D. Rの基礎知識

Rの基礎知識

- Rとは?
- エディタ:Rstudio
- •Rプログラムの基本形(関数・引数・オブジェクト)
- ワーキングディレクトリの設定
- 「生のR」と「tidyverse」の違い(応用)

Rとは



You

プログラミング言語のRとはどのような言語ですか?

(

ChatGPT

Rは統計計算やデータ解析のために設計されたプログラミング言語および環境です。以下は、Rの主な特徴です:

- 1. 統計分析とデータ処理: Rは統計学的な計算やデータ解析のために広く使用されます。統計的手法や機械学習アルゴリズムを実装し、データの可視化や処理を行うことができます。
- 2. オープンソース: Rはオープンソースであり、無料で利用できます。 これは、コミュニティによって広くサポートされ、拡張や改良が進んでいることを意味します。
- 3. **豊富なパッケージとライブラリ:** Rにはさまざまな統計的手法やデータ可視化のためのパッケージやライブラリが豊富に存在します。これらのパッケージを使用することで、特定のタスクに対する高度な機能を追加することができます。
- 4. ベクトル演算: Rはベクトル演算をサポートしており、同じ操作を複数のデータポイントに対して一度に実行できます。これにより、効率的かつ簡潔なコードの記述が可能です。
- 5. グラフィカルなデータ表示: Rはデータの視覚化に強みがあり、さまざまなグラフやプロットを作成することができます。ggplot2などのパッケージを使用することで、高度なデータ可視化が可能です。
- 6. **スクリプティング言語:** Rはスクリプティング言語であり、コードをスクリプトとして保存し、再利用することができます。これにより、繰り返しのタスクやデータ分析の自動化が容易になります。

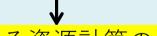
Rは主に統計家、データサイエンティスト、研究者、およびデータ解析を行う他の分野の専門家によって広く利用されています。

◆

我が国の資源評価のためのパッケージ(frasyr, frasyr23, frapmr)

資源計算、再生産関係推定、 ABC計算、結果の図の作成

一連の手順を統一スクリプトに して配布(frasyr_tool)



Rによる資源計算の演習

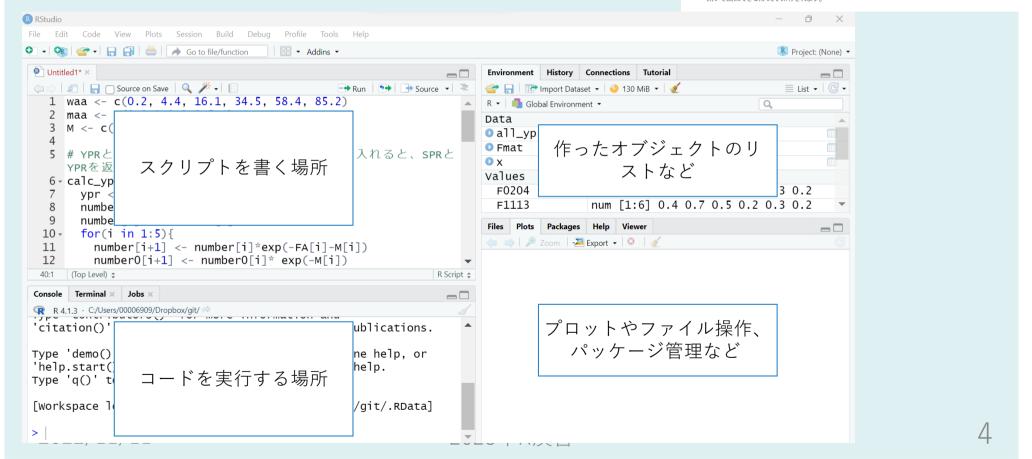
エディタ (Rstudio): スクリプトを書き実行するためのソフ

M You

スクリプトって何ですか?

ChatGPT

スクリプト(Script)は、一連のコンピュータプログラムの命令や手順がテキスト形式で記述されたファイルのことを指します。スクリプトは通常、特定のタスクや機能を実行するためのプログラムであり、解釈型の言語で書かれることがよくあります。スクリプトは通常、人が理解しやすく、テキストエディタで開いて編集できる形式で保存されます。



Rプログラムの基本形

> sumtest1 <- sum(1:10)

オブジェクト (生成されたデータ や情報が入ってい る)

〈一 代入記号 (=も可) 関数と引数(命令)

- 関数が実行されて、無事代入が終わると、「オブジェクト」ができる
- 「オブジェクト」には関数の実行結果が保存されている
- 「引数」は関数に与える様々な「オプション」
- 作ったオブジェクトを別の関数に入れて再計算を繰り返す手順=スクリプト

Rの「オブジェクト」の型

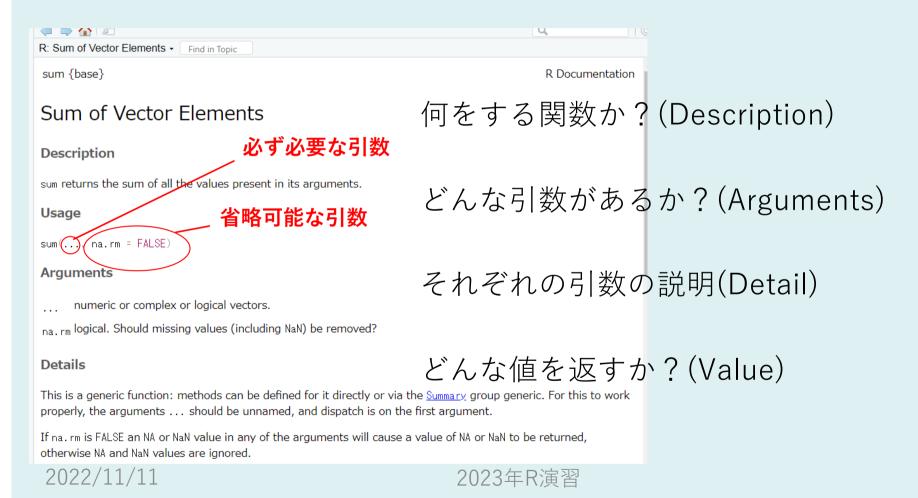
```
x1 <- c(2,3,4,5) # ベクトル
x2 <- matrix(1:6,2,3) # 行列
x3 <- data.frame(x=1:2, y=c("a","b")) # データフレーム
x4 <- tibble(x=1:2, y=c("a","b")) # データフレーム(tidyverse)
x5 <- list(x1, x2, x3, x4) # リスト
x6 <- lst(x1, x2, x3, x4) # リスト(tidyverse)
```

- 複雑な計算をする関数の返り値はたいてい「リスト」の中に、いろいろな型のオブジェクトがつまってる
- リストの中の要素を見るには、\$を使う (リストx6の中のx1を見たい → x6\$x1)
- names(x6)とすれば、リストの中にどんな名前のオブジェクトが入ってるか確認 できる

演習① 次の関数を使い、結果をオブジェクトに入れよう

- 1. sum (足し算をする)
 - •1から10の足し算をし、結果をオブジェクトに格納
- 2. rnorm (正規乱数を発生させる)
 - ・正規乱数を100個発生させ、結果をオブジェクトに格納
- 3. hist (ヒストグラムを作成する)
 - 上で発生させた正規乱数をヒストグラムとしてプロット

Helpの見方 (sum)



helpの見方 (rnorm)

The Normal Distribution

Description

Density, distribution function, quantile function and random generation for the normal distribution with mean equal to mean and standard deviation equal to sd.

Usage

```
dnorm(x, mean = 0, sd = 1, log = FALSE)
pnorm(q, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)
qnorm(p, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)
rnorm(n, mean = 0, sd = 1)
```

Arguments

```
    x, q vector of quantiles.
    p vector of probabilities.
    n number of observations. If length (n) > 1, the length is taken to be the number required.
    mean vector of means.
```

helpの見方 (hist)

Description

The generic function hist computes a histogram of the given data values. If plot = TRUE, the resulting object of class "histogram" is plotted by plot.histogram, before it is returned.

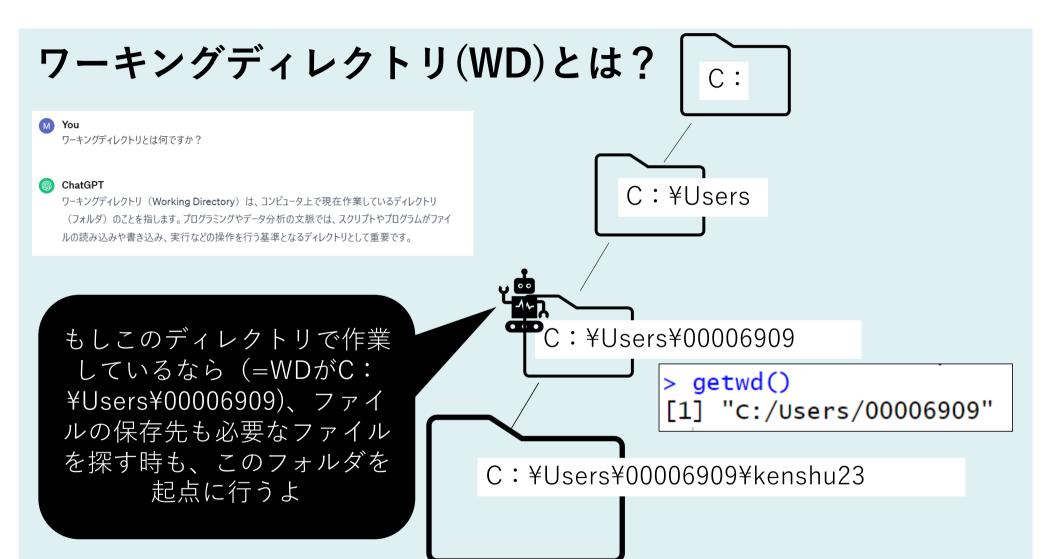
Usage

```
hist(x, ...)
## Default S3 method:
hist(x, breaks = "Sturges",
    freq = NULL, probability = !freq,
    include.lowest = TRUE, right = TRUE,
    density = NULL, angle = 45, col = "lightgray", border = NU
    main = paste("Histogram of" , xname),
    xlim = range(breaks), ylim = NULL,
    xlab = xname, ylab,
    axes = TRUE, plot = TRUE, labels = FALSE,
    nclass = NULL, warn.unused = TRUE, ...)
```

Arguments

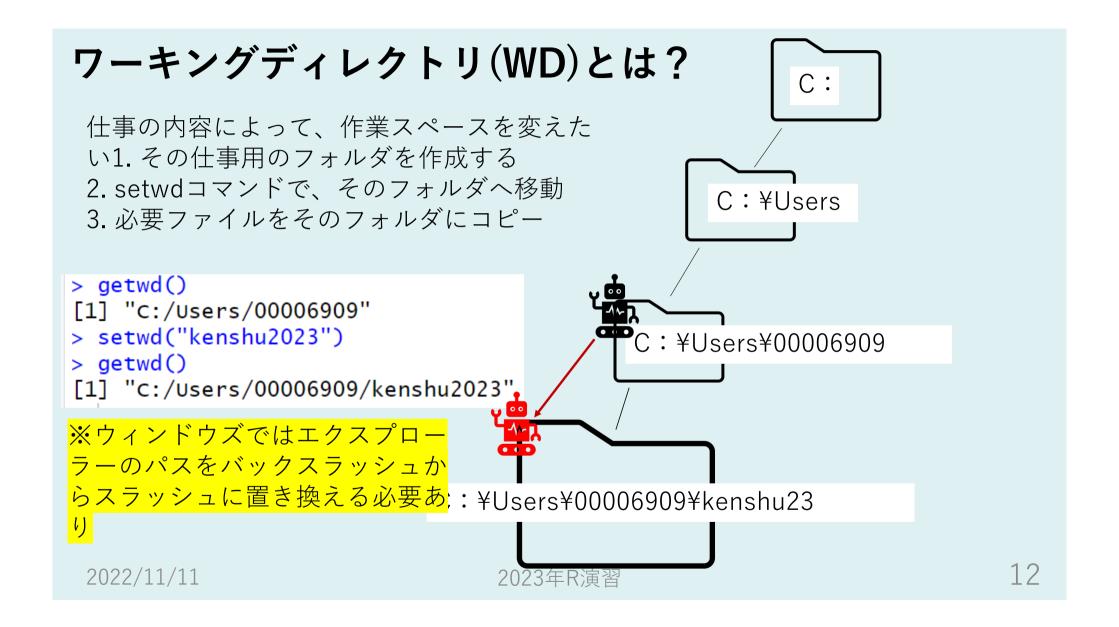
a vector of values for which the histogram is desired.

breaks one of:



2022/11/11

2023年R演習



```
# 演習2 ----
# 1. 以下の関数を編集して、自分のワーキングディレクトリを設定しましょう
# 2. また、配布した疑似データ・スクリプトをこのワーキングディレクトリ下
に保存しましょう

# wdの移動
setwd("c:/Users/00006909/kenshu2023")
# ちゃんと設定できたか確認
getwd()
# ファイルがちゃんと置けているか確認
dir()
```

(応用)「生のR」とtidyverse

M You tidyverseとは何ですか?

ChatGPT

Tidyverse(タイディバース)は、統計解析やデータ操作、可視化などのためのR言語のパッケージの集まりです。Tidyverseは、データ分析のプロセスを簡潔で効果的に行うために設計されており、一貫性のある文法やデータ構造を提供します。Hadley Wickhamとその共同開発者たちによって開発されました。

Tidyverseの主なパッケージには以下が含まれます:

- 1. ggplot2: 高品質なグラフィクスを生成するためのパッケージ。
- 2. dplyr: データのフィルタリング、整形、結合などのデータ操作を行うためのパッケージ。
- 3. tidvr: データの整形や欠損値の処理を行うためのパッケージ。
- 4. readr: テキストデータやCSVファイルなどを効率的に読み込むためのパッケージ。
- 5. purrr: データの操作や関数の適用をベクトル化するためのパッケージ。
- 6. tibble: データフレームの改良版で、表示や処理が改善されています。

これらのパッケージは、統一されたデータモデルと文法を提供し、データサイエンスや統計解析のプロセスをより効果的に行うのに役立ちます。TidyverseはR言語の生態系を豊かにし、データ分析や可視化の作業をより直感的で効率的に進めることができるようになります。

表:tidyverseが提供するパッケージと代表的な機能

パッケー ジ名	説明	代表的な関数(括弧内は、tidyverseを使わない場合のRの関数)
tibble	より柔軟な形式のデータフ レーム tibble 等を提供	データフレームを作成: tibble (data.frame), データをtibble形式に変換: as_tibble , リストを作成: 1st (list)
tidyr	データを横長データから縦 長データへ、縦長データか ら横長データへ変換する	横長データを縦長データに変換する: gather (as.data.frame.table), 縦長データを横長データに変換する: spread (tapply)
dplyr	データを整形したり、変換 したりする	列を抽出: select , 行を抽出: filter (subset), 列を追加または変更: mutate , データの並び替え: arrange (sort , order), パイプ演算子: %>% (詳細はBOX)
readr	データの読み込みと出力。 読んだデータはtibble形式と なる。	<pre>read_table (read.table), read_csv (read.csv), write_table (write.table), write_csv (write.csv)</pre>

stringr	文字列操作	str_ で始まる関数群。文字列結合: str_c (paste)
forcats	カテゴリカル変数を取扱う	fct_ で始まる関数群。他の変数を使ってカテゴリカル要素の順番 (level)を付け替える: fct_reorder, 登場頻度が多い(少ない)順に levelを付け替える: fct_infreq
purrr	ループ処理やapply系関数	map_ で始まる関数群。各要素に関数を適用: map (lapply), mapしたのちにdata.frameとして結合: map_dfr (sapply)
ggplot2	グラフ作成 (詳細は後日)	ggplot, geom_ で始まる関数群

※ 資源管理研修、R統計研修、frasyrの関数の中身は生のRとtidyverseが混在しています! https://github.com/ichimomo/main/tree/master/tidyverse#readme

演算子 %>%

matrix(0,4,4) Y sum Z 中間生成物 ほしい結果

#生のRを使う場合

Y < -matrix(0, 4, 4)

Z < - sum(Y)

生のRで中間生成物Yをつくらないやり方 Z <- sum(matrix(0, 4, 4))

余計な中間生成物 (Y)を作らずに、かつ、matrixを作った後にsumをする、 という手順がわかりやすい

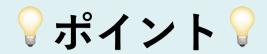
#%>%を使ったやり方

 $Z <- \frac{matrix(0, 4, 4)}{\%} \% > \% sum()$

%>%の前のオブジェクトが%>% の関数の第一引数になる

2023年R演習

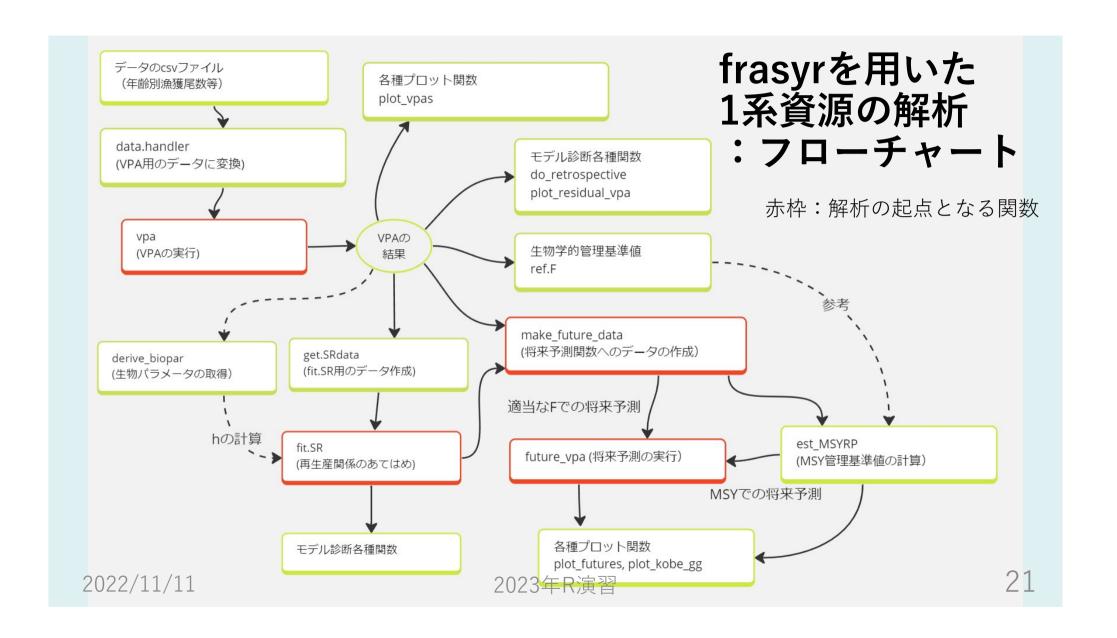
```
# 生のRとtidyverseの違い
                              $(列の指定)や「,](列、行の指定)を使った操作が多い
# 生のR
datA <- data.frame(x=1:10, y=rnorm(10)) # datAを作る
datA$z <- datA$x * 2 # zという列を足す
datA2 <- datA[datA$x>5, ] # xが5よりも大きい行をとりだす
plot(datA$x, datA$y)
                        最初にtidvverseが呼び出される
# tidyverse
library(tidyverse)
                                         %>%という演算子が登場
datB <- tibble(x=1:10, y=rnorm(10)) %>%
         mutate(z=x*2) # datBを作ってzという列を足す
datB2 <- mutate(datB, x>5) # datBでxが5よりも大きい行をとりだす
datB %>% ggplot() +
                               プロットするときはggplot() +から始ま
 geom_point(aes(x=x, y=y))
```



- 用いたい関数が何をするのか・どんな結果を返すのか・どん なオプションが可能なのかをチェックしよう
- 常にWDを意識して作業しよう
- 読みたいRコードが、生のRなのか、tidyverse構文なのか意識 しよう

D. Rを使った演習

- 1. データの読み込み
- 2. VPAのあてはめ
- 3. 再生產関係
- 4. 管理基準値と将来予測

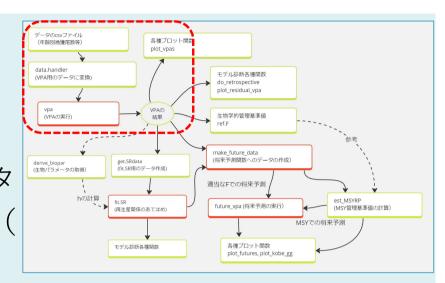


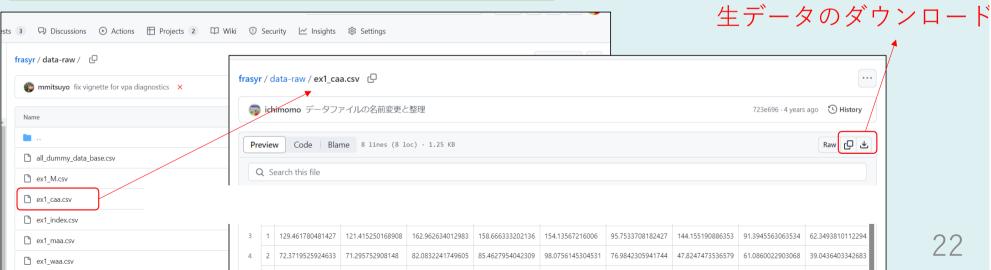
1. データの読み込み

・用意するもの:年齢別漁獲尾数・体重・ 成熟率・(あれば)資源量指数のデータ

データのフォーマット例

https://github.com/ichimomo/frasyr/tree/dev/data-raw



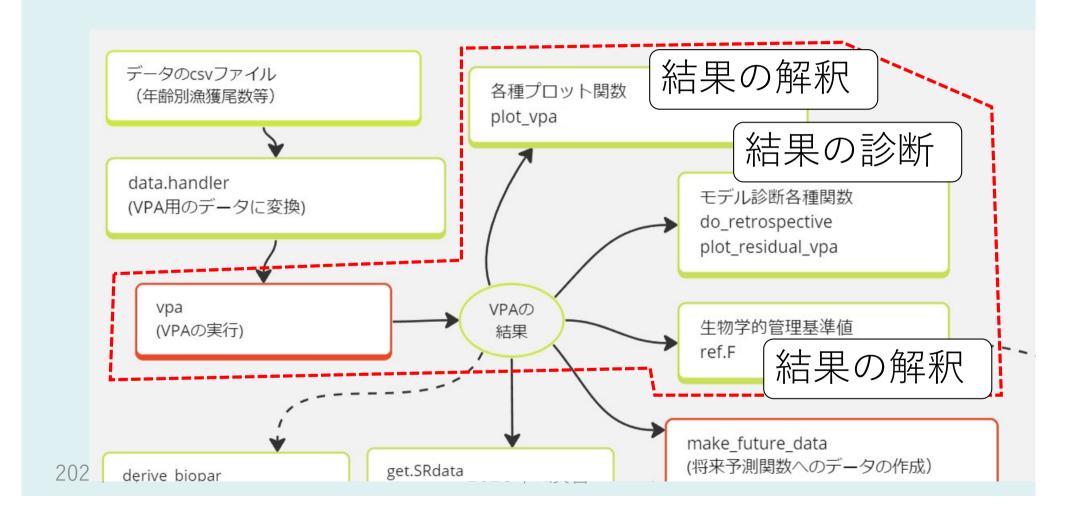


```
# 演習③ データを読み込んでVPAを一回実行しよう
# - 配布データされたcsvファイルを読み込んでVPAを一回実行してみてください
setwd("c:/Users/00006909/kenshu2023")
                                         自分の環境にあわせて変更してください
library(frasyr)
library(tidyverse)
                            必要なライブラリの読み込み
library(patchwork)
# データの読み込み ----
# 年齡別漁獲尾数
caa <- read.csv("ex1_caa.csv", row.names=1)</pre>
                                                配布したサンプルデータの読み込み
# 年齡別体重
waa <- read.csv("ex1 waa.csv", row.names=1)</pre>
# 年齡別成熟率
                                                 frasyrの関数の利用
     <- read.csv("ex1 maa.csv", row.names=1)</pre>
# 資源量指数
index <- read.csv("ex1 index.csv", row.names=1)</pre>
dat vpa <- data.handler(caa=caa, waa=waa, maa=maa, M=0.5, index=index)</pre>
res_vpa <- vpa(dat_vpa) # 引数をつけないで実行=デフォルト引数が使われる
                                                                     23
2022/11/11
                               2023年R演習
```



- 新しく自分の解析を始める場合でも、既存のデータ(動くもの)をダウンロード→変えるべき部分を変更するのが良い
- Mが年・年齢によって一定の場合には、ファイルを用意せず data.handlerの引数 (M=XXX) で指定しても良い
- ・省略可の引数を指定しない場合には、デフォルト引数が使われる→どんな引数がデフォルトとして設定されているのか、確認するのが大事

VPAの実行&プロット&モデル診断



vpa関数の全貌

```
> args(vpa)
function (dat. sel.f = NULL. tf.vear = 2008:2010. rec.new = NULL.
    rec = NULL. rec. vear = 2010. rps. vear = 2001:2010. fc. vear = 2009:2011.
    last.vear = NULL, last.catch.zero = FALSE, faa0 = NULL, naa0 = NULL,
   f.new = NULL, Pope = TRUE, p.pope = 0.5, tune = FALSE, abund = "B",
   min.age = 0, max.age = 0, link = "id", base = NA, af = NA,
    p.m = 0.5, omega = NULL, index.w = NULL, use.index = "all",
    scale = 1000, hessian = TRUE, alpha = 1, maxit = 5, d = 1e-04,
   min.caa = 0.001, plot = FALSE, plot.year = NULL, term.F = "max",
    plus.group = TRUE, stat.tf = "mean", add.p.est = NULL.
    add.p.ini = NULL, sel.update = FALSE, sel.def = "max",
    max.dd = 1e-06, ti.scale = NULL, tf.mat = NULL, eq.tf.mean = FALSE,
    no.est = FALSE, est.method = "ls", b.est = FALSE, est.constraint = FALSE,
    q.const = 1:length(abund), b.const = 1:length(abund), q.fix = NULL,
    b.fix = NULL, sigma.const = 1:length(abund), fixed.index.var = NULL,
    max.iter = 100, optimizer = "nlm", Lower = -Inf, Upper = Inf,
    p.fix = NULL, lambda = 0, beta = 2, penalty = "p",
    ssb.def = "i", ssb.lag = 0, TMB = FALSE, sel.rank = NULL,
    p.init = 0.2, sigma.constraint = 1:length(abund), eta = NULL,
    eta.age = 0, tmb.file = "rvpa_tmb", remove.abund = NULL,
    madara = FALSE, p_by_age = FALSE, penalty_age = NULL, no_eta_age = NULL,
    sdreport = FALSE, use.equ = "new", ave_S = TRUE)
```

help(vpa)

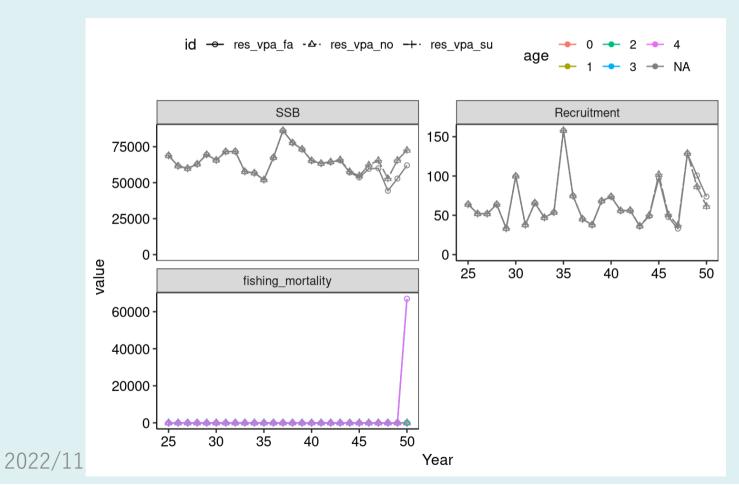


```
# 演習④ いろいろな設定のVPAを試そう
# - 配布された"dat_vpa1.rds"ファイルをreadRDSコマンドで読み込もう
# - どんなデータか確認しよう
# - いろいろなVPAの設定を試して、結果を比較しよう
#
#+
dat vpa <- readRDS("dat vpa1.rds")</pre>
#dat_vpa <- readRDS("dat_vpa2.rds")</pre>
#dat vpa <- readRDS("dat vpa3.rds")</pre>
#dat vpa <- readRDS("dat vpa4.rds")</pre>
# データの確認
par(mfrow=c(2,2))
matplot(t(dat vpa$caa), type="b")
matplot(t(dat_vpa$index[1,]), type="b")
matplot(t(dat_vpa$index[2,]), type="b")
```

まず、どんなデータをこれから解析するのか、 確認してみましょう

```
# tuningなし
res_vpa_no <- <pre>vpa(dat_vpa, tf.year=45:49, tune=FALSE,
                  p.init=0.4,Pope=FALSE,fc.year=48:50)
plot vpa(res vpa no, what.plot=c("SSB","Recruitment","fishing mortality"))
# 選択率更新法
res vpa su <- vpa(dat vpa, tf.year=45:49, sel.f=NULL,
                tune=TRUE, sel.update=TRUE,
                abund=c("N","SSB"),min.age=c(0,0), max.age=c(0,5),
                p.init=0.4, Pope=FALSE, fc.year=40:45)
plot vpa(res vpa su, what.plot=c("SSB","Recruitment","fishing mortality"))
# 全F推定
res_vpa_fa <- <pre>vpa(dat_vpa, tf.year=NULL, sel.f=NULL,
                tune=TRUE, sel.update=FALSE, term.F="all",
                abund=c("N", "SSB"), min.age=c(0,0), max.age=c(0,5),
                p.init=0.4,Pope=FALSE,fc.year=40:45)
plot vpa(res vpa fa, what.plot=c("SSB","Recruitment","fishing mortality"))
```

3つの結果の比較



```
# ### 演習⑤ モデル診断しよう
# - 残差プロット・レトロスペクティブ解析・ブートストラップ解析を行って、
# 3つのモデルを比較しよう

# 残差のプロット
gg <- plot_residual_vpa(res_vpa_su)
wrap_plots(gg[1:3], ncol=1)

gg <- plot_residual_vpa(res_vpa_fa)
wrap_plots(gg[1:3], ncol=1)
```

2022/11/11

```
# レトロスペクティブ解析
res retro no <- do retrospective vpa(res vpa no)
res retro su <- do retrospective vpa(res vpa su)
res_retro_fa <- do_retrospective_vpa(res_vpa_fa)</pre>
# Mohn's rho: 正なら過大傾向気味、負なら過小推定気味
bind rows(res retro no$mohn rho,
          res retro su$mohn rho,
          res_retro_fa$mohn_rho)
# ブートストラップ
res_boot_su <- plot_resboot_vpa(res_vpa_su, B_ite=30)</pre>
res boot fa <- <pre>plot_resboot_vpa(res_vpa_fa, B_ite=30)
wrap_plots(res_boot_su[c("plot_ssb","plot_rec","plot_biomass")],ncol=2)
wrap_plots(res_boot_fa[c("plot_ssb","plot_rec","plot_biomass")],ncol=2)
                                                                   32
 2022/11/11
                              2023年R演習
```

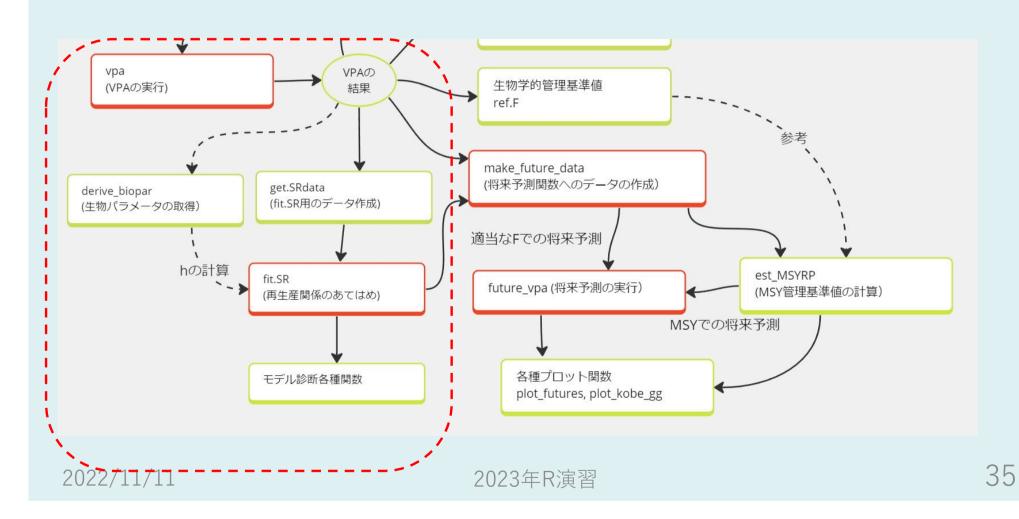
```
# 演習⑥ 資源の状態を把握しよう
# - SPR, YPRなどの管理基準値を計算し、現状の漁獲圧がどのくらいなのか
     (高いのか・低いのか) 判断してみよう
# 生物学的管理基準値の計算
F4850 <- rowMeans(res vpa su$faa[, as.character(48:50)])
res ref4850 <- ref.F(res vpa su, Fcurrent=F4850)
F4045 <- rowMeans(res_vpa_su$faa[, as.character(40:45)])
res ref4045 <- ref.F(res vpa su, Fcurrent=F4045)
res ref4850$summary
res ref4045$summary
aa <- get.SPR(res_vpa_su)</pre>
plot(aa$ysdata$perSPR)
```

2022/11/11 2023年R演習 33

- # 演習⑦ どのチューニング手法がおすすめか、判断しよう [難]
- # dat1.rdsから推定される3つのVPA結果のうち、どの結果が一番「良い」だろうか?その根拠は?[難]
- # 別データ(dat2.rds)でも同様の解析を行って、dat2.rdsの場合にはどの手法が一番良いか考えよう[難]

#

3. 再生産関係のあてはめ

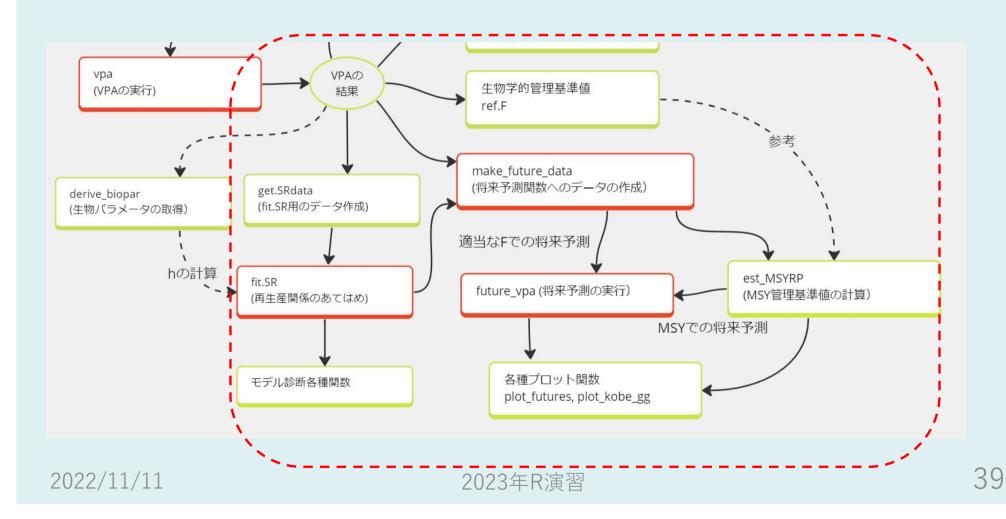


- #演習⑧ 再生産関係のあてはめと選択
- # dat1.rdsの選択率更新法の結果を用いて再生産関係の推定を行ってみよう
- # 再生産ガイドラインに沿うと、HS, RI, BHのうちどの再生産関係を選ぶのが適切だろうか?[難]
- # 再生産関係の選択によって、過去の資源のトレンドや資源状態がどのように説明できるようになるだろうか?[難]
- # dat3.rdaデータについても同様の解析を行い、dat1.rdsの結果と比較してみよう [難]

```
# hの計算に必要な生物パラメータの取り出し
biopars <- derive biopar(res vpa su, derive year=45:50)
data SR su <- get.SRdata(res vpa su)</pre>
res_allSR_su <- tryall_SR(data_SR_su,bio_par=biopars)</pre>
res allSR su %>% arrange(AICc) %>%
    mutate(deltaAICc=AICc-min(AICc)) %>% View()
res_SR_BH <- <pre>fit.SR(data_SR_su, SR="BH", method="L2", AR=0,
                     bio par=biopars, max.ssb.pred=10)
res_SR_RI <- <pre>fit.SR(data_SR_su, SR="RI", method="L2", AR=0,
                      bio par=biopars, max.ssb.pred=10)
res_SR_HS <- <pre>fit.SR(data_SR_su, SR="HS", method="L2", AR=0,
                     bio par=biopars, max.ssb.pred=10)
```

2022/11/11 2023年R演習 38

4. 管理基準值計算&将来予測



```
# 演習⑨ 将来予測をやってみよう
# - 以下のコマンドを実行し、将来予測~結果のプロット~神戸プロットの描画までを体験してみよう
# - dat3.rds、dat4.rdsについても将来予測をやってみよう。
                          それぞれどのような特徴があるだろうか?[難]
currentF <- as.numeric(unlist(res vpa su$faa))</pre>
data future <- make future data(res vpa su, nsim=100, nyear=30,
               future initial year name=50, start F year name=51,
               start biopar year name=51, start random rec year name=51,
               currentF=F4850.
               futureF=F4850,
               waa year=49:50, waa catch year=49:50, maa year=49:50,
               M year=49:50,
               faa year=NULL, start ABC year name=52,res SR=res SR HS)
res MSY <- <pre>est MSYRP(data future)
res MSY$summary
```

```
SPRmsv <- res MSY$summarv$perSPR[1]</pre>
Fmsv <- res MSY$Fvector[1,] %>% as.numeric()
SBlimit <- res MSY$summary$SSB[4]
SBban <- res_MSY$summary$SSB[3]</pre>
SBtarget <- res MSY$summary$SSB[1]</pre>
# make data futureのHCR関係の引数をMSYベースのものに変更する
data_future_msy <- redo_future(data_future,</pre>
                               input data list=list(futureF=Fmsy, HCR Bban=SBban,
                                                    HCR beta=0.8, HCR Blimit=SBlimit
                               only data=TRUE)
res future msy <- future vpa(data future msy$data, multi init=1,
                             calc SPR year name=as.character(1:100),
                             SPRtarget=SPRmsy*100)
plot futures(res vpa su,list(res future msy), Btarget=res MSY$summary$SSB[1],
             what.plot=c("Recruitment", "SSB", "biomass", "catch", "U", "Fratio"))
```

```
# 神戸プロット
kobe_ratio <- get.SPR(res_vpa_su, target.SPR=res_MSY$summary$perSPR[1]*100)$ysdata %:
    mutate(Bratio=unlist(colSums(res_vpa_su$ssb))/res_MSY$summary$SSB[1])%>%
    mutate(Fratio=get("F/Ftarget")) %>%
    mutate(year=as.numeric(colnames(res vpa su$faa)))
plot kobe gg(FBdata=kobe ratio,
             refs_base=res_MSY$summary)
```

```
# ### 演習10 将来予測でABCの不確実性を考慮してみよう[難]
# - 通常の将来予測res future msyの結果とどこが違うだろうか?
# - dat1.rdsとdat4.rdsでどのように違うだろうか?その理由は?
# 引数を変えてdata_futureを作り直す(時間がかかるため、計算回数を減らす)
data future mse <- redo future(data future msy, input data list=list(nsim=50,
nyear=10),
                            only data=TRUE)
# do MSEのオプションをTRUEにしてMSEモードにする
res_future_mse <- <pre>future_vpa(data_future_mse$data,
                          multi init=res MSY$summary$Fref2Fcurrent[1],
                          do MSE=TRUE, MSE input data=data future mse,
                          calc SPR year name=as.character(1:100),
SPRtarget=SPRmsy*100)
plot futures(res vpa su, lst(res future mse, res future msy),
           what.plot=c("Recruitment", "SSB", "biomass", "catch", "U", "Fratio"))
```