Short answers to problems (in chapter 5, 6, 4, 10 and 11) and exercises (in chapter 3 and 8): Experimental design

Chapter 5

2) T=Temperature, Co=Concentration, Ca=Catalyst

   Main effects: $L_T = 23.0$, $L_{Co} = -5.0$, and $L_{Ca} = 1.5$

   Two-factor interaction effects: $L_{T,Co} = 1.5$, $L_{T,Ca} = 10.0$, and $L_{Co,Ca} = 0.0$

   Three-factor interaction effect: $L_{T,Co,Ca} = 5.5$

   Plot the effects (both main and two-factor interactions). It seems that T has a much larger effect than Co and Ca, and that there is interaction between T and Ca. Increasing temperature has a much larger effect on yield at the high level of Catalyst than at the low level.

   Assumptions: Often we assume that the three-factor (and higher) interaction effect can be ignored and that the errors are normally distributed with expected value 0 and constant varance σ^2.

4) a) 64 runs

   b) Let σ_E^2 denote the variance of a main effect when no replicates were taken, that is, n=1. Then $\sigma_E^2 = 4\sigma^2/2^k = \sigma^2/16$.

   c) Effect$\pm z_{0.005}\sigma_E/\sqrt{n}$ =Effect$\pm 2.575\sigma_E/\sqrt{n}$, where n is the number of (genuine) repetitions of each experiment and σ_E^2 the effect variance as above in b).

   d) Compute the length of the above confidence inteval, set it equal to 4000, and solve the equation w.r.t. n. You obtain $n = 6.6$ giving the answer that one should make 7 replications of each experiment.

12) a) Eliminate or reduce variability from known sources (block factors).

   b) 2^3 with factors A, B and C. Identify the block factor as the ABC-interaction. Runs with ABC at -1 for one of the blocks and at $+1$ for the other.

   c) We number the four combinations of the two blocks (each at two levels) by 1,2,3,4 according to:

	Block 1 on -1	Block 1 on $+1$
Block 2 on -1	1	2
Block 2 on $+1$	3	4

   Then write the setup for a full 2^3 design and let Block 1=AB and Block 2=AC.

M	A	B	C	(AB) Block 1	(AC) Block 2	Block
1	-1	-1	-1	1	1	4
1	1	-1	-1	-1	-1	1
1	-1	1	-1	-1	1	3
1	1	1	-1	1	-1	2
1	-1	-1	1	1	-1	2
1	1	-1	1	-1	1	3
1	-1	1	1	-1	-1	1
1	1	1	1	1	1	4

17) Main effects: $L_1 = 3.67$, $L_2 = -1.17$, and $L_3 = 3.33$

Two factor interactions: $L_{12} = -0.29$, $L_{13} = 1.04$, and $L_{23} = 0.47$

Three-factor interaction: $L_{123} = -0.56$

Pooled estimate for the variance of the mean response variable can be estimated by $s_p^2 = \sum s_i^2/8 = 1.74^2$, where s_i's are the standard errors given in the book. Variance of an effect becomes $s_p^2/2$, and standard error $1.74/\sqrt{2} = 1.23$.

Main effects 1 and 3 are statistically significant (the test statitic is more extreme than the "rejection limit" $t_{df,\alpha/2} = 1.98$ when α=0.05). Other effects are not significant.

20) b) Main effects: $L_1 = 0.0025$, $L_2 = -0.17$, $L_3 = 0.098$, and $L_4 = 0.22$

Two factor interactions: $L_{12} = 0.005$, $L_{13} = -0.01$, $L_{14} = -0.01$, $L_{23} = -0.075$, $L_{24} = 0.01$, $L_{34} = -0.005$

Three factor interactions: $L_{123} = 0.018$, $L_{124} = -0.0075$, $L_{134} = 0.0025$, and $L_{234} = 0.0025$

Four factor interaction: $L_{1234} = 0.01$

c) (See page 201). Error variance= $\frac{1}{5}(L_{123}^2+L_{124}^2+L_{134}^2+L_{234}^2+L_{1234}^2) = 0.000095$, and standard error 0.0097.

Chapter 6

2) Generators: E=ABCD

   Defining relation: I=ABCDE

   This is a 2_V^{5-1} design.

   We have that A=BCDE, B=ACDE, C=ABDE, D=ABCE, E=ABCD, AB=CDE, AC=BDE, AD=BCE, AE=BCD, BC=ADE, BD=ACE, BE=ACD, CD=ABE,, CE=ABD, DE=ABC

   We assume that the three-factor and higher order interactions are zero.

   Main affects: $L_A = -0.25$, $L_B = 7.25$, $L_C = 5.85$, $L_D = 1.45$, and $L_E = -0.95$

   Two factor interactions: $L_{AB} = 0.95$, $L_{AC} = -1.05$, $L_{AD} = -0.05$, $L_{AE} = -0.55$, $L_{BC} = 2.75$, $L_{BD} = -0.15$, $L_{BE} = -0.35$, $L_{CD} = 0.15$, $L_{CE} = -0.35$, and $L_{DE} = -0.45$

   B and C seem to have a larger effect than the other factors.

3) a) 8 factors but 4 of them are confounded with interaction terms.
   b) $2^4 = 16$ runs
   c) 2 levels
   d) 4 independent generators, e.g. E=ABC, F=ABD, G=ACD, H=BCD.
   e) Defining relation (e.g.): I=ABCE=ABDF=ACDG=BCDH=CDEF=BDEG =ADEH=BCFG=ACFH=ABGH=AEFG=BEFH=CEGH=DFGH =ABCDEFGH
      16 words

6) Factors A, B, C, D, E, F, G, and H, 2_{IV}^{8-4}-design: E=ABC, F=ABD, G=ACD, H=BCD

   Factors A, B, C, D, E, and F, 2_{IV}^{6-2}-design: E=ABC, F=BCD

   No, 2_V^{6-2}-design does not exist

7) Factors A, B, C, D, E, and F
   a) Resolution 4
   b) Generators: E=ABC, F=BCD
   c) Defining relation: I=ABCE=BCDF=ADEF
   d) C=ABE=BDF=ACDEF
   e) AB=CE=ACDF=BDEF

16) a) 8 runs
    b) D=-ABC, E=-AC
    c) Resolution 3
    d) I=BDE
    e) Resolution 3

Chapter 3

3.2) $\bar{x}_A = 25.5$, $\bar{x}_B = 35.0$, and $\bar{x}_A - \bar{x}_B = -9.5$

From the reference distribution we can compute 17 differences of successive groups of 4 observations. None of the differences is less than (or equal to) the observed -9.5 giving p-value 0. We have evidence that B gives higher values than A.

3.4) There are 10 ("5 choose 2") ways to select a combination of 2 A's and 3 B's. For each combination we compute $\bar{x}_A - \bar{x}_B$ which gives us the reference distribution.

Observed $\bar{x}_A - \bar{x}_B = -4$.

Test the hypothesis $H_0 : \mu_A = \mu_B$ (μ denotes the expected value) against $H_1 : \mu_A \neq \mu_B$.

Randomization: None of the $\bar{x}_A - \bar{x}_B$ computed from the reference distribution is smaller than the observed -4 and 1 equals -4. Therefore, the p-value is $1/10 = 0.1$ (misprint in the Answer in the book).

T-test (usual T-test with equal variances): Distribution $T^{(3)}$, p-value=0.07 (the two-sided).

3.6) Test the hypothesis $H_0 : \mu_A = \mu_B$ against $H_1 : \mu_A \neq \mu_B$.

Randomization: $2^5 = 32$ possible (A, B) pairs. For each combination, compute the difference $\bar{x}_A - \bar{x}_B$, and compare to the observed difference -2.6. It turned out that 8 of the 32 differences are ≤ -2.6 or ≥ 2.6 giving p-value 8/32=0.25

Paired T-test: p=0.15.

3.19) p= percentage of those who obtain scores higher than 100.

Number of male applicants $n = 43$, 10 has scores higher than 100. Therefore, estimated p is 10/43

If the applicants were a random sample from the population used to standardize the test, the number of applicants with scores higher than 100, X, would have $Bin(43, 0.5)$ distribution. Therefore, to answer the question we test $H_0 : p = 0.5$ against $H_1 : p < 0.5$. The p-value is $P(X \leq 10|p = 0.5) = 0.0003$.

3.21) With continuity correction. When p=0.3: $P(Z < \frac{(y_0-1/2)-n\times0.3}{\sqrt{n\times0.3\times(1-0.3)}}) = 0.95$.
When p=0.5: $P(Z > \frac{(y_0+1/2)-n\times0.5}{\sqrt{n\times0.5\times(1-0.5)}}) = P(Z < \frac{n\times0.5-y_0+1/2}{\sqrt{n\times0.5\times0.5}}) = 0.95$.
Thus, $\frac{y_0-1/2-n\times0.3}{\sqrt{n\times0.3\times0.7}} = \frac{n\times0.5-y_0+1/2}{\sqrt{n\times0.5\times0.5}} = 1.645$ and solve for y_0 and n.

Chapter 4

1) a) Randomized block design, paint has 4 levels and site 6 levels

   b) ANOVA table:

Source	SS	df	MS
Sites	568.71	5	113.74
Paints	665.13	3	221.71
Residuals	163.13	15	10.88

   $F_{\text{Paint}} = MS_{\text{Paint}}/MS_{\text{Res}} = 20.39$ and $F_{\text{Site}} = MS_{\text{Site}}/MS_{\text{Res}} = 10.46$ giving p-values 0.000015 and 0.00018, respectively. Both the treatment (paint) and the block (site) effect are statistically significant.

   c) $t^{(15)}_{0.025} = 2.131$ (use the df for the residual) and 95% confidence interval becomes $\bar{x} \pm 2.131 s/\sqrt{6}$, where $\bar{x}$ is the mean average wear of a paint supplier, and s is the estimated σ, i.e. $s = \sqrt{MS_{\text{Res}}}$.

   Confidence intervals:

   GS: 72.33 ± 2.87

   FD: 62.5 ± 2.87

   L: 60.5 ± 2.87

   ZK: 71.5 ± 2.87

   d) Variance seems to increase with increasing wear and a right skewed distribution for the residuals (see, residuals plotted against predicted values and a histogram of the residuals).

   e) There are a couple of large residuals. (One should check the protocol for those settings.)

   f) Average wear of FD and L are about the same, and smaller than the average wear of GS and ZK, the latter two being about the same. (This can be verified by pairwise test and confidence intervals.)

   g) A large block effect. It was worhtwhile to have a block design.

Chapter 8

8.3) a) Orthogonality:

   Treatments: -0.165+0.197-0.087+0.055=0

   Poison: 0.138+0.065-0.203=0

   Interaction: Column add up to zero, rows add up to zero

   Residuals: Add up to zero in each group

Chapter 10

1) a) $\beta_0 = 0.85$ and $\beta_1 = 0$
   b) $y = 0.8385 - 0.00016x$

5) b) $\beta_1 = 26.28$ and $\beta_2 = -4.46$

9) a) 15 $\to$ -1, 20 $\to$ +1, and $x_2 =$ (feed rate $-$ 17.5)/2.5
   40 $\to$ -1, 50 $\to$ +1, and $x_3 =$ (humidity $-$ 45)/5
   b) b_0 =overall mean, $b_1 = 0.5\cdot$effect of x_1, $b_2 = 0.5\cdot$effect of x_2, $b_3 = 0.5\cdot$effect of x_3

11) Yes. The equations we would have to solve are

$$\frac{1}{8}(9 - y_6 - y_7) = 0$$

and

$$\frac{1}{8}(-1 + y_6 - y_7) = 0$$

(that is, assume that two interaction effects are zero.)

18) a, c, and e

19) a) Linear in β_1 and β_2
   b) Linear in β_3, not β_1 and β_2
   c) Linear in β_1 and β_2 and β_3
   d) Linear in none of the parameters
   e) Linear in none of the parameters

Chapter 11

2) x_1, x_2, x_3, axial and center points dublicated (8 pure replicates)
Note the printing error: Change 170 to 107 for Lack of fit.

   a) MS: $\mathrm{MS_M} = 746/9 = 82.89$, $\mathrm{MS_R} = 154/12 = 12.83$, $\mathrm{MS_L} = 107/5 = 21.4$, and $\mathrm{MS_E} = 47/7 = 6.71$
   F ratio for the lack of fit: $\mathrm{MS_L}/\mathrm{MS_E}$ =3.18
   b) Lack-of-fit: If we use the 5% limit, $F_{5,7,0.05} = 3.97$. Since the value of the test statistic is 3.18, there is not evidence of lack of fit, but...
   The p-value using a computer program: $p = P(F_{5,7} \geq 3.18) = 0.081$.
   c) Not really