From least squares: $y_i = \beta_0 + \beta_1 x_i + e_i$

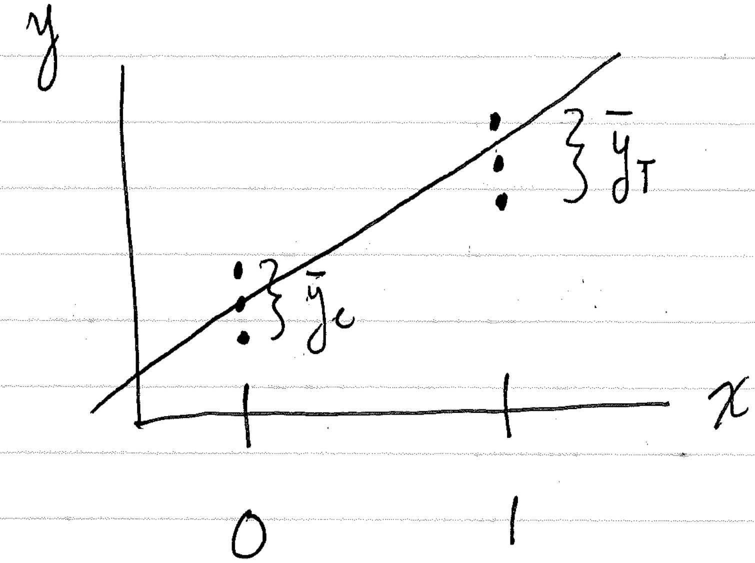
$i = 1, \ldots, n$

e_i: indep random var w/

$E(e_i) = 0$ & $Var(e_i) = \sigma^2$

$$\hat\beta_0 = \bar y - \hat\beta_1 \bar x$$

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



crudely:

$$\beta_1 = \frac{\Delta y}{\Delta x} = \frac{\bar y_T - \bar y_c}{1 - 0}$$

$$= \bar y_T - \bar y_c$$

Is this exact?

$$\hat\beta_1 = \frac{\sum_{i \in T} \left(1 - \frac{1}{2}\right)(y_i - \bar y) + \sum_{i \in C} \left(0 - \frac{1}{2}\right)(y_i - \bar y)}{\sum_{i \in T} \left(1 - \frac{1}{2}\right)^2 + \sum_{i \in C} \left(0 - \frac{1}{2}\right)^2}$$

$$= \frac{\frac{1}{2} \sum_{i \in T} y_i - \frac{1}{2} \sum_{i \in C} y_i + \frac{-1}{2} \cancel{\bar y} \cdot T + \frac{1}{2} \cancel{\bar y} \cdot C}{\left(\frac{1}{2}\right)^2 \cdot T + \left(\frac{1}{2}\right)^2 \cdot C}$$

Assume $T = C$

$$\hat{\beta}_1 = \frac{\frac{1}{2}\cdot T\,\bar{y}_T - \frac{1}{2}\cdot c\,\bar{y}_c}{\frac{1}{2^2}(T+C)}$$

$$= \frac{\frac{1}{2}T(\bar{y}_T - \bar{y}_c)}{\frac{1}{2^2}(2T)} = \frac{\frac{1}{\cancel{2}}\cancel{T}(\bar{y}_T - \bar{y}_c)}{\frac{1}{\cancel{2}}\cancel{T}}$$

$$\boxed{\hat{\beta}_1 = \bar{y}_T - \bar{y}_c}$$

Can calc $\text{var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = S_{\beta_1}^2$

Estimate σ^2 using

$$\hat{\sigma}^2 = \frac{RSS}{n-2}\ ;\ RSS = \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

t-statistic:

$$t = \frac{\hat{\beta}_1}{S_{\beta_1}}$$ same as t-stat in t-test!

Only restriction: linear model assumes equal variance!

Starting w/ $\hat{\beta}_0 = \bar{y}_c$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\bar{y} = \frac{1}{(T+C)}\left(T \cdot \bar{y}_T + C\,\bar{y}_c\right)$$

$$\bar{x} = \frac{1}{2}$$

$$\hat{\beta}_0 \overset{?}{=} \bar{y}_c$$

$$\therefore\ \hat{\beta}_1 \cdot \frac{1}{2} = \cancel{\hat{\beta}_0 - \bar{y}} = \bar{y} - \hat{\beta}_0$$

$$= \frac{\cancel{T}}{2 \cdot \cancel{T}}\left(\bar{y}_T + \bar{y}_c\right) - \bar{y}_c$$

$$\hat{\beta}_1 = \bar{y}_T + \bar{y}_c - 2\bar{y}_c$$

$$= \bar{y}_T - \bar{y}_c \quad ✓$$

DESeq & DESeq2 Size Factor

If gene i is not differentially expressed between samples j and j' then

$E(K_{ij})/E(K_{ij'}) = s_j/s_{j'}$

Generalize this to all m samples for gene i

$$\frac{E(K_{i1})}{s_1} = \frac{E(K_{i2})}{s_2} = \cdots = \frac{E(K_{im})}{s_m}$$

$$\frac{s_{ij}}{s_1} \cdot \frac{s_{ij}}{s_2} \cdot \frac{s_{ij}}{s_3} \cdots \frac{s_{ij}}{s_m} = \frac{K_{ij}}{K_{i1}} \cdot \frac{K_{ij}}{K_{i2}} \cdot \frac{K_{ij}}{K_{i3}} \cdots \frac{K_{ij}}{K_{im}}$$

$$\frac{s_{ij}^m}{\prod_{j=1}^m s_{ij}} = \frac{K_{ij}^m}{\prod_{j=1}^m K_{ij}}$$

Can set $\prod_{j=1}^m s_{ij} = 1$

Sets normalization scale!

$$\therefore \; s_{ij} = \frac{K_{ij}}{\left(\prod_{j=1}^m K_{ij}\right)^{1/m}}, \quad \prod_{j=1}^m s_{ij} = \frac{\prod_{j=1}^m K_{ij}}{\prod_{j=1}^m K_{ij}} = 1 \checkmark$$

How find i? Take median!

$$\boxed{\hat{s}_j = \text{median} \frac{K_{ij}}{\left(\prod_{j=1}^m K_{ij}\right)^{1/m}}}$$

$$\hat{\theta}_{MLE}(x) = \arg\max_{\theta} f(x|\theta)$$

Bayes' theorem

$$f(\theta|x) = \frac{f(x|\theta)\, g(\theta)}{\int_{\Theta} f(x|\varphi)\, g(\varphi)\, d\varphi}$$

Maximum a posteriori estimation estimates θ as mode of posterior distribution

$$\hat{\theta}_{MAP} = \arg\max_{\theta} f(\theta|x)$$

$$= \arg\max_{\theta} \frac{f(x|\theta)\, g(\theta)}{\int_{\Theta} f(x|\varphi)\, g(\varphi)\, d\varphi}$$

↳ >0 and does depend on θ!

$$\boxed{\hat{\theta}_{MAP} = \arg\max_{\theta} f(x|\theta)\, g(\theta)}$$

Empirical Bayes: prior distribution is estimated from the data (rather than fixed before data observed)

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$

$$\log p(x) = -\frac{1}{2}\log(2\cdot\pi\cdot\sigma^2) - (x-\mu)^2/2\sigma^2$$

$$= -\frac{1}{2}\log(2\cdot\pi) - \log\sigma - (x-\mu)^2/2\sigma^2$$

$$= -\frac{1}{2}\log(2\cdot\pi)$$

Log Normal Dist:

$X \sim$ Log normal

$E(x) = e^{\mu + \sigma^2/2}$ mean

$var(X) = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}$

$\ln(X) \sim N(\mu, \sigma^2)$

Negative ~~Bin~~ Binomial Dist:

$X \sim NB$

$E(x) = \frac{n(1-p)}{p} = \mu$ mean

$var(x) = n(1-p)/p^2 = \sigma^2$

$$\frac{1}{\sigma^2}\cdot\mu = \frac{p^{\cancel{2}\,1}}{\cancel{n(1-p)}}\cdot\frac{\cancel{n(1-p)}}{\cancel{p}} = p$$

$$\therefore \quad \boxed{p = \frac{\mu}{\sigma^2}} \qquad \boxed{p = \frac{\mu}{\sigma^2}}$$

$$\frac{n(1-p)}{p} = \mu \qquad n = \frac{p\mu}{1-p}$$

$$\cancel{\frac{p}{1-p} = \frac{\mu^{1}/\sigma^2}{(\sigma^2-\mu^2)/\sigma^2} = \frac{\mu^2}{\sigma^2-\mu^2}}$$

or

$$\mu^2 = \frac{n^2(1-p)^2}{p^2} \qquad \sigma^2 = \frac{n(1-p)}{p^2}$$

$$\frac{\mu^2}{\sigma^2} = \frac{n^{\cancel{2}\,1}(1-p)^{\cancel{2}\,1}}{\cancel{p^2}}\,\frac{\cancel{p^2}}{\cancel{n(1-p)}} = n(1-p)$$

$$\therefore \quad n = \frac{\mu^2}{\sigma^2}\,\frac{1}{1-p} = \frac{\mu^2}{\sigma^2}\,\frac{1}{1-\mu/\sigma^2}$$

$$= \frac{\mu^2}{\sigma^2-\mu} \qquad \boxed{n = \frac{\mu^2}{\sigma^2-\mu}}$$