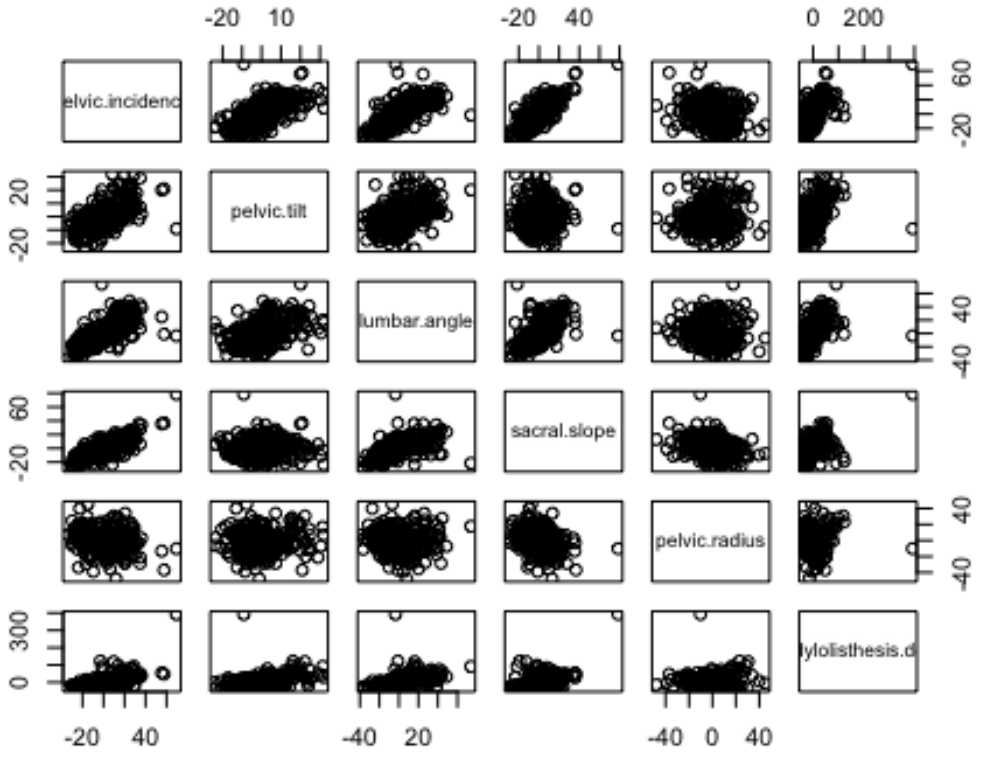
INTRODUCTION

For this project, I will analyze a biomedical data set about patients having abnormalities in their lower back. A patient is considered abnormal if they have either a disk hernia or Spondylolisthesis. The data set consists of seven features which are, pelvic incidence, pelvic tilt, lumbar angle, sacral slope, pelvic radius and grade of spondylolisthesis. The aim of this project will be to see which of these features are the most important for predicting abnormalities in the spine. Since the features are all interval variables such as degree or length, the data points have been centered by subtracting the mean of the corresponding feature.



From the graph above, we can see that some features are highly correlated with one another. This intuitively makes sense, as the lower back is interconnected and works in a mechanical way.

Model One Output (includes all features):

```
## Call:
## glm(formula = class ~ pelvic.incidence + pelvic.tilt + lumbar.angle +
##     sacral.slope + pelvic.radius + spondylolisthesis.degree,
```

```
##     family = binomial, data = vertebral)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.7317  -0.4081   0.0289   0.3639   2.2678
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)               3.216e+00  4.803e-01   6.696 2.14e-11 ***
## pelvic.incidence         -2.517e+07  4.017e+07  -0.627    0.531
## pelvic.tilt               2.517e+07  4.017e+07   0.627    0.531
## lumbar.angle             -1.794e-02  2.290e-02  -0.784    0.433
## sacral.slope              2.517e+07  4.017e+07   0.627    0.531
## pelvic.radius            -1.077e-01  2.318e-02  -4.645 3.39e-06 ***
## spondylolisthesis.degree  1.693e-01  2.335e-02   7.248 4.23e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 389.86  on 309  degrees of freedom
## Residual deviance: 177.87  on 303  degrees of freedom
## AIC: 191.87
##
## Number of Fisher Scoring iterations: 8
```

Since our response variable binary, we will be fitting a logistic regression. The first model will include all the features so we can ascertain the significant ones. From our summary, we see that pelvic tilt, pelvic incidence, lumbar angle, and pelvic radius are all not significant. In fact, pelvic tilt and sacral slope have the exact same coefficients, standard error, and p-value. While pelvic incidence also has the exact same but the negative.

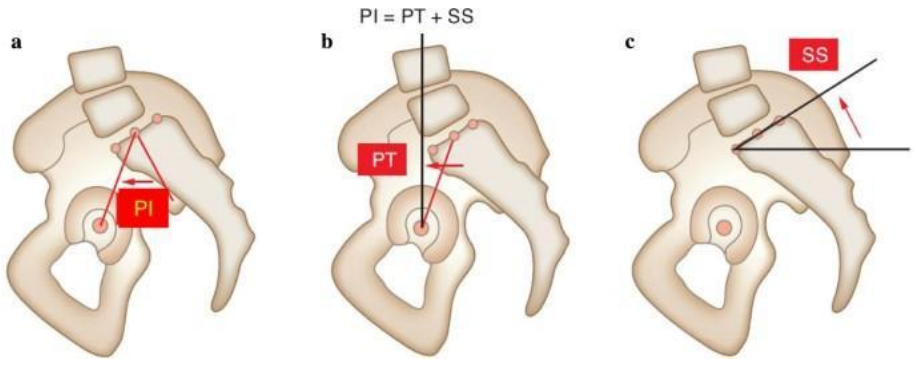


Figure 1: pelvic incidence, pelvic tilt, and sacral slope

The strange output can be explained by figure 1. It turns out, pelvic incidence is just the sum of the pelvic tilt and sacral slope. So essentially, we are measuring the same value twice making it redundant. Therefore, we will update our model by dropping pelvic incidence and lumbar angle.

Model Two Output (excludes pelvic incidence and lumbar angle):

```
## Call:
## glm(formula = class ~ pelvic.tilt + sacral.slope + pelvic.radius +
##     spondylolisthesis.degree, family = binomial, data = vertebral)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.7927  -0.4019   0.0319   0.3795   2.2198
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)                3.18805    0.47683   6.686 2.29e-11 ***
## pelvic.tilt                0.06647    0.02925   2.272   0.0231 *
## sacral.slope              -0.11495    0.02266  -5.072 3.94e-07 ***
## pelvic.radius             -0.10893    0.02279  -4.780 1.75e-06 ***
## spondylolisthesis.degree   0.16541    0.02289   7.227 4.92e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 389.86  on 309  degrees of freedom
## Residual deviance: 178.94  on 305  degrees of freedom
## AIC: 188.94
##
## Number of Fisher Scoring iterations: 8
```

As we can see, dropping these features make all the others significant. Also, to note that the parameters for pelvic tilt and sacral slope are not identical anymore.

AIC Matrix of interaction terms:

```
##                          pelvic.tilt sacral.slope pelvic.radius spondylolisthesis
## pelvic.tilt                   0.0000     190.8427      190.4010          189.9562
## sacral.slope                190.8427       0.0000      190.5201          183.8291
## pelvic.radius               190.4010     190.5201        0.0000          188.7804
## spondylolisthesis           189.9562     183.8291      188.7804            0.0000
```

However, as mentioned earlier, the lower back works in a mechanical way and therefore interaction terms may be present. When we try different combinations of features, nearly all interactions increase the AIC apart from two, which is highlighted in the AIC matrix. This is just a matrix of AIC scores. The biggest decrease in AIC happens with the inclusion of sacral slope interacting with the degree of spondylolisthesis.

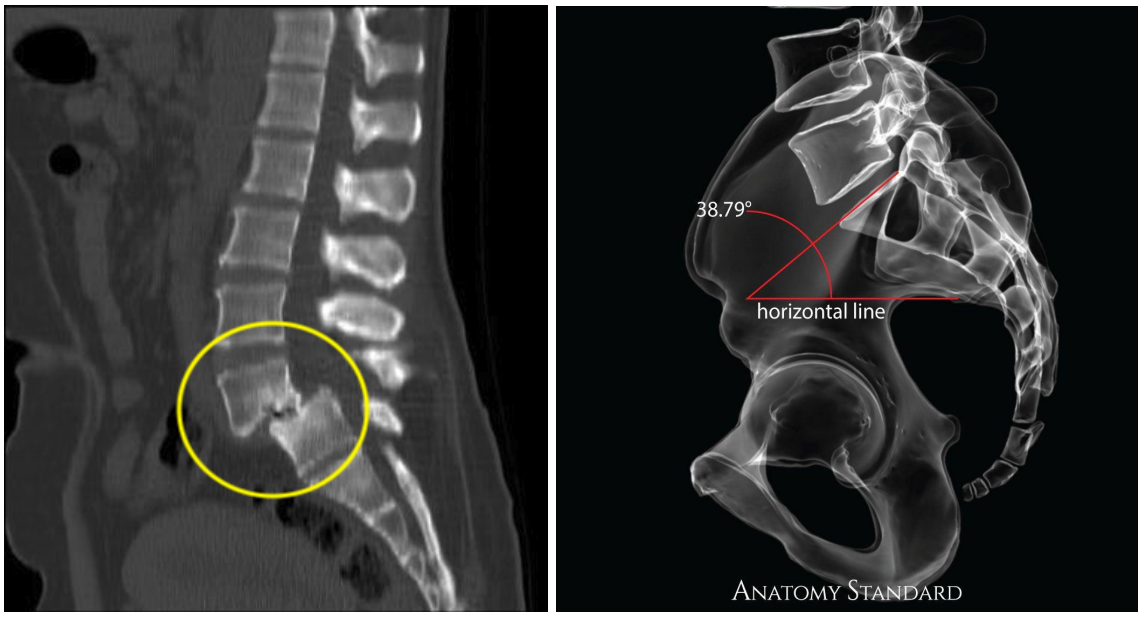


Figure 2: LHS spondylolisthesis, RHS sacral slope

Biomechanically this interaction makes sense. The LHS of figure 2 depicts spondylolisthesis. Essentially, spondylolisthesis happens when one of your lower vertebral slips out of place as shown by the yellow circle. The red lines in the RHS of figure 2 shows the sacral slope. We can see that as the sacral slope angle increases, the lower vertebral will move towards a more vertical position. Therefore, the probability of gravity pulling forward one of the disks increases and thereby getting spondylolisthesis.

Final Model Output:

```
## Call:
## glm(formula = class ~ pelvic.tilt + sacral.slope + pelvic.radius +
##     spondylolisthesis.degree + sacral.slope * spondylolisthesis.degree,
##     family = binomial, data = vertebral)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.72491  -0.39572   0.01009   0.26768   2.36282
##
## Coefficients:
##                                      Estimate Std. Error z value
Pr(>|z|)
## (Intercept)                          3.022399   0.563454   5.364
8.14e-08 ***
## pelvic.tilt                          0.071785   0.030491   2.354
0.0186 *
## sacral.slope                        -0.000337   0.054049  -0.006
0.9950
```

```
## pelvic.radius                            -0.098314   0.022756  -4.320
1.56e-05 ***
## spondylolisthesis.degree                  0.167438   0.026418   6.338
2.33e-10 ***
## sacral.slope:spondylolisthesis.degree  0.005311   0.002284   2.325
0.0200 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 389.86  on 309  degrees of freedom
## Residual deviance: 171.83  on 304  degrees of freedom
## AIC: 183.83
##
## Number of Fisher Scoring iterations: 9
```

When we look at the summary output of this model, the interaction between sacral slope and degree of spondylolisthesis is significant. Therefore, this will be our final model.

By taking the inverse logit of the intercept we get 0.95 probability of being abnormal, which is a rather high probability. However, this could be due to the underlying data having uneven balances of normal and abnormal cases, as 2/3rd of the dataset is abnormal.

When we try to interpret the coefficient by taking the exp(coefficients), it will show the increase or decrease in odds of being abnormal. This means that the positive coefficients increase log odds by their respective exp(B_i) while the negative will have a decrease in odds.

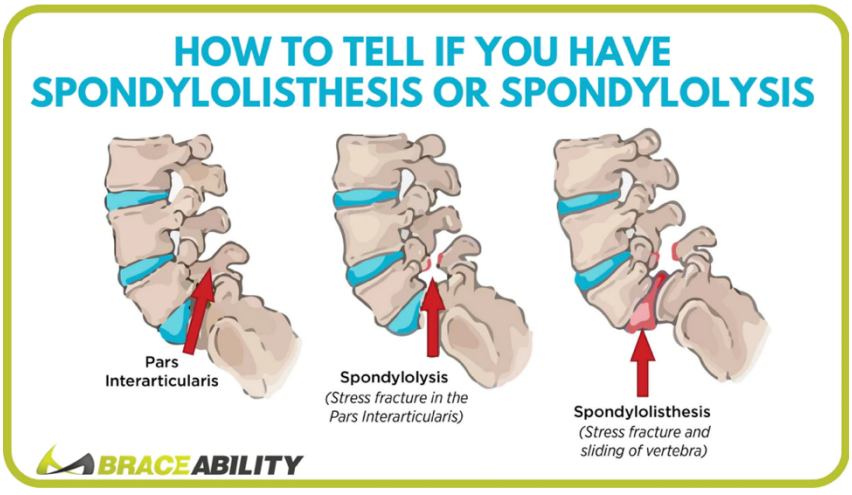


Figure 3: Spondylolisthesis

Increase in pelvic tilt will increase odds of abnormality by exp (0.0717), which means the probability of a disk to slip forward goes up as shown in figure 3.

An increase in pelvic radius will decrease abnormality. While increases in the degree of spondylolisthesis will increase probability of abnormalities, because it is a direct measure of slippage as shown previously.

All in all, increases in pelvic tilt and the curvature of the lower spine is likely to cause more abnormality. Thus, we should all strive to have proper posture to avoid lower back injuries.

CODE

```
#Load data and process
vertebral.data <- read.csv("vertebral.csv")
norm.or.abnorm = rep(0, 310)
norm.or.abnorm[vertebral.data$class == "Abnormal"] = 1
vertebral <- data.frame(pelvic.incidence = vertebral.data$pelvic_incidence -
mean(vertebral.data$pelvic_incidence),
                        pelvic.tilt = vertebral.data$pelvic_tilt -
mean(vertebral.data$pelvic_tilt),
                        lumbar.angle = vertebral.data$lumbar_lordosis_angle -
mean(vertebral.data$lumbar_lordosis_angle),
                        sacral.slope = vertebral.data$sacral_slope -
mean(vertebral.data$sacral_slope),
                        pelvic.radius = vertebral.data$pelvic_radius -
mean(vertebral.data$pelvic_radius),
                        spondylolisthesis.degree =
vertebral.data$degree_spondylolisthesis -
mean(vertebral.data$degree_spondylolisthesis),
                        class = norm.or.abnorm)

dim(vertebral.data)

## [1] 310   8

plot(vertebral[1:6])

#Fitting linear models to select right features
vertebral.fit.glm <- glm(class ~ pelvic.incidence + pelvic.tilt +
lumbar.angle + sacral.slope + pelvic.radius + spondylolisthesis.degree, data
= vertebral, family = binomial)

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

summary(vertebral.fit.glm)

#Dropping pelvic incidence and lumbar angle
vertebral.fit.glm3 <- glm(class ~ pelvic.tilt + sacral.slope + pelvic.radius
+ spondylolisthesis.degree, data = vertebral, family = binomial)

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

summary(vertebral.fit.glm3)

#AIC without interaction terms
AIC(vertebral.fit.glm3)

## [1] 188.9381

#Interaction Terms
vertebral.fit.glm4 <- glm(class ~ pelvic.tilt + sacral.slope + pelvic.radius
```

```
+ spondylolisthesis.degree + pelvic.tilt*sacral.slope, data = vertebral,
family = binomial)

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

AIC(vertebral.fit.glm4)

## [1] 190.8427

vertebral.fit.glm5 <- glm(class ~ pelvic.tilt + sacral.slope + pelvic.radius
+ spondylolisthesis.degree + pelvic.tilt*pelvic.radius, data = vertebral,
family = binomial)

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

AIC(vertebral.fit.glm5)

## [1] 190.401

vertebral.fit.glm6 <- glm(class ~ pelvic.tilt + sacral.slope + pelvic.radius
+ spondylolisthesis.degree + pelvic.tilt*spondylolisthesis.degree, data =
vertebral, family = binomial)

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

AIC(vertebral.fit.glm6)

## [1] 189.9562

vertebral.fit.glm7 <- glm(class ~ pelvic.tilt + sacral.slope + pelvic.radius
+ spondylolisthesis.degree + sacral.slope*pelvic.radius, data = vertebral,
family = binomial)

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

AIC(vertebral.fit.glm7)

## [1] 190.5201

vertebral.fit.glm8 <- glm(class ~ pelvic.tilt + sacral.slope + pelvic.radius
+ spondylolisthesis.degree + sacral.slope*spondylolisthesis.degree, data =
vertebral, family = binomial)

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

AIC(vertebral.fit.glm8)

## [1] 183.8291

vertebral.fit.glm9 <- glm(class ~ pelvic.tilt + sacral.slope + pelvic.radius
+ spondylolisthesis.degree + pelvic.radius*spondylolisthesis.degree, data =
vertebral, family = binomial)

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
AIC(vertebral.fit.glm9)
```

```
## [1] 188.7804
```

```
#Matrix of AIC interaction terms
row.names <- c("pelvic.tilt", "sacral.slope", "pelvic.radius",
"spondylolisthesis.degree")
column.names <- c("pelvic.tilt", "sacral.slope", "pelvic.radius",
"spondylolisthesis.degree")
mat <- c(0, AIC(vertebral.fit.glm4), AIC(vertebral.fit.glm5),
AIC(vertebral.fit.glm6),
        AIC(vertebral.fit.glm4), 0 , AIC(vertebral.fit.glm7),
AIC(vertebral.fit.glm8),
        AIC(vertebral.fit.glm5), AIC(vertebral.fit.glm7), 0,
AIC(vertebral.fit.glm9),
        AIC(vertebral.fit.glm6), AIC(vertebral.fit.glm8),
AIC(vertebral.fit.glm9), 0)

named_matrix <- matrix(mat, nrow=4, byrow=TRUE, dimnames=list(row.names,
column.names))
named_matrix
```

```
#Summary of best interaction term
summary(vertebral.fit.glm8)
```