# STAT 4410/8416 Homework 3

*VanSteenbergen Nicolaas*

*Due on Oct 17, 2018*

```
bigDataSample <- read.csv(file='C:/Users/Nico/Desktop/Data Science/bigDataSample.csv')
```

*1)*

*a)*

```
dat <- bigDataSample %>%
  select(contains('human'))

head(dat)
```

```
##   var_human_1_g var_human_1_p var_human_1_b var_human_1_e var_human_1_n
## 1      18.99545            21             1    21.6321136      26.03268
## 2      15.02303            34             3     0.3838458      26.92529
## 3      37.44410            28             2    33.4801022      39.30039
## 4      36.33714            26             2     2.8761174      33.75177
## 5      21.06330            25             1     3.1657313      26.19248
## 6      16.52637            35             2     5.3108922      25.07192
```

*b)*

```
col_length <- nchar(colnames(dat))
colnames(dat) <- substr(colnames(dat),col_length,col_length)
head(dat)
```

```
##          g  p b          e        n
## 1 18.99545 21 1 21.6321136 26.03268
## 2 15.02303 34 3  0.3838458 26.92529
## 3 37.44410 28 2 33.4801022 39.30039
## 4 36.33714 26 2  2.8761174 33.75177
## 5 21.06330 25 1  3.1657313 26.19248
## 6 16.52637 35 2  5.3108922 25.07192
```

*c)*

```
library(knitr)

dat <-dat %>%
  group_by(b) %>%
  summarize_all(funs(mean))

kable(dat)
```

| b | g | p | e | n |
|---|---|---|---|---|
| 0 | 28.74877 | 23.75862 | 12.214718 | 29.44332 |
| 1 | 22.47859 | 25.28302 | 10.418129 | 29.34315 |
| 2 | 23.85395 | 24.94624 | 9.615341 | 30.62800 |
| 3 | 23.81182 | 25.40909 | 10.481746 | 30.25341 |

*d)*

```
mdat <- 0
```