



R Commander を用いた統計解析の基礎(3)

岡田 昌史

1. サンプルサイズの計算 (パワーアナリシス)

「研究者が、このくらいあるはずだと考えている差を、1回の研究で確実に検出するために、十分な検出力をもった研究を行う」ために行われるのがサンプルサイズの計算、いわゆるパワーアナリシスというものです。もちろん、小さなサンプルサイズであっても、偶然良い標本が得られ、差を検出できる場合もあるわけですが、偶然に頼るようではお金と時間の無駄が多いですから、なるべく確実に、その差が本当にあるものであれば、1回の実験・調査で見つけ出したいものです。

さて、検出力(= $1.00 - \text{False Negative}$ を起こす確率(β))は何で決まっていたでしょうか。測定値の分布あるいは統計量の分布 (検定手法)、あると思われる群間の測定値の差、測定値の標準偏差 (これは往々にして先行研究からの推測になります)、サンプルサイズ、False Positive を起こす確率 (α 、通常 0.05)、以上の5つです。そして、同じ検定手法を使う限りにおいては、検出力、測定値の差、測定値の標準偏差、サンプルサイズ、 α の5つのうち、4つまでが決まれば他の1つは決定されます。このことを用いて、既存の研究計画の検出力を算出したり、あるいは期待される検出力を得るために必要なサンプルサイズを算出したりするわけです。

R の base パッケージに組み込まれているパワーアナリシスは、t 検定(平均値の差)、カイ 2 乗検定(割合の差)、分散分析(多群の平均値の差)の3つの検定手法に対応しています。ただ、これらは R コマンドではサポートされていません。しかし、使い方は非常に簡単ですので、今回は R コンソールから直接コマンドを打って算出してみましょう。"R Console" ウィンドウの赤い文字で出ている ">" のあとにコマンドを打ち込んでいき、すべて打ち込めたらエンターキーを押します。

```
power.t.test(delta=0.5, sd=2, sig.level=0.05, power=0.8, n=NULL)
```

これは、以下のような意味です。

- ・ 検定手法は t 検定
- ・ 検出力は 0.8(False Negative を起こす確率は $1-0.8=0.2$, 20%)
- ・ 測定値の群間の差 0.5 を検出したい
- ・ 測定値の標準偏差は 2 ぐらいだと思う
- ・ 有意水準(False Positive を起こす確率)は 0.05, 5%
- ・ サンプルサイズは不明 (これを求めたい)

この結果は以下ようになります(R Console ウィンドウに直接結果が出力されます)。



```

R Console
ファイル 編集 その他 パッケージ ヘルプ

また、RやRのパッケージを出版物で引用する際の形式については
'citation()'と入力してください。

'demo()'と入力すればデモをみることができます。
'help()'とすればオンラインヘルプが出ます。
'help.start()'でHTMLブラウザによるヘルプがみられます。
'q()'と入力すればRを終了します。

Tcl/Tk インターフェイス ロード中... 完了
[以前にセーブされたワークスペースを復帰します]

> power.t.test(delta=0.5, sd=2, sig.level=0.05, power=0.8,n=NULL)

Two-sample t test power calculation

      n = 252.1281
  delta = 0.5
    sd = 2
sig.level = 0.05
  power = 0.8
alternative = two.sided

NOTE: n is number in *each* group

> █

```

n,delta, sd,sig.level, power は入力した通りの値ですが、 n=252.1281(n is number in *each* group) という結果が出ています。つまり、このような測定値で 80%の検出力を得るためには、 $253*2=506$ 人程度のサンプルサイズが必要、ということになりますね。

カイ 2 乗検定の場合もやってみましょう。

```
power.prop.test(p1=0.15, p2=0.30, sig.level=0.01, power=0.85, n=NULL)
```

この場合は、2つのグループで現象の発生率の違いをみることになりますから、p1 と p2 に想定される（検出したい）確率をいれます。たとえばニコチンガム投与群と非投与群で、禁煙の成功率を比較したいような場合ですね。投与群では 30%、非投与群では 15%という結果が得られることを推測している場合、それを検出力 85%、有意水準 1%で、確実に検出するためにはどのくらいのサンプルサイズが必要か、ということです。

```

> power.prop.test(p1=0.15, p2=0.30, sig.level=0.01, power=0.85, n=NULL)

Two-sample comparison of proportions power calculation

      n = 200.3680
     p1 = 0.15
     p2 = 0.3
sig.level = 0.01
  power = 0.85
alternative = two.sided

NOTE: n is number in *each* group

> ?power.anova.test
> █

```

このように、各群で 201 名、合計 402 名のサンプルが必要であることがわかりました。

続いて、分散分析の場合。

```
power.anova.test(groups=4, between.var=1, within.var=3, sig.level=0.05, power=0.90,  
n=NULL)
```

分散分析では、群の数、群間の測定値の分散、群内の測定値の分散が必要になります (t 検定のときの標準偏差と同様、少ない情報から推測するしかない場合も多いですが)。

ここでは、4 群、群間分散は 1.0、群内分散は 3.0、有意水準 5%、検出力 90% で計算してみましよう。

```
> power.anova.test(groups=4, between.var=1, within.var=3, sig.level=0.05, power=0.90, n=NULL)
```

```
Balanced one-way analysis of variance power calculation
```

```
groups = 4  
n = 15.18834  
between.var = 1  
within.var = 3  
sig.level = 0.05  
power = 0.9
```

```
NOTE: n is number in each group
```

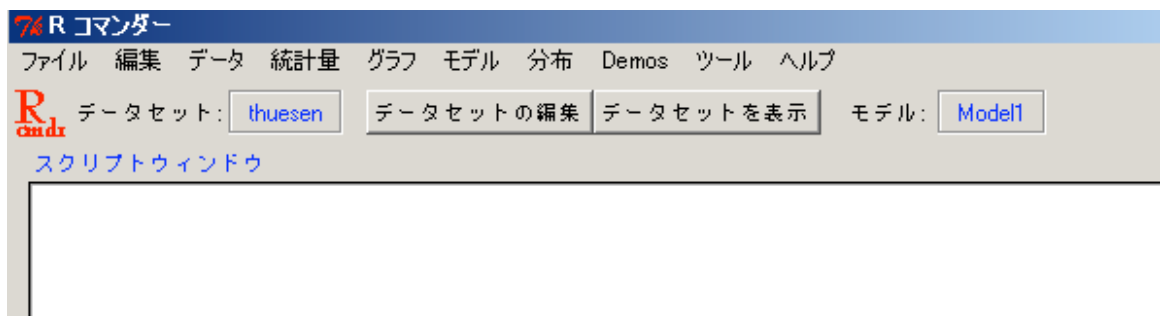
4 群ですが、各群 16 名、合計 64 名が必要なことがわかりました。

なお、コマンドーの "Demos" メニュー、"Power of the test" からは、検定方法が正規分布の場合に、検出力(Power)以外の 4 つのパラメータ(Sample Size, Standard Deviation(測定値の標準偏差), True Difference(測定値の差), Alpha Level(α))を動かすことで検出力がどのように変化するかをグラフィカルに見ることができます。

2. 回帰分析(2)

それでは、前回の回帰分析の続きに移りましょう。

コマンドーを立ち上げ、ISwR パッケージ内の thuesen データセットを呼び出してください。前回の終了時に作業スペースを保存していれば、その中に thuesen のコピーの保存されているはずですから、「アタッチされたパッケージからデータセットを読み込む...」をやらなくても、コマンドーの赤字の「アクティブデータセットなし」のところをクリックして選択するだけでアクティブデータセットとすることができます。同様に、赤字の「アクティブモデルなし」のところをクリックすれば、前回作成した「Model1」をアクティブモデルとして読み込んでおくことができますので、Model1 も読み込んでおいてください。下記のような状態になりましたか？



Model1 では、blood.glucose で short.velocity を

$$\text{short.velocity} = 1.10 + 0.0220 \times \text{blood.glucose}$$

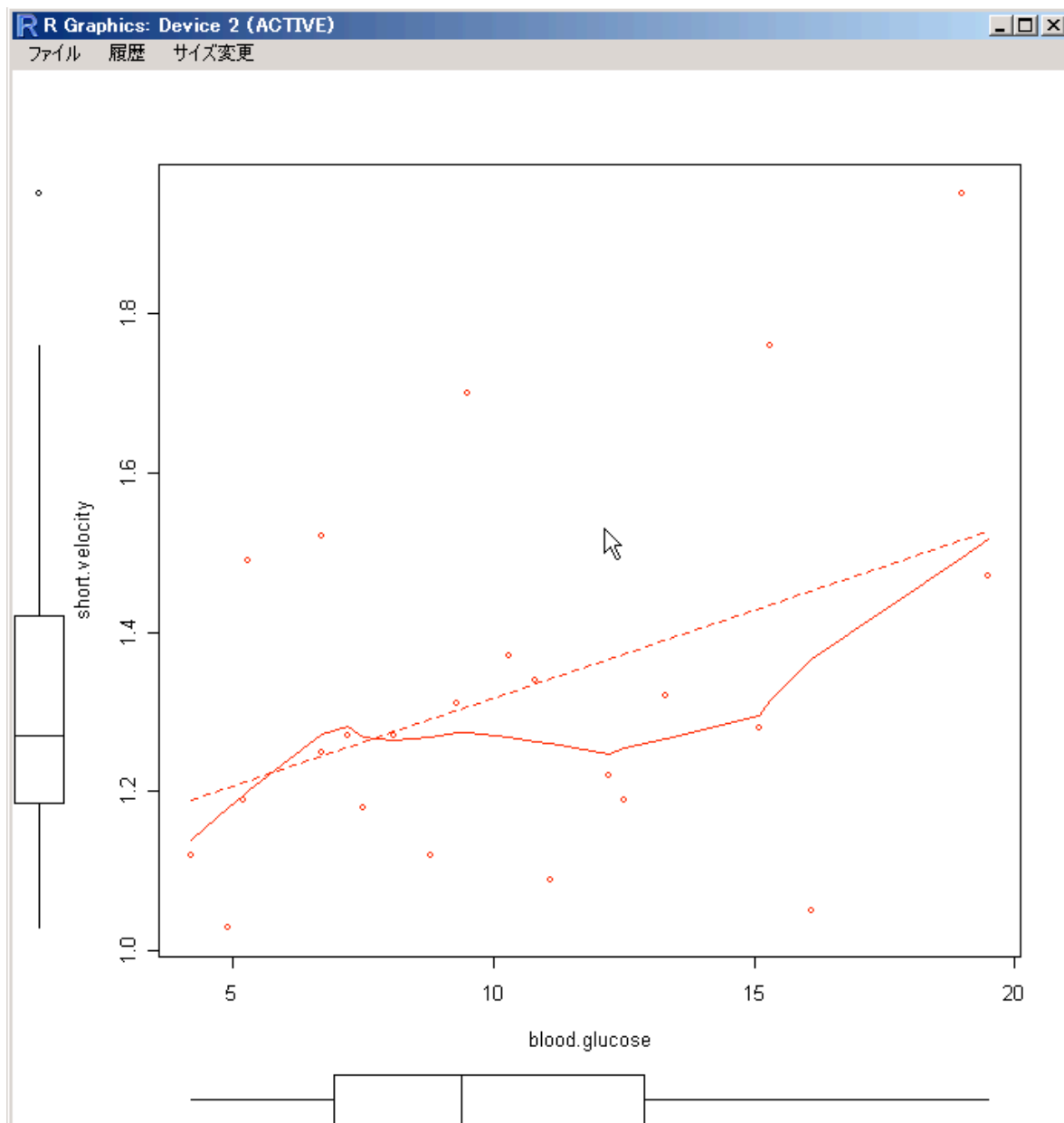
という数式によって「説明」することができる、と仮定しました。もしそういう仮定ができるのであれば、blood.glucose の値から、short.velocity を「予測」することもできますし、また、こ



の 0.0220 という回帰係数が、血糖値が心臓の収縮力に与える生理学的なメカニズムを推定する鍵になることもあるのです。

しかし、それらはすべて、この数式, Model1 が、実際の測定値に「よくあてはまっている」場合にのみいえることでしょう。では、どのくらい「よくあてはまっている」のかをどのように判断するのでしょうか。

まずは、やはりグラフです。散布図を書いてみましょう。

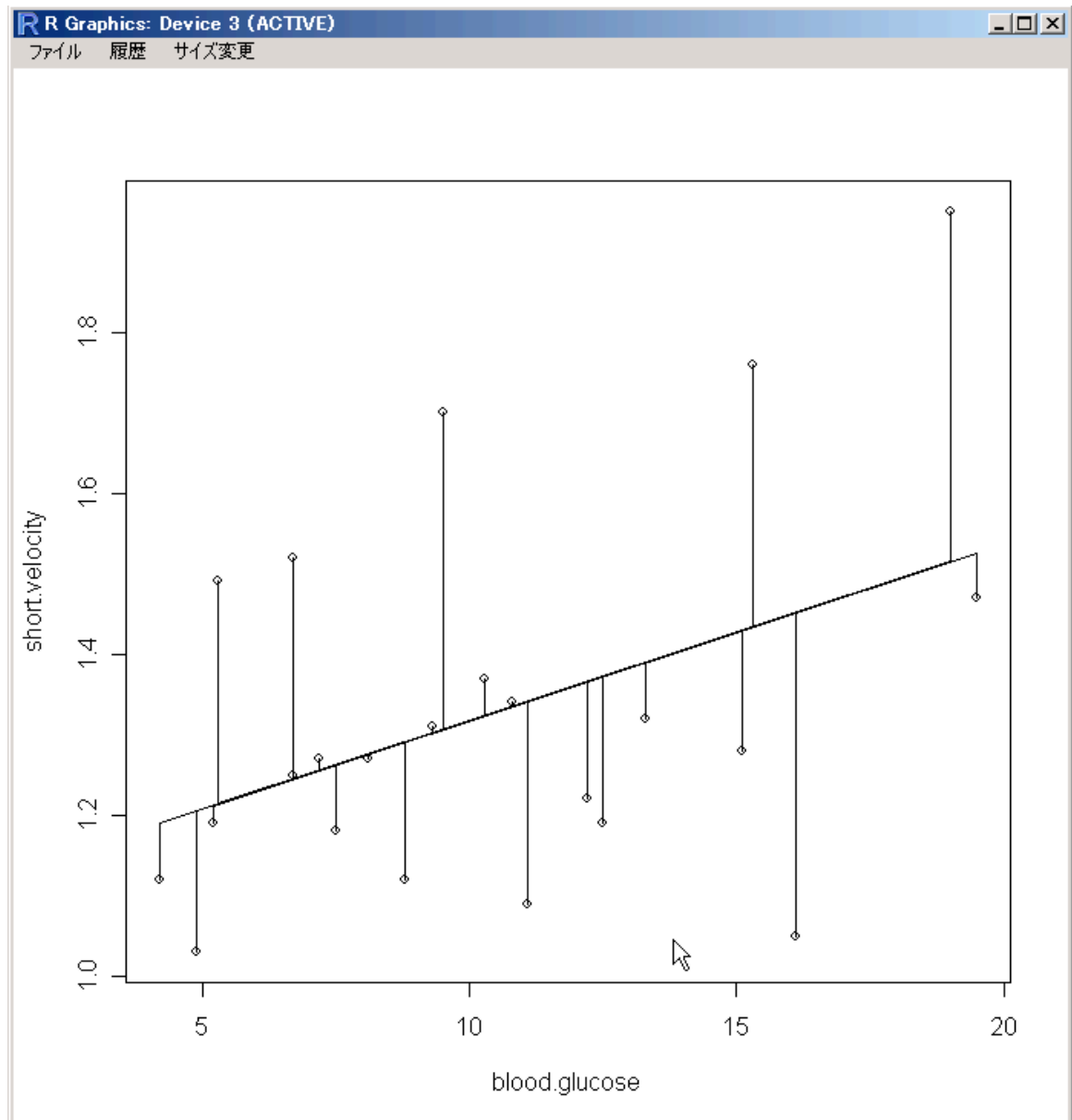


破線でかかれていますものが、

$$y = 1.10 + 0.0220x$$

の直線です。理想的に「よくあてはまっている」モデルの場合、すべてのデータはこの直線の上ののっているはずですが、そこまでよいモデルではないので、けっこう「ずれ」がありますね。

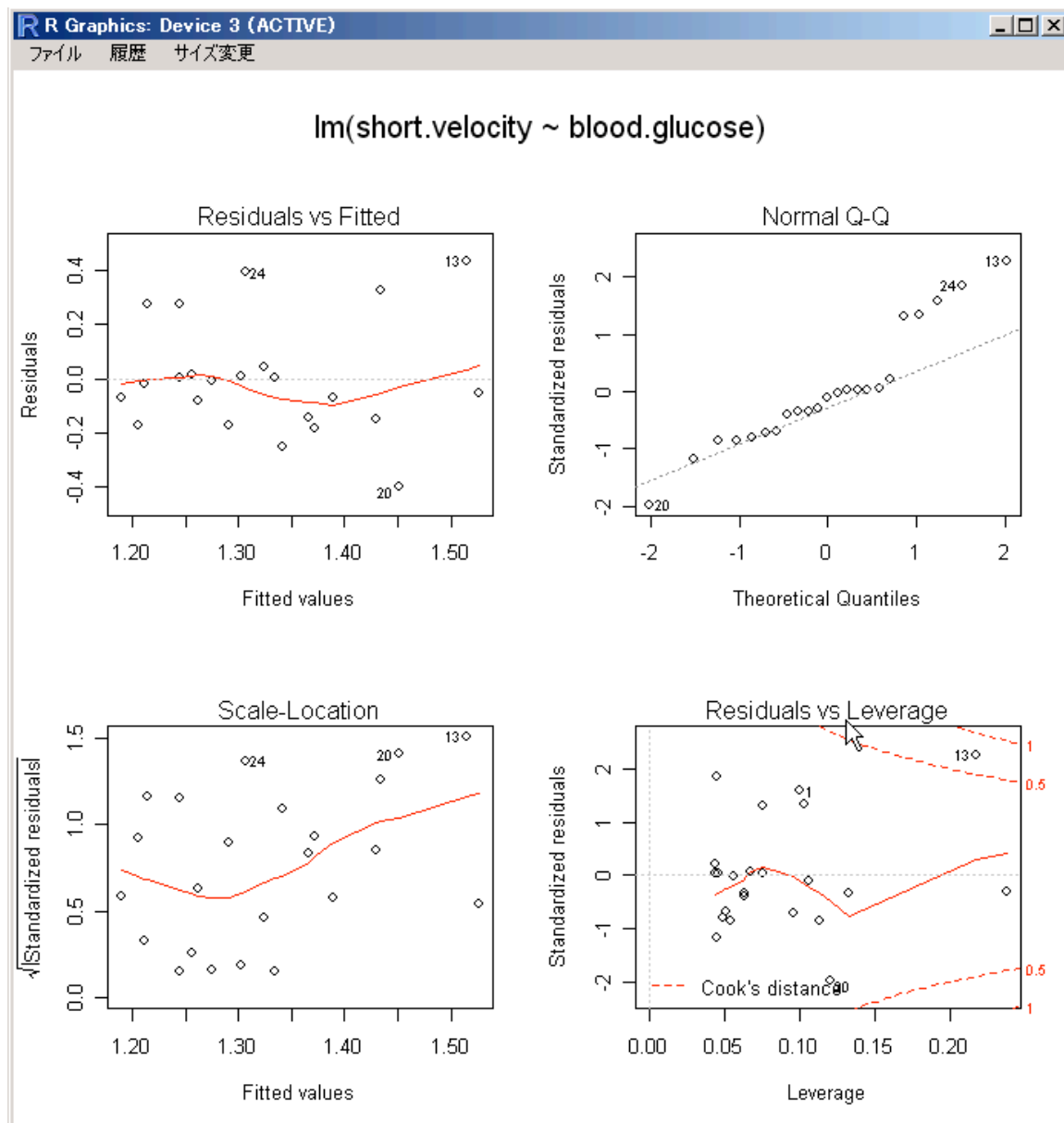
ずれ（残差）とは、以下のような、各データから回帰直線までの距離です。



これは、コマンドーからは作成できない図ですが、以下のように R Console のほうに打ち込むと描画できます。

```
> attach(thuesen)
> plot(blood.glucose,short.velocity)
> lines(blood.glucose,fitted(Model1))
> segments(blood.glucose,fitted(Model1), blood.glucose,short.velocity)
> detach()
> ■
>
```

この「ずれ」がどの程度なのかを調べ、モデルの適合度を評価することが「回帰診断」と呼ばれる作業です。まずはグラフから診断してみましょう。「モデル」メニューの「グラフ」→「基本的診断プロット」を選択してください。



このように4つのグラフが同時にプロットされます。左上のx軸は回帰直線のy軸の値（すなわち、予測された short.velocity の値）で、y軸は残差です。残差は予測された short.velocity の値とは独立に、平均ゼロの正規分布をとっているべきなので、このプロットはx軸の位置にかかわらず真ん中の水平線のまわりに一様に散らばっているべきであり、また水平線に近い点のほうが遠い点よりも多くあるべきです。今回のモデルは、理想的とまではいえませんが、ほどほど良い形をしているように見えます。右上の図は、標準化残差の正規性プロット(Q-Qプロット)と呼ばれるもので、「データが正規分布していれば、直線に近いプロットになる」というものです。これが（右上の方に行くにしたがってあやしくなってきますが）おおむね直線であることから、残差がほぼ正規分布しており、まあまあのモデルであることがわかります。

左下の図と右下の図の解釈については、かなり高度になりますので、参考書等を参照してください。

モデル適合度の数値による評価は、基本的なところはモデルを作成したときに出力ウィンドウに表示された要約表示に示されています。「モデル」→「モデルを要約」を選んでください。

```
出カウィンドウ
Call:
lm(formula = short.velocity ~ blood.glucose, data = thuesen)

Residuals:
    Min       1Q   Median       3Q      Max
-0.40141 -0.14760 -0.02202  0.03001  0.43490

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.09781    0.11748   9.345 6.26e-09 ***
blood.glucose  0.02196    0.01045   2.101  0.0479 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2167 on 21 degrees of freedom
(1 observation deleted due to missingness)
Multiple R-Squared:  0.1737,    Adjusted R-squared:  0.1343
F-statistic: 4.414 on 1 and 21 DF,  p-value: 0.0479
```

前回の説明した通り、Residuals のところに残差の記述統計量が表示されており、Residuals Standard Error に残差の標準誤差、ばらつきがでています（このばらつきの値は short.velocity の測定単位に依存します）。Multiple R-squared のところに R^2 、すなわちモデルによって、データの変動をどの程度説明できるかの割合がでてきます（これは割合なので、測定単位には依存しません）。F-statistic は、回帰係数がゼロに等しい、という帰無仮説に対して、分散分析のところででてきたのと同じ F 検定を行った結果です。ここでは $p=0.0479$ ということで、回帰係数 = 0 とはいえない、つまり、short.velocity の変化は blood.glucose とある程度直線的な関係にある、という結果となっています。なお、これは Coefficients のところにある blood.glucose の係数の t 検定と同じことをやっているだけです。

3. 分散分析と多重比較

さて、次は一度、分散分析に戻ります。3群以上で、平均値に違いがあるかどうかを検定するのがもっとも基本的な一元配置分散分析ですが、red.cell.folate データを使ってためてみましょう。

もしここで、「アタッチされたパッケージからデータセットを読み込む...」に "ISwR" パッケージが見つからなかったら、R Console のほうの「パッケージの読み込み...」で読み込んでからもういちどためてください。パッケージは一度「インストール」してしまえば、いつでも「読み込む」ことはできるようになりますが、作業スペースを保存しても、「どのパッケージが読み込まれているか」という状態は保存されません。一度 R を終了してからのちほど作業を再開する場合には、パッケージはもういちど「読み込み」をしなければならないのです。ただ、一度アクティブデータセットにしたデータについては作業スペースにコピーが作られますから、さきほどの thuesen はそのまま使うことができたのです。



	folate	ventilation
1	243	N2O+O2, 24h
2	251	N2O+O2, 24h
3	275	N2O+O2, 24h
4	291	N2O+O2, 24h
5	347	N2O+O2, 24h
6	354	N2O+O2, 24h
7	380	N2O+O2, 24h
8	392	N2O+O2, 24h
9	206	N2O+O2, op
10	210	N2O+O2, op
11	226	N2O+O2, op
12	249	N2O+O2, op
13	255	N2O+O2, op
14	273	N2O+O2, op
15	285	N2O+O2, op
16	295	N2O+O2, op
17	309	N2O+O2, op
18	241	O2, 24h
19	258	O2, 24h
20	270	O2, 24h
21	293	O2, 24h
22	328	O2, 24h

この red.cell.folate データは、心臓バイパス手術で麻酔をうけるときに笑気ガス(N₂O)と酸素(O₂)を、どれくらいの期間（手術中のみまたは連続 24 時間）与えたか(ventilation)によって、患者 22 名を 3 つの群にわけて、赤血球中の葉酸濃度(folate)がどうなったかを測定したものです。この 3 群で葉酸濃度に変化がなかった（帰無仮説）のか、あった（対立仮説）のかが興味があるところですね。

分散分析は、説明変数が blood.glucose のような量的変数ではなく、ventilation のようなカテゴリ変数である場合の回帰分析としてとらえることができます。まずは一元配置分散分析を行い、分散分析表を出力させてみましょう。「統計量」→「平均」→「一元配置分散分析...」です。

気になるチェックボックスがあると思いますが、いまはチェックせずに、グループに ventilation, 目的変数に folate を選んで Ok を押せば、出力ウィンドウに分散分析表がでてい

ずです。

```

出カウインドウ 実行
Analysis of Variance Table

Response: folate
      Df Sum Sq Mean Sq F value Pr(>F)
ventilation  2  15516    7758  3.7113 0.04359 *
Residuals  19  39716    2090
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> tapply(red.cell.folate$folate, red.cell.folate$ventilation, mean, na.rm=TRUE) # me:
N2O+O2,24h N2O+O2,op  O2,24h
  316.6250  256.4444  278.0000

> tapply(red.cell.folate$folate, red.cell.folate$ventilation, sd, na.rm=TRUE) # std.
N2O+O2,24h N2O+O2,op  O2,24h
  58.71709  37.12180  33.75648

> tapply(red.cell.folate$folate, red.cell.folate$ventilation, function(x) sum(!is.na
N2O+O2,24h N2O+O2,op  O2,24h
      8      9      5

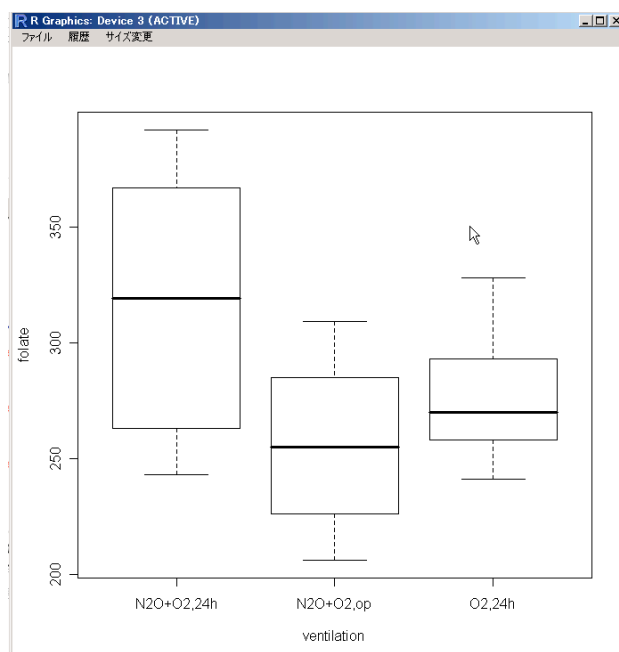
> remove(.Anova)
  
```

一番上に Analysis of Variance Table(分散分析表), その次にあるのが各群の葉酸値の平均、その次にあるのが各群の葉酸値の標準偏差、最後にあるのが各群の例数です。

分散分析表の中では、ventilation の行にある Sum Sq が群間平方和、Mean Sq が群間平均平方、F が F 統計量、Pr(>F) が F 検定の p 値です。Residuals の行にある Sum Sq が群内平方和、Mean Sq が群内平均平方ですね。

F 検定の帰無仮説は、どの群も同じ平均値と分散をもっているということ、つまり、群間分散と群内分散が同じということですが、 $p=0.04359$ というので、これは棄却されますから、群間で平均値と分散には違いがある、ということがいえます。

では、どの群がもっとも葉酸値が高いのか? ということは、実際に標本平均をみれば明らかです。N2O+O2, 24h が 316.625 というので最も高いといえます。グラフにかいてみてもいいですね。このグラフを書く方法はもう説明しなくても大丈夫ですね?





さて、これで換気方法と葉酸値には関係がある、ということはいえますが、たとえばもっとも葉酸値が高い N2O+O2, 24h と、2 番目の O2, 24h の間には有意な差があるか？ということはいえません。2 群になったんだから、t 検定をすればいいじゃないか！と思いますが、それは可能ではあるんですが、そうすると False Positive を犯す確率、 α がぐんぐん上がってしまう、という問題が発生します。これは数学的な制限ではなく、この検定で Positive, その結果をうけて次の検定でも Positive, それでさらにもういっこやって Positive, と、三段論法のような形式で「複数の検定を繰り返した結果を用いて1つの結論を導く」という論理展開自体に潜む罠です。一般に、医学研究では False Negative が出るよりも、実際にはない差を検出してしまふ False Positive のほうを重大な誤りとみなすので、 β が普通 0.20 ぐらいなのに対して α は 0.05 とキビシイのですが、じゃ、3 回検定をして False Positive をまったくおこさない確率はどうよ？というとな、 $(1-0.05) \times (1-0.05) \times (1-0.05) = 0.857$ 、つまり、3 回検定して結論を得た場合の False Positive の確率 $= 1-0.857=0.143$ つまり 14% に上がっていますね。このため、まず N2O+O2, 24h と O2, 24h の比較をして (1 回目の検定)、さらに N2O+O2, op と O2, 24h の比較をして (2 回目の検定)、最後に N2O+O2, op と N2O+O2, 24h の比較をして (3 回目) とやってしまうと、False Positive を起こす確率が 14% もあるということになります。そこで、多群を 1 つずつ比較する場合は、ちょっとそのへんに配慮した方法 (多重比較、ANOVA Post-hoc) を使おう、ということになってきたのです。

前置きが長くなりましたが、Post-hoc tests です。まずは、コマンドーの機能を使って Tukey 法を用いた場合の、各群の平均値の差の信頼区間を図示してみましょう。分散分析を行った際に気になっていたチェックボックス、「2 組ずつの平均の比較 (多重比較)」をチェックしてみましょう。



出力ウィンドウには以下の内容が追加されているはずです。

```
> confint(.Pairs)
      Simultaneous Confidence Intervals for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

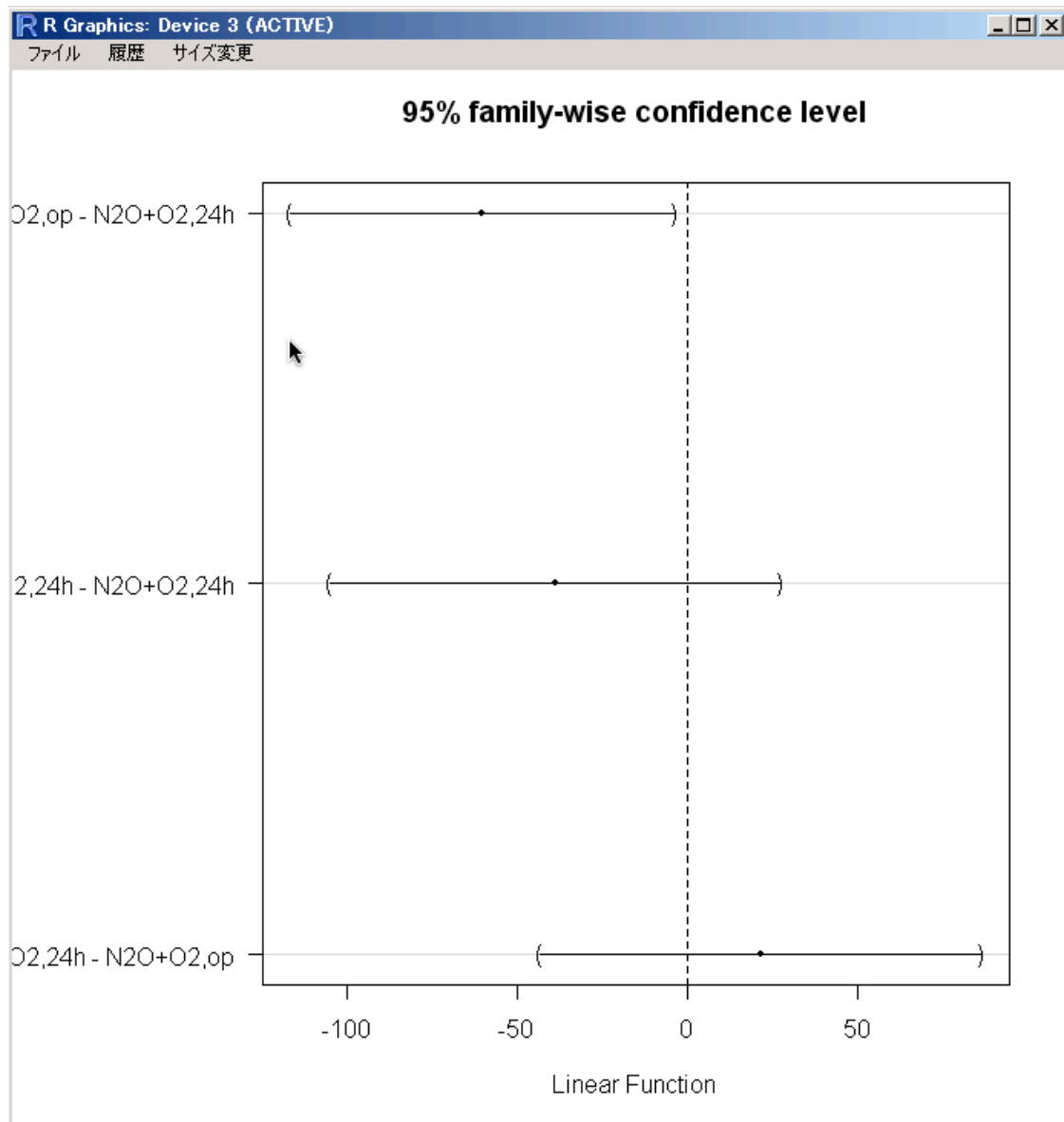
Fit: lm(formula = folate ~ ventilation, data = red.cell.folate)

Estimated Quantile = 2.5379

Linear Hypotheses:
              Estimate   lwr      upr
N2O+O2,op - N2O+O2,24h == 0 -60.1806 -116.5623  -3.7988
O2,24h - N2O+O2,24h == 0  -38.6250 -104.7738  27.5238
O2,24h - N2O+O2,op == 0   21.5556  -43.1644  86.2755

95% family-wise confidence level
```

これは、Tukey 法を用いて多重比較を行った際、各群間の平均値の差の信頼区間を示しています (Estimate が標本平均の差, lwr が 95% 信頼区間の下限, upr が上限)。この 95% 信頼区間にゼロが含まれていれば、その 2 群間に有意な差はないことになります。



同時に出るグラフのほうでも、差の95%信頼区間が示されています。破線で表現されているゼロが各群間のバーと交差していれば、群間の差があるとはいえません。

Bonferroni法を使う場合もやっておきましょう。これはR Consoleのほうからになります。

```
> attach(red.cell.folate)
> pairwise.t.test(folate,ventilation,p.adj="bonferroni")
```

Pairwise comparisons using t tests with pooled SD

data: folate and ventilation

	N2O+O2,24h	N2O+O2,op
N2O+O2,op	0.042	-
O2,24h	0.464	1.000

P value adjustment method: bonferroni

```
> detach()
> █
```



Bonferroni 法によって調整された p 値をみると、やはり N2O+O2, 24h と N2O+O2,op の間だけには有意な差が認められ、そのほかの群間には有意な差はなさそうです。

4. 重回帰分析

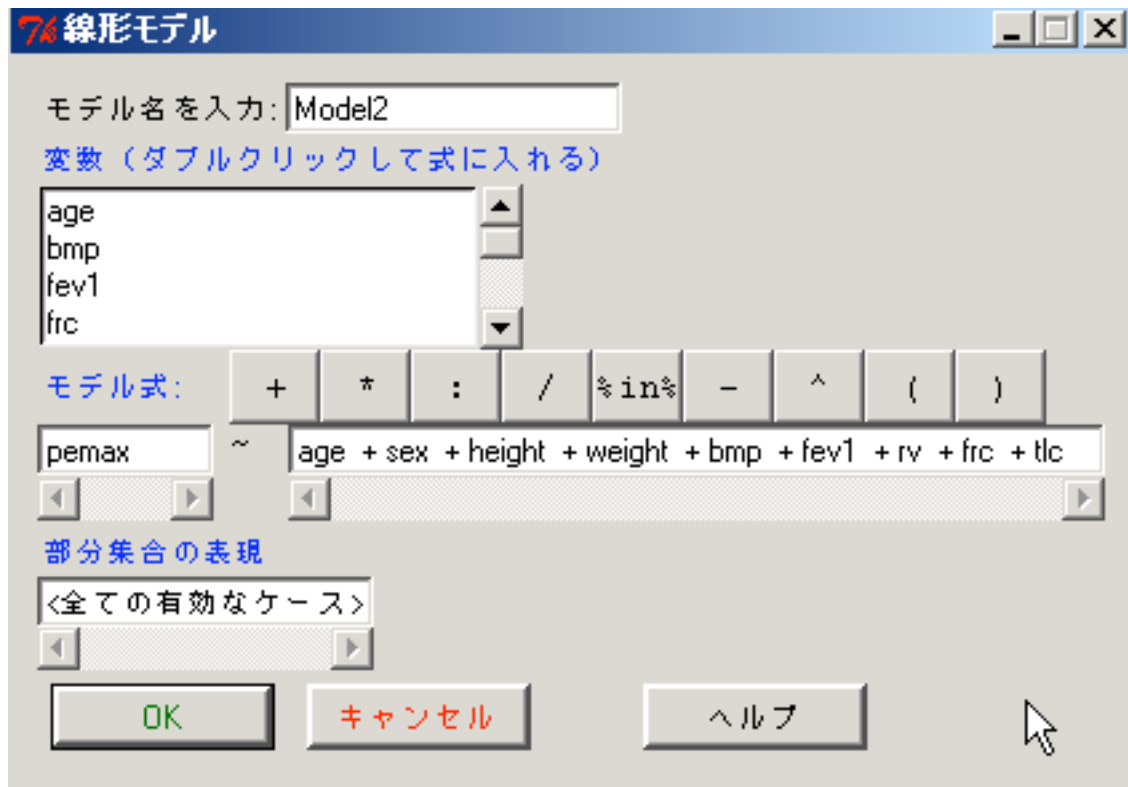
続いて重回帰分析に進みます。これは、基本的には回帰分析と同じ概念ですが、x 変数の数が増えただけです。しかしながら、x 変数が増えてくると、どの説明変数を用いて目的変数を説明するのがいいのか (**変数選択**)、モデルの選択といった問題が出てきますので、今回の講座の範囲を超えます。ここでは、モデルを作成するところまでを説明しますが、実際の解析にあたっては、変数をさまざまに増減させたり、入れかえたり、交互作用項をいれたりしながら、最もよくあてはまるモデルは何か、ということを考えていく必要がありますから、参考書等を参照してよく考えて最適なモデルを導くようにしてください。

データは "cystfibr" を用います。

	age	sex	height	weight	bmp	fev1	rv	frc	tlc	pemax
1	7	0	109	13.1	68	32	258	183	137	95
2	7	1	112	12.9	65	19	449	245	134	85
3	8	0	124	14.1	64	22	441	268	147	100
4	8	1	125	16.2	67	41	234	146	124	85
5	8	0	127	21.5	93	52	202	131	104	95
6	9	0	130	17.5	68	44	308	155	118	80
7	11	1	139	30.7	89	28	305	179	119	65
8	12	1	150	28.4	69	18	369	198	103	110
9	12	0	146	25.1	67	24	312	194	128	70
10	13	1	155	31.5	68	23	413	225	136	95
11	13	0	156	39.9	89	39	206	142	95	110
12	14	1	153	42.1	90	26	253	191	121	90
13	14	0	160	45.6	93	45	174	139	108	100
14	15	1	158	51.2	93	45	158	124	90	80
15	16	1	160	35.9	66	31	302	133	101	134
16	17	1	153	34.8	70	29	204	118	120	134
17	17	0	174	44.7	70	49	187	104	103	165
18	17	1	176	60.1	92	29	188	129	130	120
19	17	0	171	42.6	69	38	172	130	103	130
20	19	1	156	37.2	72	21	216	119	81	85
21	19	0	174	54.6	86	37	184	118	101	85
22	20	0	178	64.0	86	34	225	148	135	160
23	23	0	180	73.8	97	57	171	108	98	165
24	23	0	175	51.1	71	33	224	131	113	95
25	23	0	179	71.5	95	52	225	127	101	195

このデータは、嚢胞性肺線維症の患者 25 名について、肺機能の指標をいくつか測定したものです。全身の外分泌腺に障害を起こす病気なので、消化機能の障害から栄養失調を起こしますが、最大呼気圧をあらわす pemax が患者の栄養失調の指標になるので、これをほかの肺機能の指標を用いて説明するモデルをつくってみます。

「統計量」→「モデルへの適合」→「線形モデル」を選んでください。回帰分析のときには目的変数と説明変数が 1 つずつでしたが、今回は目的変数は 1 つですが、説明変数をたくさん使います。



今回の指示は、pemax をそのほかすべての変数で説明しようとするモデルを作ることを意味しています。完成したモデルの要約は出力ウィンドウに表示されます。

```

> summary(Model2)

Call:
lm(formula = pemax ~ age + sex + height + weight + bmp + fev1 +
    rv + frc + tlc, data = cystfibr)

Residuals:
    Min       1Q   Median       3Q      Max
-37.338 -11.532   1.081  13.386  33.405

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  176.0582   225.8912   0.779   0.448
age          -2.5420    4.8017  -0.529   0.604
sex          -3.7368   15.4598  -0.242   0.812
height      -0.4463    0.9034  -0.494   0.628
weight       2.9928    2.0080   1.490   0.157
bmp         -1.7449    1.1552  -1.510   0.152
fev1         1.0807    1.0809   1.000   0.333
rv           0.1970    0.1962   1.004   0.331
frc         -0.3084    0.4924  -0.626   0.540
tlc          0.1886    0.4997   0.377   0.711

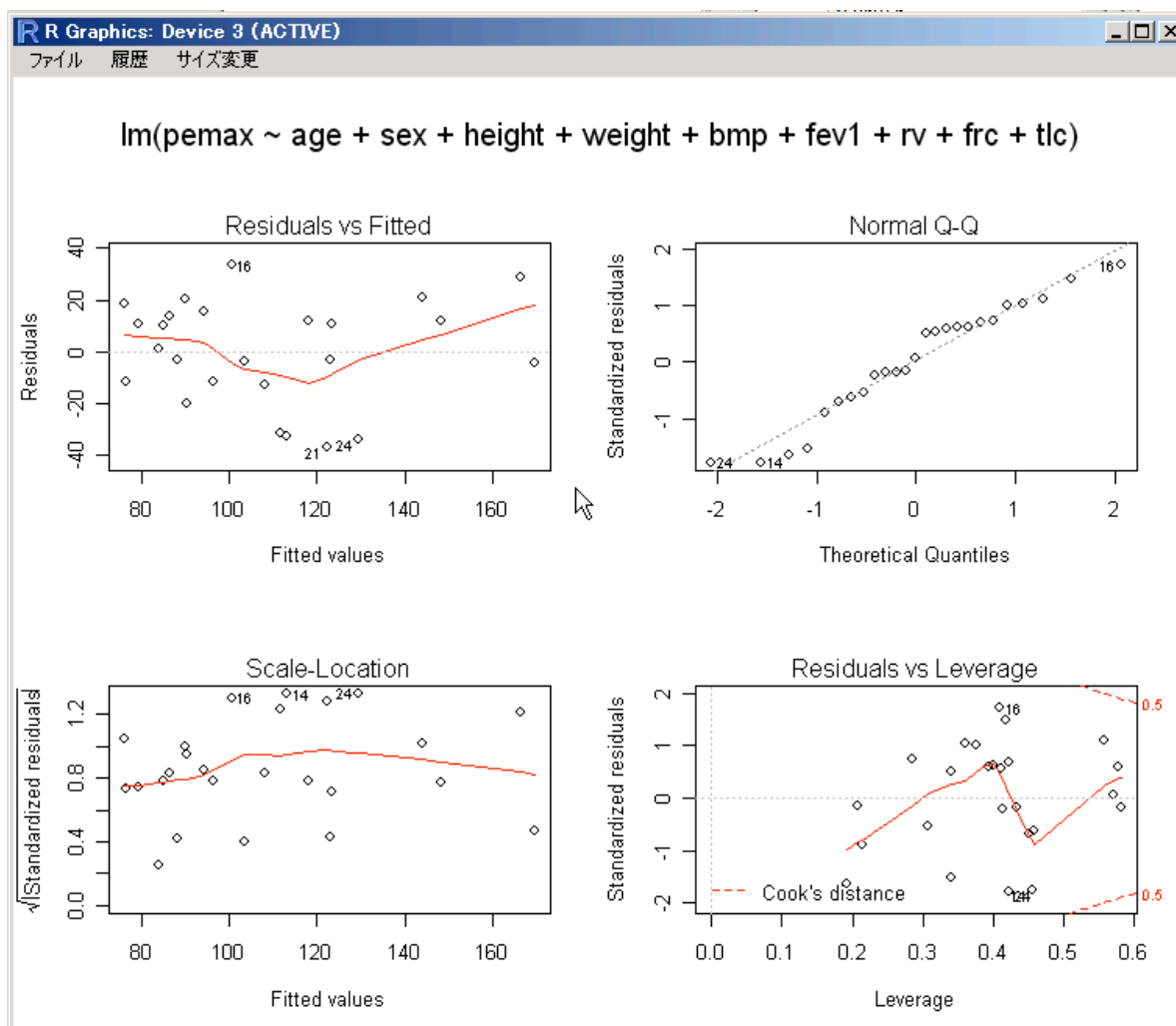
Residual standard error: 25.47 on 15 degrees of freedom
Multiple R-Squared:  0.6373,    Adjusted R-squared:  0.4197
F-statistic: 2.929 on 9 and 15 DF,  p-value: 0.03195
    
```



基本的には、この出力結果は回帰分析の場合と同じですね。Coefficients のところの変数が増えているだけです。

Coefficients の t 検定の結果は、0.05 未満となる p 値は 1 つもありませんでした。しかし、このモデルが無意味ということではありません。最後の F 検定のところを見てみると、 $p=0.03195$ と有意になっています。これはつまり、変数のうちどれかには何らかの効果があることを意味しています。各回帰係数に対する t 検定が有意でない、ということは、いいかえればその 1 つ 1 つの変数 だけ について、回帰係数が 0 であるとしてもおかしくない、ということです。しかし、1 つを取り除いたときにもまだ 8 個の変数が残るわけで、(F 検定が有意ですから) そのうちどれかが pemax を説明している、という状況は変わりません。t 検定で有意な変数が 1 つもない、ということは、ただ、モデルから絶対に取り除いてはいけない変数というのは、ない、ということではありません。一方、 R^2 については 0.637 ということで、このモデル全体としては pemax の変化をかなり説明できていると考えられます。

基本的診断プロットも出しておきましょう。



blood.glucose と short.velocity のときに比べて、いい感じのグラフになっていますね。

5. 生存時間解析

さて、次は生存時間解析です。これは医学領域に特化して使われる手法ですので、R コマンドーには含まれていません。多少面倒ですが、R Console からの直接作業になります。コマンドを書いてある通りに打ち込めば Ok ですが、もし変な出力が出てしまうようでしたら、打ち込み間違いが



ないかどうか、よく見直して下さい。カンマとピリオド、シングルクォートとダブルクォート、括弧の閉じ忘れなどがあることが多いです。

以後、R Console での作業になります。

生存時間解析は、パッケージ survival の機能として提供されています。まずは survival パッケージを読み込んでおいてください。

データとしては melanom を使います。コマンドーでは解析できませんが、読み込むことはできるので、アクティブデータセットに指定して表示させてみましょう。

	no	status	days	ulc	thick	sex
1	789	3	10	1	676	2
2	13	3	30	2	65	2
3	97	2	35	2	134	2
4	16	3	99	2	290	1
5	21	1	185	1	1208	2
6	469	1	204	1	484	2
7	685	1	210	1	516	2
8	7	1	232	1	1288	2
9	932	3	232	1	322	1
10	944	1	279	1	741	1
11	558	1	295	1	419	1
12	612	3	355	1	16	1
13	2	1	386	1	387	1
14	233	1	426	1	484	2
15	418	1	469	1	242	1
16	765	3	493	1	1256	2
17	777	1	529	1	580	2
18	61	1	621	1	706	2
19	67	1	629	1	548	2
20	819	1	659	1	773	2
21	10	1	667	1	1385	1
22	15	1	718	1	234	2
23	47	1	752	1	419	2
24	9	1	779	1	404	2
25	907	1	793	1	484	2
26	758	1	817	2	32	1
27	8	3	826	1	854	1
28	400	1	833	1	258	1
29	232	1	858	2	356	1
30	18	1	869	2	354	1

これは悪性黒色腫の手術後の患者の生存状況を示しています。no は患者の ID、status は観察終了時点での生存状況(1: 悪性黒色腫により死亡 2: 生存 3: ほかの原因により死亡)、days は観察期間 (日数)、ulc は黒色腫の潰瘍形成があったかどうか(1:あり, 2:なし)、thick は黒色腫の厚さ (1/100mm 単位)、sex は性別(1: 女性, 2: 男性)

解析は R console のほうです。以下のように入力すると、生存時間と悪性黒色腫による死亡の確率の間の関係を示すモデルが作成されます。



```

>
> data(melanom)
> attach(melanom)
> surv.all <- survfit(Surv(days,status==1))
> summary(surv.all)
Call: survfit(formula = Surv(days, status == 1))

   time  n.risk  n.event  survival  std.err  lower 95% CI  upper 95% CI
   185     201      1      0.995  0.00496    0.985    1.000
   204     200      1      0.990  0.00700    0.976    1.000
   210     199      1      0.985  0.00855    0.968    1.000
   232     198      1      0.980  0.00985    0.961    1.000
   279     196      1      0.975  0.01100    0.954    0.997
   295     195      1      0.970  0.01202    0.947    0.994
   386     193      1      0.965  0.01297    0.940    0.991
   426     192      1      0.960  0.01384    0.933    0.988

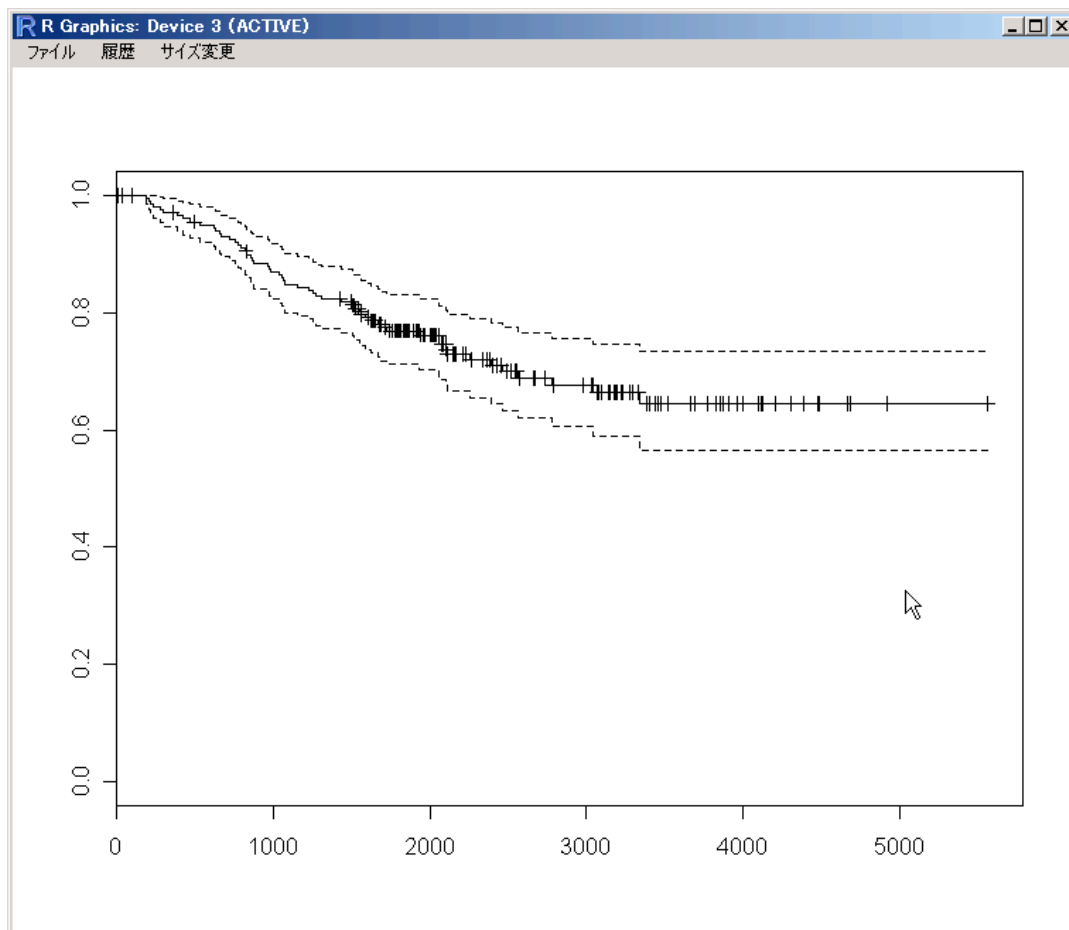
```

最後の summary(surv.all)を入力したところで、たくさんの出力結果が出たと思います。これが Kaplan-Meier 推定量です。これをプロットすれば Kaplan-Meier 曲線を描くことができます。

```

   2407      69      1      0.700  0.03019    0.633    0.773
   2565      63      1      0.689  0.03729    0.620    0.766
   2782      57      1      0.677  0.03854    0.605    0.757
   3042      52      1      0.664  0.03994    0.590    0.747
   3338      35      1      0.645  0.04307    0.566    0.735
>
>
> plot(surv.all)
>

```

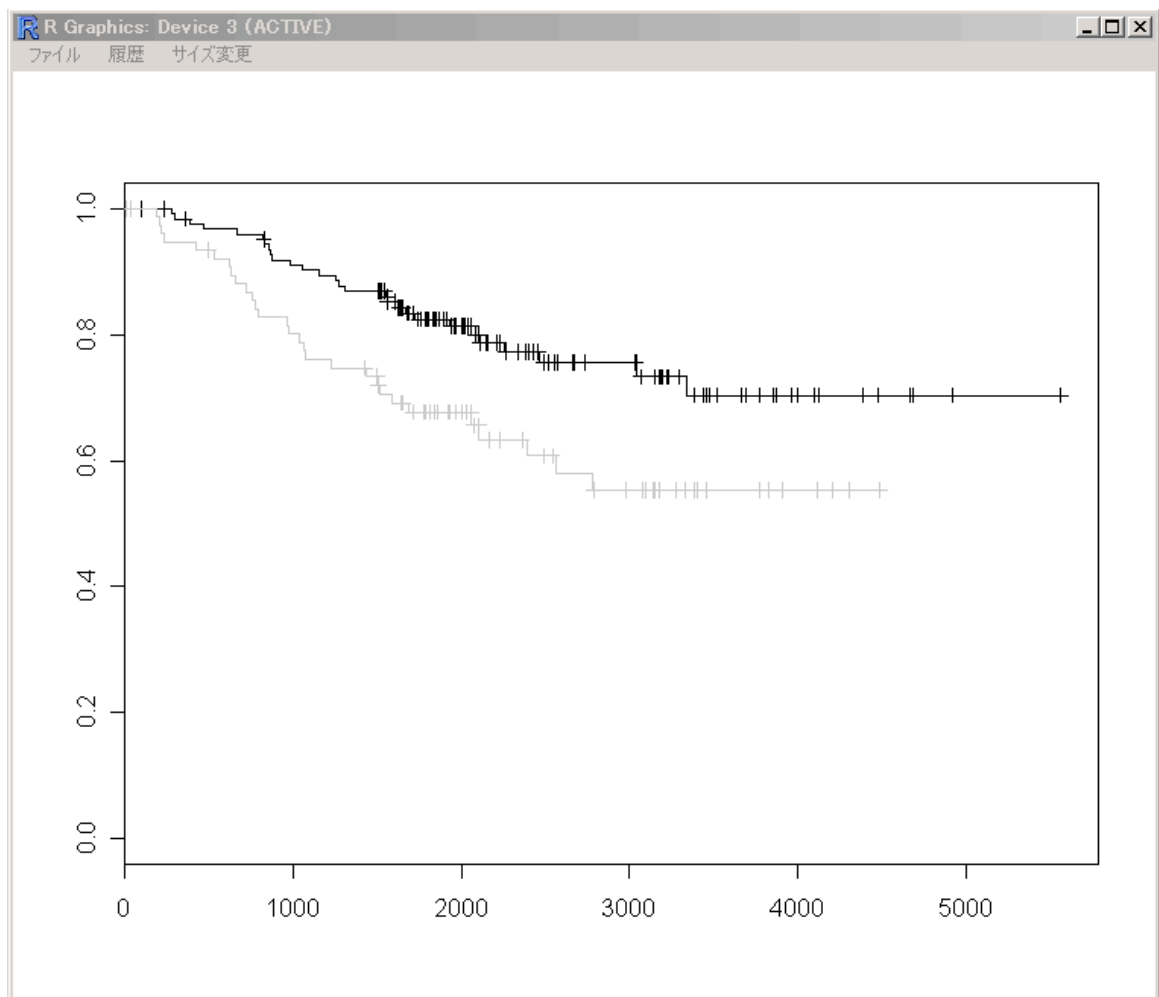


曲線の上下に出ている破線は、曲線の 95%信頼区間を示しています。

性別のデータもありますから、性別によって生存曲線に違いがあるかどうかを調べてみましょう。生存時間と悪性黒色腫による死亡の確率の間の関係を示すモデルを作成するときに、「性別で説明する」という要素を加えることができます。

```
>  
> surv.bysex <- survfit(Surv(days,status==1) ~ sex)  
> plot(surv.bysex, col=c("black","gray"))  
> █
```

status==1 が、「悪性黒色腫による死亡」を「イベント」とする、ということを示しています。col=c("black","gray")で、性別 =1(女性)を黒で、性別 =2(男性)をグレーで表示します。





この生存曲線に有意な差があるかどうかはログランク検定でしたね。

```
>
> survdiff(Surv(days,status==1) ~ sex)
Call:
survdiff(formula = Surv(days, status == 1) ~ sex)

      N Observed Expected (O-E)^2/E (O-E)^2/V
sex=1 126      28     37.1      2.25      6.47
sex=2  79      29     19.9      4.21      6.47

Chisq= 6.5  on 1 degrees of freedom, p= 0.011
> █
```

p 値が 0.011 となり、「性別ごとに作成された 2 つの生存関数は同一である」という帰無仮説は棄却されました。

6. まとめ(演習問題)

- 1) ISwR パッケージのデータセット vitcap に関して、2 つのグループ間で肺活量に差があるかどうかを t 検定を用いて分析してみましょう。
- 2) ISwR パッケージのデータセット rmr に関して、代謝率と体重との関係をグラフに示してください。また、この関係を回帰分析してみましょう。そのモデルに基づくと、体重が 70kg の場合、代謝率はどのくらいになるでしょうか？
- 3) データセット juul のうち 25 歳を超える被験者のグループにおいて、IGF-I の値の平方根と年齢の関係を回帰分析してみましょう。
- 4) 2 種類の消化性潰瘍治療薬の比較を行ったところ、下表のような結果が得られました。この 2 つの薬の効果には差がみられたでしょうか？

薬剤名	治癒	非治癒	合計	
薬 A		23	7	30
薬 B		18	13	31
合計		41	20	61

- 1) ISwR パッケージの lung データセットにおいて、3 つの測定方法は異なる結果をもたらしているでしょうか？もしそうなら、有意に他と異なる結果をもたらしている方法はどれでしょうか？
- 2) ISwR パッケージの secher データセットでは、腹部直径および児頭大横径、出生児体重をすべて対数変換すると、変数間の関係をいいモデルにすることができます。出生児体重を腹部直径と児頭大横径で予測するモデルをつくってみましょう。このモデルと、腹部直径のみ、あるいは児頭大横径のみで予測を行うモデルとはどちらがよいモデルになっているでしょうか。

7. 参考文献

より進んだ解析手法を学ぶために:

R による医療統計学, 岡田昌史(翻訳), 丸善, 2007, ISBN 978-4621078112

より進んだ R Commander の使い方を学ぶために:

R Commander ハンドブック, 舟尾暢男(著), 九天社, 2007, ISBN 978-4861671913