Stat 500: Formula sheet for Final, 2009

Multiple Regression: Model Selection

$$
\begin{aligned}
C_p &= \frac{SSE(Model)}{MSE(Full)} - (n-2p) \\
R^2_{adj} &= 1 - \frac{MS(Error)}{MS(Total)} = 1-(1-R^2)\frac{n-1}{n-p} \\
AIC &= n\ln(SSE/n) + 2p \\
BIC &= n\ln(SSE/n) + p\ln n
\end{aligned}
$$

RCB expt. design, 1 way trt design

Source	d.f.	sums of squares	Exp. MS
Blocks	$I-1$	$J\sum_i(\overline{Y}_{i.}-\overline{Y}_{..})^2$	
Trts	$J-1$	$I\sum_j(\overline{Y}_{.j}-\overline{Y}_{..})^2$	$\sigma^2+\frac{I}{J-1}\sum_j\tau_j^2$
Error	$(I-1)(J-1)$	$\sum_{i,j}(Y_{ij}-\hat{Y}_{ij})^2$	σ^2
Total	$IJ-1$	$\sum_{ij}(Y_{ij}-\overline{Y}_{..})^2$	

$Y_{ij}=\mu+\alpha_i+\beta_j+\epsilon_{ij},\ \ \epsilon_{ij}\sim N(0,\sigma^2)$
  Var $\overline{Y}_{.j}=\sigma^2/I$

2 way Factorial: a x b levels, n reps per trt

Source	d.f.	sums of squares
A	$a-1$	$nb\sum_i(\overline{Y}_{i..}-\overline{Y}_{...})^2$
B	$b-1$	$na\sum_j(\overline{Y}_{.j.}-\overline{Y}_{...})^2$
A*B	$(a-1)(b-1)$	$n\sum_{i,j}(\overline{Y}_{ij.}-\overline{Y}_{i..}-\overline{Y}_{.j.}+\overline{Y}_{...})^2$
Error	ab(n-1)	$\sum_{i,j,k}(Y_{ijk}-\overline{Y}_{ij.})^2$
Total	$abn-1$	$\sum_{i,j,k}(Y_{ijk}-\overline{Y}_{...})^2$

$Y_{ijk}=\mu+\alpha_i+\beta_j+\alpha\beta_{ij}+\epsilon_{ijk},\ \ \epsilon_{ijk}\sim N(0,\sigma^2)$
  Var $\overline{Y}_{i..}=\sigma^2/(bn)$
  Var $\overline{Y}_{.j.}=\sigma^2/(an)$
  Var $\overline{Y}_{ij.}=\sigma^2/(n)$

Contrasts in 2 way factorial:

Among	Param.	Est.	Variance
Cell means	$\sum_{ij}c_{ij}\mu_{ij}$	$\sum_{ij}c_{ij}\overline{Y}_{ij}$	$MSE\sum_{ij}c_{ij}^2/n$
Levels of A	$\sum_i c_i\mu_{i.}$	$\sum_i c_i\overline{Y}_{i.}$	$MSE\sum_i c_i^2/(nb)$
Levels of B	$\sum_j c_j\mu_{.j}$	$\sum_j c_j\overline{Y}_{.j}$	$MSE\sum_j c_j^2/(na)$

3 way Factorial: a x b x c levels

Source	d.f.	sums of squares
A	$a-1$	$nbc\sum_i(\overline{Y}_{i...}-\overline{Y}_{....})^2$
B	$b-1$	$nac\sum_j(\overline{Y}_{.j..}-\overline{Y}_{....})^2$
C	$c-1$	$nab\sum_k(\overline{Y}_{..k.}-\overline{Y}_{....})^2$
A*B	$(a-1)(b-1)$	$nc\sum_{i,j}(\overline{Y}_{ij..}-\overline{Y}_{i...}-\overline{Y}_{.j..}+\overline{Y}_{....})^2$
A*C	$(a-1)(c-1)$	$nb\sum_{i,k}(\overline{Y}_{i.k.}-\overline{Y}_{i...}-\overline{Y}_{..k..}+\overline{Y}_{....})^2$
B*C	$(b-1)(c-1)$	$na\sum_{j,k}(\overline{Y}_{.jk.}-\overline{Y}_{.j..}-\overline{Y}_{..k.}+\overline{Y}_{....})^2$
A*B*C	(a-1)(b-1)(c-1)	By difference
Error	abc(n-1)	$\sum_{i,j,k,l}(Y_{ijkl}-\overline{Y}_{ijk.})^2$
Total	$abcn-1$	$\sum_{i,j,k,l}(Y_{ijkl}-\overline{Y}_{...})^2$

Nested: One random effect (a levels, n reps)

Source	d.f.	sums of squares	Expected MS
A	$a-1$	$n\sum_i(\overline{Y}_{i.}-\overline{Y}_{..})^2$	$\sigma^2+n\sigma^2_\mu$
Error	$a(n-1)$	$\sum_{i,j}(Y_{ij}-\overline{Y}_{i.})^2$	σ^2
Total	$an-1$	$\sum_{i,j,k}(Y_{ijk}-\overline{Y}_{...})^2$	

$Y_{ij}=\mu+\alpha_i+\epsilon_{ij},\ \ \alpha_i\sim N(0,\sigma^2_\mu),\ \ \epsilon_{ij}\sim N(0,\sigma^2)$
  Var $\overline{Y}_{..}=\sigma^2_\mu/a+\sigma^2/(an)$

Nested: Two random effects (a levels, b levels, n reps)

Source	d.f.	sums of squares	Expected MS
A	$a-1$	$nb\sum_i(\overline{Y}_{i..}-\overline{Y}_{...})^2$	$\sigma^2+n\sigma^2_B+nb\sigma^2_A$
B	$a(b-1)$	$n\sum_{ij}(\overline{Y}_{ij.}-\overline{Y}_{i..})^2$	$\sigma^2+n\sigma^2_B$
Error	$ab(n-1)$	$\sum_{i,j,k}(Y_{ijk}-\overline{Y}_{ij.})^2$	σ^2
Total	$abn-1$	$\sum_{i,j,k}(Y_{ijk}-\overline{Y}_{...})^2$	

$Y_{ijk}=\mu+\alpha_i+\beta_{ij}+\epsilon_{ijk}$
  $\alpha_i\sim N(0,\sigma^2_A),\ \ \beta_{ij}\sim N(0,\sigma^2_B),\ \ \epsilon_{ijk}\sim N(0,\sigma^2)$
  Var $\overline{Y}_{...}=\sigma^2_A/a+\sigma^2_B/(ab)+\sigma^2/(abn)$

Subsampling

source	d.f.	sum of squares	Exp. MS
trt (A)	$a-1$	$\sum_i nb(\bar{Y}_{i..}-\bar{Y}_{...})^2$	$\sigma^2+n\sigma^2_B+\frac{bn}{a-1}\sum_i\alpha_i^2$
e.u. (B)	$a(b-1)$	$\sum_{ij}n(\bar{Y}_{ij.}-\bar{Y}_{i..})^2$	$\sigma^2+n\sigma^2_B$
Error	$ab(n-1)$	$\sum_{ijk}(Y_{ijk}-\bar{Y}_{ij.})^2$	σ^2
total	$abn-1$	$\sum_{ijk}(Y_{ijk}-\bar{Y}_{...})^2$	

$Y_{ijk}=\mu+\alpha_i+\beta_{ij}+\epsilon_{ijk},\ \ \beta_{ij}\sim N(0,\sigma^2_B),\ \ \epsilon_{ijk}\sim N(0,\sigma^2)$
  Var $\overline{Y}_{i..}=\sigma^2_B/b+\sigma^2/(bn)$

Split plot: n main plot reps, no blocks

Source	d.f.	Expected MS
A	$a-1$	$\sigma^2 + b\sigma_\rho^2 + \frac{bn}{a-1}\sum \alpha_i^2$
subjects	$a(n-1)$	$\sigma^2 + b\sigma_\rho^2$
B	$b-1$	$\sigma^2 + \frac{an}{b-1}\sum \beta_j^2$
A*B	$(a-1)(b-1)$	$\sigma^2 + \frac{n}{(a-1)(b-1)}\sum \alpha\beta_{ij}^2$
Error	$a(b-1)(n-1)$	σ^2
Total	$abn-1$	

$$Y_{ijk} = \mu + \alpha_j + \rho_{i(j)} + \beta_k + (\alpha\beta)_{jk} + \epsilon_{ijk},$$
$$\rho_{i(j)} \sim N(0, \sigma_\rho^2),\ \ \epsilon_{ijk} \sim N(0, \sigma^2)$$
$$\text{Var } \overline{Y}.j. = \sigma_\rho^2/n + \sigma^2/(bn)$$
$$\text{Var } \overline{Y}..k = \sigma_\rho^2/(an) + \sigma^2/(an)$$
$$\text{Var } \overline{Y}.jk = \sigma_\rho^2/n + \sigma^2/n$$

Contrasts in split-plot:

Among	Variance
Levels of A	$MSsubjects\, \Sigma_j c_j^2/(nb)$
Levels of B	$MSE\, \Sigma_k c_k^2/(na)$
Split, same main	$MSE\, \Sigma_{jk} c_{jk}^2/n$
Main, same split	$[(b-1)\, MSE + MSpair]\, \Sigma_{jk} c_{ij}^2/(nb)$
Diff main,split	$[(b-1)\, MSE + MSpair]\, \Sigma_{ij} c_{ij}^2/(nb)$

Categorical data

$$\begin{aligned}
\text{If } Y &\sim Bin(n,\pi),\\
P[Y=k] &= \frac{n!}{k!(n-k)!}\pi^k(1-\pi)^{(n-k)}\\
p &= Y/n\\
p &\pm z_{1-\alpha/2}\sqrt{p(1-p)/n}\\
n &= \frac{z_{1-\alpha/2}^2\pi(1-\pi)}{\delta^2}\\
z &= \frac{p-\pi_0}{\sqrt{\pi_0(1-\pi_0)/n}}\\
n &= \frac{(z_{1-\alpha}\sqrt{\pi_0(1-\pi_0)} + z_{1-\beta}\sqrt{\pi_a(1-\pi_a)})^2}{(\pi_0-\pi_a)^2}\\
p_1 - p_2 &\pm z_{1-\alpha/2}\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}\\
p &= (Y_1+Y_2)/(n_1+n_2)\\
z &= \frac{p_1-p_2}{\sqrt{p(1-p)(1/n_1+1/n_2)}}\\
n &= \Big(z_{1-\alpha/2}\sqrt{2\pi_0(1-\pi_0)}\\
&+ z_{1-\beta}\sqrt{\pi_1(1-\pi_1)+\pi_2(1-\pi_2)}\Big)^2/(\pi_1-\pi_2)^2\\
\chi^2 &= \sum_{i,j}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}\\
E_{ij} &= R_iC_j/N\\
\chi^2 &\sim \chi^2_{(R-1)(C-1)}\\
P[N_{11}=n_{11}] &= \frac{n_{1+}!n_{2+}!n_{+1}!n_{+2}!}{N!n_{11}!n_{12}!n_{21}!n_{22}!}\\
OR &= \phi = \frac{\pi_2(1-\pi_1)}{\pi_1(1-\pi_2)}\\
\widehat{OR} &= \frac{p_2(1-p_1)}{p_1(1-p_2)} = \frac{y_2(n_1-y_1)}{y_1(n_2-y_2)}\\
\text{Var } \ln OR &\approx \frac{1}{y_1} + \frac{1}{n_1-y_1} + \frac{1}{y_2} + \frac{1}{n_2-y_2}\\
\ln\widehat{OR} &\sim N\left(\ln\phi, \frac{1}{n_1\pi_1(1-pi_1)} + \frac{1}{n_2\pi_2(1-pi_2)}\right)\\
\text{Var } p_{1+} - p_{+1} &= \frac{1}{n}\,[p_{1+}(1-p_{1+}) + p_{+1}(1-p_{+1})\\
&\quad -2(p_{11}p_{22} - p_{12}p_{21})]\\
z &= (n_{12}-n_{21})/\sqrt{n_{12}+n_{21}}\\
n_{12} &\sim Bin(n_{12}+n_{21}, 0.5)
\end{aligned}$$