

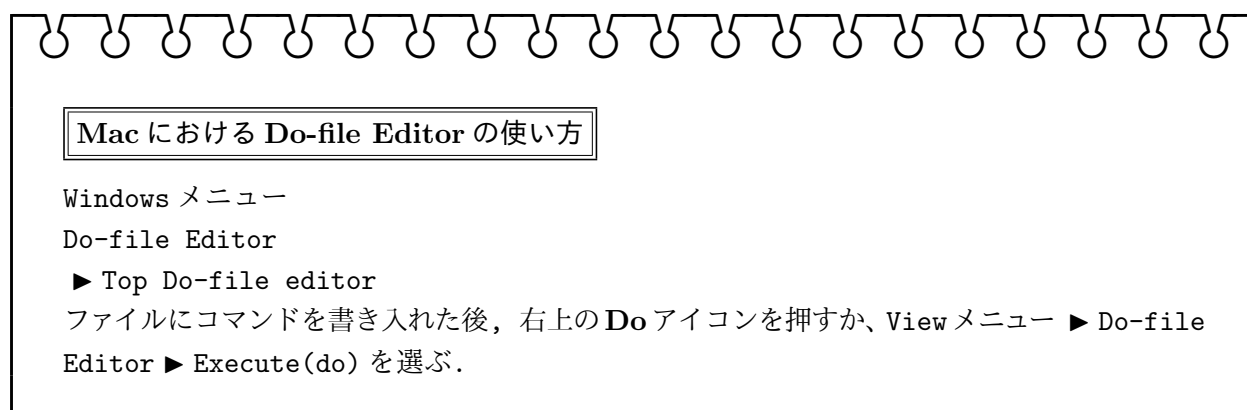
第1章 補論

2015年4月6日

1. ECCS リモートシステムの使い方.

ECCS のウェブサイト に接続し、ページ左下のリモートアクセス環境から接続をする.
(<https://ras.ecc.u-tokyo.ac.jp/guacamole/>)

2. Mac における Do-file Editor の使い方.



3. Mac における画面サイズの変更

画面左上の メニュー ▶ System Preferences...

出てきた画面で Display を選択してクリックし、さらに Resolution の Scaled をクリックして選んで 800x600 などを選ぶ.

2015年4月9日

1. データの入力のしかた (csv ファイル, エクセル・ファイル)

csv ファイルの読み込み

File メニュー

Import

▶ Tex data (delimited, *.csv,...)

画面で Browse... をクリックし, csv ファイルを選択する.

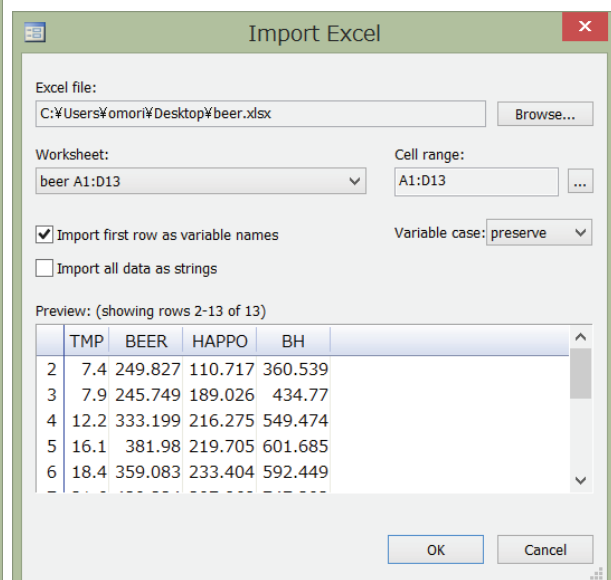
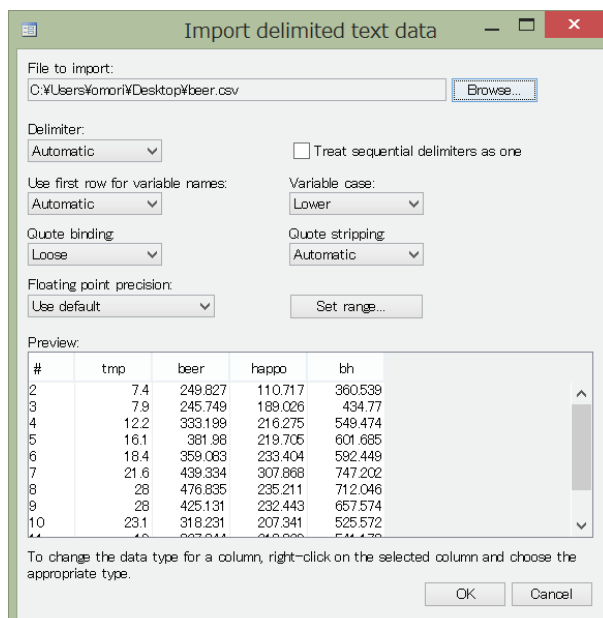
xls, xlsx ファイルの読み込み

File メニュー

Import

▶ Excel spreadsheet (*.xls, *.xlsx)

画面で Excel file: の Browse... をクリックし, xls(または xlsx) ファイルを選択する.
第1行目が変数の名前の場合には, Import first row as variable names に ✓(チェック)を入れる.



2. データの入力のしかた (Command ウィンドウを使って csv ファイルを読み込む)

(1) まず pwd と入力して (pwd:print working directory), 現在のディレクトリを表示する.

- (2) 現在のディレクトリを作業したいディレクトリに変更する． 具体的には `cd "作業ディレクトリ"` と入力するか (`cd:change directory`), File メニューから **Change Working Directory...** を選んで, ディレクトリを指定する．
- (3) `insheet using` ファイル名とすると, データが入力される．
- ※ 作業ディレクトリがわかっていれば, 直接 `insheet using "ファイルへのパス"` とすることもできる．

2015 年 4 月 20 日

1. 比率の区間推定． 比率の区間推定は以下のように行うこともできる．

比率の区間推定

Statistics メニュー

Summaries, tables, amd tests

▶ Summary and descriptive statistics

▶ Binomial calculator

画面で **Sample size** に観測値の個数 $n = 865$, **Successes** に試行の成功回数 $x = 268$ ^a を入れる．

^a $n\hat{p} = 865 \times 0.31 = 268.15$ より, $x = 268$ とわかる． デフォルトの計算方法は **Exact** だが, **Wald** を選べば正規分布に基づく近似的な信頼区間が出力される．

The screenshot shows a software window titled "cii - Confidence intervals and standard-...". It contains a "Binomial CI calculator" section with the following settings:

- Calculator type:** Binomial confidence interval calculator (selected)
- Binomial CI calculator:**
 - Sample size: 865
 - Successes: 268
 - Binomial confidence interval:**
 - Exact (selected)
 - Wald
 - Wilson
 - Agresti-Coull
 - Jeffreys
 - Confidence level: 95

At the bottom, there are buttons for "OK", "Cancel", and "Submit".

表 1.1: 比率の信頼区間

```
. display 865*0.31
268.15
. cii 865 268
```

-- Binomial Exact --					
Variable	Obs	Mean	Std. Err.	[95% Conf. Interval]	
-----+-----					
	865	.3098266	.0157228	.2791244	.3418306

```
. cii 865 268, wald (Waldを指定した場合)
```

-- Binomial Wald ---					
Variable	Obs	Mean	Std. Err.	[95% Conf. Interval]	
-----+-----					
	865	.3098266	.0157228	.2790104	.3406427

```
. prtesti 865 268 0.4, count (仮説検定を標本比率0.31でなく, 計数(count)268を使った場合)
```

One-sample test of proportion x: Number of obs = 865

-----+-----				
Variable	Mean	Std. Err.	[95% Conf. Interval]	
-----+-----				
x	.3098266	.0157228	.2790104	.3406427
-----+-----				

```

p = proportion(x)
Ho: p = 0.4
Ha: p < 0.4
Pr(Z < z) = 0.0000
Ha: p != 0.4
Pr(|Z| > |z|) = 0.0000
Ha: p > 0.4
Pr(Z > z) = 1.0000
z = -5.4135
```

※仮説検定で計数 (count) を使うには, 画面の下部にある Use integer counts instead of proportions に ✓ (チェック) をいれて, 標本比率の代わりに計数を入れればよい.

2. Video examples. ヘルプで現れる画面に, Video examples という項目がある場合には, クリックすると解説のビデオを Youtube で見ることができる (ただし, バージョンは 12).

3. 分布の裾の面積を求める

- t 分布の右裾の面積が 0.05 (5%) であるような点を求める. Command ウィンドウで display invttail(自由度, 0.05) と入れる. display invttail(自由度, 0.05) と入れる.
- t 分布において, t 値より右側の面積を求める. Command ウィンドウで display ttail(自由度, t 値) と入れる.

- 正規分布の左裾の面積が 0.95 (95%) であるような点を求める. Command ウィンドウで `display invnormal(0.95)` と入れる.
- STATA で利用可能な関数を見るには, Command ウィンドウで `help functions` と入れる. あるいは Data メニューで, Other utilities ► Hand calculator を選び, Create... を押すと, Category: に関数のリストがある.

4. その他

- Mac では, オプション・キー + ¥キー → \

2015 年 4 月 23 日

- 25 ページ. 脚注 4 を下記に修正.
対数リターン (連続複利の収益率) $(\log p_t - \log p_{t-1}) \times 100$ (たとえば, $(\log(\text{topix}) - \log(\text{L.topix})) \times 100$) を用いても良い.
- 25 ページ. 2 期前, 3 期前のラグは `L2.sony`, `L3.sony` などと表す.
- 28 ページ. 脚注 5. 図 1.24 は平均値の信頼区間で, $X_i = x_i$ が与えられた時の Y_i の期待値 $E(Y_i|X_i = x_i) = \beta_0 + \beta_1 x_i$ の 95% 信頼区間について, x_i の値を様々に変えて曲線をつなげたものとなっている. 一方, 脚注 5 にある **Confidence interval for an individual forecast** は各予測の信頼区間で, 新しく $X_i = x_i^*$ が与えられた時の $Y_i^* = \beta_0 + \beta_1 x_i^* + \epsilon_i^*$ の 95% 信頼区間であり, 前者と区別するために 95% 予測区間とも呼ばれる. 平均だけにフォーカスする前者と異なり, 誤差項 ϵ_i^* まで考慮するので区間の幅が広がる.

図 1.1: 散布図と推定された回帰直線, 95% 信頼区間 (左) と 95% 予測区間 (右) (1)

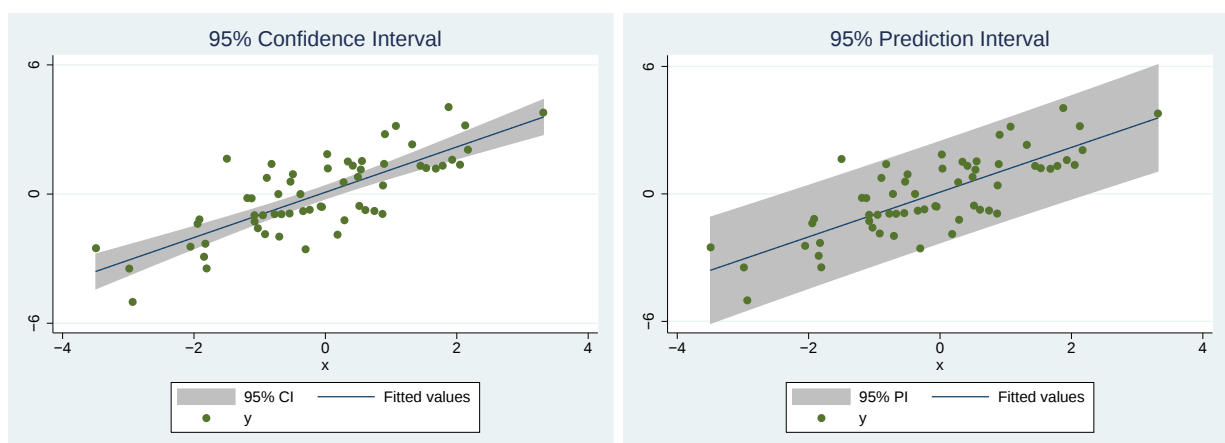
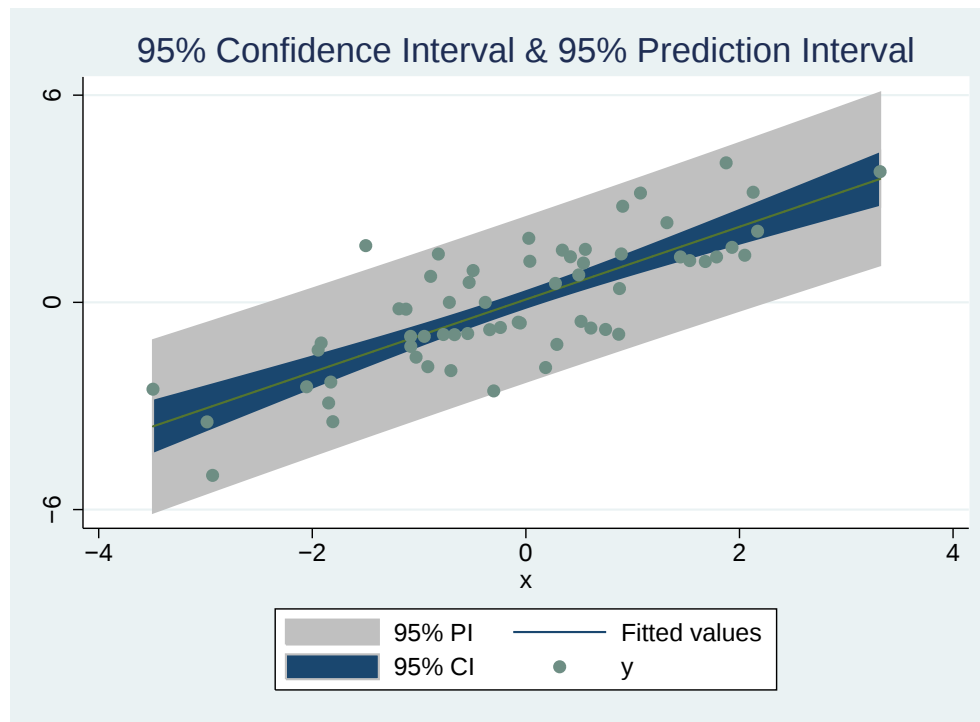


図 1.2: 散布図と推定された回帰直線, 95%信頼区間と 95%予測区間 (2)



STATA14 での変更点

1.1.

正規乱数を発生させる

データ (Data) メニュー

データの作成または変更 (Create or change data)

- ▶ その他の変数作成コマンド (Other variable creation commands)
- ▶ 正規分布から標本を抽出 (Draw sample from normal distribution)

図 1.1: 乱数を発生する

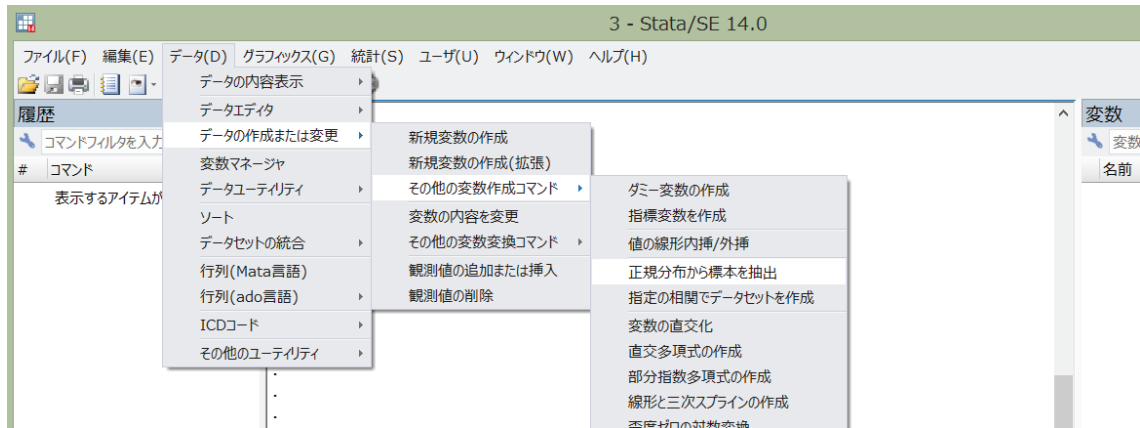
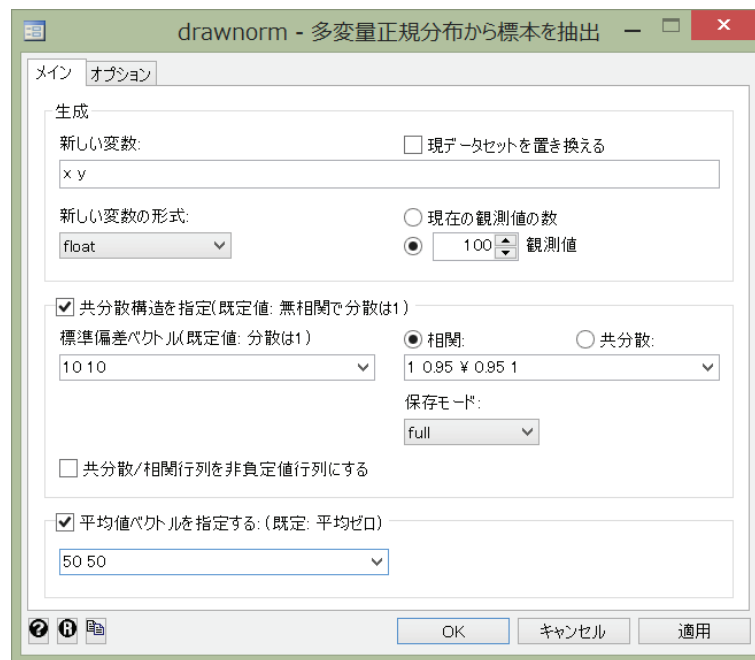


図 1.2: 乱数を発生する (2)



データを閲覧または編集する

データ (Data) メニュー

データエディタ (Data Editor)

▶ データエディタ (ブラウズ) (Data Editor (Browse))

(または ▶ データエディタ (編集) (Data Editor (Edit))

散布図を描く

グラフィックス (Graphics) メニュー

二元グラフ (散布図/折れ線など) (Twoway graph (scatter, line, etc))

図 1.3: 散布図 (1)

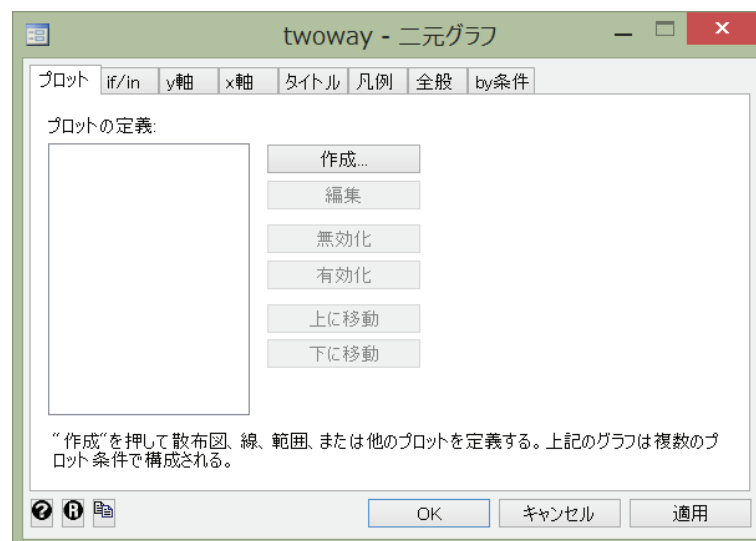
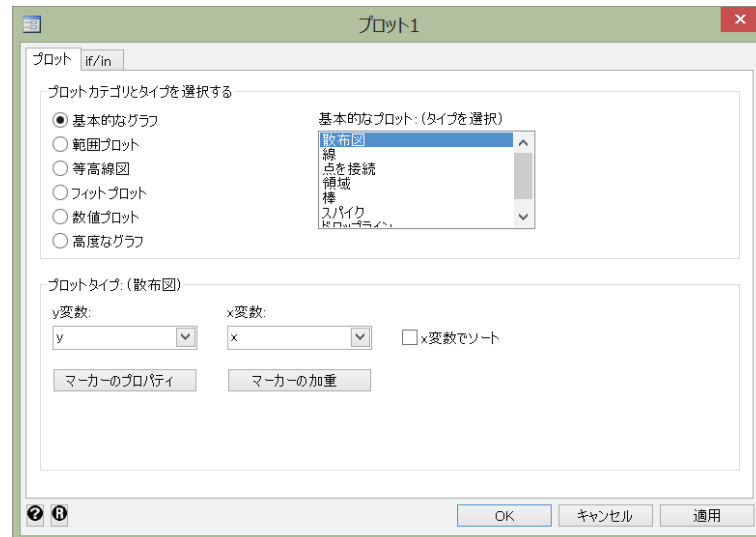


図 1.4: 散布図 (2)



基本統計量を求める

データ (Data) メニュー

データの内容表示 (Describe data)

▶ 記述統計量 (Summary Statistics)

または

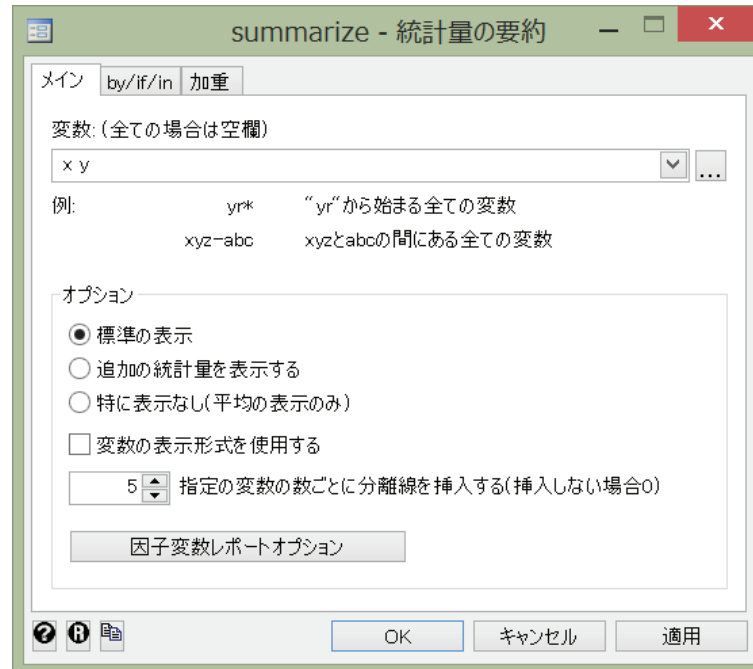
統計 (Statistics) メニュー

要約/表/検定 (Summaries, tables, and tests)

▶ 記述統計量 (Summary and descriptive statistics)

▶ 記述統計量 (Summary Statistics)

図 1.6: 基本統計量 (1)



分散, 共分散, 相関係数を求める

統計 (Statistics) メニュー

要約/表/検定 (Summaries, tables, and tests)

- ▶ 記述統計量 (Summary and descriptive statistics)
- ▶ 相関と共分散 (Correlations and covariances)

図 1.7: Command ウィンドウ

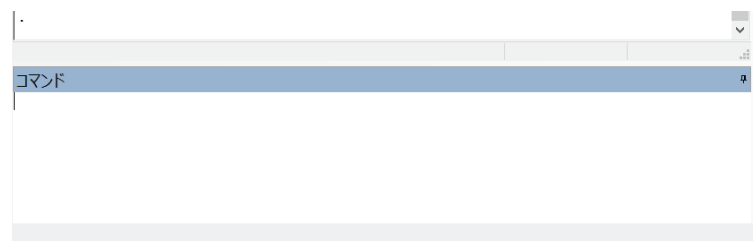


図 1.8: Do-file Editor の画面



1.2.

図 1.10: データを Excel から Data Editor にコピー (2)

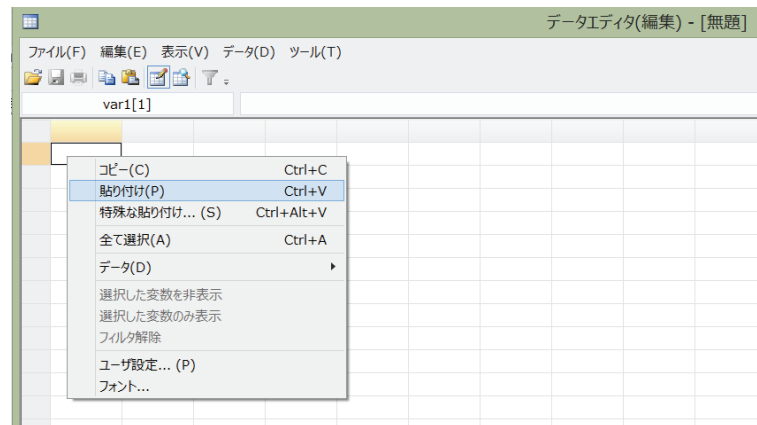


図 1.11: 詳細な要約統計量を出力する.

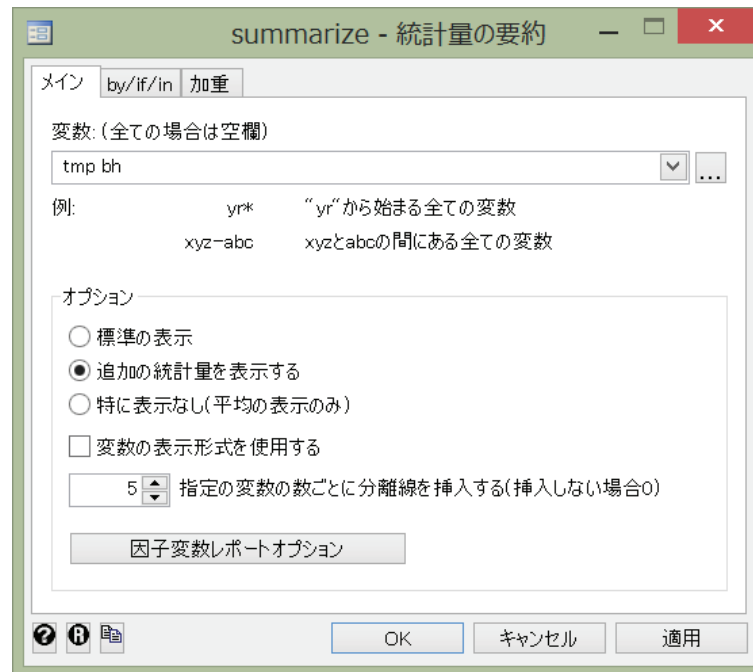
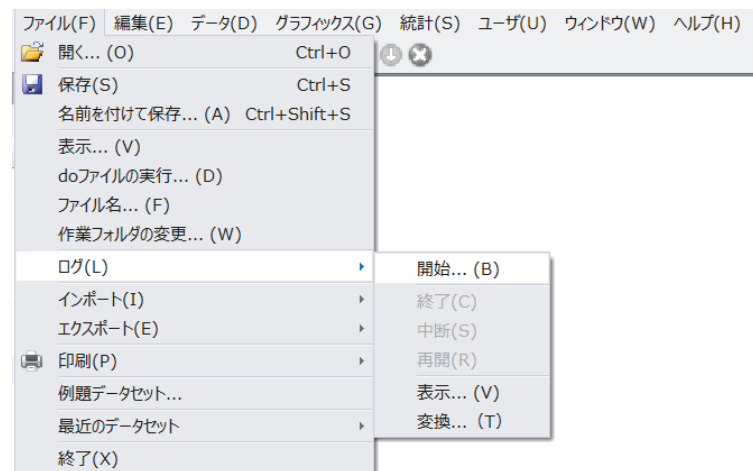


図 1.13: ログを保存する



csv ファイルの読み込み

ファイル (File) メニュー

インポート (Import)

▶ テキストデータ (デリミタ, .csv 等) (Tex data (delimited, *.csv, ...))

画面で参照 (Browse...) をクリックし, csv ファイルを選択する.

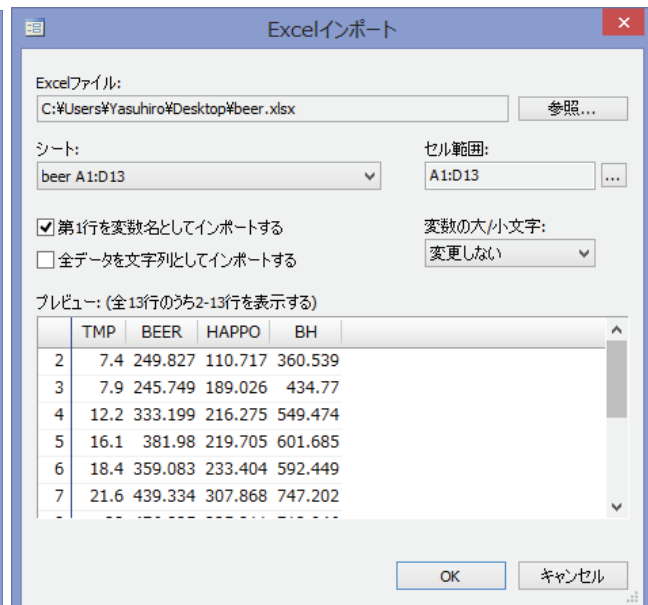
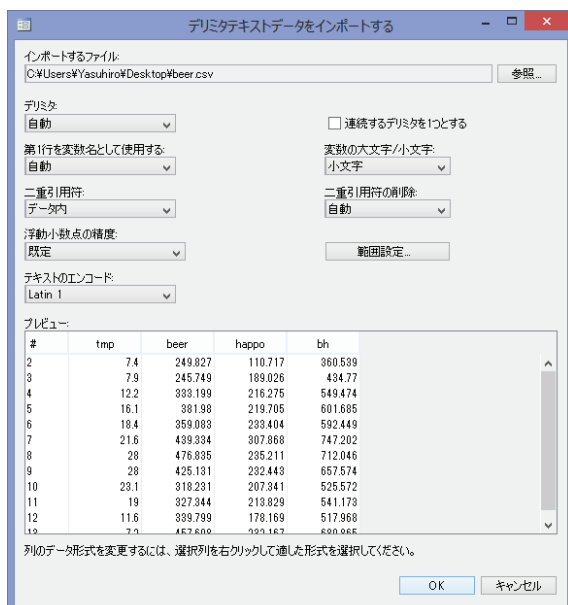
xls, xlsx ファイルの読み込み

ファイル (File) メニュー

インポート (Import)

▶ Excel シート形式 (.xls, xlsx) (Excel spreadsheet (*.xls, *.xlsx))

画面で Excel ファイル (Excel file:) の参照 (Browse...) をクリックし, xls(または xlsx) ファイルを選択する. 第 1 行目の変数の名前の場合には, 第 1 行を変数名としてインポートする (Import first row as variable names) に ✓ (チェック) を入れる.



1.3.

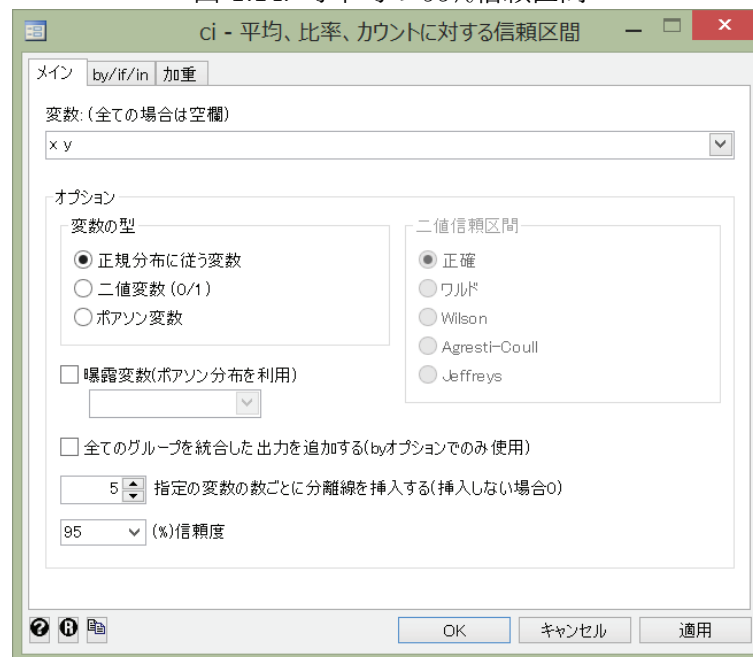
母平均の信頼区間を求める

統計 (Statistics) メニュー

要約/表/検定 (Summaries, tables, and tests)

- ▶ 記述統計量 (Summary and descriptive statistics)
- ▶ 信頼区間 (Confidence intervals)

図 1.14: 母平均の 95%信頼区間



母平均 (母平均の差) の仮説検定

統計 (Statistics) メニュー

要約/表/検定 (Summaries, tables, and tests)

- ▶ 古典的仮説検定 (Classical tests of hypotheses)
- ▶ t 検定 (平均比較検定) (t test (mean-comparison test))

図 1.15: 仮説検定:有意水準 5%

The dialog box is titled "t検定(平均比較検定)". It has a tabbed interface with "メイン" and "by/if/in" tabs. The "t検定" section has four radio buttons: "一群" (selected), "グループ変数による二群", "変数による二群", and "対応のあるデータ". The "一群の平均比較検定" section contains a "変数名:" dropdown menu with "x" selected, a "仮説平均:" text box with "50", and a "(%)信頼度" dropdown menu with "95" selected. At the bottom are buttons for "OK", "キャンセル", and "適用".

図 1.16: 母平均の差の仮説検定:有意水準 5%

The dialog box is titled "t検定(平均比較検定)". It has a tabbed interface with "メイン" and "by/if/in" tabs. The "t検定" section has four radio buttons: "一群", "グループ変数による二群", "変数による二群" (selected), and "対応のあるデータ". The "変数による二群のt検定" section contains a "第1変数:" dropdown menu with "x" selected, a "第2変数:" dropdown menu with "y" selected, a checkbox for "不等分散" (unchecked), a checkbox for "ウェルチ近似" (unchecked), and a "(%)信頼度" dropdown menu with "95" selected. At the bottom are buttons for "OK", "キャンセル", and "適用".

比率の区間推定

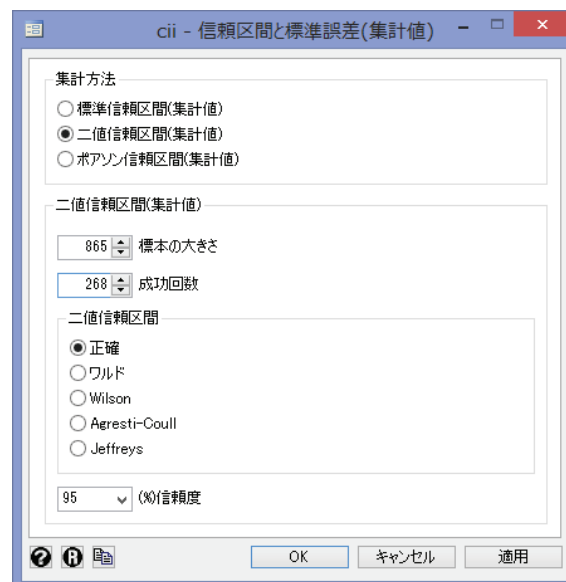
統計 (Statistics) メニュー

要約/表/検定 (Summaries, tables, and tests)

- ▶ 記述統計量 (Summary and descriptive statistics)
- ▶ 二値信頼区間 (集計値) (Binomial calculator)

画面で標本の大きさ (Sample size) に観測値の個数 $n = 865$, 成功回数 (Successes) に試行の成功回数 $x = 268$ ^a を入れる。

^a $n\hat{p} = 865 \times 0.31 = 268.15$ より, $x = 268$ とわかる。デフォルトの計算方法は正確 (Exact) だが, ワルド (Wald) を選べば正規分布に基づく近似的な信頼区間が出力される。



母比率 (母比率の差) の仮説検定

統計 (Statistics) メニュー

要約/表/検定 (Summaries, tables, and tests)

- ▶ 古典的仮説検定 (Classical tests of hypotheses)
- ▶ 比率検定 (集計値) (Proportion test calculator)

図 1.17: 比率の仮説検定:有意水準 5%

prtesti - 集計値による比率検定

集計値による比率検定

☒ 一群
☐ 二群

集計値による一群の比率検定

865 標本の大きさ
0.31 群の比率
0.4 比率(仮定)
95 (%)信頼度
☐ 比率でなく整数値で個数を示す

OK キャンセル 適用

図 1.18: 母比率の差の仮説検定:有意水準 5%

prtesti - 集計値による比率検定

集計値による比率検定

☐ 一群
☒ 二群

集計値による二群の比率検定

	標本1	標本2
サイズ	865	919
比率	0.31	0.40

95 (%)信頼度
☐ 比率でなく整数値で個数を示す

OK キャンセル 適用

1.4.

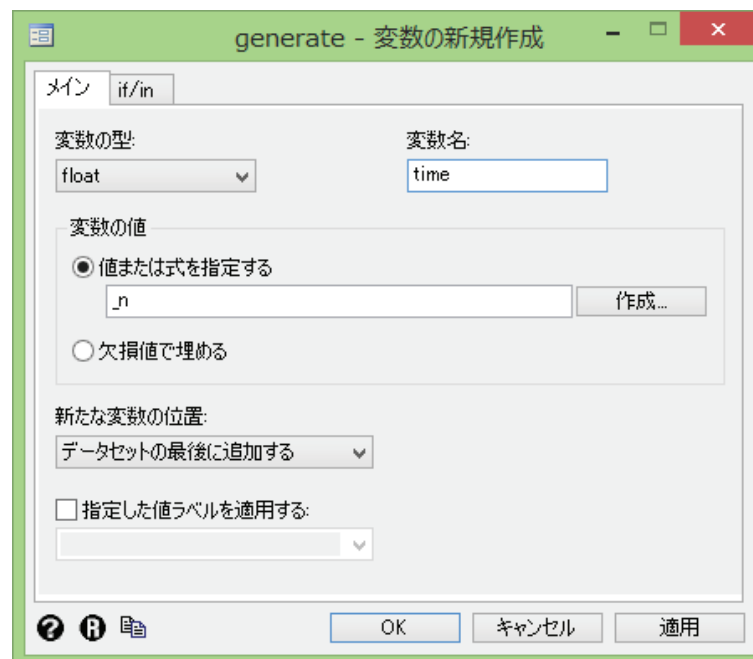
新しい変数を作成する

データ (Data) メニュー

データの作成または変更 (Create or change data)

▶ 新規変数の作成 (Create new variable)

図 1.19: 変数の定義



時間変数を指定する

統計 (Statistics) メニュー

時系列 (Time series)

▶ 設定とユーティリティ (Setup and utilities)

▶ 時系列データを定義 (Declare dataset to be time-series data)

図 1.20: 時間変数の指定

tsset - 時系列データの定義

メイン 増分

時間変数: time

パネルID変数: (任意)

時間変数の単位と表示形式

☒ 時間変数の形式を使用する

☐ 時間 ☐ 四半期

☐ 日 ☐ 半期

☐ 週 ☐ 年

☐ 月 ☐ 一般

設定...

時系列の設定をクリア

OK キャンセル 適用

単純回帰分析

統計 (Statistics) メニュー

線形モデル他 (Linear models and related)

▶ 線形回帰 (Linear regression)

図 1.21: 単純回帰分析

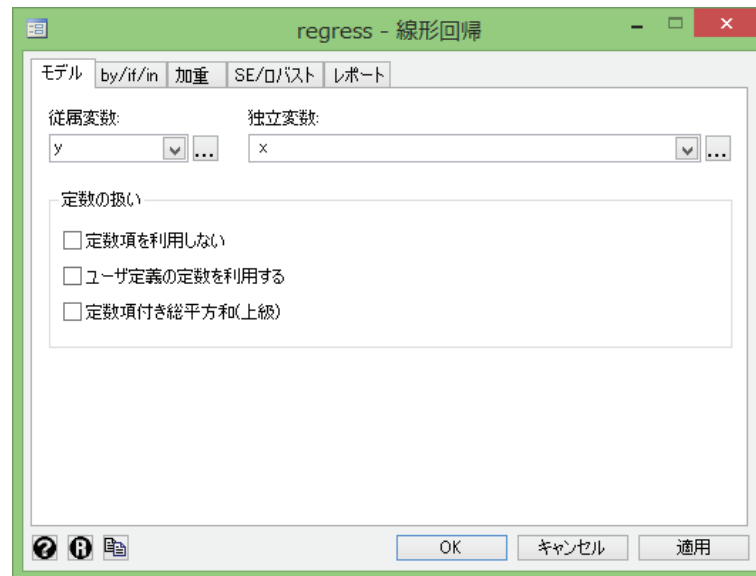


図 1.22: 推定された回帰直線と 95%信頼区間を指定する (1)

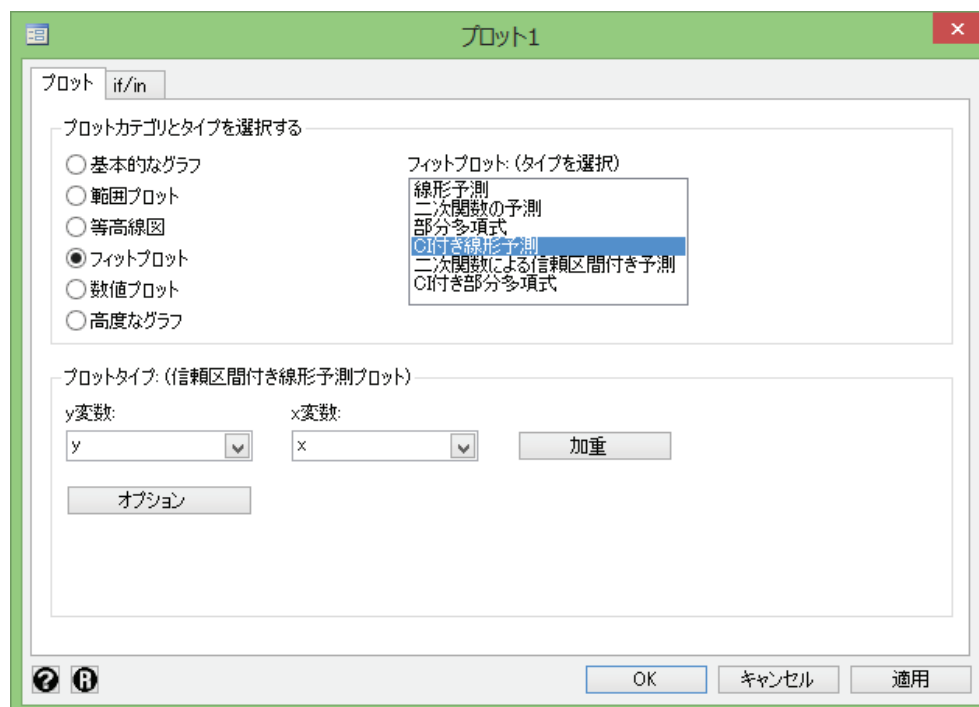


図 1.23: 推定された回帰直線と 95%信頼区間を指定する (2)

