

Baseball Research Log

Chris Comiskey

Summer 2017

June 21, 2017 - Summer Solstice!!

- Houston, we have a problem. For some reason, my package as I have it written now is eating a few observations each iteration. I don't know why. I am focusing on the first iteration, from one box to four, where I lost one observation. I think I need to record that puppy, and see which observation I am losing, see what its key attribute is, figure out how to fix it. I am using `as.image(...)`, which I think I have isolated as the step that loses the observation. Sigh.
- Oh yeah, I got a 0.6 FTE job offer from Reed. :)

Thursday, 29 June 2017

- Confidence intervals for \hat{p} for Shiny app...
 1. $\widehat{\log\left(\frac{p}{1-p}\right)} \pm m * SE\left[\widehat{\log\left(\frac{p}{1-p}\right)}\right]$
 2. Back transform
 3. $\hat{p} \pm M * SE(\hat{p})$
- This is important b/c otherwise there can be CIs for \hat{p} that include negative numbers, and thus incorrectly calibrated CI coverage.
- To check if $\widehat{\log\left(\frac{p}{1-p}\right)}$ is actually Normally distributed, a simulation study would proceed as follows.
 1. Fix $X\beta \rightarrow$ fix p .
 2. Generate Bernoulli(p) spatial data.
 3. Estimate everything; β s, thus logit(p), and thus p
 4. Iterate (2) and (3)
 5. Histograms: $\widehat{\text{logit}(p)}$ vs. \hat{p}

July 14, 2017

- My second package: `mapapp`
- **TO DO:** make add-on slides regarding IRLS and thus $SE(\beta)$, $SE(\hat{p})$, $SE(\text{logit}(\hat{p}))$. The Shiny app is a big part of what I'm doing, so likely question. See "supplementary slides" below.

Supplementary Slides

Package Questions

- How can I delete `/ResearchRepo/varyres` since I have my `varyres` repository now? Also, delete `/ResearchRepo/images`?
- How do I load my package when **not** working in package project?
- My packages seem to consist of one function each. This okay? A & C: YES.
- Plus, see Hilary Parker below!!

Hilary Parker

“...the best products are built in small steps, not by waiting for a perfect final product to be created. This concept is called the minimum viable product ([link to Wikipedia page](#)) its best to get a project started and improve it through iteration. R packages can seem like a big, intimidating feat, and they really shouldnt be. **The minimum viable R package is a package with just one function!**”

Meeting - 20 July 2017

- Use label and ref to so that figures and references to them line up
- **TO DO:** go back through Ch1 and Ch2 and add label and ref to prevent errors.
- Need to add (3 page) conclusion: what did/said, next steps
- Each chapter needs its own conclusion.
- Need to add packages
- Need to add language about packages in intro, roadmap, etc.
- Include “plan to add .gif to package” to Ch 3
- Don’t worry about Shiny app spacing; fudge it
- **TO DO:** PITCHf/x, tornado data okay to use? (PITCHf/x - basically (one more email)) (Tornado - YES)
- Defense presentation - basically my seminar talk; add `varyres(...)` with the picture I show, a slide here and there
- The previous item allows more time to sharpen package
- The package docs: (i) Show A & C pre-diss, and (ii) Include in diss.
- Modify vignette, to cater to people using their own data.
- Need to get on `mapapp`. Need to have something to iterate with.

PPM Fit specs

- 97 KNOTS =====
- 49 KNOTS =====

1 Meeting, 27 July 2017

- TO DO: email **Ch.4** to them both; hard copy also to Alix (Charlotte away this week)
- Always send Alix .pdf, b/c she wants color version always
- Should mention who invented heat maps, when they started, etc.
- Heat map history: Try Michael Friendly, "Milestones in... Statistical Graphics."
- Use "**box area**" instead of "box size" (b/c confusing with sample *size*)
- Data density is bad; data —?— instead (??)
- "Box 22" has got to go; Instead, show picture of labeled map, then use those labels.
- Add (A), (B), (C) to six map figures
- Remove numbering from equation unless referencing them later (like acronyms)
- Include some cool picture to supplement, help sell the Hosmer-Lemeshow test. Perhaps p vs. \hat{p}
- Need to cite R packages, at least the first time in each chapter
- Consult (an article in) "R Journal..." (??) for guidelines on referencing packages, or even referring to them by name.
- Ch.2 needs an introduction
- **Include pseudo-algorithm** Fun! Include: stopping rule, and subdivision method
- Include alternate stopping rules, subdivision methods in "Future VR work" section
- Add oceans to tornado example, to kill the huge overlap boxes eyesore.
- INDEPENDENCE ASSUMPTION. Need to emphasize it is an assumption, b/c correlated within pitcher, correlated within year, day, etc. Need to talk more about it when I declare it, because it's a BIG ONE. Express caution.
-

As of Now

- Jobs: ODG, Simple, Charlie!
- My edits, Betsy edits, A & C Friday deadlines
- Make packages
- Webpage
 - Talks (JSM, seminar)
 - RMarkdown, ST516 & ST517 handouts
 - Consulting reports/summaries