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:

```
data rand;
  input wings @@;
  u=ranuni(0);
  datalines;
4 4 4 4 4 4 4 4 4.75 4.75 4.75 4.75 4.75 4.75 4.75 4.75
5.5 5.5 5.5 5.5 5.5 5.5 5.5 5.5 6 6 6 6 6 6 6 6
run;
proc sort out=crd; by u;

data list; set crd;
   helicopter =_n_; flight_time='_____________';
proc print;  var wings flight_time; id helicopter;
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
```

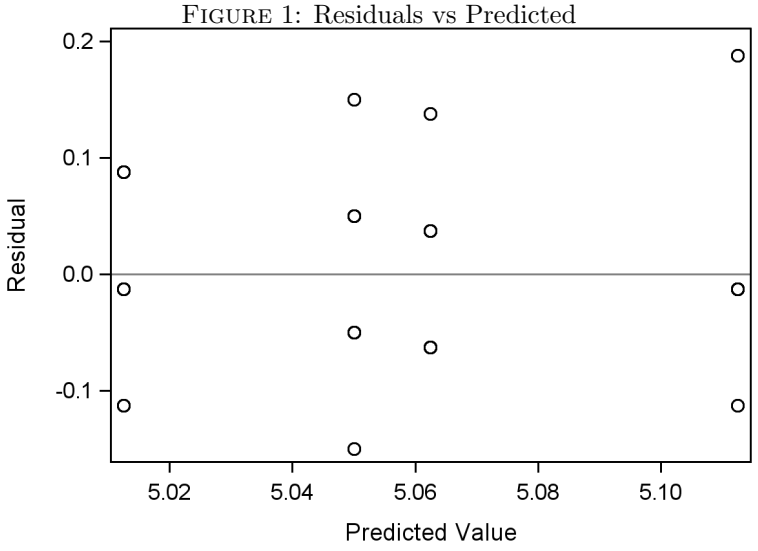
h. The ANOVA table produced below shows that the P-value is approximately .15 and is above the significance level $\alpha = .05$. We therefore fail to reject the null hypothesis and conclude that no statistically significant differences exist between the mean flight times for the four wing sizes (factor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

```
ods graphics on/width=4.0in height=3.0in;
proc glm data=crd plots=diagnostics(unpack);
  class wings;
  model flight_time=wings/solution;
run;
ods graphics off;
```

```
Dependent Variable: flight_time
                                Sum of
Source              DF         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.

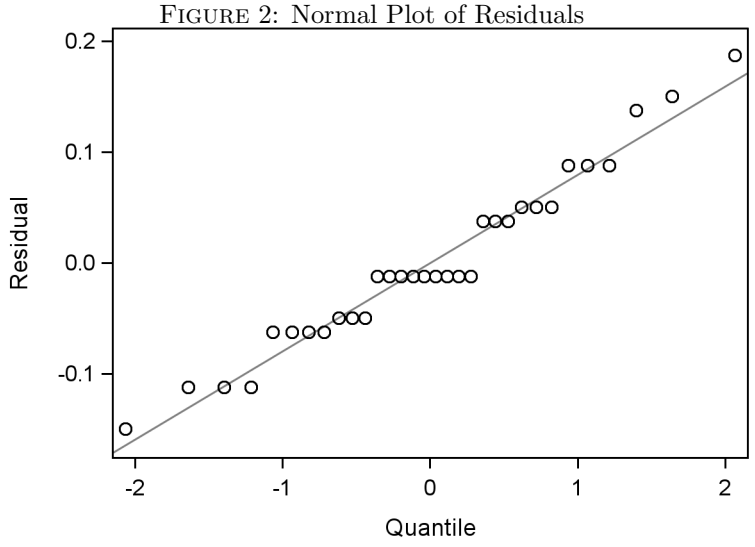
FIGURE 1: Residuals vs Predicted



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.

```
proc rank data=s normal=vw; var resid; ranks zscore;
proc gplot;
plot resid*zscore;
symbol1 i=none v=dot c=black;
run;
```

FIGURE 2: Normal Plot of Residuals



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

```
proc iml;
t={4 4.75 5.5 6};
C=orpol(t);
print C;
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
```

```
proc glm data=crd;
  class wings;
  model flight_time=wings;
  estimate 'Linear Term' wings -0.701068 -0.206197 0.2886751 0.6185896;
  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

a. 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.

b. 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.

c. The response is the time it takes for the file to be downloaded.

d. 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 exists 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

a. The dough for each biscuit is an experimental unit since it was randomly assigned one of four factor levels.

b. The ANOVA table resulting from the experiment is shown below. Since P-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.

```
data bread;
  input tsp h1-h4;
  height=h1; output;
  height=h2; output;
  height=h3; output;
  height=h4; output;
  keep tsp height;
datalines;
.25 11.4 11 11.3 9.5
.5 27.8 29.2 26.8 26
.75 47.6 47 47.3 45.5
1 61.6 62.4 63 63.9
run;

proc glm;
class tsp; model height=tsp/solution; run;
```

```
Dependent Variable: height
                              Sum of
Source              DF        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.

```
proc iml;
t={.25 .5 .75 1};
C=orpol(t);
print C;
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
```

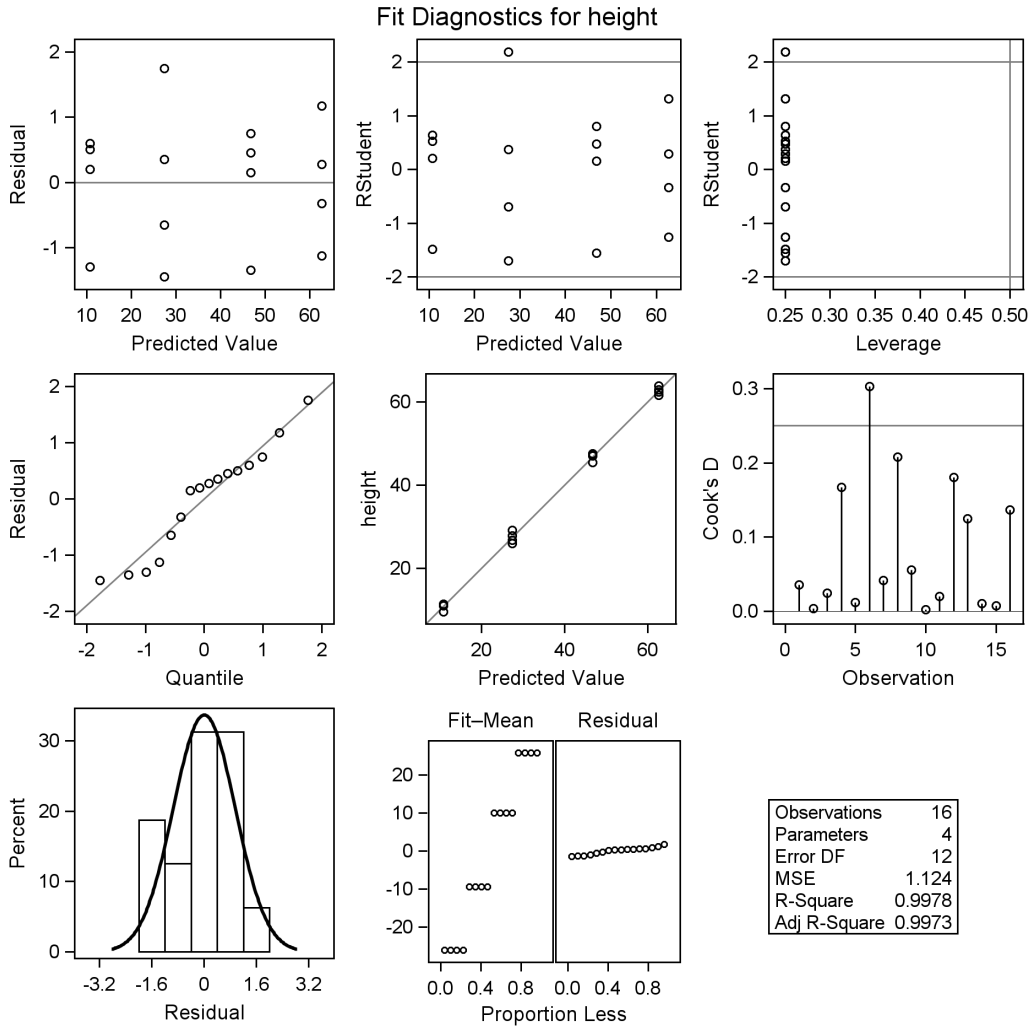
```
proc glm data=bread;
  class tsp;
  model height=tsp;
  estimate 'Linear Trend' tsp -0.67082 -0.223607 0.223607 0.67082;
  output out=s r=resid p=yhat;
run;
```

```
                                 Standard
Parameter            Estimate        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.

```
data asp;
  input trt$ y1-y5;
  yield=y1; output;
  yield=y2; output;
  yield=y3; output;
  yield=y4; output;
  yield=y5; output;
  keep trt yield;
datalines;
Control 94.7 96.1 86.5 98.5 94.9
IAA 89.9 94 99.1 92.8 99.4
ABA 96.8 87.8 89.1 91.1 89.4
GA3 99.1 95.3 94.6 93.1 95.7
CPPU 104.4 98.9 98.9 106.5 104.8
run;
```

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

```
proc glm data=asp;
 class trt;
 model yield=trt/solution;
 lsmeans trt/pdiff adjust=tukey;
 means trt/dunnett('Control');
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 `dunnettu`) 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:

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

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

```
data Power;
do r=2 to 10;
   nu1=4-1; * df for numerator;
   nu2=4*(r-1); * df for denomonator;
   alpha=.05;
   Fcrit=finv(1-alpha,nu1,nu2); *F critical value;
   nc=2*r;*noncentrality parameter for noncentral F;
   power=1-probf(Fcrit,nu1,nu2,nc);
   output;
end;
keep r nu1 nu2 nc power;
title Power Calculation in Data Step;
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:

```
data case;
input trt meanht;
datalines;
1    -1
2     0
3     0
4     1
proc glmpower;
  class trt;
  model meanht=trt;
  power
    stddev=1
    ntotal=8 to 40 by 4
    power = .;
run;
```

Computed Power			
	N	Error	
Index	Total	DF	Power
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.