

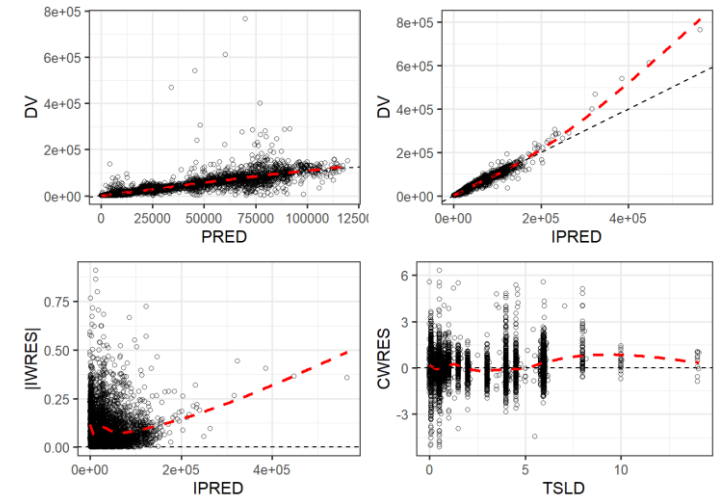
GOFプロット作成と視覚的共変量探索

Contents

- GOFプロット
 - 必要な関数の説明
 - 演習-1
- 視覚的共変量探索
 - 共変量候補の相関関係
 - 必要な関数・パッケージの説明
 - 演習-2
 - ETA(変量効果)と共変量の関係
 - 演習-3

Goodness-of-fit (GOF) プロット

- 母集団解析で構築したモデルの診断プロットの1つ
 - 実測値 (DV) vs 母集団予測値 (PRED)
 - 実測値 (DV) vs 個別予測値 (IPRED)
 - 条件付き重み付き残差 (CWRES) vs PRED
 - CWRES vs Time
 - 個別重み付き残差の絶対値 ($|IWRES|$) vs IPRED など



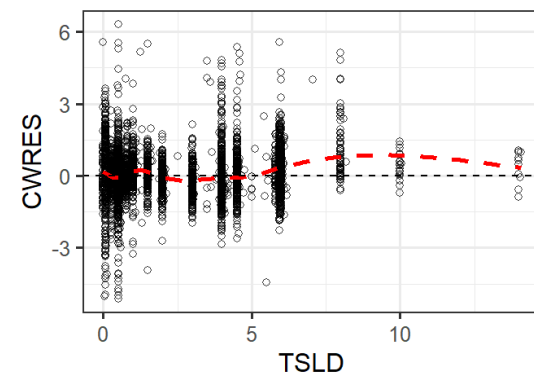
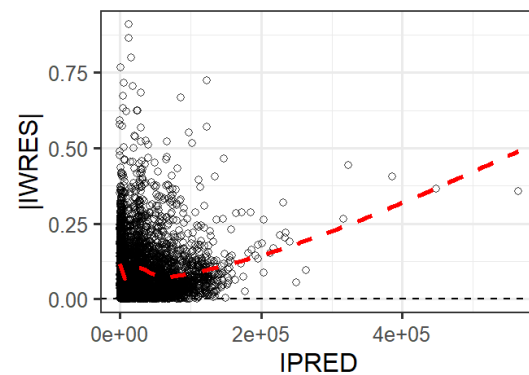
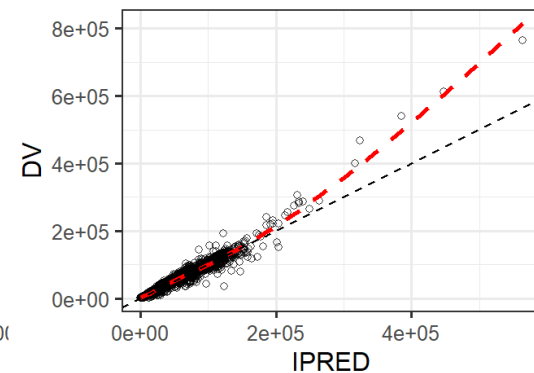
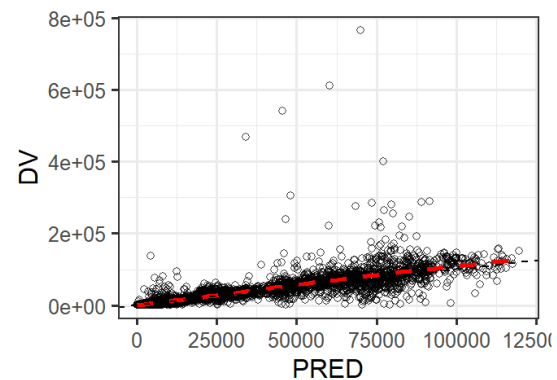
GOFプロットの作成手順



sdtab (GOFプロット作成に必要な変数を出力したファイル)

GOFプロット

ID	DRUG	TIME	TSLD	IPRED	IWRES	CWRES	DV	PRED
1001	1	0.08	0.08	28659.037	0.2262266	1.9951971	35142.7	21061.817
1001	1	0.25	0.25	74049.641	-0.0535400	0.4345582	70085.0	56645.410
1001	1	0.48	0.48	115050.120	-0.2447800	-0.8001000	86888.0	91746.698
1001	1	0.75	0.75	73646.393	-0.0714910	-0.2390400	68381.3	65146.467
1001	1	1.00	1.00	52299.705	0.0411351	0.2186856	54451.1	48906.945
1001	1	1.50	1.50	33480.929	0.0037355	-0.0824990	33606.0	31857.561
1001	1	2.00	2.00	24345.973	0.0304608	0.1652561	25087.6	22487.875
1001	1	3.00	3.00	13816.528	0.0589955	0.4711499	14631.7	11805.423
1001	1	4.50	4.50	6003.747	0.0627425	0.5179446	6380.5	4558.547
1001	1	6.05	6.05	2538.807	-0.0263040	-0.0950420	2472.0	1706.605



GOFプロット作成に必要な関数

- `geom_point`: 散布図を作成する

- 例: `p <- ggplot(data = sdtab, aes(y=DV, x=PRED)) + geom_point(size=1, alpha=0.8, shape=21)`
- `size`: 点の大きさ
- `alpha`: 点の透過性 ($0 < \alpha \leq 1$)、0に近いほど透明、1に近いほど不透明
- `shape`: 点の形 (0~25)、お勧めはshape=21「○」、詳細はリンク参照 http://www.cookbook-r.com/Graphs/Shapes_and_line_types/

- `stat_smooth`: 平滑化曲線を追加する

- 例: `p <- p + stat_smooth(method="loess", linetype="dashed", colour="red", se=FALSE)`
- `method`: 直線/曲線の算出方法 ("loess", "glm", "lm", "gam")
- `linetype`: 線の種類 ("solid", "dashed", "dotted" ...)
- `size`: 線の太さ
- `colour`: 線の色 (デフォルトは "black")
- `se`: 信頼区間を表示させるか否か (TRUE, FALSE)

- `geom_abline`: 対角線 ($y=x$) を追加する

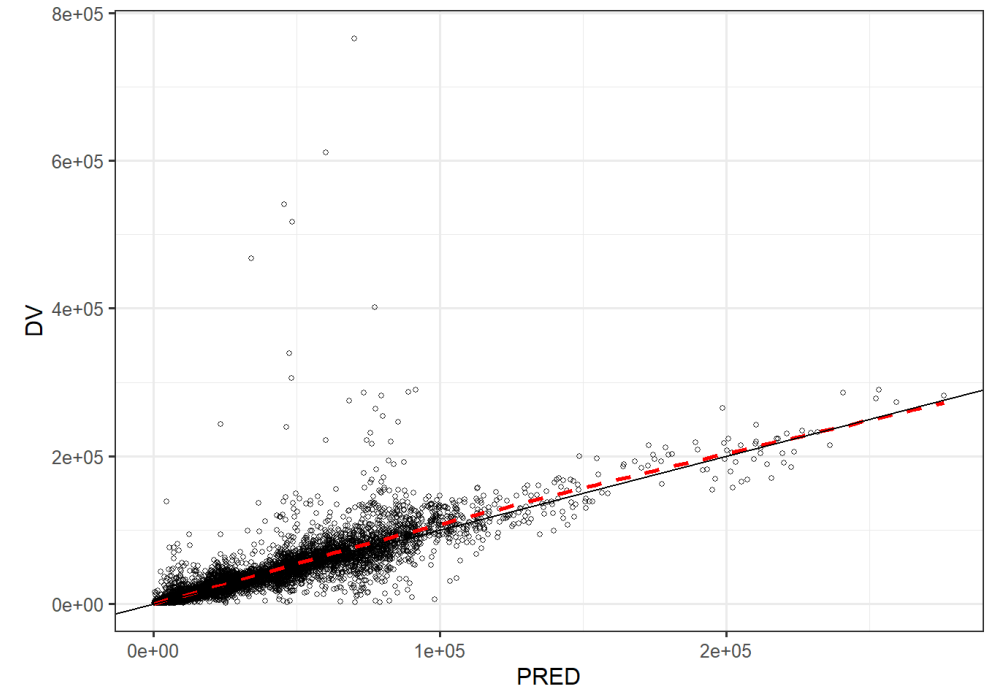
- 例: `p <- p + geom_abline(linetype="solid")`

- `geom_hline`: 対角線 ($y=0$) を追加する

- 例: `p <- p + geom_hline(yintercept=0, linetype="solid")`
- `yintercept`: y切片の値を指定

- `theme`: グリッドと背景を指定する

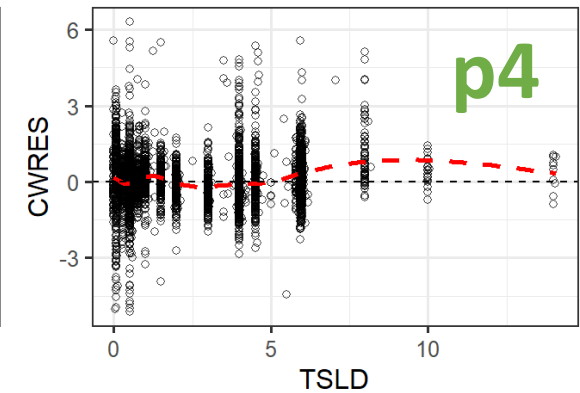
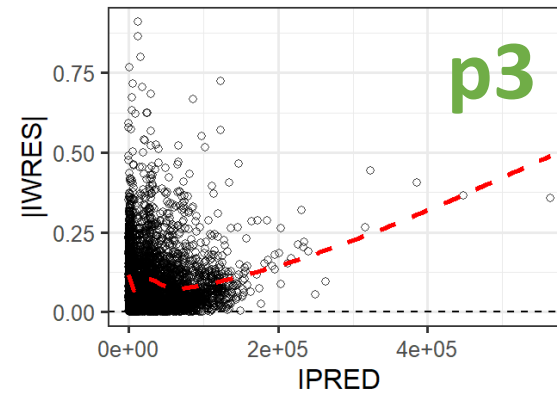
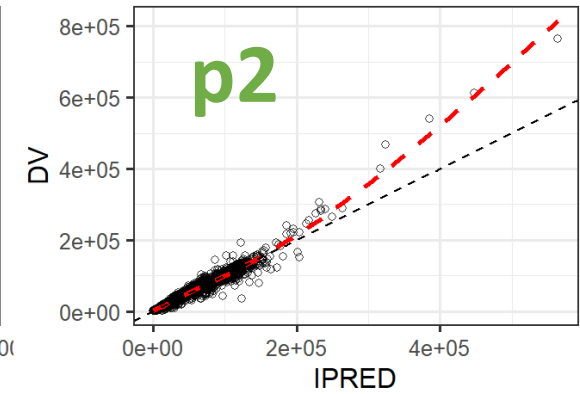
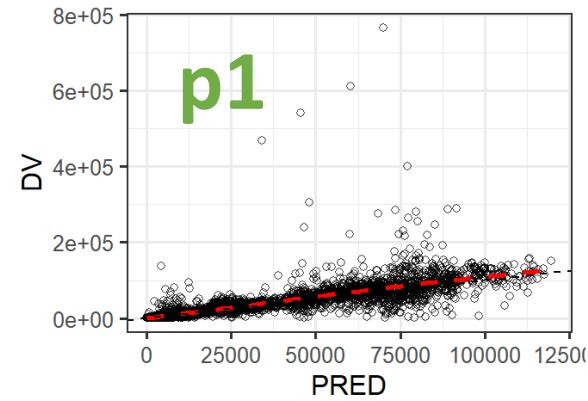
- 例: `p <- p + theme_bw(base_size=12)`
- `theme_bw`: 白背景に灰色のグリッド (デフォルトは `theme_gray`)
- `base_size`: 軸ラベル、タイトル等のフォントサイズ



複数の図をまとめて表示する

- パッケージ“gridExtra”

- 例: `grid.arrange(p1, p2, p3, p4, nrow=2)`
- `nrow`: 分割する行数を指定
- `ncol`: 分割する列数を指定



演習-1

- Data folder内のsdtab60を読み込み、Drug=1, DV>0のデータについて、以下の図を作成してください。
 - 実測値(DV)と母集団予測値(PRED)のプロットを平滑化曲線と対角線($y=x$)付きで作成してください。
 - CWRESと直近の投与後時間(TSLD)のプロットを平滑化曲線と直線($y=0$)付きで作成してください。
 - 上記で作成した図を1行2列で並べて1つの図として表示してください。

※sdtab60ファイル中には、DRUG=1、2のデータが含まれています。

演習-1: 回答コード例

データ読み込み

```
sdtab <- read_table(paste0(path, "/sdtab60"), skip = 1)
```

DV vs PRED

```
p <- ggplot(data=sdtab %>% filter(DV > 0) %>% filter(DRUG == 1), aes(x =
```

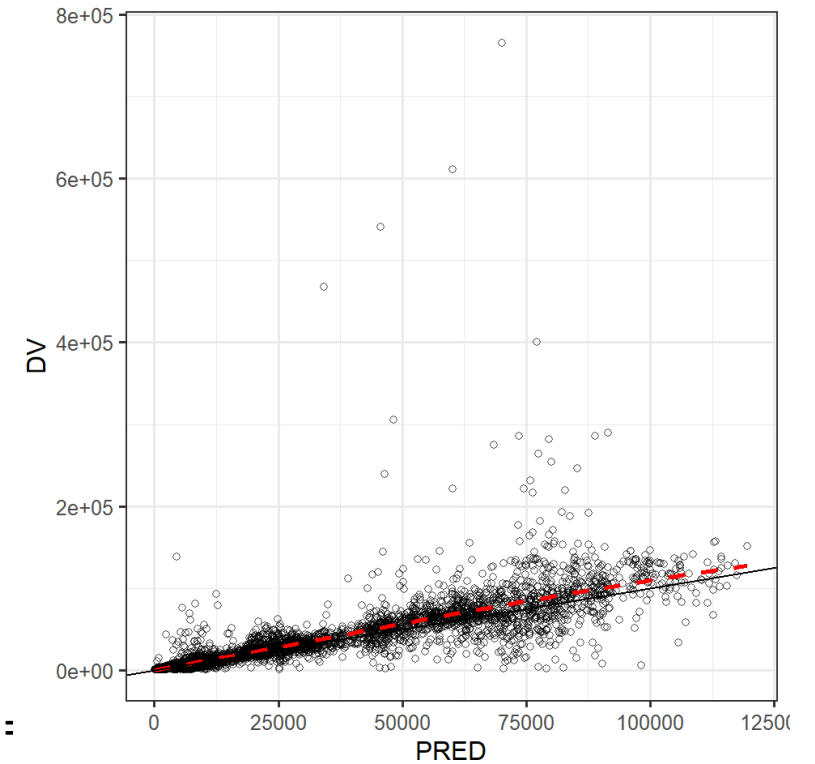
```
p <- p + geom_point(alpha=0.7, shape=21)
```

```
p <- p + stat_smooth(method="loess", linetype = "dashed", colour = "red", se = FALSE)
```

```
p <- p + geom_abline()
```

```
p <- p + theme_bw(base_size=12)
```

```
p1 <- p
```



演習-1: 回答コード例

CWRES vs TSLD

```
p <- ggplot(sdtab %>% filter(DV > 0) %>% filter(DRUG == 1), aes(x = TSLD, y = CWRES))
```

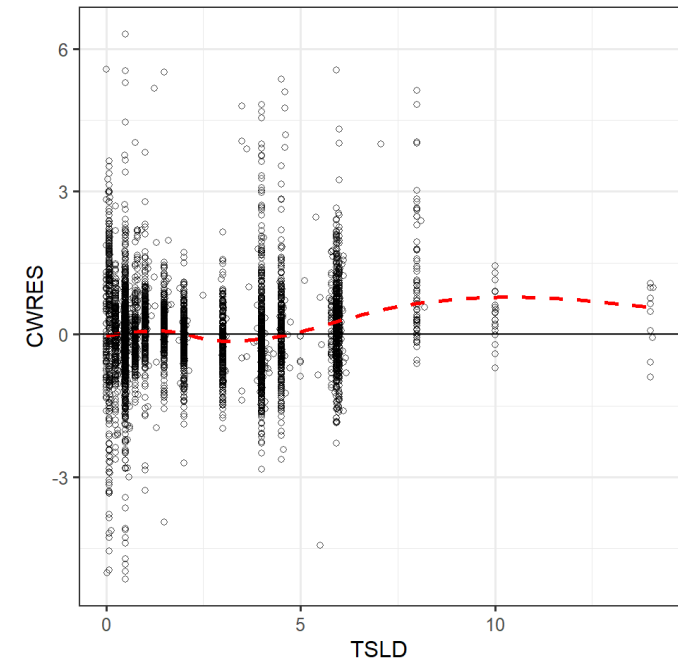
```
p <- p + geom_point(alpha=0.7, shape=21)
```

```
p <- p + stat_smooth(method="loess", linetype = "dashed", colour = "red", se = FALSE)
```

```
p <- p + geom_hline(yintercept = 0)
```

```
p <- p + theme_bw(base_size=12)
```

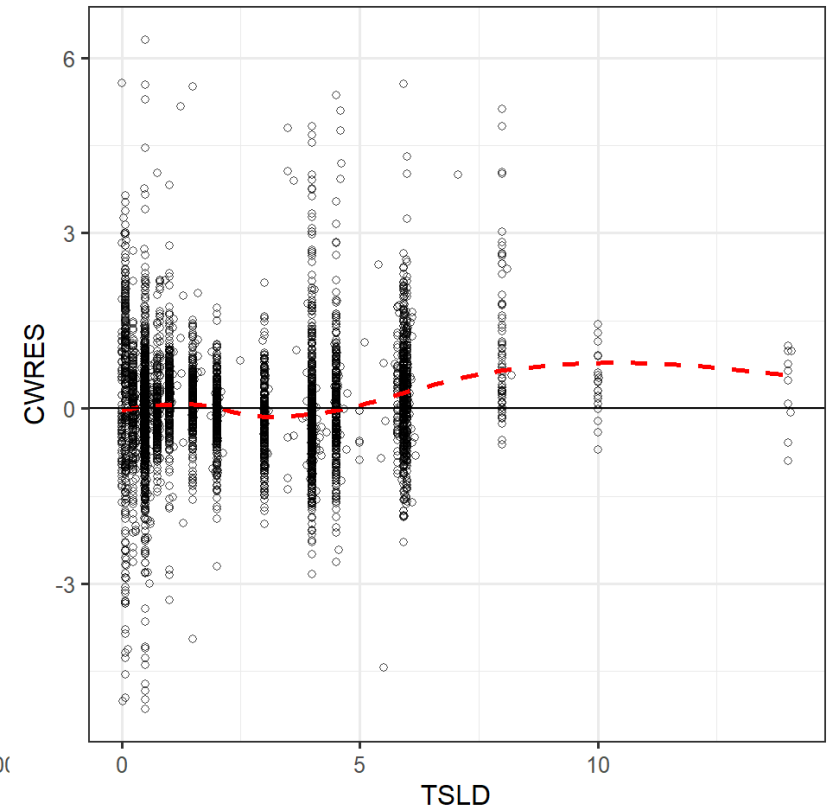
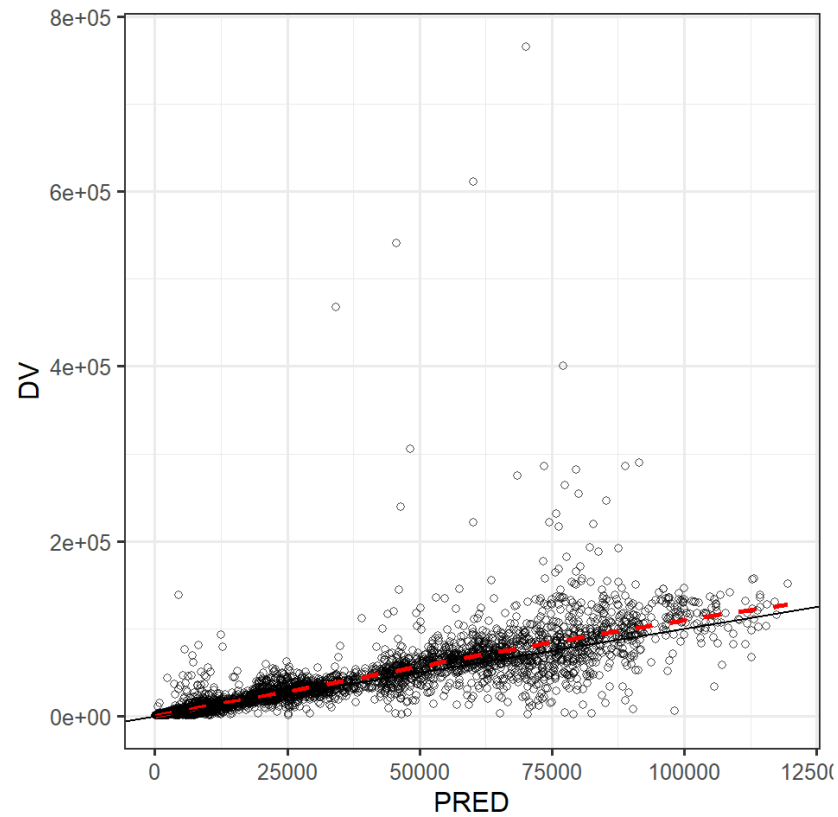
```
p2 <- p
```



演習-1: 回答コード例

図をまとめて表示

```
grid.arrange(p1, p2, ncol=2)
```



視覚的共変量探索

- 共変量
 - 薬物動態や薬力学に影響を及ぼす要因
 - 内因性の要因: 体重、性別、年齢、臨床検査値、遺伝子多型など
 - 外因性の要因: 併用薬、合併症、喫煙の有無など
 - 試験デザイン: 製剤、食事の条件など



共変量探索の流れ

- 患者背景の確認
 - 要約統計量の把握
 - 共変量候補の相関関係の確認(散布図行列)
- Base model構築
- ETA(変量効果)のEBEと共変量の相関を確認(散布図、ボックスプロット)
- NONMEMで共変量探索の実行

※実際に演習を行っていただく部分は青字で示しています。

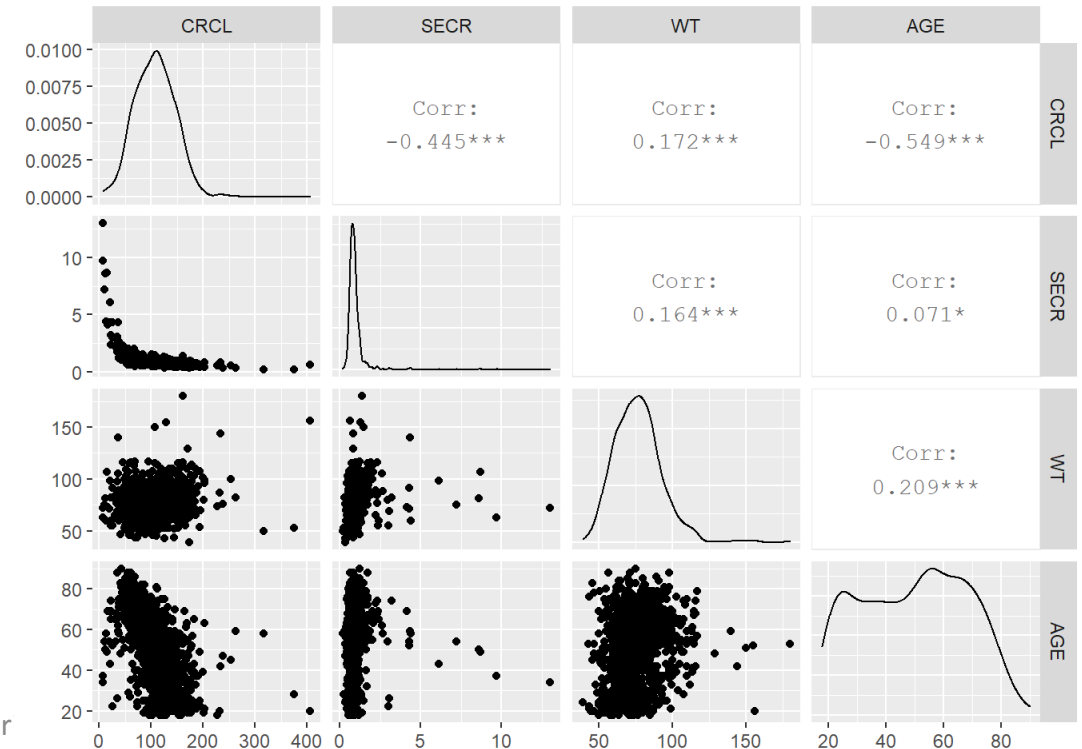
要約統計量の把握

- 収集した患者背景データで共変量探索可能かどうか確認する
 - 連続変数: 平均、標準偏差、データの範囲、分布の形は？
 - 離散変数: 全体に占める割合は？

共変量候補の相関関係の確認

各共変量候補が互いに独立しているか確認する

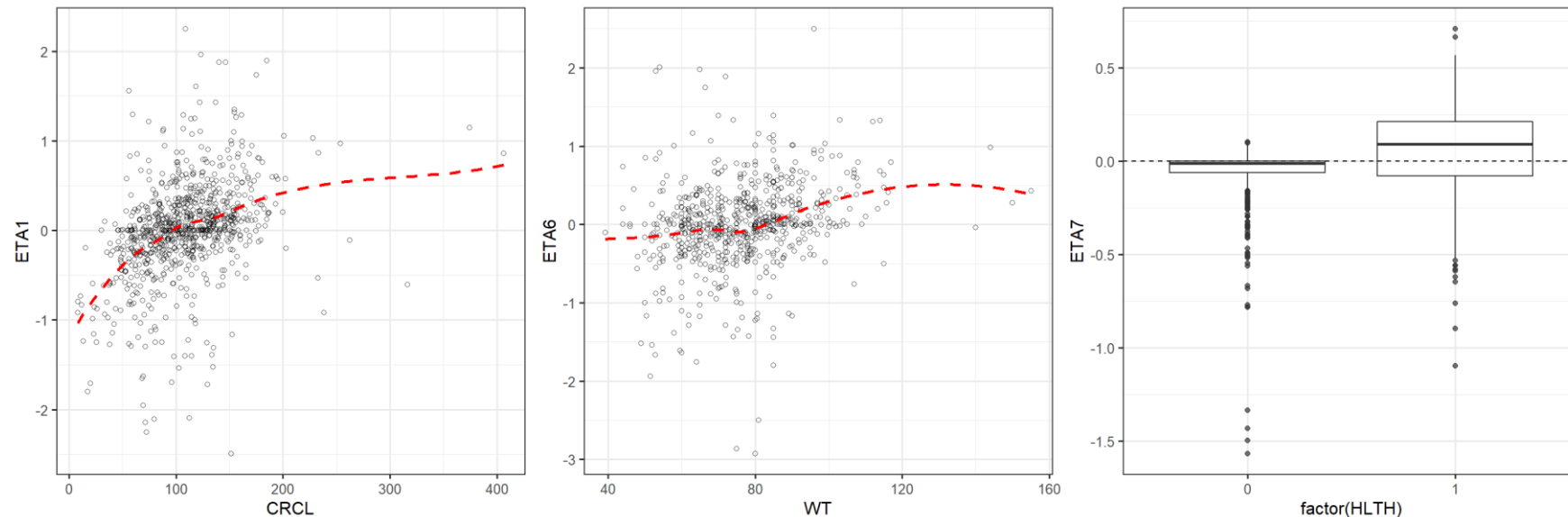
- 相関が強い共変量を複数同時に組み込むとパラメータを適切に推定できない
- あらかじめ共変量の相関を確認し、相関の強い共変量を複数同時にパラメータに組み込まないように注意する



ETA（変量効果）のEBEと共変量の相関を確認

Base modelのETAと共変量に相関関係があるか確認する

- 共変量探索前に組み込まれそうな共変量に当たりをつける
- 共変量を組み込む際の式を検討する（比例的な増加か？べき乗的な増加か？頭打ちか？）



散布図行列作成の前処理に必要な関数

<dplyr>

- **filter**: 列を指定し、条件式に該当する行を抜き出す
 - 例: `nmdata2 <- nmdata %>% filter(DOSE==500)`
 - 意味: 列「DOSE」=500の行を抜き出す
- **distinct**: 列を指定し、重複した値を持つ行を削除する
 - 例: `nmdata2 <- nmdata %>% distinct(ID, .keep_all = TRUE)`
 - 意味: 「ID」列を指定し、重複した値を持つ行を削除する(初出を残し、2回目以降は削除)
 - `.keep_all`: 指定した列以外を残すか否か(TRUE/FALSE)
- **select**: 指定した列を抜き出す
 - 例: `nmdata2 <- nmdata %>% select(ID, DOSE)`
 - 意味: 「ID」「DOSE」列を抜き出す
- **mutate**: 新たに列を作成する(既にある列を指定すると、データが置き換わる)
 - 例: `nmdata2 <- nmdata %>% mutate(IBW = 22 * (HT/100)^2)`
 - 意味: 新たに $IBW (= 22 * (HT/100)^2)$ という列を作成
 - 例: `nmdata2 <- nmdata %>% mutate(MALE = as.factor(MALE))`
 - 意味: 「MALE」列の型を因子(factor)に変更する

共変量の散布図行列作成に必要なパッケージ

<GGally>: 「ggplot2」を利用して散布図行列を作成するパッケージ

• 例: `p <- ggpairs(data=nmdata)`

- 非対角要素(lower)
 - 連続変数×連続変数: 散布図
 - 連続変数×離散変数: ヒストグラム
 - 離散変数×離散変数: 棒グラフ
- 対角要素(diag)
 - 連続変数: ヒストグラム(密度)
 - 離散変数: 棒グラフ
- 非対角要素(upper)
 - 連続変数×連続変数: 相関係数
 - 連続変数×離散変数: ボックスプロット
 - 離散変数×離散変数: 割合

• 例: `p <- ggpairs(data=nmdata, aes(alpha=0.7, colour=MALE))`

- 「MALE」列で色分けして表示



ggpairs関数の詳細はリンク参照 <https://ggobi.github.io/ggally/articles/ggpairs.html>

演習-2

- Data folder内のPSP4-8-748-s012.csvをnmdataとして読み込み、共変量の散布図行列作成のためのデータセットをnmdata2として作成してください。
 - 被験者番号(ID)が重複した行は削除し、各IDにつき1行のデータセットとしてください。
 - 「CRCL」「AGE」「MALE」「HLTH」の列を抜き出してください。
 - 「CRCL」の値が欠損しているID(CRCL=-99)は削除してください。
 - 「MALE」「HLTH」の型を因子(factor)にしてください。
- データセット「nmdata2」を用いて共変量の散布図行列を作成してください。
 - 「CRCL」「AGE」「MALE」「HLTH」の散布図行列を、「HLTH」で色分けして作成してください。

演習-2: 回答コード例

データ処理

```
nmdata <- read_csv(paste0(path, "/PSP4-8-748-s012.csv"), skip = 0)
```

```
nmdata2 <- nmdata %>%
```

```
  distinct(ID, .keep_all = TRUE) %>%
```

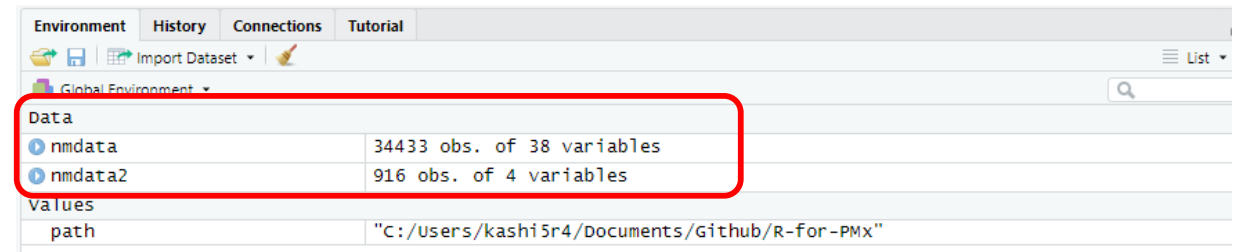
```
  select(CRCL, AGE, MALE, HLTH) %>%
```

```
  filter(CRCL != -99) %>%
```

```
  mutate(MALE = as.factor(MALE)) %>%
```

```
  mutate(HLTH = as.factor(HLTH))
```

```
kable(head(nmdata2, 10))
```



The screenshot shows the RStudio Environment pane. A red rectangle highlights the 'Data' section, which contains two objects: 'nmdata' (34433 obs. of 38 variables) and 'nmdata2' (916 obs. of 4 variables). Below this, the 'Values' section shows the 'path' variable with the value 'C:/Users/kashi5r4/Documents/Github/R-for-PMx'.

Environment	History	Connections	Tutorial
Global Environment			
Data			
nmdata	34433 obs. of 38 variables		
nmdata2	916 obs. of 4 variables		
Values			
path	"C:/Users/kashi5r4/Documents/Github/R-for-PMx"		

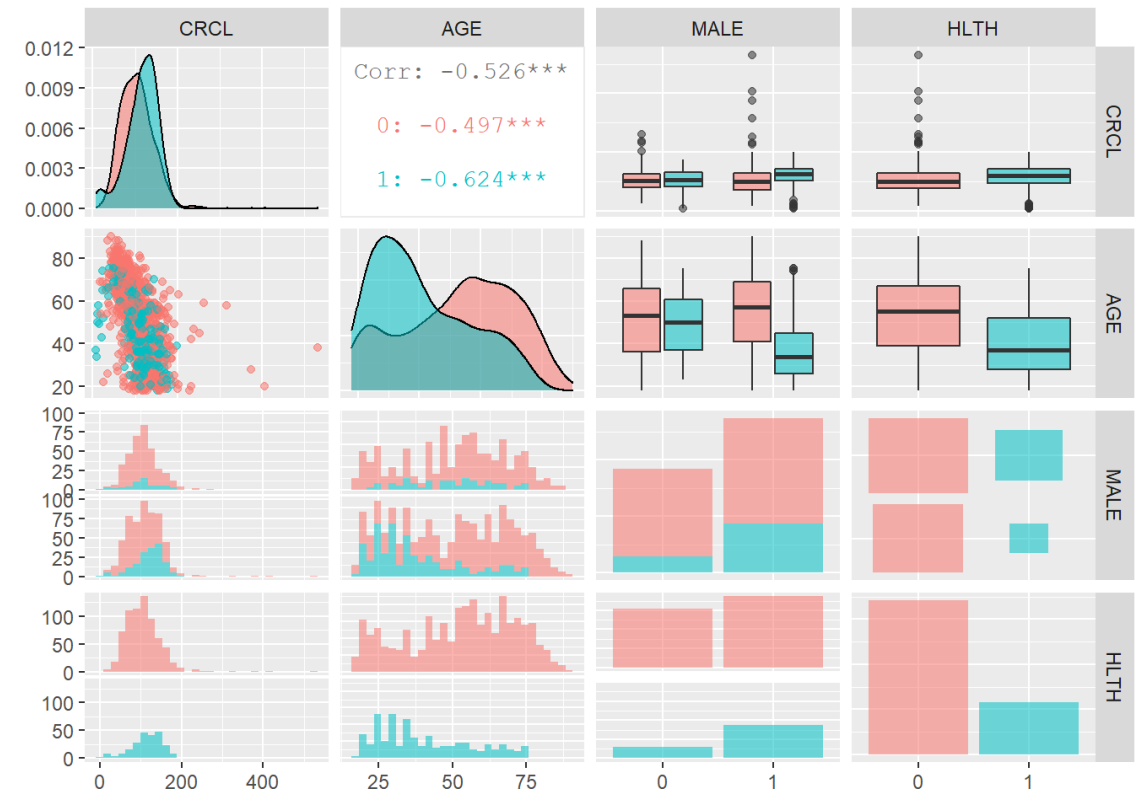
CRCL	AGE	MALE	HLTH
113.5	29	1	1
152.6	23	1	1
145.3	22	1	1
145.7	40	1	1
177.0	19	1	1
88.3	39	1	1
147.5	34	1	1
110.0	40	1	1
136.7	38	1	1
160.4	22	1	1

演習-2: 回答コード例

散布図行列

```
p <- ggpairs(data=nmdata2, aes(alpha=0.7, colour=HLTH))
```

```
print(p)
```



演習-3

- Data folder内の`patab61`、`catab61`、`cotab61`、`sdtab61`について
 - `patab61`、`catab61`、`cotab61`は被験者番号 (ID) が重複した行を削除して、`patab`、`catab`、`cotab`を作成してください。
 - `sdtab61`はID及びDRUGが重複した行を削除して、`sdtab`を作成してください。
 - `patab`、`catab`、`cotab`、`sdtab`をIDをkeyとしてマージした`nmdata`を作成してください。
- `nmdata`を用いてETA1とCRCLの相関プロットを作成してください。
- `nmdata`を用いてETA1とHLTHのボックスプロットを作成してください。

演習-3に用いるデータセットの説明

patab61: 患者個別パラメータ(ETA)を出力したファイル

catab61: 共変量候補のうち、離散変数を出力したファイル

cotab61: 共変量候補のうち、連続変数を出力したファイル

sdtab61: GOFプロット作成に必要な変数を出力したファイル(今回は「DRUG」列をマージするために使用)

演習-3: 回答コード例

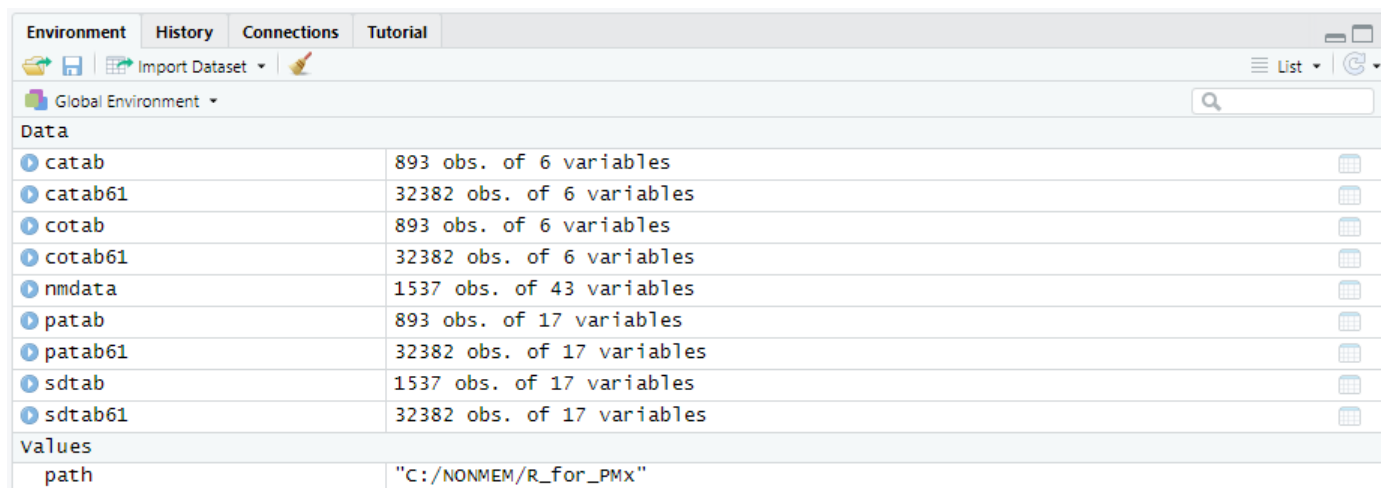
データ読み込み

```
cotab61 <- read_table(paste0(path, "/cotab61"), skip = 1)
catab61 <- read_table(paste0(path, "/catab61"), skip = 1)
patab61 <- read_table(paste0(path, "/patab61"), skip = 1)
sdtab61 <- read_table(paste0(path, "/sdtab61"), skip = 1)
```

データ処理

```
patab <- patab61 %>% distinct(ID, .keep_all=TRUE)
catab <- catab61 %>% distinct(ID, .keep_all=TRUE)
cotab <- cotab61 %>% distinct(ID, .keep_all=TRUE)
sdtab <- sdtab61 %>% distinct(ID, DRUG, .keep_all=TRUE)

nmdata <- left_join(sdtab, patab, by="ID")
nmdata <- left_join(nmdata, catab, by="ID")
nmdata <- left_join(nmdata, cotab, by="ID")
```



The screenshot shows the RStudio Environment pane. The 'Global Environment' is selected, displaying a list of data frames and their dimensions. The data frames are: catab (893 obs. of 6 variables), catab61 (32382 obs. of 6 variables), cotab (893 obs. of 6 variables), cotab61 (32382 obs. of 6 variables), nmdata (1537 obs. of 43 variables), patab (893 obs. of 17 variables), patab61 (32382 obs. of 17 variables), sdtab (1537 obs. of 17 variables), and sdtab61 (32382 obs. of 17 variables). Below the data frames, the 'Values' section shows the path variable set to 'C:/NONMEM/R_for_PMX'.

Object	Dimensions
catab	893 obs. of 6 variables
catab61	32382 obs. of 6 variables
cotab	893 obs. of 6 variables
cotab61	32382 obs. of 6 variables
nmdata	1537 obs. of 43 variables
patab	893 obs. of 17 variables
patab61	32382 obs. of 17 variables
sdtab	1537 obs. of 17 variables
sdtab61	32382 obs. of 17 variables

Variable	Value
path	"C:/NONMEM/R_for_PMX"

演習-3: 回答コード例

ETA vs covariateプロット

```
p <- ggplot(data=nmdata %>% filter(CRCL != -99 & DRUG == 1), aes(x=CRCL, y=ETA1))  
p <- p + geom_point(alpha=0.7, shape=21)  
p <- p + stat_smooth(method="loess", linetype = "dashed", colour = "red", se = FALSE)  
p <- p + theme_bw(base_size=12)  
p1 <- p  
  
p <- ggplot(data=nmdata %>% filter(CRCL != -99 & DRUG == 1), aes(x=factor(HLTH),  
y=ETA1))  
p <- p + geom_boxplot(alpha=0.7, shape=21)  
p <- p + geom_hline(yintercept=0, linetype="dashed")  
p <- p + theme_bw(base_size=12)  
p2 <- p
```

```
grid.arrange(p1, p2, ncol=2)
```

