The Normal Equations (for linear regression)

- Consider the simplest linear relationship

$$y = \beta_0 + \beta_1 x,$$

- Now form the Sum of the Squared Residuals:

$$SSR = \sum_{i=1}^{m} (y_i - (\beta_0 + \beta_1 x_i))^2.$$

- Minimization of SSR is now achieved by differentiating w.r.t. the 2 (unknown) model parameters and setting =0 to find the extrema…

The Normal Equations (for linear regression)

$$\frac{dSSR}{d\beta_0} = \sum_{i=1}^{m} -2(y_i - (\beta_0 + \beta_1 x_i)) = 0,$$

$$\frac{dSSR}{d\beta_1} = \sum_{i=1}^{m} -2x_i(y_i - (\beta_0 + \beta_1 x_i)) = 0.$$

After simplification of the above two equations, we get

$$m\beta_0 + \beta_1 \sum_{i=1}^{m} x_i = \sum_{i=1}^{m} y_i,$$

$$\beta_0 \sum_{i=1}^{m} x_i + \beta_1 \sum_{i=1}^{m} x_i^2 = \sum_{i=1}^{m} x_i y_i.$$

Equations (2.34) and (2.35) are referred to as the **normal equations**.

The Normal Equations (for linear regression)

- Those two equations can be expressed as a single matrix equation:

$$\begin{bmatrix} m & \sum_{i=1}^{m} x_i \\ \sum_{i=1}^{m} x_i & \sum_{i=1}^{m} x_i^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{m} y_i \\ \sum_{i=1}^{m} x_i y_i \end{bmatrix}$$

- Yielding the solution:

$$\beta_1 = \frac{m \sum_{i=1}^{m} x_i y_i - \sum_{i=1}^{m} x_i \sum_{i=1}^{m} y_i}{m \sum_{i=1}^{m} x_i^2 - \left(\sum_{i=1}^{m} x_i\right)^2}$$

and

$$\beta_0 = \bar{y} - \beta_1 \bar{x},$$

The Normal Equations (for linear regression)

where the means of x and y are defined as

$$\bar{x} = \frac{1}{m}\sum_{i=1}^{m} x_i \quad \text{and} \quad \bar{y} = \frac{1}{m}\sum_{i=1}^{m} y_i.$$

In fact, β_1 can be expressed in somewhat simpler form

$$\beta_1 = \frac{\sum_{i=1}^{m}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{m}(x_i - \bar{x})^2}.$$

The Normal Equations (for linear regression)

- Now how about fitting a general second-order polynomial to a data set?

$$y = \beta_0 + \beta_1 x + \beta_2 x^2.$$

- Now the <u>objective function</u> to be minimized is:

$$SSR = \sum_{i=1}^{m} \left(y_i - (\beta_0 + \beta_1 x_i + \beta_2 x_i^2)\right)^2$$

- Once again we will take the derivative of the SSR w.r.t. the 3 model parameters β_i and set the derivatives =0 to find the extrema…

The Normal Equations (for linear regression)

$$\frac{dSSR}{d\beta_0} = \sum_{i=1}^{m} -2(y_i - (\beta_0 + \beta_1 x_i + \beta_2 x_i^2)) = 0,$$

$$\frac{dSSR}{d\beta_1} = \sum_{i=1}^{m} -2x_i(y_i - (\beta_0 + \beta_1 x_i + \beta_2 x_i^2)) = 0,$$

$$\frac{dSSR}{d\beta_2} = \sum_{i=1}^{m} -2x_i^2(y_i - (\beta_0 + \beta_1 x_i + \beta_2 x_i^2)) = 0.$$

Multiplying this out we get…

The Normal Equations (for linear regression)

$$m\beta_0 + \beta_1 \sum_{i=1}^{m} x_i + \beta_2 \sum_{i=1}^{m} x_i^2 = \sum_{i=1}^{m} y_i,$$

$$\beta_0 \sum_{i=1}^{m} x_i + \beta_1 \sum_{i=1}^{m} x_i^2 + \beta_2 \sum_{i=1}^{m} x_i^3 = \sum_{i=1}^{m} x_i y_i,$$

$$\beta_0 \sum_{i=1}^{m} x_i^2 + \beta_1 \sum_{i=1}^{m} x_i^3 + \beta_2 \sum_{i=1}^{m} x_i^4 = \sum_{i=1}^{m} x_i^2 y_i,$$

Or, in matrix form…

The Normal Equations (for linear regression)

$$\begin{bmatrix} m & \sum_{i=1}^{m} x_i & \sum_{i=1}^{m} x_i^2 \\ \sum_{i=1}^{m} x_i & \sum_{i=1}^{m} x_i^2 & \sum_{i=1}^{m} x_i^3 \\ \sum_{i=1}^{m} x_i^2 & \sum_{i=1}^{m} x_i^3 & \sum_{i=1}^{m} x_i^4 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{m} y_i \\ \sum_{i=1}^{m} x_i y_i \\ \sum_{i=1}^{m} x_i^2 y_i \end{bmatrix}$$

These the normal equations for a quadratic function.

The Normal Equations (for linear regression)

- We need a more general formulation, to handle any linear combination of functions of x:

$$y = \beta_1 f_1(x) + \beta_2 f_2(x) + \cdots + \beta_n f_n(x)$$

- Like example models such as these:

$$y = \frac{\beta_1}{x} + \beta_2 \tan x,$$
$$y = \sqrt{x}(\beta_1 \sin x + \beta_2 \cos x),$$
$$y = \beta_1 \ln x + \beta_2 x^3 + \beta_3 x^{1/3}.$$

The Normal Equations (for linear regression)

- First, let's add the model function of "1", to include a general constant β_0:

$$y = \beta_0 + \beta_1 f_1(x) + \cdots + \beta_n f_n(x).$$

- If we have m data points, then we have a (m)x(n) matrix form:

$$\boldsymbol{A} = \begin{bmatrix} 1 & f_1(x_1) & \ldots & f_n(x_1) \\ 1 & f_1(x_2) & \ldots & f_n(x_2) \\ & & \vdots & \\ & & \cdot & \\ 1 & f_1(x_m) & \ldots & f_n(x_m) \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \cdot \\ y_m \end{bmatrix},$$

The Normal Equations (for linear regression)

We seek the minimum of

$$\|\boldsymbol{r}\|^2 = \boldsymbol{r}^{\mathrm{T}}\boldsymbol{r} = (\boldsymbol{y} - \hat{\boldsymbol{y}})^{\mathrm{T}}(\boldsymbol{y} - \hat{\boldsymbol{y}}) = (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{c})^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{c}),$$

$$\|\boldsymbol{r}\|^2 = \boldsymbol{y}^{\mathrm{T}}\boldsymbol{y} - \boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{c} - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{c}.$$

- Note the transpose rule: $(\boldsymbol{A}\boldsymbol{B})^{\mathrm{T}} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}$
- Again we differentiate w.r.t. the model parameters and set the partial derivatives =0…

The Normal Equations (for linear regression)

$$\|\boldsymbol{r}\|^2 = \boldsymbol{y}^{\mathrm{T}}\boldsymbol{y} - \boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{c} - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{c}.$$

$$\Downarrow$$

$$\nabla_c(\boldsymbol{r}^{\mathrm{T}}\boldsymbol{r}) = 0 - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{c} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{c} = 0,$$

where

$$\nabla_c = \left(\frac{\partial}{\partial\beta_0}, \frac{\partial}{\partial\beta_1}, \ldots, \frac{\partial}{\partial\beta_n}\right)$$

Thus, $\boxed{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{c} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}.}$

The Normal Equations (for linear regression)

Two main steps are involved when applying the normal equations (Equation 2.46):

(1) construction of $\boldsymbol{A}$, and
(2) solution of $\boldsymbol{c} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}$ in MATLAB as `c = (A*A)\A*y`.

Q1: which is NOT a way to create a column vector in MATLAB?

A. v = [1 2 3]’
B. v = [1 2 3]
C. v = [1,2,3]’
D. v = [1;2;3]

Q2: The MATLAB command ones(m) creates a square matrix of size m x m, with elements all =1. The command eye(n) creates an n x n matrix with ones along the diagonal and the other elements =0 (the identity matrix I).

How could you use these two commands to create the matrix:

A. ones(3) + 2*eye(3)
B. ones(2) + 3*eye(2)
C. ones(3) + 3*eye(2)
D. ones(2) + 2*eye(3)
E. ones(3) + 3*eye(3)

D =

3	1	1
1	3	1
1	1	3

Q3: At the MATLAB prompt, the statement A(2,1) would return which value?

A. 7
B. 5
C. 1
D. 4
E. ??? Index exceeds matrix dimensions.

$$\boldsymbol{A} = \begin{bmatrix} 4 & 7 & 2 \\ 5 & 1 & 8 \end{bmatrix}$$

Q4: For the row vector x = (4 0.5), which of the following statements would NOT calculate the norm?

A. sqrt(x*x’)
B. norm(x)
C. x(1)*x(1) + x(2)*x(2)
D. sqrt(sum(x.^2))

Q5: The following matrix equation represents a linear system of 3 coupled equations.

$$\boldsymbol{A}\boldsymbol{x} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \boldsymbol{b}.$$

Which of the following IS a valid mathematical expression?

A. $a_{11}b_1 + a_{12}b_2 + a_{13}b_3 = x_1$
B. $a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = b_2$
C. $a_{11}x_1 + a_{21}x_2 + a_{31}x_3 = b_1$
D. $a_{11} + a_{12} + a_{13} + \mathbf{x} = \mathbf{b}$

Q6: Which one of the following matrices CAN be inverted in MATLAB using the inv command?

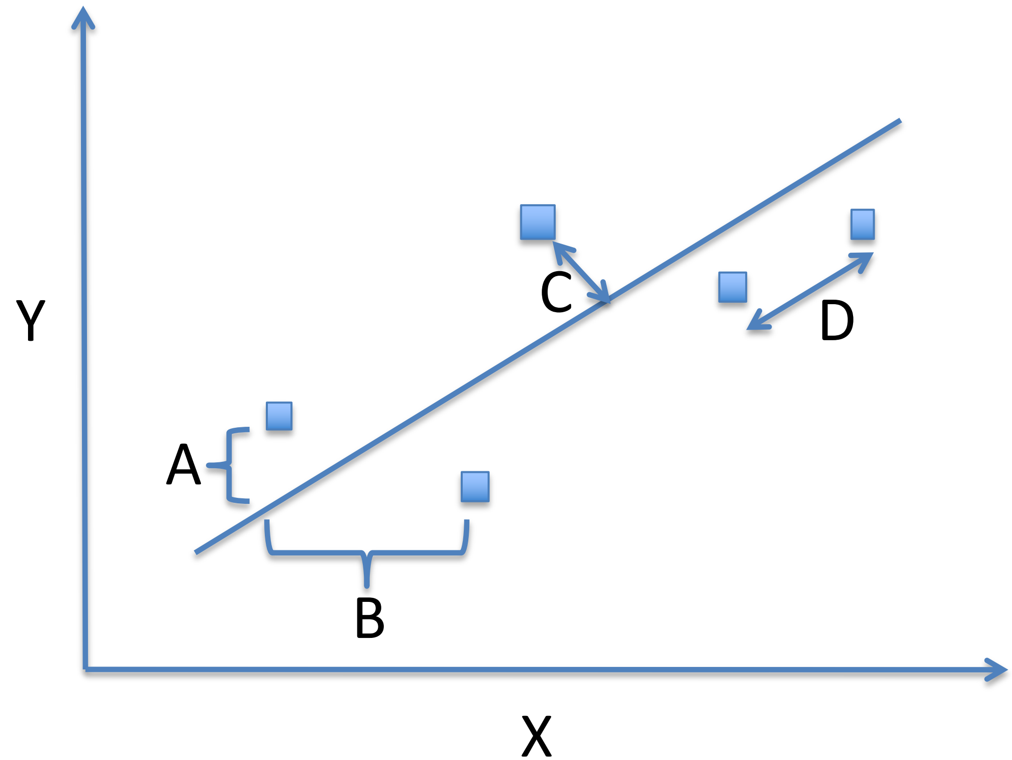
A. $$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 3 \\ 2 & 2 & 2 \end{pmatrix}$$

B. $$\begin{pmatrix} 1 & 1 & 2 \\ 4 & 3 & 2 \end{pmatrix}$$

C. $$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 5 & 5 & 5 \end{pmatrix}$$

D. $$\begin{pmatrix} 1 & 1 & 2 \\ 0 & 0 & 0 \\ 2 & 3 & 4 \end{pmatrix}$$

Q7: In least squares linear regression, which error distance is minimized?



Q8: True or False: outlier data points dominate the best fit slope of a line.

A. True
B. False

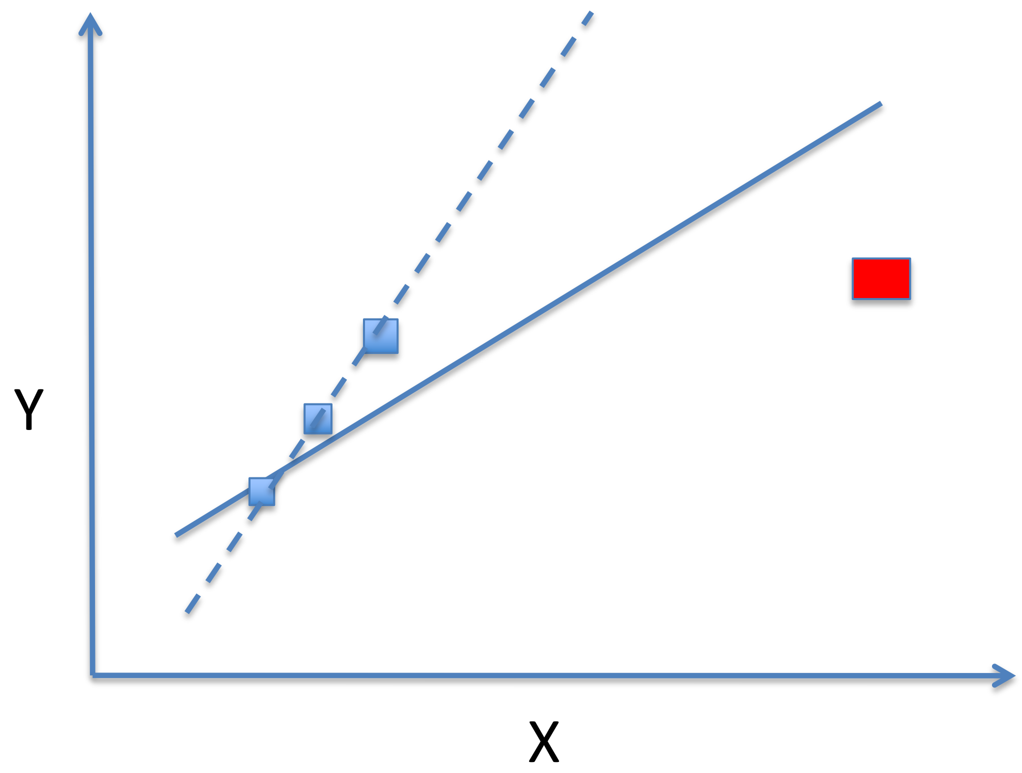
Q8: True or False: outlier data points dominate the best fit slope of a line.

A. True

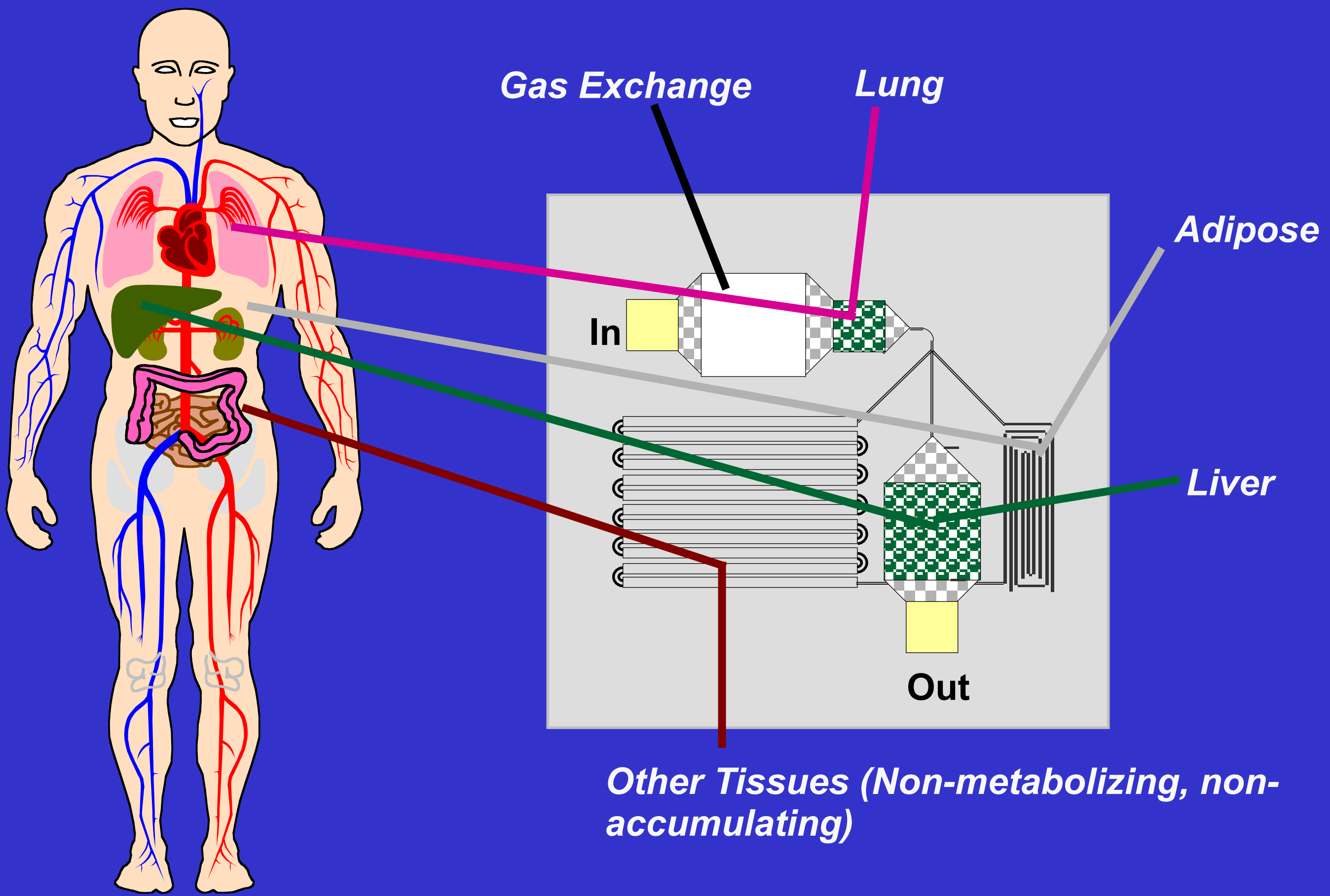
B. False

TRUE!

SSR = $1^2 + 1^2 + 1^2 + 3^2 = 12$



The original vision: an "animal on a chip" (i.e., an *in vitro*, multi-tissue, microfluidic, cell-based assay platform for improved pharmacological / toxicological prediction)



Working together with bioengineers and experts in liver microstructure we developed microfluidic, cell-based biochips

- Individual compartments contain cultures of living cells of different organs
- Heterogeneous cell types mimic different organs or tissues of an animal (and humans)
- Compartments fluidically interconnected
- Fluid and compounds recirculate

  as in a living system

(See ***Nature***, **435**: 12-13, May 5, 2005;
Forbes, August 15, 2005, pp. 53-54;
The Observer, September 25, 2005, p.7;
Newsweek, October 10, 2005, p.59
Nature, **471**: 661-665, March 31, 2011)

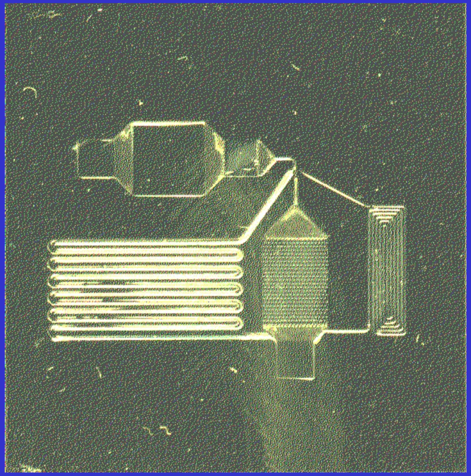


Photo of early prototype silicon biochip

Hµrel (human relevant) Corporation
A new model in private-private partnerships
www.hurelcorp.com

H*µ*REL*static*™
liver-simulative cellular co-culture; highest functionality available; convenience of arriving ready for use "out of the box" in standard labware; robotics-compatible, established workflow-compatible (available Q3-Q4 2010)

H*µ*REL*flow*™
contract research service performed by Hurel.high-functioning liver-simulative cellular co-culture with additional attribute of flow; flow affords cellular functionality higher still by multifolds (available Q3-Q4 2010)

H*µ*REL*flow*™
customer-operable version of H*µ*REL*flow*™. high-value, consumable biochips that arrive with H*µ*REL*static*™ cells pre-cultured on them (Q1 2012)

"Allergy Test on a Chip™
skin allergenicity test to replace Local Lymph Node Assay ("LLNA"), animal-based test widely used in cosmetics and industrial and consumer product toxicology (prototype 4Q11; production version 2013)

follow-on application areas
multi-tissue; Hepatitis C and other virology; immunology; environmental testing, biodefense testing