

SOLUTIONS MANUAL FOR

Design and Analysis
of Experiments with SAS

by

Seyed M. Mottaghinejad
John Lawson

SOLUTIONS MANUAL FOR

Design and Analysis of Experiments with SAS

_____ by _____

Seyed M. Mottaghinejad
John Lawson



CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

Chapman & Hall/CRC
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2011 by Taylor and Francis Group, LLC
Chapman & Hall/CRC is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works

Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number: 978-1-4398-5601-7 (Paperback)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (<http://www.copyright.com/>) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Visit the Taylor & Francis Web site at
<http://www.taylorandfrancis.com>

and the CRC Press Web site at
<http://www.crcpress.com>

Contents

1	Introduction	1
2	Completely Randomized Designs with One Factor	3
3	Factorial Designs	14
4	Randomized Blocked Designs	29
5	Designs to Study Variances	41
6	Fractional Factorial Designs	53
7	Incomplete and Confounded Block Designs	94
8	Split-Plot Designs	110
9	Crossover and Repeated Measures Designs	130
10	Response Surface Designs	143
11	Mixture Experiments	183
12	Robust Parameter Design Experiments	206

1 Introduction

PROBLEM 1.1

- a. Each person, or more specifically each pair of hands from the person, is the experimental unit.
- b. The factors or independent variables are (1) water temperature and (2) detergent concentration.
- c. The response is bacterial count on the palms of the subjects' hands.

PROBLEM 1.2

A sub-sample is a subset of the experimental unit. In this example, the palm of the subjects' right hand could be considered a sub-sample (whereas the *pair* of hands make up the experimental unit), and the palm from the left hand could be another sub-sample. Taking the average response on the two sub-samples would then give us the responses for each of the experimental units.

PROBLEM 1.3

A duplicate in this case would arise when measuring the response (bacterial count) for the same run more than once. The resulting variation in the measurements for each run could then be seen as measurement error.

PROBLEM 1.4

I like to run a 5K three days a week. There's been a lot of speculation as to whether sports drinks and energy drinks enhance your performance and, if so, by how much. In an observational study, runners would be asked if they drink sports drinks and their running performance (speed or timing) would be measured. Thus it would be possible to see if there is a statistically significant difference between the running performance of runners who use sports drinks and those who don't. However, even in the presence of such difference, there is no way to determine whether it is actually due to energy drinks or whether, for example, runners who perform better are more motivated to buy energy drinks (lurking variable).

An experiment would be conducted as follows: (1) We seek to know if sports drinks cause athletes to perform better. We are thus interested in a cause-and-effect relationship and will use a designed experiment. Our experimental units are athletes participating in the study.

Each one will be randomly assigned a treatment (energy drink or placebo) and their athletic performance will be measured. The dependent variable will consist of the time it takes the athlete (in this case a runner) to run a 5K. A decrease in time of 5 percent or less should be considered a significant one and would confirm the hypothesis that energy drinks do indeed increase performance.

PROBLEM 1.5

The experimental units in this case consist of groups of subjects, and not the subjects themselves, as it is the groups that are receiving random assignments. For example, the districts in the breast cancer screening example are the experimental units. The individual measurements on the patients thus represent sub-units, because together they form the clusters, i.e. the units.

2 Completely Randomized Designs with One Factor

PROBLEM 2.1

Objectives: The objective would be to determine how changing the time-to-rise effects the height of bread dough.

Experimental Unit: The experimental unit in would be the dough in a single loaf pan.

Response or Dependent Variable: The response would be the measured height of the risen dough in a loaf pan.

Independent variables and lurking variables: The independent variable would be the time the dough is allowed to rise, chosen by the experimenter. Lurking variables could be differences in the amount of yeast from loaf to loaf, caused by nonuniform mixing of the ingredients or a temperature gradient in the room causing some loaves to be slightly warmer or cooler than others.

Pilot Test: A pilot test could be run to determine if the rise time could be accurately controlled (by setting multiple timers for example) and whether the loaf height can be measured objectively.

Bias could enter this experiment if factor levels (independent variables) are not randomly assigned to loaves and if lurking variables, such as a temperature gradient, could influence the risen dough height. The ability to measure the risen dough height objectively could also cause problems. Careful attention to a pilot test and randomization could avoid these problems.

PROBLEM 2.2

- a. The experimental unit is a piece of paper combined with the state of the air the helicopter will fly through.
- b. We have replicates when we make and drop more than one helicopter with the same wing length. Duplicates, on the other hand, occur when several people stand at the ready and simultaneously measure the flight time of the same helicopter for the same drop.

- c. The treatment factor consists of two or more levels in which each level corresponds to a given wing length (or a given trimming size).
- d. An updraft or a down-draft could have a considerable impact during the flight. The paper stiffness, mass, and the reaction time of the people dropping and timing the helicopter can also affect the flight time.
- e. Randomization would offset any potential influence caused by a lurking variable. In other words, the lurking variables described above create a sort of built-in bias which can skew the results of our study. However, by randomly picking experimental units and assigning them to treatment groups, we offset the effect of this inconsistency by having it be "evenly spread" amongst the treatment groups.
- f. Here's the code to create it:

```

1 data rand;
2   input wings @@;
3   u=ranuni(0);
4   datalines;
5 4 4 4 4 4 4 4 4 4.75 4.75 4.75 4.75 4.75 4.75 4.75 4.75
6 5.5 5.5 5.5 5.5 5.5 5.5 5.5 5.5 6 6 6 6 6 6 6
7 run;
8 proc sort out=crd; by u;
9
10 data list; set crd;
11   helicopter =_n_; flight_time='-----';
12 proc print; var wings flight_time; id helicopter;
13 run;

```

It gives the following result:

helicopter	wings	flight_time
1	6.00	-----
2	6.00	-----
3	5.50	-----
SOME OUTPUT OMITTED		
31	4.00	-----
32	4.00	-----

- g. The experiment was conducted and the resulting data are shown below:

Obs	wings	time
1	4.00	5.2
2	5.50	5.0
3	4.75	5.1
SOME OUTPUT OMITTED		
30	4.75	5.0
31	6.00	4.9
32	4.00	5.0

- h. The ANOVA table produced below shows that the P-value is approximately .15 and is above the significance level $\alpha = .05$. We therefore fail to reject the null hypothesis and conclude that no statistically significant differences exist between the mean flight times for the four wing sizes (factor levels).

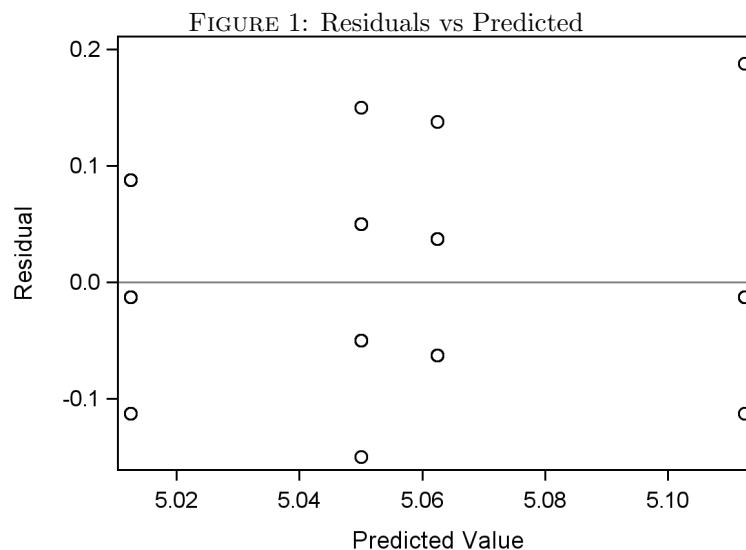
```

1 ods graphics on/width=4.0in height=3.0in;
2 proc glm data=crd plots=diagnostics(unpack);
3   class wings;
4   model flight_time=wings/solution;
5 run;
6 ods graphics off;

```

Dependent Variable: flight_time					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	0.04093750	0.01364583	1.95	0.1449
Error	28	0.19625000	0.00700893		
Corrected Total	31	0.23718750			

- i. To check the assumption of equal variances across factor levels we look at the residuals vs. predicted value plot, which was created by the `plots=diagnostics(unpack);` option on the `proc glm` statement above and is shown in Figure 1. The plot shows the residual variation is approximately equal at each predicted value or factor level.

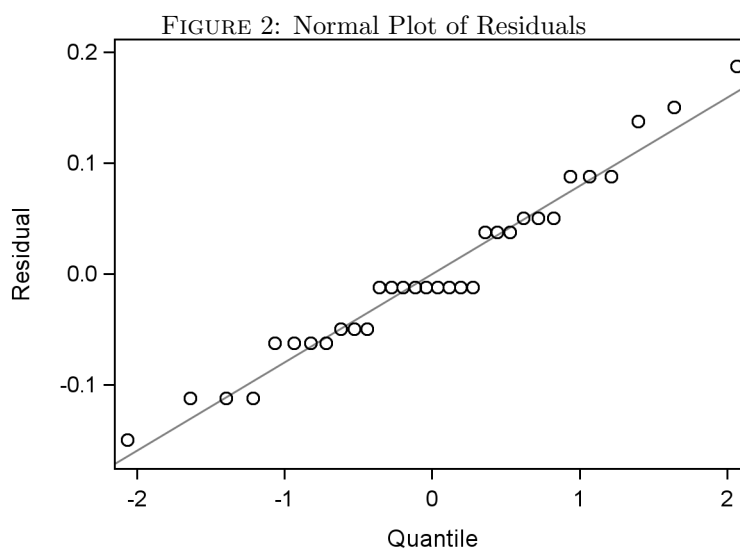


Furthermore, to see if the residuals are approximately normal, we look at the normal probability plot of the residuals. The plot in Figure 2 shows that the residuals and their normal scores more or less fall along a straight line. Thus it is safe to assume normality.

```

1 proc rank data=s normal=vw; var resid; ranks zscore;
2 proc gplot;
3 plot resid*zscore;
4 symbol1 i=none v=dot c=black;
5 run;

```



j. The results below show that neither the linear nor the quadratic terms are significant:

```

1 proc iml;
2 t={4 4.75 5.5 6};
3 C=orpol(t);
4 print C;
5 quit;

```

```

C
0.5 -0.701068 0.4750737 -0.181131
0.5 -0.206197 -0.609528 0.5796188
0.5 0.2886751 -0.376473 -0.724524
0.5 0.6185896 0.5109283 0.3260356

```

```

1 proc glm data=crd;
2   class wings;
3   model flight_time=wings;
4   estimate 'Linear Term' wings -0.701068 -0.206197 0.2886751 0.6185896;
5   estimate 'Quadratic Term' wings 0.4750737 -0.609528 -0.376473 0.5109283; run;

```

Parameter	Estimate	Standard Error	t Value	Pr > t
Linear Term	-0.00773239	0.02959926	-0.26	0.7958
Quadratic Term	-0.05030844	0.02959923	-1.70	0.1003

PROBLEM 2.3

- The experimental unit consists of the state of the network (i.e., amount of traffic) at the point in time when the file is downloaded.
- The treatment factor is a list of web sites, i.e. each factor level is a different web site from which one or several files will be downloaded.
- The response is the time it takes for the file to be downloaded.
- The experimental error is the difference of each observed download time from the long run average download time at a particular web site. Experimental error can be categorized into two types of error: bias error and random error. Bias error is caused by the effect of lurking variables and can hence be offset by the random assignment of experimental units to factor levels. Random error, on the other hand, is caused by the inability to repeat the same result (or download time) within a given factor level (or web site). An estimate for the random experimental error can only be obtained when replicates exist for each factor level (i.e. at least two units per factor level). Thus, randomization reduces bias error while replicates provide an estimate of the random error.

PROBLEM 2.4

- The dough for each biscuit is an experimental unit since it was randomly assigned one of four factor levels.
- The ANOVA table resulting from the experiment is shown below. Since $P\text{-value} < .0001 < \alpha$, we reject the null hypothesis and conclude that at least one treatment level

mean is significantly different from the other ones, i.e. for at least one level of baking powder, the mean rise in the dough is significantly different from the rest.

```

1 data bread;
2   input tsp h1-h4;
3   height=h1; output;
4   height=h2; output;
5   height=h3; output;
6   height=h4; output;
7   keep tsp height;
8 datalines;
9 .25 11.4 11 11.3 9.5
10 .5 27.8 29.2 26.8 26
11 .75 47.6 47 47.3 45.5
12 1 61.6 62.4 63 63.9
13 run;
14
15 proc glm;
16 class tsp; model height=tsp/solution; run;

```

Dependent Variable: height					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	6145.731875	2048.577292	1822.65	<.0001
Error	12	13.487500	1.123958		
Corrected Total	15	6159.219375			

- c. To test for a proportional increase of the mean response for increases in the factor levels we can see if the linear term is significant. The results below show that it is in fact significant.

```

1 proc iml;
2 t={.25 .5 .75 1};
3 C=orpol(t);
4 print C;
5 quit;

```

```

C
0.5 -0.67082 0.5 -0.223607
0.5 -0.223607 -0.5 0.6708204
0.5 0.2236068 -0.5 -0.67082
0.5 0.6708204 0.5 0.2236068

```

```

1 proc glm data=bread;
2   class tsp;
3   model height=tsp;
4   estimate 'Linear Trend' tsp -0.67082 -0.223607 0.223607 0.67082;
5   output out=s r=resid p=yhat;
6 run;

```

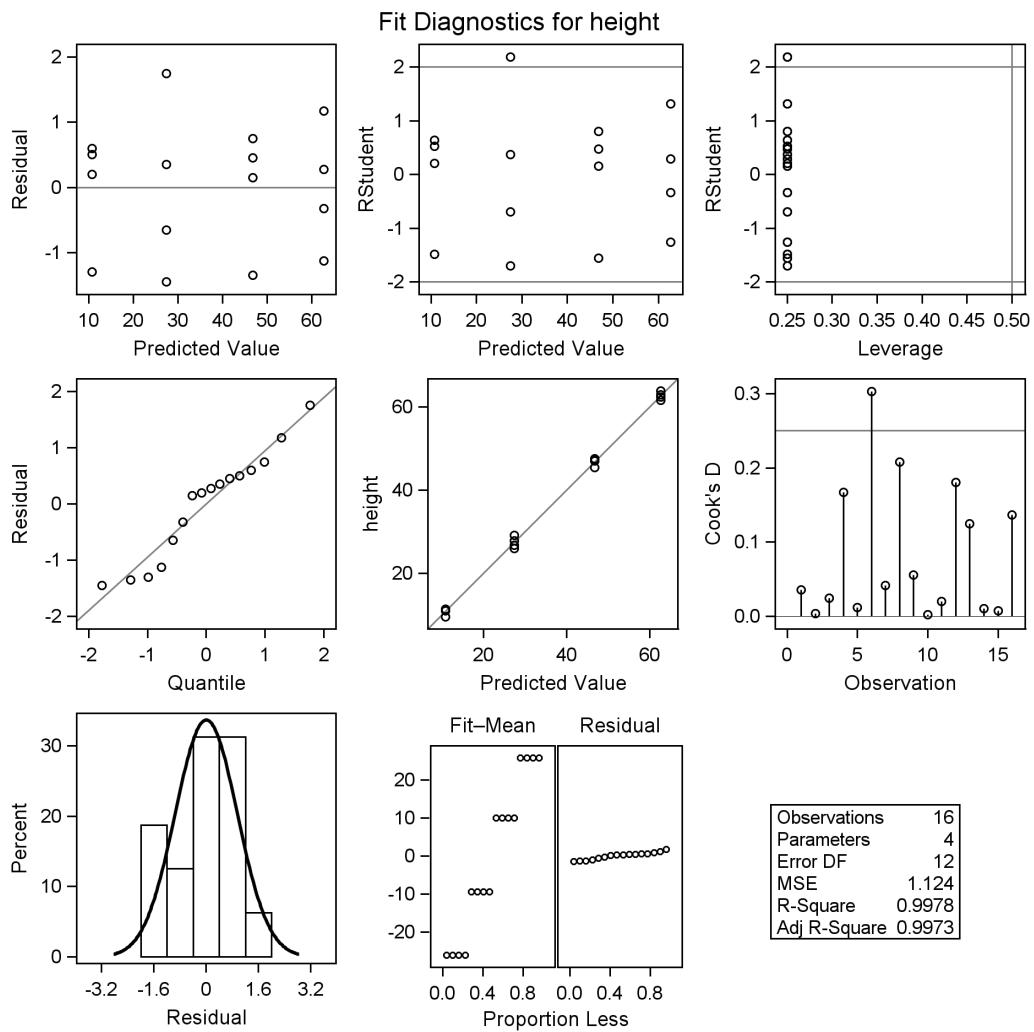
Parameter	Estimate	Standard Error	t Value	Pr > t
Linear Trend	39.1703043	0.53008427	73.89	<.0001

d. The variance of the experimental error is estimated by the mean square error from the ANOVA table, which is approximately 1.124.

```

1 ods graphics on;
2 proc glm plots=diagnostic;
3 class tsp; model height=tsp/solution; run;
4 ods graphics off;
    
```

FIGURE 3: Diagnostic Panel



- e. By including the `ods graphics on;` statement and the `plots=diagnostic` option on the model statement (shown above), `proc glm` produces the diagnostic panel of residual plots shown in Figure 3. The residuals vs. predicted values plot is shown in the upper left of the panel and indicates that the residuals have overall about the same variance, although the second group from the left seems to have slightly larger variation than the other groups. However, with as few observations as there are in this data set, there is not strong evidence against homogeneity of variance. It would be safe to assume the equal variance assumption is satisfied. The normal probability plot of the residuals is shown in second row of the diagnostic panel. The residuals and their normal scores seem to be satisfactorily aligned with each other. It is therefore safe enough to assume normality.
- f. Yes; if the four biscuits from each factor level were put in the oven together, then our experimental unit would consist of the batches of dough and the dough for individual biscuits would be subunits. Thus our experiment would not have replicates and we would not be able to estimate the experimental error. In a case like this we might use the variance between the subsamples within each unit as an estimate of the experimental error, but we would do this at our own risk since this measure could be highly biased.

PROBLEM 2.5

- a. The data was entered into SAS as shown below, where the yield for the four groups were put into one column.

```

1 data asp;
2   input trt$ y1-y5;
3   yield=y1; output;
4   yield=y2; output;
5   yield=y3; output;
6   yield=y4; output;
7   yield=y5; output;
8   keep trt yield;
9 datalines;
10 Control 94.7 96.1 86.5 98.5 94.9
11 IAA 89.9 94 99.1 92.8 99.4
12 ABA 96.8 87.8 89.1 91.1 89.4
13 GA3 99.1 95.3 94.6 93.1 95.7
14 CPPU 104.4 98.9 98.9 106.5 104.8
15 run;

```

An analysis of variance was performed in `proc glm` and the results are shown here:

```

1 proc glm data=asp;
2   class trt;
3   model yield=trt/solution;
4   lsmeans trt/pdiff adjust=tukey;
5   means trt/dunnett('Control');
6 run;

```

Dependent Variable: yield

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	377.4936000	94.3734000	6.98	0.0011
Error	20	270.2680000	13.5134000		
Corrected Total	24	647.7616000			

It is evident from the results that we must reject the null hypothesis of equal treatment means and conclude that there is at least one treatment level whose mean is significantly different from the rest. The P-value for the test is .0011.

- b. The `lsmeans` option was used in `proc glm` to find which pairwise comparisons are significant using Tukey's HSD. The results below show that the CPPU treatment level had a significantly larger mean than the other four treatment levels. No other significant differences were detected.

Least Squares Means for effect trt					
Pr > t for H0: LSMean(i)=LSMean(j)					
Dependent Variable: yield					
i/j	1	2	3	4	5
1		0.0005	0.6230	0.2881	0.3973
2	0.0005		0.0114	0.0425	0.0266
3	0.6230	0.0114		0.9717	0.9948
4	0.2881	0.0425	0.9717		0.9994
5	0.3973	0.0266	0.9948	0.9994	

- c. The results below are similar to what Tukey's HSD gave us, i.e. only the CPPU group is shown to significantly improve yield compared to the control group. Furthermore, we are 95 percent confident that the increase will be between 2.396 and 14.724. The other three groups are not significantly different from the control group. A one-way test of the hypothesis (using `dunnett`) lead to the same conclusion.

Comparisons significant at the 0.05 level are indicated by ***.					
Difference					
trt		Between	Simultaneous 95%		
Comparison		Means	Confidence	Limits	
CPPU	- Control	8.560	2.396	14.724 ***	
GA3	- Control	1.420	-4.744	7.584	
IAA	- Control	0.900	-5.264	7.064	
ABA	- Control	-3.300	-9.464	2.864	

PROBLEM 2.6

- a. When Δ is a multiple of σ , then we get:

$$\lambda = \frac{r}{\sigma^2} \sum_{i=1}^4 (\mu_i - \bar{\mu})^2 = \frac{r}{\sigma^2} \frac{\Delta^2}{2} = 2r$$

which we adjusted in the code below to obtain the power:

```

1 data Power;
2 do r=2 to 10;
3   nu1=4-1; * df for numerator;
4   nu2=4*(r-1); * df for denomonator;
5   alpha=.05;
6   Fcrit=finv(1-alpha,nu1,nu2); *F critical value;
7   nc=2*r;*noncentrality parameter for noncentral F;
8   power=1-probf(Fcrit,nu1,nu2,nc);
9   output;
10 end;
11 keep r nu1 nu2 nc power;
12 title Power Calculation in Data Step;
13 proc print; run;

```

Obs	r	nu1	nu2	nc	power
1	2	3	4	4	0.16980
2	3	3	8	6	0.33906
3	4	3	12	8	0.50370
4	5	3	16	10	0.64423
5	6	3	20	12	0.75459
6	7	3	24	14	0.83613
7	8	3	28	16	0.89360
8	9	3	32	18	0.93258
9	10	3	36	20	0.95819

- b. The results from `proc glmpower` are shown below and tally with the results from the SAS Analyst Tool:

```

1 data case;
2 input trt meanht;
3 datalines;
4 1 -1
5 2 0
6 3 0
7 4 1
8 proc glmpower;
9   class trt;
10  model meanht=trt;
11  power
12  stddev=1
13  ntotal=8 to 40 by 4
14  power = .;
15 run;

```

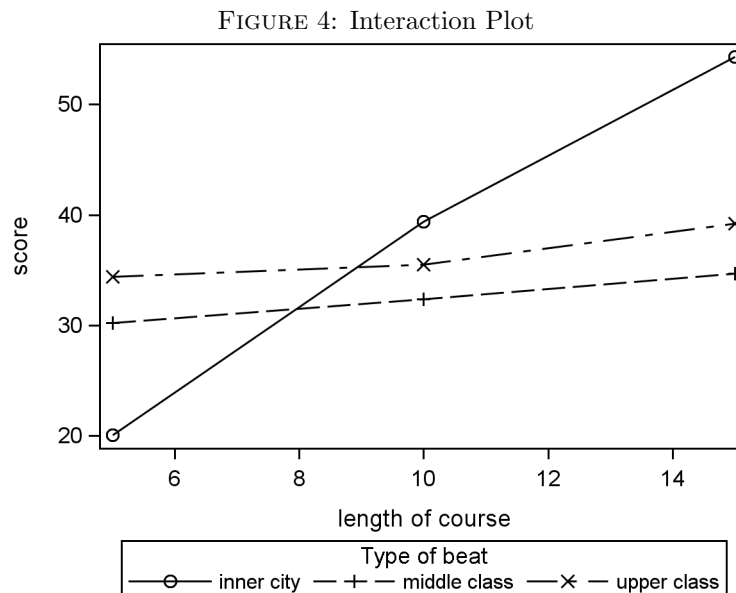
Index	Computed Power		Power
	N	Error	
Total	DF		
1	8	4	0.169803
2	12	8	0.339058
3	16	12	0.503705
4	20	16	0.644233
5	24	20	0.754586
6	28	24	0.836129
7	32	28	0.893598
8	36	32	0.932577
9	40	36	0.958186

- c. To obtain at least 90 percent power we need 9 replicates in each treatment level.
- d. The number of replicates would change to 11 and 7 respectively.

3 Factorial Designs

PROBLEM 3.1

- a. The interaction plot is shown here:



- b. As course length goes up, it causes an increase in the test scores. This increase is approximately the same for officers assigned to the middle-class beat and the upper-class beat. However, for officers assigned to the lower-class beat, the score increases approximately six times as much as it does for officers assigned to the other two beats.

PROBLEM 3.2

- a. The condition of the catapult (i.e. elasticity of the rubber band, position of the arm and friction with the other parts, and the force used to push the ball into the cup) at the time the start and top angles are set would represent the experimental unit in this case. By randomizing this order, we can offset the possible effect of the lurking variables that represents the friction of the catapult after every use and loss of elasticity of the rubber band.
- b. The following code was used; the result is shown in the table below (the variables "Start" and "Stop" represent start angle and stop angle respectively).

```

1 proc factex;
2 factors Stop Start/nlev=3;
3 model estimate=(Stop|Start);
4 output out=exp randomize designrep=2
5 Stop nvals=(1 2 3) Start nvals=(1 2 3);
6 run;
7 proc print data=exp;
8 run;

```

Obs	Stop	Start
1	1	2
2	3	2
3	2	2
4	3	3
5	1	3
6	3	1
7	2	1
8	2	3
9	1	1
10	2	2
11	1	1
12	3	1
13	1	2
14	3	3
15	2	1
16	1	3
17	3	2
18	2	3

c. The results show that we reach the required power with six replicates.

```

1 data power;
2 do r=2 to 6;
3   nu1=9-1; * degrees of freedom for numerator;
4   nu2=9*(r-1); *degrees of freedom for denominator;
5   alpha=0.05;
6   Fcrit=finv(1-alpha,nu1,nu2); * Fcritical value;
7   sigma2=12;
8   css=50; * (10**2/2);
9   nc=r*(css)/sigma2; * noncentrality parameter for noncentral F;
10  power=1-probf(Fcrit,nu1,nu2,nc);
11  output;
12 end;
13 keep r nu1 nu2 nc power;
14 proc print; run;

```

Obs	r	nu1	nu2	nc	power
1	2	8	9	8.3333	0.25151
2	3	8	18	12.5000	0.50644
3	4	8	27	16.6667	0.71674
4	5	8	36	20.8333	0.85463
5	6	8	45	25.0000	0.93190

- d. Because of increased power due to hidden replications, the results below show that two replicates per cell will be sufficient.

```

1 data power;
2   do r=2 to 4;
3     nu1=3-1; * df for numerator;
4     nu2=9*(r-1); * df for denominator;
5     alpha=.05;
6     Fcrit=finv(1-alpha,nu1,nu2); *F critical;
7     sigma2=12;
8     css=288; * (12**2/2);
9     nc=3*r*(css)/sigma2; *noncentrality parameter for noncentral F;
10    power=1-probf(Fcrit,nu1,nu2,nc);
11    output;
12  end;
13 keep r nu1 nu2 nc power;
14 proc print; run;

```

Obs	r	nu1	nu2	nc	power
1	2	2	9	144	1
2	3	2	18	216	1
3	4	2	27	288	1

- e. The experiment was conducted and the data recorded as follows:

```

1 proc factex;
2 factors Stop Start/nlev=3;
3 model estimate=(Stop|Start);
4 output out=exp randomize designrep=2
5 Stop nvals=(1 2 3) Start nvals=(1 2 3);
6 run;
7 data new; set exp;
8 input y @@;
9 datalines;
10 6.6 7.6 4.4 7.0 4.2 4.2 4.9 5.4 6.5 4.7 5.1 7.9 6.7 6.8 5.7 4.3 6.6 5.6
11 run;

```

Obs	Stop	Start	y
1	3	1	6.6
2	3	2	7.6
3	2	2	4.4
4	1	2	7.0
5	2	1	4.2
6	3	3	4.2
7	1	3	4.9
8	1	1	5.4
9	2	3	6.5
10	2	3	4.7
11	2	2	5.1
12	3	1	7.9
13	1	1	6.7
14	3	2	6.8
15	1	3	5.7

16	2	1	4.3
17	1	2	6.6
18	3	3	5.6

- f. The overall test as shown here tells us that significant factor effects and or interaction effects exist. Furthermore, the type I sums of squares tells us that the stop angle is a significant factor, but the start angle main effect and the interaction are not significant at $\alpha = .05$.

```

1 proc glm;
2 class Stop Start;
3 model y=Stop|Start/solution;
4 means Start/snk;
5 run;

```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	19.45777778	2.43222222	4.16	0.0239
Error	9	5.26000000	0.58444444		
Corrected Total	17	24.71777778			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Stop	2	8.13444444	4.06722222	6.96	0.0149
Start	2	2.93444444	1.46722222	2.51	0.1360
Stop*Start	4	8.38888889	2.09722222	3.59	0.0516

Means with the same letter are not significantly different.

SNK Grouping	Mean	N	Stop
A	6.4500	6	3
A	6.0500	6	1
B	4.8667	6	2

- g. The results show that when the stop angle is either 1 or 3, the distance traveled is significantly higher than when it is set to 2. However, there are no significant differences between stop angles 1 and 3.

PROBLEM 3.3

- a. The interaction plot is shown in Figure 5:

```

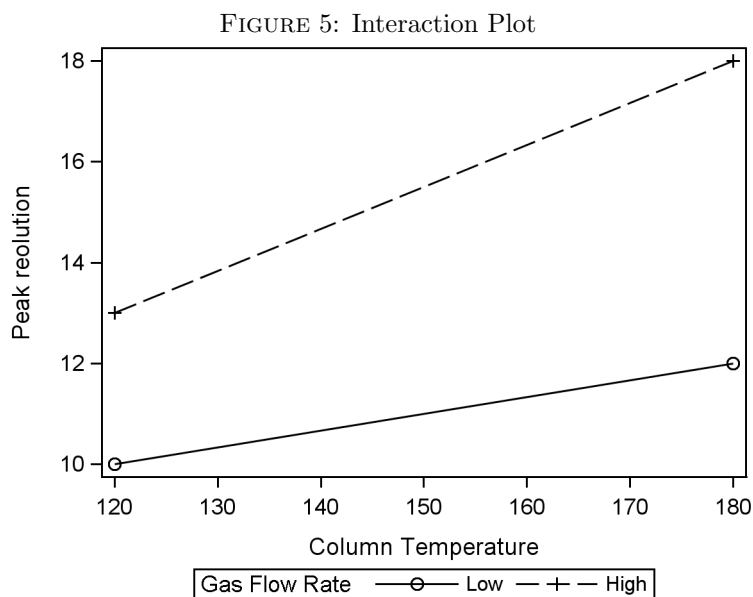
1 data res;
2 input temp gas $ resol;
3 datalines;
4 120 Low 10

```

```

5 120 High 13
6 180 Low 12
7 180 High 18
8 run;
9 ods graphics on/ border=off width=4.0in height=3.0in;
10 proc sgplot ;
11 series x= temperature y=resol/group=gas markers;
12 yaxis label='Peak resolution';
13 xaxis label='Column Temperature';
14 keylegend/ title='Gas Flow Rate';
15 run;
16 ods graphics off;

```



- b. As temperature goes up from 120 to 180 degrees, it causes the average peak resolution to increase. This is true whether the gas flow rate is low or high. However, the increase in the average peak resolution is greater when the gas flow rate is high than when it is low.

PROBLEM 3.4

- a. The experimental unit consists of the 1/3 cup of popcorn kernels that we put in the microwave each run.
- b. The standard deviation of our pilot experiment turns out to be approximately equal to .14. The minimum corrected sums of squares for marginal means that differ by 0.25 is

$css = \left(\frac{.25^2}{2}\right) = .03125$. If three levels of Brand were used and two levels each for Time and Power, the numerator degrees of freedom for Brand would be 2 and the numerator degrees of freedom for Time and Power would be 1. Assuming all interactions will be included in the model, the denominator degrees of freedom is $12(r - 1)$. Modifying the code on page 70 in the text, we see that there will be greater than 90% power for detecting marginal differences of 0.25 in Brands with 3 replicates per cell and greater than 90% power for detecting a marginal difference of 0.25 for Time and Power with 2 replicates per cell (if the standard deviation of replicates remains 0.14).

```

1 data Power;
2   do r=2 to 6;
3     nu1TP=1; *numerator degrees of freedom for Time and Power;
4     nu1B=2; *numerator degrees of freedom for Brand;
5     nu2=12*(r-1); * denominator degrees of freedom ;
6     alpha=0.05;
7     FcritTP=finv(1-alpha,nu1TP,nu2); * F critical for Time & Power;
8     FcritB=finv(1-alpha,nu1B,nu2); * F critical for Brand;
9     css = 0.03124;
10    sigma2=(0.14)**2;
11    ncTP=6*r*(css)/sigma2; * noncentrality for Time & Power;
12    ncB=4*r*(css)/sigma2; * noncentrality for Brand;
13    powerTP=1-probf(FcritTP,nu1TP,nu2,ncTP);
14    powerB=1-probf(FcritB,nu1B,nu2,ncB);
15    output;
16  end;
17  keep r nu1TP nu1B nu2 ncTP ncB powerTP powerB;
18 proc print; run;

```

Obs	r	nu1TP	nu1B	nu2	ncTP	ncB	powerTP	powerB
1	2	1	2	12	19.1265	12.7510	0.97945	0.80810
2	3	1	2	24	28.6898	19.1265	0.99924	0.96486
3	4	1	2	36	38.2531	25.5020	0.99997	0.99437
4	5	1	2	48	47.8163	31.8776	1.00000	0.99920
5	6	1	2	60	57.3796	38.2531	1.00000	0.99990

- c. We created 3 levels of popcorn brand, and two levels of both the power and the time the popcorn is left in the microwave. The randomized order in which the runs are to be conducted is shown here:

```

1 proc factex;
2   factors Time Power;
3   output out=TP Time cvals=('Short' 'Long')
4     Power cvals=('Low' 'High');
5   run;
6   factors Brand/nlev=3;
7   output out=popcorn designrep=3 pointrep=TP
8   Brand cvals=('Brand A' 'Brand B' 'Brand C');
9 data popcorn; set popcorn;

```