

<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2017**のファイルと、今回あらかじめ河田が作成した**打者成績2017**のファイルを用いる。

【課題 12】講義用 HP から、**打者成績 2017** のファイルをダウンロードしてみよう。

📖 手順

- ① 検索エンジンで、「河田研究室」と入力し検索すると、「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)
- ② 「統計学基礎」をクリックし、「第13回 11月13日(月)」の配布資料にある、「打者成績2017」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2017**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2017**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:D51)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」のところに、「血液型」をドラッグして「行」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、「集計方法」を「個数」とする。
- ④ ここまでで分割表は一応完成であるが、「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、「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、

ピボットテーブルのフィールド... ▼ ×

レポートに追加するフィールドを選択してください: [設定]

番号
 月
 日
 血液型

その他のテーブル...

次のボックス間でフィールドをドラッグしてください:

フィルター	列
	月 ▼
行	Σ 値
血液型 ▼	合計 / 番号 ▼

レイアウトの更新を保留する [更新]

B型というように変更していけば、次の表のようになる。(列ラベル、行ラベルも同様に変更が必要かもしれない。)

血液型と誕生日	誕生日												総計
血液型	1	2	3	4	5	6	7	8	9	10	11	12	
A型	2	1	1	1		1	1			2	2		11
B型	1	1	1		1		1			3		3	11
O型	1	2			1	3	1	2	1	2			13
AB型	1		1			1		1	1		2	1	8
不明	1		1		2			2	1				7
総計	6	4	4	1	4	5	3	5	3	7	4	4	50

【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、血液型も誕生日も質的変量であり、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した状態で、「挿入」タブをクリックして、「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、集合縦棒の右端の「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、「グラフ要素を追加」のタブをクリックし、「グラフタイトル」、「軸ラベル」-「第1横軸」、「第1縦軸」、「奥行き」のボタンをそれぞれクリックする。そして、

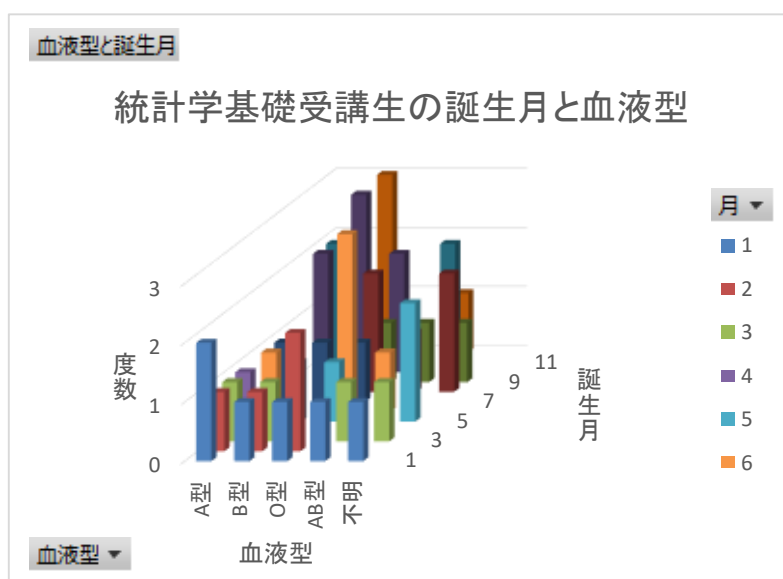
タイトル：「統計学基礎受講生の誕生日と血液型」と記入する。

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：「誕生日」と記入する。

さらに、「度数」と「誕生日」を縦書きにする。軸ラベルを右クリックし、「軸ラベルのオプション」-「文字のオプション」をクリックし、テキストボックスのマークをクリックする。ここで、「文字列の方向」を「縦書き」にすると、上図のようになる。上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

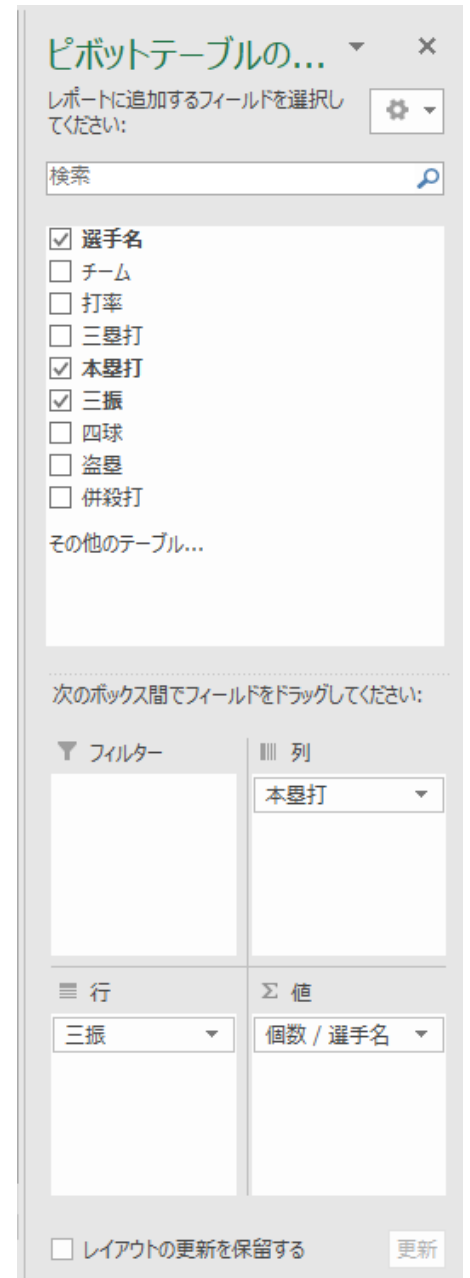
2. では質的変量どうしを、ピボットテーブルで分割表にまとめた。次にこの応用として、とりうる値の多い離散変量どうしについて、2次元度数分布表を作成する。打者成績2017のファイルを用い、本塁打数と三振数の関係を2次元度数分布表にまとめてみよう。

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、グラフは縦棒どうしがくっついた3-Dグラフとなる。

📖 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:I29)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、2次元度数分布表を作成する左上のセルを指定する。
- ③ 右図のようなウインドウが開くので、「本塁打」をドラッグして「列」のところに、「三振」をドラッグして「行」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、「列ラベル」のセルの下の「0」と書かれているセルをクリックし、「フィールドのグループ化」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」と出るのをクリックする)
- ⑤ そこで、「先頭の値」を0、「末尾の値」を39、「単位」を10とすると、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「行ラベル」の下のセルをクリックし、「グループフィールド」のボタンをクリックして、「先頭の値」を40、「末尾の値」を159、「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。(列ラベルを「本塁打」に、行ラベルを「三振」に変更している。)

回数 / 選手名	本塁打					
三振	0-9	10-19	20-29	30-39	総計	
40-59			1		1	
60-79	6	2			8	
80-99	1	3	2	2	8	
100-119	2	3	2	1	8	
120-139	1		1		2	
140-159			1		1	
総計	10	9	6	3	28	



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:F29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「クイックレイアウト」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、

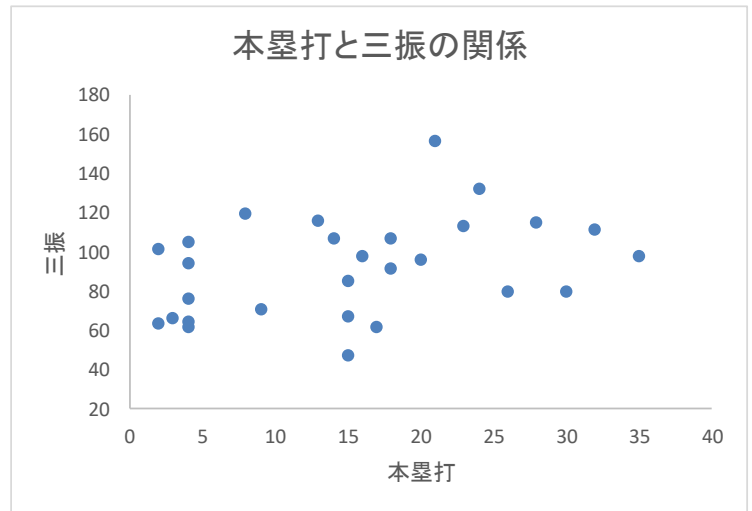
タイトル:「本塁打と三振の関係」と記入する。

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 2017**の方は河田まで提出すること。なお、

- ① **本塁打と三振以外に 1 つ以上散布図**を描いていること。
- ② 別のシート(多くの方は"Sheet4"のはず)に、**相関係数行列**が計算されていること。

に注意しよう。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 20 日(月) 9:15** とする。

なお、ファイル名は「**打者成績 2017 E47-○○○**」のように、**学籍番号をつける**こと。