STATISTICS 642 - ASSIGNMENT #7 Solution

Problem I (15 points) : For a 2^{n-p} fractional factorial design we have the following results:

a. The fraction of the full design is 2^{-p}.

b. There are p generators.

c. The number of Generalized Interactions (Implicit Generators) equals the number of newly created contrasts by given p generators. So, the number of generalized interactions is $\sum_{k=2}^{p}\binom{p}{k} = 2^p - p - 1$ for $p \geq 2$ and 0 for $p = 1$.

d. The number of effects in each of the $2^{n-p}-1$ alias sets is given by $1+p+(2^p-p-1) = 2^p$. Therefore, each effect has $2^p - 1$ aliases, that is, effect is confounded with $2^p - 1$ effects.

e. The number of experimental units required to conduct the experiment is 2^{n-p} .

Design	a.	b.	c.	d.	e.
2^{6-2}	$\frac{1}{4}$	2	1	3	16
2^{7-3}	$\frac{1}{8}$	3	4	7	16
2^{7-4}	$\frac{1}{16}$	4	11	15	8

Problem II (15 points)

Using the SAS progran fractional,main.sas, a 1/8 fraction of a 2^8 factorial experiment with resolution IV is obtained using the +1 levels of $ABFG$, $ACFH$ and $ABCDEF$, the 3 generators.

TRT	A	B	C	D	E	F	G	H	$ABFG$	$ACFH$	$ABCDEF$
(I)	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1
efgh	-1	-1	-1	-1	1	1	1	1	1	1	1
dfgh	-1	-1	-1	1	-1	1	1	1	1	1	1
de	-1	-1	-1	1	1	-1	-1	-1	1	1	1
cfg	-1	-1	1	-1	-1	1	1	-1	1	1	1
ceh	-1	-1	1	-1	1	-1	-1	1	1	1	1
cdh	-1	-1	1	1	-1	-1	-1	1	1	1	1
cdefg	-1	-1	1	1	1	1	1	-1	1	1	1
bfh	-1	1	-1	-1	-1	1	-1	1	1	1	1
beg	-1	1	-1	-1	1	-1	1	-1	1	1	1
bdg	-1	1	-1	1	-1	-1	1	-1	1	1	1
bdeh	-1	1	-1	1	1	1	-1	1	1	1	1
bch	-1	1	1	-1	-1	-1	1	1	1	1	1
bcef	-1	1	1	-1	1	1	-1	-1	1	1	1
bcdf	-1	1	1	1	-1	1	-1	-1	1	1	1
bcdegh	-1	1	1	1	1	-1	1	1	1	1	1
af	1	-1	-1	-1	-1	1	-1	-1	1	1	1
aegh	1	-1	-1	-1	1	-1	1	1	1	1	1
adgh	1	-1	-1	1	-1	-1	1	1	1	1	1
adef	1	-1	-1	1	1	1	-1	-1	1	1	1
acg	1	-1	1	-1	-1	-1	1	-1	1	1	1
acefh	1	-1	1	-1	1	1	-1	1	1	1	1
acdfh	1	-1	1	1	-1	1	-1	1	1	1	1
acdeg	1	-1	1	1	1	-1	1	-1	1	1	1
abh	1	1	-1	-1	-1	-1	-1	1	1	1	1
abefg	1	1	-1	-1	1	1	1	-1	1	1	1
abdfg	1	1	-1	1	-1	1	1	-1	1	1	1
abdeh	1	1	-1	1	1	-1	-1	1	1	1	1
abcfgh	1	1	1	-1	-1	1	1	1	1	1	1
abce	1	1	1	-1	1	-1	-1	-1	1	1	1
abcd	1	1	1	1	-1	-1	-1	-1	1	1	1
abcdefgh	1	1	1	1	1	1	1	1	1	1	1

Problem III (15 points) Problem 12.6:

a. $2^{-2} = 1/4$ fraction of the possible 32 ($2^5 = 32$) treatments or 8 treatments will be included in the experiment for both sets of generators

b. The experimenter should use $I = ABCD = BCE$, Design i, because this will yield the implied generator $I = ADE$. The minimum length of the two generators and the implied generator is 3 factors so the design resolution will be 3 (RES.=III).

   If the experimenter selected generators $I = ABCDE = ABCD$ then the implied generator would be $I = E$. Thus, the main effect of factor E would be non-estimable, a not very useful design.

Problem IV (15 points) Problem 12.11:

a. The design has resolution = V because we have a 2^{5-1} factorial design ($16 = 2^4 = 2^{5-1}$) with generator $I = ABCDE$. We know this is the generator because the 16 selected treatments all have $ABCDE = +1$.

b. E=ABCD, that is, the value for E in any given row is obtained by multiplying the values of A, B, C, and D in that row.

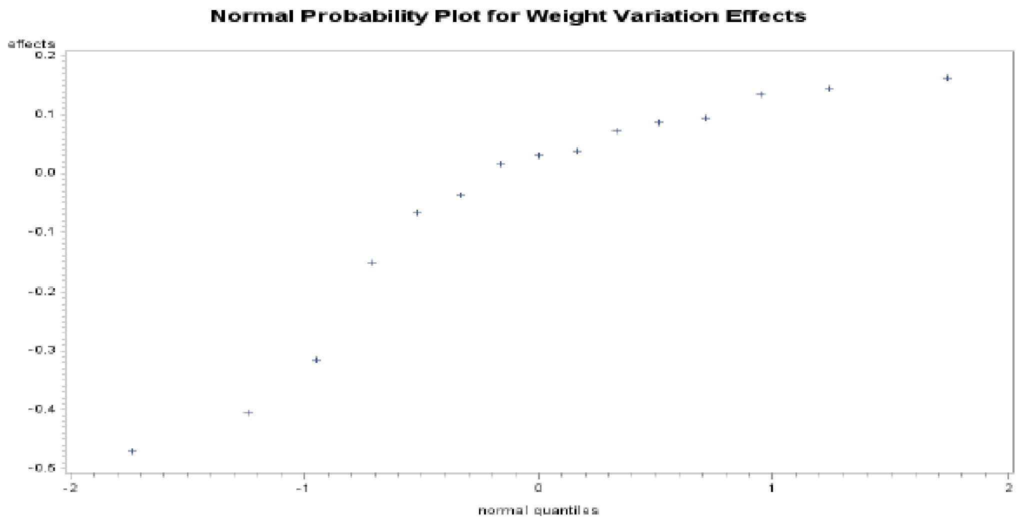
c. The alias structure is identical to the alias structure in Table 4 on page 13 in Handout 10.

d. Since main effects are confounded with four-factor interactions and two-factor interactions are confounded with three-factor interactions, it is necessary to assume that three-factor or higher interactions are negligible in order to be able to estimate main effects and two-factor interactions free of any other effects.

e. The standard errors cannot be estimated because there were no df for error. The estimates are given below:

A(.145)	B(.0875)	C(.0375)	D(-.0375)	E(-.47)
AB(.015)	AC(.095)	AD(.03)	AE(-.1525)	BC(-.0675)
BD(.1625)	BE(-.405)	CD(.0725)	CE(.135)	DE(-.315)

f. From normal probability plot, it would appear that E, BE, and DE are effects which may impact the weight variation.



g. From the 16 runs shown in the given design, a 1/4 fraction of a 2^5 factorial experiment with defining equations $I = -ABC = -ABCDE$ can be obtained by taking those treatments having $ABC = -1$

Run	A	B	C	D	E	ABC	y
12	-1	-1	-1	-1	-1	-1	1.13
1	-1	-1	-1	1	1	-1	0.78
16	-1	1	1	-1	-1	-1	1.18
9	-1	1	1	1	1	-1	0.76
4	1	-1	1	-1	-1	-1	1.28
2	1	-1	1	1	1	-1	1.10
3	1	1	-1	-1	-1	-1	1.70
10	1	1	-1	1	1	-1	0.62

- However, note that the treatments listed above constitute a 2^{5-2} design (quarter fraction of the complete 2^5 design) with defining generators $I = ABCDE = ABC$ and implied generator $I = DE$. Thus the above design has resolution 2 (RES=II) and the main effects of factors D and E confounded. The researcher would have better off to start over and use the generators $I = ABD = CDE$ with implied generator $I = ABCE$. The resolution would then be 3 (RES=III).

V. (20 points) **Problem 16.2 from Textbook**

a. There is some balance in the experiment in that each treatment follows every other treatment one time in the sequences; and each treatment is observed once in each time period and once on each subject.

b. Model: Let y_{ijk} be the systolic blood pressure of subject j in sequence i recorded during period k

$y_{ijk} = \mu + \alpha_i + b_{j(i)} + \gamma_k + \tau_{d(i,k)} + \lambda_{c(i,k-1)} + e_{ijk}$ with

α_i the fixed effect of sequence i with $\alpha_6 = 0$; γ_k the fixed effect of time period k with $\gamma_3 = 0$

$b_{j(i)}$ the random effect of subject j nested within sequence i having iid $N(o, \sigma_b^2)$ distributions;

$\tau_{d(i,k)}$ is the fixed direct effect of treatment d administered in period k of sequence i;

$\lambda_{c(i,k-1)}$ is the fixed carryover effect of treatment d administered in period $k-1$ of sequence i;

e_{ijk} the random effect of subject j nested within sequence i during period k having correlated $N(o, \sigma_e^2)$ distributions with the correlations satisfying the H-F condition;

c. S=SEQUENCE; SUB=SUBJECT; P=PERIOD; D=DRUG; C=CARRYOVER

Source	DF	Type I SS	Mean Square	F Value	Pr > F
S	5	3888.90278	777.78056	2.78	0.0294
SUB(S)	18	27066.41667	1503.68981	5.38	<.0001
P	2	1189.19444	594.59722	2.13	0.1319
D	2	7050.98779	3525.49389	12.61	<.0001
C	2	1167.08357	583.54178	2.09	0.1367
Error	42	11742.06753	279.57304		

There is not significant evidence (p-value=.1367) of a carryover effect and hence we can test for drug differences using all 3 periods data. There is significant (p-value< .0001) evidence of a difference in the 3 treatments with respect to their average systolic blood pressure in the subjects.

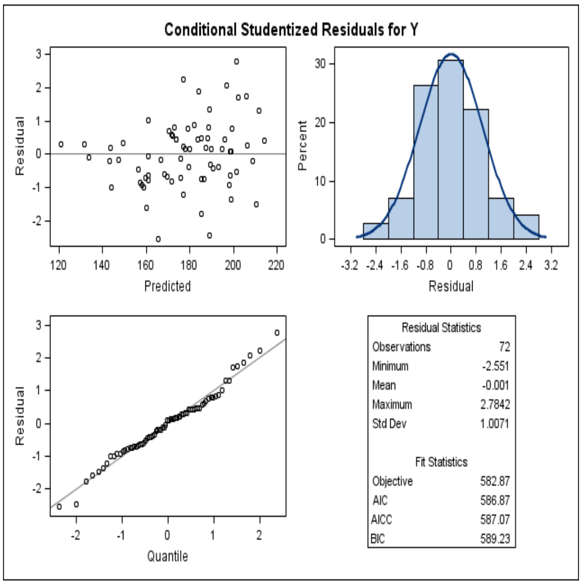
d. The contrasts comparing Treatment C to Drug A and Drug B are given below:

Contrasts

Label	Num DF	Den DF	F Value	Pr > F
DRUG C VS A+B	1	44	18.77	<.0001
A VS B	1	44	5.62	0.0222

Based on the above analysis, there is significant evidence that Drug C is different from the average response of Drugs A and B. Also, there is significant evidence that the means for Drugs A and B are different.

e. Based on the following residuals, there appears to be a slight increase in the variance as the mean response increases. There does not appear to be a violation of the normality condition:



VI. (20points) **Problem 17.1 from Textbook:**

a. model: $Y_{ij} = \mu + \beta_i(X_{ij} - \bar{X}_{..}) + \tau_i + e_{ij},\ i = 1,2,3,\ j = 1,\cdots,6,$ where τ_i is fixed effect of ith Alloy, X_{ij} is weld diameter, β_i is the slope for linear regression of weld strength on weld diameter for Alloy i, and e_{ij} is experimental error. We assume that 1) alloys do not affect the value of the weld diameter X_{ij}, 2) the weld diameters are measured without error, and 3) $e_{ij} \sim iid\ N(0, \sigma_e^2)$.

b. (1) ANOVA tables:

i. Let $I_{1i} = 1$ if Weld i was on Alloy1 , 0 if Weld i was on Alloy2 or Alloy3
Let $I_{2i} = 1$ if Weld i was on Alloy2 , 0 if Weld i was on Alloy1 or Alloy3
Fitting model I: $Y_i = \beta_0 + \beta_1 I_{1i} + \beta_2 I_{2i} + \beta_3 X_i + \beta_4 I_{1i} X_i + \beta_5 I_{2i} X_i + e_i$.

Source	df	SS	MS	F	$Pr > F$
T	2	122.257995	61.128997	1.60	0.2417
X	1	2213.861	2213.861	58.02	$< .0001$
X*T	2	258.265	129.132	3.38	0.0683
Error	12	457.899	38.158		

(2) Tests:

i. From the SAS output, the slopes for the three Alloys are
$\beta_{A1} = 4.1892 - .2692 = 3.92,\ \beta_{A2} = 4.1892 + 3.1614 = 7.3506,\ \beta_{A3} = 4.1892$
H_0 : Alloys have common Slopes vs H_1 : Alloys have different Slopes.
$\Leftrightarrow H_0 : \beta_4 = \beta_5 = 0$ vs $H_1 : \beta_4 \neq 0$ and/or $\beta_5 \neq 0$:
$\Leftrightarrow H_0 : Q_{X*T} = 0$ vs $H_1 : Q_{X*T} \neq 0$:
The test statistic is $F = \frac{129.132}{38.158} = 3.38$ with $df = 2, 12 \Rightarrow p - value = 0.0683 < 0.25$ (See the discussion in part h. of this problem)
Therefore, reject H_0 and we conclude there is significant evidence of a difference in the treatment slopes.

ii. Based on our conclusion in part i. we conclude that the covariate is significant.

iii. Based on the conclusion that the slopes are different for the three alloys, the assessment of the difference in the adjusted Alloy means must be done separately at specified values of the Weld Diameters. A global assessment of differences is not possible. See the results in part c.

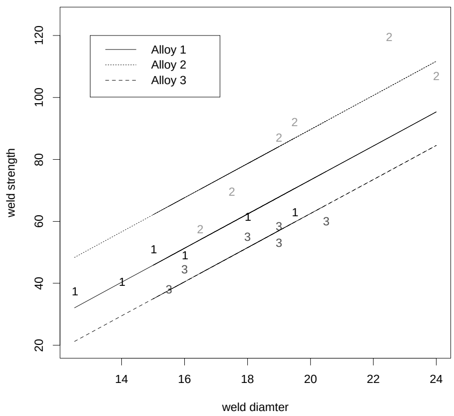
c. Adjusted treatment means using a model with unequal slopes at specified values of X.

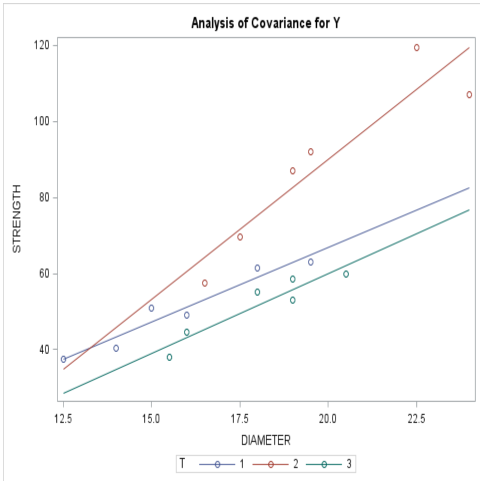
The entry X = 17.88889 is the mean value of X in the data set and is the value used when a value is not specified in the LSMEANS statement.

Alloy	Least Squares Line	X_o	$\hat{\mu}_{ix_o}^{ADJ}$	$\widehat{SE}(\hat{\mu}_{ix_o}^{ADJ})$
1	$\hat{Y} = -11.650 + 3.9200X$	17.8889	58.474	3.3461
2	$\hat{Y} = -57.037 + 7.3506X$	17.8889	74.457	3.1318
3	$\hat{Y} = -23.905 + 4.1892X$	17.8889	51.035	2.5269
1	$\hat{Y} = -11.650 + 3.9200X$	14	43.230	3.1949
2	$\hat{Y} = -57.037 + 7.3506X$	14	45.872	6.1154
3	$\hat{Y} = -23.905 + 4.1892X$	14	34.743	6.2739
1	$\hat{Y} = -11.650 + 3.9200X$	22	74.590	7.0634
2	$\hat{Y} = -57.037 + 7.3506X$	22	104.676	3.2622
3	$\hat{Y} = -23.905 + 4.1892X$	22	68.257	6.2739

- There are very large differences of $\hat{\mu}_{ix_o}^{ADJ}$ depending on the value of x_o which implies that attempting to fit regression lines with the same slope would not yield reasonable values for the mean strength of the welds.

d. The regression lines, the plot on the left is with equal slopes and the plot on the right is with unequal slopes:





e. Efficiency with a Covariate: $\sum_{i=1}^{t}\sum_{j=1}^{r_i}(X_{ij}-\bar{X}_{i.})^2 = 93.667$, $\sum_{i=1}^{t}\sum_{j=1}^{r_i}(\bar{X}_{i.}-\bar{X}_{..})^2 = 48.111$, $MSE_2 = 51.155$ and $MSE_{without\ covariate} = 237.006$. Thus,

$$E = \frac{237.006}{51.155\left(1 + \frac{48.111}{(2)93.667}\right)} = 3.686.$$

f. Test of homogeneous regressions: See b.(2)-i on this problem.

g. In comparing the average weld strengths at the three values of x_o we obtain different groupings of the three Alloys:

At $x_o = 14$, $G_1 = \{A1, A2, A3\}$

At $x_o = 17.8889$, $G_1 = \{A1, A3\}$, $G_2 = \{A2\}$

At $x_o = 22$, $G_1 = \{A1, A3\}$, $G_2 = \{A2\}$

We conclude that there is a significant Alloy effect. From the groupings, we conclude that the size and significance of the difference in the mean weld strength for the three Alloys depends on the specific value of the weld diameter. The efficiency of the covariance adjustment is $E = 3.686$. Thus, without covariance adjustment for weld diameter, about four times as many EU's would be required for this study to obtain the same precision on the estimated Alloy means.