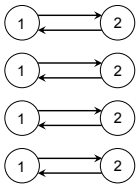
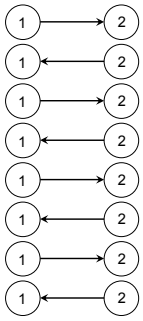
Choosing between Competing Experimental Designs

2/23/2009

Design 1 for Comparing Two Treatments



Design 2 for Comparing Two Treatments



Design 1: Mixed Linear Model for the Observations from a Single Gene

$Y_{1111}=\mu+\tau_1+\delta_1+s_1+b_{11}+e_{1111}$ → $Y_{2211}=\mu+\tau_2+\delta_2+s_1+b_{21}+e_{2211}$

$Y_{1221}=\mu+\tau_1+\delta_2+s_2+b_{11}+e_{1221}$ ← $Y_{2121}=\mu+\tau_2+\delta_1+s_2+b_{21}+e_{2121}$

$Y_{1132}=\mu+\tau_1+\delta_1+s_3+b_{12}+e_{1132}$ → $Y_{2232}=\mu+\tau_2+\delta_2+s_3+b_{22}+e_{2232}$

$Y_{1242}=\mu+\tau_1+\delta_2+s_4+b_{12}+e_{1242}$ ← $Y_{2142}=\mu+\tau_2+\delta_1+s_4+b_{22}+e_{2142}$

$Y_{1153}=\mu+\tau_1+\delta_1+s_5+b_{13}+e_{1153}$ → $Y_{2253}=\mu+\tau_2+\delta_2+s_5+b_{23}+e_{2253}$

$Y_{1263}=\mu+\tau_1+\delta_2+s_6+b_{13}+e_{1263}$ ← $Y_{2163}=\mu+\tau_2+\delta_1+s_6+b_{23}+e_{2163}$

$Y_{1174}=\mu+\tau_1+\delta_1+s_7+b_{14}+e_{1174}$ → $Y_{2274}=\mu+\tau_2+\delta_2+s_7+b_{24}+e_{2274}$

$Y_{1284}=\mu+\tau_1+\delta_2+s_8+b_{14}+e_{1284}$ ← $Y_{2184}=\mu+\tau_2+\delta_1+s_8+b_{24}+e_{2184}$

$$Y_{ijkl}=\mu+\tau_i+\delta_j+s_k+b_{il}+e_{ijkl}$$

Design 2: Mixed Linear Model for the Observations from a Single Gene

$Y_{1111}=\mu+\tau_1+\delta_1+s_1+b_{11}+e_{1111}$ → $Y_{2211}=\mu+\tau_2+\delta_2+s_1+b_{21}+e_{2211}$

$Y_{1222}=\mu+\tau_1+\delta_2+s_2+b_{12}+e_{1222}$ ← $Y_{2122}=\mu+\tau_2+\delta_1+s_2+b_{22}+e_{2122}$

$Y_{1133}=\mu+\tau_1+\delta_1+s_3+b_{13}+e_{1133}$ → $Y_{2233}=\mu+\tau_2+\delta_2+s_3+b_{23}+e_{2233}$

$Y_{1244}=\mu+\tau_1+\delta_2+s_4+b_{14}+e_{1244}$ ← $Y_{2144}=\mu+\tau_2+\delta_1+s_4+b_{24}+e_{2144}$

$Y_{1155}=\mu+\tau_1+\delta_1+s_5+b_{15}+e_{1155}$ → $Y_{2255}=\mu+\tau_2+\delta_2+s_5+b_{25}+e_{2255}$

$Y_{1266}=\mu+\tau_1+\delta_2+s_6+b_{16}+e_{1266}$ ← $Y_{2166}=\mu+\tau_2+\delta_1+s_6+b_{26}+e_{2166}$

$Y_{1177}=\mu+\tau_1+\delta_1+s_7+b_{17}+e_{1177}$ → $Y_{2277}=\mu+\tau_2+\delta_2+s_7+b_{27}+e_{2277}$

$Y_{1288}=\mu+\tau_1+\delta_2+s_8+b_{18}+e_{1288}$ ← $Y_{2188}=\mu+\tau_2+\delta_1+s_8+b_{28}+e_{2188}$

$$Y_{ijkl}=\mu+\tau_i+\delta_j+s_k+b_{il}+e_{ijkl}$$

Design 2: Mixed Linear Model for the Observations from a Single Gene

$Y_{1111}=\mu+\tau_1+\delta_1+s_1+b_{11}+e_{1111}$ → $Y_{2211}=\mu+\tau_2+\delta_2+s_1+b_{21}+e_{2211}$

$Y_{1222}=\mu+\tau_1+\delta_2+s_2+b_{12}+e_{1222}$ ← $Y_{2122}=\mu+\tau_2+\delta_1+s_2+b_{22}+e_{2122}$

$Y_{1133}=\mu+\tau_1+\delta_1+s_3+b_{13}+e_{1133}$ → $Y_{2233}=\mu+\tau_2+\delta_2+s_3+b_{23}+e_{2233}$

$Y_{1244}=\mu+\tau_1+\delta_2+s_4+b_{14}+e_{1244}$ ← $Y_{2144}=\mu+\tau_2+\delta_1+s_4+b_{24}+e_{2144}$

$Y_{1155}=\mu+\tau_1+\delta_1+s_5+b_{15}+e_{1155}$ → $Y_{2255}=\mu+\tau_2+\delta_2+s_5+b_{25}+e_{2255}$

$Y_{1266}=\mu+\tau_1+\delta_2+s_6+b_{16}+e_{1266}$ ← $Y_{2166}=\mu+\tau_2+\delta_1+s_6+b_{26}+e_{2166}$

$Y_{1177}=\mu+\tau_1+\delta_1+s_7+b_{17}+e_{1177}$ → $Y_{2277}=\mu+\tau_2+\delta_2+s_7+b_{27}+e_{2277}$

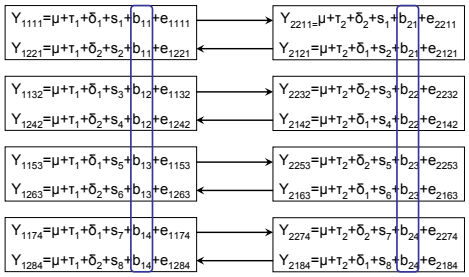
$Y_{1288}=\mu+\tau_1+\delta_2+s_8+b_{18}+e_{1288}$ ← $Y_{2188}=\mu+\tau_2+\delta_1+s_8+b_{28}+e_{2188}$

Note that b and e are completely confounded in Design 2. Thus we would use only one random residual term for both factors, but we write the terms separately here for the sake of comparison with Design 1.

Test of Interest

- $H_0 : \tau_1 = \tau_2$ vs. $H_A : \tau_1 \neq \tau_2$
- Equivalent to $H_0 : \tau_1 - \tau_2 = 0$ vs. $H_A : \tau_1 - \tau_2 \neq 0$
- We estimate $\tau_1 - \tau_2$ by $\overline{Y}_{1\cdots} - \overline{Y}_{2\cdots}$
- $\overline{Y}_{1\cdots} - \overline{Y}_{2\cdots} = \tau_1 - \tau_2 + \overline{b}_{1\cdot} - \overline{b}_{2\cdot} + \overline{e}_{1\cdots} - \overline{e}_{2\cdots}$

Design 1 Estimate of τ_1-τ_2

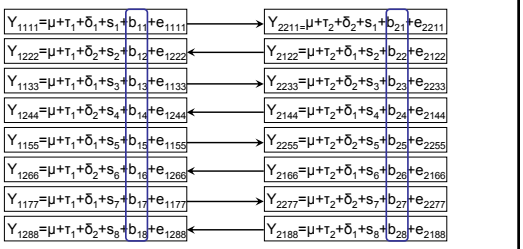


$\overline{Y}_{1\cdots}=\mu+\tau_1+\overline{\delta}_{\cdot}+\overline{s}_{\cdot}+\overline{b}_{1\cdot}+\overline{e}_{1\cdots}$ $\quad$ $\overline{Y}_{2\cdots}=\mu+\tau_2+\overline{\delta}_{\cdot}+\overline{s}_{\cdot}+\overline{b}_{2\cdot}+\overline{e}_{2\cdots}$

Average over 4 effects

$\overline{Y}_{1\cdots} - \overline{Y}_{2\cdots} = \tau_1 - \tau_2 + \overline{b}_{1\cdot} - \overline{b}_{2\cdot} + \overline{e}_{1\cdots} - \overline{e}_{2\cdots}$

Design 2 Estimate of τ_1-τ_2



$\overline{Y}_{1\cdots}=\mu+\tau_1+\overline{\delta}_{\cdot}+\overline{s}_{\cdot}+\overline{b}_{1\cdot}+\overline{e}_{1\cdots}$ $\quad$ $\overline{Y}_{2\cdots}=\mu+\tau_2+\overline{\delta}_{\cdot}+\overline{s}_{\cdot}+\overline{b}_{2\cdot}+\overline{e}_{2\cdots}$

Average over 8 effects

$\overline{Y}_{1\cdots} - \overline{Y}_{2\cdots} = \tau_1 - \tau_2 + \overline{b}_{1\cdot} - \overline{b}_{2\cdot} + \overline{e}_{1\cdots} - \overline{e}_{2\cdots}$

Variance of the Estimated Difference

- $\overline{Y}_{1\cdots} - \overline{Y}_{2\cdots} = \tau_1 - \tau_2 + \overline{b}_{1\cdot} - \overline{b}_{2\cdot} + \overline{e}_{1\cdots} - \overline{e}_{2\cdots}$
- $\mathrm{Var}(\overline{Y}_{1\cdot\cdot} - \overline{Y}_{2\cdot\cdot}) = \mathrm{Var}(\overline{b}_{1\cdot} - \overline{b}_{2\cdot} + \overline{e}_{1\cdots} - \overline{e}_{2\cdots})$
- For Design 1: $\frac{\sigma_b^2 + \sigma_b^2}{4} + \frac{\sigma_e^2 + \sigma_e^2}{8} = \frac{2\sigma_b^2 + \sigma_e^2}{4}$

  Design 1 variance is never smaller than Design 2 variance.
- For Design 2: $\frac{\sigma_b^2 + \sigma_b^2 + \sigma_e^2 + \sigma_e^2}{8} = \frac{\sigma_b^2 + \sigma_e^2}{4}$

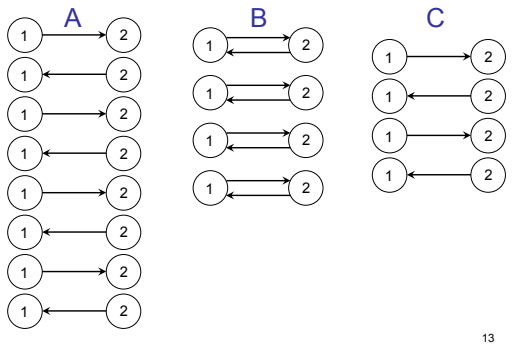
Design 2 is Preferred over Design 1

- The variance of the estimated treatment difference for Design 2 will always be lower than or equal to the variance for Design 1.
- The standard errors for Design 2 will tend to smaller than for Design 1.
- The *t*-statistics for Design 2 will tend to be more extreme than the *t*-statistics for Design 1 when genes are truly differentially expressed.
- The *p*-values for differentially expressed genes will tend be smaller with Design 2 than with Design 1.
- Design 2 has more power for detecting differential expression than Design 1.

Some General Microarray Experimental Design Advice

- Use as much biological replication as is affordable.
- If the number of microarray slides or GeneChips is the limiting factor, measure each sample only once. Measuring any one sample more than once reduces the degree of biological replication that is possible, and this reduces the power to detect differential expression.
- If the number of biological replications is the limiting factor, measuring each experimental unit multiple times can improve precision, but this technical replication is no substitute for biological replication.

For example, Design A > Design B > Design C



13

An Analysis Based on Red – Green Differences

$Y_{1111}=\mu+\tau_1+\delta_1+s_1+b_{11}+e_{1111}$ → $Y_{2211}=\mu+\tau_2+\delta_2+s_1+b_{21}+e_{2211}$ d_1

$Y_{1222}=\mu+\tau_1+\delta_2+s_2+b_{12}+e_{1222}$ ← $Y_{2122}=\mu+\tau_2+\delta_1+s_2+b_{22}+e_{2122}$ d_2

$Y_{1133}=\mu+\tau_1+\delta_1+s_3+b_{13}+e_{1133}$ → $Y_{2233}=\mu+\tau_2+\delta_2+s_3+b_{23}+e_{2233}$ d_3

$Y_{1244}=\mu+\tau_1+\delta_2+s_4+b_{14}+e_{1244}$ ← $Y_{2144}=\mu+\tau_2+\delta_1+s_4+b_{24}+e_{2144}$ d_4

$Y_{1155}=\mu+\tau_1+\delta_1+s_5+b_{15}+e_{1155}$ → $Y_{2255}=\mu+\tau_2+\delta_2+s_5+b_{25}+e_{2255}$ d_5

$Y_{1266}=\mu+\tau_1+\delta_2+s_6+b_{16}+e_{1266}$ ← $Y_{2166}=\mu+\tau_2+\delta_1+s_6+b_{26}+e_{2166}$ d_6

$Y_{1177}=\mu+\tau_1+\delta_1+s_7+b_{17}+e_{1177}$ → $Y_{2277}=\mu+\tau_2+\delta_2+s_7+b_{27}+e_{2277}$ d_7

$Y_{1288}=\mu+\tau_1+\delta_2+s_8+b_{18}+e_{1288}$ ← $Y_{2188}=\mu+\tau_2+\delta_1+s_8+b_{28}+e_{2188}$ d_8

14

An Analysis Based on Red – Green Differences

$Y_{1111}=\mu+\tau_1+\delta_1+s_1+b_{11}+e_{1111}$ → $Y_{2211}=\mu+\tau_2+\delta_2+s_1+b_{21}+e_{2211}$ d_1

$Y_{1222}=\mu+\tau_1+\delta_2+s_2+b_{12}+e_{1222}$ ← $Y_{2122}=\mu+\tau_2+\delta_1+s_2+b_{22}+e_{2122}$ d_2

$Y_{2211}-Y_{1111}=(\mu+\tau_2+\delta_2+s_1+b_{21}+e_{2211})-(\mu+\tau_1+\delta_1+s_1+b_{11}+e_{1111})$

$Y_{2211}-Y_{1111}=\delta_2-\delta_1+\tau_2-\tau_1+b_{21}-b_{11}+e_{2211}-e_{1111}$

$d_1 = \mu_1 + r_1$

$Y_{1222}-Y_{2122}=\delta_2-\delta_1+\tau_1-\tau_2+b_{12}-b_{22}+e_{1222}-e_{2122}$

$d_2 = \mu_2 + r_2$

15

Test for Differential Expression Using a 2-Sample *t*-Test

$d_1 = \mu_1 + r_1$
$d_2 = \mu_2 + r_2$
$d_3 = \mu_1 + r_3$
$d_4 = \mu_2 + r_4$
$d_5 = \mu_1 + r_5$
$d_6 = \mu_2 + r_6$
$d_7 = \mu_1 + r_7$
$d_8 = \mu_2 + r_8$

$\mu_1 = \delta_2-\delta_1+\tau_2-\tau_1$ $\mu_2 = \delta_2-\delta_1+\tau_1-\tau_2$

$\mu_1 = \mu_2$
$\delta_2-\delta_1+\tau_2-\tau_1 = \delta_2-\delta_1+\tau_1-\tau_2$
$\tau_2-\tau_1 = \tau_1-\tau_2$
$\tau_1 = \tau_2$

A 2-sample *t*-test of H_0:$\mu_1=\mu_2$ using d_1,d_3,d_5,d_7 as one sample and d_2,d_4,d_6,d_8 as the other sample is a test of H_0: $\tau_1=\tau_2$.

16

An Analysis Based on Red – Green Differences

$Y_{1111}=\mu+\tau_1+\delta_1+s_1+b_{11}+e_{1111}$ → $Y_{2211}=\mu+\tau_2+\delta_2+s_1+b_{21}+e_{2211}$ d_1

$Y_{1222}=\mu+\tau_1+\delta_2+s_2+b_{12}+e_{1222}$ ← $Y_{2122}=\mu+\tau_2+\delta_1+s_2+b_{22}+e_{2122}$ d_2

$Y_{2211}-Y_{1111}=(\mu+\tau_2+\delta_2+s_1+b_{21}+e_{2211})-(\mu+\tau_1+\delta_1+s_1+b_{11}+e_{1111})$

$Y_{2211}-Y_{1111}=\delta_2-\delta_1+\tau_2-\tau_1+b_{21}-b_{11}+e_{2211}-e_{1111}$

$Y_{2211}-Y_{1111}=\delta_2-\delta_1+(\tau_1-\tau_2)(-1)+b_{21}-b_{11}+e_{2211}-e_{1111}$

$d_1 = \beta_0 + \beta_1 * x_1 + r_1$

$Y_{1222}-Y_{2122}=\delta_2-\delta_1+(\tau_1-\tau_2)(1)+b_{12}-b_{22}+e_{1222}-e_{2122}$

$d_2 = \beta_0 + \beta_1 * x_2 + r_2$

Note that test of interest becomes $H_0: \beta_1 = 0$ vs. $H_A: \beta_1 \neq 0$.

17

Test for Differential Expression
Using Simple Linear Regression

$d_1 = \beta_0 + \beta_1 x_1 + r_1$
$d_2 = \beta_0 + \beta_1 x_2 + r_2$
$d_3 = \beta_0 + \beta_1 x_3 + r_3$
$d_4 = \beta_0 + \beta_1 x_4 + r_4$
$d_5 = \beta_0 + \beta_1 x_5 + r_5$
$d_6 = \beta_0 + \beta_1 x_6 + r_6$
$d_7 = \beta_0 + \beta_1 x_7 + r_7$
$d_8 = \beta_0 + \beta_1 x_8 + r_8$

18

Test for Differential Expression Using Simple Linear Regression

$$
\begin{aligned}
d_1 &= \beta_0 + \beta_1(-1) + r_1 \\
d_2 &= \beta_0 + \beta_1(1) + r_2 \\
d_3 &= \beta_0 + \beta_1(-1) + r_3 \\
d_4 &= \beta_0 + \beta_1(1) + r_4 \\
d_5 &= \beta_0 + \beta_1(-1) + r_5 \\
d_6 &= \beta_0 + \beta_1(1) + r_6 \\
d_7 &= \beta_0 + \beta_1(-1) + r_7 \\
d_8 &= \beta_0 + \beta_1(1) + r_8
\end{aligned}
$$

19

Same Equations in Matrix Form

$$
\begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \\ d_6 \\ d_7 \\ d_8 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \\ 1 & -1 \\ 1 & 1 \\ 1 & -1 \\ 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \\ r_6 \\ r_7 \\ r_8 \end{pmatrix}
$$

$$\mathbf{d} = \mathbf{X}\boldsymbol{\beta} + \mathbf{r}$$

20

Results from Multiple Regression

Note that the elements of $\mathbf{r}$ are independent and normally distributed with variance $\sigma^2 = 2\sigma_b^2 + 2\sigma_e^2$ according the our original linear model. Thus

- The *best linear unbiased estimate* of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{d},$$

- $\hat{\boldsymbol{\beta}}$ is normally distributed with mean $\boldsymbol{\beta}$ and var $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$,
- $\hat{\sigma}^2 = (\mathbf{d}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{d}-\mathbf{X}\hat{\boldsymbol{\beta}})/(n-p)$ is an unbiased estimate of σ^2 where n=length of $\mathbf{d}$ and p=length of $\boldsymbol{\beta}$,
- and $t=(\mathbf{g}'\hat{\boldsymbol{\beta}} - \mathbf{g}'\boldsymbol{\beta})/(\hat{\sigma}^2\mathbf{g}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{g})^{0.5} \sim$ t with n-p d.f.

21

Results from Multiple Regression for Our Example

$$(\mathbf{X}'\mathbf{X}) = \begin{pmatrix} 8 & 0 \\ 0 & 8 \end{pmatrix} \qquad (\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 1/8 & 0 \\ 0 & 1/8 \end{pmatrix}$$

$$\mathbf{X}'\mathbf{d} = \begin{pmatrix} d_1+d_2+d_3+d_4+d_5+d_6+d_7+d_8 \\ -d_1+d_2-d_3+d_4-d_5+d_6-d_7+d_8 \end{pmatrix} = \begin{pmatrix} Y_{\cdot 2\cdot\cdot} - Y_{\cdot 1\cdot\cdot} \\ Y_{1\cdot\cdot\cdot} - Y_{2\cdot\cdot\cdot} \end{pmatrix}$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{d} = \begin{pmatrix} \overline{Y}_{\cdot 2\cdot\cdot} - \overline{Y}_{\cdot 1\cdot\cdot} \\ \overline{Y}_{1\cdot\cdot\cdot} - \overline{Y}_{2\cdot\cdot\cdot} \end{pmatrix}$$

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2 \begin{pmatrix} 1/8 & 0 \\ 0 & 1/8 \end{pmatrix} = (2\sigma_b^2 + 2\sigma_e^2) \begin{pmatrix} 1/8 & 0 \\ 0 & 1/8 \end{pmatrix} = \begin{pmatrix} \frac{\sigma_b^2+\sigma_e^2}{4} & 0 \\ 0 & \frac{\sigma_b^2+\sigma_e^2}{4} \end{pmatrix}$$

22

A 2 x 2 Factorial Experiment

- Suppose we wish to study gene expression in chickens in response to Salmonella infection.
- We have only 6 two-color microarray slides and 12 chickens to work with.
- We wish to consider two factors:
  - infection type : mock or Salmonella
  - breed : resistant (r) or susceptible (s)

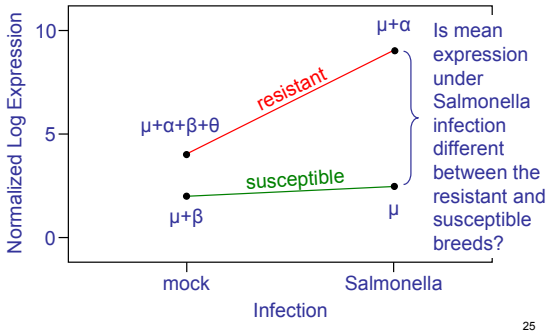
23

There are 4 Possible Treatments

Treatments		Alternative ways to code their means		
1.	mock r	μ_1	$\mu+\alpha_1+\beta_1+(\alpha\beta)_{11}$	$\mu+\alpha+\beta+\theta$
2.	mock s	μ_2	$\mu+\alpha_1+\beta_2+(\alpha\beta)_{12}$	$\mu+\beta$
3.	Salmonella r	μ_3	$\mu+\alpha_2+\beta_1+(\alpha\beta)_{21}$	$\mu+\alpha$
4.	Salmonella s	μ_4	$\mu+\alpha_2+\beta_2+(\alpha\beta)_{22}$	μ

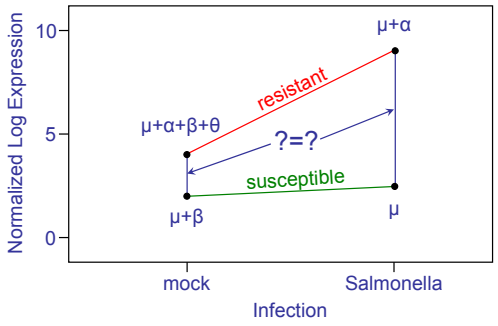
24

Tests of Interest



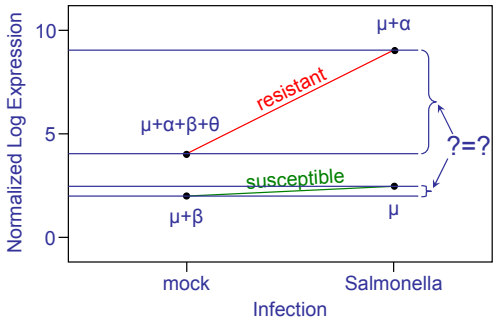
25

Is the difference that we see between the breeds the same under Salmonella infection as it is under mock infection?



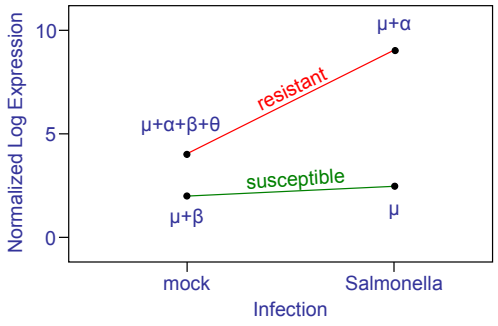
26

Does the difference between mock and Salmonella infected chickens depend on the breed?



27

Write the null hypothesis for each test of interest in terms of the treatment mean parameters.



28

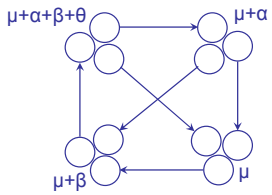
How would you assign treatments to chickens and pair chickens on slides to best answer our questions of interest?

Recall that we have 12 chickens, 6 slides, and 4 treatments.

29

A Comparison of Competing Designs

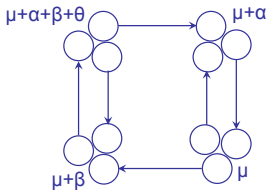
Design 1



30

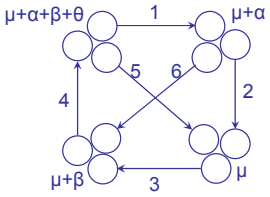
A Comparison of Competing Designs

Design 2



31

Red – Green Differences for Design 1



Slide	Mean Difference
1	$\delta_2-\delta_1-\beta-\theta$
2	$\delta_2-\delta_1-\alpha$
3	$\delta_2-\delta_1+\beta$
4	$\delta_2-\delta_1+\alpha+\theta$
5	$\delta_2-\delta_1-\alpha-\beta-\theta$
6	$\delta_2-\delta_1-\alpha+\beta$

32

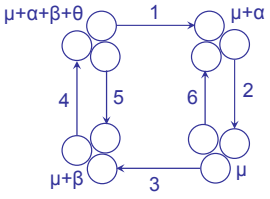
X Matrix for Design 1

Slide	Mean Difference
1	$\delta_2-\delta_1-\beta-\theta$
2	$\delta_2-\delta_1-\alpha$
3	$\delta_2-\delta_1+\beta$
4	$\delta_2-\delta_1+\alpha+\theta$
5	$\delta_2-\delta_1-\alpha-\beta-\theta$
6	$\delta_2-\delta_1-\alpha+\beta$

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & -1 & -1 \\ 1 & -1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 0 \end{pmatrix} \qquad \boldsymbol{\beta} = \begin{pmatrix} \delta_2-\delta_1 \\ \alpha \\ \beta \\ \theta \end{pmatrix}$$

33

Red – Green Differences for Design 2



Slide	Mean Difference
1	$\delta_2-\delta_1-\beta-\theta$
2	$\delta_2-\delta_1-\alpha$
3	$\delta_2-\delta_1+\beta$
4	$\delta_2-\delta_1+\alpha+\theta$
5	$\delta_2-\delta_1-\alpha-\theta$
6	$\delta_2-\delta_1+\alpha$

34

X Matrix for Design 2

Slide	Mean Difference
1	$\delta_2-\delta_1-\beta-\theta$
2	$\delta_2-\delta_1-\alpha$
3	$\delta_2-\delta_1+\beta$
4	$\delta_2-\delta_1+\alpha+\theta$
5	$\delta_2-\delta_1-\alpha-\theta$
6	$\delta_2-\delta_1+\alpha$

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & -1 & -1 \\ 1 & -1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & -1 & 0 & -1 \\ 1 & 1 & 0 & 0 \end{pmatrix} \qquad \boldsymbol{\beta} = \begin{pmatrix} \delta_2-\delta_1 \\ \alpha \\ \beta \\ \theta \end{pmatrix}$$

35

Comparing Variances of the Competing Designs

$$\boldsymbol{\beta} = \begin{pmatrix} \delta_2-\delta_1 \\ \alpha \\ \beta \\ \theta \end{pmatrix}$$

Recall that our tests of interest are $H_0 : \alpha = 0$ and $H_0 : \theta = 0$.

The variance of our estimate of **β** is given by $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$.

$(\mathbf{X}'\mathbf{X})^{-1}$ for Design 1

$$\begin{pmatrix} 0.2 & 0.10 & 0.00 & 0.0 \\ 0.1 & \boxed{0.55} & 0.25 & -0.5 \\ 0.0 & 0.25 & 0.50 & -0.5 \\ 0.0 & -0.50 & -0.50 & 1.0 \end{pmatrix}$$

$(\mathbf{X}'\mathbf{X})^{-1}$ for Design 2

$$\begin{pmatrix} 0.1875 & -0.0625 & -0.0625 & 0.125 \\ -0.0625 & \boxed{0.4375} & 0.1875 & -0.375 \\ -0.0625 & 0.1875 & 0.6875 & -0.375 \\ 0.1250 & -0.3750 & -0.3750 & 0.750 \end{pmatrix}$$

Design 2 has a lower variance for the estimate of α.

36

Comparing Variances of the Competing Designs

$$\boldsymbol{\beta} = \begin{pmatrix} \delta_2-\delta_1 \\ \alpha \\ \beta \\ \theta \end{pmatrix}$$

Recall that our tests of interest are $H_0 : \alpha = 0$ and $H_0 : \theta = 0$.

The variance of our estimate of $\boldsymbol{\beta}$ is given by $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$.

$(\mathbf{X}'\mathbf{X})^{-1}$ for Design 1

$$\begin{pmatrix} 0.2 & 0.10 & 0.00 & 0.0 \\ 0.1 & 0.55 & 0.25 & -0.5 \\ 0.0 & 0.25 & 0.50 & -0.5 \\ 0.0 & -0.50 & -0.50 & 1.0 \end{pmatrix}$$

$(\mathbf{X}'\mathbf{X})^{-1}$ for Design 2

$$\begin{pmatrix} 0.1875 & -0.0625 & -0.0625 & 0.125 \\ -0.0625 & 0.4375 & 0.1875 & -0.375 \\ -0.0625 & 0.1875 & 0.6875 & -0.375 \\ 0.1250 & -0.3750 & -0.3750 & 0.750 \end{pmatrix}$$

Design 2 has a lower variance for the estimate of θ.

37

Dominance

- Design 2 is said to *dominate* Design 1 with respect to the tests of interest because the variances of the parameter estimates are lower for each parameter of interest when using Design 2 as compared to Design 1.
- Let v_{ik} denote the variance of the estimate of the k^{th} parameter of interest using Design i. Design 2 is said to dominate Design 1 if $v_{2k} \le v_{1k}$ for all k of interest and $v_{2k} < v_{1k}$ for at least one k of interest.

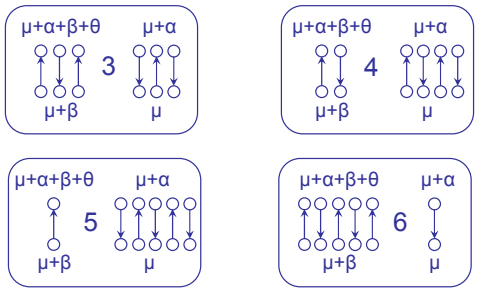
38

Admissibility

- A design is said to be *admissible* within a class of designs if there is no design in the class that dominates it.
- A design that is dominated by another design in its class is said to *inadmissible*.
- In our example, Design 1 is inadmissible among the class of designs that use 12 chickens and 6 slides because it is dominated by Design 2.
- Design 2 is also inadmissible in the class of designs that use 12 chickens and 6 slides. Can you find a design that dominates it?

39

Alternative Designs



40

The concept of admissibility for two-color microarray experiments was introduced by

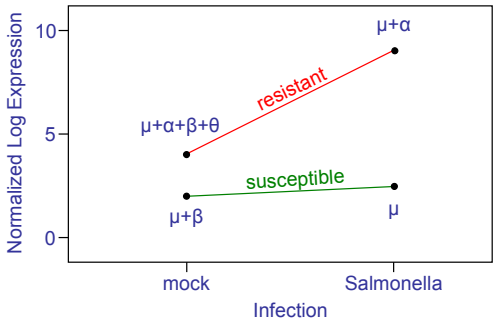
Glonek, G. F. V. and Solomon, P. J. (2004). Factorial and time course designs for cDNA microarray experiments. *Biostatistics*, **5**, 89-111.

41

Which of Design 1 or Design 2 would be better if our primary goal was to...

1. test for a dye effect?
2. test for a difference between mock and Salmonella infection for the susceptible breed?
3. test for a difference between the resistant and susceptible breeds under mock infection?
4. test for infection type main effects?
5. test for breed main effects?

42



43

Questions Restated in Terms of Model Parameters

1. $\delta_2-\delta_1=0$? (1 0 0 0) $\boldsymbol{\beta} = 0$?
2. $\beta=0$? (0 0 1 0) $\boldsymbol{\beta} = 0$?
3. $\alpha+\theta=0$? (0 1 0 1) $\boldsymbol{\beta} = 0$?
4. $\beta+\theta/2=0$? (0 0 1 .5) $\boldsymbol{\beta} = 0$?
5. $\alpha+\theta/2=0$? (0 1 0 .5) $\boldsymbol{\beta} = 0$?

$$\boldsymbol{\beta} = \begin{pmatrix} \delta_2-\delta_1 \\ \alpha \\ \beta \\ \theta \end{pmatrix}$$

44

Choice of Parameterization is Not Important

- The definition of the four treatment means as $\mu+\alpha+\beta+\theta$, $\mu+\beta$, $\mu+\alpha$, and μ and the choice of $\boldsymbol{\beta}=(\delta_2-\delta_1, \alpha, \beta, \theta)'$ were just convenient choices for the sake of illustration.
- Any other equivalent parameterization would lead to the same conclusions.
- It is not necessary to use an X matrix that has full column rank. Calculations could be based on comparisons of $\mathbf{g}'(X'X)^{-}\mathbf{g}$, where $(X'X)^{-}$ is an generalized inverse of X'X.

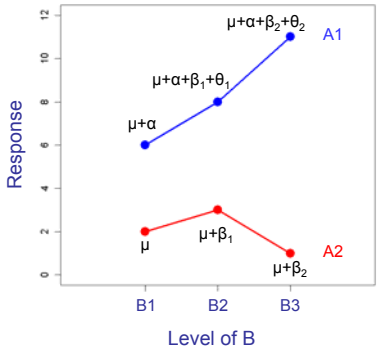
45

2 x 3 Factorial Experiment

- Suppose factor A has levels A1 and A2.
- Suppose factor B has levels B1, B2, and B3.
- Suppose we have 12 experimental units and 6 two-color microarray slides.
- What design should we use if our primary objective is to test for interaction between factors A and B?

46

Means for the 2 x 3 Factorial Experiment



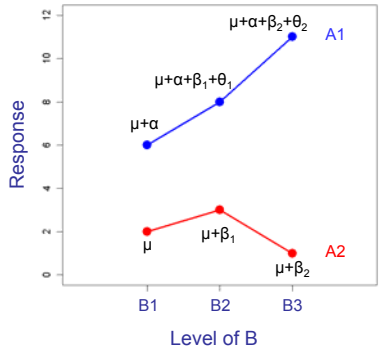
H_0 : No Interaction

H_0 : $\alpha+\theta_1=\alpha+\theta_2$

H_0 : $\theta_1=\theta_2=0$

47

Means for the 2 x 3 Factorial Experiment



H_0 : No Interaction

H_0 : $\beta_1=\beta_1+\theta_1$ and $\beta_2-\beta_1=\beta_2-\beta_1+\theta_2-\theta_1$ and $\beta_2=\beta_2+\theta_2$

H_0 : $\theta_1=\theta_2=0$

48

Multiple Regression Parameterization

$\mathbf{d} = X\boldsymbol{\beta} + \mathbf{r}$

vector of red – green differences

design matrix

parameter vector

vector of residuals

$$\boldsymbol{\beta} = \begin{pmatrix} \delta_2-\delta_1 \\ \alpha \\ \beta_1 \\ \beta_2 \\ \theta_1 \\ \theta_2 \end{pmatrix}$$

$H_0 : \theta_1=\theta_2=0$

$H_0 : M\boldsymbol{\beta}=\mathbf{0}$ where

$$M = \begin{pmatrix} 0&0&0&0&1&0 \\ 0&0&0&0&0&1 \end{pmatrix} \quad \mathbf{0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

49

$$\boldsymbol{\beta} = \begin{pmatrix} \delta_2-\delta_1 \\ \alpha \\ \beta_1 \\ \beta_2 \\ \theta_1 \\ \theta_2 \end{pmatrix}$$

$H_0 : \theta_1=\theta_2=0$

$H_0 : M\boldsymbol{\beta}=\mathbf{0}$ where

$$M = \begin{pmatrix} 0&0&0&0&1&0 \\ 0&0&0&0&0&1 \end{pmatrix} \quad \mathbf{0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

The interaction parameters θ_1 and θ_2 are estimated by $M\hat{\boldsymbol{\beta}}$.

$$\mathrm{Var}(M\hat{\boldsymbol{\beta}})=\sigma^2 M(X'X)^{-1}M'$$

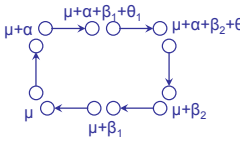
50

$$\mathrm{Var}(M\hat{\boldsymbol{\beta}})=\sigma^2 M(X'X)^{-1}M'$$

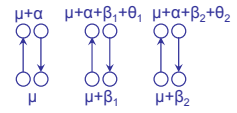
Find design X so that the determinant of $M(X'X)^{-1}M'$ is minimized.

51

Which design is preferred?



$$X_1 = \begin{pmatrix} 1&0&1&0&1&0 \\ 1&0&-1&1&-1&1 \\ 1&-1&0&0&0&-1 \\ 1&0&1&-1&0&0 \\ 1&0&-1&0&0&0 \\ 1&1&0&0&0&0 \end{pmatrix}$$

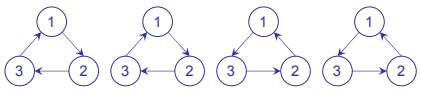


$$X_2 = \begin{pmatrix} 1&1&0&0&0&0 \\ 1&-1&0&0&0&0 \\ 1&1&0&0&1&0 \\ 1&-1&0&0&-1&0 \\ 1&1&0&0&0&1 \\ 1&-1&0&0&0&-1 \end{pmatrix}$$

$$\boldsymbol{\beta} = \begin{pmatrix} \delta_2-\delta_1 \\ \alpha \\ \beta_1 \\ \beta_2 \\ \theta_1 \\ \theta_2 \end{pmatrix}$$

52

Three Treatment CRD with Loops



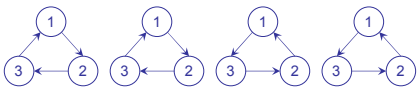
fixed

$Y = \mathrm{trt}\ \mathrm{dye}\ \mathrm{slide}\ xu$

random

53

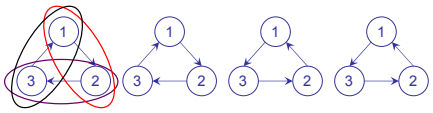
Three Treatment CRD with Loops



$$Y_{ijkl} = \mu+\tau_i+\delta_j+s_k+b_l+e_{ijkl}$$

54

Consider Red – Green Differences for Each Slide



$$\begin{aligned} Y_{2212} &= \mu+\tau_2+\delta_2+s_1+b_2+e_{2212} \\ -\,Y_{1111} &= -(\mu+\tau_1+\delta_1+s_1+b_1+e_{1111}) \end{aligned}$$

$$Y_{2212}-Y_{1111} = \delta_2-\delta_1+\tau_2-\tau_1+b_2-b_1+e_{2212}-e_{1111}$$

$$Y_{3223}-Y_{2222} = \delta_2-\delta_1+\tau_3-\tau_2+b_3-b_2+e_{3223}-e_{2222}$$

$$Y_{1131}-Y_{3233} = \delta_2-\delta_1+\tau_1-\tau_3+b_1-b_3+e_{1131}-e_{3233}$$

Variance of Red-Green Differences is Constant

$$\begin{aligned} \mathrm{Var}(Y_{2212}-Y_{1111}) &= \mathrm{Var}(\delta_2-\delta_1+\tau_2-\tau_1+b_2-b_1+e_{2212}-e_{1111}) \\ &= \mathrm{Var}(b_2-b_1+e_{2212}-e_{1111}) \\ &= \sigma_b^2+\sigma_b^2+\sigma_e^2+\sigma_e^2 \\ &= 2\sigma_b^2+2\sigma_e^2 \end{aligned}$$

This variance is the same for the difference from each slide.

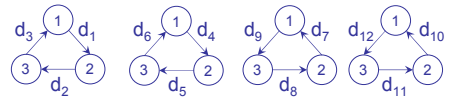
Covariance between Red-Green Differences from the Same Loop

$$\begin{aligned} &\mathrm{Cov}(Y_{2212}-Y_{1111}\,,\, Y_{3223}-Y_{2222}) \\ &\quad = \mathrm{Cov}(b_2-b_1+e_{2212}-e_{1111}\,,\, b_3-b_2+e_{3223}-e_{2222}) \\ &\quad = \mathrm{Cov}(b_2\,,\,-b_2) \\ &\quad = -\sigma_b^2 \end{aligned}$$

This covariance is the same for all pairs of differences that come from the same loop.

The covariance between differences that come from different loops is 0.

We can model the differences using a multiple linear regression model with correlated residual random effects.



$$\begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \\ d_6 \\ d_7 \\ d_8 \\ d_9 \\ d_{10} \\ d_{11} \\ d_{12} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \\ 1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \delta_2-\delta_1 \\ \tau_2-\tau_1 \\ \tau_3-\tau_2 \end{pmatrix} + \begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \\ r_6 \\ r_7 \\ r_8 \\ r_9 \\ r_{10} \\ r_{11} \\ r_{12} \end{pmatrix}$$

$$\mathbf{d} = \mathbf{X\beta} + \mathbf{r}$$

$\mathrm{Var}(\mathbf{d}) = \mathrm{Var}(\mathbf{r}) = V$ where

$$V = \begin{pmatrix} A & B & B & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ B & A & B & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ B & B & A & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & A & B & B & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & B & A & B & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & B & B & A & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & A & B & B & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & B & A & B & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & B & B & A & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & A & B & B \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & B & A & B \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & B & B & A \end{pmatrix}$$

$A = 2\sigma_b^2+2\sigma_e^2$

$B = -\sigma_b^2$

If V is known

The *best linear unbiased estimate* of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (X'V^{-1}X)^{-1}\,X'V^{-1}\mathbf{d}.$$

- $\hat{\boldsymbol{\beta}}$ is normally distributed with mean $\boldsymbol{\beta}$ and var $(X'V^{-1}X)^{-1}$
- Designs could be compared by examining $(X'V^{-1}X)^{-1}$ for various choices of X.
- Note however that the assessment of a design might depend on the variance components in V.
- In reality these variance component parameters are unknown and must be estimated from the data.