11} variable selection is a form of data snooping, so "inference is not valid after selecting a model using variable selection." Can use the model for description.

If possible do a pilot study with variable selection, then do a bigger study using the model selected in the pilot study as the candidate full model. Inference from the bigger study can be done.

12} Often building a good full model also uses data snooping, then inference is no longer valid.

13} p 392-3
"For a given data set, may have to use data snooping to find a model that fits the data reasonably well. The fitted model tends to fit the training data $(Y_1, \underline{x}_1), \ldots, (Y_n, \underline{x}_n)$ better than future test data, and "inference can not be justified by large sample theory."

(don't look at several candidate models)

14) If the variable selection is done by computer, eg use I_I or I_{min}, and data snooping was not used to build the full model, it is be possible to develop valid inference for the selected model. It is known that I_{min} overfits, asymptotically, see p 448.

Skim §12.2

Ignore cross validation on p403-6, §12.3.3-4

Skim § 12.4.3 on stepwise selection:
Terms that were deleted can come back in the model.

§12.5 15) Suppose $\underline{Y} = \underline{X}\underline{\beta} + \underline{\varepsilon}$ and $p > \frac{n}{5}$ or there is collinearity, Shrinkage methods shrink some $|\hat{\beta}_i|$ values towards 0.

ex) Forward selection and backwards elimination make $\hat{\beta}_i = 0$ for the omitted variables.

Skip §12.5.1

16) p423 The ridge regression estimator

is $\hat{\underline{\beta}}(\lambda) = \underbrace{(\underline{X}'\underline{X} + \lambda \underline{I})^{-1}}_{p \times p} \underline{X}'\underline{Y}$, $\lambda \geq 0$

nonsingular even if $\underline{X}'\underline{X}$ is singular for $\lambda > 0$

$\hat{\underline{\beta}}(\lambda)$ minimizes the criterion

$$RSS(\lambda) = (\underline{Y} - \underline{X}\underline{\beta})^T(\underline{Y} - \underline{X}\underline{\beta}) + \lambda \underline{\beta}_S^T \underline{\beta}_S$$

$$= RSS(\underline{\beta}) + \lambda \underline{\beta}_S^T \underline{\beta}_S = RSS(\underline{\beta}) + \lambda \sum_{i=1}^{p-1} \beta_i^2$$

where $\underline{\beta} = \begin{pmatrix} \beta_0 \\ \underline{\beta}_S \end{pmatrix}$. This estimator can be fit even if $p \geq n$. Typically use

$\hat{\underline{\beta}}(\hat{\lambda})$ where λ is estimated by $\hat{\lambda}$, (GCV CV etc)

(Let $\underline{w} = \underline{X}\underline{x}$ then $\underline{x}'\underline{X}'\underline{X}\underline{x} = \underline{w}'\underline{w} = \sum_{i=1}^{p} w_i^2 \geq 0$ so $\underline{X}'\underline{X} \geq 0$

so $\underline{x}'(\underline{X}'\underline{X} + \lambda \underline{I})\underline{x} = \underline{x}'\underline{X}'\underline{X}\underline{x} + \lambda \underbrace{\underline{x}'\underline{x}}_{\sum_{i=1}^{p} x_i^2 > 0 \text{ unless } \underline{x} = \underline{0}.} > 0$ for $\lambda > 0$ unless $\underline{x} = \underline{0}$. So $\underline{X}'\underline{X} + \lambda \underline{I} > 0$ for $\lambda > 0$)

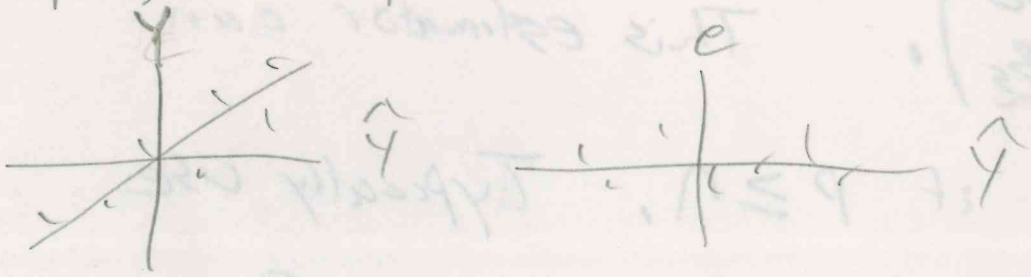
Skip §12.5.3, 12.6

Skim §12.7,

Skip §12.8, 12.9 except Skim §12.9.1

Ch 8 1] p187 In experimental design models or design of experiments DOE, the entries of $\underline{X}$ are coded, often $-1, 0, 1$. Often the $\underline{X}$ matrix is not a full rank matrix.

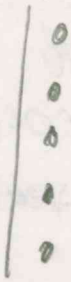
2] Some DOE models, like the 2^k factorial model, have one Y_i per $\underline{x}_i$, lots of $\underline{x}_i$'s, and the response and residual plots look like those for MLR.



3] Some DOE models have n_i Y_i's for each distinct $\underline{x}_i$. Then the response and residual plots no longer look like those for MLR.

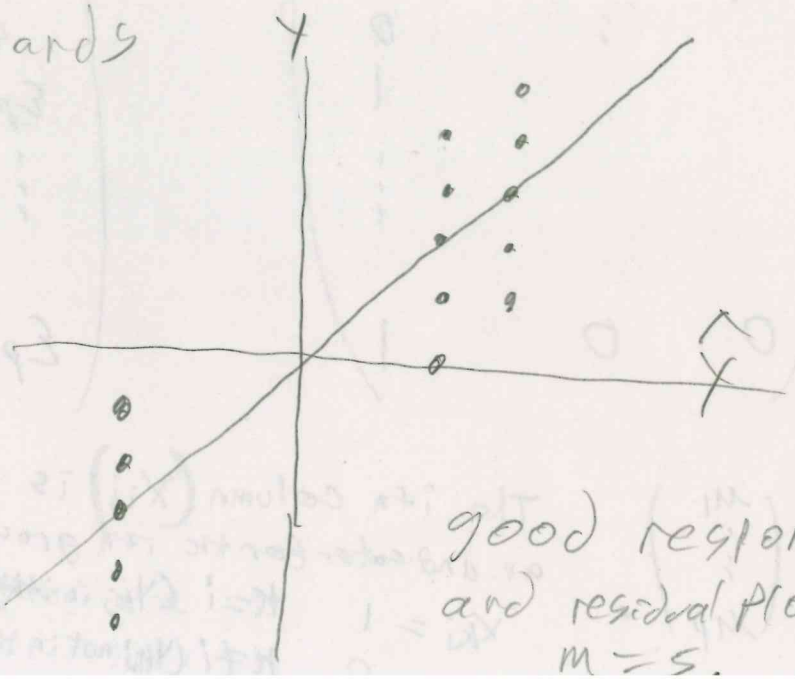
4) Suppose there are p distinct $\underline{x}_i$, called treatments, $n = n_1 + \cdots + n_p$, and $n_i \equiv m = \frac{n}{p}$. A dot plot of $Z_1, \ldots, Z_m$ consists of

an axis and
m points corresponding to $z_1, \ldots, z_m$.

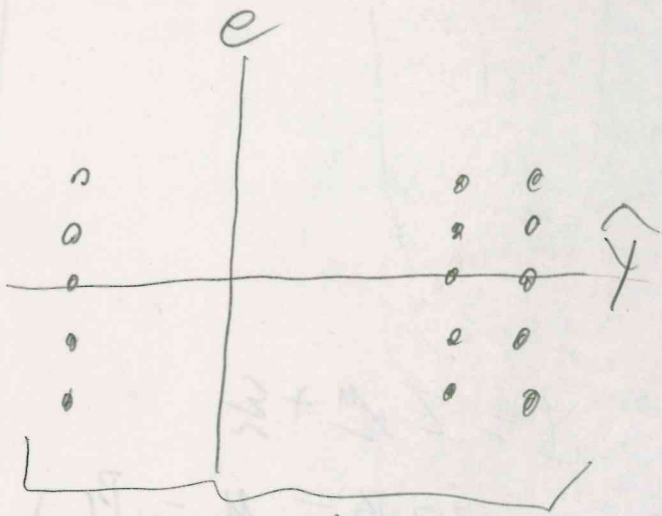


If $m \geq 5$ and p is small,
the response plot consists of
p dot plots, one for each treatment.
If $m \geq 10$, the p dot plots should
have roughly the same shape and spread.

The residual plot also consists of
p dot plots. The points in the response
and residual plots should scatter about
the identity and $e = 0$ lines, but the
scatter need not be in evenly populated
bands



good response and residual plots with $m = 5$.



3 dot plots

§8.2 One way ANOVA model

p188

					Sample mean
group or pop	1	Y_{11}	Y_{12}	Y_{1n_1}	$\bar{Y}_{1\cdot}$
or treatment	2	Y_{21}	Y_{22}	Y_{2n_2}	$\bar{Y}_{2\cdot}$
	⋮				
	p	Y_{p1}	Y_{p2}	Y_{pn_p}	$\bar{Y}_{p\cdot}$

Cell means model

$$Y_{ij} = \mu_i + \varepsilon_{ij} \qquad i = 1,\dots,p;\ j = 1,\dots,n_i$$

$$E(\varepsilon_{ij}) = 0 \qquad V(\varepsilon_{ij}) = \sigma^2, \ \varepsilon_{ij} \text{ are iid.}$$

$$\underset{\sim}{Y} = \begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ Y_{21} \\ \vdots \\ Y_{2n_2} \\ \vdots \\ Y_{p1} \\ \vdots \\ Y_{pn_p} \end{pmatrix} = \underset{=}{X}\underset{\sim}{\mu} + \underset{\sim}{\varepsilon} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & & 0 \\ 0 & 1 & & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 1 & & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & 1 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1n_1} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{2n_2} \\ \vdots \\ \varepsilon_{p1} \\ \vdots \\ \varepsilon_{pn_p} \end{pmatrix}$$

$$\underline{Y} = X\underline{\mu} + \underline{\varepsilon}$$

$$\text{So } \beta = \underline{\mu} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix}.$$

The ith column (X_{ij}) is an indicator for the ith group:

$X_{kj} = 1$ if $k = i$ (Y_{kj} in ith group)

0 if $k \neq i$ (Y_{kj} not in ith group)

6) p189 For the cell means model,

$\underset{\sim}{X}$ is full rank, $\underset{\sim}{1}$ is not a column of $\underset{\sim}{X}$, but $\underset{\sim}{1} \in C(\underset{\sim}{X})$

Since if $\underset{\sim}{X} = (\underset{\sim}{v}_1, \ldots, \underset{\sim}{v}_p)$, then $\underset{\sim}{1} = \sum_{i=1}^{p} \underset{\sim}{v}_i$.

$$\underset{p \times n \; n \times p}{\underset{\sim}{X}'\underset{\sim}{X}} = \begin{pmatrix} 1 & \cdots & 1 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 1 & \cdots & 1 \end{pmatrix} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & & \\ 0 & 1 & & \\ \vdots & \vdots & & \\ 0 & 1 & & \\ \vdots & \vdots & & 0 \\ 0 & 0 & & 1 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & 1 \end{bmatrix}$$

$$= \begin{pmatrix} n_1 & 0 & \cdots & 0 \\ 0 & n_2 & 0 & \cdots & 0 \\ \vdots & & & & \\ 0 & \cdots & & 0 & n_p \end{pmatrix}.$$

$$(\underset{\sim}{X}'\underset{\sim}{X})^{-1} = \operatorname{diag}\left(\frac{1}{n_1}, \ldots, \frac{1}{n_p}\right).$$

$$\underset{\sim}{X}'\underset{\sim}{Y} = \left(\sum_{j=1}^{n_1} Y_{1j} \quad \sum_{j=1}^{n_2} Y_{2j} \quad \cdots \quad \sum_{j=1}^{n_p} Y_{pj}\right)' = \left(Y_{10} \; Y_{20} \cdots Y_{p0}\right)^T$$

$$\text{So } \underset{\sim}{\hat{\mu}} = \underset{\sim}{\hat{\beta}} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{Y} = \operatorname{diag}\left(\frac{1}{n_1}, \ldots, \frac{1}{n_p}\right)\begin{pmatrix} Y_{10} \\ \vdots \\ Y_{p0} \end{pmatrix} = \begin{pmatrix} \bar{Y}_{10} \\ \vdots \\ \bar{Y}_{p0} \end{pmatrix}$$

$$\underline{\hat{Y}} = X(X'X)^{-1}X'\underline{Y} = X\underline{\hat{\mu}} =$$

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ \vdots & & & \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & & \\ 0 & 1 & \cdots & 0 \\ \vdots & & & \\ 0 & \cdots & 0 & 1 \\ \vdots & & & \\ 0 & \cdots & 0 & 1 \end{pmatrix} \begin{pmatrix} \bar{Y}_{10} \\ \bar{Y}_{20} \\ \vdots \\ \bar{Y}_{p0} \end{pmatrix} = \begin{pmatrix} \bar{Y}_{10} \\ \vdots \\ \bar{Y}_{10} \\ \bar{Y}_{20} \\ \vdots \\ \bar{Y}_{20} \\ \vdots \\ \bar{Y}_{p0} \\ \vdots \\ \bar{Y}_{p0} \end{pmatrix}$$

So $\hat{Y}_{ij} = \bar{Y}_{i0}, \quad i = 1, \ldots, p, \; j = 1, \ldots, n_i.$

7) Hence the dot plot for the jth treatment crosses the identity line at $\bar{Y}_{j0}$ in the response plot.