Department of Mathematics and Computer Science
University of Southern Denmark, Odense December 8, 2020

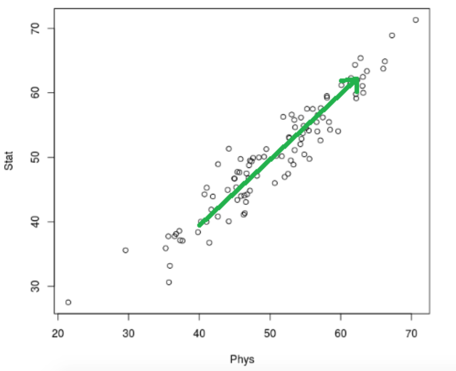
# DM561 — Linear Algebra with Applications

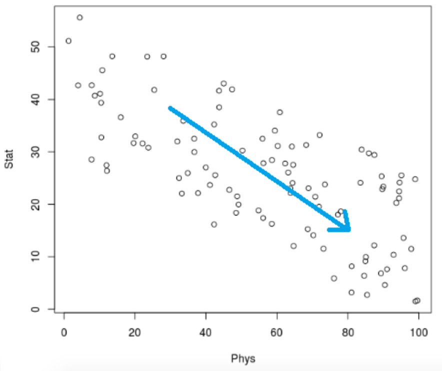
## Sheet 8, Fall 2020

Author: Robert Rasmussen

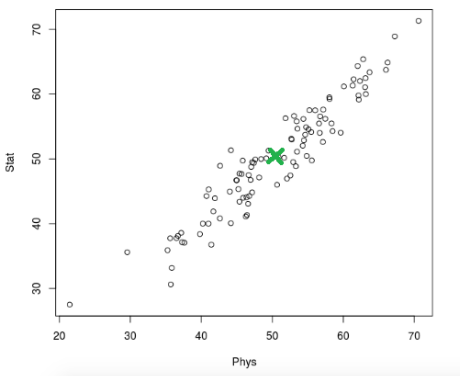
## Exercise 1

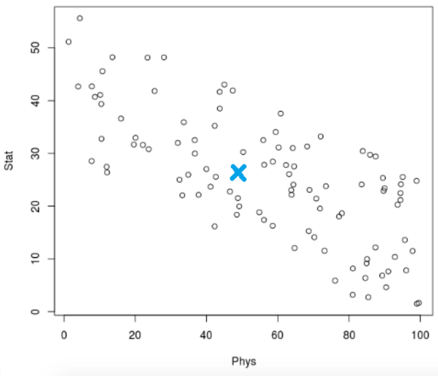
1. The direction the data varies the most are marked below






2. The point representing the mean value is marked below (roughly):






## Exercise 2

1. The matrix is

$$X' = \begin{bmatrix} 90 & 60 & 90 \\ 90 & 90 & 30 \\ 60 & 60 & 60 \\ 60 & 60 & 90 \\ 30 & 30 & 30 \end{bmatrix}$$

   The i'th row represents student $i$ (1-indexed), ie. the feature values for that student. The first column holds math grades, the second english grades, and the third art grades.

2. The averages for the features are

- Math: $\frac{1}{5} \cdot (90 + 90 + 60 + 60 + 30) = 66$
- English: $\frac{1}{5} \cdot (90 + 90 + 60 + 60 + 30) = 60$
- Art: $\frac{1}{5} \cdot (90 + 90 + 60 + 60 + 30) = 60$

For each entry we subtract the corresponding mean for the column. We obtain

$$X = \begin{bmatrix} 66-90 & 60-60 & 60-90 \\ 66-90 & 60-90 & 60-30 \\ 66-60 & 60-60 & 60-60 \\ 66-60 & 60-60 & 60-90 \\ 66-30 & 60-30 & 60-30 \end{bmatrix} = \begin{bmatrix} -24 & 0 & -30 \\ -24 & -30 & 30 \\ 6 & 0 & 0 \\ 6 & 0 & -30 \\ 36 & 30 & 30 \end{bmatrix}$$

It trivial to observe the mean of each column (each feature) is 0.

3. The covariance is probably

   a) positive are relatively high. It seems that there are correlation between the grade for Math and English; when one is high the other is also high.

   b) close to 0. It seems that there are no tendencies indicating a correlation between grades for English and Art.

4. We should use the formulae $C_X = \frac{1}{n} X^T X$. Each row of $X$ correspond to all $m = 3$ measurements (a Mathgrade , an English grade, and an Art grade) from one particular trial (a student). Each column corresponds to all $n = 5$ measurements of a particular type (either Math grades, English grades, or Art grades).

   Remark: It important that $X$ is mean centered.

5. Let's compute the covariance without Bessel's correction. Let $M = \{90, 90, 60, 60, 30\}$ be the Math grades, $E = \{60, 90, 60, 60, 30\}$ the English grades, and $A = \{90, 30, 60, 90, 30\}$ the Art grades. The mean of

   - $M$ is $\bar{m} = 66$
   - $E$ is $\bar{e} = 60$
   - $A$ is $\bar{a} = 60$

   cf. (2).

   Now we can compute

   - $\sigma^2_{MM} = \frac{1}{5} \sum_{i=1}^{5} (m_i - \bar{m})^2 = 504$
   - $\sigma^2_{EE} = \frac{1}{5} \sum_{i=1}^{5} (e_i - \bar{e})^2 = 360$
   - $\sigma^2_{AA} = \frac{1}{5} \sum_{i=1}^{5} (a_i - \bar{a})^2 = 720$
   - $\sigma^2_{ME} = \frac{1}{5} \sum_{i=1}^{5} (m_i - \bar{m})(e_i - \bar{e}) = 360$
   - $\sigma^2_{MA} = \frac{1}{5} \sum_{i=1}^{5} (m_i - \bar{m})(a_i - \bar{a}) = 180$
   - $\sigma^2_{EA} = \frac{1}{5} \sum_{i=1}^{5} (e_i - \bar{e})(a_i - \bar{a}) = 0$

   It would of course have been easier just to use the formulae from (4). Let's do that in (6). Here we also use Bessel's correction.

6. We have that

   a) the covariance matrix with Bessel's correction is

$$C_X = \frac{1}{5-1} X^T X = \begin{bmatrix} 630 & 450 & 225 \\ 450 & 450 & 0 \\ 225 & 0 & 900 \end{bmatrix}$$

   b) the covariance matrix without Bessel's correction is

$$C_X = \frac{1}{5} X^T X = \begin{bmatrix} 504 & 360 & 180 \\ 360 & 360 & 0 \\ 180 & 0 & 720 \end{bmatrix}$$

   For $X$ we use the mean centered matrix from (2).

# Exercise 3

1) We construct a transformation matrix $W$ of the form

$$W = [w_1 \mid w_2]$$

   Afterwards, we make the projection onto the new feature space by computing $Y = X \cdot W$.

   This is not quite the formulae from [1, p. 7] but it's close. By transposing each side we get

$$Y = X \cdot W \Leftrightarrow Y^T = W^T \cdot X^T$$

   The matrix $W^T$ corresponds to the matrix $P$ from [1, p. 7] as the row vectors of $P$ should be the new basis vectors for expressing the columns of $X^T$. The new feature space's basis vectors are the vectors $w_1$ and $w_2$. These are the column vectors of $W$; thus, they are the row vectors of $W^T$. The matrix $X^T$ has features as rows and samples as columns (since $X$ had samples as rows and different features as columns). The matrix $Y^T$ is the new representation of the data with principle components as rows (in a sense "new features") and samples as columns.

2) The size of $W$ is $4 \times 2$. The size of $X$ is $150 \times 4$. Since $Y$ is computed as $X \cdot W$ its size becomes $150 \times 2$ (in a sense we have the same samples but now the features have changed to the principle components).

# Exercise 4

1) Just do it!

2) Just do it!

3) Just do it!

4) Add this after cell 8:

```
# First image of test dataset X_test. Note: A row of X_test is an image.
v = X_test[0]

# Project first image of X_test onto eigenface-space (space spanned by the
    150 principle components).
v_proj = pca.transform([v])
print(v_proj)
```

# References

[1] Daniel Merkle. Principal component analysis and eigenfaces. URL https://dm561.github.io/assets/DM561-DM562-PCA-Eigenfaces.pdf, 2020.