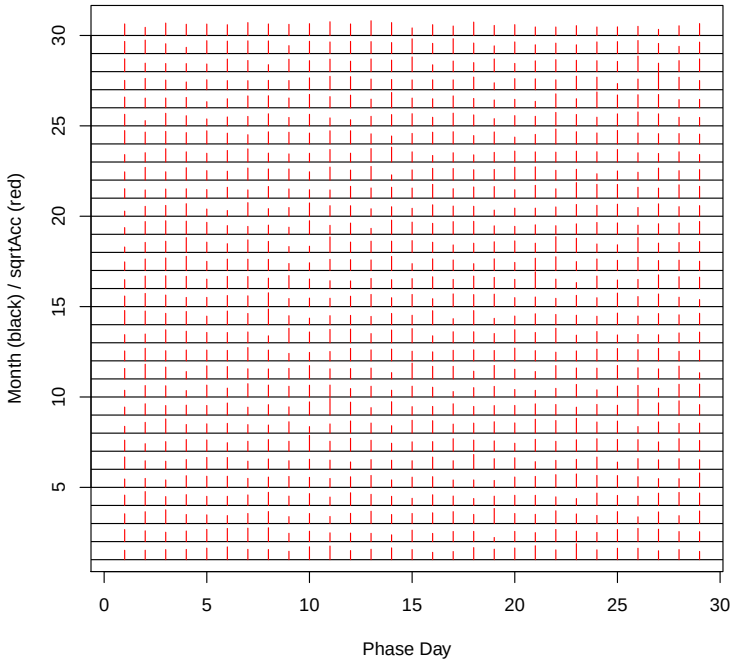
36-402/608 Homework #5 Solutions 2/18

1. Moon phases and behavior (20 points, 5 each)

   There have been many studies of moon phases and behavior. The data for this problem represent the daily accident rate for the US national highway system over 30 lunar months of 29 days each. (Ignore the complication that a lunar month is 29.5 days). Because accidents are counts and because the monthly boxplots showed skewed distributions, the outcome is recoded as the square root of the daily accident rate (sqrtAcc). Here is the raw data:



   (a) Here is an ANOVA comparing the first three weeks of the lunar month (full moon through new moon to first quarter) to the last week (first quarter to full moon):

```
last8Days = factor(day>=22)
summary(aov(sqrtAcc~last8Days))
#              Df Sum Sq Mean Sq F value  Pr(>F)
# last8Days     1    611  611.40   5.982 0.01465 *
# Residuals   868  88717  102.21
fit.contrast(aov(sqrtAcc~last8Days), "last8Days", rbind(LminusF=c(-1,1)))
#                   Estimate Std. Error  t value   Pr(>|t|)
# last8DaysLminusF 1.875634  0.7668781 2.445805 0.01465055
```

What is the standard interpretation of these results?

We reject the null hypothesis that the daily accident rate is the same for the first 3 weeks of a month compared to the last 8 days (p=0.015). Our best estimate is that the square root of the number of accidents is 1.88 higher for the last 8 days. (A CI is a better way to express what we have learned from the study. The CI, standard error, and p-value are all only approximately correct because we have violated the independent errors assumption.)

(b) Now I will tell you that 25 ANOVAs were run, each with a different split of the month, starting with lunar days (1 through 2) vs. (3 through 29), then (1 through 3) vs. (4 through 29), and continuing up to (1 through 26) vs. (27 through 29). Only the ANOVA with the smallest p-value was reported. Why is this procedure inappropriate?

Each additional test gives another chance for a type-1 error, so the overall type-1 error rate is much larger than 0.05.

(c) Knowing that 25 tests were done, state what multiple comparisons correction approach is most appropriate, how you would apply it, and what your new conclusion is.

We can use the Bonferroni correction for multiple testing by using 0.05/25=0.002 as the cutoff p-value for "statistically significant". This will assure that our type-1 error rate is no larger than 0.05. Because $0.015 \geq 0.002$ we retain the null hypothesis that the accident rate does not depend on the phase of the moon.

(Note: These data were simulated completed randomly.)

2. Brick strength (20 points, 5 each)

Bricks were prepared in 10 batches of 10 bricks each with each batch treating two brick with each of five additives, A through E.

```
br=aov(strength~additive+batch, brick)
summary(br)
#             Df  Sum Sq Mean Sq F value    Pr(>F)
# additive     4  605.12 151.280 12.0812 7.752e-08 ***
# batch        9   86.24   9.583  0.7653    0.6485
# Residuals   86 1076.89  12.522

planned = rbind(CDEvsAB=c(-1/2,-1/2,1/3,1/3,1/3),
                AvsB=c(1,-1,0,0,0),
                EvsCD=c(0,0,-1/2,-1/2,1),
                CvsD=c(0,0,1,-1,0))
library(gmodels)
```

```
round(fit.contrast(br, "additive", planned, conf.int=0.95), 3)
#                 Estimate Std. Error t value Pr(>|t|) lower CI upper CI
# additiveCDEvsAB    4.667      0.722   6.461    0.000    3.231    6.103
# additiveAvsB      -0.564      1.119  -0.504    0.616   -2.789    1.661
# additiveEvsCD      2.438      0.969   2.515    0.014    0.511    4.364
# additiveCvsD       0.010      1.119   0.009    0.993   -2.214    2.235

TukeyHSD(br, "additive", ordered=TRUE)
#       diff        lwr      upr     p adj
# B-A 0.5640 -2.5541336 3.682134 0.9867666
# D-A 4.1310  1.0128664 7.249134 0.0034999
# C-A 4.1415  1.0233664 7.259634 0.0033938
# E-A 6.5740  3.4558664 9.692134 0.0000008
# D-B 3.5670  0.4488664 6.685134 0.0166799
# C-B 3.5775  0.4593664 6.695634 0.0162306
# E-B 6.0100  2.8918664 9.128134 0.0000065
# C-D 0.0105 -3.1076336 3.128634 1.0000000
# E-D 2.4430 -0.6751336 5.561134 0.1959570
# E-C 2.4325 -0.6856336 5.550634 0.1995642
```

(a) What are your conclusions about the planned contrasts?

We retain the null hypotheses that the population means of brick strength are equal for additive A vs. B (p=0.616) and additives C vs. D (p=0.993). We reject the null hypothesis of equal strength for the average of CDE vs. AB (p<0.0005) and conclude that the strength of bricks with additives C, D or E is 3.23 to 6.10 units higher than for additives A or B (95% CI). Also, we reject the null hypothesis of equal strength for the average of E vs. CD (p=0.014) and conclude that the strength of bricks with additives E is 0.51 to 4.36 units higher than for additives C or D (95% CI).

Note: if the p-value had been equal to 0.005 or greater, R would have rounded the p-value to 0.001.

(b) What additional contrasts can be reported as statistically significant? D vs. A, C vs. A, E vs. A, D vs. B, C vs. B, and E vs. B.

(c) What p-values should you report for the comparisons of groups B vs. A, D vs. C, and D vs A?

B vs. A: 0.616 and D vs. C: 0.993 from the fit.contrast() results, because they are planned.
D vs. A: 0.0035 from the Tukey table because it is unplanned.

(d) If it now looks interesting to compare AB vs CD, what correction method should you use? Scheffe is appropriate when the possible set of contrasts considered is all linear contrasts.

3. Gene array (10 points, 5 each) A gene array experiment is performed using an analysis that consists of 500,000 t-tests.

   (a) If you were to use a Bonferroni correction, what p-value would you need to be less than to call a comparison statistically significant?

   0.05/500,000=0.0000001 (1e-7)

   (b) What specific problem would this approach lead to? Very low power (so rather than try to protect type 1 error and have a low chance of any false positives, we use the false detection rate approach and control the average fraction of positives that are false positives.

4. Math test (50 points)

   What is the square root of 9? 3