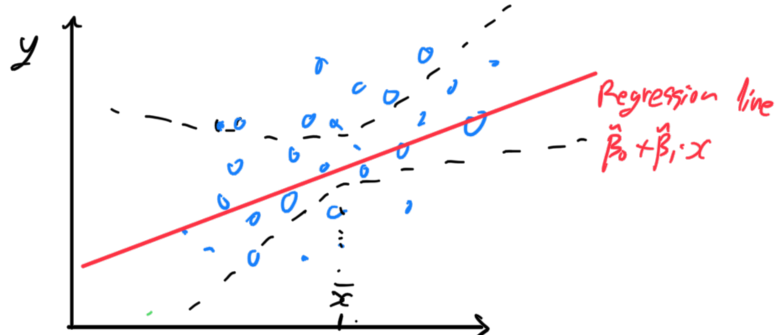
Lecture 9

$\boxed{\text{Lecture 9}}$ 17/5/2021

Recap: Simple Regression $y_i = \beta_0 + \beta_1 \cdot x_i + \varepsilon_i$
$i = 1, \dots, n$

- Confidence Bands for $\beta_0 + \beta_1 \cdot x$



$$\underline{\beta_0 + \beta_1 \cdot x} \underset{\text{w.p. } 1-\alpha}{\in} \left(\hat{\beta}_0 + \hat{\beta}_1 x \pm \underline{t_{n-2}^{1-\alpha/2}} \, s \sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}} \right)$$

In general:

$$z_0^T \beta \underset{\text{w.p. } 1-\alpha}{\in} \left(z_0^T \hat{\beta} \pm t_{n-p}^{1-\alpha/2} \cdot s \sqrt{z_0^T (Z^T Z)^{-1} z_0} \right)$$

$z_0 \in \mathbb{R}^p$

$$s^2 = \frac{1}{n-p} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \frac{\|\varepsilon\|^2}{n-p}$$

- <u>Prediction Bands</u> (in HW)

We want to predict y_{n+1} given x_{n+1} and $(y_i, x_i)_{i=1}^n$.

$$y_{n+1} \underset{\text{w.p. } 1-\alpha}{\in} \left(\hat{\beta}_0 + \hat{\beta}_1 x_{n+1} \pm t_{n-2}^{1-\alpha/2} \cdot s \sqrt{1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{S_{xx}}} \right)$$

We get a band if we evaluate this for every $x_{n+1} \in \mathbb{R}$

- We can take avg. of m measurements

at the same x_{n+1}. In this case:

$$\bar{y}_{n+1} \in \left(\hat{\beta}_0 + \hat{\beta}_1 x_{n+1} \pm t_{n-2}^{1-\alpha/2} \cdot s \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(x_{n+1}-\bar{x})^2}{S_{xx}}}\right)$$

w.p. $1-\alpha$

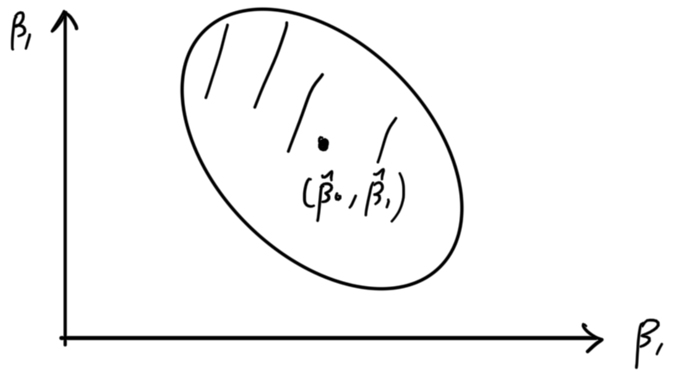
Simultaneous Bands

- Contain $(\beta_0 + \beta_1 \cdot x)_{x \in \mathbb{R}}$ with prob. $1-\alpha$

- In p dim, contain $(z_0^T \beta)_{z_0 \in \mathbb{R}^p}$
- From the distribution of $\hat{\beta}$:

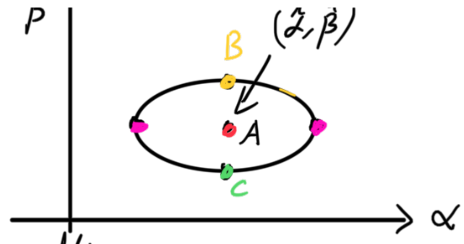
$$Pr\left((\hat{\beta}-\beta)^T (Z^T Z)^{-1} (\hat{\beta}-\beta) \leq s^2 \cdot p \cdot F_{p,n-p}^{1-\alpha} \right) = 1-\alpha$$

This define an ellipsoid in $\mathbb{R}^p$

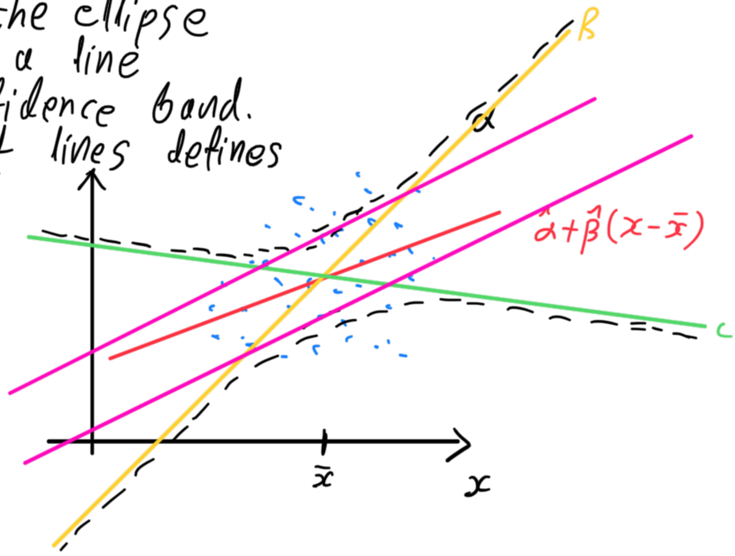


- For $p=2$, β_0 and β_1 lay in an ellipse with prob. $1-\alpha$

- If we use $\overset{\beta_0}{\alpha} + \beta(x-\bar{x})$ for the regression line, then $Z^T Z$ becomes diagonal and the ellipse's axes are aligned with the x & y axes:



each point in the ellipse corresponds to a line within the confidence band. The collection of lines defines the band



Working Hotelling Bands:

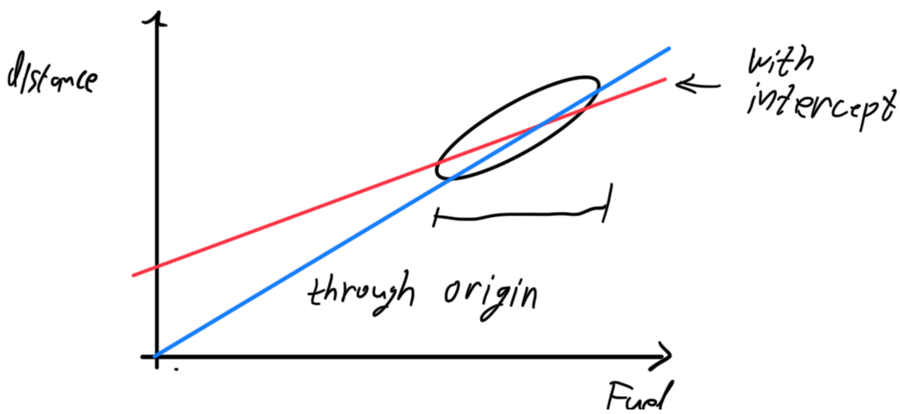
Confidence: $\hat\beta_0+\hat\beta_1 x \pm \sqrt{2F^{1-\alpha}_{2,n-2}}\cdot s\sqrt{\frac{1}{n}+\frac{(x-\bar x)^2}{S_{xx}}}$

Prediction: $\hat\beta_0+\hat\beta_1 \cdot x \pm \sqrt{2F^{1-\alpha}_{2,n-2}}\cdot s\sqrt{1+\frac{1}{n}+\frac{(x-\bar x)^2}{S_{xx}}}$

Avg. of m predictions: $\hat\beta_0+\hat\beta_1 x \pm \sqrt{2F^{1-\alpha}_{2,n-2}}\cdot s\sqrt{\frac{1}{m}+\frac{1}{n}+\frac{(x-\bar x)^2}{S_{xx}}}$

(these are wider than earlier-mentioned bands)

- Regression Through the Origin



- This is dangerous if there is no data near the origin
- Also, no interpertation of R^2 in this case.

Muliple Regression:

$x_i \in \mathbb{R}^d$

Suppose: $$Z = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1d} \\ 1 & \vdots & & & \\ \vdots & & & & \vdots \\ 1 & x_{n1} & \cdots & & x_{nd} \end{bmatrix} \in \mathbb{R}^p$$

$p = d+1$

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

(fraction of explained variance)

- Note: when using R^2 you must have an intercept term
- R^2 is not additive $0 \leq R \leq 1$
- No rule of thumb for how large R^2 must be
- Increasing the number of predictors never decreases R^2

Some Considerations:

- A "true" β_j:
  - Depends on what variables

are included.

  - A "true" β_j exists for each and every subset of the covariates used.

- Naive Face-Value Interpertation:
  - With $y_i = \sum_{j=1}^{p} x_{ij}\beta_j + \varepsilon_i$. We may about β_j as $\frac{\partial y}{\partial x_j}$ or $\frac{\partial \mathbb{E}[y]}{\partial x_j}$.
  - This is unreliable. For example if $x_2 = x_1^2$, then changing one predictor alone is not possible.

- Wrong Signs:
  - may be due to correlation within the data.

Correlation, Association, Causality

- Regression with observational data captures correlation/association, not causality
- It could be $x \to y$, $y \to x$, or $z \to (x, y)$

Interplay Between Variables

- Competing Variables:
  - $\hat{\beta}_1$ is significant if x_2 is not in the model
  - $\hat{\beta}_2$ is significant if x_1 is not in the model.

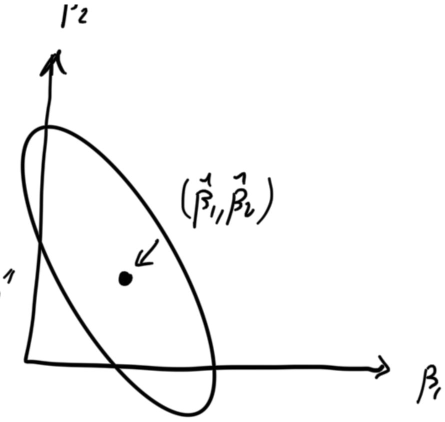
- In the house prices example:

  OverallQual & OverallCond have this property.

- Graphically:

You can reject "$\beta_1 = \beta_2 = 0$"

but not "$\beta_1 = 0$ or $\beta_2 = 0$"



- Occures when there is positive correlation between x_1 & x_2

collaborating variables:

- $\hat{\beta}_1$ is sig. if x_2 is in the model
- $\hat{\beta}_2$ is sig if x_1 is in the model

- This is much more rare than competition

- Example:

x_1 = air pressure in loc. A
x_2 = air pressure in loc. B
y = wind intensity $A \to B$
(proportional to $X_1 - X_2$

Simpson's Paradox

Two hospitals A, B

Mild cases

Hospital	S	D
A	50	0
B	80	10

Critical cases

	S	D
A	25	25
B	0	10

Total:

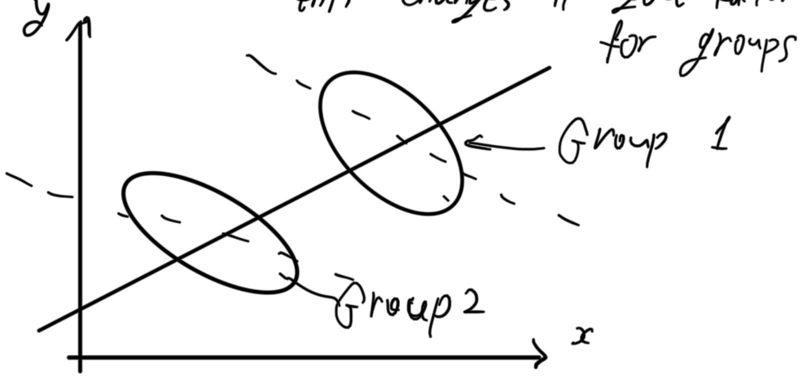
	S	D
A	75	25

B | 80 | 20

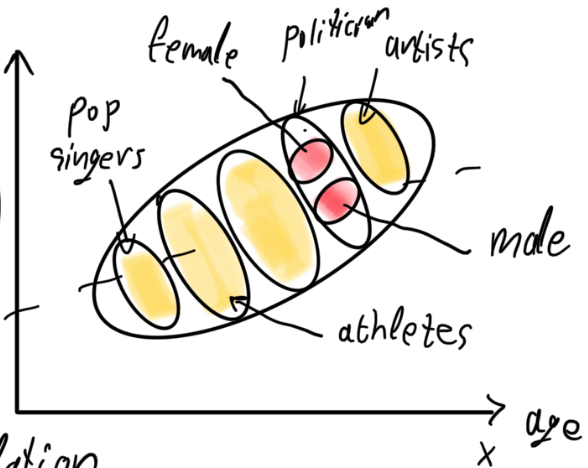
- B looks better overall, although it losses in each category individually.

- With $y = \beta_0 + \beta_1 \cdot Age$, $\beta_1 \geq 0$, but this changes if you factor for groups

media references



Factoring by diff. groups lead to diff. sign of reg. coefficient



Partial Correlation

- We want to look for the connection between height and spelling ability of kids from ages 7-12.
  We get positive correlation, but kids height & spelling ability change as they grow.

- We regress height on age, and get residuals
  These residuals are "age-adjusted height"

- We do the same for spelling_level, getting "age-adjusted spelling"

  The correlation between the two sets of residuals is the partial correlation of height and spelling_level, adjusted for age

Def. <u>Partial Correlation $\rho_{ij|k}$</u>

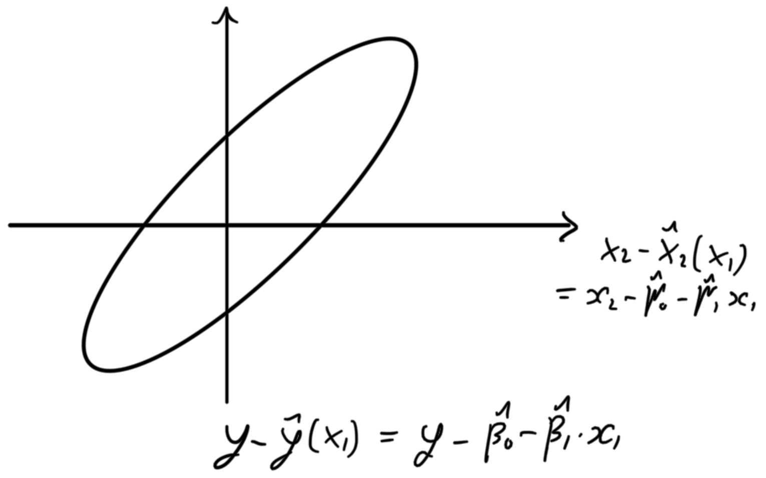
Partial correlation of x_i x_j, adjusting for x_k is the correlation of resid. for x_i on x_k and resid. for x_j on x_k

For Gaussian data:

$$\rho_{ij|k} = \frac{\rho_{ij} - \rho_{ik}\rho_{jk}}{\sqrt{(1-\rho_{ik}^2)(1-\rho_{jk}^2)}}$$

(with or without hats)

$$\hat{\rho}_{ij} = \frac{x_i^T x_j}{\|x_i\| \|x_j\|} \qquad \rho_{ij} = \frac{Cov(X_i, X_j)}{\sqrt{Var(X_i) Var(X_j)}}$$



- We can find partial corr. of (x_1, y), adjusting for $x_2, \ldots, x_k$: we regress those predictors out of both x_1 and y and find the partial correlation.