

<2変量データの記述>

1. 準備

今回の実習では、以前使用した**誕生日血液型2011**のファイルと、今回あらかじめ河田が作成した**打者成績2011**のファイルを用いる。

【課題 12】講義用 HP から、**打者成績 2011** のファイルをダウンロードしてみよう。

📖 手順

- ① 徳山大学のトップページ (<http://www.tokuyama-u.ac.jp>) で、「在学生の皆様へ」のメニューから、「教員紹介」を選び、河田の教員紹介を開く。
- ② そのページの下の方にある、「個人サイト」のところをクリックすると、「河田研究室」のページにジャンプする。(ここまでの手順は、<http://www2.tokuyama-u.ac.jp/kawada> とアドレスを直接入力してもよい)
- ③ 「統計学基礎」をクリックし、「第13回 11月14日(月)」の配布資料にある、「打者成績2011」をクリックし、自分の使いやすい場所に保存する。

※ **誕生日血液型2011**のファイルがない者は、同様の方法で、**第6回**のところにある**誕生日血液型2011**のファイルをダウンロードすること。

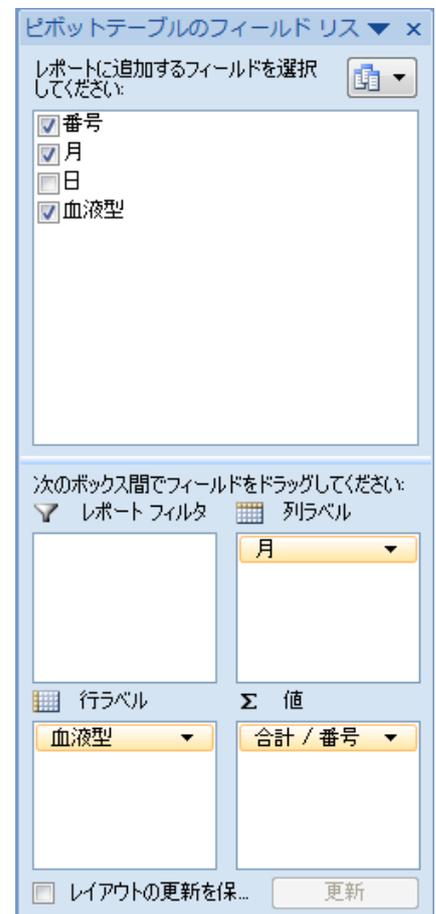
2. 分割表と2次元ヒストグラム

【課題 13】この講義の履修者を、血液型と誕生月の 2 項目で分類し、分割表にまとめよう。

Excel で分割表を作成する場合、**ピボットテーブル**を用いる。

📖 手順

- ① まず、分割表にまとめるデータの範囲(ここでは、A1:D54)をドラッグし、**ピボットテーブル**のボタン(挿入タブにある)をクリックする。
- ② すると、「ピボットテーブルの作成」というウインドウが開くので、「ピボットテーブルレポートを配置する場所」として、「既存のワークシート」のボタンをクリックし、分割表を作成する左上のセルを指定する。
- ③ 右図のようなウインドウが開くので、「月」をドラッグして「列ラベル」のところに、「血液型」をドラッグして「行ラベル」のところに、「番号」をドラッグして「値」のところにそれぞれ配置する。そして、「合計/番号」となっている右側の▼をクリックし、「値フィールドの設定」を選び、「集計の方法」を「データの個数」とする。
- ④ ここまでで分割表は一応完成であるが、「データの個数/番号」となっているところを、「血液型と誕生月」に変えよう。これは、「データの個数/番号」と書かれているセルをクリックし、数式バーで「血液型と



誕生月」と入力する。同様に、血液型もA型、B型というように変更していけば、下の表のようになる。(列ラベル、行ラベルも同様に変更が必要かもしれない。)

誕生月と血液型	月												総計
血液型	1	2	3	4	5	6	7	9	10	11	12	総計	
A型	2			1		1	3	2	1	4	3	17	
B型			1	1	1	2	1	1	2		2	11	
O型	1	2		1	1	1	5	2	3		3	19	
AB型	1		1			1			2		1	6	
総計	4	2	2	3	2	5	9	5	8	4	9	53	

【課題 14】この分割表を2次元ヒストグラムであらわそう。

この場合、血液型も誕生月も質的変量であり、2次元ヒストグラムは通常の3-D縦棒グラフ¹となる。

手順

- ① ピボットテーブルのどこかをクリックすると、最上部に「ピボットテーブルツール」と出るので、これをクリックし、「ピボットグラフ」のボタンをクリックする。
- ② 「グラフの挿入」というウィンドウが開くので、右上の「3-D縦棒」を選択すれば、大体完成である。

- ③ 細かな修正として、タイトルと軸ラベルの設定をおこなう。グラフをアクティブにして、「レイアウト」のタブをクリックし、「グラフタイトル」、「軸ラベル」のボタンでそれぞれ設定する。主縦軸と奥行き軸の「軸ラベル」は垂直に配置しよう。そして、

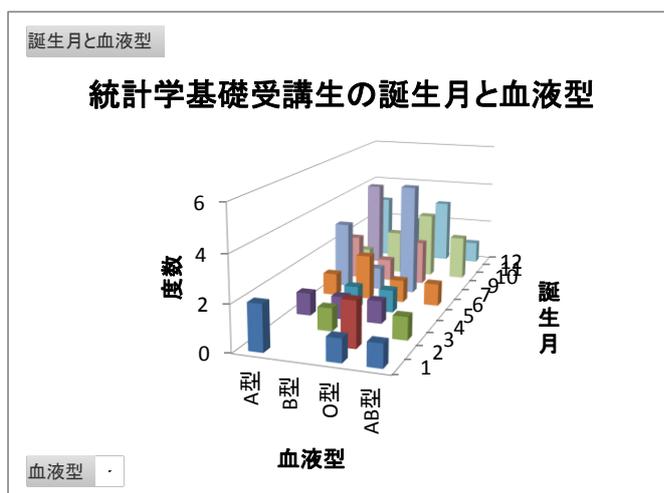
タイトル：「統計学基礎受講生の誕生月と血液型」と記入する。

主横軸：「血液型」と記入する。

主縦軸：「度数」と記入する。

奥行き軸：「誕生月」と記入する。

すると、右図のようになる。右図はグラフの大きさを多少調整してある。



3. 2次元度数分布表

2. では質的変量どうしを、ピボットテーブルで分割表にまとめた。次にこの応用として、とりうる値の多い離散変量どうしについて、2次元度数分布表を作成する。打者成績2011のファイルを用い、本塁打数と三振数の関係を2次元度数分布表にまとめてみよう。

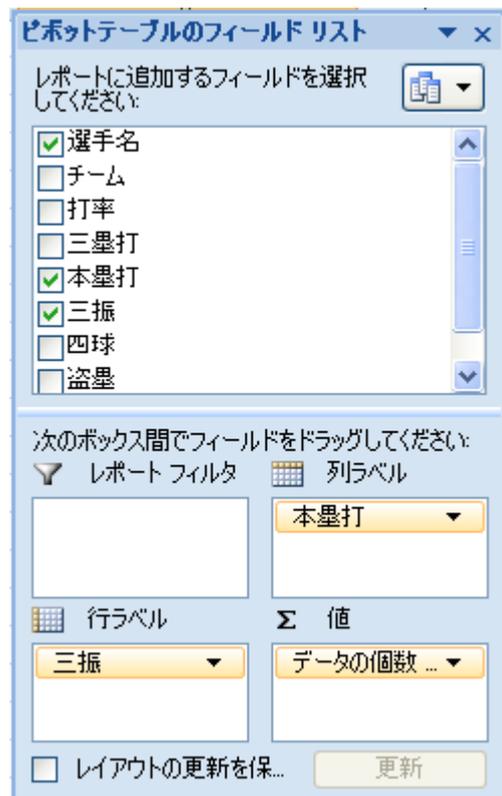
手順

- ① まず、2次元度数分布表にまとめるデータの範囲(ここでは、A1:I25)をドラッグし、ピボットテーブルのボタン(挿入タブにある)をクリックする。
- ② すると、「ピボットテーブルの作成」というウィンドウが開くので、「ピボットテーブルレポートを配置する場所」として、「既存のワークシート」のボタンをクリックし、2次元度数分

¹身長と体重のように、連続データの場合は、ピボットテーブル作成のときに、少し工夫が必要であり、グラフは縦棒どころか、くっついた3-Dグラフとなる。

布表を作成する左上のセルを指定する。

- ③ 右図のようなウィンドウが開くので、「本塁打」をドラッグして「列ラベル」のところに、「三振」をドラッグして「行ラベル」のところに、「選手名」をドラッグして「値」のところにそれぞれ配置する。
- ④ この時点では、本塁打1本ごと(ただし度数のある者のみ)に集計されている。ここで、「本塁打」と書かれているセルをクリックし、「グループフィールド」のボタンをクリックする。(ない場合は、最上部に「ピボットテーブルツール」が出るのでそれをクリックする)
- ⑤ そこで、「先頭の値」を0、「末尾の値」を49、「単位」を10とする(おそらくデフォルトでこうなっている)と、本塁打数が10本単位の階級で分類される。
- ⑥ 三振の方も、同様に「グループフィールド」のボタンをクリックし、「先頭の値」を20、「末尾の値」を139、「単位」を20とすると、下図のような2次元度数分布表が作成される。



データの個数 / 選手名	本塁打				
三振	0-9	10-19	20-29	30-39	総計
20-39	1				1
40-59	2				2
60-79	7	3	1		11
80-99	1	4	1		6
100-119	1	1	1		3
120-139				1	1
総計	12	8	3	1	24

4. 散布図

【課題 15】打者成績のファイルについて、本塁打数と三振数の相関をみるために、散布図を描いてみよう。

📖 手順

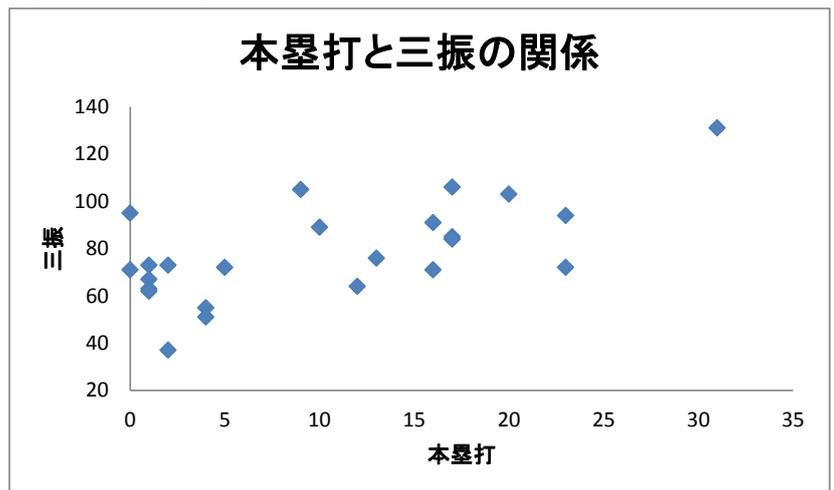
- ① 最初にグラフに描く範囲を範囲指定する。E2:F25をドラッグし、範囲指定する。
- ② グラフを作成するには、**挿入タブ**をクリックすることで、リボン内にグラフのグループが表示される。ここでは、**散布図**のボタンをクリックする。
- ③ すると散布図グラフのフォーマット(型式)メニューが出るので、左上の散布図(マーカーのみ)をクリックする。
- ④ この時点でグラフのサンプルが自動的に描かれている。これを修正していく。まずレイアウトを変更する。リボンの中の「グラフのレイアウト」を展開し、レイアウト1(左上)をクリックし、タイトルや軸ラベルなどが書き込めるようにする。そして、

タイトル:「本塁打と三振の関係」と記入する。

X 軸:「本塁打」と記入する。

Y 軸:「三振」と記入する。

- ⑤ さらにいくつかの細かい修正を加えたものが右図である。右図のようにするには、
- i. 凡例の消去
 - ii. 「軸の書式設定」において、最大値、最小値の変更
 - iii. 目盛線の消去
 - vi. グラフの大きさ変更をおこなっている。



☒ 演習問題

他の指標のくみあわせについても、いろいろ散布図を描いてみよう。隣接しない 2 変量は、1 変量をドラッグした後、**Ctrl** キーを押しながらもう 1 つの変量ドラッグすることで範囲指定できる。

また、複数の変量間の相関係数をいっぺんに計算するには、統計分析を行うための分析ツールを用いればよい。

分析ツールを最初に使用する場合には、アドイン(有効にすること)しなくてはならない。分析ツールのアドインは次のようにおこなう。

- ① 「ファイル」のタブをクリックし、下にある「オプション」のボタンをクリックする。
- ② 「Excel のオプション」のウインドウが開くので、左側の「アドイン」をクリックする。
- ③ 一番下に表示される「Excel アドイン」の右の**設定**ボタンを押す。
- ④ 「分析ツール」にチェックをつけ、OK ボタンをクリックする。

すると、データタブの中に「データ分析」のボタンが出てくるので、下のほうにある、「相関」を選び、ウィザードの要求に従ってデータ範囲を指定すれば、相関係数行列が計算できる。

☒ 本日実習したファイルのうち、**打者成績 2011**の方は河田まで提出すること。なお、本塁打と三振以外もいくつか散布図を描いてみる。提出は

1. Webclass 経由
2. E-mail に添付ファイルとして(送付先アドレスは kawada@tokuyama-u.ac.jp)のいずれでもかまわない。

締め切りは **11月24日(木)9:00** とする。

なお、ファイル名は「打者成績 2011 E41-○○○」のように、**学籍番号**をつけること。