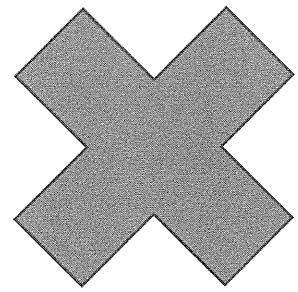
捏造(fabrication)

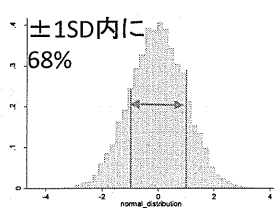


- データの統計学的検討から、捏造は暴かれる
- 遍界不曾蔵(へんかいかってかくさず)

(事例)Kyoto Heart Study糖尿病サブ解析報告
登録時非糖尿病患者群の患者背景

N	2224
HbA1c(%)	5.4±1.8
血清Na	143±30
血清K	4.4±6.8

平均±標準偏差



興梠貴英、山崎力　Kyoto Heart Study不正発覚のきっかけ
日本医事新報　2014/6/7、(No 4702) 45-48

20

(事例)麻酔科学会事件

術後嘔気嘔吐に対する制吐薬使用時の頭痛発症例

	群1	群2	群3	群4	群5	群6	P
研究1	2/30	2/30	2/30	2/30			0.0103
研究2	2/40	2/40	2/40	2/40			0.0066
研究3	3/45	3/45	3/45	3/45	3/45	3/45	0.0103

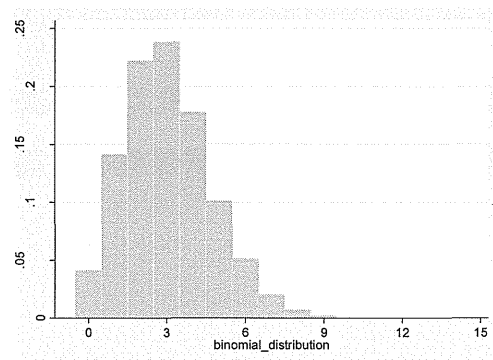
日本麻酔科学会　○○氏論文に対する調査特別委員会報告書　2012年6月
Anesth Analg　2000;90:1000　より抜粋

すべての群で同数の有害事象発症数となる確率(真の発症率が0.0785と仮定)

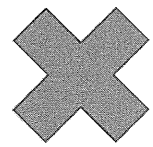
二項分布から予想される発症回数

- 二項分布:重要な離散分布
- 赤玉がpの割合で入っている壺からn個取ることを繰り返した時の赤玉がでる個数の分布。
- n人の患者がいて発症確率pの時の発症者数の分布。

- 45人の対象者で発症率が1/15の時の発症回数の分布

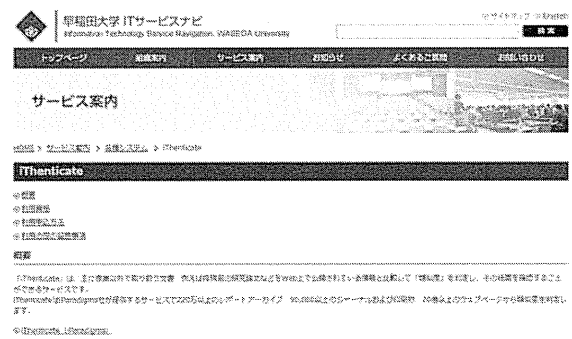


剽窃(plagiarism)



- いわゆる盗作
- 自分自身の過去論文からのコピーペーストも自己剽窃(self-plagiarism)といわれ、行うべきではない

剽窃検知ツール
(CrossCheck、
iThenticateなど)が普及



二重出版(duplicate publication)

- 二重出版は、版権の侵害。
- そもそも二重投稿ではないと申告して雑誌に掲載されているはず。
- 異なる研究のように偽るとすれば、読者の信頼を裏切る。

- 許容される場合:学会抄録の論文化など二次的な出版が容認される場合がある。出版社や関係者の許諾のもとで透明性に配慮してなされるべき。

著者資格(authorship)

- 誰が著者になるのか、どの順で著者名を記載するか問題。
- 研究グループ内で論文執筆前に相談。

- 国際医学雑誌編集者委員会の勧告を参照。
  1）構想、計画、データ取得、解析、解釈に対する実質的貢献
  2）重要な内容の執筆や改訂
  3）記載内容に対する最終承認
  4）研究の正確性(accuracy)、統一性(integrity)に関する説明責任

共著者の責任が問われた事例

- 医科歯科大教授に停職処分　iPS虚偽発表問題で　（日本経済新聞2012/12/28より抜粋）

- 森〇尚〇氏がiPS細胞の臨床研究をしたと虚偽の発表をした問題で、東京医科歯科大は28日、森〇氏の元指導教官の〇〇〇〇教授について、研究内容を検証せずに共著者となった論文が20本あったとする調査結果を発表した。〇〇教授を26日付で停職2カ月としたほか、森〇氏に支払った総額130万円余りの経費について〇〇教授への返還請求を決めた。

生物統計の基礎と多変量解析入門

大久保 智哉

独立行政法人 大学入試センター 研究開発部
試験評価解析研究部門

Contents

1 はじめに **5**

1.1 統計モデルによってデータを分析する必要性 5

1.2 課題 8

2 記述統計量 **9**

2.1 平均値・中央値・四分位点 9

2.2 記述統計量の解釈 11

2.3 相関性・共変性の検討 12

2.4 相関係数と回帰係数 14

2.5 偏相関係数 14

2.6 課題 16

3 検定統計量 **17**

3.1 統計的仮説検定の原理 17

3.2 統計的仮説検定の考え方 18

3.3 検定をおこなう前に 19

3.4 サンプルサイズと p 値の関係 20

3.5 課題 22

4 変量の検定 **23**

4.1 統計的仮説検定の過誤 23

4.2 サンプルサイズの計算 23

4.3 仮説検定法 24

4.4 Welch の t 検定 26

4.5 対応のある t 検定 27

4.6 分散分析 28

4.7 F 検定 30

4.8 相関の検定 31

4.9 偏相関の検定 32

4.10 Wilcoxon の順位和検定 (Mann-Whitney の U 検定) 33

4.11 Kruskal-Wallis の検定 34

4.12 Spearman の順位相関係数の検定 35

4.13 Fisher の正確検定 (χ^2 検定) 36

4.14 χ^2 検定 37

4.15 McNemar の検定 39

4.16 Cochran の Q 検定 40

4.17 課題 40

A　R と Rcmdr　41
A.1 R のインストール 41
A.2 Rcmdr のインストール 43
A.3 Rcmdr の起動と終了 46
A.4 パッケージのインストール 47
A.5 パッケージの読み込み 48
A.6 R による単純計算 49
A.7 R によるオブジェクトへの付値 49
A.8 R におけるパスの設定 50
A.9 オプション・自作関数の読み込み 51
A.10 課題 52

1

はじめに

§ 1.1
統計モデルによってデータを分析する必要性

ここでは，統計モデルを用いて分析することの重要性について説明をします．時折，「統計分析を使って，小難しい分析をしてもクライアントに説明が出来ないし，クライアントも理解できない．それよりも単純な集計データを効果的に見せて理解してもらう方が重要で，分析者の適切な態度である」というロジックに出くわすことがあります．これでは，誤った知見に導かれてしまうことになります．

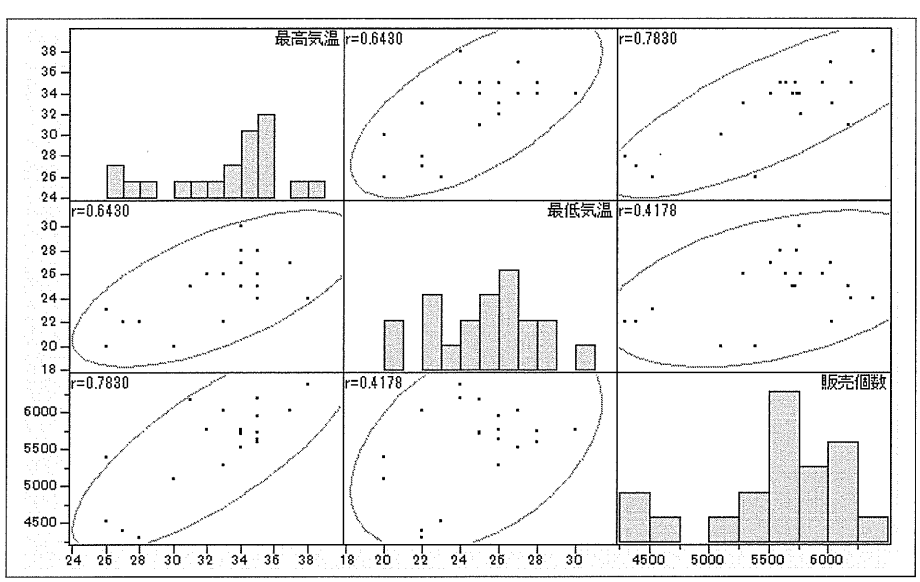
統計モデルの意義は単に「数学的表現の付与」や「情報の可視化」，「データの縮約表現」ではありません．実際の利用状況において，おそらく最も重要な点は「交絡要因の除去」です．実際に得られるデータは関心の対象となる事象に対して多くの交絡要因を含んでいます．

例を用いて考えます．次のデータは，20 日分のアイスクリームの販売個数とその日の最高気温と最低気温です．データに基づいて「気温が販売にどのような影響を及ぼすか」を検討することとします．

1.1 アイスクリーム売上データ

最高気温	最低気温	販売個数
35	28	5590
35	25	5723
34	27	5517
38	24	6369
35	26	5946
32	26	5762
27	22	4393
33	22	6027
28	22	4301
34	30	5758
34	28	5735
35	26	5635
34	25	5702
26	23	4527
35	24	6190
37	27	6020
26	20	5388
33	26	5280
31	25	6169
30	20	5099

「単純な集計」で導かれることを以下の図に表現してみました．変数 (販売個数，最高気温，最低気温) ごとのヒストグラムに加えて，変数間の散布図が描かれていますので,「単純な集計」で抽出できる最大限の情報がこの図に表れています．



1.1 アイスクリーム売上データの多変量散布図

少なくともこの図から導かれる事実は以下の点です．

1. 最高気温が高い日は最低気温も高い傾向にある．(最高気温と最低気温の散布図より)
2. 最高気温が高い日は販売個数も高い傾向にある．(最高気温と販売個数の散布図より)
3. 最低気温が高い日は販売個数も高い傾向にある．(最低気温と販売個数の散布図より)

しかし，統計分析をすると次の表現のうち１つは間違いであることがわかります．

1. 最高気温が上がれば，販売個数も増える傾向にある．
2. 最低気温が上がれば，販売個数も増える傾向にある．

なお，統計分析の結果は次のようなものでした．

1.2 分析モデルの推定値

切片	最高気温	最低気温
1588.08	145.91	−31.78

実際には，これは線形モデルと呼ばれる分析をした結果です．線形モデルについては後の章で説明されますが，

$$\text{販売個数} = \beta_0(\text{切片}) + \beta_1 \times \text{最高気温} + \beta_2 \times \text{最低気温} + e \tag{1.1}$$

という１次式によって，販売個数に対する最高気温と最低気温の関係性を記述しようとするモデルです．

統計モデリングとは，上記のようなモデルを設定し，β_0，β_1，β_2，e の値を定める作業であると言い換えることも出来ます．

さて，上記線形モデルにおける推定値の解釈は次のようにおこないます．

> 説明側の他の変数の値は変わらずに，当該変数の値が１単位上がったときに，被説明側の変数の値がいくつ上がることが期待されるか．

したがって，先の分析結果の解釈は

> 最低気温が同じ日の比較の場合，最高気温が１度高いと，おおよそ 146 個販売個数が多い．一方，最高気温が同じ日の間の比較の場合，最低気温が１度高くなると，おおよそ 32 個売れなくなる．

となります．

つまり，「最低気温が上がれば，販売個数も増える傾向にある.」というのは誤りということになります.「誤り」と言い切る理由は後述します．

「最低気温が高い日は販売個数も高い傾向にある」と「最低気温が上がれば，販売個数も増える傾向にある」という表現の違いは，前者の表現では，最低気温と販売個数の共変性について説明をしたもので，実際には両者の間に直接の関係性が無くとも，最低気温と関連性のある別の変数が販売個数との間により強い共変関係を持つことによる間接的な影響による関係性をも含む表現です．具体例としては，最低気温と関連性のある最高気温という要因があり，最高気温がより強く販売個数との共変関係を持っていた場合には，最低気温と販売個数の間に共変性がない場合でも，最低気温と関連性のある最高気温の販売個数との共変性を引きずる形で最低気温と販売個数の間に見せかけの関連性が出てきます．一方，後者は最低気温と販売個数の間のみの共変性についての表現となっています．統計学の分野では前者を「回帰係数の解釈」，後者を「偏回帰係数の解釈」と言います．

実際，分析結果と既存の情報から導かれる知見は次のようなものです．

1. 最高気温と最低気温は共変する傾向にある．
2. １日の気温 (最高気温や最低気温で表される) が高いとアイスクリームは売れる傾向にある．
3. 最高気温が高いのに最低気温が低い日は，放射冷却 [*1] のメカニズムが働いた晴天の日だと考えられる．一方，最高気温が高くて最低気温も高い日は曇りの日である．
4. アイスクリームは最高気温が同じであれば最低気温が高い日 (すわなち曇りの日) より，最低気温は低い日 (すわなち晴天) の方がよく売れる．

[*1] 放射冷却とは，地面が熱を放射することで気温が下がる現象のこと．曇りの日よりも晴れの日に度合いが強くなる．

5. アイスクリームの売り上げ要因として，気温に加え，日射の状況が要因として考えられることが示唆された．

データ解析において重要なことは，データから得られる情報に既存の情報を加えて，包括的にデータの背後にあるメカニズムを理解することで適切な知見を得ることです．

複雑な統計モデルを用いたから，より深い知見が得られるかというとそうではありません．複雑な統計モデルは，データが特別な構造をしている場合に，その特殊性に対応するために作られたものです．一般的なデータに対しては，単純な統計モデルで十分にデータから情報を引き出せます．データ解析において目指すところは，複雑なモデルを使いこなすことよりも，分析によって得られた情報と当該分野の既存の情報を組み合わせていかに深い知見を引き出せるか．になります．

先ほどの例では，データ分析の結果に加えて，自然メカニズムに対する理解 (放射冷却) があってはじめて適切な解釈が可能になりました．

宇宙人が人間を調査すると「人間は平均的に１つの卵巣を持つ」という誤った理解をする

統計学上の (簡単な) 知識だけではなく，その分野に関する知識や深い理解が不可欠なものであることがわかります．統計モデリングはあくまで道具なのです．

§ 1.2
課題

次の論の展開において間違いがあります．どのような誤りの可能性があるか考えてください．

1. 高校生の喫煙と学力の関係について調査をおこなった．
2. ある高校の全校生徒に対して同じ数学のテストを実施した．
3. 高校生を喫煙群と非喫煙群に別け，それぞれにおいてテストの平均得点を算出した．
4. その結果，喫煙群の方がテストの平均得点が高かった．
5. 同様のことを，物理でも英語でも実施したところ同様の結果が出た．また，学校を変えても同じであった．
6. したがって，喫煙は (喫煙によるリラックス効果などによって) 学力の向上に資する．

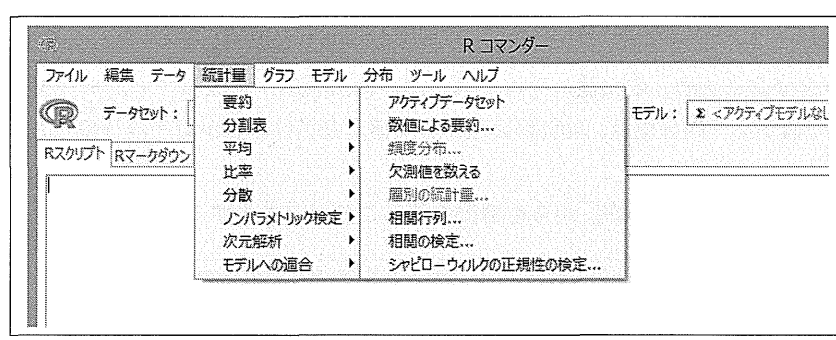
2

記述統計量

§ 2.1
平均値・中央値・四分位点

ここでは，Ice_Cream01.csv を使います．読み込んでください．データのオブジェクト名は Ice01 としておきます．

1. データセットをアクティブにします (【データセット：】からデータを選択).
2. メニューバーの【統計量】→【要約】→【アクティブデータセット】を選択します.



2.1

出力

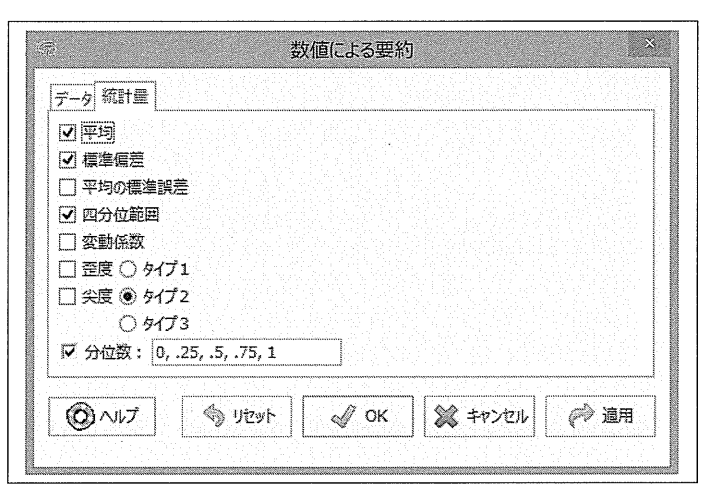
```
> summary(Ice01)
    MaxTemp         MinTemp         Quantity
 Min.   :26.00   Min.   :20.00   Min.   :4301
 1st Qu.:30.75   1st Qu.:22.75   1st Qu.:5361
 Median :34.00   Median :25.00   Median :5712
 Mean   :32.60   Mean   :24.80   Mean   :5557
 3rd Qu.:35.00   3rd Qu.:26.25   3rd Qu.:5964
 Max.   :38.00   Max.   :30.00   Max.   :6369
```

2.2

```
> summary(Ice01)
    MaxTemp         MinTemp         Quantity
 Min.   :26.00   Min.   :20.00   Min.   :4301
 1st Qu.:30.75   1st Qu.:22.75   1st Qu.:5361
 Median :34.00   Median :25.00   Median :5712
 Mean   :32.60   Mean   :24.80   Mean   :5557
 3rd Qu.:35.00   3rd Qu.:26.25   3rd Qu.:5964
 Max.   :38.00   Max.   :30.00   Max.   :6369
```

上から，最小値，第一四分位点 (値)，中央値，平均，第三四分位点 (値)，最大値です．データを昇順に並べ，一番初めの値が最小値，25%目の値が第一四分位点，50%目の値が中央値，75%目の値が第三四分位点，最後の値が最大値となります．

他にも，以下のように【数値による要約】からも情報を得ることができます．



1. メニューバーの【統計量】→【要約】→【数値による要約】を選択します．

```
> numSummary(Ice01[,c("MaxTemp", "MinTemp", "Quantity")], statistics=c("mean", "sd",
+  "IQR", "quantiles"), quantiles=c(0,.25,.5,.75,1))
            mean         sd    IQR   0%     25%    50%     75% 100%  n
MaxTemp    32.60   3.515380   4.25   26   30.75   34.0   35.00   38 20
MinTemp    24.80   2.687202   3.50   20   22.75   25.0   26.25   30 20
Quantity 5556.55 584.943225 603.50 4301 5361.00 5712.5 5964.50 6369 20
```

sd とは標準偏差 (Srandard Deviation) のことで，散らばり具合を表す指標です．また，IQR とは Inter Quartile Rang の略で，第三四分位値から第一四分位値を引いたもので，散らばり具合を表しています．

変数 $x(i = 1, 2, \ldots, N)$ の平均 $m(x)$ は

$$m(x) = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{2.1}$$

で表され，分散は $var(x)$

$$var(x) = \frac{1}{N}\sum_{i=1}^{N}(x_i - m(x))^2 \tag{2.2}$$

その平方根を取ったものが標準偏差と呼ばれます．

$$sd(x) = \sqrt{var(x)} \tag{2.3}$$

さらに，2 変量間の共分散を

$$cov(x, y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - m(x))(y_i - m(y)) \tag{2.4}$$

と定義し，基準化した

$$cor(x, y) = \frac{cov(x, y)}{sd(x)sd(x)} \tag{2.5}$$

を相関係数と呼びます．

§ 2.2
記述統計量の解釈

記述統計量とは，ある変数の分布を縮約的に表現したものです．詳細な情報を見せる必要があれば，分布そのものをヒストグラムなどで見せるのが効果的です．変数間の比較を見せたのであれば，箱ひげ図などが効率的でしょう．

しかし，もっと単純に分布の情報を提供するためには記述統計量を報告することになります．また，ヒストグラムや箱ひげ図では細かな数値までは読み取れないので，その点からも記述統計量の報告が求められる状況は多いです．

記述統計量の報告パタンは 2 つあります．ほとんどの場合，いずれかに当てはまります．

1. データが正規分布に近い形状を取る場合：代表値としては，平均値を選び，散らばりの程度を表す指標として標準偏差 (分散) を示す．
2. それ以外の場合：代表値として，中央値を選び，第一四分位点と第三四分位点を併せて示す．

ここで，例として，2 つの国 (A 国と B 国) の年収の分布を考えてみます．

実際のデータを得るのではなく，今回は乱数からデータを発生させてみます．乱数を用いていますので，記述統計量は当然，ヒストグラムの形状も毎回異なるはずです．

次のコードを R コンソールか Rcmdr のスクリプトに張り付けて実行してみて下さい.

```
# 自由度 7 のカイ二乗分布から乱数を 10,000 個発生させ, 60 をかける.
IncomeA <- rchisq(10000, 7)* 60

# 平均 450, 標準偏差 180 の正規分布から乱数を 10,000 個発生させる. ただし, 0 未満は 0 に置き換える.
IncomeB <- rnorm(10000, 450, 180)
IncomeB <- ifelse(IncomeB < 0, 0, IncomeB)

# 図の表示画面を縦 2, 横 1 の図が入るように設定を変更する.
par(mfcol = c(2,1))

# ヒストグラムの描画
hist(IncomeA, main = "", xlab="Annual income of A", ylab="Frequency",
     xlim = c(0, 1200), ylim=c(0, 1100), breaks=seq(0,2500,50))
hist(IncomeB, main = "", xlab="Annual income of B", ylab="Frequency",
     xlim = c(0, 1200), ylim=c(0, 1100), breaks=seq(0,2500,50))

# 図の表示画面を縦 1, 横 1 の図が入るように設定を戻しておく.
par(mfcol = c(1,1))

# 数値の要約
numSummary(IncomeA, statistics=c("mean", "sd", "quantiles"), quantiles=c(0.25, 0.5, 0.75))
numSummary(IncomeA, statistics=c("mean", "sd", "quantiles"), quantiles=c(0.25, 0.5, 0.75))
```

```
> numSummary(IncomeA, statistics=c("mean", "sd", "quantiles"), quantiles=c(0.25, 0.5, 0.75))
     mean       sd      25%      50%      75%     n
 420.4841 224.3602 257.7351 382.2292 540.1652 10000
> numSummary(IncomeA, statistics=c("mean", "sd", "quantiles"), quantiles=c(0.25, 0.5, 0.75))
    mean       sd      25%      50%      75%     n
 452.9429 180.0066 328.2933 452.8467 576.8801 10000
```

§ 2.3
相関性・共変性の検討

まず, 相関について説明をする前に相関と因果は異なるものであるということを理解してください.

相関係数 $cor(x,y)$ は, 2 つの変数間の共変性の強さを示したもので, $-1 \leq cor(x,y) \leq +1$ の値を取ります. 一つの指標として, $0 \leq |cor(x,y)| < 0.2$ を「相関がない」$0.2 \leq |cor(x,y)| < 0.4$ を「弱い (正の or 負の) 相関がある」$0.4 \leq |cor(x,y)| < 0.7$ を「中程度の (正の or 負の) 相関がある」$0.7 \leq |cor(x,y)|$ を「強い (正の or 負の) 相関がある」と表現します.

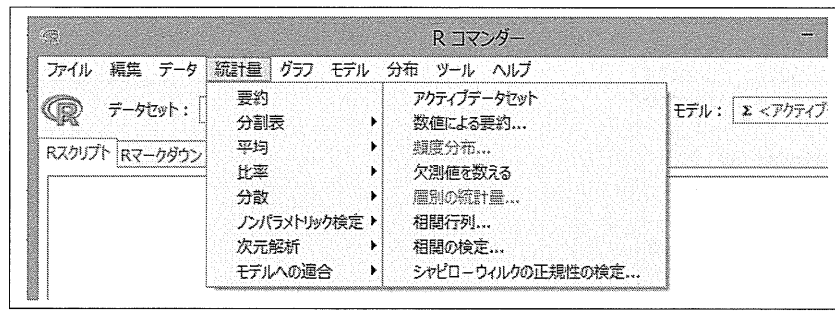
ただし, これはあくまで参考です. データ解析では, 数値から「大きい・小さい」の解釈を含んだ瞬間に読み手ごとに異なる印象を与えます. 結果の報告の際には, このような言葉での報告は避けるようにして下さい. 解釈として, 最大限注意を払った上で表現をおこなうようにして下さい.

それでは, Ice01 のデータで相関係数を計算してみます.

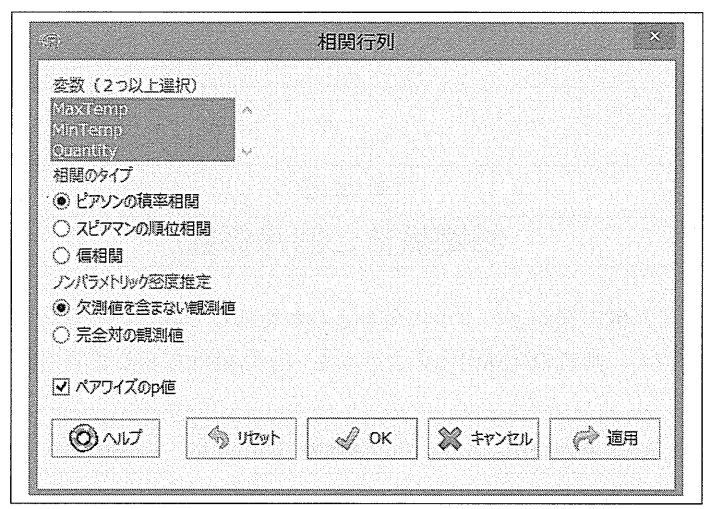
1. データセットをアクティブにします (【データセット：】からデータを選択).
2. メニューバーの【統計量】→【要約】→【相関行列】を選択します.
3. 【変数】の中から変数を 2 つ以上入れます.
4. 【相関のタイプ】については, 連続量の場合にはピアソンの積率相関係数を選び, 順序尺度の場合にはスピアマン

の順位相関係数を選びます

5. 【ペアワイズのｐ値】にチェックを入れ,【OK】を押します.



2.3



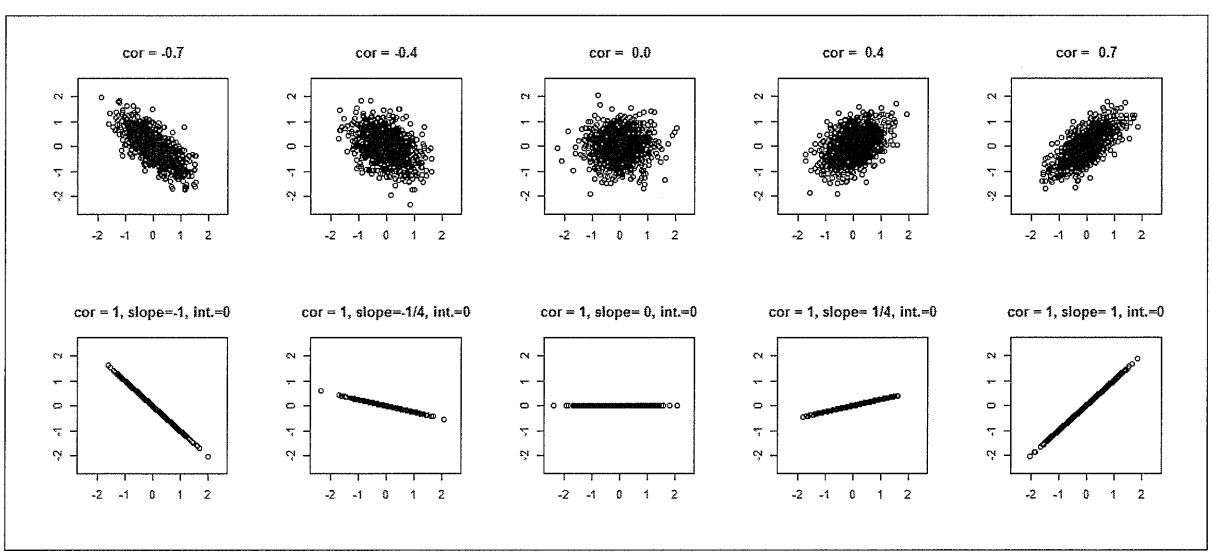
2.4

Rcmdr のスクリプトに以下のように実行され，結果が表示されます.

```
> rcorr.adjust(Ice01[,c("MaxTemp","MinTemp","Quantity")], type="pearson", use="complete")

 Pearson correlations:
          MaxTemp MinTemp Quantity
MaxTemp     1.000  0.6430   0.7830
MinTemp     0.643  1.0000   0.4178
Quantity    0.783  0.4178   1.0000

 Number of observations: 20

 Pairwise two-sided p-values:
          MaxTemp MinTemp Quantity
MaxTemp           0.0022  <.0001
MinTemp   0.0022          0.0668
Quantity <.0001   0.0668

 Adjusted p-values (Holm's method)
          MaxTemp MinTemp Quantity
MaxTemp           0.0045  0.0001
MinTemp   0.0045          0.0668
Quantity  0.0001  0.0668
```

§ 2.4
相関係数と回帰係数

ここでは，相関係数によって，散布図がどのように変わるのかを示してみたいと思います (図上段)．さらに，後に出てくる「回帰係数」も条件に加え，両者の関係がどのようなものかも併せて示します (図下段)．



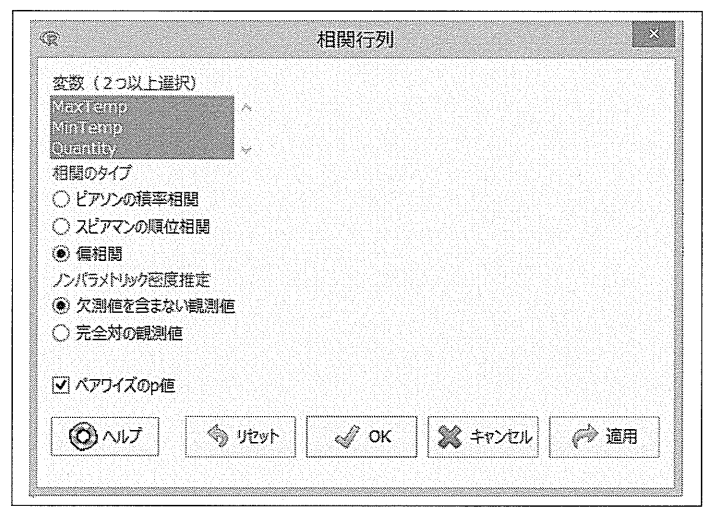
§ 2.5
偏相関係数

偏相関係数とは，多変量データにおいて関心下の 2 変数間の相関をそのほかの変数による影響を除外した形で算出さ

れる相関係数です．関心下の変数を x, y とし，局外変数 z の影響を除外した偏相関係数 $cor(x,y)_z$ は次のように定義されます．

$$cor(x,y)_z = \frac{cor(x,y) - cor(x,z)cor(y,z)}{\sqrt{1-cor(x,z)^2}\sqrt{1-cor(x,z)^2}} \tag{2.6}$$

Rcmdr では，以下の手順によって出力します．



1. 【要約】→【相関行列】を選択する．
2. 【相関のタイプ】で偏相関を選択し，【OK】を押す．

Rcmdr では，次のようなスクリプトが生成され，併せて結果も表示されます．結果を見ると，やはり「最低気温」と「販売個数」の偏相関は負になっていることが見て取れます．

```
> partial.cor(IceCream[,c("MaxTemp","MinTemp","Quantity")], tests=TRUE, use="complete")

 Partial correlations:
          MaxTemp  MinTemp Quantity
MaxTemp   0.00000  0.55887  0.73924
MinTemp   0.55887  0.00000 -0.17982
Quantity 0.73924 -0.17982  0.00000

 Number of observations: 20

 Pairwise two-sided p-values:
          MaxTemp MinTemp Quantity
MaxTemp           0.0129  0.0003
MinTemp   0.0129          0.4613
Quantity 0.0003  0.4613

 Adjusted p-values (Holm's method)
          MaxTemp MinTemp Quantity
MaxTemp           0.0257  0.0009
MinTemp   0.0257          0.4613
Quantity 0.0009  0.4613
```

§ 2.6
課題

課題 1

1. Smoke03.csv を読み込み，score に関して分布を確認した上で，代表値と散らばり具合を報告して下さい．

3

検定統計量

§ 3.1
統計的仮説検定の原理

ここでは，統計的仮説検定の原理について単純に説明します．

今，手元にあるコインが歪んでいます．このコインを用いてゲームをするのですが，不公平がないように「このコインの裏表の出る可能性は等しい」ということを検討したいと思います．

1. そこで「このコインで表の出る確率が裏の出る確率に等しい」という仮説を立てることにしました．
2. 次に，実際に 20 回コインを投げてデータを取ったところ，表が 5 回 (裏は 15 回) でした (比率の推定値 =0.25)．
3. 表と裏が同等に出るのであれば，表が 5 回以下もしくは 15 回以上出る確率 (推定値よりも極端な値が出る確率) は 0.0295 です．
4. つまり，歪んでいないコインであれば，3% 程度しか起こりえないことが，このコインでは起こったことになります．5% くらいの偶然性であれば許容するつもりでした．
5. そこで，許容する確率よりも小さい確率で起こる事象が起こったことになってしまったので，最初に立てた仮説「このコインで表 (裏) の出る確率は 1/2 である」自体が誤っていたと考えます．

統計的仮説検定の枠組みでは，上記の内容を次のように表現します．

1. 帰無仮説を「コインの表が出る確率 = 1/2」とする．
2. 標本データから比率の母数 (母比率) の推定値を計算する (= 5/20)．
3. 帰無仮説 (コインの表の出る確率 = 1/2) のもとで，p-value(p 値) を計算する．
4. p-value(p 値) があらかじめ設定した有意水準 (5%) を下回るかどうかを判定する．
5. p 値が有意水準よりも小さかったので，帰無仮説 (このコインで表 (裏) の出る確率は 1/2 である) を棄却し，「このコインで表 (裏) の出る確率は 1/2 であるとは言えない」と結論づけます．

```
> round(dbinom(0:20, 20, 0.5), 5) # round(, 5); 小数点第5位での四捨五入. # dbinom(); 2項分布 (x, size, prob)
 [1] 0.00000 0.00002 0.00018 0.00109 0.00462 0.01479 0.03696 0.07393 0.12013 0.16018 0.17620
[12] 0.16018 0.12013 0.07393 0.03696 0.01479 0.00462 0.00109 0.00018 0.00002 0.00000
>  barplot(dbinom(0:20, 20, 0.5), names.arg=c(0:20)) # barplot(, ラベル)
```