

# BOXPLOTプロシジャの 臨床データへの適用事例

キリンビール株式会社 医薬カンパニー

開発本部 臨床データ統括部

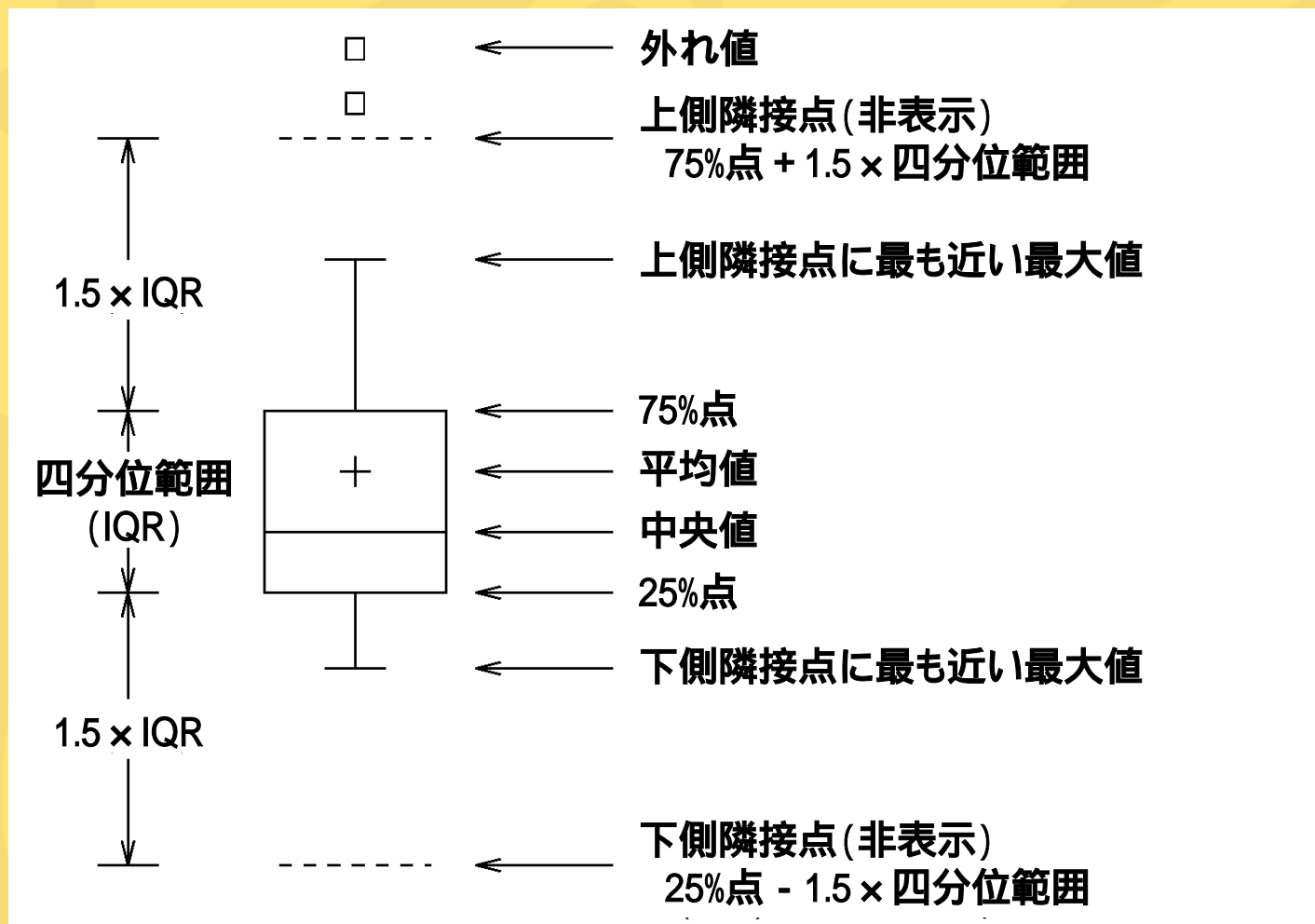
長谷川 要

[hasegawak@kirin.co.jp](mailto:hasegawak@kirin.co.jp)

# 要旨

- 計量データの分布を要約し、その分布の特徴を図示する方法としてしばしば箱ひげ図が用いられる
- SASでは従来から箱ひげ図を作図するプロシジャとしてBase SASプロダクトにUNIVARIATEプロシジャが用意されていたが、その出力はキャラクターベースのアウトプット画面に作成されるため、実用には適さない出力であった
- 近年SASはバージョンアップされ、SAS System 8からは箱ひげ図を作図するための専用のBOXPLOTプロシジャがSAS/STATプロダクトに追加され、見栄えの良いグラフィックベースの箱ひげ図が容易に作図できるようになった
- 本論文では、このBOXPLOTプロシジャの利用方法を紹介し、臨床試験データの解析で利用される箱ひげ図の作成事例を紹介する

# 箱ひげ図の解説



BOXPLOTプロシジャのBOXSTYLE=SCHEMATICオプションで作成される箱ひげ図の解説

# BOXPLOTプロシジャの使用方法

```
PROC BOXPLOT <options>;
```

```
    PLOT analysis-variable*group-variable
```

```
        <(block-variable)>
```

```
        <=symbol-variable>
```

```
    </options>;
```

```
    BY variables;
```

```
    ID variables;
```

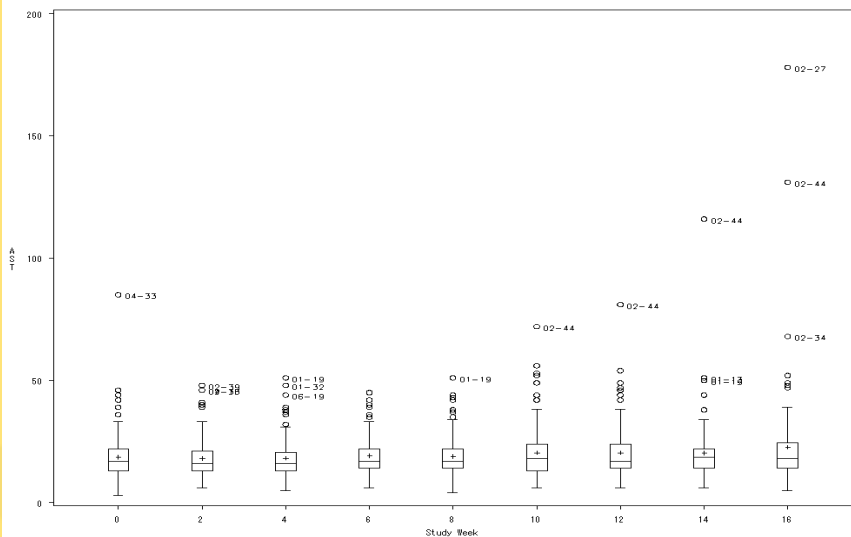
# PLOTステートメント

- analysis-variable
  - ・ 箱ひげ図の作図対象となる数値型の変数
- group-variable
  - ・ グループ分けを行うための変数
- block-variable
  - ・ group-variableで作図された複数の箱ひげ図を分類するための変数
- symbol-variable
  - ・ SYMBOLnステートメントで定義した出力を行わせるための変数
- /options
  - ・ 各種出力様式を指定するためのオプションリスト

# 臨床データへの適用事例 (論文集へ収載した事例)

- 事例1. 一群の臨床検査値の推移の箱ひげ図
- 事例2. 二群の検査値の推移の箱ひげ図
- 事例3. 複数の箱ひげ図を分類した箱ひげ図
- 事例4. 個別プロットを付加した箱ひげ図
- 事例5. 代表値を線で結んだ箱ひげ図
- 事例6. データ数の重みを考慮した箱ひげ図

# 一群の臨床検査値の推移の箱ひげ図



```
PROC BOXPLOT DATA=LAB_DATA;  
    PLOT AST*WEEK  
        /BOXSTYLE=SCHEMATICIDFAR  
        IDSYMBOL=CIRCLE;  
    ID USUBJID;  
RUN;
```

**ポイント:**

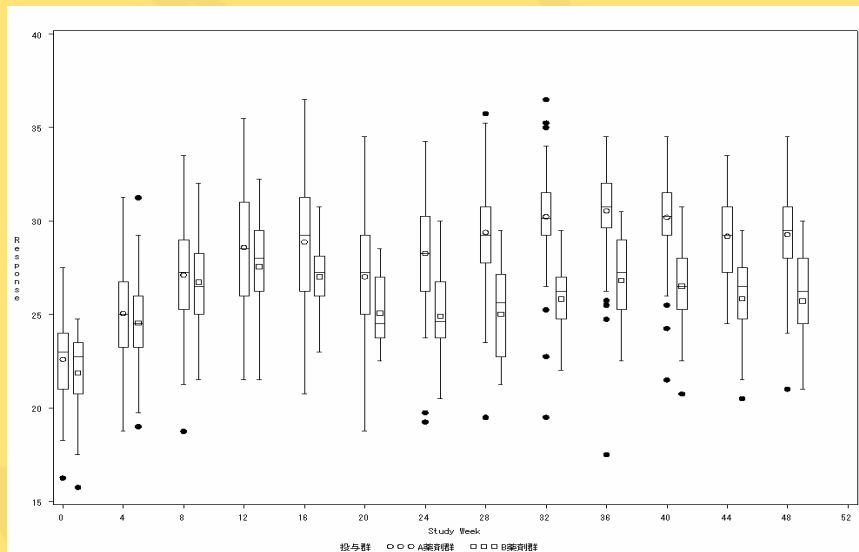
外れ値のシンボル表示及び外れ値のデータラベル  
を表示させるために、

**BOXSTYLE=SCHEMATICIDFARオプション  
及びIDステートメントを組み合わせで指定する。**



# 事例2.

## 二群の検査値の推移の箱ひげ図



### ポイント:

群毎にシンボルを指定するために、=symbol-variableを用いる。

1ページ中に全ての測定時点を出力するために、

TOTPANELS=1

横軸に指定した変数を連続変数と認識させる、

CONTINUOUS

オプションをPLOTステートメントで指定する。



```
DATA PLOT_DATA;
    SET LAB_DATA;
    IF GROUP EQ 'B' THEN WEEK=WEEK+1;
RUN;
```

```
PROC BOXPLOT DATA=PLOT_DATA;
    PLOT RESPONSE*WEEK=GROUP
        /BOXSTYLE=SCHEMATIC
        IDSAMPLE=DOT TOTPANELS=1
        CONTINUOUS
```

```
HAXIS=0 TO 52 BY 4;
```

```
SYMBOL1 V=CIRCLE C=BLACK;
```

```
SYMBOL2 V=SQUARE C=BLACK;
```

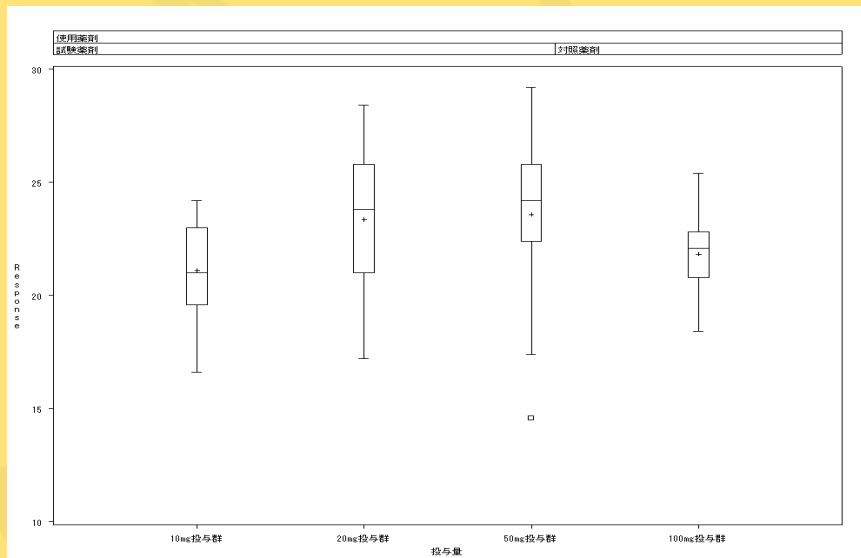
```
LEGEND1 label=('投与群');
```

```
RUN;
```



## 事例3.

# 複数の箱ひげ図を分類した箱ひげ図



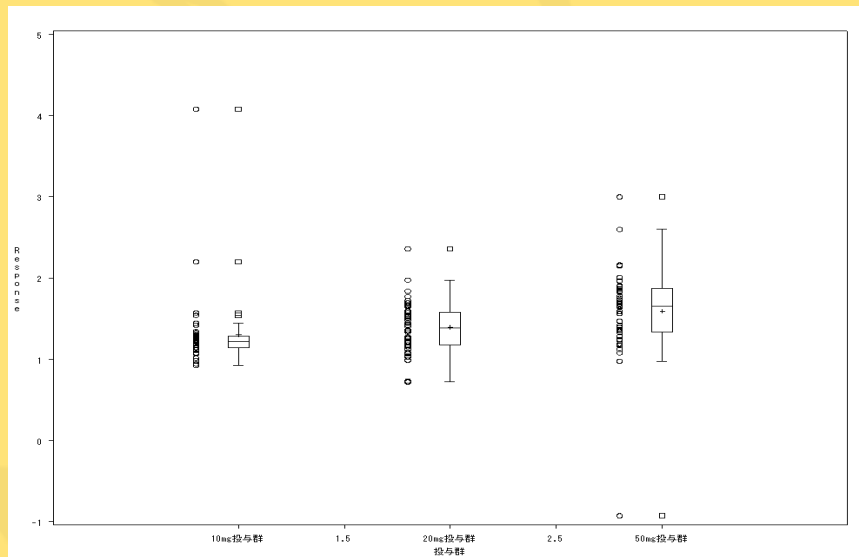
```
PROC BOXPLOT DATA=RESPONSE_DATA;  
    PLOT RESPONSE*DOSE(GROUP)  
        /BOXSTYLE=SCHEMATIC;  
RUN;
```

### ポイント:

PLOTステートメント中で、group-variableに  
block-variableを( )内に指定する。



# 事例4. 個別プロットを付加した箱ひげ図



ポイント:

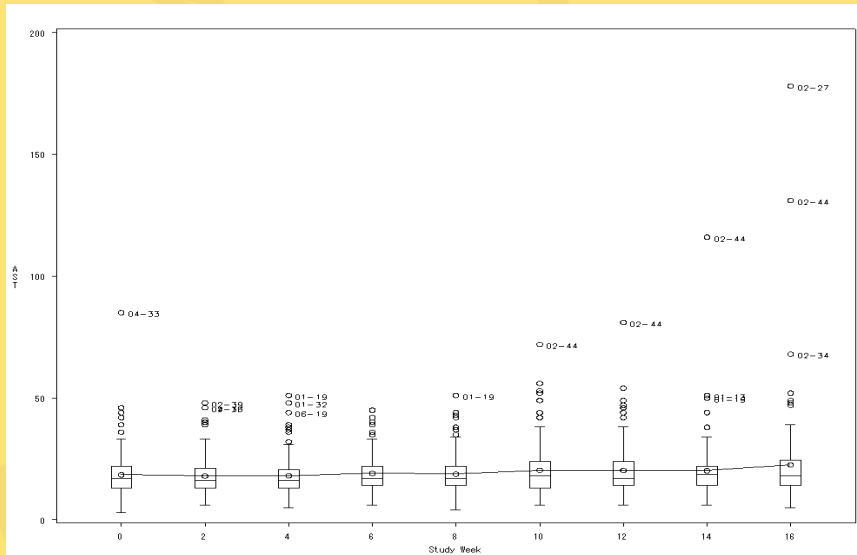
個別プロットの作成はBOXPLOTプロシジャの機能に存在しないことから、適切なANNOTATEデータセットを作成し、作図を行う。



```
DATA ANNO_DATA;  
    SET RESPONSE_DATA;  
    XSYS='2';  
    YSYS='2';  
    X=DOSE-0.2;  
    Y=RESPONSE;  
    TEXT='CIRCLE';  
    FUNCTION='SYMBOL';  
    OUTPUT;  
RUN;
```

```
PROC BOXPLOT DATA=RESPONSE_DATA;  
    PLOT RESPONSE*DOSE  
        /BOXSTYLE=SCHEMATIC  
        ANNOTATE=ANNO_DATA  
        CONTINUOUS;  
RUN;
```

## 代表値を線で結んだ箱ひげ図



```
PROC BOXPLOT DATA=LAB_DATA;  
    PLOT AST*WEEK  
        /BOXSTYLE=SCHEMATICIDFAR  
        IDSYMBOL=CIRCLE  
        BOXCONNECT=MEAN;  
    ID USUBJID;  
RUN;
```

**ポイント:**

PLOTステートメントのBOXCONNECT=オプションの指定で、箱ひげ図と箱ひげ図の間で代表値を用いて結線が可能となる。

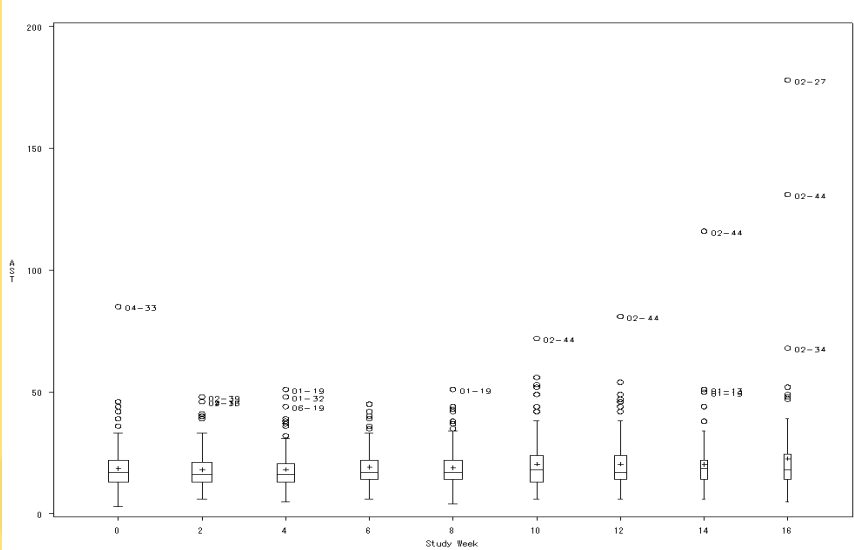
指定可能な代表値は、

MEAN, MEDIAN, MIN, MAX, Q1, Q3

である。



## 事例6.



```
PROC BOXPLOT DATA=LAB_DATA;  
    PLOT AST*WEEK  
        /BOXSTYLE=SCHEMATICIDFAR  
        IDSYMBOL=CIRCLE  
        BOXWIDTHSCALE=1;  
    ID USUBJID;  
RUN;
```

**ポイント:**

作図に用いたデータ数を箱ひげ図の箱の幅に反映させるために、BOXWIDTHSCALE=オプションをPLOTステートメントに指定する。

- =1: サンプルサイズの $n$ で重み付け
- =0.5: サンプルサイズの  $n$ で重み付け
- =0: サンプルサイズの $\log(n)$ で重み付け



# 論文集へ収載しなかった追加事例 (応用編)

- 論文集では触れなかったグラフの出力デバイスと、その他の応用編について追加事例を以下に紹介する
- SAS/GRAPHプロダクトでは多種多様な出力デバイスが用意されているが、検討した結果PSLLA4を用いたPostScriptファイルへの出力が実用的であった (Adobe Acrobat DistillerによりPDFファイルへ変換可能)
- 作図のラベルに用いるフォントはHWPSL009が有用であった
- 箱ひげ図の作図に用いたデータ数を併記することは有用と考え、ANNOTATEデータセットを活用した出力方法を検討した

# 追加事例: サンプルプログラム

\* ----- 出力用の各種設定;

```
%let gfile=c:\temp\sugi2006.ps;
filename outgraph "&gfile.";
options noxwait;
x "del &gfile";
goptions gaccess=gsasfile gsfmode=append
        gsfname=outgraph device=PSLLA4;
```

\* ----- 時点ごとのポイント数を算出;

```
proc sort data=lab_data;
  by week;
run;
proc univariate data=lab_data noprint;
  by week; var ast; output out=n n=n;
run;
```

\* ----- annotateデータ作成;

```
data anno;
  set n;
  xsys='2'; ysys='2'; x=week; y=5;
  position='5'; size=1.2;
  style='HWPSL009';
  text='n=' || trim(left(put(n,best.)));
  function='label';
  output;
run;
```

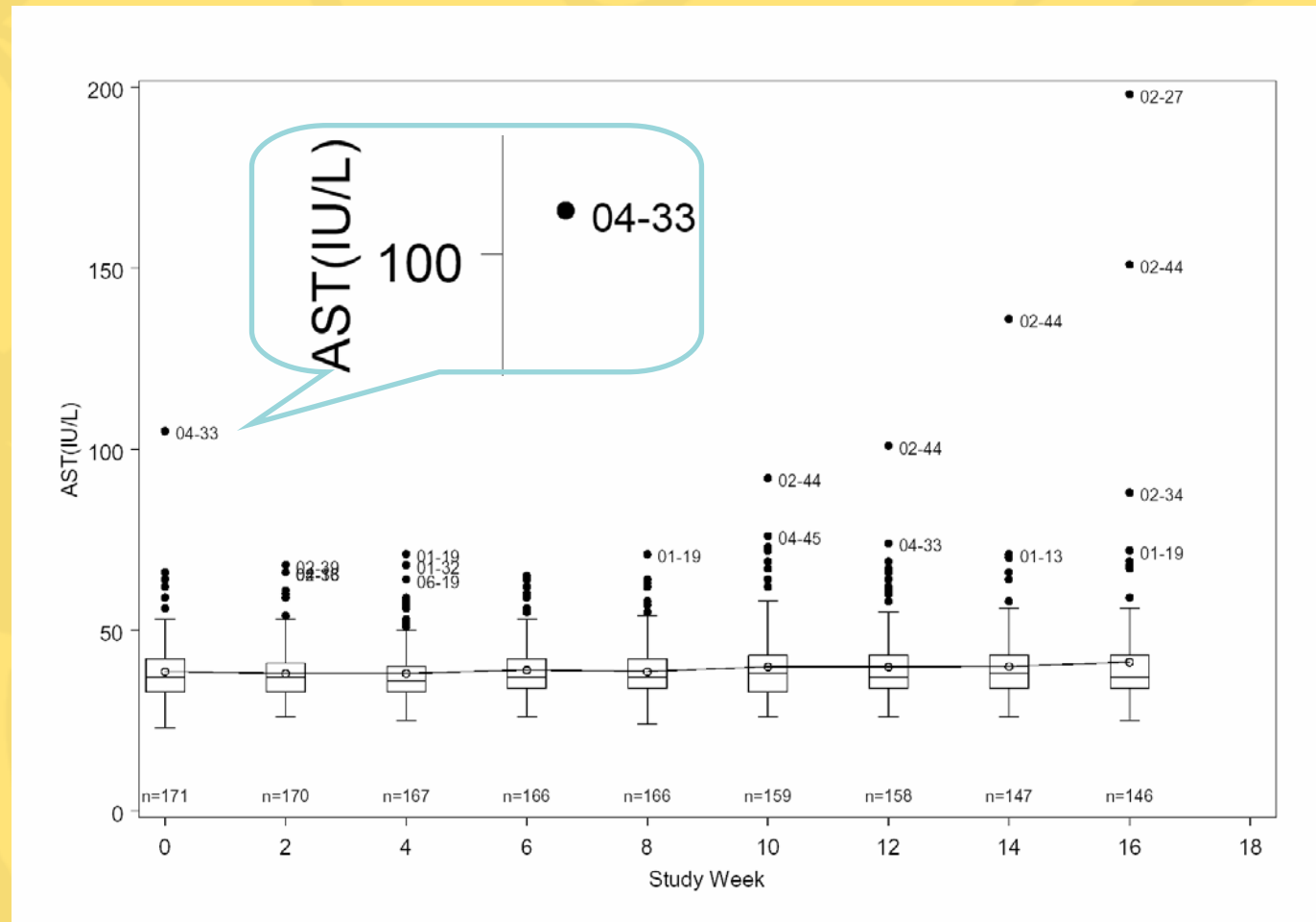
\* ----- 作図処理;

```
proc boxplot data=lab_data annotate=anno;
  plot value*week
        /boxstyle=schematicidfar
        cboxes=black idsymbol=dot
        continuous
        boxconnect=mean
        vaxis=axis1 haxis=axis2
        idheight=2 idfont=HWPSL009
        totpanels=1;
  axis1 order=0 to 200 by 50
        value=(f=HWPSL009 h=1.5)
        label=(f=HWPSL009 r=-90 h=1.5
               "AST(IU/L)");
  axis2 order=0 to 18 by 2
        value=(f=HWPSL009 h=1.5)
        label=(f=HWPSL009 h=1.5 "Study Week");
  symbol v=circle c=black;
  id usubjid;
run;
quit;
```

# 追加事例:出力結果 (PSファイル PDFファイル変換後)

51,774 Bytes

14,259 Bytes



# おわりに

- 本論文ではSAS System 8のSAS/STATプロダクトに追加されたBOXPLOTプロシジャの利用方法及び活用事例の紹介を行った
- 一般的な臨床試験データのまとめに際して実用に十分耐えうるプロシジャと考えられた
- ANNOTATEデータセットを併用することにより、BOXPLOTプロシジャ単独では作図困難なグラフも容易に作図が可能である
- 出力デバイスについては本論文では触れなかったが、本ポスター発表にて追加報告した方法により、比較的美しい出力結果がPDFファイルとして作成が可能である