experimental if it was randomly assigned to a treatment, and observational, otherwise.

3] For linear models $y = \underset{\sim}{x}^T\underset{\sim}{\beta} + e$, the pdf of e is $f(y)$, and the pdf of $y|\underset{\sim}{x}^T\underset{\sim}{\beta}$ is $f(y - \underset{\sim}{x}^T\underset{\sim}{\beta})$.

location family with $V(Y|\underset{\sim}{x}^T\underset{\sim}{\beta}) \equiv \sigma^2$

MLR and DOE are linear models.

The constant variance assumption is used.

4] In an experiment, the researcher randomly assigns n_i units to each treatment for $i = 1, \ldots, p$ (so p treatments).

$n = n_1 + \cdots + n_p$.

Let $n_i \equiv m = \frac{n}{p}$. If $p > 8$ and $m = 1$ then the response and residual plots for DOE are often interpreted just like those for MLR.

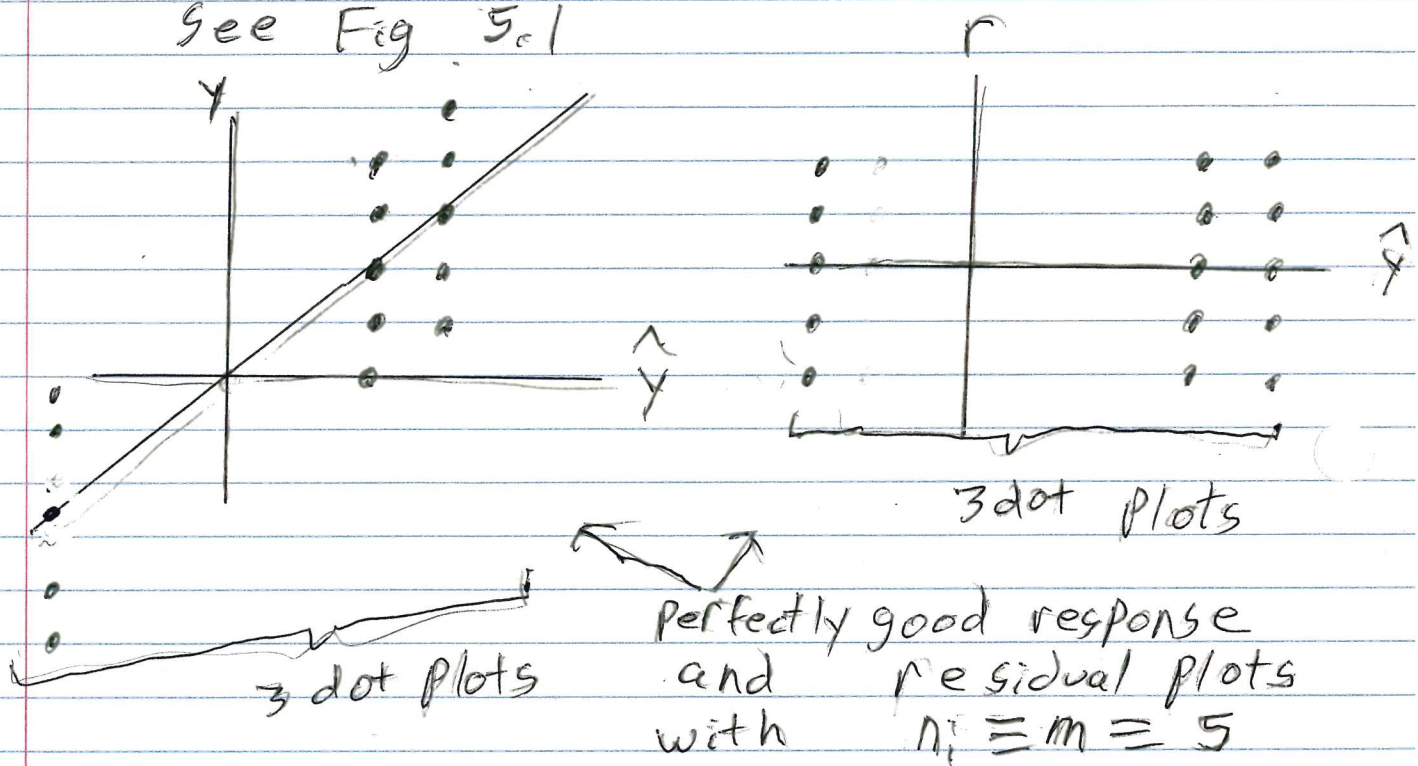
5] * If $m \geq 5$ and p small, the interpretation is different. A dot plot of $z_1, \ldots, z_m$ consists of an axis and m points corresponding to $z_1, \ldots, z_m$.

The response plot consists of p dot plots, one for each treatment. If $m \geq 10$

then 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 scatter about the identity and $r=0$ lines, but the scatter need not be in evenly populated bands.

See Fig 5.1



6] Know for final p183 4 step fixed effects one way Anova F test

Source	df	SS	MS	F	Pval for
Treatment	$p-1$	SSTR	MSTR	$F_0 = \frac{MSTR}{MSE}$	$H_0: \mu_1 = \cdots = \mu_p$
Error	$n-p$	SSE	MSE		

i) $H_0: \mu_1 = \mu_2 = \cdots = \mu_p$ H_A not H_0

ii) $F_0 = \frac{MSTR}{MSE}$

iii) $pval = P(F_{p-1, n-p} > F_0)$

} usually from output

iv) If $pval < \delta$, reject H_0. Conclude the mean response depends on the factor level. If $pval \geq \delta$, fail to reject H_0, conclude mean response does not depend on factor level

Write the conclusion as a nontechnical sentence.

ex] Phlebites is an ear infection that can occur when giving shots. Each rabbit got a shot in one ear. y = temp difference for treated ear − untreated ear.

Amiodarone + vehicle solution to carry drug	vehicle solution only	saline solution = placebo
2.2	0.3	0.1
1.6	0.0	0.1
0.8	0.6	0.2
1.8	0.0	−0.4
1.4	−0.3	0.3
0.4	0.2	0.1
0.6		0.1
1.5		−0.5
0.5		
y_{i0} 10.80	0.80	0.00
$\bar{y}_{i0}$ 1.20	0.13	0.00

ANOVA

Source	df	SS	MS	F	Pval
Trt	2	7.2162	3.6081	16.58	0.000
error	20	4.3533	0.2177		
Total	22	11.5696			

perform the Anova F test

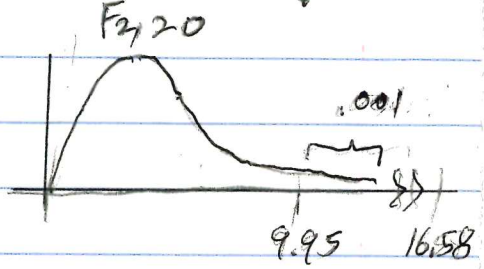
i) $H_0: \mu_1 = \mu_2 = \mu_3$ H_A not H_0

ii) $F_0 = 16.58$

iii) pval $= 0$

iv) reject H_0, the mean temp differences depend on the shot.

Also see ex 5.5 b)



7] Know p 178 The cell means model

for the fixed effects one way Anova is $Y_{ij} = \mu_i + e_{ij}$ where Y_{ij} is the value of the response variable for the jth trial of the ith treatment.

$E[e_{ij}] = 0$ $\quad V[e_{ij}] \equiv \sigma^2$

$i = 1, \ldots, p$ $\quad j = 1, \ldots, n_i$.

The e_{ij} are iid from a location family with pdf $f_Z(z)$. For the normal cell means model, the e_{ij} are iid $N(0, \sigma^2)$ for $i = 1, \ldots, p$ and $j = 1, \ldots, n_i$. p.178 writes the cell means model as a linear model.

8] Know $\hat{Y}_{ij} = \hat{\mu}_i = \overline{Y}_{i\cdot} = \frac{Y_{i\cdot}}{n_i}$

dot means sum over j

$= \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}$. The "dot notation"

means sum over the subscript corresponding to 0, eg j. So $Y_{00} = \sum_{i=1}^{p} \sum_{j=1}^{n_i} Y_{ij}$.

is the sum of all of the Y_{ij} = grand total for the response.

$\bar{Y}_{i0} = \frac{Y_{i0}}{n_i}$ = sample mean for ith factor level.

$\bar{Y}_{00} = \frac{Y_{00}}{n} = \frac{\sum_{i=1}^{p} \sum_{j=1}^{n_i} Y_{ij}}{n}$ = sample mean of all Y_{ij}.

$n_1 + \cdots + n_p = n$.

ex) see ex5a)

ex] In last ex, find $\hat{\mu}_i$.

Soln $\hat{\mu}_1 = 1.2$, $\hat{\mu}_2 = 0.13$, $\hat{\mu}_3 = 0.0$

For ex, $\hat{\mu}_2 = \frac{Y_{20}}{n_2} = \frac{0.80}{6} =$

$\frac{0.3 + 0.0 + 0.6 + 0.0 - 0.3 + 0.2}{6} = \bar{Y}_{20} = 0.13$

9] Anova Tables Know

i) me source df ss ms F pvalue
treatment
error

ii) R source DF Sumsq meansq Fvalue Pr(>F)
factor
residual

iii) SAS source DF sum of squares mean square F value Pr>F
model
error

iv) Arc Source df SS MS F P-value
regression
residual

v) Minitab Source DF SS MS F P
design
error

10] Know for E3, final using R to randomly assign treatments to units (randomization)
see ex 5.1.

ex] 1 Haneet, 2 Josh, 3 Ward, 4 Daniel, 5 Brian
YU 6, Damon 7, Gerard 8, Kursad 9

assign numbers in order listed

Assign the above people to 3 groups using R output
sample(9)

[1] 9 2 5 8 7 3 4 1 6
(1st 3, 2nd 3, 3rd 3)

trt

A		B		C	
9	Kursad	8	Gerard	4	Daniel
2	Josh	7	Damon	1	Haneet
5	Brian	3	Ward	6	YU

11] * p176 In an observational study, (the individuals decide which "trt" to take), the researchers simply observe a response,

12) Know Randomization of units into 2 or more treatment groups is an extremely important step for an experiment. Each group is given a different trt. Use random numbers (Sample in R) to decide which unit gets which trt.

13) Randomization i) makes the sample means $\bar{Y}_{10}, \ldots, \bar{Y}_{p0}$ such that inference like F tests, t tests and confidence intervals can be used,
ii) makes the p groups "as alike as possible" except for the different treatments
iii) minimizes the effects of lurking variables (a variable not included in the study that may affect the relationship among the variables in the study).
iv) Helps prevent biases of the experimenter from influencing the outcome of the study.

14) Know The purpose of an experiment is to study the changes in the response variable as the values of the explanatory variable(s) change: study $Y|\underset{\sim}{x}$.

15) P176 Replication means that for each trt,

50,5

$Y_{i1}, \ldots, Y_{in_i}$ are approximately iid.

Read ex 5.2 on p 177. ← I should do this

16] * p 177 The one way fixed effects Anova model has one qualitative predictor variable (factor) with p categories $a_1, \ldots, a_p$. $Y | (W = a_i) \sim f_Z(y - \mu_i)$ where the location family has 2nd moments $E(Y | W = a_i) = \mu_i$ and $V(Y | W = a_i) = \sigma^2$. All p distributions come from the same location family with different location parameter μ_i and the same variance σ^2. The p pdfs have the same shape and spread. $Y | (W = a_i) \sim N(\mu_i, \sigma^2)$ is the one way fixed effects normal model.

17] * p 179-180 For one way Anova, the response and residual plots consist of p dot plots. They should have roughly the same shape and spread. If the identity line is clearly superior to any horiz line, then the means differ. If the response plot looks like the residual plot, then there is not much difference in the means.

for output pvalues to be ≈ correct

18] Outliers can have large vertical distances from the identity or $r = 0$ line. But large outliers often have small residuals. Look for a large gap between the bulk of the data and a large outlier.