

STAA 554 Homework 3

Contents

1.) 4pts ANOVA procedure	1
2.) 5pts ML estimate	1
3.) 3pts REML procedure	2
4.) 10pts (2pts each) Likelihood Ratio Test via parametric bootstrap.	2
5.) 0 pts (Practice, Solutions will be posted) Analysis of Egg production	3

1.) 4pts ANOVA procedure

Consider the data: pulp.csv

Paper brightness is explored in relation to the shift operator.

Analyze this data as a fixed effects one-way ANOVA using anova estimators. Paste a summary of your model below. (Not shown here, but you can verify the results are equivalent to the ANOVA estimators treating the operator effect as random.)

a. (1 pt) Test the null hypothesis that there is no difference in brightness across shift operators. Report the test statistic and p-value.

b. (1 pt) Provide an estimate of the residual variance.

c. (2 pt) Provide an estimate of the variance of the operator effects. (Hint, consider EMS).

2.) 5pts ML estimate

Estimate the same model as in #1, but using ML and **treating operator as a random effect**. Paste a summary of your output below.

a. (1 pt) Test the null hypothesis that there is no difference in brightness across shift operators. Report the test statistic and p-value.

b. (2 pt) Although the verbiage I used in 2a is the same as in 1a, the test carried out addresses a slightly different null hypothesis. Describe the difference in the null hypothesis statements for the tests carried out in 1a and 2a.

c. (1 pt) Provide an estimate of the residual variance.

d. (1 pt) Provide an estimate of the variance of the operator effects. How does this estimate compare to Part 1c? Comment briefly.

3.) 3pts REML procedure

Estimate the same model as in #1, but using REML and treating operator as a random effect. Paste a summary of your output below.

a. (1 pt) Test the null hypothesis that there is no difference in brightness across shift operators. Report the test statistic and p-value.

b. (1 pt) Provide an estimate of the residual variance.

c. (1 pt) Provide an estimate of the variance of the operator effects. How does this estimate compare to Part 1c? Comment briefly.

4.) 10pts (2pts each) Likelihood Ratio Test via parametric bootstrap.

We can use the parametric bootstrap approach to obtain a more accurate p-value. We need to estimate the probability, given that the null hypothesis is true, of observing an LRT of our observed value or greater. Under the null hypothesis, $y \sim N(\mu, \sigma^2)$. A simulation approach generates data under this model, fits the null and alternative models and computes the LRT statistic. The process is repeated a large number of times and the proportion of LRT statistics exceeding the observed value is used to estimate the p-value. In practice, we do not know the true values of μ and σ , but we can use the estimated values; this distinguishes the parametric bootstrap from the purely simulation approach. The simulate function makes it simple to generate a sample from a model:

- Our null model would be:

```
nullmod = lm(bright~ 1, pulp)
```

- Our alternative model would be as in question 3.
- Calculate the observed LRT:

```
as.numeric(2*(logLik(REMLmod, REML = TRUE) - logLik(nullmod, REML=TRUE)))
```

Use a parametric bootstrap as outlined below to determine the null distribution for the LRT statistic calculated.

1.) Simulate data under the null

```
y = unlist(simulate(nullmod))
```

2.) Fit the null model to the data generated in 1.

3.) Fit the alternative model to the data generated in 1, (use operators designated from pulp data set)

4.) Calculate and store the observed LRT statistic comparing results from 2 and 3.

5.) Repeat steps 1 -4 1000 times.

a. Include a histogram of the LRT statistics generated.

b. Include a summary of the LRT statistics generated (min, 1st quartile, median, mean, 3rd qu., max)

c. What proportion of LRT statistics generated under the null fall above the observed test statistic for the pulp data? (i.e. your p-value).

d. You likely got an error “boundary (singular) fit: see ?isSingular”. Look this up. Why is it not surprising to see this in this setting?

e. Load (install) the package RLRsim. Then run the command `exactRLRT(REMLmod)`, replacing `REMLmod` with the appropriate name of your model. Compare with part c. (Note, this command is only useful for testing random effects).

5.) 0 pts (Practice, Solutions will be posted) Analysis of Egg production

The `eggprod.csv` dataset concerns an experiment where six pullets were placed into each of 12 pens. Four blocks were formed from groups of three pens based on location. Three treatments were applied. The number of eggs produced was recorded.

a. Make suitable plots of the data and comment.

b. Fit a fixed effects model for the number of eggs produced with the treatments and blocks as predictors. Determine the significance of the two predictors and perform a basic diagnostic check.

c. Fit a model for the number of eggs produced with the treatments as fixed effects and the blocks as random effects. Which treatment is best in terms of maximizing production according to the model? Are you sure it is better than other two treatments?

d. Use the Kenward-Roger procedure for an F-test to check for differences between the treatments. How does the result compare to the fixed effects result?

e. Perform the same test of the fixed effect for treatment, but using a bootstrap method with LRT as outlined in question 4. How do the results compare to part d?

f. The parametric bootstrap method can be implemented using `PBmodcomp(modFULL, modREDUCED)`. Run this command and compare the results.

g. Test for the significance of the blocks. Does the outcome agree with the fixed effects result? (Use `exactRLRT()` as in 4e).