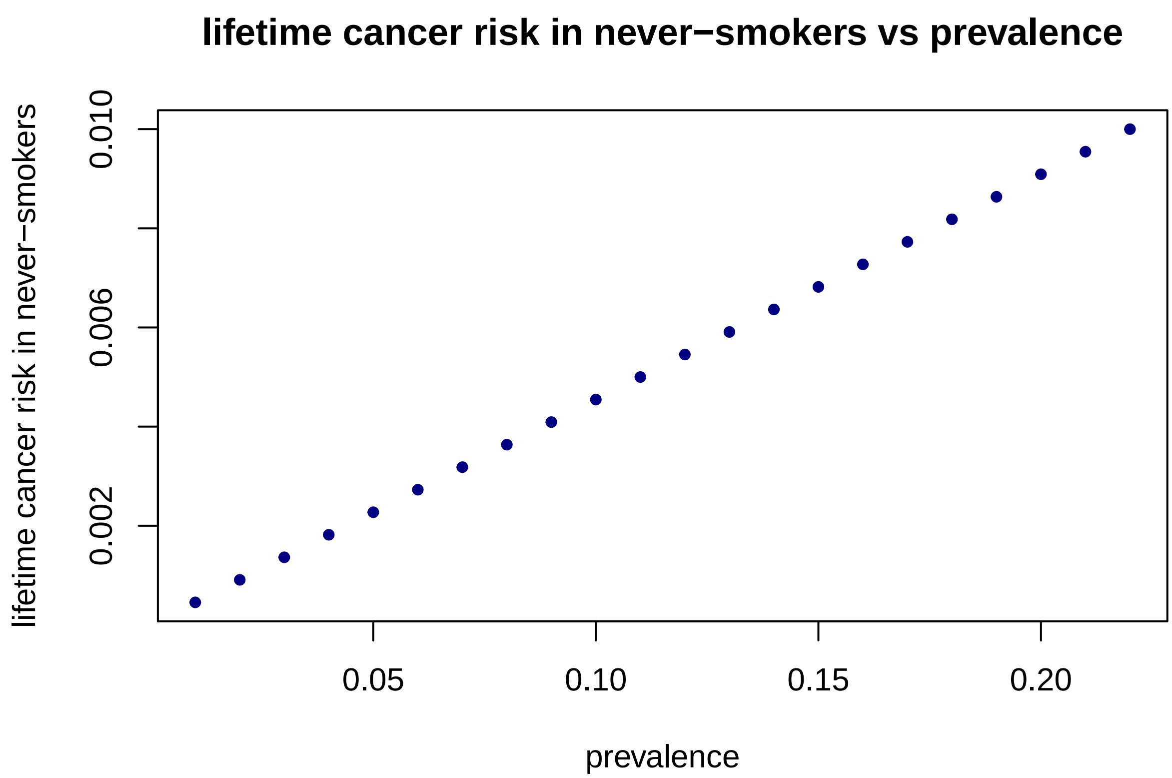
```r
seq_OR
```

```
##  [1]  12.23596  12.82451  13.47191  14.18746  14.98252  15.87112  16.87079
##  [8]  18.00375  19.29856  20.79257  22.53558  24.59551  27.06742  30.08864
## [15]  33.86517  38.72071  45.19476  54.25843  67.85393  90.51311 135.83146
## [22] 271.78652
```

```r
seq_p_LC
```

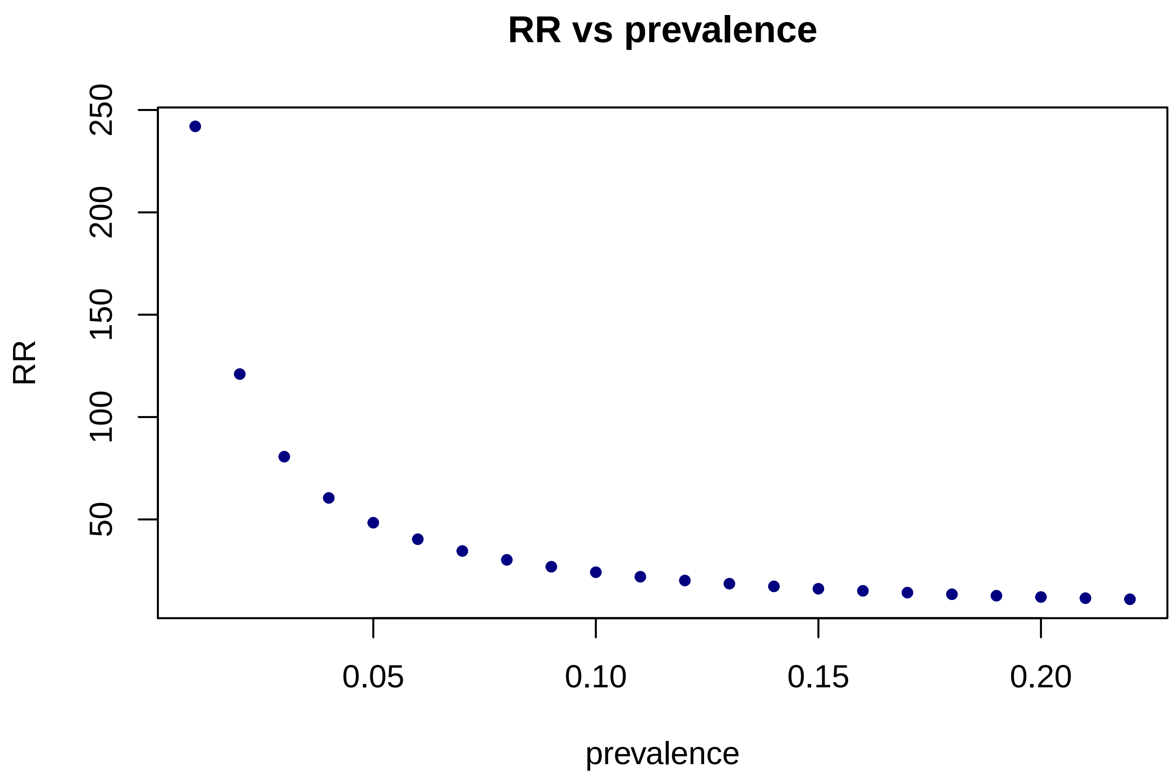
```
##  [1] 0.032000000 0.030640909 0.029272727 0.027895455 0.026509091 0.025113636
##  [7] 0.023709091 0.022295455 0.020872727 0.019440909 0.018000000 0.016550000
## [13] 0.015090909 0.013622727 0.012145455 0.010659091 0.009163636 0.007659091
## [19] 0.006145455 0.004622727 0.003090909 0.001550000
```

```r
plot(seq_p_s, seq_p_LC_ns,
     pch=20, col="navyblue", xlab="prevalence",
     ylab="lifetime cancer risk in never-smokers",
     main="lifetime cancer risk in never-smokers vs prevalence")
```

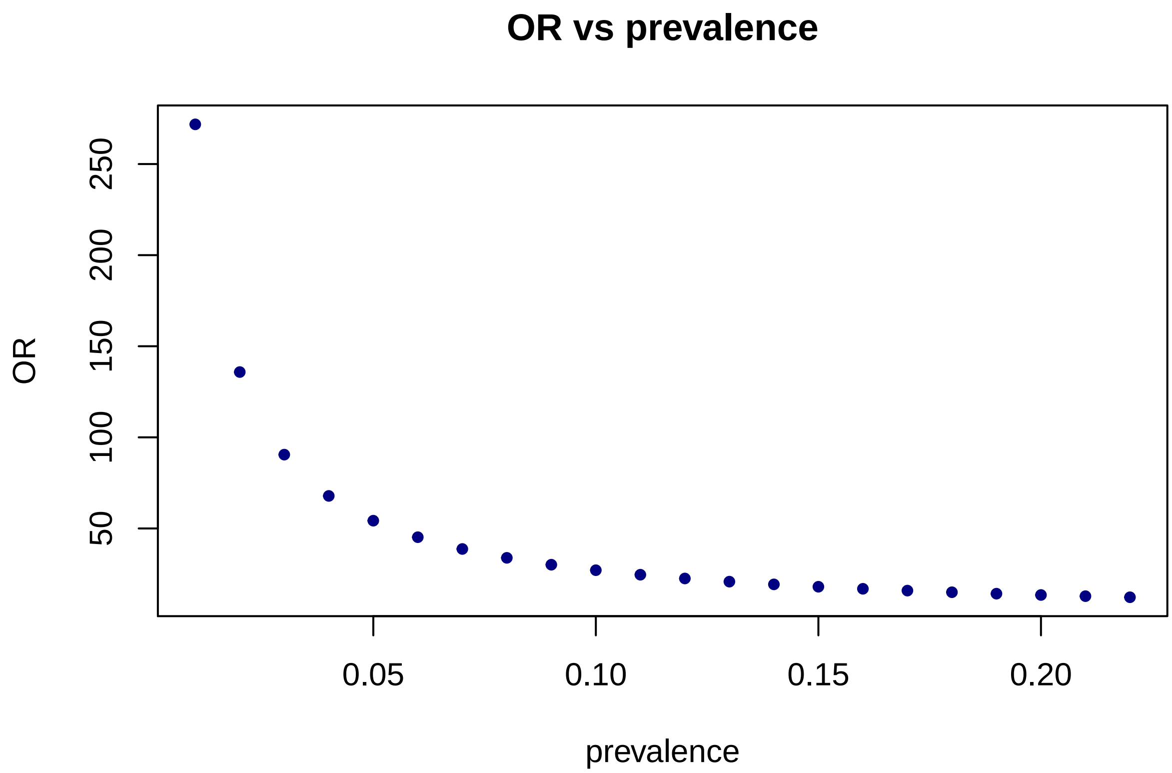



```r
# that pattern we see is due to the definition of chi and since we are keeping
# chi constant, the cancer risk in never-smokers must decrease when the
# prevalence decreases

plot(seq_p_s, seq_RR,
     pch=20, col="navyblue", xlab="prevalence",
     ylab="RR",
```

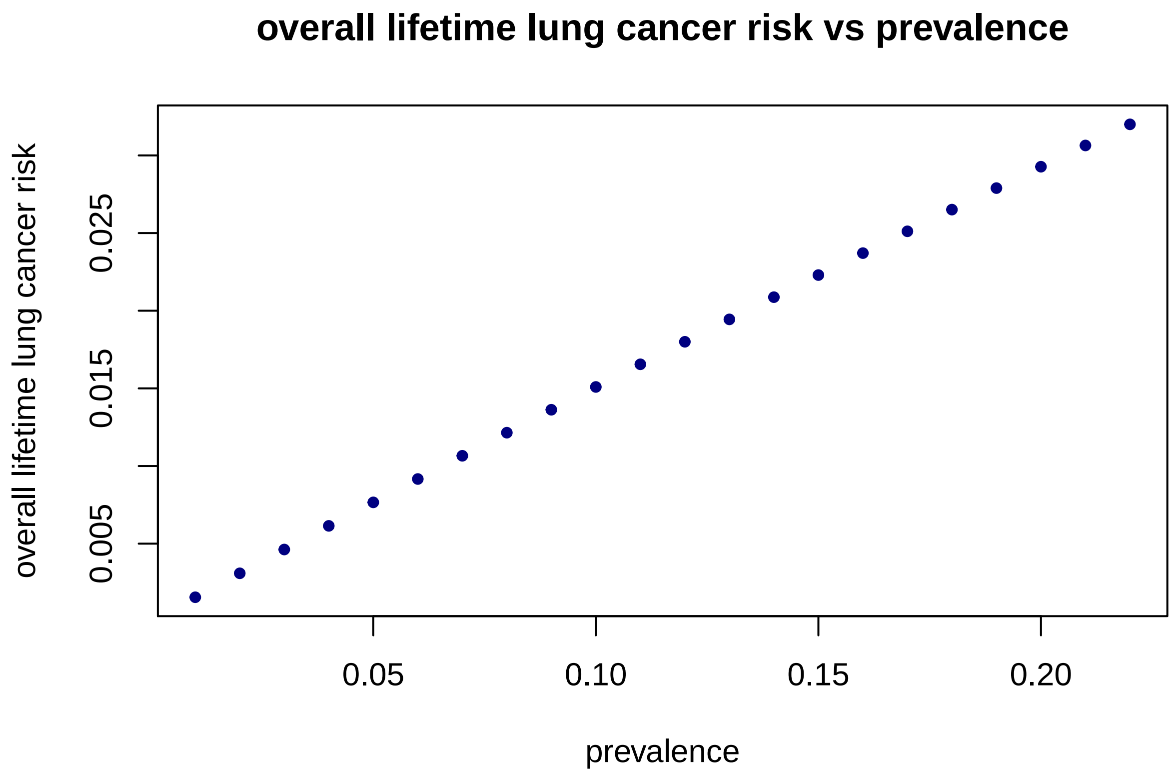
```
      main="RR vs prevalence")
```




```
# the RR increases for decreasing prevalence. THis is because for decreasing
# prevalence, we get a smaller cancer risk of non-smokers, which is the
# denominator of the RR.

plot(seq_p_s, seq_OR,
     pch=20, col="navyblue", xlab="prevalence",
     ylab="OR",
     main="OR vs prevalence")
```




```
# similar as for RR

plot(seq_p_s, seq_p_LC,
     pch=20, col="navyblue", xlab="prevalence",
     ylab="overall lifetime lung cancer risk",
     main="overall lifetime lung cancer risk vs prevalence")
```




```
# here we see that the overall lifetime lung cancer risk increases for increasing
# prevalence which seems reasonable.


# approximation of RR through OR
seq_RR - seq_OR
```

```
##  [1]  -1.235955  -1.300696  -1.371910  -1.450621  -1.538077  -1.635823
##  [7]  -1.745787  -1.870412  -2.012841  -2.177182  -2.368914  -2.595506
## [13]  -2.867416  -3.199750  -3.615169  -4.149278  -4.861423  -5.858427
## [19]  -7.353933  -9.846442 -14.831461 -29.786517
```

```
# hence the approximation gets worse for decreasing prevalence
```

Exercise 3

```
# Exercise 9.3

library(RcppAlgos)

po_1 <- c(14,0,1,2,3,1,10,9)
po_0 <- c(13,6,4,5,6,6,8,8)
# number of permutations
factorial(8) / (factorial(5) * factorial(3))
```

data <- data.frame("Y1" = c(14,0,1,2,3,1,10,9), "Y2" = c(13,6,4,5,6,6,8,8))

```
## [1] 56
```

num <- choose(8,3) # number of assignments

combinations <- combn(8,3)
mean.fx <- matrix(0, num)
median.fx <- matrix(0, num)

```r
# there are 56 permutations

# vector of treatment permutations
tr_per <- permuteGeneral(1:0, freq=c(3,5))
tr_per
```

```
##       [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
##  [1,]    1    1    1    0    0    0    0    0
##  [2,]    1    1    0    1    0    0    0    0
##  [3,]    1    1    0    0    1    0    0    0
##  [4,]    1    1    0    0    0    1    0    0
##  [5,]    1    1    0    0    0    0    1    0
##  [6,]    1    1    0    0    0    0    0    1
##  [7,]    1    0    1    1    0    0    0    0
##  [8,]    1    0    1    0    1    0    0    0
##  [9,]    1    0    1    0    0    1    0    0
## [10,]    1    0    1    0    0    0    1    0
## [11,]    1    0    1    0    0    0    0    1
## [12,]    1    0    0    1    1    0    0    0
## [13,]    1    0    0    1    0    1    0    0
## [14,]    1    0    0    1    0    0    1    0
## [15,]    1    0    0    1    0    0    0    1
## [16,]    1    0    0    0    1    1    0    0
## [17,]    1    0    0    0    1    0    1    0
## [18,]    1    0    0    0    1    0    0    1
## [19,]    1    0    0    0    0    1    1    0
## [20,]    1    0    0    0    0    1    0    1
## [21,]    1    0    0    0    0    0    1    1
## [22,]    0    1    1    1    0    0    0    0
## [23,]    0    1    1    0    1    0    0    0
## [24,]    0    1    1    0    0    1    0    0
## [25,]    0    1    1    0    0    0    1    0
## [26,]    0    1    1    0    0    0    0    1
## [27,]    0    1    0    1    1    0    0    0
## [28,]    0    1    0    1    0    1    0    0
## [29,]    0    1    0    1    0    0    1    0
## [30,]    0    1    0    1    0    0    0    1
## [31,]    0    1    0    0    1    1    0    0
## [32,]    0    1    0    0    1    0    1    0
## [33,]    0    1    0    0    1    0    0    1
## [34,]    0    1    0    0    0    1    1    0
## [35,]    0    1    0    0    0    1    0    1
## [36,]    0    1    0    0    0    0    1    1
## [37,]    0    0    1    1    1    0    0    0
## [38,]    0    0    1    1    0    1    0    0
## [39,]    0    0    1    1    0    0    1    0
## [40,]    0    0    1    1    0    0    0    1
## [41,]    0    0    1    0    1    1    0    0
## [42,]    0    0    1    0    1    0    1    0
## [43,]    0    0    1    0    1    0    0    1
## [44,]    0    0    1    0    0    1    1    0
## [45,]    0    0    1    0    0    1    0    1
## [46,]    0    0    1    0    0    0    1    1
## [47,]    0    0    0    1    1    1    0    0
```

```
## [48,]    0    0    0    1    1    0    1    0
## [49,]    0    0    0    1    1    0    0    1
## [50,]    0    0    0    1    0    1    1    0
## [51,]    0    0    0    1    0    1    0    1
## [52,]    0    0    0    1    0    0    1    1
## [53,]    0    0    0    0    1    1    1    0
## [54,]    0    0    0    0    1    1    0    1
## [55,]    0    0    0    0    1    0    1    1
## [56,]    0    0    0    0    0    1    1    1
```

```
sample_mean <- 1:56
sample_median <- 1:56

for (i in 1:56) {
  p <- tr_per[i,]
  mean0 <- mean(po_0[which(p==0)])
  mean1 <- mean(po_1[which(p==1)])
  sample_mean[i] <- mean1 - mean0
  med0 <- median(po_0[which(p==0)])
  med1 <- median(po_1[which(p==1)])
  sample_median[i] <- med1 - med0
}
sample_mean
```

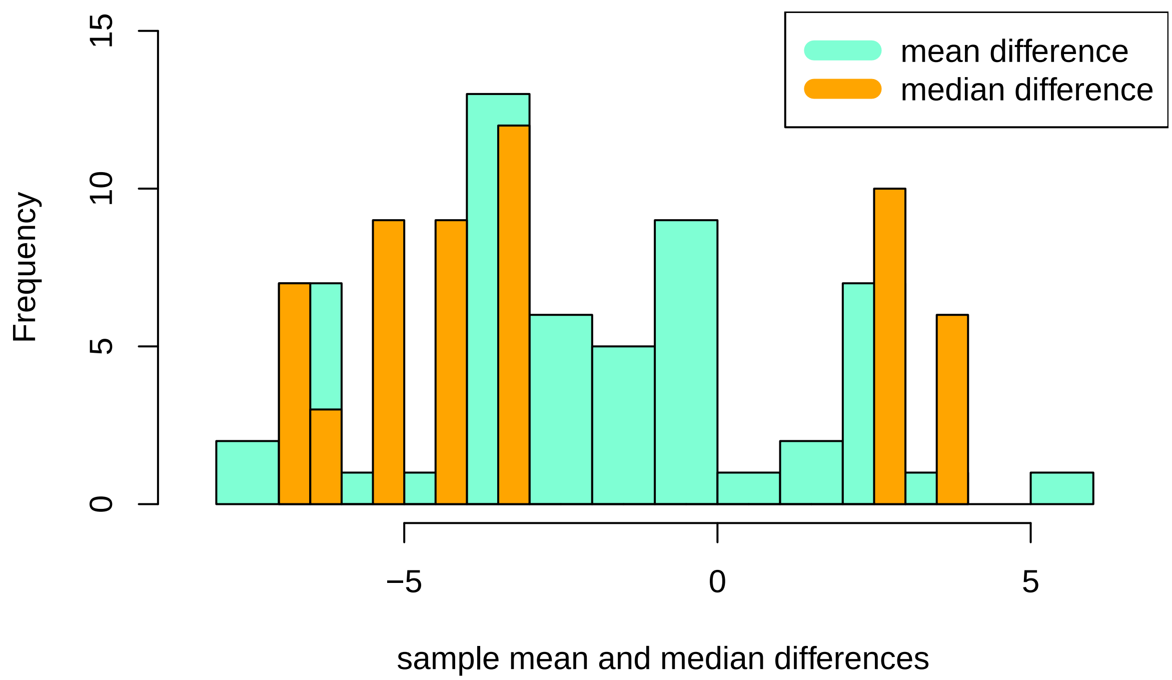
```
##  [1] -1.60000000 -1.06666667 -0.53333333 -1.20000000  2.20000000  1.86666667
##  [7] -1.13333333 -0.60000000 -1.26666667  2.13333333  1.80000000 -0.06666667
## [13] -0.73333333  2.66666667  2.33333333 -0.20000000  3.20000000  2.86666667
## [19]  2.53333333  2.20000000  5.60000000 -7.20000000 -6.66666667 -7.33333333
## [25] -3.93333333 -4.26666667 -6.13333333 -6.80000000 -3.40000000 -3.73333333
## [31] -6.26666667 -2.86666667 -3.20000000 -3.53333333 -3.86666667 -0.46666667
## [37] -6.20000000 -6.86666667 -3.46666667 -3.80000000 -6.33333333 -2.93333333
## [43] -3.26666667 -3.60000000 -3.93333333 -0.53333333 -5.80000000 -2.40000000
## [49] -2.73333333 -3.06666667 -3.40000000  0.00000000 -2.53333333 -2.86666667
## [55]  0.53333333 -0.13333333
```

```
sample_median
```

```
##  [1] -5 -4 -3 -5  4  3 -4 -3 -5  4  3 -3 -4  4  3 -3  4  3  4  3  4 -7 -7 -7 -5
## [26] -5 -6 -7 -4 -4 -7 -3 -3 -5 -5  3 -6 -7 -4 -4 -7 -3 -3 -5 -5  3 -6 -3 -3 -4
## [51] -4  3 -3 -3  3  3
```

```
min_mean <- min(sample_mean)
min_median <- min(sample_median)
low <- min(min_mean, min_median)

max_mean <- max(sample_mean)
max_median <- max(sample_median)
high <- max(max_mean, max_median)

hist(sample_mean,
     xlim = c(low-1, high+1),
     ylim = c(0, 15),
     xlab = "sample mean and median differences",
     freq = TRUE, col = "aquamarine", breaks = 16, main="")
hist(sample_median, freq=TRUE, col="orange", breaks=16, add=TRUE)
legend("topright", c("mean difference", "median difference"),
```

```
      col=c("aquamarine", "orange"), lwd=10)
```




```
# we see that the sample median is never zero or +- 1
# whereas the sample mean is dispersed along the whole x-range
```

Assignment 3

1 conditional exchangeability for each value of L

$L=0$

$$P[Y^a=1 \mid A=1, L=1] = \frac{0}{4} = 0$$
$$P[Y^a=1 \mid A=0, L=1] = \frac{2}{2} = 1$$
not equal

$$P[Y^a=0 \mid A=1, L=1] = \frac{3}{4}$$
$$P[Y^a=0 \mid A=0, L=1] = \frac{1}{2}$$
not equal

$$P[Y^a=1 \mid A=1, L=0] = \frac{2}{3}$$
$$P[Y^a=1 \mid A=0, L=0] = \frac{2}{3}$$
equal

$$P[Y^a=0 \mid A=1, L=0] = \frac{1}{3}$$
$$P[Y^a=0 \mid A=0, L=0] = \frac{1}{3}$$
equal

hence in the subset $L=1$, exchangeability does not hold

in the subset $L=0$, exchangeability holds

Thus conditional exchangeability does not hold

2

example on slide 17

- stratum specific risk difference

$\underline{L=1}$

causal

$$P(Y^{a=1}=1 \mid L=1) - P(Y^{a=0}=1 \mid L=1)$$

You should compute the causal risk difference using the potential outcomes

$$= P(Y=1 \mid L=1, A=1) - P(Y=1 \mid L=1, A=0)$$

$$= \frac{6}{9} - \frac{2}{3}$$

$$= 0$$

associational

$$P(Y=1 \mid A=1, L=1) - P(Y=1 \mid A=0, L=1)$$

$$= \frac{6}{9} - \frac{2}{3}$$

$$= 0$$

$\underline{L=1:}$

causal

$$P(Y^{a=1}=1 \mid L=0) - P(Y^{a=0}=1 \mid L=0)$$

$$\overset{(*)}{=} P(Y=1 \mid L=0, A=1) - P(Y=1 \mid L=0, A=0)$$

$$= \frac{1}{4} - \frac{1}{4}$$

$$= 0$$

associational (*)

You should check this equality

- stratum specific odds ratio

L=0:

causal

$$\frac{P(Y^{a=1}=1 \mid L=0) \,/\, P(Y^{a=1}=0 \mid L=0)}{P(Y^{a=0}=1 \mid L=0) \,/\, P(Y^{a=0}=0 \mid L=0)}$$

$$\overset{(*)}{=} \frac{P(Y=1 \mid L=0, A=1) \,/\, (1 - P(Y=1 \mid L=0, A=1))}{P(Y=0 \mid L=0, A=0) \,/\, (1 - P(Y=0 \mid L=0, A=0))}$$

$$= \frac{\frac{1}{4} \,/\, \frac{3}{4}}{\frac{3}{4} \,/\, \frac{1}{4}}$$

$$= 1$$

associational: see (*)

L=1:

causal

$$\frac{P(Y^{a=1}=1 \mid L=1) \,/\, P(Y^{a=1}=0 \mid L=1)}{P(Y^{a=0}=1 \mid L=1) \,/\, P(Y^{a=0}=0 \mid L=1)}$$

$$\overset{(*)}{=} \frac{P(Y=1 \mid L=1, A=1) \,/\, (1 - P(Y=1 \mid L=1, A=1))}{P(Y=0 \mid L=1, A=0) \,/\, (1 - P(Y=0 \mid L=1, A=0))}$$

$$= \frac{6/9 \,/\, 3/9}{1/3 \,/\, 2/3}$$

$$= 1$$

associational: see (*)

hence for both measure, the causal form is correctly evaluated by the associational form

Compute it using the potential outcomes, then the associational risk and check if the two results are the same

3

Consider the following two conditions:

(1) $Y^a \perp\!\!\!\perp L \quad \forall a$

(2) $Y^a \perp\!\!\!\perp A | L \quad \forall a, l$

claim: (1) and (2) imply marginal exchangeability

proof: marginal exchangeability is satisfied if $Y^a \perp\!\!\!\perp A \quad \forall a$

from (2) we get:

$$P[Y^a = 1 | A = 1, L = l] = P[Y^a = 1 | A = 0, L = l] \quad \forall a, l \quad (*)$$

from (1) we get

No, this is Y^a conditionally independent from L given A

$$P[Y^a = 1 | A = x, L = l] = P[Y^a = 1 | A = x, L = l'] \quad \forall x \in A \; \forall l, l' \in L \; (*)$$

combining $(*)$ and $(*)$ it follows

$$P[Y^a = 1 | A = 1, L = l] \overset{(*)}{=} P[Y^a = 1 | A = 0, L = l]$$

$$\overset{(*)}{=} P[Y^a = 1 | A = 0, L = l']$$

$$\overset{(*)}{=} P[Y^a = 1 | A = 0, L = l]$$

$$\Rightarrow P[Y^a = 1 | A = 1] = P[Y^a = 1 | A = 0]$$

You should marginalize (sum) over L to show this

$$\Rightarrow Y^a \perp\!\!\!\perp A \quad \forall a$$

●

authors argue that it is not possible to estimate effect of obesity on mortality without consistency

obesity depends on many factors and it is impossible to control all these

Exercise 4

What is the causal question?

Does obesity shorten life? effect of obesity on mortality

Which causal question did each of the (three) randomised experiment answer?

1. Does intense exercise of 1h per day influence BMI distribution and mortality rate?
2. Does intake of calories and carbohydrates influence BMI distribution?
3. Does the combination of exercise and dietary intervention influence BMI distribution?

Which identifiability condition does the causal question about obesity violate?

consitency definition: the observed outcomes are the potential outcomes

Why is consistency a trivial condition for randomised experiments but not for observational ones?

Consistency is the property according to which for every study unit i the observed outcome coincides with the potential outcome corresponding to the actual treatment received. This is trivial in randomised experiments because we know the treatment and can observe the outcome.

here, the treatment is not-well defined because there are different paths with different effects
=> we do not know to which cause to attribute the effect

In an observational study, the potential outcome is not necessarily equal to the observed outcome, because we do not know the procedure which led to the outcome.
The counterfactual outcome is a very vague concept then.

Does the observational study answer a valid question?

What does it mean that a potential outcome is vague and what are the implications for any causal contrasts?

It means that we do not know the procedure which led to the outcome.
Vague counterfactual outcomes lead to ill-defined causal contrasts involving that counterfactual outcome.

In which way may violations of consistency complicate the achievement of conditional exchangeability?

Lack of consistency makes it hard to avoid confounding. We would need to measure all possible confounding factors in order to achieve conditional exchangeability.

In which way may violations of consistency lead to lack of positivity and what is it meant with lack of generalizability?

Confounding can lead to the situation where some strata defined by the confounders do not fullfill the requirement of positivity.

Which types of exposures pose higher challenges to inform policy making?

Those for which the relevant interventions are not clear or those which are not easy to measure.

Assignment 4

$\boxed{1}$

slide 16 ,

L: prognostic factor
A: treatment
Y: outcome

$P(Y^{a=1}=1)$

We use standardization here

$$P(Y^{a=1}=1) = \sum_{\ell} P(Y=1 \mid L=\ell, A=1)\, P(L=\ell)$$

$$= \underbrace{P(Y=1 \mid L=0, A=1)}_{=3/8} \cdot \underbrace{P(L=0)}_{16/40} + \underbrace{P(Y=1 \mid L=1, A=1)}_{12/18} \cdot \underbrace{P(L=1)}_{=24/40}$$

$$= \frac{3}{\cancel{8}} \cdot \frac{\overset{2}{\cancel{16}}}{40} + \frac{12}{18} \cdot \frac{24}{40}$$

$$= \frac{6}{40} + \frac{2}{\cancel{3}} \cdot \frac{\cancel{3}}{5}$$

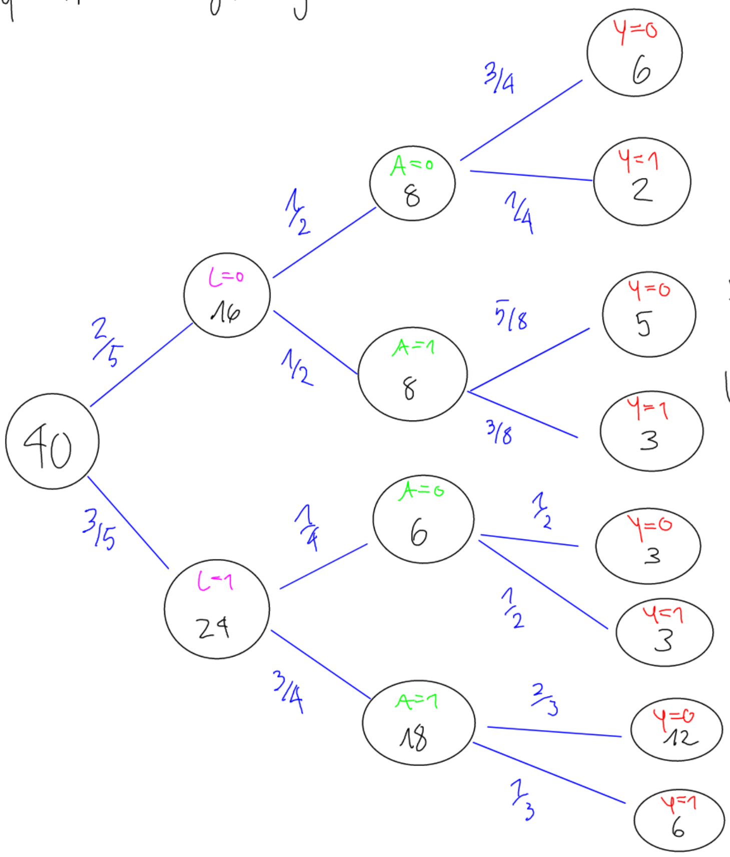
$$= \frac{3}{20} + \frac{2}{5}$$

$$= \frac{3}{20} + \frac{8}{20}$$

$$= \frac{11}{20}$$

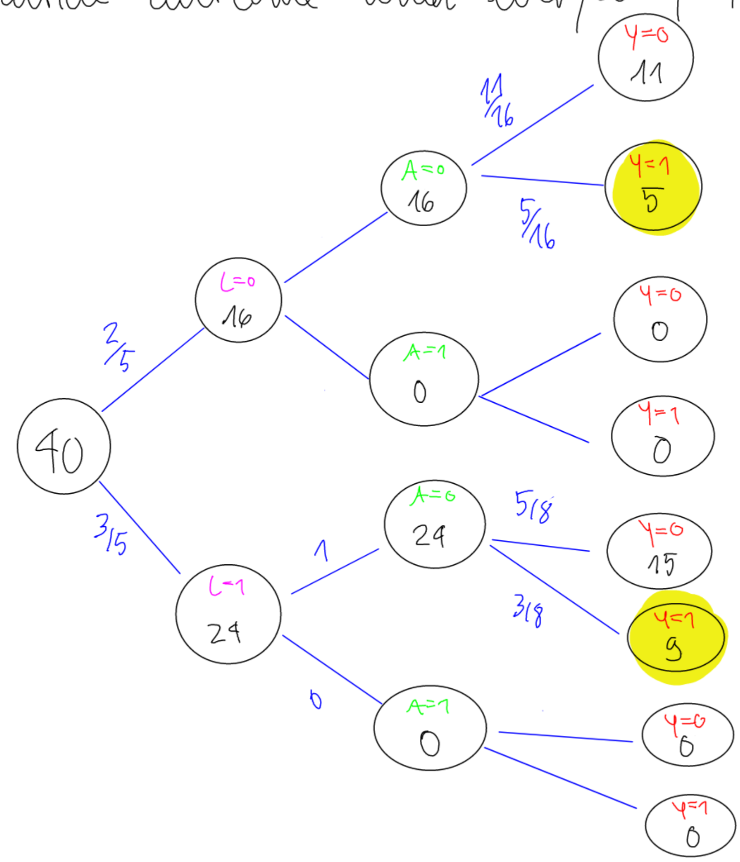
$P(Y^{a=0}=1)$

by IP-weighting



cond. exch. ensures that if the treated had been left untreated, they would experience same risk of adverse events as the units actually left untreated

potential outcome when everybody is untreated



conditional exchangeability ensures that the probabilities on the edges branching off from the nodes corresponding to state A of the treatment assignment are left unchanged

hence $P(Y^{a=0}=1) = \frac{5+9}{40} = \frac{14}{40} = \frac{7}{20}$

Algebra error; use IPW formula

$\boxed{2}$

Roman population on slide 16

standardisation:

$$P(Y^{a=1} = 1 \mid V=0) = \sum_{\ell} P(Y=1 \mid L=\ell, A=1, V=0)\, P(L=\ell)$$

$$= \underbrace{P(Y=1 \mid L=0, A=1, V=0)}_{=\frac{2}{4}} \underbrace{P(L=0 \mid V=0)}_{=\frac{8}{20}} + \underbrace{P(Y=1 \mid L=1, A=1, V=0)}_{=\frac{6}{9}} \underbrace{P(L=1 \mid V=0)}_{=\frac{12}{20}}$$

$$= \frac{2}{4} \cdot \frac{8}{20} + \frac{6}{9} \cdot \frac{12}{20}$$

$$= \frac{1}{2} \cdot \frac{2}{5} + \frac{2}{3} \cdot \frac{3}{5}$$

$$= \frac{1}{5} + \frac{2}{5}$$

$$= \frac{3}{5}$$

$= 0.6$ hence the same result as on slide 20 ✓

$$P(Y^{a=0} = 1 \mid V=0) = \sum_{\ell} P(Y=1 \mid L=\ell, A=0, V=0)\, P(L=\ell)$$

$$= \underbrace{P(Y=1 \mid L=0, A=0, V=0)}_{=\frac{1}{4}} \underbrace{P(L=0 \mid V=0)}_{=8/20} + \underbrace{P(Y=1 \mid L=1, A=0, V=0)}_{=\frac{1}{3}} \underbrace{P(L=1 \mid V=0)}_{=\frac{12}{20}}$$

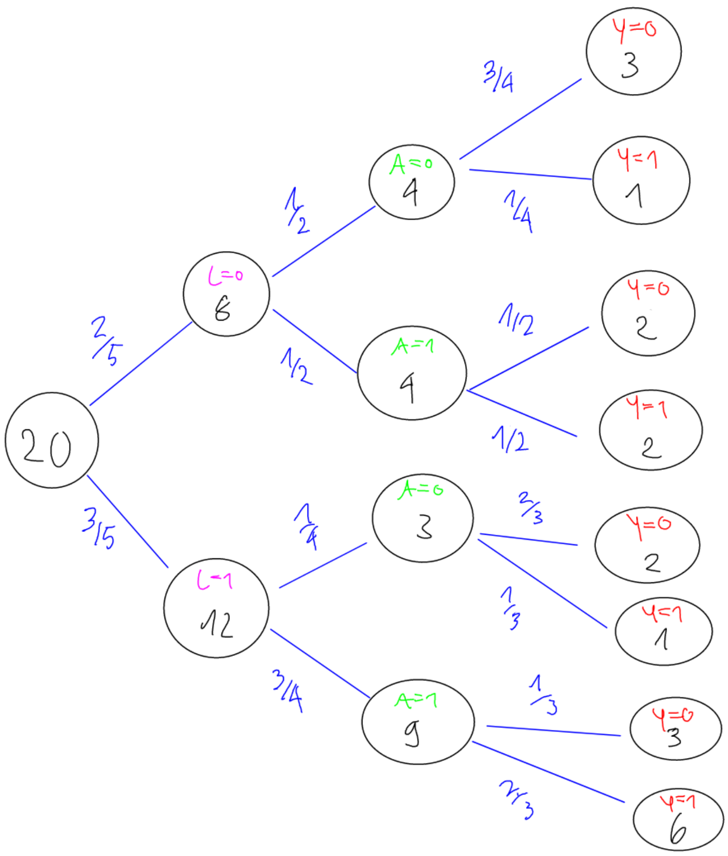
$$= \frac{1}{4} \cdot \frac{8}{20} + \frac{1}{3} \cdot \frac{12}{20}$$

$$= \frac{1}{10} + \frac{2}{10}$$

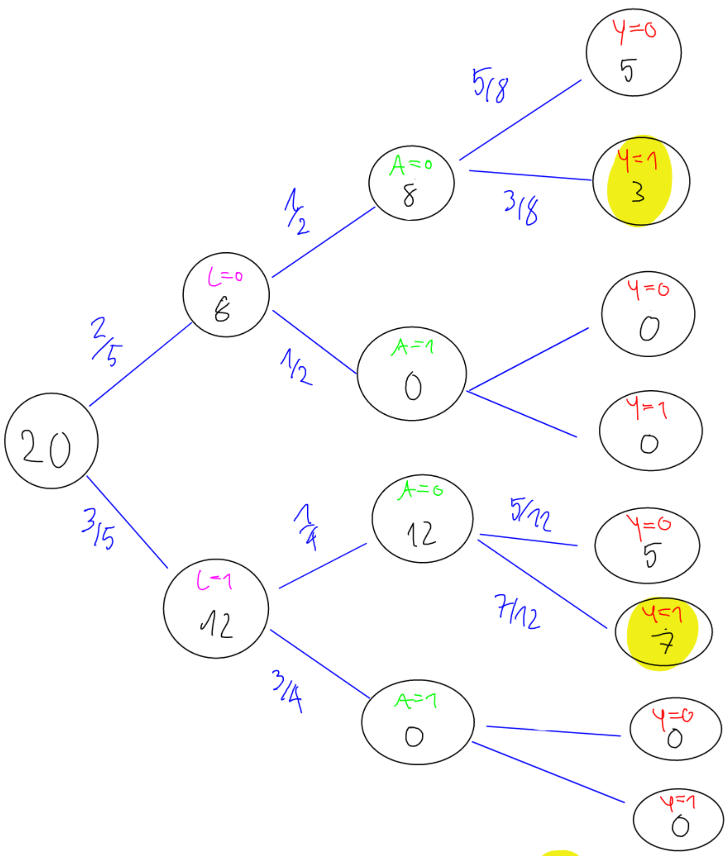
$$= \frac{3}{10}$$

$= 0.3$ hence the same result as on slide 20 ✓

IP-weighting

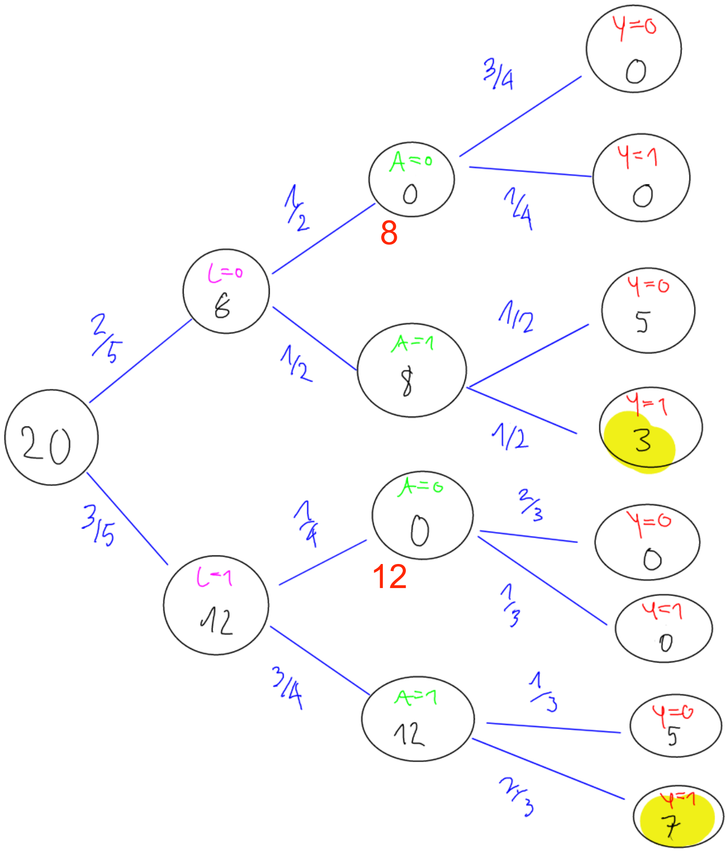


potential outcome when everybody is untreated



hence $P(Y^{a=0}=1) = \frac{7+3}{20} = \frac{14}{40} = \frac{7}{20}$

potential outcome when everybody is treated



hence $P(Y^{a=0}=1) = \frac{7+3}{20} = \frac{14}{40} = \frac{7}{20}$

↳ both my IP calculations do <u>not</u> yield the numbers of slide 20.

I could not find my mistake ??

The tree representation is meant to explain the rationale behind IPW, it's easy to make mistakes: to compute the causal effects use the formula population with twice as many Roman Gods as Greek Gods,

we would get the same

causal risk ratio because:

- the probabilities are frequencies, hence do no not
  But you are doubling only a subset of the population
  change when doubling

- if we double the number of Roman Gods, then
  the total population will increase by $\frac{1}{3}$

Hence the causal risk ratio will stay the same

no, we are not increasing the population uniformly by 1/3

But the causal risk difference changes because

it is a sum No

3

claim 1: If $P(Y=1 \mid A=1, Z=z) > P(Y=1 \mid A=0, Z=z)$, then the treatment is harmful in units with $Z=z$

answer: False

$P(Y=1 \mid A=1, Z=z) > P(Y=1 \mid A=0, Z=z)$ implies the presence of an association

but only in marginally randomised studies association implies causal effect.

claim 2: If $Y^a \perp\!\!\!\perp A \mid Z$ then $P(Y=1 \mid A=1) > P(Y=1 \mid A=0)$ implies that the treatment is harmful in the population

answer: False

see claim 5

claim 3: If $P(Y=1 \mid A=1) > P(Y=1 \mid A=0)$ then the treatment is harmful

answer: FALSE

See answer to claim 1

claim 4: The treatment is harmful in units with $Z=z$ if

$$P(Y^{a=1}=1 \mid Z=z) > P(Y^{a=0}=1 \mid Z=z)$$

answer: TRUE

$P(Y^{a=1}=1 \mid Z=z) > P(Y^{a=0}=1 \mid Z=z)$ means there is causal effect in the population with $Z=z$

claim 5: If $Y^a \perp\!\!\!\perp A \mid Z$ then $P(Y=1 \mid A=1, Z=z) > P(Y=1 \mid A=0, Z=z)$ implies that the treatment is harmful in units with $Z=z$

answer: TRUE

$$
\begin{aligned}
P(Y^{a=1}=1 \mid Z=z) &= P(Y=1 \mid A=1, Z=z) \\
&\overset{\text{cond. exch.}}{=} P(Y=1 \mid A=1) \\
&> P(Y=1 \mid A=0 \\
&= P(Y^{a=0}=1 \mid Z=z)
\end{aligned}
$$

claim 6: The treatment is harmful for the population if $P(Y^{a=1}=1) > P(Y^{a=0}=1)$

answer: TRUE

here we have a population level causal effect

claim 7: If $Y^a \perp\!\!\!\perp A \mid Z$ then $P(Y=1 \mid A=1, Z=z)$ takes the same value for all levels of $Z=z$

answer: TRUE X We can have effect modification claim 8: It is possible for a treatment to be beneficial for all levels $Z=z$ and harmful for the whole population

answer: False

claim 9: It is possible for a treatment to have non-null effects in several levels of $Z=z$ and have no effect in the whole population

answer: TRUE

the effects in different levels of $Z=z$ can neutralize each other

claim 10: If we observe a treatment imbalance $P(Z=z \mid A=1) \neq P(Z=z \mid A=0)$ in certain levels of $Z=z$, then marginal exch. does not hold

answer: TRUE

the concerned levels $Z=z$ are not exchangeable

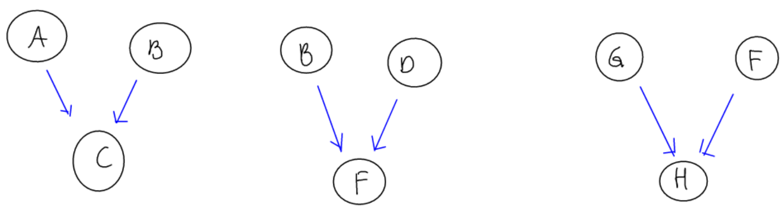
No, adjusting for Z takes this imbalance into account

Assignment 5

1

- Γ is a DAG since it is a directed graph and does not contain any cycles

- a v-structure is a special type of collider when its parents are not adjacent



hence 3 v-structures

- $Pa(H) = \{G, F\}$

- $nbd(C) = \{A, B, E\}$

joint distribution $P(A, B, C, \dots, H)$ which obeys the local Markov Property described in I

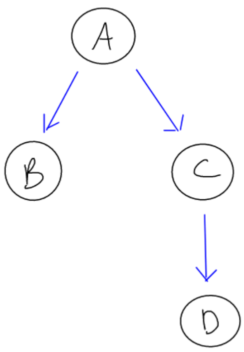
- $G \perp\!\!\!\perp C \mid A$ due to local Markow property
  $C \perp\!\!\!\perp G \mid A$ ✓
  What about not conditional on A?
- $A \not\!\perp\!\!\!\perp H$ but $H \perp\!\!\!\perp A \mid G$
- $B \not\!\perp\!\!\!\perp F \mid C, D$ since $F \notin Nd(B)$
- X $G \perp\!\!\!\perp F \mid A, B, H$ since $F \in Nd(G)$ and $Pa(G) = \{A\}$ and $A \in \{A, B, H\}$
  But H is a collider
- $P(A, B, C, \dots, H) = \prod_{i=A}^{H} P(i \mid Pa(i))$

$= P(A) \cdot P(B) \cdot P(C \mid A, B) \cdot P(D) \cdot P(E \mid C) \cdot P(F \mid B, D) \cdot P(G \mid A) \cdot P(H \mid G, F)$

2

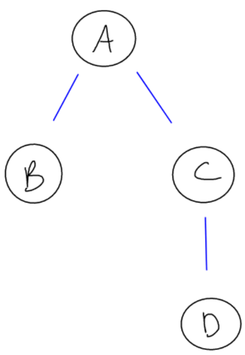
$P(A,B,C,D) = P(B|A) \cdot P(C|A) \cdot P(D|C) \cdot P(A)$

- DAG with respect to which P satisfies local Markov property



here we do not have any v-structures

- CPDAG



Two DAG's are in the same equivalence class if they share a common skeleton and the same set of v-structures (0 in this case)

in the associated Markov equivalence class we have $2^3 = 8$ DAG's since we have 3 edges all of which have two possible directions

But some of these 8 DAGs have v-structures… There are only 4 in this case

•

claim: $(A \perp\!\!\!\perp D \mid C)_P$

proof: $P(A, D \mid C) = \dfrac{P(A, C, D)}{P(C)}$

$$= \frac{P(C|A) \cdot P(D|C) \cdot P(A)}{P(C)}$$

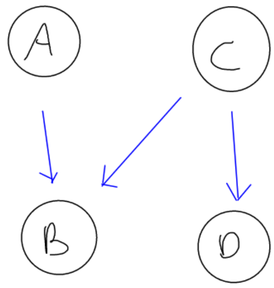
$$= \frac{P(C|A) \cdot P(A)}{P(C)} \cdot P(D|C)$$

$$\overset{\text{Bayes}}{=} P(A|C) \cdot P(D|C)$$

$$\Longrightarrow (A \perp\!\!\!\perp D \mid C)_P$$

$$P(A,B,C,D) = P(A)\,P(B|A,C)\,P(C)\,P(D|C)$$

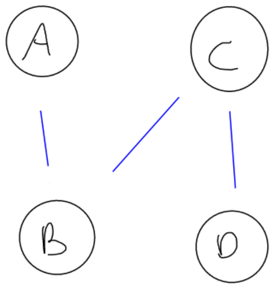
- DAG with respect to which P satisfies local Markov property



no v-structures

A-B-C is a v-structure!

- CPDAG



no v-structures possible

hence the associated Markov equivalence class contains

$2^3 = 8$ DAGs

Not without creating additional v-structures

claim: A and D are marginally independent

proof:

$$P(A,D) = \sum_b \sum_c P(A,B,C,D)$$

$$= \sum_b \sum_c P(A)\,P(B|A,C)\,P(C)\,P(D|C)$$

$$= P(A) \cdot \underbrace{\sum_b P(B|A,C)}_{=1} \underbrace{\sum P(D|C)\,P(C)}_{=P(D)}$$

$$= P(A) \cdot P(D)$$

$\Longrightarrow$ A and D are marginally independent

$\boxed{3}$

$$X \leftarrow Z \leftarrow Y \qquad P(X,Y,Z) = P(Y)\,P(Z|Y)\cdot P(X|Z)$$

$X \perp\!\!\!\perp Y \mid Z$:

$$P(X,Y \mid Z) = \frac{P(X,Y,Z)}{P(Z)} = \frac{P(Y)\,P(Z|Y)\cdot P(X|Z)}{P(Z)} \overset{\text{Bayes}}{=} P(Y|Z)\cdot P(X|Z)$$

$$X \rightarrow Z \rightarrow Y \qquad P(X,Y,Z) = P(Y|Z)\cdot P(Z|X)\cdot P(X)$$

$Y \perp\!\!\!\perp X \mid Z$

$$P(Y,X \mid Z) = \frac{P(X,Y,Z)}{P(Z)} = \frac{P(Y)\,P(Z|Y)\cdot P(X|Z)}{P(Z)} \overset{\text{Bayes}}{=} P(Y|Z)\cdot P(X|Z)$$

$$X \leftarrow Z \rightarrow Y \qquad P(X,Y,Z) = P(X|Z)\cdot P(Y|Z)\cdot P(Z)$$

Z has no non-descendants

$X \perp\!\!\!\perp Y \mid Z$ and $Y \perp\!\!\!\perp X \mid Z$

$$P(X,Y \mid Z) = \frac{P(X,Y,Z)}{P(Z)} = \frac{P(X|Z)\cdot P(Y|Z)\cdot P(Z)}{P(Z)} = P(X|Z)\cdot P(Y|Z)$$

Show that these distributions are equivalent

- $X \longrightarrow Z \longleftarrow Y$ $\quad P(X,Y,Z) = P(X) \cdot P(Y) \cdot P(Z \mid X, Y)$

$X \perp\!\!\!\perp Y \;,\; Y \perp\!\!\!\perp X \;,\; Z \perp\!\!\!\perp Y \mid X \;,\; Z \perp\!\!\!\perp X$

$Z \perp\!\!\!\perp X \mid Y \;,\; Z \perp\!\!\!\perp Y$

Independence is always symmetric, no need to be redundant

hence we do not have the same distribution

Assignment 6

1

1. $\overset{A}{X} = D_1$, $Z = D_3$

priors: $P(Y = D_1) = P(Y = D_2) = P(Y = D_3) = \frac{1}{3}$

likelihood: $P(Z = D_3 \mid Y = D_1) = \frac{1}{2}$ This doesn't really make sense, how it this 0.5?

$P(Z = D_3 \mid Y = D_2) = 1$

normalizing constant: $P(Z = D_3) = 1$

posterior: $P(Y = D_1 \mid Z = D_3) = \dfrac{P(Z = D_3 \mid Y = D_1) \cdot P(Y = D_1)}{P(Z = D_3)} = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$

The probability of my initial choice being correct must be 1/3…

$$P(Y = D_2 \mid Z = D_3) = \frac{P(Z = D_3 \mid Y = D_2) \cdot P(Y = D_2)}{P(Z = D_3)} = 1 \cdot \frac{1}{3} = \frac{1}{3}$$

Hence we should switch doors

2.

$$\begin{tikzcd} X \arrow[dr] & & Y \arrow[dl] \\ & Z & \end{tikzcd}$$

$P(X, Y, Z) = P(X) \cdot P(Y) \cdot P(Z \mid X, Y)$

Z is a collider
hence once we know the value of Z, all probabilities become conditional on this information. Thus X and Y are not independent anymore.
This is causeless correlation

2

gender:

$$
\begin{tikzcd}
 & \text{Gender} \arrow[dl] \arrow[dr] & \\
\text{treatment} \arrow[rr] & & \text{outcome}
\end{tikzcd}
$$

causal risk diff.: we use IP-weighting

$$P(Y^{a=1}=1) = \sum_{l} \frac{P(Y=1,\ A=1,\ L=l)}{P(A=1 \mid L=l)}$$

$$= \frac{P(Y=1, A=1, L=1)}{P(A=1 \mid L=1)} + \frac{P(Y=1, A=1, L=0)}{P(A=1 \mid L=0)}$$

why only 6? there are 81

$$= \frac{6/357}{87/357} + \frac{71/343}{263/343}$$

X

The numbers are wrong

$$= \frac{6}{87} + \frac{71}{263}$$

$$\approx 0.34$$

$L=0$: female
$L=1$: male

$$P(Y^{a=1}=0) = \sum_{l} \frac{P(Y=1,\ A=0,\ L=l)}{P(A=0 \mid L=l)}$$

$$= \frac{P(Y=1, A=0, L=1)}{P(A=0 \mid L=1)} + \frac{P(Y=1, A=0, L=0)}{P(A=0 \mid L=0)}$$

$$= \frac{36/357}{270/357} + \frac{25/343}{80/343}$$

Same

$$= \frac{36}{270} + \frac{25}{80}$$

$$= \frac{2}{15} + \frac{5}{16} \approx 0.45$$

hence the causal risk difference is

$$P(Y^{a=1}=1) - P(Y^{a=0}=1) = 0.34 - 0.45 = -0.11$$ X

blood pressure

$$\begin{tikzcd} & \text{blood pressure} \arrow[dr] & \\ \text{treatment} \arrow[ur] \arrow[rr] & & \text{outcome} \end{tikzcd}$$

You shouldn't adjust for blood pressure, both paths combined represent the causal effect

$L=0$: low BP
$L=1$: high BP

causal risk diff.

standardisation:

$$P(Y^{a=1}=1) = \sum_{\ell} P(Y=1 \mid A=1, L=\ell) \cdot P(L=\ell)$$

$$= P(Y=1 \mid A=1, L=0) \cdot P(L=0) + P(Y=1 \mid A=1, L=1) \cdot P(L=1)$$

$$= \frac{36}{270} \cdot \frac{357}{700} + \frac{25}{80} \cdot \frac{343}{700}$$

$$\approx 0.068 + 0.15$$

$$\approx 0.22$$

$$P(Y^{a=0}=1) = \sum_{\ell} P(Y=1 \mid A=0, L=\ell) \cdot P(L=\ell)$$

$$= P(Y=1 \mid A=0, L=0) \cdot P(L=0) + P(Y=1 \mid A=0, L=1) \cdot P(L=1)$$

$$= \frac{6}{87} \cdot \frac{357}{700} + \frac{71}{263} \cdot \frac{343}{700}$$

$$\approx 0.035 + 0.13$$

$$\approx 0.17$$

hence $P(Y^{a=1}=1) - P(Y^{a=0}=1) = 0.22 - 0.17 = 0.05$ X

3

- Diet quality $\perp\!\!\!\perp$ Amount of free time and thus also d-separated
  since the only path goes through the collider "freq. of ex."

- Am. of free time $\perp\!\!\!\perp$ level of chol. | BMI cond. on BMI has no effect on the path
  since the only path goes through the collider "freq. of ex."

local Markow property: d-sep => cond. indep.
d-sep is property of DAG, cond.indep. is property of distribution

local Mark.cond. + faithfulness:
then cond. indep <=> d-sep

- if we condition on the non-collider diet quality, then every path from health consc. to BMI is blocked correct until here
  hence diet quality is the smallest set that d-separates health. cons. and BMI

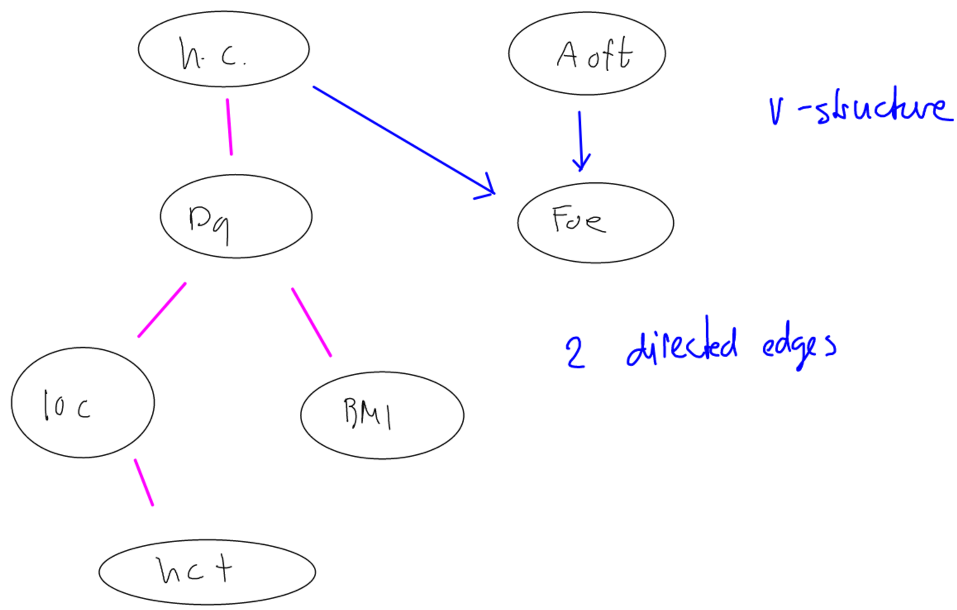
  faithfulness: guarantees that conditional independence relations entailed in G also hold in P.

  So if P were not faithful to the DAG, this could change the answer because d-separation would not imply independence

X

No, d-separation always implies cond. independence (this is local markow property)

- CPDAG



correct

- it would be compatible with the given structure because we only have association between frequency of exercise and cholesterol test, but we do not know in which direction.

we have an open path connecting high chol. test and freq. of exercise, hence yes

- there could be a zero-causal-effect because from the diagram we only have association, not necessarily causation

we even expect a zero causal effect

- if we are interested in particular in diet quality as pot. causal effect, we could ignore amount of free time and frequency of exercise

Also Health consciousness

hence 3 variables

- interventions on diet quality are hard to define because they are not easy to quantify and realize consequently

Modified causal diagram

- diet quality $\perp\!\!\!\perp$ amount of free time since collider freq. of ex. on the path

  correct

- Am. of free time $\not\perp\!\!\!\perp$ level of alcohol. | BMI via conditioning the collider BMI is opened

  correct

- health consc. $\perp\!\!\!\perp$ BMI | Freq. of ex., Diet quality

  P(health consc., BMI | Freq. of ex, Diet quality)

  = P(health consc. | Freq. of ex, Diet quality)

  · P(BMI | Freq. of ex, Diet quality)

  true

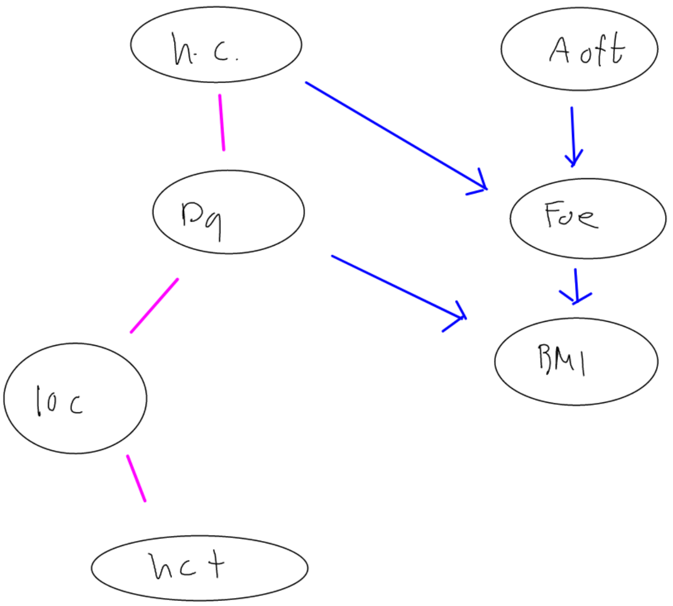
  conditioning on diet quality and freq.of exercise

X if P were not faithful to G, we could not infer anything on the distribution

No, we can still infer all the CI relationships that are implied by d-seperation

same as before

- CPDAG



2 v-structures

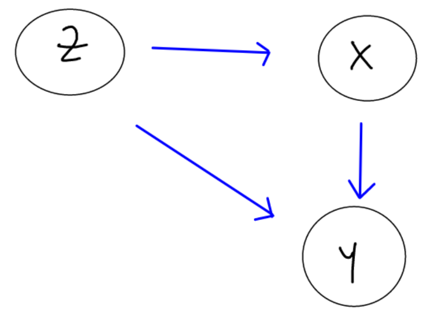
4 directed edges

- same answers as before
- same answers as before
- we could ignore Amount of free time

correct

Assignment 7

1



a. $P(x,y,z) = P(X|Z) \cdot p(Y|Z,X) \cdot P(Z)$

now we look at the probabilities.

	value	probability
Z:	z_1	r
	z_2	$1-r$
X:	$x_1 \mid z_2$	q_1
	$x_1 \mid z_1$	q_2
	$x_2 \mid z_2$	$1-q_1$
	$x_2 \mid z_1$	$1-q_2$
Y:	$y_1 \mid x_2, z_2$	p_1
	$y_1 \mid x_1, z_2$	p_2
	$y_1 \mid x_1, z_1$	p_4
	$y_1 \mid x_2, z_1$	p_3
	$y_2 \mid x_2, z_2$	$1-p_1$
	$y_2 \mid x_1, z_2$	$1-p_2$
	$y_2 \mid x_1, z_1$	$1-p_4$
	$y_2 \mid x_2, z_1$	$1-p_3$

now we can write down the combinations

1: $P(y_1 | x_2, z_2) \cdot P(x_2 | z_2) \cdot P(z_2) = p_1 \cdot (1-q_1) \cdot (1-r)$

2: $P(y_1 | x_1, z_2) \cdot P(x_1 | z_2) \cdot P(z_2) = p_2 \cdot q_1 \cdot (1-r)$

3: $P(y_1 | x_2, z_1) \cdot P(x_2 | z_1) \cdot P(z_1) = p_3 \cdot (1-q_2) \cdot r$

4: $P(y_1 | x_1, z_1) \cdot P(x_1 | z_1) \cdot P(z_1) = p_4 \cdot q_2 \cdot r$

5: $P(y_2 | x_2, z_2) \cdot P(x_2 | z_2) \cdot P(z_2) = (1-p_1) \cdot (1-q_1) \cdot (1-r)$

$P(y_2 | x_1, z_2) \cdot P(x_1 | z_2) \cdot P(z_2) = (1-p_2) \cdot q_1 \cdot (1-r)$

$P(y_2 | x_2, z_1) \cdot P(x_2 | z_1) \cdot P(z_1) = (1-p_3) \cdot (1-q_2) \cdot r$

$P(y_2 | x_1, z_1) \cdot P(x_1 | z_1) \cdot P(z_1) = (1-p_4) \cdot q_2 \cdot r$

b. $RD = P(Y=y_1 | X=x_1) - P(Y=y_1 | X=x_2)$

$Z=z_1$: $RD_1 = P(Y=y_1 | X=x_1, Z=z_1) - P(Y=y_1 | X=x_2, Z=z_1) = p_4 - p_3$

$Z=z_2$: $RD_2 = P(Y=y_1 | X=x_1, Z=z_2) - P(Y=y_1 | X=x_2, Z=z_2) = p_2 - p_1$

whole population: $P(Y=y_1 | X=x_1) = \frac{P(Y=y_1, X=x_1, Z)}{P(X=x_1, Z)} = \frac{\sum_z P(Y=y_1 | X=x_1, z) \cdot P(X=x_1 | z) \cdot P(z)}{\sum_z P(X=x_1 | z) \cdot P(z)}$

$$= \frac{p_4 \cdot q_2 \cdot r + p_2 \cdot q_1 \cdot (1-r)}{q_1 \cdot (1-r) + q_2 \cdot r}$$

$P(Y=y_1 | X=x_0) = \frac{P(Y=y_1, X=x_0, Z)}{P(X=x_0, Z)} = \frac{\sum_z P(Y=y_1 | X=x_0, z) \cdot P(X=x_0 | z) \cdot P(z)}{\sum_z P(X=x_0 | z) \cdot P(z)}$

$$= \frac{p_1 \cdot (1-q_1) \cdot (1-r) + p_3 \cdot (1-q_2) \cdot r}{(1-q_1)(1-r) + (1-q_2) r}$$

$$\Rightarrow RD_3 = \frac{p_4 \cdot q_2 \cdot r + p_2 \cdot q_1 \cdot (1-r)}{q_1 \cdot (1-r) + q_2 \cdot r}$$

$$- \frac{p_1 \cdot (1-q_1) \cdot (1-r) + p_3 \cdot (1-q_2) \cdot r}{(1-q_1)(1-r) + (1-q_2) r}$$

Simpson: find a combination of parameters which exhibit Simpson's reversal

here, the syndrome Z is a confounder between the outcome and the treatment

we need $RD_1 > 0$, $RD_2 > 0$ but $RD_3 < 0$

or $RD_1 < 0$, $RD_2 < 0$ but $RD_3 > 0$

in order to achieve that, we can take the following parametrization: $p_1 = 0.1$, $p_2 = 0$, $p_3 = 0.3$, $p_4 = 0.2$

$q_1 = 0$, $q_2 = 1$, $r = 0.1$

then $RD_1 = p_4 - p_3 = 0.2 - 0.3 = -0.1$

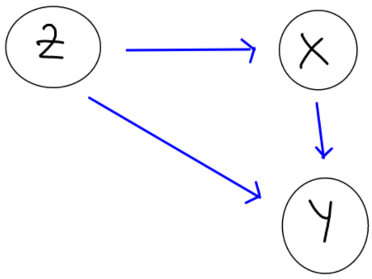
$RD_2 = p_2 - p_1 = 0 - 0.1 = -0.1$

$$RD_3 = \frac{p_4 \cdot q_2 \cdot r + p_2 \cdot q_1 \cdot (1-r)}{q_1 \cdot (1-r) + q_2 \cdot r}$$

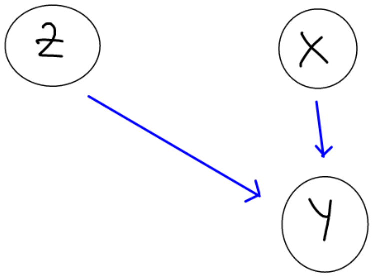
$$- \frac{p_1 \cdot (1-q_1) \cdot (1-r) + p_3 \cdot (1-q_2) \cdot r}{(1-q_1)(1-r) + (1-q_2) r}$$

$$= 0.1$$

2



(a) now to study $P(Y=y \mid do(X=x))$, we need to eliminate the incoming arcs of X



$$P(Y=y \mid do(X=x)) = P_m(Y=y \mid X=x)$$

$$= \sum_z P_m(Y=y \mid Z=z, X=x) \cdot P_m(Z=z \mid X=x)$$

$Z \perp\!\!\!\perp X$ in P_m since only connected via collider

$$= \sum_z P_m(Y=y \mid Z=z, X=x) \cdot P_m(Z=z)$$

$P_m(Z) = P(Z)$ since no parents in both cases, similar for $P(Y \mid Z, X)$

$$= \sum_z P(Y=y \mid Z=z, X=x) \cdot P(Z=z) \quad (*)$$

X takes 2 values and Y takes 2 values

therefore we have 4 combinations

$Y=y_1, X=x_1$ $\quad P(Y=y_1 \mid do(X=x_1)) \overset{(*)}{=} P(Y=y_1 \mid Z=z_1, X=x_1) \cdot P(Z=z_1) + P(Y=y_1 \mid Z=z_2, X=x_1) \cdot P(Z=z_2$

$$= r \cdot p_4 + (1-r) \cdot p_2$$

$Y=y_1, X=x_2$ $\quad P(Y=y_1 \mid do(X=x_2)) \overset{(*)}{=} P(Y=y_1 \mid Z=z_1, X=x_2) \cdot P(Z=z_1) + P(Y=y_1 \mid Z=z_2, X=x_2) \cdot P(Z=z_2$

$$= r p_3 + (1-r) p_1$$

$Y=y_2, X=x_1$ | $P(Y=y_2 \mid do(X=x_1)) = 1 - P(Y=y_1 \mid do(X=x_1))$

$$= 1 - (r \cdot p_4 + (1-r) \cdot p_2)$$

$Y=y_2, X=x_2$ | $P(Y=y_2 \mid do(X=x_2)) = 1 - P(Y=y_1 \mid do(X=x_2))$

$$= 1 - rp_3 + (1-r)\, p_1$$

(b) according to slide 362 23, we have the following adjustment formula:

$$P(Y=y \mid do(X=x)) = \sum_z P(Y=y \mid Z=z, X=x) \cdot P(Z=z)$$

this is the same as in part (a)

hence we get the same results

(c) we can use slide 16:

$$ACE = P(Y=y_1 \mid do(X=x_1)) - P(Y=y_1 \mid do(X=x_2))$$

$$\overset{\text{part (a)}}{=} r \cdot p_4 + (1-r)\, p_2 - rp_3 - (1-r)\, p_1$$

we compare it to RD_3 from Ex. 1 (b)

we see that $ACE \neq RD_3$

the difference is described on slide 7:

ACE measures the effect on Y from an intervention that forces X to be x_1

RD measures the effect on Y from observing a change from x_2 to x_1

d) we already found a comb. of parameters in Ex. 1 c)

what is disegregated data ?

“Disaggregated” here means stratified

Exercise 3

Source:
Miguel A Hernán, David Clayton, Niels Keiding, The Simpson's paradox unraveled, *International Journal of Epidemiology*, Volume 40, Issue 3, June 2011, Pages 780–785, https://doi.org/10.1093/ije/dyr041

How would you describe Simpson's paradox to a friend?

Consider a population and two variables under study in this population. Now it can happen that the association between the two variables can emerge, disappear or reverse when we divide the population into subpopulations and study the same variables in the subpopulations.

What considerations should drive the choice between a marginal or conditional analysis?

The choice between marginal or conditional analysis depends on the research setting (p. 780-781).
Page 781: "From a purely statistical standpoint, no general rule seems to exist as to whether the conditional association or the marginal association should be preferred."

Was there a reversal of the effect in Simpson's original paper?

No, see page 782: "This reversal of association, though not present in Simpson's article [...]".

Can Simpson's paradox be explained purely in terms of confounding?

No, see page 782: " However, equating Simpson's paradox with confounding misses Simpson's main point: statistical reasoning is insufficient to choose between the marginal and the conditional association measure."

Also: "Equating Simpson's paradox and confounding not only takes credit away from earlier authors, but also detracts from Simpson's most important message: the realization that statistical information needs to be supplemented with expert knowledge for causal inference from observational data"

Assignment 8

Exercise 1

NYT news article

Do you find the wording appropriate to convey the intended message?

How would you phrase the causal question the headline (article) appears to answer?

original press release

Which causal question does the original headline appear to be answering?

Compared to what baseline?

Does the consumption of olive oil lower cardiovascular and coronary heart disease risk?

Can you think of and describe a target trial (providing suggestions for the key components) the investigators may have wished to conduct?

An RCT would probalby have been desirable.

Who should we include?

- People without coronary or cardiovascular diseases

What treatment or exposure strategies should we compare?

- One group consumes no olive oil, and the other consumes olive oil. Apart from olive oil, both groups have the exactly same diet. The groups are randomized.

How long should we follow the people in our study for?

- The study goes over 5 decades.

What outcome should we compare and what question about the outcome do we really want to answer?

- cardiovascular and coronary heart disease
  - there we need to specify certain thresholds for some measurable values

- or number of heart attacks

Does it sound like a realistic and useful study?

This is an ideal, unrealistic scenario. A more realistic one would be an observational study where people record their daily intake of foods/drinks.

In an observational study, we do not randomize. This means:
- We have to worry about confounding.
- It can only be used to estimate the per-protocol effect

After forming your own opinion confront it with a hot take, on the topic of epidemiological studies dealing with nutrition questions, as elaborated by Epidemiologist Ellie Murray at the link below (also providing insights about target trials) ?

RCT's in the field of nutritional epidemiology are unrealistic. We can do observational studies. To design such a study, the target trial helps.

Explain better what other issues the observational or randomised studies face

Assignment 9

1

a. A set of covariates L satisfies the backdoor criterion if all backdoor paths between X and Y are blocked by conditioning on L and L contains no variables that are descendants of treatment X

sets of 2 nodes: $\{Z, A\}$, $\{Z, B\}$, $\{Z, C\}$, $\{Z, D\}$

" 3 " : $\{Z, A, B\}$, $\{Z, A, C\}$, $\{Z, A, D\}$, $\{Z, B, C\}$, $\{Z, B, D\}$, $\{Z, C, D\}$

" 4 " : $\{Z, A, B, C\}$, $\{Z, A, B, D\}$, $\{Z, A, C, D\}$, $\{Z, B, C, D\}$

" 5 " : $\{Z, A, B, C, D\}$

b. all sets of part a with 2 nodes are minimal
in all sets with more than 2 nodes, we can remove a node and the backdoor criterion still holds

c. causal effect of D on Y
we are looking for a set that blocks all backdoor paths from D to Y.
$\{C\}$ is such a set, hence every other set containing c is not minimal

if the set does not contain C, it must contain Z,
because otherwise we have an open path $Y \leftarrow Z \leftarrow C \rightarrow D$

Z alone is not a minimal set because the path
$Y \leftarrow W \leftarrow X \leftarrow A \leftarrow B \rightarrow Z \leftarrow C \rightarrow D$ is open
to block this path we must add A, B, X or W
so there are 4 more minimal sets:
$\{Z, A\}, \{Z, B\}, \{Z, X\}, \{Z, W\}$

- effect of $\{W, D\}$ on Y:

$\{Z\}$ is such a set

other sets not containing Z must contain X and C

hence there are two minimal sets: $\{Z\}$ and $\{X, C\}$

d. We check the definition of the Front Door criterion on slide 25.

A set of variables Z satisfies the front door criterion relative to an ordered pair of variables (X, Y) if:

- Z intercepts all directed paths from X to Y
- there is no backdoor path from X to Z
- X blocks all backdoor paths from Z to Y

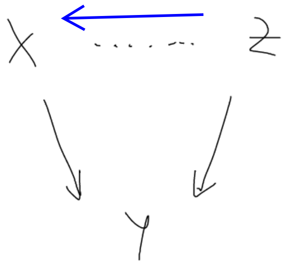
$\{W\}$ is such a set

if $P(x, w) > 0$ then the causal effect of X on Y is identifiable and the adjustment formula takes the following form:

$$P(Y=y \mid do(X=x)) = \sum_{w} P(W=w \mid X=x) \cdot \sum_{x'} P(Y=y \mid Z=z, X=x') \cdot P(X=x')$$

2

causal diagram



..... : induced relationship between X and Z

correlation between Z and X ist just given by the blue arrow because Y is a collider

hence Y does not influence the correlation

when we condition on Y = 1 , the correlation changes because we get association between X and Z via X-Y-Z, this influence is strong enough to change the sign of the correlation (see results for correlation)

Causal Inference

Assignment 9

Benedikt Schmidt

04 May 2022

Exercise 2

```
# exercise 2

N <- 10000
U <- sample(1:100, N,TRUE)
Z <- (U - 50) /30
a_0 <- 0.25
a_z <- 0.3
b_0 <- -2
b_z <- 4
b_x <- 2
e <- rnorm(N, mean=0, sd=1)
X <- a_0 + a_z*Z + e
p <- exp(b_0 + b_z*Z + b_x*X) / (1 + exp(b_0 + b_z*Z + b_x*X))
Y <- rbinom(N,1,p)

# correlation between X and Z for the entire simulated dataset:
c1 <- cor(X,Z)
c1
```

```
## [1] 0.9934316
```

0.27 would be correct

```
# correlation between X and Z restricted to the observations with Y = 1

X1 <- X[which(Y==1)]
Z1 <- Z[which(Y==1)]
c2 <- cor(X1, Z1)
c2
```

correlation does not tell anything about the steepness of the linear relationship
it is a measure for the closeness of the points to to line

```
## [1] 0.9743801
```

-0.19 would be correct

```
# the two different numbers for the different correlations are pretty similar
# the difference is:
abs(c2-c1)
```

```
## [1] 0.01905151
```

```
# we see that the correlation between X and Z does not change much when
# conditioning on Y = 1

# it could be that we have Y as a collider in the diagram (see handwritten notes)
# if we condition on Y, then we induce an association between X and Z
```

```
#

# The article talks about collider bias. If two factors influence being selected,
# they collide on selection => collider bias
# An association of the two factors can be introduced even though they are
# independent
```