**How to generate effort and catch estimates using the “CreelEstimates” code repository**

*Background*

This document outlines the steps for estimation of freshwater fishery angler effort and catch from survey data collected during Washington Department of Fish and Wildlife (WDFW) freshwater sportfishing monitoring (creel) projects, using a scripted analysis in the CreelEstimates [code repository](https://github.com/wdfw-fp/CreelEstimates), hosted on the WDFW fish program’s Github account. It is applicable to creel projects which adhere to study design guidelines outlined in the [WDFW Statewide Protocol for Roving Creel Surveys](https://stateofwa.sharepoint.com/:w:/r/sites/DFW-TeamFPFWCreelMonitoringProgram163/_layouts/15/Doc.aspx?sourcedoc=%7B83F2B81F-39FC-49E0-955B-B6E9145AB8B7%7D&file=Protocol%20-%20Statewide%20roving%20creel%20surveys%20(Part%201%20of%202)%20-%20STUDY%20DESIGN%20(in%20progress).docx&action=default&mobileredirect=true).

CreelEstimates is a code repository with a template script that provides a workflow to fetch raw observed data, generate intermediate data summaries, produce expanded point estimates with associated uncertainty measures, and provide supporting tables and figures. Data used in the scripted analysis is fetched from publicly available data views at [data.wa.gov](https://data.wa.gov/browse?q=creel) and documented in the [WDFW creel database data dictionary](https://dfw-fp-r5.s3.us-west-2.amazonaws.com/data-dictionaries/creel_data_dictionary.html#Database_Overview).

*Methods*

Data used in the analysis is specified by fishery name, a database field which describes the area, start and end dates, and one or more fish species of interest. The process of generating creel estimates relies on three primary data components: 1) index effort counts, 2) census effort counts, and 3) interviews. These observed data are used to generate estimates of angler effort and catch, stratified by angler-type, day-type, section, and time-period, for open legal fishing days within the specified monitoring period (table 1). 1) Index effort count data are collected at a discrete set of locations in the fishery area, either with direct counts of anglers or indirect counts of objects assumed to represent anglers (e.g., vehicles, boat trailers). 2) Census effort count data are collected on a subset of survey days as a continuous count of anglers observed within a section, simultaneous to a scheduled index effort count. Census effort count data are used to assess the degree of bias (undercount or overcount of anglers) in angler counts derived from index effort counts. 3) Interviews are conducted at any location within a section where anglers may be encountered and are used to estimate their catch rates of all encountered fish species.

CreelEstimates currently provides two analysis options to generate creel estimates: a 1) Point Estimate (PE) method, which uses fixed (i.e., deterministic) equations as a model to estimate catch, and 2) a Bayesian State Space (BSS) method, which uses a state-space model to estimate catch while accounting for observation error.

*Point Estimate method*

The point estimate (PE) method for generating estimates of angler effort and catchfor stratum with modified “direct expansion estimators” (Pollock et al. 1994; Thompson 2002), where a sample mean with a sample size *n* is expanded to a population total *N*. The method is computationally simple and fast, producing estimates of effort and catch using a modified form of established methods for analyzing creel data (Pollock et al. 1994; Hahn et al. 2000). Because this method does not use numerical approximation and produce associated measures of model validation, it will provide “warning-free” estimates of effort and catch for “low catch” datasets, a use case where the frequency of catch for a given catch group is rare (i.e., reported catch < 30 fish). It should be noted that estimates generated in this context should be evaluated with caution and in consultation with a study design lead. The study design and associated levels of stratification inherent in the roving survey creel survey protocol limit the current utility of the PE method to provide accurate and unbiased measures of uncertainty (insert help from TB/KB).

Issues with uncertainty measures for PE:

* + Confidence interval widths biased low
  + ignore error propagation from census surveys (and other sources?)
  + sample size constraints due to stratification (e.g., day-type and time-period) cause instability in sample variance or inability to calculate variance (i.e., *n* = 1)

*Bayesian State Space (BSS) Method*

The functions used to summarize, aggregate, and model effort and catch in CreelEstimates require formatting consistent with WDFW’s [creel database](https://dfw-fp-r5.s3.us-west-2.amazonaws.com/data-dictionaries/creel_data_dictionary.html#Database_Overview) and “proofed” data (i.e., reviewed for data collection / entry errors). Before proceeding to the instructions portion of this document, review and complete the checklist of analysis preparation steps (table ).

*Instructions*

The following steps provide an overview of how to generate estimates of effort and catch from creel surveys using the CreelEstimates tool. If additional guidance is needed, contact your Study Design Lead.

1. **Sync Teams folder(s) and map to T: network drive**

Sync the Team “DFW-Team FP FW Creel Monitoring Program – General” (and other Teams folders as desired) to your computer with the following steps. This process is required when saving output from the CreelEstimates script to a Teams network drive location.

* + Navigate to the General channel in the team “DFW-Team FP FW Creel Monitoring Program – General” in Teams. Click on “Files” and then click “Sync”. This will create a synced copy of the Teams folder in the “Washington State Executive Branch Agencies” location within “C:\Users\username”
  + Go to "This PC" in File Explorer and click "Map network drive"
  + In the prompt, choose the letter T, and in the "Folder" box enter \\localhost\C$\Users\youruser\Washington State Executive Branch Agencies
  + Click okay to complete things. You can confirm completion by opening Rstudio or a terminal window and typing "T:/" then hit tab - you should get a little popup of the available synced directories from Teams.

1. **Clone the “CreelEstimates” repository**
   * The most up-to-date version of CreelEstimates is located on GitHub [here](https://github.com/wdfw-fp/CreelEstimates) and is stored as a code [repository](https://docs.github.com/en/repositories/creating-and-managing-repositories/about-repositories)
   * To use CreelEstimates, the repository must be cloned to your local computer (for further details see “[Clone the Repository with RStudio](https://resources.github.com/github-and-rstudio/)” sub-section). The repository can be located anywhere on your computer since you’ll interact with R script (.R) and R Markdown (.Rmd) files that run within a RStudio project file (.Rproj) that use relative filepaths within the cloned project folder.
   * Before a repository can be downloaded, a user must have a GitHub account (see “[Prerequisites](https://resources.github.com/github-and-rstudio/)” subsection) and have Git and RStudio installed on their computer (see “[Install Git and RStudio](https://resources.github.com/github-and-rstudio/)”)
2. **Download Rstan package and C++ Toolchain**
   * The Bayesian state-space creel model is run in [stan](https://mc-stan.org/)
   * Running stan models in R(Studio) requires the package Rstan and a C++ toolchain software to be installed
   * Follow the instructions provide on the [RStan Getting Started](https://github.com/stan-dev/rstan/wiki/RStan-Getting-Started) webpage
   * Pay particular attention to the “Verifying Installation” and dealing with compilation errors related to versions of R and Rtools
3. **Create a new analysis project** 
   * Open the “CreelEstimates” folder
   * Open the “CreelEstimates.Rproj” file. This opens an instance of R Studio within the self-contained working environment of the “CreelEstimates” folder with associated scripts and data.
   * Open the “establish\_analysis.R” file. This script creates a renamed copy of the fw\_creel.Rmd analysis script and saves it within a folder for the specified analysis\_name (fishery\_name, est\_date\_start, est\_date\_end) within a folder specifying the project\_name (a database field that groups creel studies, generally by work unit).
     1. Enter the values for the following parameters:
        1. project\_name
        2. fishery\_name
        3. est\_date\_start
        4. est\_date\_end
        5. output\_location\_filepath
        6. output\_teams\_name – default value = “"DFW-Team FP FW Creel Monitoring Program - General"
     2. run the script.
4. **Run the analysis script** 
   * Navigate to the new folder for your analysis at: project\_scripts/*project\_name*/*analysis\_name*

and open the file *analysis\_name*.Rmd. This file is a parameterized report, where most user inputs are entered as control parameters. Consequently, the script does not require, and by design discourages, direct editing of code chunks in the R Markdown file and .R scripts containing source functions .

* + Specify the metadata parameters of the script:In the [YAML](https://yaml.org/spec/) metadata section at the top of the script, enter parameter values which specify the data and analysis control options for the .Rmd script, including the same values for the parameters used in the establish\_analysis.R script (table ). After updating parameters, click on the Session drop down menu and click “Restart R”. This step is required to update the params used in the script.
  + Interactively run the script
    1. The fw\_creel.Rmd script is broken up into code chunks.
    - The code is designed to be run from top to bottom. The user can do this “chunk by chunk”, running each chunk in successive order. Alternatively, the entire script can be run with the “knit” button, which runs all code chunks, exports output files, and renders a report in .html format.
    - The main script seen in the .Rmd file uses functions contained in .R scripts in the R\_functions folder. These scripts “run behind the scenes” to fetch, summarize, and generate estimates from creel data. Feel free to open these source files to understand how CreelEstimates is working. This code should not be edited unless you establish a formal plan to do this using a distributed version control system (i.e., create a branch to CreelEstimates, make edits to the code, commit, and push the changes to your branch of the repo, and submit a pull request to tell collaborators about the changes in your branch
  + **Run the analysis**