

統計学(基礎)

第3回

基本統計量とグラフ出力、JASPの出力利用

データの整理・要約

基本統計量

データの整理・要約

- 基本統計量の算出
 - 度数の算出
 - 平均値、標準偏差、中央値、最大値、最小値などの算出
- 表・グラフによる整理
 - 棒グラフ、円グラフ、帯グラフ、折れ線グラフ、散布図
 - ヒストグラム
 - クロス集計表
- 記述統計でも推測統計でも行う

使うファイル

- JASPで開く
「開く」-「データライブラリ」-「1. Descriptives」
Plot layout.csv
- jamoviで開く
「開く」-「このPC」-「ブラウズ」
C:¥Program Files (x86)¥JASP¥resources¥Data
Sets¥Data Library¥1. Descriptives
Plot layout.csv
※Macの場合は Data Library以降

Plot layout: Plot Builder (beta)

- 説明
 - 南極のパーマー諸島(Palmer Archipelago)の3つの島で、2007～2009年に科学者が収集したペンギンのデータ
- 参考文献
 - Horst A. M., Hill A. P., & Gorman K. B. (2020). palmerpenguins: Palmer Archipelago (Antarctica) penguin data. R package version 0.1.0.

Plot layout: Plot Builder (beta)

- 変数

- species(名義): Adelie(アデリーペンギン)、Chinstrap(ヒゲペンギン)、Gentoo(ジェンツーペンギン)
- island(名義): データが収集された島(Biscoe, Dream, Torgersen)
- bill_length_mm (スケール): くちばしの長さ[mm]
- bill_depth_mm (スケール): くちばしの深さ(高さ)[mm]
- flipper_length_mm(スケール): ひれ(フリッパー)の長さ[mm]
- body_mass_g(スケール): 体重[g](5g単位に四捨五入)
- sex(名義): 性別(male / female)
- year(順序): 調査年(2007, 2008, 2009)

おまけ

- ペンギンについて
 - アデリーペンギン、ヒゲペンギン、ジェンツーペンギンはみな、アデリーペンギン属に属するペンギン、南極大陸に生息
- <https://nagoyaaqua.jp/study/column/27729/>
 - 名古屋港水族館(公益財団法人名古屋みなと振興財団)
スタッフコラム2025.09.16

JASPの記述統計

Plot layout*

データの編集

記述統計量

t検定

分散分析

混合モデル

回帰

度

記述統計

雨雲プロット

時系列記述統計

フレックスプロット

Plot Builder (beta)

species

island

bill_length_mm

bill_depth_mm

flipper_length_mm

body_mass_g

sex

year

変数

分割

記述表を転置する

統計量

サンプルサイズ

☒ 妥当

☒ 欠損値

分位数

☐ 四分位数

☐ カットポイント: 4 等分のグループ

☐ パーセンタイル値:

代表値

☐ 最頻値

☐ 中央値

☒ 平均値

ばらつき

☒ 標準偏差

☐ 変動係数

☐ 中央絶対偏差

☐ ロバスト中央値絶対偏差

☐ 四分位範囲

☐ 分散

☐ 範囲

☒ 最小値

☒ 最大値

分布

☐ 歪度

☐ 尖度

☐ シャピロ-ウィルク検定

☐ 総和

表

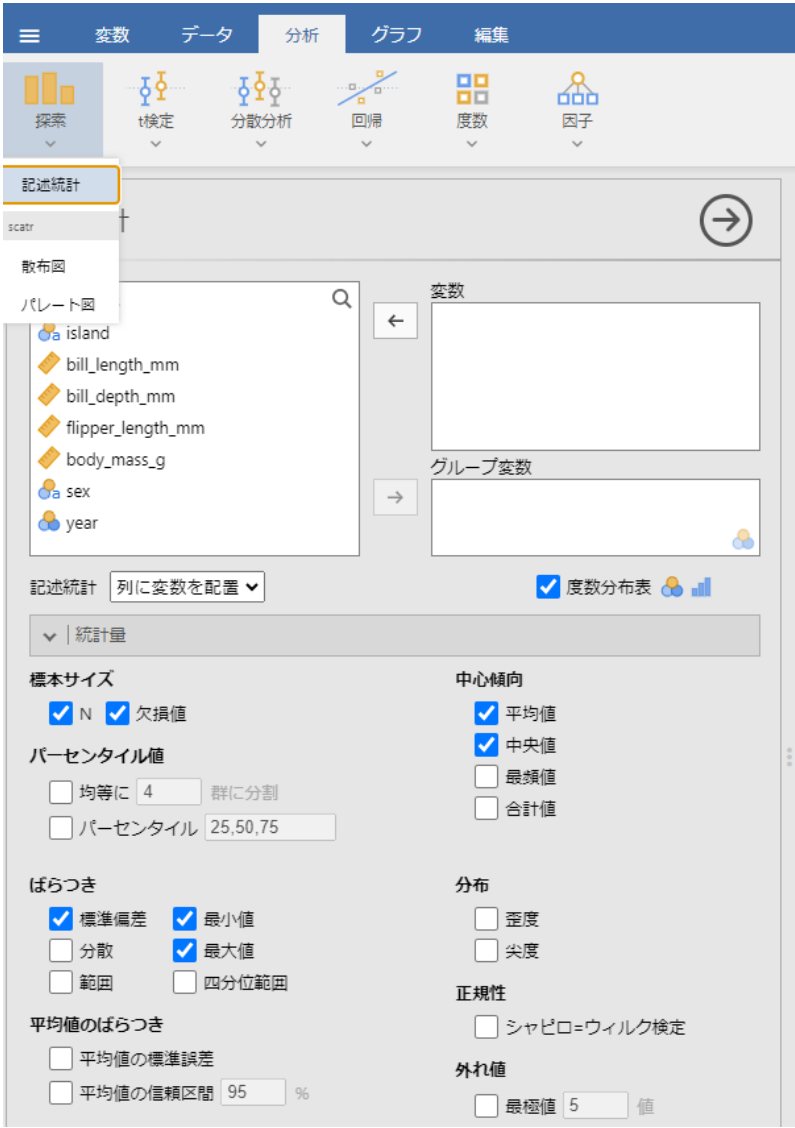
☒ 度数分布表

異なる値の最大値 10

☐ 幹葉図

スケール 1

jamoviの記述統計



記述統計出力

結果

記述統計

記述統計

	species	bill_length_mm
妥当	333	333
欠損値	0	0
平均値		43.99
標準偏差		5.469
最小値		32.10
最大値		59.60

度数分布表

species の頻度

species	頻度	パーセント	有効パーセント	累積パーセント
Adelie	146	43.8	43.8	43.8
Chinstrap	68	20.4	20.4	64.3
Gentoo	119	35.7	35.7	100.0
欠損値	0	0.0		
合計	333	100.0		

注 以下の変数には10 点の重複値が含まれるので、省略されています：
bill_length_mm.

結果

記述統計

記述統計

	species	bill_length_mm
N	333	333
欠損値	0	0
平均値		44.0
中央値		44.5
標準偏差		5.47
最小値		32.1
最大値		59.6

度数

speciesの度数分布表

species	度数	全体%	累積%
Adelie	146	43.8 %	43.8 %
Chinstrap	68	20.4 %	64.3 %
Gentoo	119	35.7 %	100.0 %

基本統計量

質的変数(名義・順序)

- 度数
- 相対度数(度数割合)
- 累積度数
- 累積相対度数(累積度数割合)

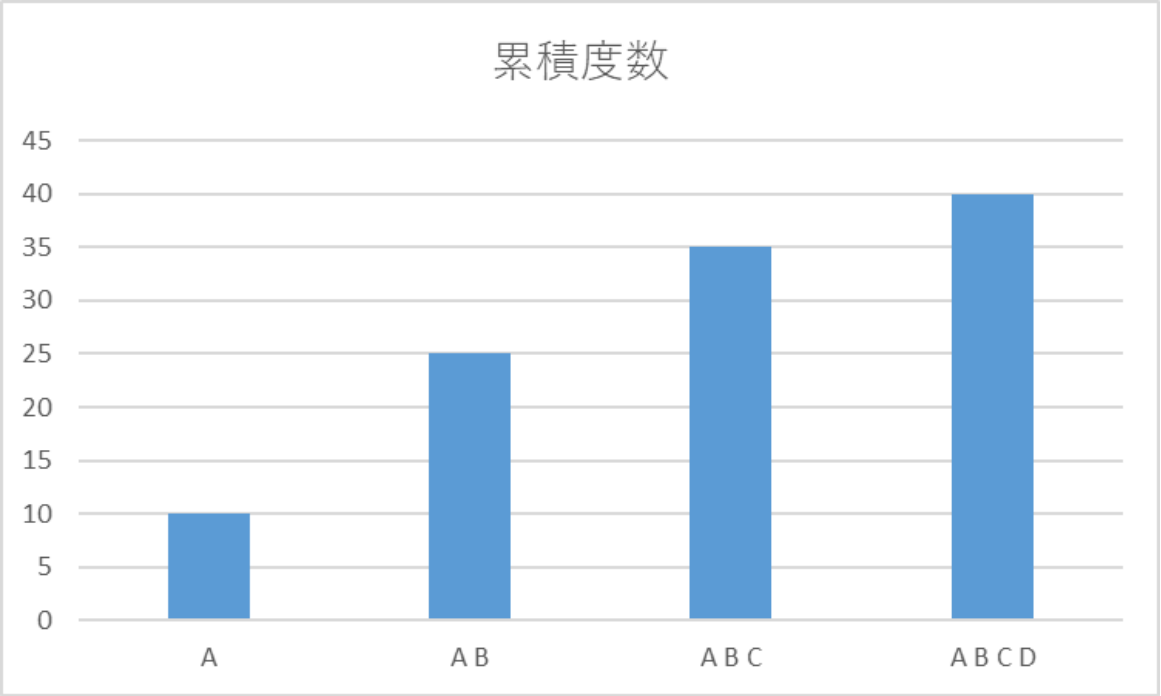
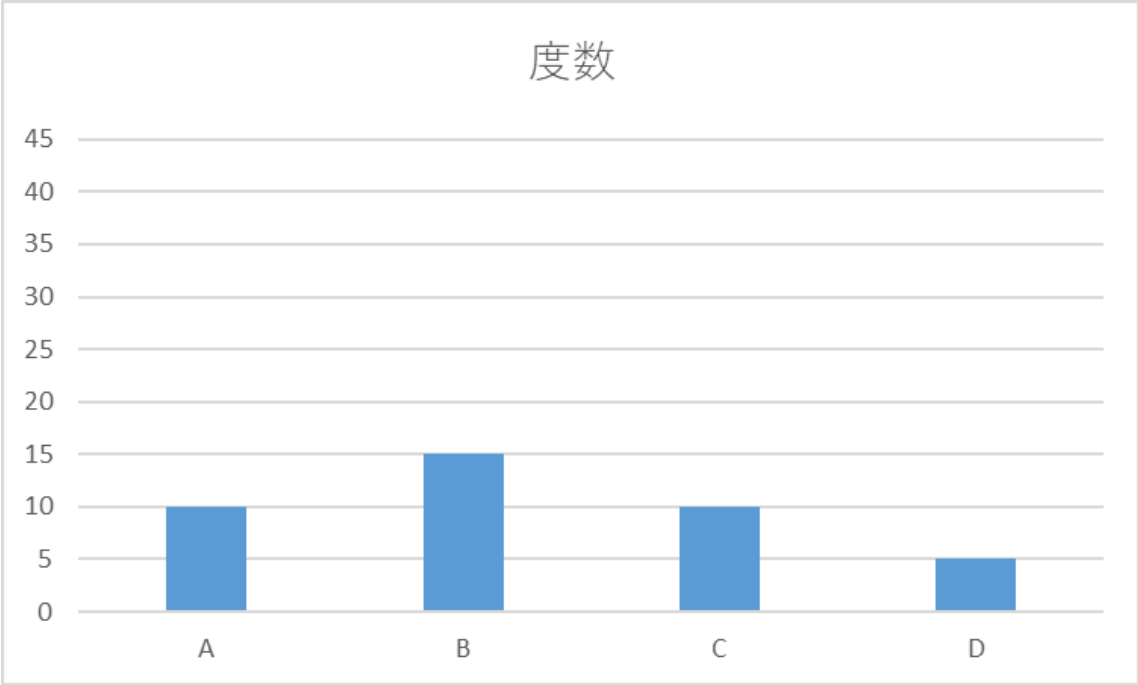
量的変数

- 平均
- 中央値
- 最大値
- 最小値
- 分散
- 標準偏差
- など

質的変数用の基本統計量

- 度数
 - カテゴリに該当する数
- 相対度数(度数割合)
 - 合計に対する各度数の割合
- 累積度数
 - カテゴリを順に累積した度数
- 累積相対度数(累積度数割合)
 - 合計に対する各累積度数の割合

度数と累積度数(イメージ)



質的変数用の基本統計量

- 度数
 - カテゴリに該当する数
 - 相対度数(度数割合)
 - 合計に対する各度数の割合
 - 累積度数
 - カテゴリを順に累積した度数
 - 累積相対度数(累積度数割合)
 - 合計に対する各累積度数の割合
- 名義尺度、順序尺度
のどちらでも
- 順序尺度でないとい
意味が無い

量的変数の基本統計量

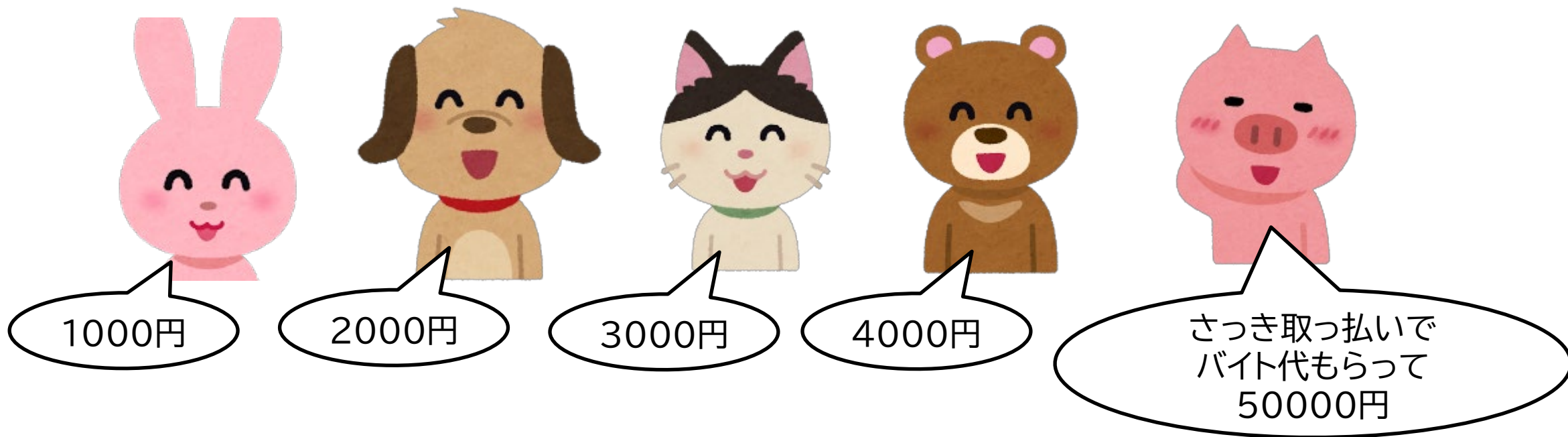
- 平均値
- 最大値
- 最小値
- 中央値
- 分散
- 標準偏差

平均値

- データをすべて足して、足した数で割る
 - 極端に大きな値、小さな値があるとその影響を受ける
 - 真ん中ではない

平均値→真ん中ではない

- 例 今お財布にいくら入ってる？



※取っ払い→現金支払いのこと

かわいいフリー素材集 いらすとや <https://www.irasutoya.com/> 1/7/61

平均値→真ん中ではない

- 例 今お財布にいくら入ってる？



1000



2000



3000



4000



50000

合計は、

 $1000 + 2000 + 3000 + 4000 + 50000$ $= 60000$ 平均は、 $60000 \div 5 = 12000$

でも、平均より多い人は1人しかいない

最大値・最小値・中央値

- 最大値
 - 一番大きな値
- 最小値
 - 一番小さな値
- 中央値
 - 真ん中の値

最大値・最小値・中央値

- 例 今お財布にいくら入ってる？



1000 ← 最小値



2000



3000 ← 中央値 中央値は真ん中なので、必ずそれより大きい
人と小さい人が同じだけいる



4000



50000 ← 最大値

中央値 データが偶数の時

- 例 今お財布にいくら入ってる？



1000



2000



3000



4000

←この間が中央値 この場合は2500

真ん中の2つのデータを足して2で割る。
中央値は真ん中なので、必ずそれより大きい人と小さい人が同じだけいる。

標準偏差

- 標準偏差は平均値からのばらつきの平均(のようなもの)
 1. 各値の平均値との差を出す
 2. 「各値の平均値との差」を2乗する
 3. 「各値の平均値との差の2乗」を全部足す
 4. 「各値の平均値との差の2乗を全部足したもの」をデータ数で割る
(ここまでが分散)
 5. 「各値の平均値との差の2乗を全部足したものをデータ数で割ったもの」の平方根($\sqrt{\quad}$)を求める

標準偏差をなぜ求めるのか

- A・Bの2クラスで100点満点のテストをしました。

Aクラス



Bクラス



かわいいフリー素材集 いらすとや <https://www.irasutoya.com/> 23/61

標準偏差

Aクラス



点数

30	40	50	60	70
----	----	----	----	----

Bクラス



点数

48	49	50	51	52
----	----	----	----	----

標準偏差をなぜ求めるのか

- A 30 40 50 60 70
平均値50 中央値50
- B 48 49 50 51 52
平均値50 中央値50
- 平均値と中央値だけなら同じ集団と言えてしまう
 - 今回は5つずつしかデータがないから見て違いもわかるけど、データが100個以上あるときなどは気づかない

標準偏差

- 違いに気づきたい→平均値からの差を出す

A	点数	30	40	50	60	70
	平均値との差	-20	-10	0	10	20
B	点数	48	49	50	51	52
	平均値との差	-2	-1	0	1	2

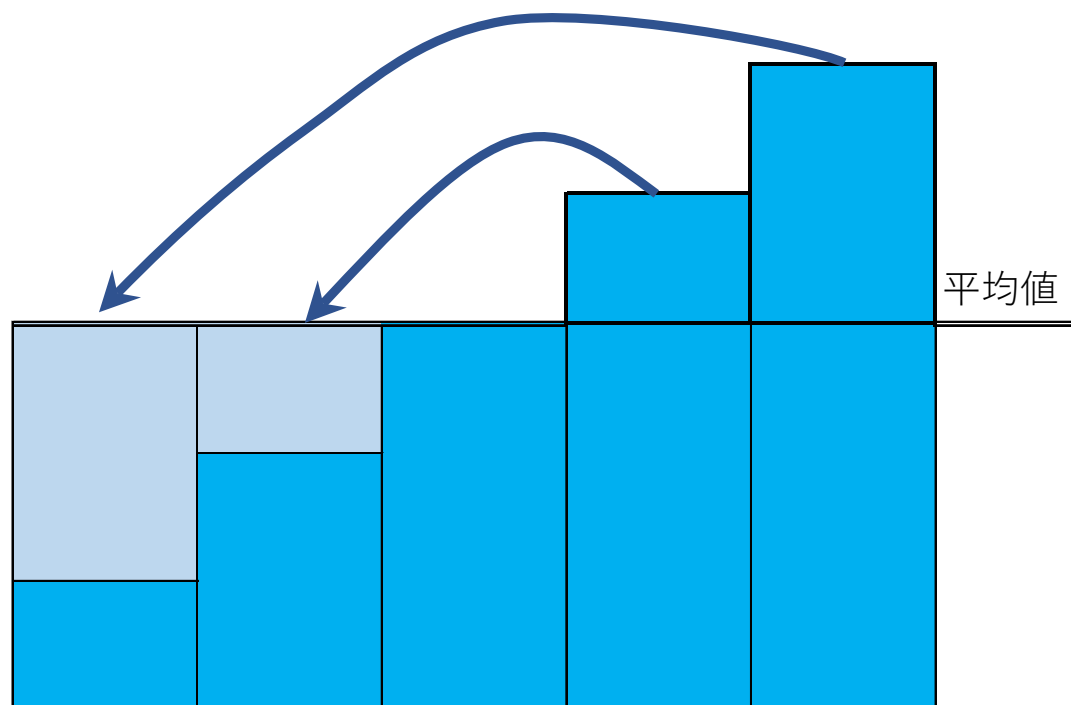
標準偏差

- 「平均値からの差」で違いが言えそうなので、「平均値からの差」の平均を出したいが、合計が0になる

A	点数	30	40	50	60	70	合計0
	平均値との差	-20	-10	0	10	20	
B	点数	48	49	50	51	52	合計0
	平均値との差	-2	-1	0	1	2	

平均値の考え方

- 各値と平均値の差を合計すると0になる



- 平均値の差からのオーバー部分とアンダー部分は同じ
- 平均値は全体の凸凹を平らにしたもの
- なので、平均値からの差の合計は0

標準偏差

- どうか「平均値からの差」を使いたい
- マイナスを取るために2乗する

A	点数	30	40	50	60	70	合計 1000
	平均値との差	-20	-10	0	10	20	
	↑の2乗	400	100	0	100	400	
B	点数	48	49	50	51	52	合計 10
	平均値との差	-2	-1	0	1	2	
	↑の2乗	4	1	0	1	4	

標準偏差

- 2乗したものの合計をデータ数で割る

A

点数	30	40	50	60	70
平均値との差	-20	-10	0	10	20
↑の2乗	400	100	0	100	400

2乗したものの合計
1000

$1000 \div 5 = 200 \leftarrow \text{これが「分散」}$

B

点数	48	49	50	51	52
平均値との差	-2	-1	0	1	2
↑の2乗	4	1	0	1	4

2乗したものの合計
10

$10 \div 5 = 2 \leftarrow \text{これが「分散」}$

分散はあんまり便利じゃない

A

点数	30	40	50	60	70
平均値との差	-20	-10	0	10	20
↑の2乗	400	100	0	100	400

B

点数	48	49	50	51	52
平均値との差	-2	-1	0	1	2
↑の2乗	4	1	0	1	4

- Aの分散は200
- Bの分散は2
- Bは平均50で最小値が48、最大値が52なので分散が2でもなんとなくわかる
- Aは最小値30、最大値70で分散が200といわれてもピンとこない

標準偏差

- 分散を平方根にして見やすくした

A

点数	30	40	50	60	70
平均値との差	-20	-10	0	10	20
↑の2乗	400	100	0	100	400

分散 $1000 \div 5 = 200$

その平方根

$\sqrt{200} = \sqrt{2 \times 10^2} = 10\sqrt{2}$
 $\div 14.142 \leftarrow$ 標準偏差

B

点数	48	49	50	51	52
平均値との差	-2	-1	0	1	2
↑の2乗	4	1	0	1	4

分散 $10 \div 5 = 2$

その平方根

$\sqrt{2} \div 1.414 \leftarrow$ 標準偏差

標準偏差

Aクラス



点数

30	40	50	60	70
----	----	----	----	----

平均値 50
中央値 50
標準偏差 14.1

Bクラス



点数

48	49	50	51	52
----	----	----	----	----

平均値 50
中央値 50
標準偏差 1.4

かわいいフリー素材集 いらすとや
<https://www.irasutoya.com/> 33/61

標準偏差とは

- 平均値からのばらつきの平均(のようなもの)
 1. 各値の平均値との差を出す
 2. 「各値の平均値との差」を2乗する
 3. 「各値の平均値との差の2乗」を全部足す
 4. 「各値の平均値との差の2乗を全部足したもの」をデータ数で割る(ここまでが分散)
 5. 「各値の平均値との差の2乗を全部足したものをデータ数で割ったもの」の平方根($\sqrt{\quad}$)を求める

標本標準偏差

1. 各値の平均値との差を出す
2. 「各値の平均値との差」を2乗する
3. 「各値の平均値との差の2乗」を全部足す
4. 「各値の平均値との差の2乗を全部足したもの」をデータ数-1で割る(ここまでが不偏分散)
5. 「各値の平均値との差の2乗を全部足したものをデータ数-1で割ったもの」の平方根($\sqrt{\quad}$)を求める

標準偏差はどっちだ

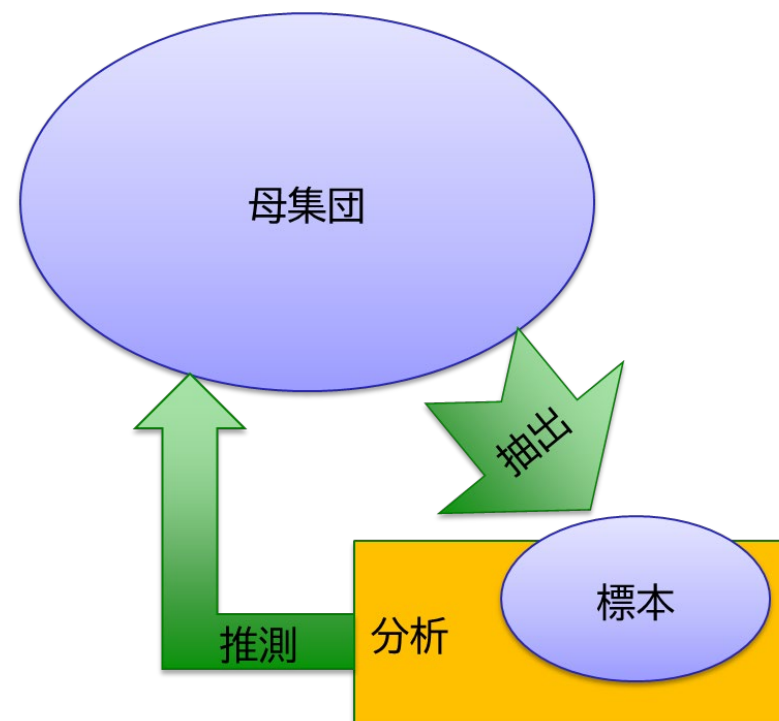
- 「各値の平均値との差の2乗を全部足したものをデータ数で割ったもの」の平方根($\sqrt{\quad}$)を求める
 - 標準偏差
- 「各値の平均値との差の2乗を全部足したものをデータ数-1で割ったもの」の平方根($\sqrt{\quad}$)を求める
 - 標本標準偏差→標本データから求めた母集団の標準偏差の推定値

記述統計では考えない

推定

推定

- 抽出した実測データから母集団のデータを推定する
 - 点推定
 - 区間推定



点推定

- 母平均の推定値
- 母分散の推定値
- 平均値の標準誤差

母平均の推定値

- 母平均の推定値＝標本平均
 - 標本の平均を、そのまま母集団の平均値の推定値と見なす
 - 母集団の平均値＝母平均

母分散と不偏分散 母標準偏差と標本標準偏差

- 不偏分散(標本分散)は、母分散の不偏推定値
- データ数が多ければ、不偏分散と母分散は同じと見なせる
- 母標準偏差と、標本標準偏差も同じ考え

- 手持ちのデータから推測を行うときの考え方

母集団についての補足

- 「母集団(Population)」の「母」は、性別を表すものではない
- ここでの「母」は「もとになる」「起点となる」という意味
 - 例: 母国語＝もととなる言語、母校＝出身校
- 統計学では、調査や実験の対象となる全体の集団を表す既定の専門用語として「母集団」という語を広く用いている
 - したがって、この言葉はジェンダー的な意味合いを持つものではなく、統計学上の慣用表現として理解してください。

出力されているのは

- JASPやjamoviで出力される標準偏差は、標本標準偏差($n-1$ で割っている方)
 - 用途として推測統計に使うのが圧倒的だから
 - 統計解析アプリは「統一的な計算基準を保つ」目的で、標本標準偏差($n-1$)に揃えてあることがほとんど
 - Excelの関数だと、`stdev.p`と`stdev.s`で別になっている

出力の使い方

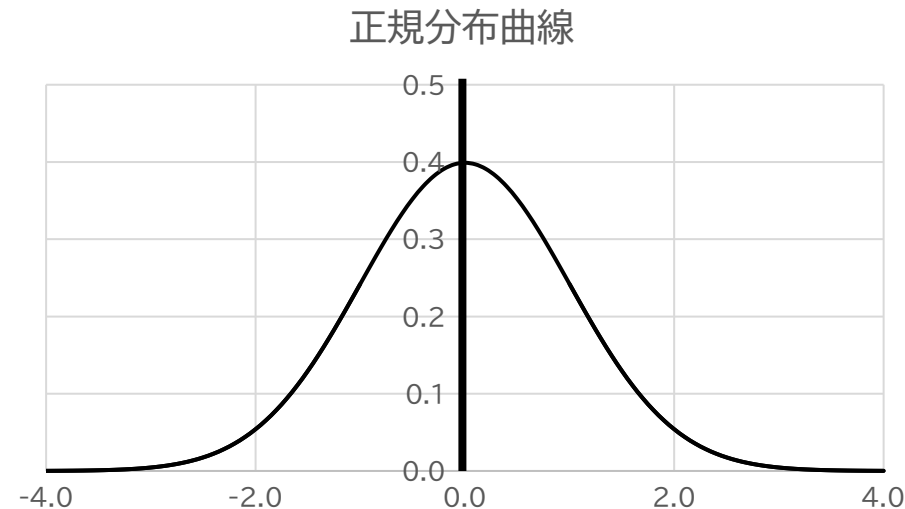
- 行政や地域保健のデータのように、母集団全体を把握している(=全数調査)場合は、理論的には「母標準偏差(分母 = N)」が正しい
- 実際の分析では $n-1$ 補正をしても数値上の差はごくわずか
 - 特に自治体レベルで数千～数万世帯の規模になると、両者の差は小数点3～4桁以下となり、政策判断に影響するようなレベルではなくなる
 - 数理的厳密さを求めると「母標準偏差」
 - 実務・分析手法の整合性を保つなら「標本標準偏差」でも問題なし。

(平均値の)標準誤差

- $\frac{\delta}{\sqrt{n}}$
 1. 標本標準偏差を求める(本来は母集団の標準偏差の推定値)
 2. 1.をデータ数の平方根($\sqrt{\quad}$)で割る
- 標準誤差は推定の精度
 - 標準誤差が小さくなる条件は
 - 分散が小さい
 - データ数が多い

区間推定

- 母平均値を範囲で推定する
 - 標本抽出を多数回繰り返したときに、標本平均の平均は正規分布という分布に従う→中心極限定理



区間推定

- 平均値の95%信頼区間
 - 95%の確率で母平均が取り得る範囲
- 平均値 \pm (t分布の2.5%点 \times 標準誤差)が95%信頼区間
平均値 $-$ t分布の2.5%点 \times 標準誤差 $<$ 平均値 $<$ 平均値 $+$ t分布の2.5%点 \times 標準誤差

グラフ

記述統計量のグラフ JASP

☒ 最大値

推論

☐ 標準誤差

☐ 平均値の信頼区間

幅 %

方法 T model

☐ 標準偏差の信頼区間

幅 %

方法 Analytical (chi-square)

☐ 分散の信頼区間

幅 %

方法 Analytical (chi-square)

連関行列

☐ 共分散

☐ 相関

使用 全て

▼ 基本的な図

☒ 分布図

☐ 相関のプロット

☐ 密度を表示する

☐ ラグマークを表示する

ピン幅タイプ スタージェス

ピンの数

☐ 区間プロット

☐ Q-Qプロット

☐ 円グラフ

☐ ドットプロット

カテゴリカルプロット

☐ バレートプロット

☐ バレートルール %

☐ リッカートプロット

☐ すべての変数は同じ水準を共有していると仮定

縦軸用の調節可能なフォントサイズ 正規

▶ カスタマイズ可能なプロット

▼ 表

☒ 度数分布表

☐ 幹葉図

異なる値の最大値

スケール

分布のプロット ▼

species ▼

species	count
Adelie	145
Chinstrap	70
Gentoo	120

species

bill_length_mm

bill_length_mm

49/61

川崎市立看護大学大学院 看護学研究科 博士前期課程

©Ryota Takayanagi 2025

記述統計のグラフ JASP

▼ 基本的な図

☒ 分布図

☐ 相関のプロット

☐ 密度を表示する

☐ ラグマークを表示する

ピン幅タイプ

スタージェス

ピンの数

30

☐ 区間プロット

☐ Q-Qプロット

☒ 円グラフ

☐ ドットプロット

カテゴリカルプロット

☐ パレートプロット

☐ パレートルール

95

 %

☐ リッカープロット

☐ すべての変数は同じ水準を共有していると仮定

縦軸用の調節可能なフォントサイズ

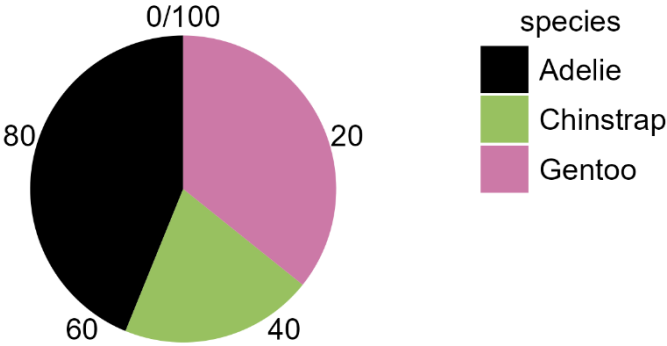
正規

▶ カスタマイズ可能なプロット

注 以下の変数には10 点の重複値が含まれるので、省略されています：
bill_length_mm.

円グラフ

species



記述統計量のグラフ JASP

▼ カスタマイズ可能なプロット

カラーパレット カラーブラインド ▼

☒ ボックスプロット

☒ 箱ひげ図 要素 ☐ カラーパレットを使用

☐ パイオリン 要素 ☐ 外れ値をラベル

☐ Jitter 要素

☐ 散布図

散布図の上のグラフ

☐ 密度

☐ ヒストグラム

☐ なし

☐ 回帰直線の追加

☒ 線形

☐ 平滑化

☐ 信頼区間を表示 95 %

散布図の右のグラフ

☐ 密度

☐ ヒストグラム

☐ なし

☐ 凡例を表示する

☐ 度数図

island sex year

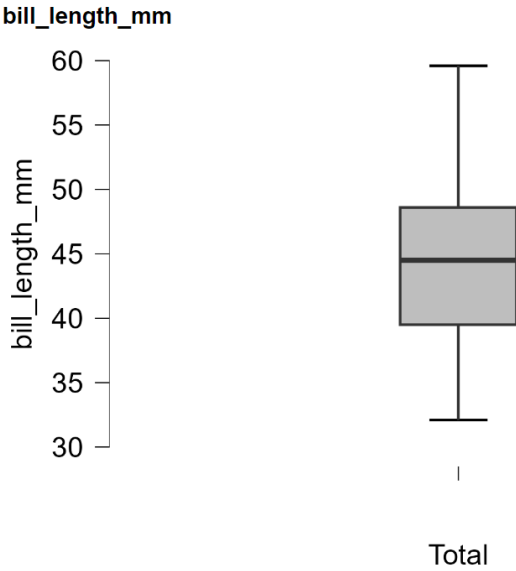
↓

度数を分離:

Gentoo	119	35.7	35.7	100.0
欠損値	0	0.0		
合計	333	100.0		

注 以下の変数には10 点の重複値が含まれるので、省略されています :
bill_length_mm.

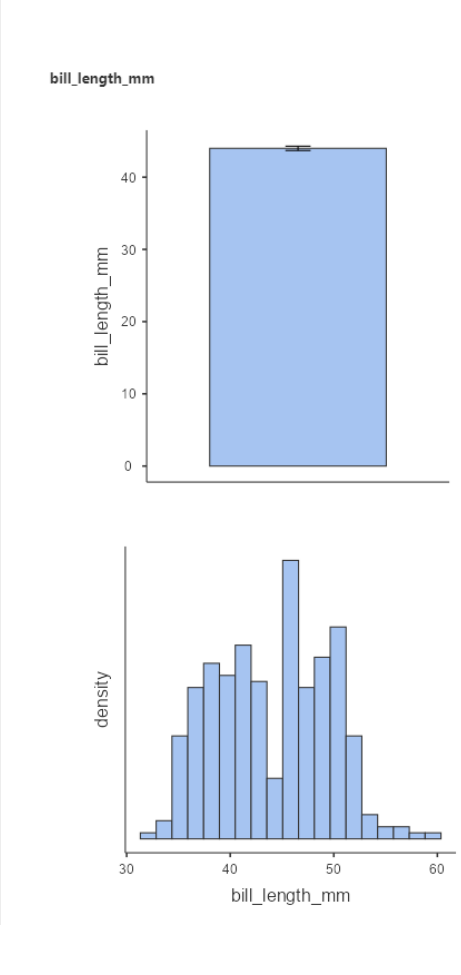
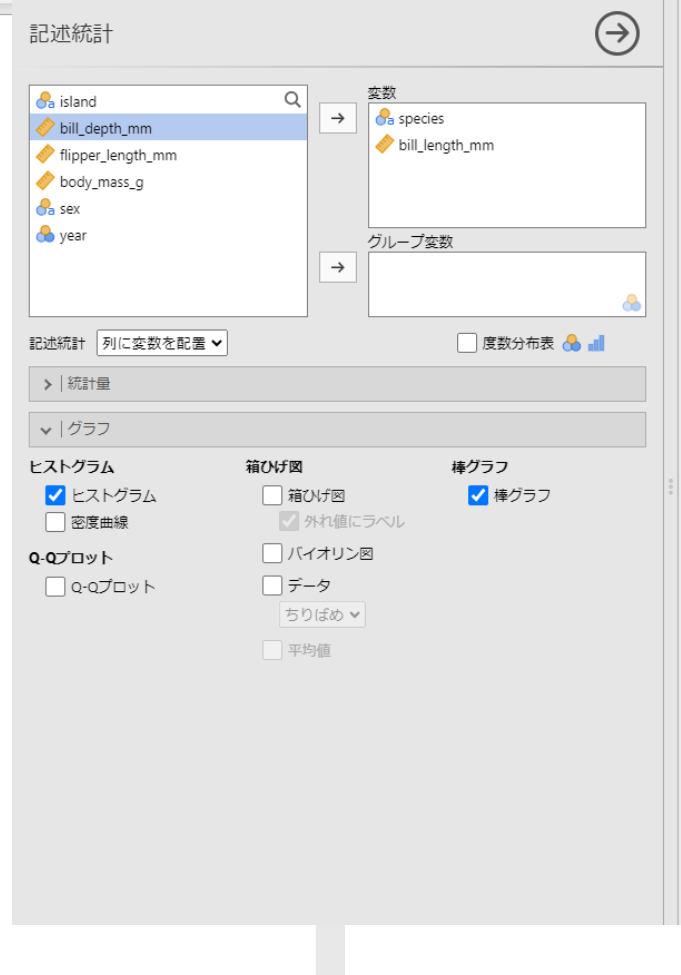
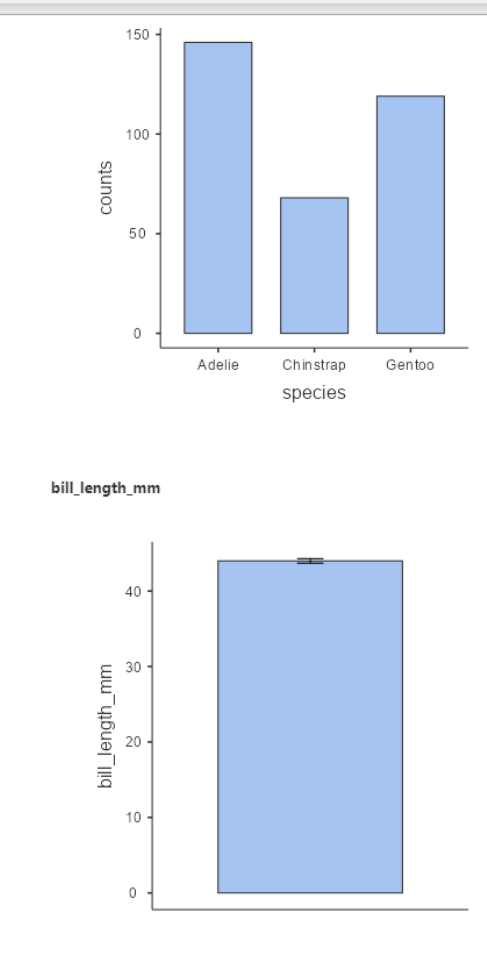
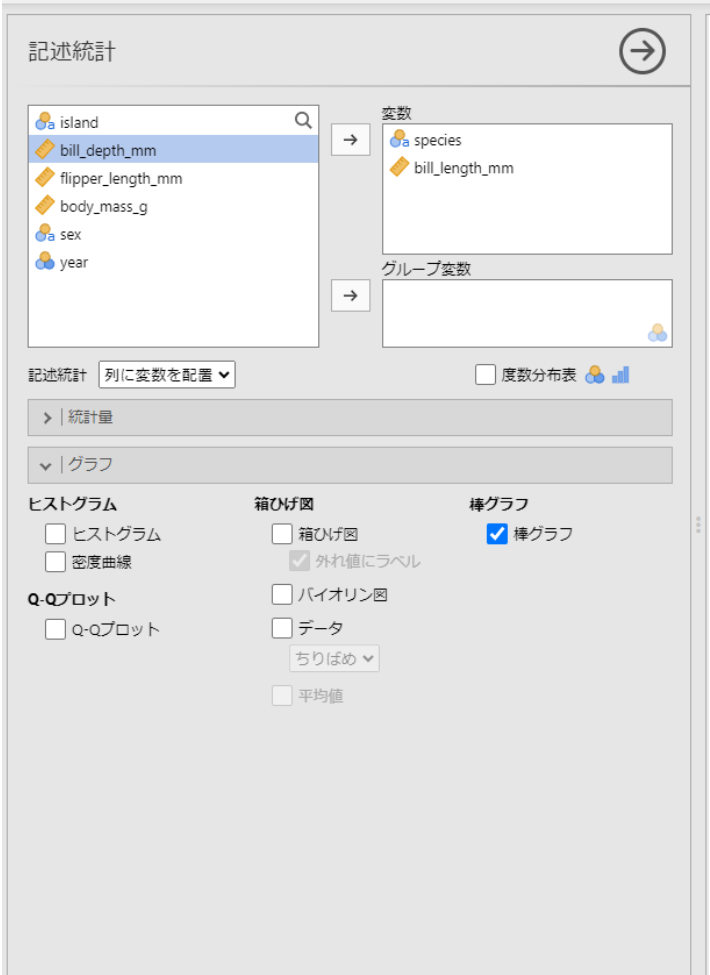
ボックスプロット



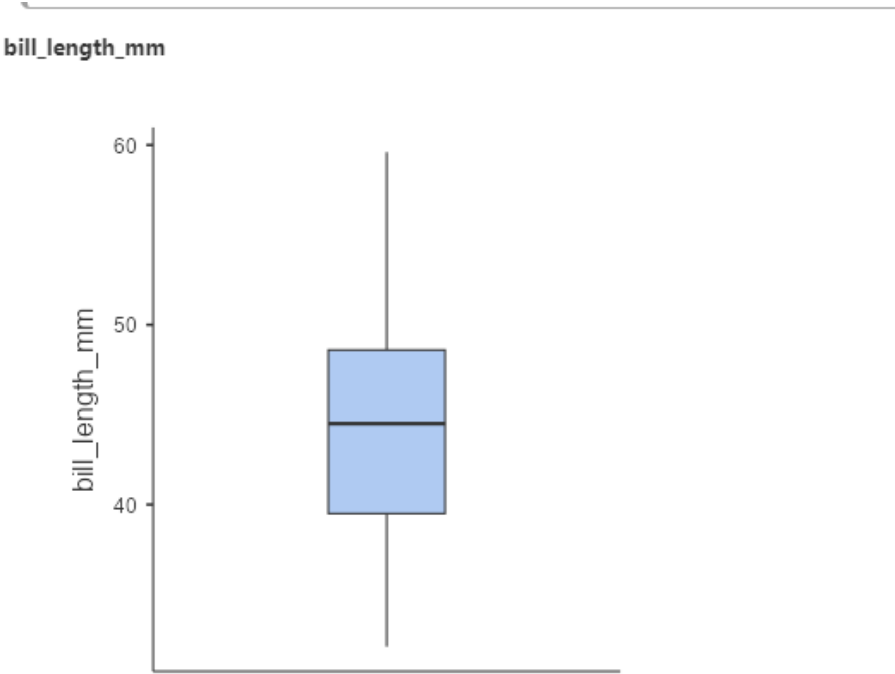
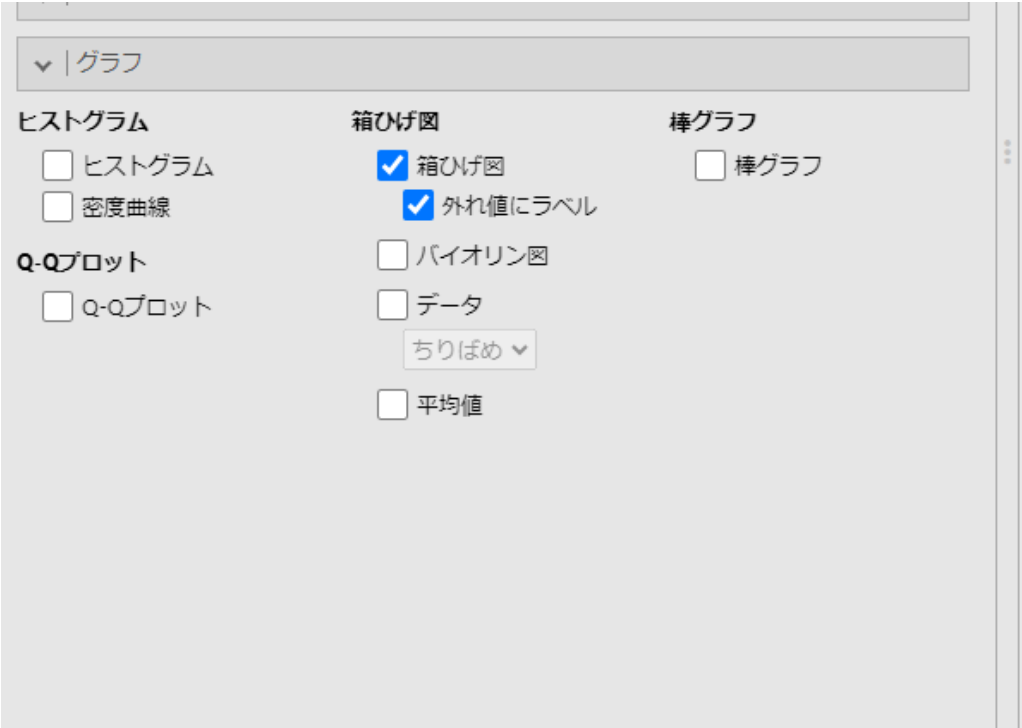
記述統計量のグラフ JASP



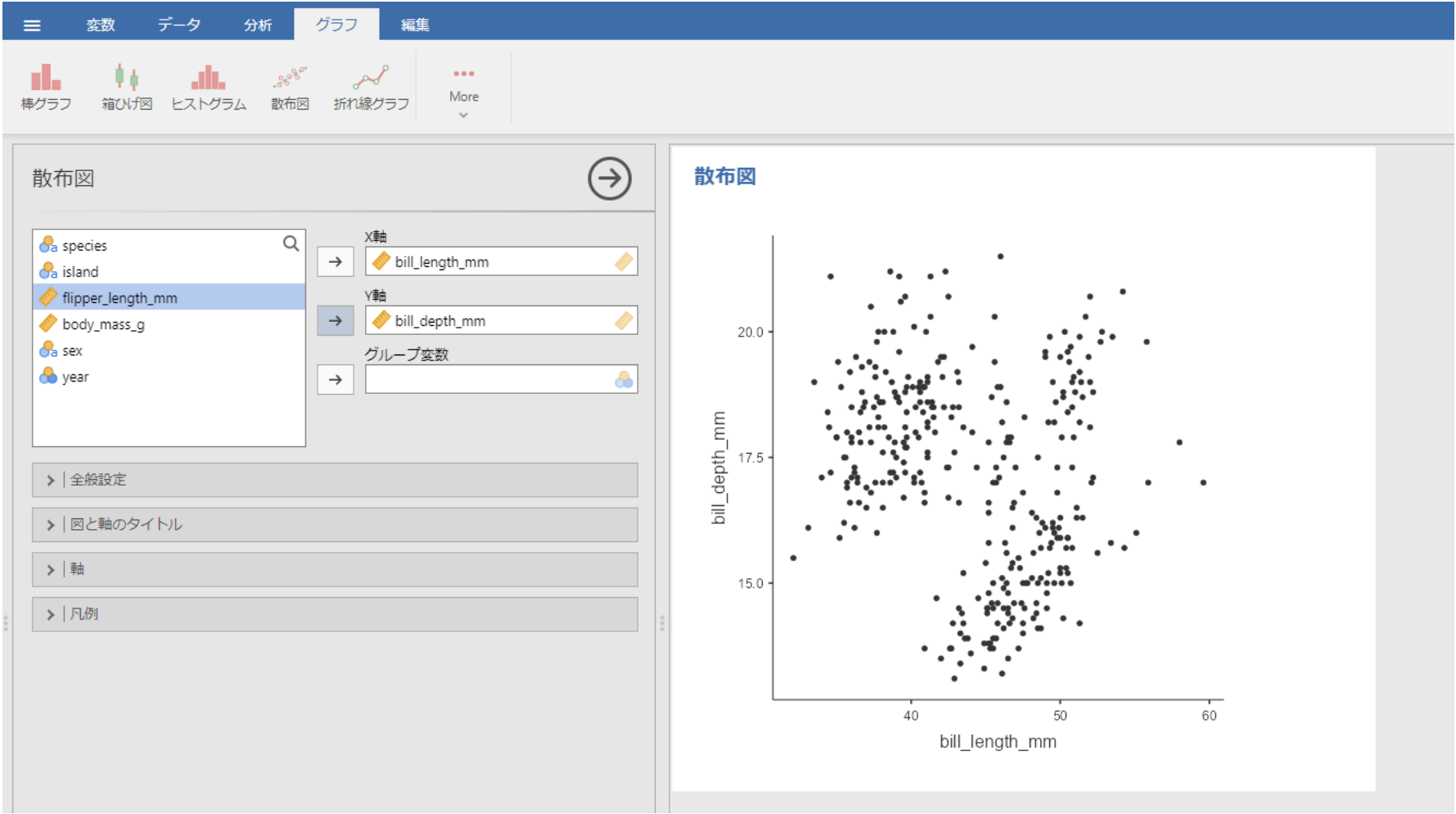
記述統計量のグラフ jamovi



記述統計量のグラフ jamovi



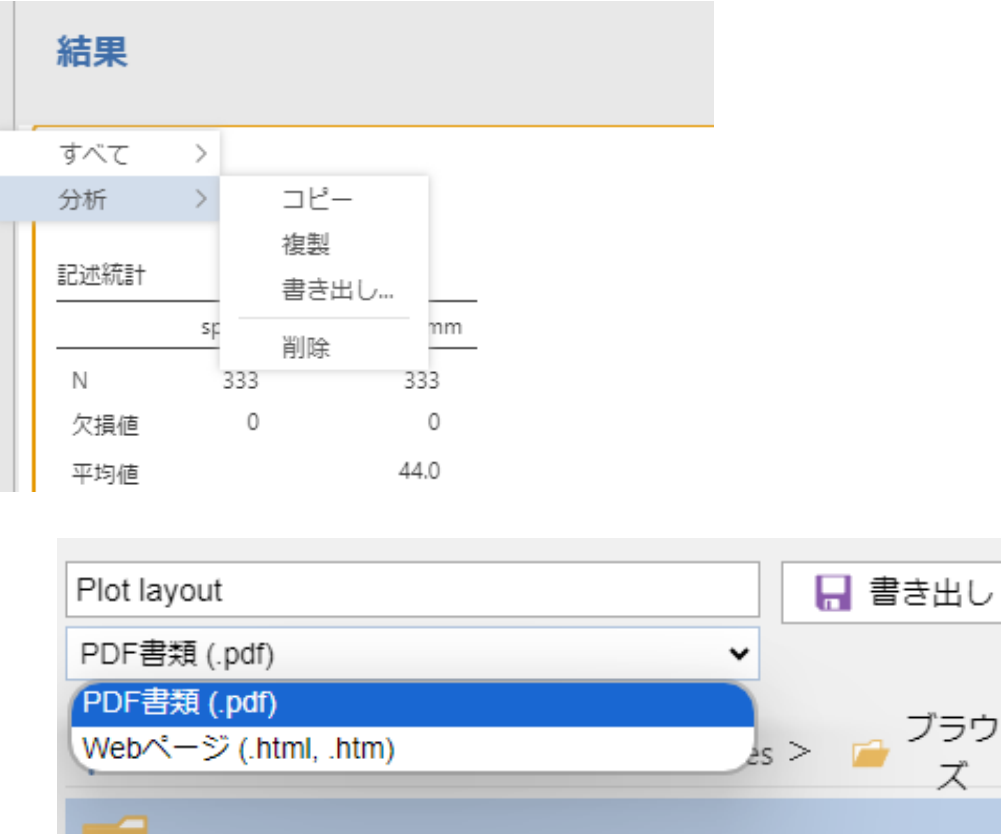
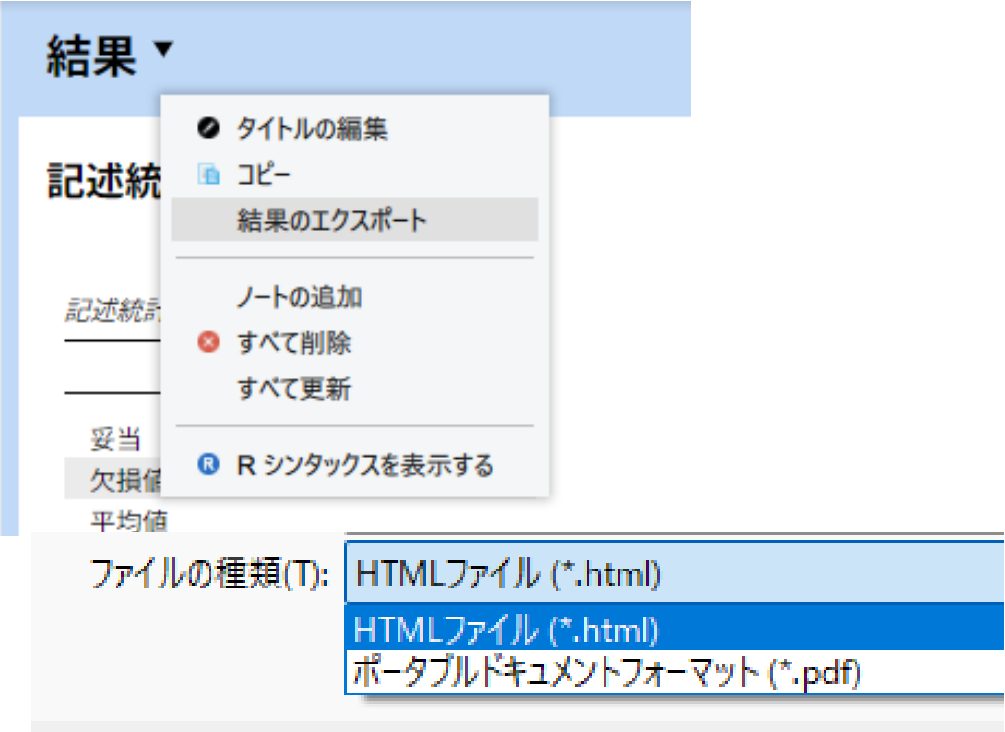
記述統計量のグラフ jamovi



出力の扱い

出力の利用

- 全体
 - HTMLかPDF



出力の利用

・ 部分

・ 個別

度数分布表 ▼

同居の頻度

同居	頻度	パーセント	有効パーセント	累積パーセント
独居	0	0.0	60.0	60.0
同居	40	40.0	40.0	100.0
欠損値	0	0.0		
合計	100	100.0		

朝食の頻度

朝食	頻度	パーセント	有効パーセント	累積パーセント
食べる	50	50.0	50.0	50.0
食べない	50	50.0	50.0	100.0
欠損値	0	0.0		
合計	100	100.0		

同居の頻度 ▼

同居	頻度	パーセント	有効パーセント	累積パーセント
独居	0	0.0	60.0	60.0
同居	40	40.0	40.0	100.0
欠損値	0	0.0		
合計	100	100.0		

結果

すべて >

分析 >

記述統計

	sf	mm
N	333	333
欠損値	0	0
平均値		44.0

コピーをするとHTML形式

- Wordへの貼り付け

度数分布表

同居の頻度				
同居	頻度	パーセント	有効パーセント	累積パーセント
独居	60	60.0	60.0	60.0
同居	40	40.0	40.0	100.0
欠損値	0	0.0		
合計	100	100.0		

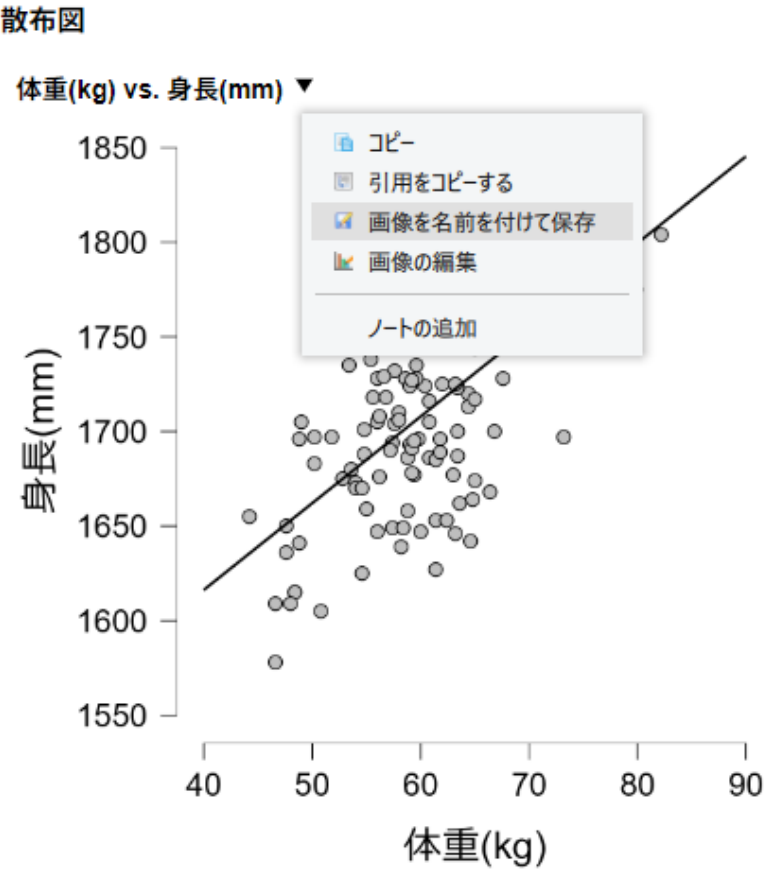
朝食の頻度

朝食	頻度	パーセント	有効パーセント	累積パーセント
食べる	50	50.0	50.0	50.0
食べない	50	50.0	50.0	100.0
欠損値	0	0.0		
合計	100	100.0		

Ctrl

グラフ

- 形式



ファイル名(N):

ファイルの種類(T):

フォルダーの非表示

- ポータブルネットワークグラフィックス(*.png)
- ポータブルネットワークグラフィックス(*.png)
- ポータブルドキュメント形式(*.pdf)
- エンカプセルドポストスクリプト (*.eps)
- 300 dpi タグ付きイメージ ファイル (*.tiff)
- PowerPoint (*.pptx)
- スケーラブルベクターグラフィックス(*.svg)

Plot layout

PDF書類 (.pdf)

PDF書類 (.pdf)

PNG画像 (.png)

SVG画像 (.svg)

EPS画像 (.eps)

PowerPoint Slide (.pptx)

書き出し

ブラウズ

※全体の一括保存を選択したときにグラフがあると、png
ファイルとなって、HTML内で埋め込まれる

グラフ出力形式

- png
- tiff
- pdf
- pptx(PowerPoint)

