Principal Components and Factor Analysis: a tale of two dimension reduction techniques

Andrew Blandino

RTG – University of California, Davis

May 17th, 2019

Parkinsons example

- [Sakar et al.2019] collected audio recordings of 188 patients with Parkinson's Disease (PD) (107 men and 81 women).

Parkinsons example

- [Sakar et al.2019] collected audio recordings of 188 patients with Parkinson's Disease (PD) (107 men and 81 women).
- Data was also collected on a control group of 64 healthy patients (23 men and 41 women).

Parkinsons example

- [Sakar et al.2019] collected audio recordings of 188 patients with Parkinson's Disease (PD) (107 men and 81 women).
- Data was also collected on a control group of 64 healthy patients (23 men and 41 women).
- Each subject was recorded three times.

Parkinsons example

- [Sakar et al.2019] collected audio recordings of 188 patients with Parkinson's Disease (PD) (107 men and 81 women).
- Data was also collected on a control group of 64 healthy patients (23 men and 41 women).
- Each subject was recorded three times.
- The variables of interest consist of 754 auditory measurements for each patient.

Parkinsons example

- The data can be expressed in the following table:

ID	gender	status	PPE	DFA	RPDE	...
1	M	$+$	0.852	0.718	0.572	...
2	M	$+$	0.767	0.695	0.540	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	...	

Parkinsons example

- The data can be expressed in the following table:

ID	gender	status	PPE	DFA	RPDE	...
1	M	+	0.852	0.718	0.572	...
2	M	+	0.767	0.695	0.540	...
⋮	⋮	⋮	⋮	⋮	...	

- Time to plot the data ...

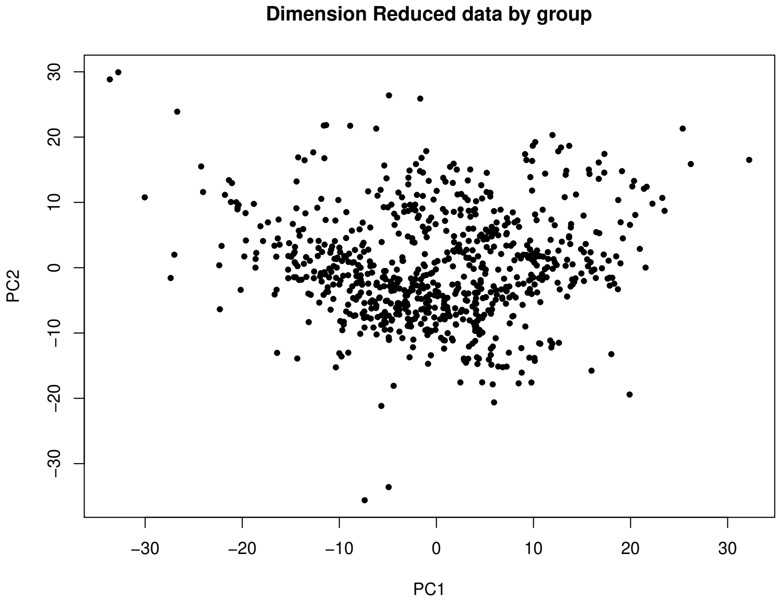
Parkinsons example

- The data can be expressed in the following table:

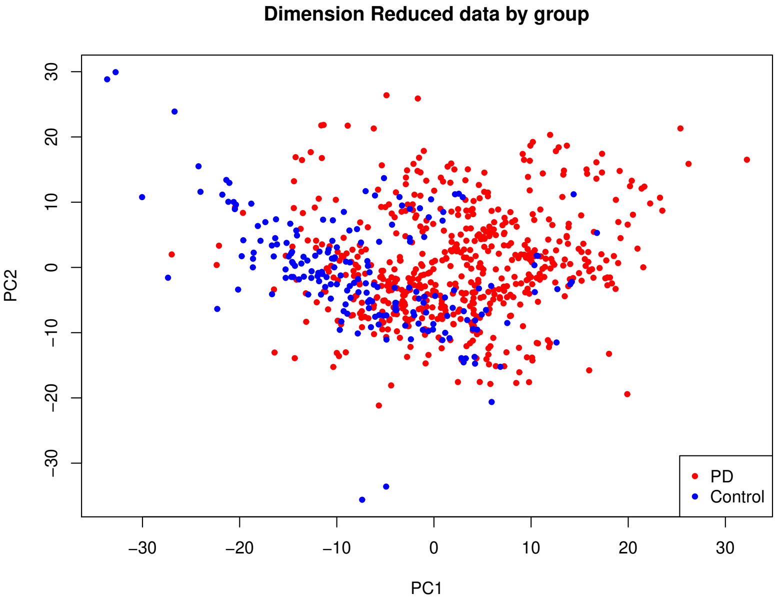
ID	gender	status	PPE	DFA	RPDE	...
1	M	+	0.852	0.718	0.572	...
2	M	+	0.767	0.695	0.540	...
⋮	⋮	⋮	⋮	⋮	...	

- Time to plot the data ...
- With Principal components analysis (PCA), we can obtain the following *two*-dimensional plot of the data.

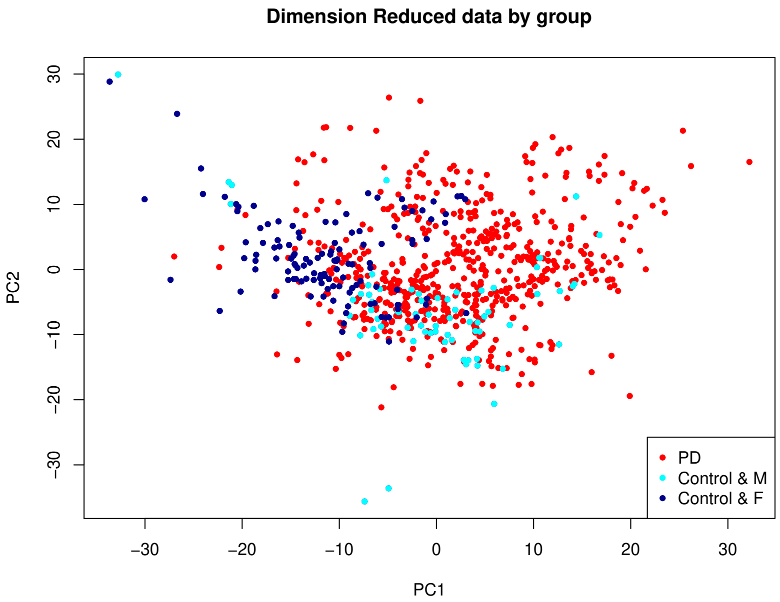
Parkinsons example



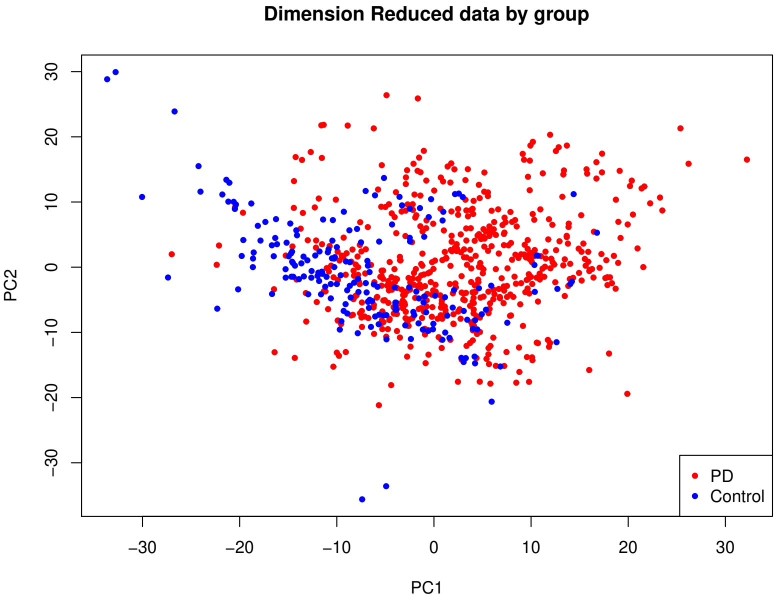
Parkinsons example



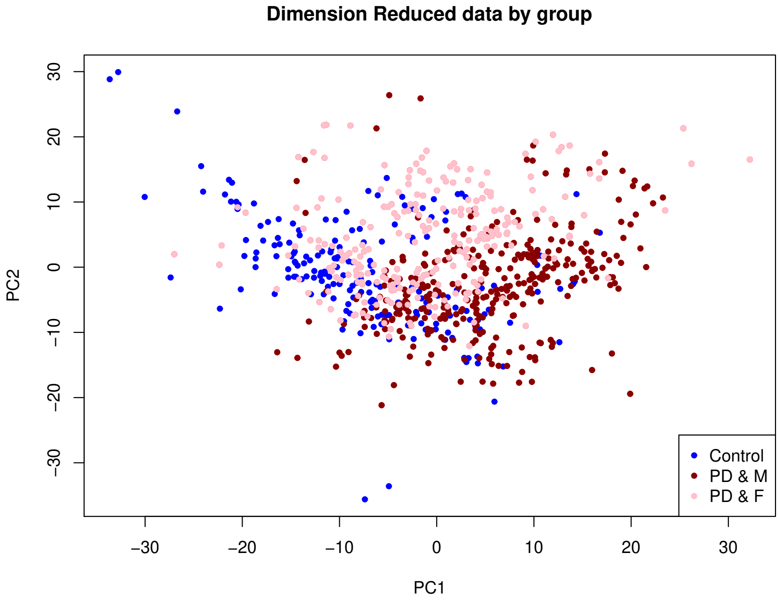
Parkinsons example



Parkinsons example



Parkinsons example



PCA and projections

- We've seen what PCA *can* do.

PCA and projections

- We've seen what PCA *can* do.
- But how is it carried out? What's going on?

PCA and projections

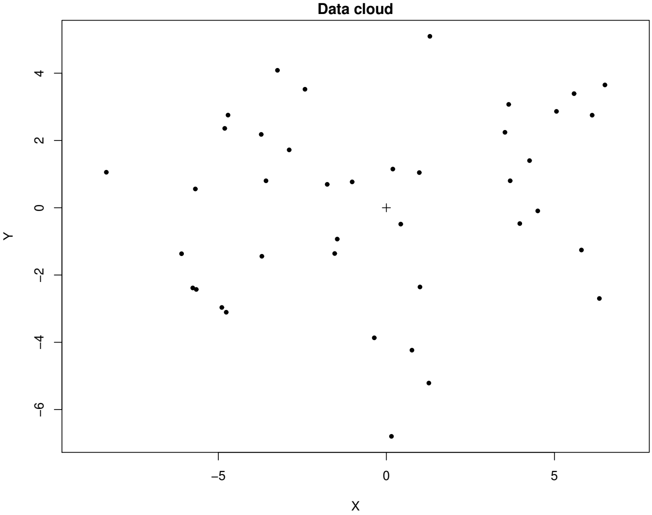
- We've seen what PCA *can* do.
- But how is it carried out? What's going on?
- To get an intuitive sense we'll need to understand:
  - Projecting data into lower dimension

PCA and projections

- We've seen what PCA *can* do.
- But how is it carried out? What's going on?
- To get an intuitive sense we'll need to understand:
  - Projecting data into lower dimension
  - and how to maximize the variation of the result

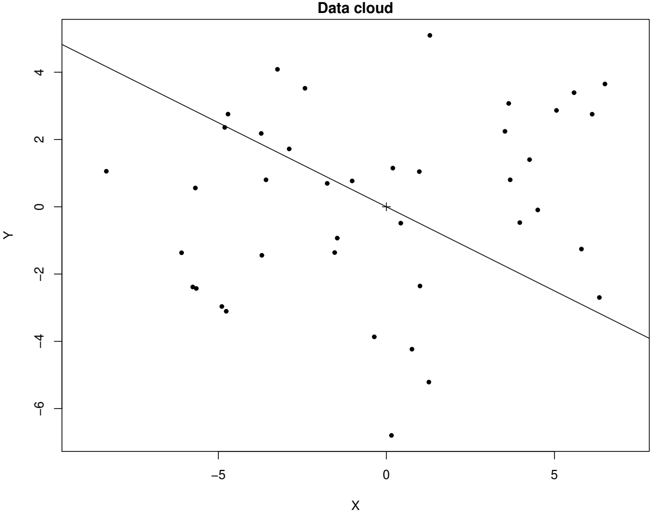
Projections of data

- We have the following (example) data cloud.



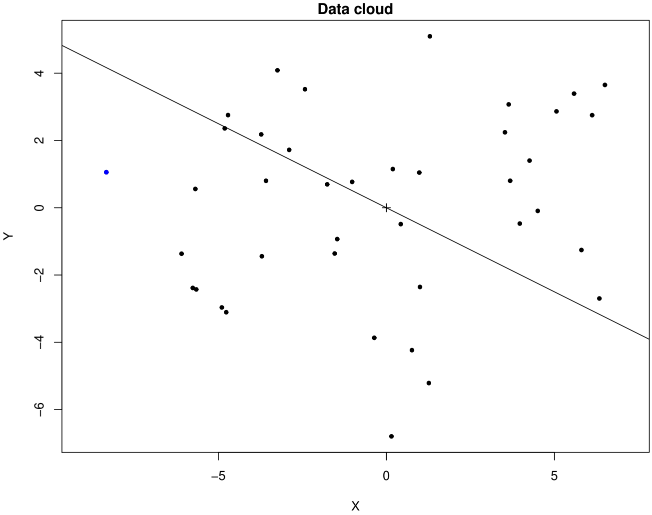
Projections of data

- Suppose we want to *project* our data onto the line.



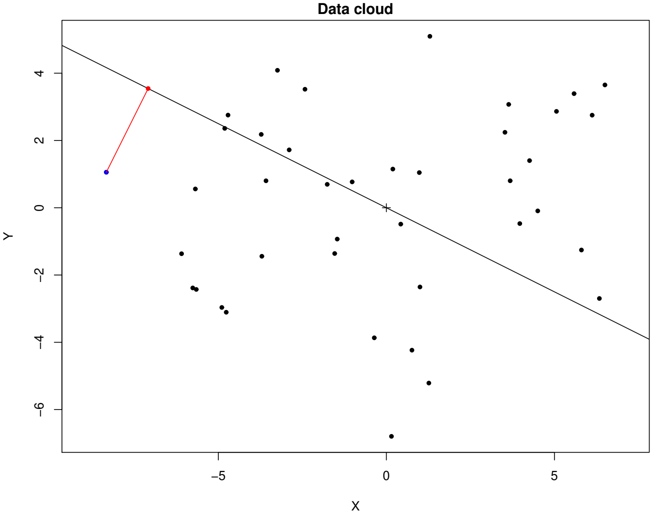
Projections of data

- Let's focus on projecting this point (in blue).



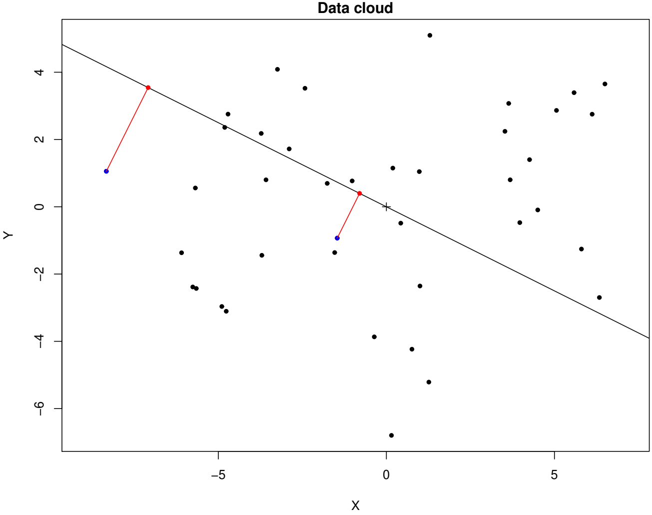
Projections of data

- The red point is nearest to the blue (on the line).



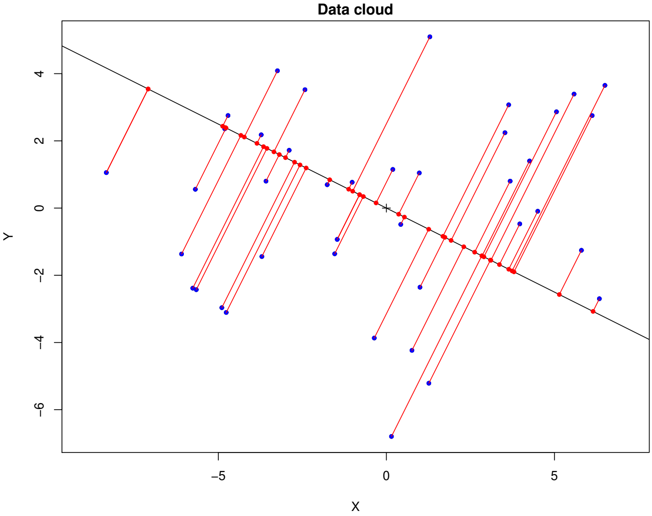
Projections of data

- Similarly for another point



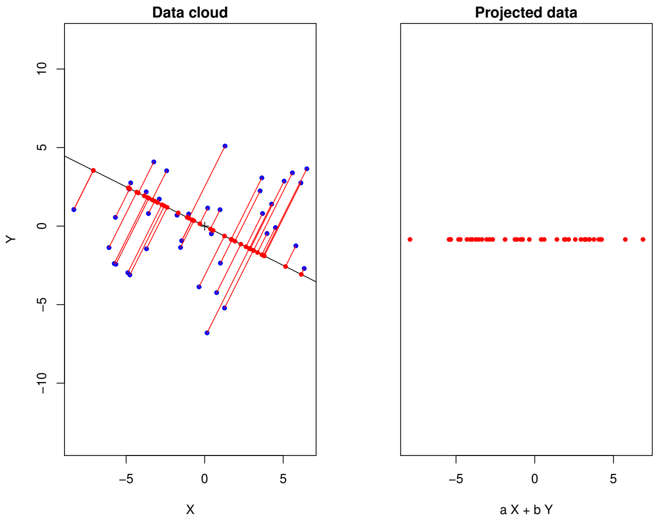
Projections of data

- And finally for all points.



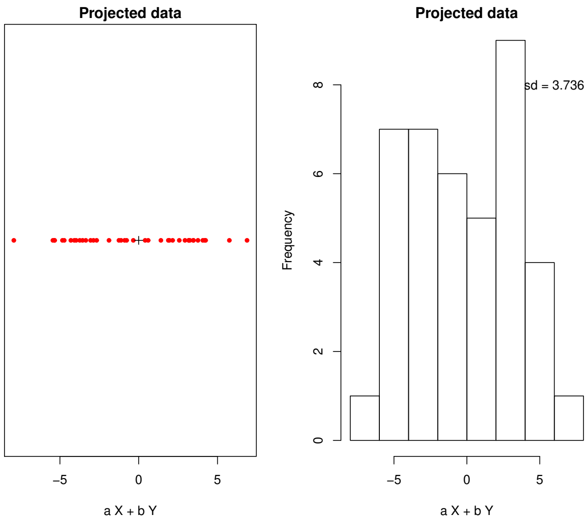
Projections of data

- Data can be treated as univariate while on the line.



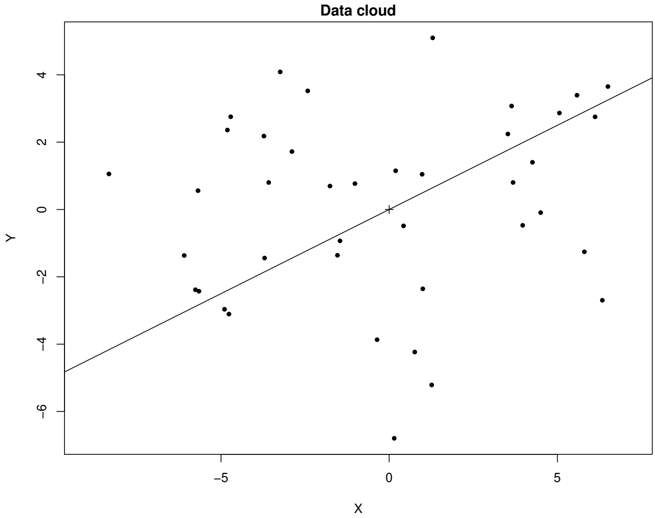
Projections of data

- With histogram



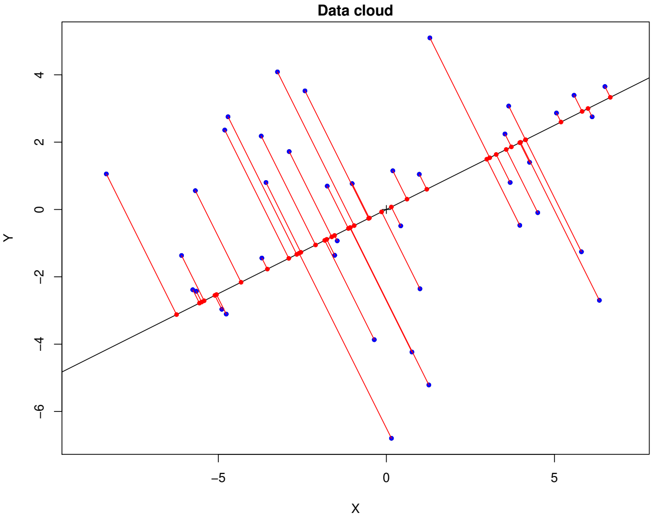
Projections of data

- Let's do this again but with a *new* line



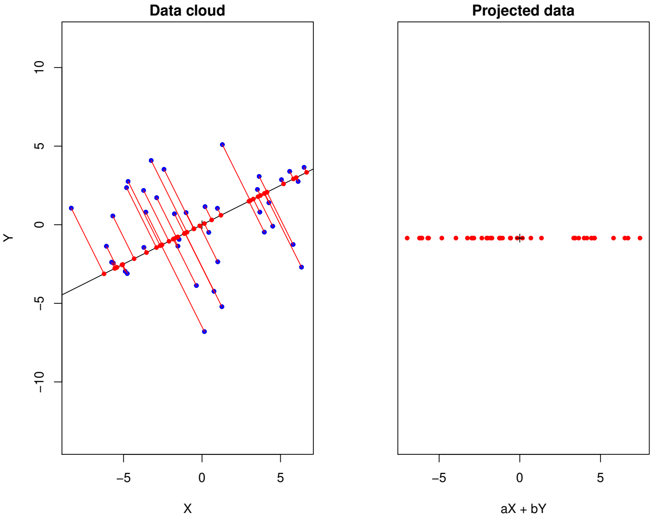
Projections of data

- Results in new *projected* data.



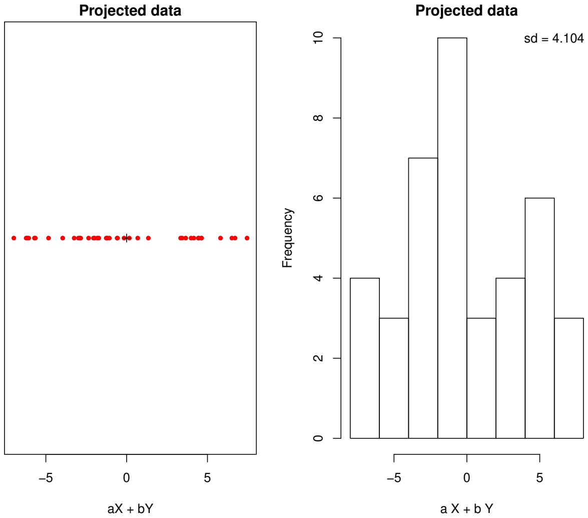
Projections of data

- Which, again, can be considered univariate.



Projections of data

- With a new histogram.



Projections of data

- Notice a change in the *units* of the projected data:

$$aX + bY$$

Projections of data

- Notice a change in the *units* of the projected data:

$$aX + bY$$

- The projected data is a *weighted* average of the original

Projections of data

- Notice a change in the *units* of the projected data:

$$aX + bY$$

- The projected data is a *weighted* average of the original
- Weights determined by:
  - $\frac{b}{a} =$ slope of line
  - $a^2 + b^2 = 1$ (length $= 1$)

Projections of data

- Notice a change in the *units* of the projected data:

$$aX + bY$$

- The projected data is a *weighted* average of the original
- Weights determined by:
  - $\frac{b}{a}$ = slope of line
  - $a^2 + b^2 = 1$ (length = 1)
- The weights can be *positive* or *negative*

Projections of data

- Notice a change in the *units* of the projected data:

$$aX + bY$$

- The projected data is a *weighted* average of the original
- Weights determined by:
  - $\frac{b}{a}$ = slope of line
  - $a^2 + b^2 = 1$ (length $= 1$)
- The weights can be *positive* or *negative*
- Notice $-a$ and $-b$ also work.

Projections of data

- Notice a change in the *units* of the projected data:

$$aX + bY$$

- The projected data is a *weighted* average of the original
- Weights determined by:
  - $\frac{b}{a}$ = slope of line
  - $a^2 + b^2 = 1$ (length = 1)
- The weights can be *positive* or *negative*
- Notice $-a$ and $-b$ also work.
- The std. dev of this projected data depends on the weights / projection line.

Projections of data

- What line we project onto determines the std. dev.

Projections of data

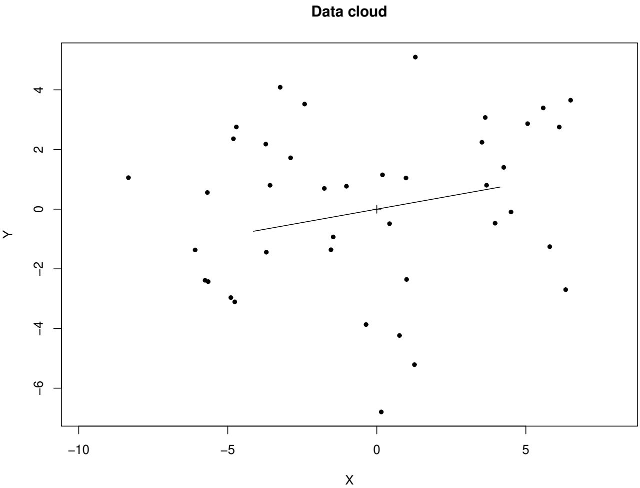
- What line we project onto determines the std. dev.
- Is there a maximum?

Projections of data

- What line we project onto determines the std. dev.
- Is there a maximum?
- What projection determines it?

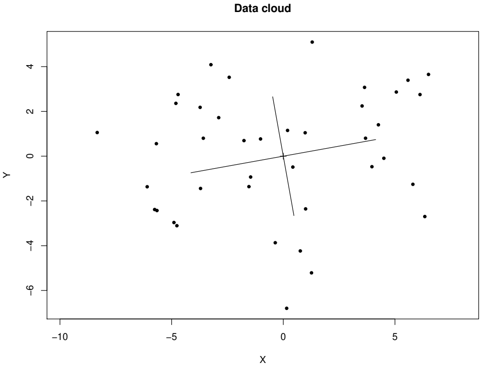
Projections of data

- This projection line *maximizes* the std. dev with sd $= 4.20$ (length $= 2 \times$ sd)



Projections of data

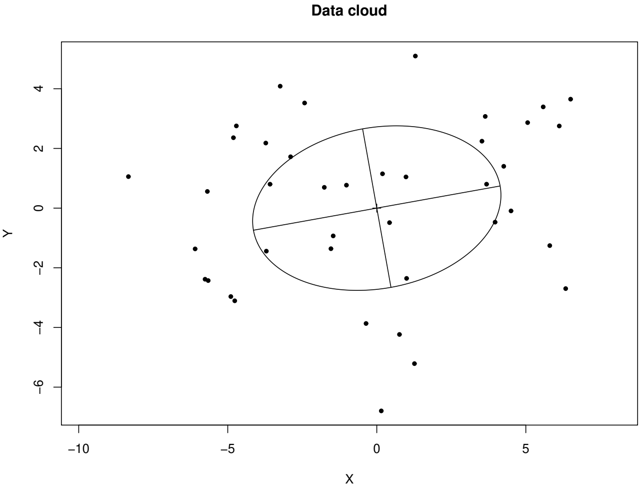
- The additional projection line *minimizes* the std. dev with sd $= 2.69$ (length is $2 \times$ sd)



- Notice that this is orthogonal (perpendicular) to the line for the *maximum*.

Projections of data

- Which motivates the *ellipse*

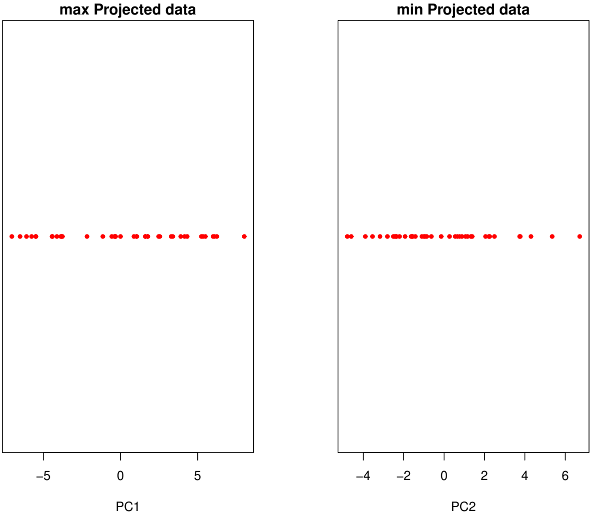


Projections of data

- Now we understand (part of) the sign out front of the building

PCA and Projections

- The maximum and minimum projections give us two sets of univariate data.



PCA and Projections

- These univariate data have popular names:

PCA and Projections

- These univariate data have popular names:
- maximum $\rightarrow$ **first Principal Component** i.e. PC1.

PCA and Projections

- These univariate data have popular names:
- maximum $\rightarrow$ **first Principal Component** i.e. PC1.
- minimum $\rightarrow$ **second Principal Component** i.e. PC2.

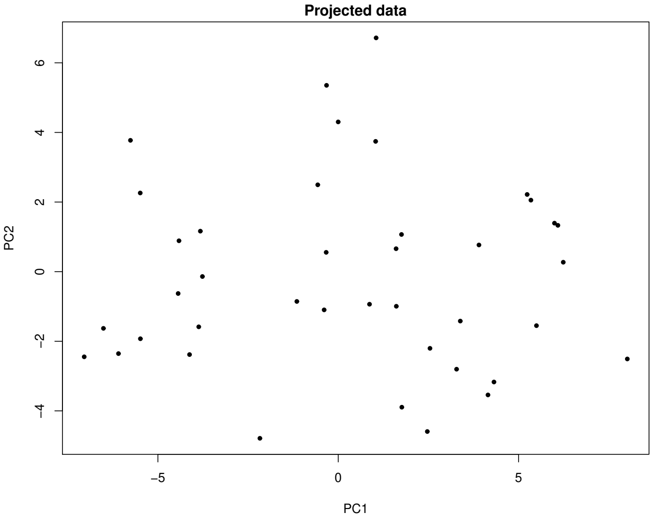
PCA and Projections

- These univariate data have popular names:
- maximum $\rightarrow$ **first Principal Component** i.e. PC1.
- minimum $\rightarrow$ **second Principal Component** i.e. PC2.
- And remember that each is given by

$$\text{PC1} = a^* X + b^* Y,$$
$$\text{PC2} = c^* X + d^* Y.$$

PCA and Projections

- Plotting both gives



PCA and Projections

- What do we gain by using $(\mathrm{PC1}, \mathrm{PC2})$ instead of (X, Y)?

PCA and Projections

- What do we gain by using $(\mathrm{PC1}, \mathrm{PC2})$ instead of (X, Y)?
- Comparing correlations

$$\mathrm{corr}(X, Y) = 0.1570813,$$
$$\mathrm{corr}(\mathrm{PC1}, \mathrm{PC2}) = -4.230633 \times 10^{-18}.$$

PCA and Projections

- What do we gain by using $(\text{PC1}, \text{PC2})$ instead of (X, Y)?
- Comparing correlations

$$\text{corr}(X, Y) = 0.1570813,$$
$$\text{corr}(\text{PC1}, \text{PC2}) = -4.230633 \times 10^{-18}.$$

- Hence, the projected data is *uncorrelated* (a consequence of the orthogonality we saw before).

PCA and Projections

- What do we gain by using $(\mathrm{PC1}, \mathrm{PC2})$ instead of (X, Y)?
- Comparing correlations

$$\mathrm{corr}(X, Y) = 0.1570813,$$
$$\mathrm{corr}(\mathrm{PC1}, \mathrm{PC2}) = -4.230633 \times 10^{-18}.$$

- Hence, the projected data is *uncorrelated* (a consequence of the orthogonality we saw before).
- Obviously, no dimension reduction happens in this case.

PCA and Projections

- We could use the transformed data, PC1, as our new dataset.

PCA and Projections

- We could use the transformed data, PC1, as our new dataset.
- This accomplishes dimension-reduction.

PCA and Projections

- We could use the transformed data, PC1, as our new dataset.
- This accomplishes dimension-reduction.
- Why not use PC2 as our new dataset instead?

PCA and Projections

- We could use the transformed data, PC1, as our new dataset.
- This accomplishes dimension-reduction.
- Why not use PC2 as our new dataset instead?
- One popular method decides PC1 over PC2.

PCA and Projections

- It turns out that

$$\text{var}(\text{PC1}) + \text{var}(\text{PC2}) = \text{var}(X) + \text{var}(Y) = \text{total variation}$$

PCA and Projections

- It turns out that

$$\text{var}(\text{PC1}) + \text{var}(\text{PC2}) = \text{var}(X) + \text{var}(Y) = \text{total variation}$$

- We also know

$$\text{var}(\text{PC1}) \geq \text{var}(\text{PC2})$$

  from before.

PCA and Projections

- Therefore, PC1 accounts for more of the variation present in the original data.

PCA and Projections

- Therefore, PC1 accounts for more of the variation present in the original data.
- As a proportion

$$\frac{\text{var}(\text{PC1})}{\text{var}(\text{PC1}) + \text{var}(\text{PC2})} \geq \frac{\text{var}(\text{PC2})}{\text{var}(\text{PC1}) + \text{var}(\text{PC2})}$$

  where each proportion is the percentage of variability explained of that PC.

PCA and Projections

- Therefore, PC1 accounts for more of the variation present in the original data.
- As a proportion

$$\frac{\text{var}(\text{PC1})}{\text{var}(\text{PC1}) + \text{var}(\text{PC2})} \geq \frac{\text{var}(\text{PC2})}{\text{var}(\text{PC1}) + \text{var}(\text{PC2})}$$

  where each proportion is the percentage of variability explained of that PC.
- This motivates choosing PC1 over PC2.

PCA and Projections

- With *only* dimension reduction in mind, this argument is intuitive (heuristically).

PCA and Projections

- With *only* dimension reduction in mind, this argument is intuitive (heuristically).
- Total variation $\approx$ amount of 'information'.

PCA and Projections

- With *only* dimension reduction in mind, this argument is intuitive (heuristically).
- Total variation $\approx$ amount of 'information'.
- Accounting for more variation $\approx$ less 'information' lost.

PCA and Projections

- With *only* dimension reduction in mind, this argument is intuitive (heuristically).
- Total variation $\approx$ amount of ‘information’.
- Accounting for more variation $\approx$ less ‘information’ lost.
- This argument may break down when we have other goals.

PCA and Projections

- With *only* dimension reduction in mind, this argument is intuitive (heuristically).
- Total variation $\approx$ amount of 'information'.
- Accounting for more variation $\approx$ less 'information' lost.
- This argument may break down when we have other goals.
- 'Information' depends on context, goals, etc.

PCA and Projections

- We now have a firm grip on PCA in two-dimensions.

PCA and Projections

- We now have a firm grip on PCA in two-dimensions.
- Extending to higher-dimensions follows from our two-dimensional case.

PCA and Projections

- We now have a firm grip on PCA in two-dimensions.
- Extending to higher-dimensions follows from our two-dimensional case.
- Projecting multivariate-data, $(X_1, \ldots, X_p)$, onto a line results in

$$a_1 X_1 + \cdots + a_p X_p$$

with weights or **loadings** (positive or negative) $a_1, \ldots, a_p$ with length 1.

PCA and Projections

- We now have a firm grip on PCA in two-dimensions.
- Extending to higher-dimensions follows from our two-dimensional case.
- Projecting multivariate-data, $(X_1, \ldots, X_p)$, onto a line results in
$$a_1 X_1 + \cdots + a_p X_p$$
with weights or **loadings** (positive or negative) $a_1, \ldots, a_p$ with length 1.
- Note that $-a_1, \ldots, -a_p$ also works.

PCA and Projections

- We can find a line to project onto that achieves maximal variation. This univariate data is, again, denoted by

$$\mathrm{PC1} := a_1^* X_1 + \cdots + a_p^* X_p.$$

PCA and Projections

- We can find a line to project onto that achieves maximal variation. This univariate data is, again, denoted by

$$\mathrm{PC1} := a_1^* X_1 + \cdots + a_p^* X_p.$$

- We now have PC1 based on some projection line.

PCA and Projections

- Next we look for a second projection line

PCA and Projections

- Next we look for a second projection line
  - with maximal variation
  - and is *orthogonal* to the first projection line.

PCA and Projections

- Next we look for a second projection line
  - with maximal variation
  - and is *orthogonal* to the first projection line.
- This can be found and results in univariate data which is denoted by

$$\mathrm{PC2} := b_1^* X_1 + \cdots + b_p^* X_p$$

PCA and Projections

- We iterate this process where at each step we look for a projection line

PCA and Projections

- We iterate this process where at each step we look for a projection line
  - with maximal variation
  - and is *orthogonal to all* previous projection lines.

PCA and Projections

- We iterate this process where at each step we look for a projection line
  - with maximal variation
  - and is *orthogonal to all* previous projection lines.
- This finishes after p projection lines are found.

PCA and Projections

- We iterate this process where at each step we look for a projection line
  - with maximal variation
  - and is *orthogonal to all* previous projection lines.
- This finishes after p projection lines are found.
- The process results in p sets of univariate data

$$(\text{PC}1, \ldots, \text{PC}p)$$

PCA and Projections

- All pairs of PC's are *uncorrelated* (from *orthogonality*).

PCA and Projections

- All pairs of PC's are *uncorrelated* (from *orthogonality*).
- Total variation property holds

$$\mathrm{var}(\mathrm{PC1}) + \cdots + \mathrm{var}(\mathrm{PC}p) = \mathrm{var}(X_1) + \cdots + \mathrm{var}(X_p)$$

PCA and Projections

- All pairs of PC's are *uncorrelated* (from *orthogonality*).
- Total variation property holds

$$\mathrm{var}(\mathrm{PC}1) + \cdots + \mathrm{var}(\mathrm{PC}p) = \mathrm{var}(X_1) + \cdots + \mathrm{var}(X_p)$$

- Decreasing variation

$$\mathrm{var}(\mathrm{PC}1) \geq \cdots \geq \mathrm{var}(\mathrm{PC}p).$$

PCA and Projections

- Dimension reduction is achieved by using a (smaller) subset of the projected data in place of the original.

PCA and Projections

- Dimension reduction is achieved by using a (smaller) subset of the projected data in place of the original.
- Many subsets are possible, how do we decide?

PCA and Projections

- Dimension reduction is achieved by using a (smaller) subset of the projected data in place of the original.
- Many subsets are possible, how do we decide?
- The previous method, as in the bivariate case, gives one possibility.

PCA and Projections

- The proportion of variability explained by each PC

PCA and Projections

- The proportion of variability explained by each PC

$$\frac{\text{var(PC1)}}{\text{total variation}} \geq \cdots \geq \frac{\text{var(PC}p\text{)}}{\text{total variation}}$$

PCA and Projections

- The proportion of variability explained by each PC

$$\frac{\text{var}(\text{PC}1)}{\text{total variation}} \geq \cdots \geq \frac{\text{var}(\text{PC}p)}{\text{total variation}}$$

- Suppose we want to best explain variability of the original data

PCA and Projections

- The proportion of variability explained by each PC

$$\frac{\text{var(PC1)}}{\text{total variation}} \geq \cdots \geq \frac{\text{var(PC}p\text{)}}{\text{total variation}}$$

- Suppose we want to best explain variability of the original data
- Then the optimal choice for each size is

$$\begin{aligned}
\text{univariate} &\implies \text{PC1} \\
\text{bi-variate} &\implies (\text{PC1}, \text{PC2}) \\
\text{tri-variate} &\implies (\text{PC1}, \text{PC2}, \text{PC3}) \\
&\vdots
\end{aligned}$$

PCA and Projections

- So to best account for variation we substitute the original data with $(\text{PC}1, \ldots, \text{PC}q)$, in order, for some q.

PCA and Projections

- So to best account for variation we substitute the original data with $(\mathrm{PC}1, \ldots, \mathrm{PC}q)$, in order, for some q.
- But how do we choose q?

PCA and Projections

- So to best account for variation we substitute the original data with $(\mathrm{PC}1, \ldots, \mathrm{PC}q)$, in order, for some q.
- But how do we choose q?
- This depends on what goals we have for the data analysis.

Example in R

- Let's show how to do PCA on a simple example

Example in R

- Let's show how to do PCA on a simple example
- In R, we an use `princomp()`

Example in R

- Let's show how to do PCA on a simple example
- In R, we an use `princomp()`
- We use the following data

	X1	X2	X3	X4	X5
1	1.43	0.44	-0.08	-0.53	-1.25
2	2.59	1.33	-1.20	-0.66	0.44
3	1.55	0.70	1.05	-0.79	1.54
⋮	⋮	⋮	⋮	⋮	⋮

Example in R

- Using the following code

```
X = read.table("example_data.csv")
X_pca = princomp(X)
print( X_pca$loadings, cutoff = 0 )
```

Example in R

- Using the following code

```
X = read.table("example_data.csv")
X_pca = princomp(X)
print( X_pca$loadings, cutoff = 0 )
```

- Gives the following output

Example in R

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
X1	0.89	0.40	0.03	0.22	0.00
X2	0.42	-0.78	-0.38	-0.24	0.09
X3	-0.13	0.39	-0.91	-0.06	-0.01
X4	-0.05	0.05	0.01	0.11	0.99
X5	-0.10	-0.28	-0.17	0.94	-0.10
Proportion of Var	0.43	0.25	0.17	0.09	0.06
Cumulative Var	0.43	0.68	0.85	0.94	1.00

Example in R

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
X1	0.89	0.40	0.03	0.22	0.00
X2	0.42	-0.78	-0.38	-0.24	0.09
X3	-0.13	0.39	-0.91	-0.06	-0.01
X4	-0.05	0.05	0.01	0.11	0.99
X5	-0.10	-0.28	-0.17	0.94	-0.10
Proportion of Var	0.43	0.25	0.17	0.09	0.06
Cumulative Var	0.43	0.68	0.85	0.94	1.00

- then PC1 is given by

$$\mathrm{PC1} := 0.89X_1 + 0.42X_2 - 0.13X_3 - 0.05X_4 - 0.10X_5$$

Example in R

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
X1	0.89	0.40	0.03	0.22	0.00
X2	0.42	-0.78	-0.38	-0.24	0.09
X3	-0.13	0.39	-0.91	-0.06	-0.01
X4	-0.05	0.05	0.01	0.11	0.99
X5	-0.10	-0.28	-0.17	0.94	-0.10
Proportion of Var	0.43	0.25	0.17	0.09	0.06
Cumulative Var	0.43	0.68	0.85	0.94	1.00

- then PC1 is given by

$$PC1 := 0.89X_1 + 0.42X_2 - 0.13X_3 - 0.05X_4 - 0.10X_5$$

- which accounts for 43% of the variation in the data.

Example in R

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
X1	0.89	0.40	0.03	0.22	0.00
X2	0.42	-0.78	-0.38	-0.24	0.09
X3	-0.13	0.39	-0.91	-0.06	-0.01
X4	-0.05	0.05	0.01	0.11	0.99
X5	-0.10	-0.28	-0.17	0.94	-0.10
Proportion of Var	0.43	0.25	0.17	0.09	0.06
Cumulative Var	0.43	0.68	0.85	0.94	1.00

Example in R

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
X1	0.89	0.40	0.03	0.22	0.00
X2	0.42	-0.78	-0.38	-0.24	0.09
X3	-0.13	0.39	-0.91	-0.06	-0.01
X4	-0.05	0.05	0.01	0.11	0.99
X5	-0.10	-0.28	-0.17	0.94	-0.10
Proportion of Var	0.43	0.25	0.17	0.09	0.06
Cumulative Var	0.43	0.68	0.85	0.94	1.00

- and PC2 is given by

$$\text{PC2} := 0.40X_1 - 0.78X_2 + 0.39X_3 + 0.05X_4 - 0.28X_5$$

Example in R

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
X1	0.89	0.40	0.03	0.22	0.00
X2	0.42	-0.78	-0.38	-0.24	0.09
X3	-0.13	0.39	-0.91	-0.06	-0.01
X4	-0.05	0.05	0.01	0.11	0.99
X5	-0.10	-0.28	-0.17	0.94	-0.10
Proportion of Var	0.43	0.25	0.17	0.09	0.06
Cumulative Var	0.43	0.68	0.85	0.94	1.00

- and PC2 is given by

$$PC2 := 0.40X_1 - 0.78X_2 + 0.39X_3 + 0.05X_4 - 0.28X_5$$

- which accounts for 25% of the variation in the data.

Example in R

- Using PC1 and PC2 accounts for 68% of the variation

Example in R

- Using PC1 and PC2 accounts for 68% of the variation
- The new data from PC's, called **scores**, is also calculated and can be accessed by `X_pca$scores`

Example in R

- Using PC1 and PC2 accounts for 68% of the variation
- The new data from PC's, called **scores**, is also calculated and can be accessed by `X_pca$scores`
- This completes PCA (for calculation).

Using PCA

- Now that we understand the how PCA achieves dimension reduction, how can it be used?

Using PCA

- Now that we understand the how PCA achieves dimension reduction, how can it be used?
- Some examples:
    - Data visualization / Exploratory data analysis

Using PCA

- Now that we understand the how PCA achieves dimension reduction, how can it be used?
- Some examples:
  - Data visualization / Exploratory data analysis
  - PC's for data interpretation

Using PCA

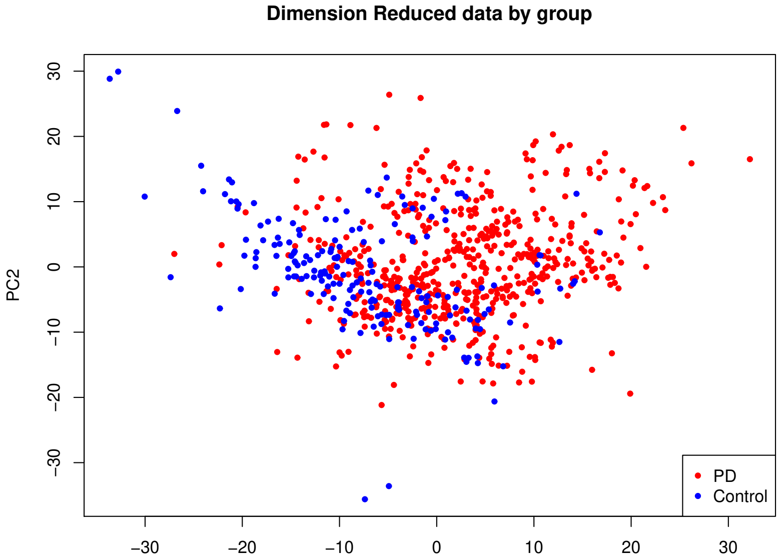
- Now that we understand the how PCA achieves dimension reduction, how can it be used?
- Some examples:
  - Data visualization / Exploratory data analysis
  - PC's for data interpretation
  - Combination with other methods (e.g. Regression)

Using PCA

- Now that we understand the how PCA achieves dimension reduction, how can it be used?
- Some examples:
  - Data visualization / Exploratory data analysis
  - PC's for data interpretation
  - Combination with other methods (e.g. Regression)
  - And many more ...

PCA and Data visualization

- We've already seen this at the beginning



PCA and Data visualization

- PCA was done on all **numeric** data, excluding PD status and gender.

PCA and Data visualization

- PCA was done on all **numeric** data, excluding PD status and gender.
- PC data can then by colored for each individual by group

PCA and Data visualization

- PCA was done on all **numeric** data, excluding PD status and gender.
- PC data can then by colored for each individual by group
- Can see separation, or lack of separation, between groups

PCA and Data visualization

- PCA was done on all **numeric** data, excluding PD status and gender.
- PC data can then by colored for each individual by group
- Can see separation, or lack of separation, between groups
- Easier to spot data anomalies or outliers

PCA and Data visualization

- PCA was done on all **numeric** data, excluding PD status and gender.
- PC data can then by colored for each individual by group
- Can see separation, or lack of separation, between groups
- Easier to spot data anomalies or outliers
- How to 'make sense' of the plot will depend on the subject, etc.

PCA for Interpretation

- [Nash et al.1994] gives a dataset of Abalone measurements.

PCA for Interpretation

- [Nash et al.1994] gives a dataset of Abalone measurements.
- 2835 Abalone specimen composed of Male and Female (1528 and 1307, resp).

PCA for Interpretation

- [Nash et al.1994] gives a dataset of Abalone measurements.
- 2835 Abalone specimen composed of Male and Female (1528 and 1307, resp).
- 7 anatomical measurements

PCA for Interpretation

- [Nash et al.1994] gives a dataset of Abalone measurements.
- 2835 Abalone specimen composed of Male and Female (1528 and 1307, resp).
- 7 anatomical measurements
- Let's do PCA and inspect the results.

PCA for Interpretation

Group	Male		Female	
Component	1	2	1	2
length	-0.17	-0.12	0.16	-0.02
diameter	-0.14	-0.14	0.13	-0.06
height	-0.05	-0.09	0.05	-0.08
whole weight	-0.85	-0.14	0.85	-0.15
shucked weight	-0.39	0.76	0.38	0.76
viscera weight	-0.18	-0.02	0.18	0.00
shell weight	-0.22	-0.60	0.23	-0.62
Cumulative	0.97	0.98	0.96	0.98

PCA for Interpretation

Group	Male		Female	
Component	1	2	1	2
length	-0.17	-0.12	0.16	-0.02
diameter	-0.14	-0.14	0.13	-0.06
height	-0.05	-0.09	0.05	-0.08
whole weight	-0.85	-0.14	0.85	-0.15
shucked weight	-0.39	0.76	0.38	0.76
viscera weight	-0.18	-0.02	0.18	0.00
shell weight	-0.22	-0.60	0.23	-0.62
Cumulative	0.97	0.98	0.96	0.98

- Hmm, looks like PCs 1 and 2 only load a few components.

PCA for Interpretation

Group	Male		Female	
Component	1	2	1	2
length	-0.17	-0.12	0.16	-0.02
diameter	-0.14	-0.14	0.13	-0.06
height	-0.05	-0.09	0.05	-0.08
whole weight	-0.85	-0.14	0.85	-0.15
shucked weight	-0.39	0.76	0.38	0.76
viscera weight	-0.18	-0.02	0.18	0.00
shell weight	-0.22	-0.60	0.23	-0.62
Cumulative	0.97	0.98	0.96	0.98

- Hmm, looks like PCs 1 and 2 only load a few components.
- Is this a real effect? Or is something else happening?

PCA for Interpretation

- The PCA used data *as-is*, without standardization

PCA for Interpretation

- The PCA used data *as-is*, without standardization
- The sample standard deviations are given in the following table

	length	diameter	height	whole weight	shucked weight	viscera weight	shell weight
M	2.95	2.42	1.00	13.52	6.41	3.01	3.76
F	2.15	1.77	1.00	10.76	4.97	2.44	3.14

- Values have been ‘normalized’ by dividing by the minimum standard deviation.

PCA for Interpretation

- The PCA used data *as-is*, without standardization
- The sample standard deviations are given in the following table

	length	diameter	height	whole weight	shucked weight	viscera weight	shell weight
M	2.95	2.42	1.00	13.52	6.41	3.01	3.76
F	2.15	1.77	1.00	10.76	4.97	2.44	3.14

- Values have been 'normalized' by dividing by the minimum standard deviation.
- Whole weight **dominates** the others in terms of variation.

PCA for Interpretation

- The PCA used data *as-is*, without standardization
- The sample standard deviations are given in the following table

	length	diameter	height	whole weight	shucked weight	viscera weight	shell weight
M	2.95	2.42	1.00	13.52	6.41	3.01	3.76
F	2.15	1.77	1.00	10.76	4.97	2.44	3.14

- Values have been 'normalized' by dividing by the minimum standard deviation.
- Whole weight **dominates** the others in terms of variation.
- Hence, PC1 considers it as the most important.

PCA for Interpretation

- To demonstrate the point, suppose we have bi-variate data where

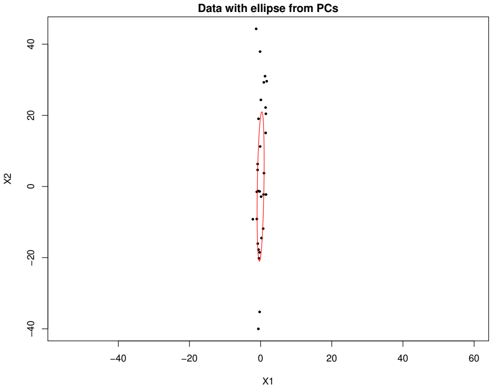
$$\mathrm{sd}(X_1) = 0.991,$$
$$\mathrm{sd}(X_2) = 21.002.$$

PCA for Interpretation

- To demonstrate the point, suppose we have bi-variate data where

$$\mathrm{sd}(X_1) = 0.991,$$
$$\mathrm{sd}(X_2) = 21.002.$$

- Then the ellipse based on max / min directions is



PCA for Interpretation

- PC's will simply select those variables with large std. dev.

PCA for Interpretation

- PC's will simply select those variables with large std. dev.
- Examples like this motivate first standardizing your data, then doing PCA.

PCA for Interpretation

- PC's will simply select those variables with large std. dev.
- Examples like this motivate first standardizing your data, then doing PCA.
- Lets re-do PCA on this example but with standardized data.

PCA for Interpretation

	Male		Female	
Component	1	2	1	2
length	-0.38	-0.00	-0.40	0.14
diameter	-0.38	0.07	-0.40	0.12
height	-0.36	0.77	-0.27	-0.96
whole weight	-0.39	-0.22	-0.41	0.10
shucked weight	-0.38	-0.47	-0.39	0.16
viscera weight	-0.38	-0.29	-0.39	0.12
shell weight	-0.37	0.20	-0.38	0.02
Cumulative	0.90	0.94	0.84	0.93

Table: PCA on standardized Abalone data

PCA for Interpretation

	Male		Female	
Component	1	2	1	2
length	-0.38	-0.00	-0.40	0.14
diameter	-0.38	0.07	-0.40	0.12
height	-0.36	0.77	-0.27	-0.96
whole weight	-0.39	-0.22	-0.41	0.10
shucked weight	-0.38	-0.47	-0.39	0.16
viscera weight	-0.38	-0.29	-0.39	0.12
shell weight	-0.37	0.20	-0.38	0.02
Cumulative	0.90	0.94	0.84	0.93

Table: PCA on standardized Abalone data

- The first component can be interpreted as an overall-size component.

PCA for Interpretation

	Male		Female	
Component	1	2	1	2
length	-0.38	-0.00	-0.40	0.14
diameter	-0.38	0.07	-0.40	0.12
height	-0.36	0.77	-0.27	-0.96
whole weight	-0.39	-0.22	-0.41	0.10
shucked weight	-0.38	-0.47	-0.39	0.16
viscera weight	-0.38	-0.29	-0.39	0.12
shell weight	-0.37	0.20	-0.38	0.02
Cumulative	0.90	0.94	0.84	0.93

Table: PCA on standardized Abalone data

- The first component can be interpreted as an overall-size component.
- The second component contrasts between different measurements.

PCA with regression

- Many different flavors of PCA & regression out there [Jolliffe2011]

PCA with regression

- Many different flavors of PCA & regression out there [Jolliffe2011]
- A popular version is to first do PCA on predictors and then carry out regression.

PCA with regression

- Many different flavors of PCA & regression out there [Jolliffe2011]
- A popular version is to first do PCA on predictors and then carry out regression.
- By using a smaller number of predictors, via PCA, we may obtain better a regression model (avoid over-fitting).

PCA with regression: an Example

- [Hamidieh2018] provides a dataset of physical measurements used to predict superconductor critical temperature.

PCA with regression: an Example

- [Hamidieh2018] provides a dataset of physical measurements used to predict superconductor critical temperature.
- 81 predictors: e.g. mean atomic mass, mean Density, etc.

PCA with regression: an Example

- [Hamidieh2018] provides a dataset of physical measurements used to predict superconductor critical temperature.
- 81 predictors: e.g. mean atomic mass, mean Density, etc.
- Performed PCA to use only 10 predictors.

PCA with regression: an Example

- [Hamidieh2018] provides a dataset of physical measurements used to predict superconductor critical temperature.
- 81 predictors: e.g. mean atomic mass, mean Density, etc.
- Performed PCA to use only 10 predictors.
- Let's look at the the results of regression.

PCA with regression: an Example

- If we use the first 10 PC's, we obtain the following results

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	34.42	0.15	223.83	0.00
PC1	4.00	0.03	146.22	0.00
PC2	-1.99	0.05	-37.67	0.00
PC3	2.40	0.06	43.38	0.00
PC4	1.25	0.06	20.58	0.00
PC5	-2.57	0.07	-36.42	0.00
PC6	-3.43	0.09	-39.04	0.00
PC7	-0.76	0.09	-8.51	0.00
PC8	0.58	0.10	5.93	0.00
PC9	1.91	0.11	17.18	0.00
PC10	-0.09	0.12	-0.75	0.45

PCA with regression: an Example

- If we use the first 10 PC's, we obtain the following results

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	34.42	0.15	223.83	0.00
PC1	4.00	0.03	146.22	0.00
PC2	-1.99	0.05	-37.67	0.00
PC3	2.40	0.06	43.38	0.00
PC4	1.25	0.06	20.58	0.00
PC5	-2.57	0.07	-36.42	0.00
PC6	-3.43	0.09	-39.04	0.00
PC7	-0.76	0.09	-8.51	0.00
PC8	0.58	0.10	5.93	0.00
PC9	1.91	0.11	17.18	0.00
PC10	-0.09	0.12	-0.75	0.45

- and $R^2_{adj} = 0.5714$.

PCA with regression: an Example

- If we instead use the 10 PC's which have the *most statistically significant coefficients*, we obtain the following results

PCA with regression: an Example

- If we instead use the 10 PC's which have the *most statistically significant coefficients*, we obtain the following results

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	34.42	0.14	245.57	0.00
PC1	4.00	0.02	160.43	0.00
PC2	-1.99	0.05	-41.33	0.00
PC3	2.40	0.05	47.59	0.00
PC5	-2.57	0.06	-39.96	0.00
PC6	-3.43	0.08	-42.83	0.00
PC23	13.21	0.31	42.95	0.00
PC25	11.53	0.35	32.88	0.00
PC20	8.11	0.25	32.15	0.00
PC11	3.08	0.12	26.72	0.00
PC15	4.01	0.16	25.52	0.00

PCA with regression: an Example

- If we instead use the 10 PC's which have the *most statistically significant coefficients*, we obtain the following results

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	34.42	0.14	245.57	0.00
PC1	4.00	0.02	160.43	0.00
PC2	-1.99	0.05	-41.33	0.00
PC3	2.40	0.05	47.59	0.00
PC5	-2.57	0.06	-39.96	0.00
PC6	-3.43	0.08	-42.83	0.00
PC23	13.21	0.31	42.95	0.00
PC25	11.53	0.35	32.88	0.00
PC20	8.11	0.25	32.15	0.00
PC11	3.08	0.12	26.72	0.00
PC15	4.01	0.16	25.52	0.00

- and $R^2_{adj} = 0.644$.

PCA with regression

- Our model fit improved using PC's outside of the first 10.

PCA with regression

- Our model fit improved using PC's outside of the first 10.
- PCA says nothing about association with some other variable!

PCA with regression

- Our model fit improved using PC's outside of the first 10.
- PCA says nothing about association with some other variable!
- It is possible for later PC's to be better predictors!

PCA with regression

- Our model fit improved using PC's outside of the first 10.
- PCA says nothing about association with some other variable!
- It is possible for later PC's to be better predictors!
- **Lesson**: Always remember PCA only explains variation within a dataset.

PCA with regression

- Our model fit improved using PC's outside of the first 10.
- PCA says nothing about association with some other variable!
- It is possible for later PC's to be better predictors!
- **Lesson**: Always remember PCA only explains variation within a dataset.
- Side note: since the PC dataset are uncorrelated (by construction), dropping variables by t-tests is reasonable.

What is Factor Analysis?

- Before going into details of Factor Analysis (FA), we'll give some motivation.

What is Factor Analysis?

- Before going into details of Factor Analysis (FA), we'll give some motivation.
- Researchers are often interested in **unobservable** quantities or effects (or **latent variables**). Some examples:

What is Factor Analysis?

- Before going into details of Factor Analysis (FA), we'll give some motivation.
- Researchers are often interested in **unobservable** quantities or effects (or **latent variables**). Some examples:
  - Intelligence (verbal, mathematical, etc.)

What is Factor Analysis?

- Before going into details of Factor Analysis (FA), we'll give some motivation.
- Researchers are often interested in **unobservable** quantities or effects (or **latent variables**). Some examples:
  - Intelligence (verbal, mathematical, etc.)
  - Personality (extroversion, agreeableness, etc.)

What is Factor Analysis?

- Before going into details of Factor Analysis (FA), we'll give some motivation.
- Researchers are often interested in **unobservable** quantities or effects (or **latent variables**). Some examples:
  - Intelligence (verbal, mathematical, etc.)
  - Personality (extroversion, agreeableness, etc.)
  - Medical diagnoses (depression, anxiety, etc.)

What is Factor Analysis?

- Before going into details of Factor Analysis (FA), we'll give some motivation.
- Researchers are often interested in **unobservable** quantities or effects (or **latent variables**). Some examples:
  - Intelligence (verbal, mathematical, etc.)
  - Personality (extroversion, agreeableness, etc.)
  - Medical diagnoses (depression, anxiety, etc.)
- Researchers collect data with **measurable** quantities that, hopefully, associate with hypothesized **unobservable** quantities.

Factor Analysis model

- Suppose our dataset consists of p variables $(X_1, \ldots, X_p)$ of interest.

Factor Analysis model

- Suppose our dataset consists of p variables $(X_1, \ldots, X_p)$ of interest.
- The Factor model supposes k unobservable and **random** variables, denoted by $(f_1, \ldots, f_k)$, explain each variable.

Factor Analysis model

- Suppose our dataset consists of p variables $(X_1, \dots, X_p)$ of interest.
- The Factor model supposes k unobservable and **random** variables, denoted by $(f_1, \dots, f_k)$, explain each variable.
- Specifically, it supposes each variable can be expressed by

$$X_i = a_1^{(i)} f_1 + \cdots + a_k^{(i)} f_k + e_i.$$

Factor Analysis model

$$X_i = a_1^{(i)} f_1 + \cdots + a_k^{(i)} f_k + e_i.$$

- Each variable is a *weighted* average of the **same** underlying factors.

Factor Analysis model

$$X_i = a_1^{(i)} f_1 + \cdots + a_k^{(i)} f_k + e_i.$$

- Each variable is a *weighted* average of the **same** underlying factors.
- Each variable has it's own weights (or **loadings**)

Factor Analysis model

$$X_i = a_1^{(i)} f_1 + \cdots + a_k^{(i)} f_k + e_i.$$

- Each variable is a *weighted* average of the **same** underlying factors.
- Each variable has it's own weights (or **loadings**)
- Each variable has residual e_i.

Factor Analysis model

- That is the basic Factor Model (omitting other technical assumptions)

Factor Analysis model

- That is the basic Factor Model (omitting other technical assumptions)
- This method has historically been confused with PCA.

Factor Analysis model

- That is the basic Factor Model (omitting other technical assumptions)
- This method has historically been confused with PCA.
- One possible reason is they can give similar **loadings**

FA and PCA loadings

- It is possible for FA and PCA to have similar **loadings**

FA and PCA loadings

- It is possible for FA and PCA to have similar **loadings**
- But they are loading on very different quantities. For PCA

$$\text{PC1} := a^{(1)}X_1 + \cdots + a^{(p)}X_p.$$

FA and PCA loadings

- It is possible for FA and PCA to have similar **loadings**
- But they are loading on very different quantities. For PCA

$$\mathrm{PC1} := a^{(1)}X_1 + \cdots + a^{(p)}X_p.$$

- Whereas for FA

$$\begin{aligned} X_1 &= a^{(1)}f_1 + e_1, \\ &\vdots \\ X_p &= a^{(p)}f_1 + e_p. \end{aligned}$$

FA and PCA loadings

- It is possible for FA and PCA to have similar **loadings**
- But they are loading on very different quantities. For PCA

$$\text{PC1} := a^{(1)}X_1 + \cdots + a^{(p)}X_p.$$

- Whereas for FA

$$\begin{aligned} X_1 &= a^{(1)}f_1 + e_1, \\ &\vdots \\ X_p &= a^{(p)}f_1 + e_p. \end{aligned}$$

- The loadings/weights may be similar but their interpretation is not.

How does FA compare to PCA?

- This table summarizes some of the big distinctions between the two

	Method	
Criterion	FA	PCA
Modeling assumption	Very much so	(almost) none
Changing dimension	Major changes	No changes
Estimation	Complex	Simple
Goal	Latent factors	simplify by transformation
Emphasis	Correlation	Variance
Loadings	for Latent factors	for Original variables

Code and Slide materials

- Slides and code used at github
  https://github.com/AndoBlando/PCA_Fa_Talk

Bibliography I

Hamidieh, K. (2018).
A data-driven statistical model for predicting the critical temperature of a superconductor.
Computational Materials Science, 154:346 – 354.

Jolliffe, I. (2011).
Principal component analysis.
Springer.

Nash, W. J., Sellers, T. L., Talbot, S. R., Cawthorn, A. J., and Ford, W. B. (1994).
The population biology of abalone (haliotis species) in tasmania. i. blacklip abalone (h. rubra) from the north coast and islands of bass strait.
Sea Fisheries Division, Technical Report, 48.

Bibliography II

Sakar, C. O., Serbes, G., Gunduz, A., Tunc, H. C., Nizam, H., Sakar, B. E., Tutuncu, M., Aydin, T., Isenkul, M. E., and Apaydin, H. (2019).
A comparative analysis of speech signal processing algorithms for parkinson's disease classification and the use of the tunable q-factor wavelet transform.
Applied Soft Computing, 74:255 – 263.