DS 2006 Midterm 2

NAME: Olivia Park

NOTE: The `.rmd` version of the file is available here: (link)

Instructions

Please prepare reponses/solutions for the following questions.

Allowable resources

You may use the solutions you've prepared for the prep questions during the exam. You are welcome to use your homeworks, deliverables, or class notes. You are permitted to access the internet for publicly available content. You are not allowed to communicate with anyone via the internet or any other means during the exam. This includes, but is not limited to:

- No messaging, emailing, or using social media to contact others.
- No posting questions or seeking answers on forums, chat rooms, chat bots (including large language models like ChatGPT), or any collaborative platforms.
- No sharing or discussing exam content with peers through any online or electronic medium.

You may **NOT** discuss any aspect of the exam or prep questions with anyone other than the instructor or TA. You may **NOT** share code or documents.

Submission instructions

1. Within your course repo, create a folder called `Midterm2`
2. Within the folder, create the script file `exam.rmd` with your solutions. Create a rendered report in `.pdf` output.
3. Add, commit, and push to your repo on github.com.

Questions

Exam questions are organized into sections cooresponding to the learning outcomes of the course.

Section 1. Tools of the data scientists

> Learning objective: Use the tools of data scientists
>
> Learning objective: Implement best programming/coding practices

1.1 [5 pts] Write your name at the prompt above (line 6 in the script).

1.2 [5 pts] When you are done with the exam, please render this report as a pdf document.

1.3 [5 pts] The following is a schematic of a project folder, with subfolders and files.

```
project
|
|---code
|       script.rmd
|
|---data
|       survey-responses.csv
|
|---docs
```

Supposing the `code` subfolder is the designated working directory, write the command to be included in the `script.rmd` file which will read the `survey-responses.csv` data, avoiding absolute file paths?

```
survey <- read.csv("~/data/survey-responses.csv")
```

Section 2. Probability & Diagnostics

Learning objective: Compare and contrast different definitions of probability, illustrating differences with simple examples

Learning objective: Express the rules of probability verbally, mathematically, and computationally.

Learning objective: apply cross table framework to the special case of binary outcomes

2.1 [5 pts] In a particular town, 1% of all individuals have a certain rare disease. There's a test for this disease that correctly identifies a sufferer 99% of the time (true positive rate) but also falsely identifies the disease in 2% of the healthy population (false positive rate).

HINT: False positive rate = $P(T+|D-)$

Complete the following table of cell, row, and column probabilites based on the information about the prevalence, true positive rate, and false positive rate. You are welcome to use an excel spreadsheet (link) which will automatically create the table for any combination of prevalence, sensitivity, and specificity. (It might save time to insert a screen shot of the table rather than manually creating the table.)

	Disease +	Disease -	
Test +			
→ cell		.0198	.000198
→ row	.99	.01	
→ col		.02	
Test -			
→ cell		.97	
→ row			
→ col		.98	
	.01	.99	1

set specificity to .98 set sensitivity to .99

		D+	D-	
T +	cell	0.010	0.010	0.020
	row	0.497	0.503	
	col	0.980	0.010	
T -	cell	0.000	0.980	0.980
	row	0.000	1.000	
	col	0.020	0.990	
		0.010	0.990	

2.2 [5 pts] Suppose that a new test is developed and approved 98% true positive rate and 1% false positive rate. As a consumer, which would you prefer? Be specific about your reasoning and the quantities that you are using to make a decision.

		D+	D-	
T +	cell	0.010	0.020	0.030
	row	0.333	0.667	
	col	0.990	0.020	
T -	cell	0.000	0.970	0.970
	row	0.000	1.000	
	col	0.010	0.980	
		0.010	0.990	

2.3 [10 pts] An audit of an email filtering system resulted in a dataset of 10000 emails, each manually verified as spam or not spam. In addition to the type of email, the dataset indicates if the filter sent the email to the inbox or the junk folder.

The following command reads the dataset into memory. From the data, generate an estimate of the positive predictive value and the negative predictive value of the spam filter.

HINT: Specificity in this example = $P(\text{Junk folder}|\text{Spam})$

```
d1 <- readRDS(url("https://tgstewart.cloud/spam-data.RDS"))
table(d1)
```

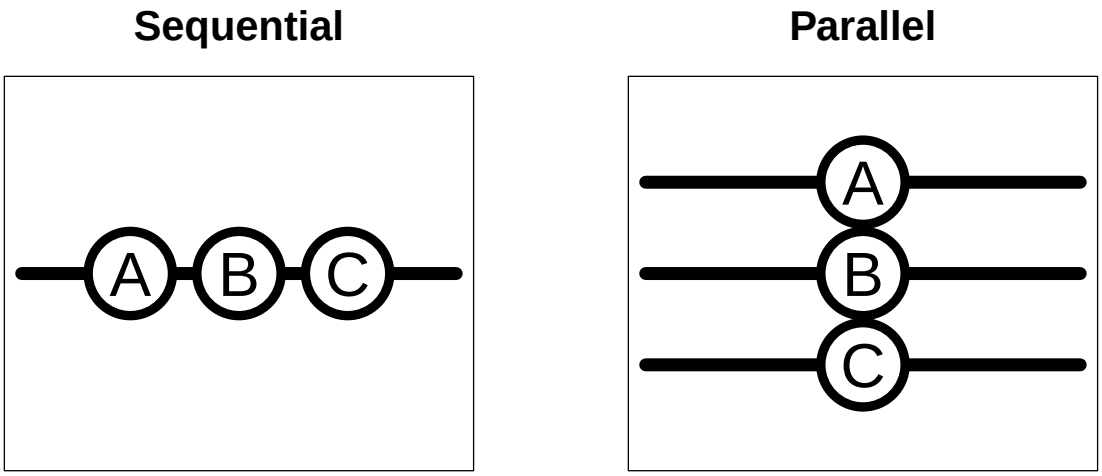
```
##           Folder
## Type       Inbox Junk
##   Not spam  7589  422
##   Spam       197 1792
```

Section 3. Simulation

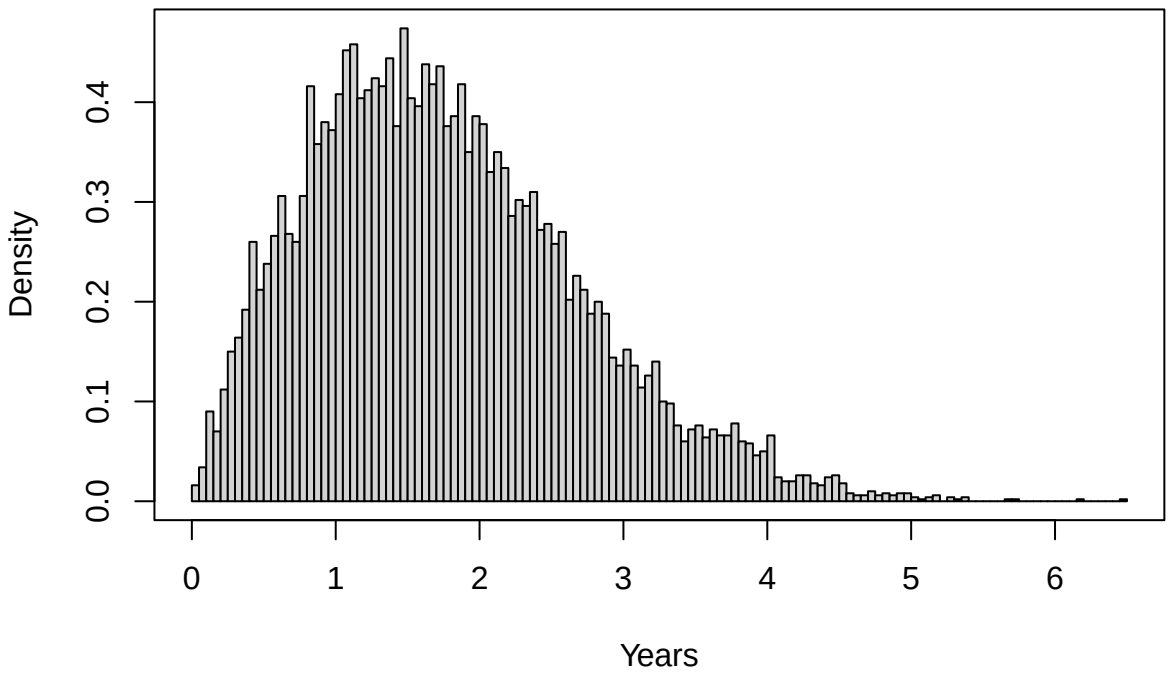
Learning objective: Use probability models to build simulations of complex real world processes to answer research questions

3.1 [5 pts] Consider two systems of three components (say A, B, and C). In the first, a failure of any component leads to a failure of the entire system. In the second, the components are redundent, and a failure only occurs if all three components fail.

```
## pdf
##   2
```



Suppose the failure time of an individual component is a random variable with the following distribution.



The following function `sysfail` generates replicates of the time to system failure (years) for the sequential and parallel systems. The input parameter `R` is the number of replicates that the fuction will return.

```
sysfail <- function(R){
    A <- array(rweibull(R*3,2,2),dim = c(R,3))
    data.frame(sequential = apply(A,1,min), parallel = apply(A,1,max))
}
```

The following provides an estimate of how much longer the parallel system will last compared to the sequential system by simulating 25 different system failures.

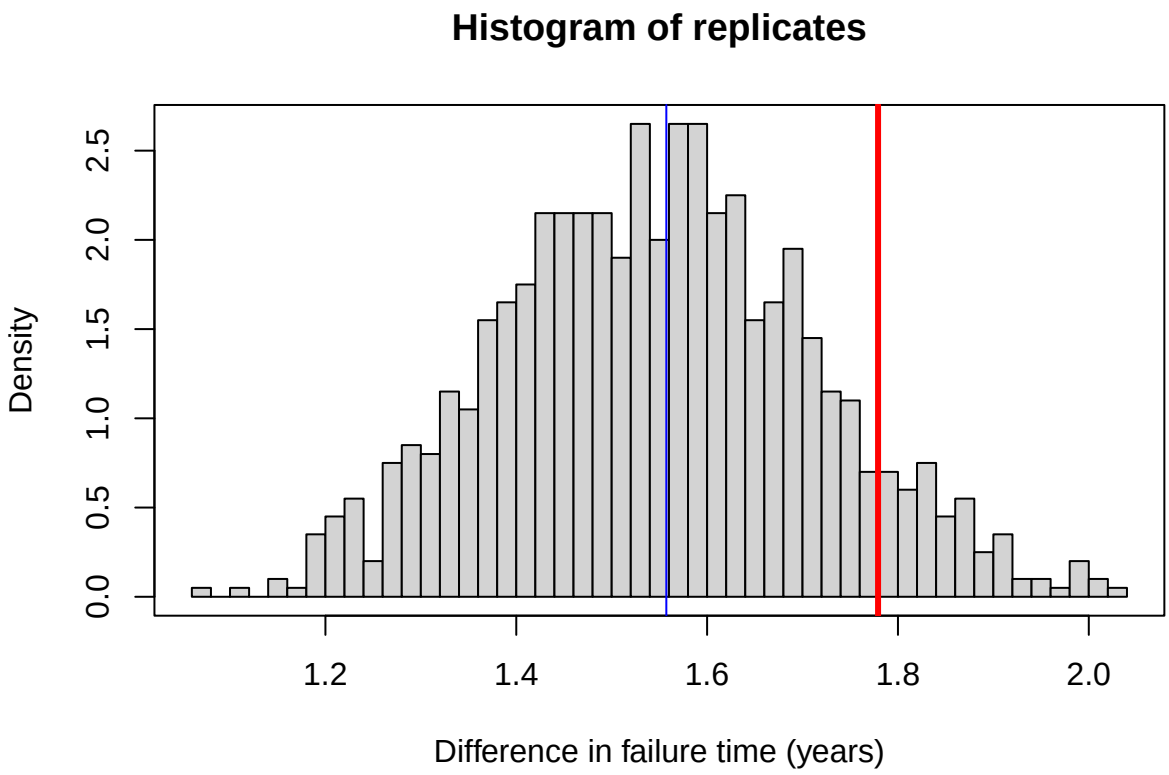
```
set.seed(230583)
a1 <- sysfail(25)
a2 <- colMeans(a1)
diff(a2)
```

```
## parallel
## 1.779168
```

The calculated difference is based on pseudo-random data. The process can be repeated many times. The following code creates 1000 estimates. The redline refers to the single estimate generated above.

```
R <- 25
replicates <- replicate(1000, sysfail(R) |> colMeans() |> diff())
hist(replicates, breaks = 50, freq = FALSE, xlab = "Difference in failure time (years)"); box()
```

```r
abline(v=diff(a2), lwd = 3, col = "red")
mean_reps <- mean(replicates <- replicate(1000, sysfail(R) |> colMeans() |> diff()))
abline(v=mean_reps, col = "blue")
```



Add to the figure blue reference line for the mean of the 1000 estimates. (Simply edit the code chunk above. You do not need to create a second copy.)

3.2 [5 pts] What is the range (min and max) of the 1000 values you generated for the improved failure time estimate?

```
#min and max
min(replicates)
```

```
## [1] 1.01974
```

```
max(replicates)
```

```
## [1] 2.181636
```

3.3 [5 pts] What is the average absolute error of the 1000 estimates? Use the mean calculated in 3.1 as the "true" value.

```
absolute_error <- function(out, mean_reps){
  sum_of_error <- 0
  for(i in 1:length(out)){
    sum_of_error <- sum_of_error + abs(out[i] - mean_reps)
  }
  absolute_error <- sum_of_error / length(out)
}
abs_err <- absolute_error(replicates, mean_reps)

print(abs_err)
```

```
##  parallel
## 0.1391898
```

3.4 [5 pts] What is the average relative error of the 1000 estimates? Use the mean calculated in 3.1 as the "true" value.

```
rel_err <- abs_err/mean_reps
print(rel_err)
```

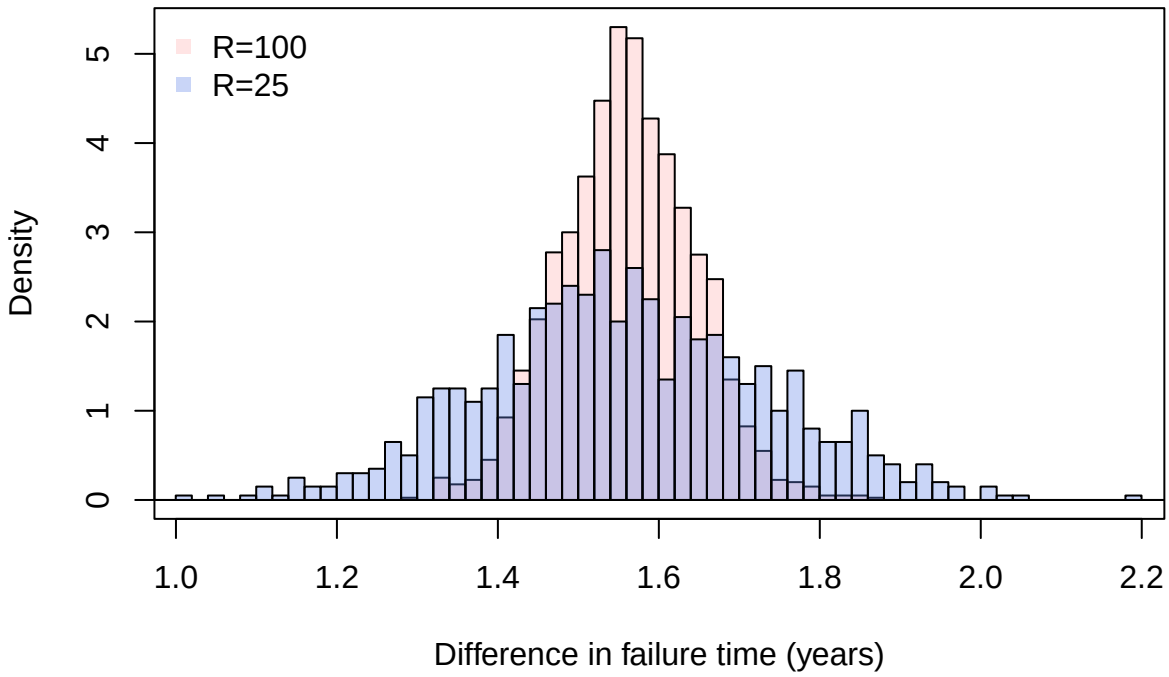
```
##   parallel
## 0.08938152
```

3.5 [5 pts] If you wanted to reduced the error by half, how many replicates (R) would you need to use? you would want to use 100 replicates to half the error.

3.6 [5 pts] Generate a plot of overlapping histograms to show the difference between R=25 and your R from the previous problem.

```
R2 <- 100 # Change this
replicates2 <- replicate(2000, sysfail(R2) |> colMeans() |> diff())

b1 <- seq(min(c(replicates,replicates2))-.1, max(c(replicates,replicates2))+0.1,by=0.02)
hist(replicates2, breaks = b1, col = "#ffabab50", freq = FALSE, main = "", xlim = range(replicates), xla
legend("topleft", legend = c(paste0("R=",R2),"R=25"), col = c("#ffabab50","#6488ea59"),bty = "n", pch =
hist(replicates, breaks = b1, add=TRUE, col = "#6488ea59", freq = FALSE)
box()
```



3.7 [5 pts] Calculate the average absolute error of the 1000 estimates generate with the new choice of R? Did it change as you expected it to?

The absolute error is half with the new R, which is what we expected when we multiplied the R by 4.

```
mean_reps2 <- mean(replicates2)

absolute_error <- function(out, mean_reps2){
  sum_of_error <- 0
  for(i in 1:length(out)){
    sum_of_error <- sum_of_error + abs(out[i] - mean_reps)
  }
  absolute_error <- sum_of_error / length(out)
}
abs_err <- absolute_error(replicates2, mean_reps2)

print(abs_err)
```

```
##   parallel
## 0.06522258
```

Section 4. Confounding vs Causal Pathway

> Learning objective: define/describe confounding variables, Simpson's paradox, DAGs, and the causal pathway

4.1 [10 pts] The following function generates data from a cohort of individuals who agreed to be studied about heart disease. In the dataset, there is exercise level at age 20 (below average, above average), blood pressure at age 25 (low, normal, high), and heart disease at age 30 (present, absent).

Generate 10000 draws from the function and create the cross table for exercise level and heart disease. Calculate a summary of the effect of exercise by calculating the risk ratio:

$$RR = \frac{P(\text{heart disease present}|\text{below average exercise})}{P(\text{heart disease present}|\text{above average exercise})}$$

I t

```
heart_data <- function(R){
    ex <- rbinom(R,1,.5)
    bp <- rnorm(R,-ex+1/2,1)
    bp <- cut(bp,c(-Inf,-1,1,Inf), labels = FALSE)
    hd <- 1*(rnorm(R,bp-3,1.8)>0)
    data.frame(
        exercise = factor(ex,0:1,c("below average","above average"))
      , blood_pressure = factor(bp,1:3,c("low","normal","high"))
      , heart_disease = factor(hd, 0:1, c("absent","present"))
    )
}


set.seed(20240329); d1 <- heart_data(1000000)

calc_delta <- function(df){
  num_ex <- 0
  ex_and_hd <- 0

  for(i in 1:nrow(df)){
    if(df[i,1] == "exsericise"){
      num_ex <- num_ex + 1
      if(df[i,3] == "heart disease"){
        vax_and_recov <- ex_and_hdv + 1
      }
    }
  }
  p_hd_given_ex = ex_and_hd / num_ex

  num_noex <- 0
  noex_and_nohd <- 0

  for(i in 1:nrow(df)){
    if(df[i,1] == "noexcerise"){
      num_noex <- num_noex + 1
      if(df[i,3] == "nohd"){
        noex_and_nohd <- noex_and_nohd + 1
```

```
      }
    }
  }
   p_hd_given_ex = ex_and_hd / num_ex

  delta = p_hd_given_ex - noex_and_nohd
}

Ra <- 1000
exer <- heart_data(Ra)

print(exer)
```

```
##           exercise blood_pressure heart_disease
## 1    above average         normal        absent
## 2    below average         normal        absent
## 3    above average         normal       present
## 4    above average         normal        absent
## 5    above average         normal        absent
## 6    below average           high       present
## 7    below average         normal        absent
## 8    below average         normal       present
## 9    below average           high       present
## 10   above average         normal        absent
## 11   below average         normal        absent
## 12   below average         normal        absent
## 13   below average         normal        absent
## 14   below average            low        absent
## 15   above average         normal        absent
## 16   below average           high        absent
## 17   above average         normal       present
## 18   above average            low        absent
## 19   above average            low        absent
## 20   below average         normal        absent
## 21   above average            low        absent
## 22   below average           high       present
## 23   below average           high       present
## 24   above average            low       present
## 25   below average         normal       present
## 26   above average         normal        absent
## 27   below average         normal        absent
## 28   below average         normal       present
## 29   above average         normal       present
## 30   above average         normal       present
## 31   above average         normal        absent
## 32   above average            low        absent
## 33   above average            low        absent
## 34   above average         normal        absent
## 35   above average            low        absent
## 36   below average            low        absent
## 37   above average         normal        absent
## 38   above average         normal       present
## 39   below average           high        absent
```

```
## 40   below average         normal       present
## 41   below average         normal        absent
## 42   below average         normal        absent
## 43   below average           high        absent
## 44   below average         normal        absent
## 45   below average         normal       present
## 46   above average         normal        absent
## 47   above average            low       present
## 48   above average         normal        absent
## 49   above average         normal        absent
## 50   above average         normal       present
## 51   below average           high       present
## 52   above average            low       present
## 53   above average         normal       present
## 54   above average         normal       present
## 55   below average         normal        absent
## 56   below average         normal        absent
## 57   below average         normal       present
## 58   below average         normal       present
## 59   above average         normal        absent
## 60   below average            low        absent
## 61   above average         normal        absent
## 62   above average         normal       present
## 63   above average           high       present
## 64   above average         normal        absent
## 65   below average           high        absent
## 66   below average         normal        absent
## 67   above average         normal        absent
## 68   below average         normal       present
## 69   below average            low        absent
## 70   below average            low        absent
## 71   above average         normal        absent
## 72   below average         normal        absent
## 73   above average            low        absent
## 74   above average         normal        absent
## 75   above average            low        absent
## 76   below average           high       present
## 77   below average         normal        absent
## 78   above average            low        absent
## 79   above average         normal        absent
## 80   below average         normal       present
## 81   above average            low        absent
## 82   above average         normal       present
## 83   below average           high        absent
## 84   below average         normal        absent
## 85   above average           high        absent
## 86   below average         normal        absent
## 87   above average            low        absent
## 88   below average         normal       present
## 89   below average         normal        absent
## 90   below average           high       present
## 91   above average            low        absent
## 92   above average         normal        absent
## 93   above average         normal       present
```

```
## 94   below average         normal       present
## 95   above average         normal       present
## 96   below average           high       present
## 97   above average            low        absent
## 98   below average           high       present
## 99   above average         normal        absent
## 100  below average           high        absent
## 101  below average            low        absent
## 102  below average         normal       present
## 103  below average           high        absent
## 104  below average         normal       present
## 105  below average         normal        absent
## 106  below average         normal       present
## 107  above average         normal        absent
## 108  above average            low       present
## 109  below average         normal        absent
## 110  below average         normal        absent
## 111  below average            low        absent
## 112  below average           high        absent
## 113  below average         normal       present
## 114  above average         normal        absent
## 115  below average         normal        absent
## 116  below average         normal        absent
## 117  above average            low        absent
## 118  above average            low        absent
## 119  below average         normal        absent
## 120  above average         normal        absent
## 121  above average         normal       present
## 122  above average         normal        absent
## 123  below average           high        absent
## 124  above average            low        absent
## 125  above average         normal        absent
## 126  above average           high       present
## 127  above average            low        absent
## 128  above average            low        absent
## 129  above average            low       present
## 130  above average         normal        absent
## 131  above average         normal        absent
## 132  below average         normal        absent
## 133  above average            low        absent
## 134  below average         normal       present
## 135  above average            low        absent
## 136  above average         normal        absent
## 137  above average            low        absent
## 138  above average            low        absent
## 139  above average         normal        absent
## 140  below average           high        absent
## 141  above average         normal        absent
## 142  below average         normal        absent
## 143  above average         normal        absent
## 144  below average         normal       present
## 145  below average         normal        absent
## 146  above average         normal        absent
## 147  below average         normal        absent
```

```
## 148  above average         normal       present
## 149  below average           high        absent
## 150  above average         normal        absent
## 151  above average         normal        absent
## 152  below average         normal        absent
## 153  above average           high        absent
## 154  below average         normal        absent
## 155  below average         normal       present
## 156  below average           high       present
## 157  below average           high       present
## 158  below average            low        absent
## 159  below average           high       present
## 160  below average         normal       present
## 161  below average            low        absent
## 162  above average         normal        absent
## 163  above average         normal       present
## 164  above average            low        absent
## 165  above average         normal        absent
## 166  above average            low       present
## 167  below average         normal        absent
## 168  above average         normal        absent
## 169  above average         normal       present
## 170  below average         normal        absent
## 171  above average           high       present
## 172  above average            low       present
## 173  above average         normal        absent
## 174  below average         normal        absent
## 175  below average         normal        absent
## 176  below average         normal        absent
## 177  below average         normal       present
## 178  above average         normal        absent
## 179  below average           high        absent
## 180  below average         normal       present
## 181  above average         normal       present
## 182  below average         normal        absent
## 183  below average         normal        absent
## 184  below average           high       present
## 185  above average         normal        absent
## 186  below average           high        absent
## 187  above average         normal       present
## 188  above average         normal        absent
## 189  above average         normal       present
## 190  above average         normal        absent
## 191  above average         normal        absent
## 192  above average            low        absent
## 193  below average           high       present
## 194  below average         normal        absent
## 195  below average         normal       present
## 196  below average           high        absent
## 197  above average            low        absent
## 198  below average         normal        absent
## 199  above average            low        absent
## 200  above average         normal        absent
## 201  above average         normal       present
```

```
## 202  above average         normal        absent
## 203  above average         normal        absent
## 204  above average            low        absent
## 205  below average           high        absent
## 206  above average         normal        absent
## 207  above average         normal        absent
## 208  above average            low        absent
## 209  below average         normal        absent
## 210  above average            low       present
## 211  above average         normal        absent
## 212  above average         normal        absent
## 213  below average         normal        absent
## 214  above average         normal        absent
## 215  below average           high        absent
## 216  above average            low        absent
## 217  above average         normal       present
## 218  above average         normal        absent
## 219  below average         normal        absent
## 220  below average           high       present
## 221  below average         normal        absent
## 222  above average         normal        absent
## 223  above average         normal        absent
## 224  below average         normal        absent
## 225  above average         normal        absent
## 226  above average         normal        absent
## 227  below average            low       present
## 228  below average         normal        absent
## 229  below average         normal       present
## 230  below average            low        absent
## 231  above average         normal        absent
## 232  below average         normal        absent
## 233  below average           high       present
## 234  below average         normal        absent
## 235  above average         normal        absent
## 236  below average         normal        absent
## 237  below average           high       present
## 238  above average            low        absent
## 239  below average         normal        absent
## 240  below average           high       present
## 241  above average         normal        absent
## 242  above average            low        absent
## 243  below average         normal        absent
## 244  above average         normal        absent
## 245  above average         normal        absent
## 246  above average         normal        absent
## 247  above average         normal       present
## 248  above average           high       present
## 249  above average            low        absent
## 250  above average           high        absent
## 251  below average           high        absent
## 252  below average         normal        absent
## 253  below average           high       present
## 254  below average         normal       present
## 255  above average            low        absent
```

```
## 256  below average         normal        absent
## 257  above average            low        absent
## 258  below average           high        absent
## 259  below average           high        absent
## 260  above average         normal        absent
## 261  above average         normal       present
## 262  below average           high       present
## 263  below average         normal        absent
## 264  below average            low        absent
## 265  above average         normal        absent
## 266  above average           high       present
## 267  below average           high        absent
## 268  below average           high       present
## 269  above average         normal        absent
## 270  above average            low        absent
## 271  below average           high        absent
## 272  above average            low        absent
## 273  above average            low        absent
## 274  below average         normal       present
## 275  below average         normal        absent
## 276  below average         normal        absent
## 277  above average            low        absent
## 278  above average            low        absent
## 279  above average         normal        absent
## 280  above average         normal        absent
## 281  above average            low       present
## 282  below average         normal       present
## 283  below average         normal        absent
## 284  below average         normal        absent
## 285  above average         normal       present
## 286  below average         normal       present
## 287  above average           high       present
## 288  below average         normal        absent
## 289  above average         normal        absent
## 290  above average           high        absent
## 291  below average         normal        absent
## 292  above average            low        absent
## 293  below average         normal       present
## 294  below average         normal        absent
## 295  above average         normal        absent
## 296  above average         normal        absent
## 297  below average           high       present
## 298  above average         normal       present
## 299  above average           high       present
## 300  above average         normal       present
## 301  below average           high        absent
## 302  above average            low        absent
## 303  above average         normal        absent
## 304  below average           high        absent
## 305  above average         normal        absent
## 306  below average           high        absent
## 307  above average         normal        absent
## 308  below average         normal        absent
## 309  above average            low        absent
```

```
## 310  above average         normal       present
## 311  above average         normal        absent
## 312  above average         normal        absent
## 313  below average         normal       present
## 314  below average           high       present
## 315  above average         normal        absent
## 316  above average         normal        absent
## 317  below average           high        absent
## 318  above average            low        absent
## 319  above average            low        absent
## 320  above average         normal       present
## 321  below average         normal        absent
## 322  above average         normal        absent
## 323  below average           high        absent
## 324  above average           high       present
## 325  below average         normal       present
## 326  above average         normal        absent
## 327  below average         normal        absent
## 328  below average         normal        absent
## 329  above average         normal        absent
## 330  above average         normal        absent
## 331  above average         normal        absent
## 332  above average         normal        absent
## 333  below average            low        absent
## 334  below average           high        absent
## 335  below average           high       present
## 336  above average            low       present
## 337  above average         normal        absent
## 338  below average         normal        absent
## 339  below average         normal        absent
## 340  below average         normal        absent
## 341  below average           high       present
## 342  above average            low        absent
## 343  above average         normal        absent
## 344  above average         normal        absent
## 345  below average         normal       present
## 346  above average         normal       present
## 347  below average         normal       present
## 348  below average         normal       present
## 349  below average            low        absent
## 350  above average            low        absent
## 351  above average            low       present
## 352  below average         normal        absent
## 353  below average         normal        absent
## 354  above average            low       present
## 355  below average         normal        absent
## 356  above average         normal        absent
## 357  above average            low       present
## 358  above average         normal        absent
## 359  below average         normal        absent
## 360  below average         normal        absent
## 361  above average         normal       present
## 362  below average           high        absent
## 363  below average           high       present
```

```
## 364  below average           high        absent
## 365  below average           high        absent
## 366  below average           high        absent
## 367  above average         normal        absent
## 368  below average         normal        absent
## 369  above average         normal        absent
## 370  below average            low       present
## 371  above average            low        absent
## 372  below average           high       present
## 373  below average         normal        absent
## 374  above average         normal        absent
## 375  below average           high        absent
## 376  above average         normal        absent
## 377  below average         normal        absent
## 378  below average         normal        absent
## 379  below average         normal        absent
## 380  above average            low        absent
## 381  above average         normal        absent
## 382  below average           high       present
## 383  below average         normal       present
## 384  above average           high       present
## 385  below average           high       present
## 386  above average            low        absent
## 387  below average         normal       present
## 388  below average           high       present
## 389  below average         normal        absent
## 390  below average         normal       present
## 391  above average         normal       present
## 392  above average            low        absent
## 393  above average            low        absent
## 394  below average         normal        absent
## 395  below average           high       present
## 396  above average            low       present
## 397  below average         normal        absent
## 398  above average            low        absent
## 399  below average         normal        absent
## 400  above average            low       present
## 401  above average         normal        absent
## 402  below average            low       present
## 403  below average         normal        absent
## 404  above average         normal        absent
## 405  below average         normal        absent
## 406  below average           high        absent
## 407  below average         normal       present
## 408  above average         normal        absent
## 409  below average           high       present
## 410  above average         normal        absent
## 411  above average         normal        absent
## 412  above average            low        absent
## 413  below average            low        absent
## 414  below average         normal       present
## 415  above average            low        absent
## 416  below average         normal        absent
## 417  above average         normal        absent
```

```
## 418  above average           low        absent
## 419  below average        normal        absent
## 420  below average        normal        absent
## 421  above average        normal       present
## 422  above average           low        absent
## 423  below average        normal        absent
## 424  below average        normal        absent
## 425  below average        normal        absent
## 426  below average          high        absent
## 427  below average          high        absent
## 428  below average        normal        absent
## 429  below average          high        absent
## 430  above average           low        absent
## 431  above average        normal       present
## 432  below average        normal        absent
## 433  above average          high       present
## 434  below average        normal        absent
## 435  above average        normal        absent
## 436  below average        normal        absent
## 437  below average        normal        absent
## 438  below average        normal       present
## 439  below average        normal        absent
## 440  above average        normal        absent
## 441  above average        normal       present
## 442  below average        normal        absent
## 443  below average        normal        absent
## 444  below average        normal       present
## 445  below average        normal        absent
## 446  below average        normal        absent
## 447  above average           low        absent
## 448  below average           low        absent
## 449  above average        normal        absent
## 450  above average        normal       present
## 451  below average        normal        absent
## 452  above average          high       present
## 453  above average        normal       present
## 454  above average        normal        absent
## 455  below average        normal        absent
## 456  below average        normal        absent
## 457  above average        normal        absent
## 458  above average        normal       present
## 459  above average        normal       present
## 460  above average           low        absent
## 461  below average          high       present
## 462  below average          high        absent
## 463  above average          high       present
## 464  above average        normal        absent
## 465  above average        normal       present
## 466  above average        normal       present
## 467  above average        normal        absent
## 468  above average        normal       present
## 469  below average          high       present
## 470  below average          high       present
## 471  above average        normal       present
```

```
## 472  below average         normal        absent
## 473  below average         normal        absent
## 474  below average         normal       present
## 475  below average         normal        absent
## 476  above average            low        absent
## 477  above average         normal       present
## 478  above average         normal       present
## 479  above average         normal       present
## 480  below average         normal       present
## 481  above average         normal        absent
## 482  below average         normal        absent
## 483  above average            low        absent
## 484  below average           high       present
## 485  above average         normal        absent
## 486  below average           high        absent
## 487  above average         normal       present
## 488  above average         normal       present
## 489  below average           high        absent
## 490  below average         normal        absent
## 491  above average            low        absent
## 492  above average         normal       present
## 493  below average         normal        absent
## 494  below average         normal        absent
## 495  below average         normal        absent
## 496  below average           high       present
## 497  above average            low        absent
## 498  below average         normal       present
## 499  below average           high       present
## 500  above average         normal        absent
## 501  below average         normal       present
## 502  above average         normal        absent
## 503  below average           high       present
## 504  below average           high       present
## 505  above average         normal        absent
## 506  below average         normal       present
## 507  above average         normal        absent
## 508  below average         normal        absent
## 509  below average         normal       present
## 510  above average         normal       present
## 511  above average         normal        absent
## 512  above average         normal        absent
## 513  above average            low        absent
## 514  below average           high        absent
## 515  above average         normal        absent
## 516  below average         normal       present
## 517  below average         normal        absent
## 518  above average            low       present
## 519  below average         normal        absent
## 520  above average            low       present
## 521  below average           high        absent
## 522  above average            low        absent
## 523  above average         normal        absent
## 524  above average         normal       present
## 525  below average           high        absent
```

```
## 526  above average            low       present
## 527  below average           high       present
## 528  below average         normal        absent
## 529  above average         normal        absent
## 530  above average         normal        absent
## 531  above average         normal        absent
## 532  below average         normal        absent
## 533  below average           high        absent
## 534  below average         normal        absent
## 535  below average           high        absent
## 536  below average         normal        absent
## 537  below average           high       present
## 538  above average         normal       present
## 539  above average         normal        absent
## 540  below average           high       present
## 541  above average         normal        absent
## 542  above average            low        absent
## 543  below average         normal        absent
## 544  below average         normal        absent
## 545  above average         normal        absent
## 546  below average         normal        absent
## 547  above average            low        absent
## 548  below average           high       present
## 549  below average           high       present
## 550  above average           high       present
## 551  above average            low        absent
## 552  above average            low        absent
## 553  below average         normal        absent
## 554  above average            low        absent
## 555  above average         normal        absent
## 556  above average           high       present
## 557  below average         normal        absent
## 558  above average         normal        absent
## 559  below average         normal       present
## 560  above average         normal        absent
## 561  below average           high       present
## 562  above average         normal       present
## 563  below average         normal        absent
## 564  above average            low        absent
## 565  below average           high        absent
## 566  above average           high        absent
## 567  below average         normal        absent
## 568  below average         normal        absent
## 569  below average         normal        absent
## 570  below average         normal        absent
## 571  below average         normal        absent
## 572  below average           high       present
## 573  above average            low        absent
## 574  below average         normal        absent
## 575  below average         normal        absent
## 576  below average            low        absent
## 577  above average           high        absent
## 578  below average         normal       present
## 579  below average         normal       present
```

```
## 580  above average         normal        absent
## 581  below average         normal        absent
## 582  below average         normal       present
## 583  above average         normal        absent
## 584  above average         normal        absent
## 585  above average            low        absent
## 586  above average         normal       present
## 587  below average         normal        absent
## 588  above average            low        absent
## 589  below average         normal        absent
## 590  above average            low        absent
## 591  below average         normal        absent
## 592  above average         normal        absent
## 593  above average         normal        absent
## 594  below average            low        absent
## 595  above average         normal        absent
## 596  above average         normal        absent
## 597  above average            low        absent
## 598  above average         normal        absent
## 599  above average         normal       present
## 600  above average         normal       present
## 601  above average            low        absent
## 602  below average         normal       present
## 603  below average           high        absent
## 604  above average            low        absent
## 605  below average         normal        absent
## 606  below average         normal        absent
## 607  above average            low        absent
## 608  above average            low        absent
## 609  above average         normal        absent
## 610  above average            low        absent
## 611  below average         normal       present
## 612  above average         normal        absent
## 613  above average         normal       present
## 614  below average         normal        absent
## 615  above average            low        absent
## 616  above average            low        absent
## 617  above average         normal       present
## 618  below average         normal        absent
## 619  below average           high       present
## 620  above average         normal        absent
## 621  above average            low        absent
## 622  above average            low        absent
## 623  above average            low       present
## 624  below average         normal        absent
## 625  below average         normal        absent
## 626  above average         normal        absent
## 627  above average         normal       present
## 628  below average         normal        absent
## 629  below average         normal        absent
## 630  below average         normal        absent
## 631  below average           high       present
## 632  below average           high        absent
## 633  above average           high        absent
```

```
## 634  below average         normal        absent
## 635  below average           high       present
## 636  below average           high        absent
## 637  below average           high       present
## 638  below average           high       present
## 639  above average            low        absent
## 640  below average         normal        absent
## 641  above average         normal        absent
## 642  below average           high       present
## 643  above average         normal       present
## 644  below average         normal        absent
## 645  below average            low        absent
## 646  above average         normal       present
## 647  above average           high        absent
## 648  below average           high       present
## 649  above average         normal       present
## 650  below average           high       present
## 651  above average         normal        absent
## 652  above average         normal       present
## 653  below average         normal        absent
## 654  below average           high       present
## 655  above average            low        absent
## 656  above average         normal        absent
## 657  below average           high        absent
## 658  below average         normal       present
## 659  above average         normal       present
## 660  above average           high       present
## 661  above average         normal        absent
## 662  above average         normal        absent
## 663  above average         normal        absent
## 664  above average         normal       present
## 665  above average            low        absent
## 666  above average            low       present
## 667  below average           high        absent
## 668  below average         normal        absent
## 669  below average         normal        absent
## 670  below average         normal        absent
## 671  below average           high        absent
## 672  below average           high        absent
## 673  above average         normal       present
## 674  above average            low        absent
## 675  above average         normal        absent
## 676  below average           high       present
## 677  above average         normal        absent
## 678  above average         normal       present
## 679  above average         normal        absent
## 680  above average            low       present
## 681  below average         normal       present
## 682  above average         normal        absent
## 683  above average            low        absent
## 684  below average           high        absent
## 685  below average         normal        absent
## 686  above average         normal        absent
## 687  below average         normal        absent
```

```
## 688  below average         normal       present
## 689  below average         normal        absent
## 690  below average            low       present
## 691  above average         normal        absent
## 692  below average         normal        absent
## 693  below average         normal        absent
## 694  below average         normal        absent
## 695  above average         normal        absent
## 696  below average           high       present
## 697  below average         normal        absent
## 698  below average           high        absent
## 699  below average           high       present
## 700  below average           high        absent
## 701  above average         normal        absent
## 702  below average         normal       present
## 703  above average         normal        absent
## 704  above average            low        absent
## 705  below average         normal        absent
## 706  above average            low        absent
## 707  below average           high       present
## 708  below average         normal        absent
## 709  below average         normal        absent
## 710  below average         normal        absent
## 711  below average           high        absent
## 712  above average         normal       present
## 713  above average            low        absent
## 714  above average         normal        absent
## 715  below average         normal       present
## 716  below average         normal        absent
## 717  below average         normal       present
## 718  above average         normal       present
## 719  below average           high       present
## 720  above average         normal       present
## 721  below average         normal        absent
## 722  above average         normal        absent
## 723  below average           high        absent
## 724  below average         normal       present
## 725  above average         normal        absent
## 726  above average         normal        absent
## 727  above average         normal        absent
## 728  below average            low       present
## 729  above average         normal        absent
## 730  above average         normal       present
## 731  below average         normal        absent
## 732  below average           high       present
## 733  below average         normal       present
## 734  above average            low        absent
## 735  above average         normal        absent
## 736  below average         normal        absent
## 737  below average           high       present
## 738  below average         normal        absent
## 739  above average         normal       present
## 740  below average         normal        absent
## 741  below average            low        absent
```

```
## 742  above average         normal        absent
## 743  below average           high       present
## 744  below average         normal        absent
## 745  below average           high        absent
## 746  above average           high       present
## 747  below average         normal        absent
## 748  below average         normal        absent
## 749  below average         normal        absent
## 750  above average            low        absent
## 751  below average           high        absent
## 752  below average           high       present
## 753  above average            low        absent
## 754  above average            low        absent
## 755  below average           high       present
## 756  above average            low       present
## 757  below average         normal       present
## 758  below average         normal        absent
## 759  above average         normal        absent
## 760  below average           high       present
## 761  below average         normal       present
## 762  below average         normal       present
## 763  above average         normal       present
## 764  above average         normal        absent
## 765  below average           high        absent
## 766  below average           high       present
## 767  below average         normal       present
## 768  below average           high       present
## 769  below average         normal       present
## 770  above average         normal       present
## 771  below average         normal        absent
## 772  below average           high       present
## 773  above average           high        absent
## 774  above average         normal       present
## 775  below average         normal        absent
## 776  below average         normal       present
## 777  below average         normal        absent
## 778  above average         normal       present
## 779  below average         normal        absent
## 780  above average         normal        absent
## 781  above average         normal        absent
## 782  below average         normal       present
## 783  above average           high       present
## 784  below average         normal        absent
## 785  below average           high        absent
## 786  below average         normal        absent
## 787  above average         normal        absent
## 788  above average            low        absent
## 789  above average           high       present
## 790  below average           high       present
## 791  above average            low        absent
## 792  above average         normal        absent
## 793  above average         normal        absent
## 794  below average            low       present
## 795  below average           high        absent
```

```
## 796  above average         normal        absent
## 797  below average            low        absent
## 798  above average         normal        absent
## 799  below average            low        absent
## 800  below average         normal        absent
## 801  above average            low        absent
## 802  above average         normal        absent
## 803  below average         normal        absent
## 804  above average         normal       present
## 805  below average         normal        absent
## 806  below average         normal        absent
## 807  above average         normal        absent
## 808  below average         normal        absent
## 809  below average           high        absent
## 810  above average            low        absent
## 811  below average         normal        absent
## 812  below average         normal       present
## 813  below average            low        absent
## 814  below average         normal        absent
## 815  above average           high        absent
## 816  above average         normal        absent
## 817  above average         normal       present
## 818  below average           high       present
## 819  above average         normal        absent
## 820  below average         normal       present
## 821  below average         normal        absent
## 822  below average         normal        absent
## 823  below average           high       present
## 824  above average         normal        absent
## 825  below average         normal       present
## 826  above average         normal        absent
## 827  above average         normal       present
## 828  below average           high       present
## 829  below average         normal        absent
## 830  above average            low        absent
## 831  above average         normal        absent
## 832  below average           high       present
## 833  above average         normal       present
## 834  below average         normal        absent
## 835  above average            low        absent
## 836  above average         normal        absent
## 837  below average         normal        absent
## 838  above average            low        absent
## 839  above average         normal        absent
## 840  above average         normal        absent
## 841  below average         normal        absent
## 842  below average           high        absent
## 843  below average           high       present
## 844  below average         normal        absent
## 845  below average         normal        absent
## 846  above average         normal        absent
## 847  below average           high       present
## 848  above average            low       present
## 849  above average            low       present
```

```
## 850  below average         normal       present
## 851  above average         normal       present
## 852  below average           high       present
## 853  below average           high        absent
## 854  above average            low        absent
## 855  below average           high        absent
## 856  above average            low        absent
## 857  below average         normal        absent
## 858  above average            low        absent
## 859  above average            low        absent
## 860  above average         normal        absent
## 861  above average            low        absent
## 862  above average         normal        absent
## 863  above average           high        absent
## 864  below average         normal        absent
## 865  above average         normal       present
## 866  above average         normal        absent
## 867  above average         normal       present
## 868  below average         normal        absent
## 869  above average         normal        absent
## 870  below average         normal        absent
## 871  below average         normal        absent
## 872  above average         normal        absent
## 873  above average         normal        absent
## 874  below average         normal        absent
## 875  below average         normal        absent
## 876  below average         normal        absent
## 877  below average         normal        absent
## 878  below average         normal        absent
## 879  below average           high        absent
## 880  above average         normal        absent
## 881  above average         normal       present
## 882  above average         normal        absent
## 883  below average         normal       present
## 884  above average         normal       present
## 885  above average            low        absent
## 886  above average            low        absent
## 887  above average         normal       present
## 888  above average            low        absent
## 889  above average           high       present
## 890  above average            low        absent
## 891  below average         normal       present
## 892  above average         normal        absent
## 893  above average         normal       present
## 894  below average           high        absent
## 895  above average            low        absent
## 896  below average           high       present
## 897  above average         normal       present
## 898  below average           high       present
## 899  above average         normal        absent
## 900  below average           high        absent
## 901  below average         normal        absent
## 902  below average         normal        absent
## 903  above average         normal        absent
```

```
## 904  above average            low        absent
## 905  below average         normal        absent
## 906  above average           high       present
## 907  below average         normal        absent
## 908  below average         normal       present
## 909  above average         normal       present
## 910  below average           high       present
## 911  below average         normal        absent
## 912  below average         normal        absent
## 913  below average         normal        absent
## 914  above average         normal        absent
## 915  above average         normal       present
## 916  below average         normal        absent
## 917  below average         normal       present
## 918  below average         normal       present
## 919  below average         normal       present
## 920  above average         normal        absent
## 921  below average           high        absent
## 922  above average         normal        absent
## 923  below average         normal        absent
## 924  above average           high        absent
## 925  below average         normal        absent
## 926  above average         normal       present
## 927  below average         normal       present
## 928  below average           high       present
## 929  above average           high        absent
## 930  above average           high        absent
## 931  below average         normal       present
## 932  below average           high       present
## 933  above average         normal        absent
## 934  below average         normal        absent
## 935  below average           high       present
## 936  below average         normal        absent
## 937  below average         normal        absent
## 938  below average           high       present
## 939  below average         normal        absent
## 940  below average         normal       present
## 941  below average         normal       present
## 942  below average         normal        absent
## 943  below average         normal        absent
## 944  above average            low        absent
## 945  above average            low        absent
## 946  above average         normal        absent
## 947  above average         normal        absent
## 948  below average         normal        absent
## 949  above average         normal       present
## 950  below average         normal       present
## 951  above average         normal        absent
## 952  below average         normal       present
## 953  above average         normal        absent
## 954  below average           high        absent
## 955  above average         normal       present
## 956  below average           high       present
## 957  above average         normal       present
```

```
## 958  below average            low        absent
## 959  above average           high        absent
## 960  above average            low       present
## 961  below average           high       present
## 962  above average         normal        absent
## 963  above average         normal        absent
## 964  below average         normal        absent
## 965  below average           high        absent
## 966  above average         normal        absent
## 967  below average         normal        absent
## 968  above average         normal        absent
## 969  above average            low       present
## 970  below average           high        absent
## 971  below average         normal       present
## 972  below average         normal       present
## 973  above average            low        absent
## 974  above average         normal        absent
## 975  above average         normal       present
## 976  above average         normal       present
## 977  above average         normal        absent
## 978  above average         normal        absent
## 979  above average         normal        absent
## 980  below average         normal        absent
## 981  below average         normal        absent
## 982  below average         normal       present
## 983  below average         normal        absent
## 984  above average         normal        absent
## 985  above average         normal        absent
## 986  above average            low        absent
## 987  above average         normal        absent
## 988  below average           high       present
## 989  below average            low        absent
## 990  below average            low       present
## 991  above average           high        absent
## 992  below average           high       present
## 993  above average         normal       present
## 994  below average         normal        absent
## 995  below average           high       present
## 996  below average         normal       present
## 997  below average         normal        absent
## 998  above average         normal        absent
## 999  below average         normal       present
## 1000 above average         normal       present
```

```
delta <- calc_delta(exer)
print(delta)
```

```
## [1] NaN
```

4.2 [10 pts] Stratify the table by blood pressure. As in the previous problem, calculate the same treatment effect in each strata.

4.3 [5 pts] Based on the summary of the treatment effect that you observed in the combined and stratified tables, is exercise associated with lower rates of heart disease?

It would depend on whether or not the values were postive or negative.

4.4 [5 pts] Which measure of treatment effect is most persuasive? The combined estimate or the stratified estimates? Which estimate(s) should you rely on? Explain why, creating a DAG to represent relationship between the variables.

assuming the startfied tables reveal infomrtion, a DAG could look like:

Exersise amount –> Blood pressure (D) –> heart disease (R) | +——————————————————> heart disease (R)