Stat 541 Homework #3

Kenny Flagg

February 1, 2016

2. (a) H_0: $\tau_1 = \tau_2 = \tau_3 = \tau_4$; All effects are equal.
   H_a: $\tau_i \neq \tau_j$ for some $i \neq j$; Some effects differ.

   (b) SAS produced the following output. (See code in the appendix.)

```
                      PORTLAND CEMENT PROBLEM

                         The GLM Procedure

                  Dependent Variable: strength

                                        Sum of
Source                      DF         Squares     Mean Square    F Value    Pr > F

Model                        3     489740.1875     163246.7292      12.73    0.0005

Error                       12     153908.2500      12825.6875

Corrected Total             15     643648.4375


            R-Square     Coeff Var      Root MSE    strength Mean

            0.760881      3.862817      113.2506         2931.813


Source                      DF     Type III SS     Mean Square    F Value    Pr > F

technique                    3     489740.1875     163246.7292      12.73    0.0005
```

   Since the `technique` row has p-value = 0.0005, we reject H_0 with $\alpha = 0.05$. There is evidence that the mean tensile strength is not the same for all mixing techniques.

(c) Fisher's LSD output from SAS:

```
                         PORTLAND CEMENT PROBLEM

                            The GLM Procedure

                       t Tests (LSD) for strength

NOTE: This test controls the Type I comparisonwise error rate, not the
                       experimentwise error rate.


                 Alpha                                0.05
                 Error Degrees of Freedom               12
                 Error Mean Square                12825.69
                 Critical Value of t               2.17881
                 Least Significant Difference       174.48


      Means with the same letter are not significantly different.


         t Grouping          Mean      N    technique

                  A       3156.25      4    2

                  B       2971.00      4    1
                  B
                  B       2933.75      4    3

                  C       2666.25      4    4
```

All pairs of means are significantly different except for the means of techniques 1 and 3.

(d) Bonferroni's procedure would result in fewer significant differences because the minimum significant difference would be larger than that used by Fisher's procedure. In particular, techniques 1 and 2 may no longer be significantly different because their difference (185.25 lb/in^2) is only slightly higher than Fisher's least significant difference.

(e) Since the measurements differ in the number of significant digits, I would ask if the measurements were made by the same individual and the same equipment. Different individuals and equipment would be sources of dependency between observations that this analysis does not account for.

3. (a) H_0: $\sigma^2_\tau = 0$; There is no variation in the mean calcuim concentration between batches.
H_a: $\sigma^2_\tau > 0$; There is variation in the mean calcuim concentration between batches.

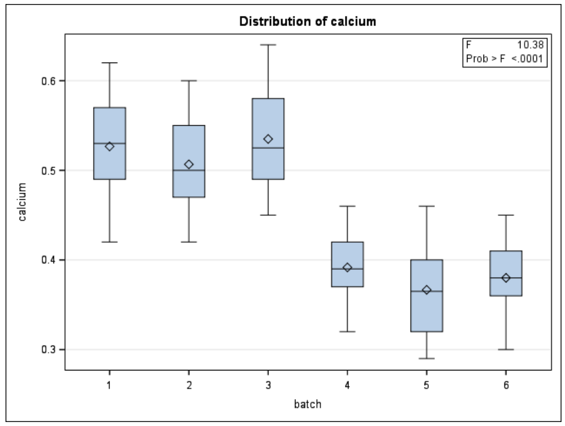
(b) ANOVA table from SAS:

```
                         BATCH VARIABILITY PROBLEM

                           The GLM Procedure

                        Dependent Variable: calcium

                                         Sum of
Source                        DF        Squares    Mean Square    F Value    Pr > F

Model                          5     0.18932222     0.03786444      10.38    <.0001

Error                         30     0.10943333     0.00364778

Corrected Total               35     0.29875556


             R-Square     Coeff Var      Root MSE    calcium Mean

             0.633703      13.38846      0.060397        0.451111


Source                        DF    Type III SS    Mean Square    F Value    Pr > F

batch                          5     0.18932222     0.03786444      10.38    <.0001
```

(c) The REML estimates from `PROC VARCOMP` are $\widehat{\sigma}_\tau^2 = 0.0057028$ and $\widehat{\sigma}^2 = 0.0036478$.

```
                     BATCH VARIABILITY PROBLEM

              Variance Components Estimation Procedure

                      Class Level Information

                 Class         Levels    Values

                 batch              6    1 2 3 4 5 6


              Number of Observations Read          36
              Number of Observations Used          36


                   Dependent Variable:    calcium


                           REML Iterations

     Iteration          Objective          Var(batch)          Var(Error)

             0    -184.7778274868        0.0057027778        0.0036477778
             1    -184.7778274868        0.0057027778        0.0036477778


      Convergence criteria met.


                           REML Estimates

                      Variance
                      Component        Estimate

                      Var(batch)     0.0057028
                      Var(Error)     0.0036478


             Asymptotic Covariance Matrix of Estimates

                             Var(batch)      Var(Error)

             Var(batch)      0.00001595      -1.4785E-7
             Var(Error)      -1.4785E-7      8.87086E-7
```

(d) Boxplot of calcium concentration by batch, produced by SAS:



All of the batches have somewhat similar spreads. Batches 1, 2, and 3 have similar means. These are noticeably higher than the means of batches 4, 5, and 6, which are themselves quite similar.

(e) We do not have enough information to assess the assumption of independence between batches. The pattern in the boxplot could occur if one person or device measured the first three batches and another person or device measured the second three batches. We need to know what equipment and operator made each measurement.

4. Let μ_0 be the mean for the sham treatment, and let μ_i be the mean for the i h/day treatment, $i = 1, 2, 3$.

Ignoring the sham treatement, orthogonal linear and quadratic contrasts for the treatment means are:

$$\Gamma_L = 0\mu_0 - 1\mu_1 + 0\mu_2 + 1\mu_3$$
$$\Gamma_Q = 0\mu_0 + 1\mu_1 - 2\mu_2 + 1\mu_3$$

The estimates of the treatment means are

$$\widehat{\mu}_0 = 5.6725 \qquad \widehat{\mu}_1 = 6.1655 \qquad \widehat{\mu}_2 = 5.4780 \qquad \widehat{\mu}_3 = 6.3505$$

so the estimates of the contrasts are

$$\widehat{\Gamma}_L = 0 \times 5.6725 - 1 \times 6.1655 + 0 \times 5.4780 + 1 \times 6.3505 = 0.185$$
$$\widehat{\Gamma}_Q = 0 \times 5.6725 + 1 \times 6.1655 - 2 \times 5.4780 + 1 \times 6.3505 = 1.560$$

Appendix: SAS Code

Problem 2

```
options center nodate nonumber linesize=80 formdlim='*';
ods listing file='hw3p2.lis';

TITLE 'PORTLAND CEMENT PROBLEM';

DATA in;
  DO technique = 1 TO 4;
    DO rep = 1 TO 4;
      INPUT strength @@;
      OUTPUT;
    END;
  END;
DATALINES;
3129 3000 2865 2890  3200 3300 2975 3150
2800 2900 2985 3050  2600 2700 2600 2765
;

PROC GLM DATA=in;
  CLASS technique;
  MODEL strength = technique / SS3;
  MEANS technique / LSD;

RUN;
QUIT;
```

Problem 3

```
options center nodate nonumber linesize=80 formdlim='*';
ods listing file='hw3p3.lis';

TITLE 'BATCH VARIABILITY PROBLEM';

DATA in;
  DO batch = 1 TO 6;
    DO rep = 1 TO 6;
      INPUT calcium @@; OUTPUT;
    END;
  END;
DATALINES;
0.51 0.55 0.62 0.42 0.49 0.57  0.60 0.47 0.42 0.52 0.55 0.48
0.51 0.64 0.45 0.54 0.49 0.58  0.32 0.40 0.37 0.46 0.42 0.38
0.29 0.46 0.37 0.32 0.40 0.36  0.30 0.41 0.36 0.45 0.39 0.37
;
```

```
PROC GLM DATA=in;
  CLASS batch;
  MODEL calcium = batch / SS3;
  RANDOM batch / TEST;

PROC VARCOMP DATA=in METHOD=REML;
  CLASS batch;
  MODEL calcium = batch / FIXED=0;

RUN;
QUIT;
```