☠ ☠ ☠ ☠ ☠ ☠ ☠ ☠ ☠ ☠ ☠ ☠ ☠ ☠

MORT

GREETINGS, I AM DEATH. I AM A BUSY PROFESSIONAL. I SELECT WHOSE LIFE TO TAKE CAREFULLY. I STUDIED MORE THAN 21 000 SENIOR CITIZENS LIVING AROUND CENTRAL EUROPE IN 2007. I TOOK 1200 OF THEM TO THE OTHERSIDE BY 2013. THIS IS ONE WAY OF PICKING THE RIGHT PEOPLE.

YOU, MY YOUNG APPRANTICE, FOLLOW ALONG IF YOU WANT TO BECOME AN ARCHANGEL ONE DAY...

WHAT IS THE MORTALITY RATE IN THIS SAMPLE?

```
Death...
             Freq    Percent
nope        20595    94.265
yep          1253     5.735
Total       21848    100
```

YOU MIGHT WANT TO LOOK AT THEIR SPORTS HABITS

THIS IS HOW THE PEOPLE I TOOK TO THE OTHER SIDE WERE DOING SPORTS. GROUP 1 DOES A LOT OF SPORTS. 2 AND 3 ARE FREQUENTERS. GROUP 4 THE FOOLS, ALMOST NEVER, BIG MISTAKE...
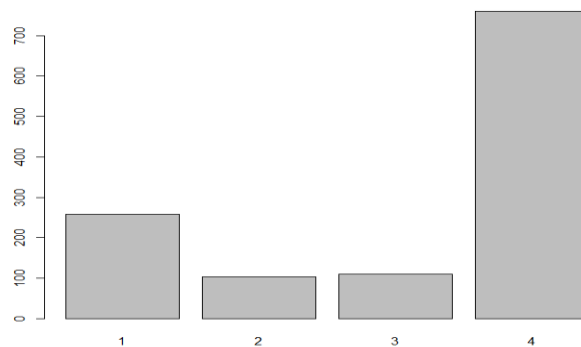

```
================================================
          DEPENDENT VARIABLE:
          --------------------------
               DECEASED
-------------------------------------------
SPORTS_ALOT          -0.059***
                     (0.004)

SPORTS_OFTEN         -0.059***
                     (0.005)

SPORTS_SOME          -0.042***
                     (0.006)

CONSTANT              0.092***
                     (0.004)

-------------------------------------------
OBSERVATIONS          21,848
R2                    0.014
================================================
          NOTE:       ***P<0.01
```

These results show differences in MORTality rates among people with different sports habits in observational data. The six-year mortality probability is **9.2%** among the people who don't do sports. Those who DO sports *sometime* have 4.2 percentage points less mortality than inactive people. Those who do sports *often and alot* have 5.9 percentage points less mortality than inactive people. We can say doing sport often and a lot are almost the same thing in this data for the senior people.

These differences are statistically significant at the 0.1% level: the pvalue is less than 0.1%.
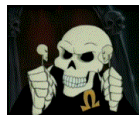
Without the ability to run experiments we try to get closer to the causal effect by controlling for confounders in the regression. Our data has information on gender, age, education and the income group they belong to in their countries. This can get us closer to the causal effect we are after.

```
==================================================
                DEPENDENT VARIABLE:
                ------------------------
                    DECEASED
--------------------------------------------------------------------------------
    SPORTS_ALOT                          -0.044***
                                         (0.004)

    SPORTS_OFTEN                         -0.045***
                                         (0.005)

    SPORTS_SOME                          -0.037***
                                         (0.005)


    EVER_SMOKED, FEMALE, AGE, EDU, INCOME    YES

--------------------------------------------------------------------------------
            OBSERVATIONS          21,848
            R2                    0.049
==================================================
                NOTE:        ***P<0.01
```

Those who "*sport sometime*" have 3.7 percentage points less MORTality than "*inactive people*", for those of the same gender, age, years of education and income group. Checking the same for "*sports-often and sports-a lot*" show 4.5 to 4.4 percentage point less mortality rate than the inactive people. Although a bit decreased, these differences are like our first model, supporting the idea of sport's effect on mortality, still significant. We have higher coefficients in the first regression due to the contribution of the other variables we didn't take into account.

Coefficients are again statistically significant at the 0.1% level: the pvalue is less than 0.1%.

There are still lots of potential variables which can influence the mortality outcome (such as my boss the God's Will). We might get closer to the causal interpretation, but we are still not there.
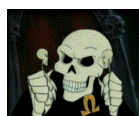
Looking into Logit Marginal Differencess

```
Logitmfx formula = deceased ~ sports_alot + sports_often + sports_some

                Marginal Effects:
                        dF/dx  Std. Err.
        sports_alot   -0.051  0.003 ***
        sports_often  -0.043  0.003 ***
        sports_some   -0.028  0.003 ***
                ***p<0.01 **p<0.05
```

Average difference in the probablity of MORTality decreases by 5.1 percentage point for the people who sport a lot compared to the people who never exercise. The same logic applies to the "sport often" and "sport some" groups for 4.3 and 2.8 percentage point.

```
logitmfx formula = deceased ~ sports_alot + sports_often + sports_some +
                   female + age + eduyears_mod + income10g

                    Marginal Effects:
                            dF/dx    Std. Err.
              sports_alot   -0.037   0.004***
              sports_often  -0.034   0.005***
              sports_some   -0.025   0.006***
              female        -0.038   0.005***
              age            0.004   0.000***
              eduyears_mod  -0.002   0.001***
              income10g     -0.001   0.001**
                    ***p<0.01 **p<0.05
```

Adding our socio-economic confounders one more time. These variables are very much likely to be correlated with life expectancy and sports habits.

Average difference in the probablity of MORTality decreases by 3.7 percentage point for people of same gender, age, years of education and income group who sport "*a lot*" compared to people who never exercise with the same conditions. The same logic applies to "sports often" and "sports sometimes" groups for 3.4 and 2.5 percentage points.

If we look at the first marginal differences table these differences were higher: 5.1, 4.3 and 2.8 respectively. When we controlled for additional variables these have reduced. Because these additional variables are also related to mortality. This has happened for LPM also which is another similarity.

These probability differences are substantial, as we should not forget that the mortality rate in this sample was 5.7%

LOGIT GIVES A BETTER FIT THAN THE LINEAR PROBABILITY MODELS. LPMS HAVE THE ISSUE OF GIVING RESULTS GREATER THAN 1 OR LESS THAN 0. PROBABILITIES AT THAT LEVEL ARE SUCH EVENTS ONLY MY BOSS THE GOD CAN DEAL WITH. SO IT IS BETTER FOR YOUNG ANGELS TO USE LOGIT WHICH WILL ALWAYS GIVE PROBABILITIES WITHIN THE LIMITS OF SPACE-TIME FABRIC.
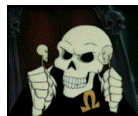
MORT

ON THE OTHER SIDE WE TAKE DECISIONS BASED ON DATA ONLY. BECAUSE IT IS A MATTER OF LIFE AND DEATH YOU KNOW...

☠ ☠ ☠ ☠ ☠ ☠ ☠ ☠ ☠ ☠ ☠ ☠ ☠ ☠