統計学(基礎)

第14回
ベイズ統計

ベイズ統計とは

- ベイズ統計という手法があるわけではない
- 結果を解釈するための指標に対する考え方

  仮説の評価

    頻度流：p値

    ベイズ：ベイズファクター(BF)

頻度流(frequentist)

- いわゆる従来の方法(p値で見る方法)
  - 仮説を先に固定する
  - その仮説の下でこの結果がどれくらいの確率で起こりそうか(p:有意確率)を見る
    - 「差がない」という仮説の下で、このデータが得られる確率はいくつ(どれくらい珍しい)か？

ベイズ統計(Bayesian Statistics)

- ベイズファクター(BF)で仮説に対する比率を見る
  - 結果(データ)を先に固定する
  - その結果に対してどちらの仮説がどれくらい当てはまりがよいか(尤度)を見る
    - データが得られたとき、「差がある仮説(対立仮説)」と「差がない仮説(帰無仮説)」ではどちらの方が説明しやすいか?

頻度流統計とベイズ統計の違い

- 頻度論：（帰無）仮説の下で、分析結果になるのはどれくらいの確率であるかを見る
  - 差が無いという仮説の下で、この値が得られる確率はどれくらい？
- ベイズ統計：得られた結果が、どちらの仮説の下の方がより起こりそうか（尤度）
  - この値は、どちらの仮説の方がより起こりやすい？

ベイズ統計の簡単な考え方

- 事後オッズ＝事前オッズ×ベイズファクター

- 事前オッズ(対立仮説と帰無仮説の確率比)はわからないので、とりあえず1としておく
- 分析をしてベイズファクターを出す

  結果＝事前オッズ×ベイズファクターなので、次回同じ分析をするときの事前オッズは更新されている

ベイズファクター(BF)

- 対立仮説が起こりそうな尤度(likelifood)と、帰無仮説が起こりそうな尤度の比

$$BF_{10} = 対立仮説 \div 帰無仮説$$
$$BF_{01} = 帰無仮説 \div 対立仮説$$

  - BF_{10}を使うことの方が多い
    - 帰無仮説よりも対立仮説が何倍　という説明にしやすい
    - $BF_{10} > 1$なら対立仮説の方が起こりやすく、$BF_{10} < 1$なら対立仮説の方が起こりにくい

頻度流は悪くもないし無くなりもしない

- p値(有意確率)がα(有意水準)を超えたか超えていないかという閾値で考えるとおかしくなる
  - αを下回ったかどうかでなく、pを直接記述する
  - 閾値($p<0.05$とか)ではなく、連続量として考える
  - 発症の有無や効果の有無などは頻度流のほうが説明しやすい

アメリカ統計学会の声明

- The ASA’s Statement on p-Values: Context, Process, and Purpose
  - アメリカ統計学会によるP値に関する声明:文脈経緯目的
  - Ronald L. Wasserstein & Nicole A. Lazar
    The American Statistician, Vol.70, No.2 (2016), pp.129–133.

アメリカ統計学会の声明

1. P値は、データが帰無仮説とどの程度一致しているかを示す指標である。
2. P値は、仮説が正しい確率でも、結果が偶然に得られた確率でもない。
3. 科学的結論や政策決定をP値だけで判断してはならない。
4. 適切な推論には、研究設計前提データ品質効果量既存知識などの文脈が必要である。
5. ある特定のP値(例:0.05)を境に“有意”/“非有意”と二分する慣習は避けるべきである。
6. P値は透明性と完全な報告の一部として扱うべきである。

その後の声明（2019）

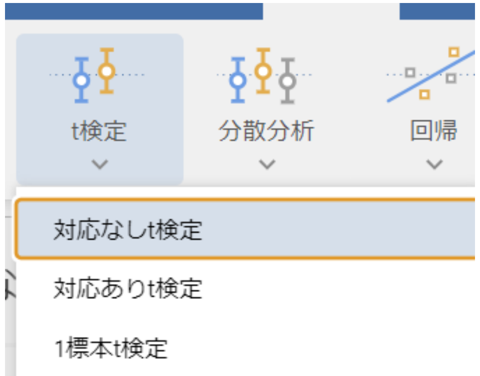
- 「$p < 0.05$」を閾値として使う慣習をやめよう
- Statistically significant” という言葉も避けよう
  - 代わりに：
    - 効果量（effect size）
    - 信頼区間（confidence interval）
    - 事前知識や理論的根拠
    - ベイズ推論再現性の重視
  - *Moving to a World Beyond ‘p < 0.05’* Wasserstein, Schirm & Lazar, *The American Statistician*, 2019

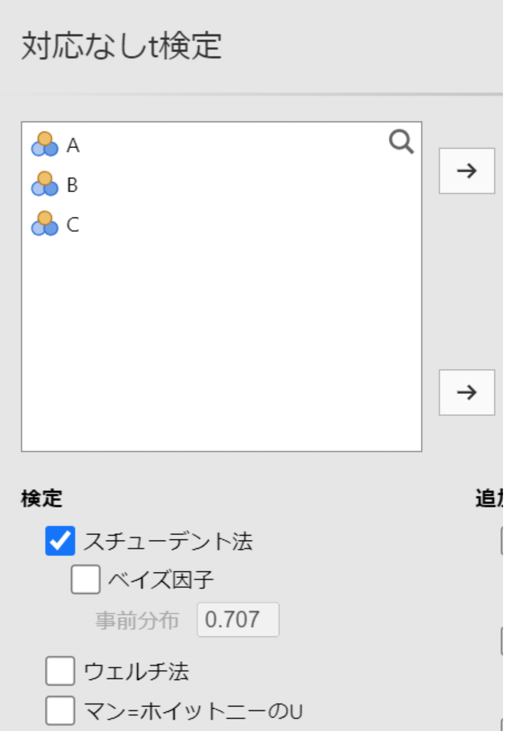
2群の差でやってみよう

同じ検定を頻度流とベイズで解釈する

その前に

- JASPにはデフォルトでベイズ統計メニューがある
- jamoviはちょっと一癖ある



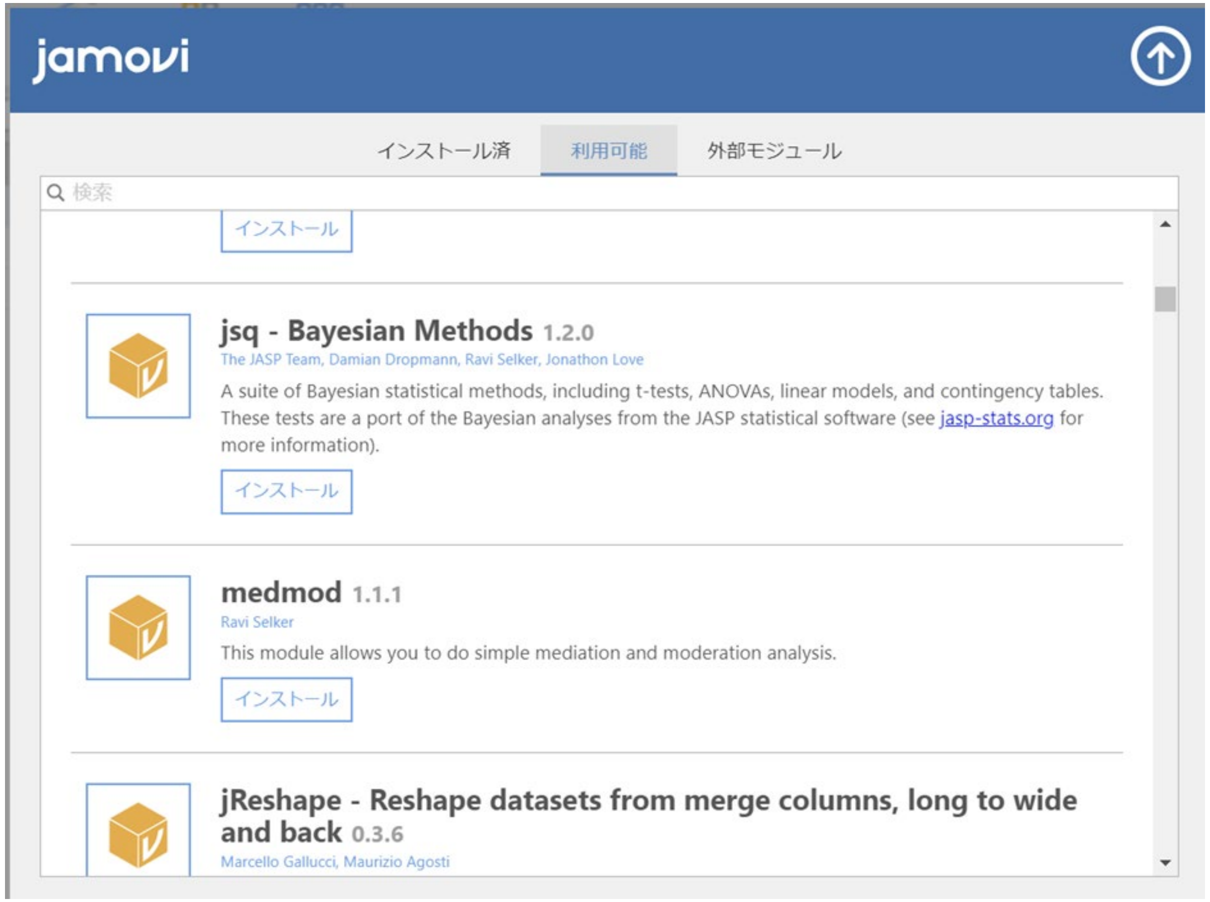


jamoviのベイズ統計

- 分析メニューでモジュールを追加できる

モジュールの追加

- jamoviライブラリから「利用可能」タブを選択して「jsq – Baysean Methods」を選択

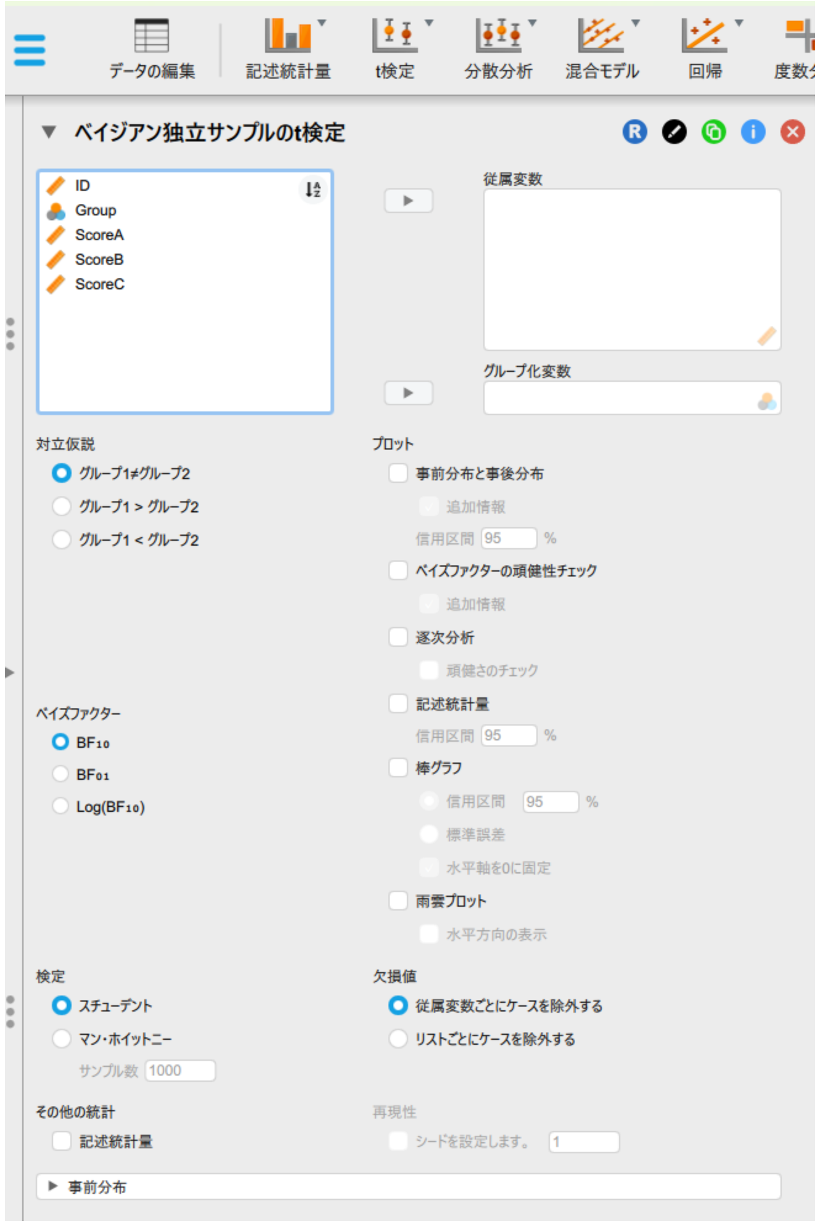


モジュールの追加

- インストールすると「モジュール」の「インストール済みモジュール」に表示される
- メニューにも追加される
  - 追加モジュールは日本語版がないことが多い

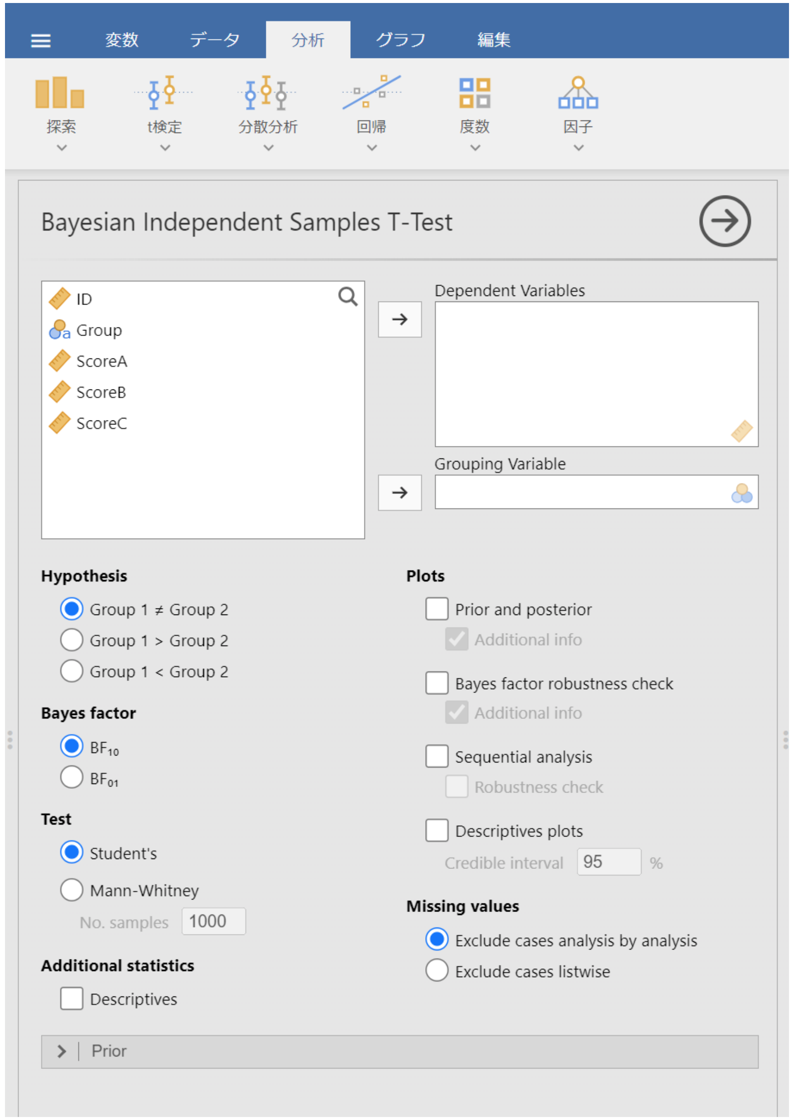
jsq Bayesian Methods

  - The JASP Team, Damian Dropmann, Ravi Selker, Jonathon Love
- A suite of Bayesian statistical methods, including t-tests, ANOVAs, linear models, and contingency tables.
- These tests are a port of the Bayesian analyses from the JASP statistical software (see jasp-stats.org for more information).

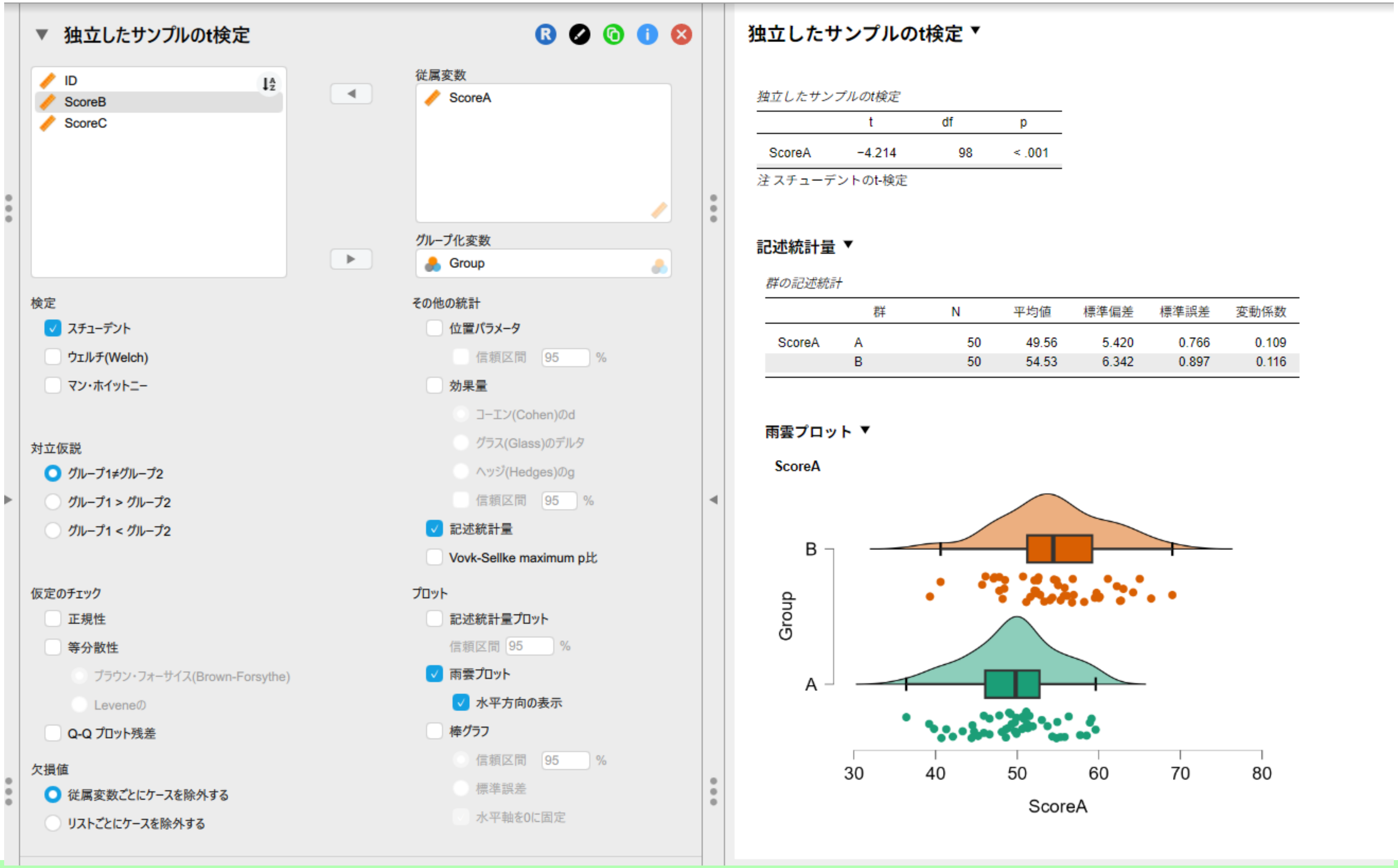


ほぼおなじ

だって、移植してるから



t検定(頻度流 JASP)


独立したサンプルのt検定
ID
ScoreB
ScoreC
従属変数
ScoreA
グループ化変数
Group
検定
スチューデント
ウェルチ(Welch)
マン・ホイットニー
対立仮説
グループ1≠グループ2
グループ1 > グループ2
グループ1 < グループ2
仮定のチェック
正規性
等分散性
ブラウン・フォーサイス(Brown-Forsythe)
Leveneの
Q-Q プロット残差
欠損値
従属変数ごとにケースを除外する
リストごとにケースを除外する
その他の統計
位置パラメータ
信頼区間 95 %
効果量
コーエン(Cohen)のd
グラス(Glass)のデルタ
ヘッジ(Hedges)のg
信頼区間 95 %
記述統計量
Vovk-Sellke maximum p比
プロット
記述統計量プロット
信頼区間 95 %
雨雲プロット
水平方向の表示
棒グラフ
信頼区間 95 %
標準誤差
水平軸を0に固定
独立したサンプルのt検定
独立したサンプルのt検定
t df p
ScoreA -4.214 98 < .001
注 スチューデントのt-検定
記述統計量
群の記述統計
群 N 平均値 標準偏差 標準誤差 変動係数
ScoreA A 50 49.56 5.420 0.766 0.109
B 50 54.53 6.342 0.897 0.116
雨雲プロット
ScoreA
B
A
Group
30 40 50 60 70 80
ScoreA

t検定（頻度流 jamovi）

t検定(ベイズ JASP)

ベイジアン独立サンプルのt検定

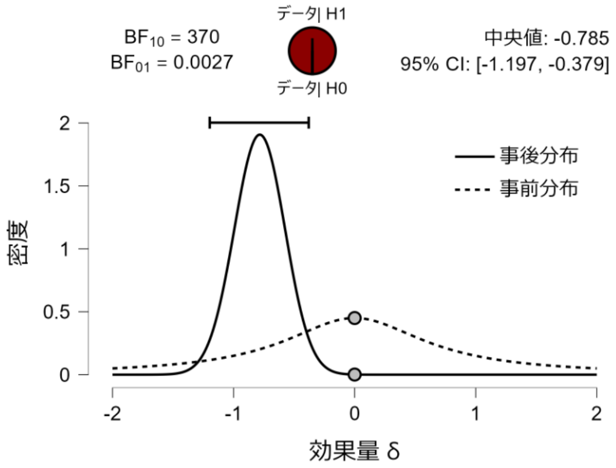
ベイジアン独立サンプルのt検定

	BF_{10}	エラー %
ScoreA	369.8	2.267×10^{-9}

推論プロット

ScoreA

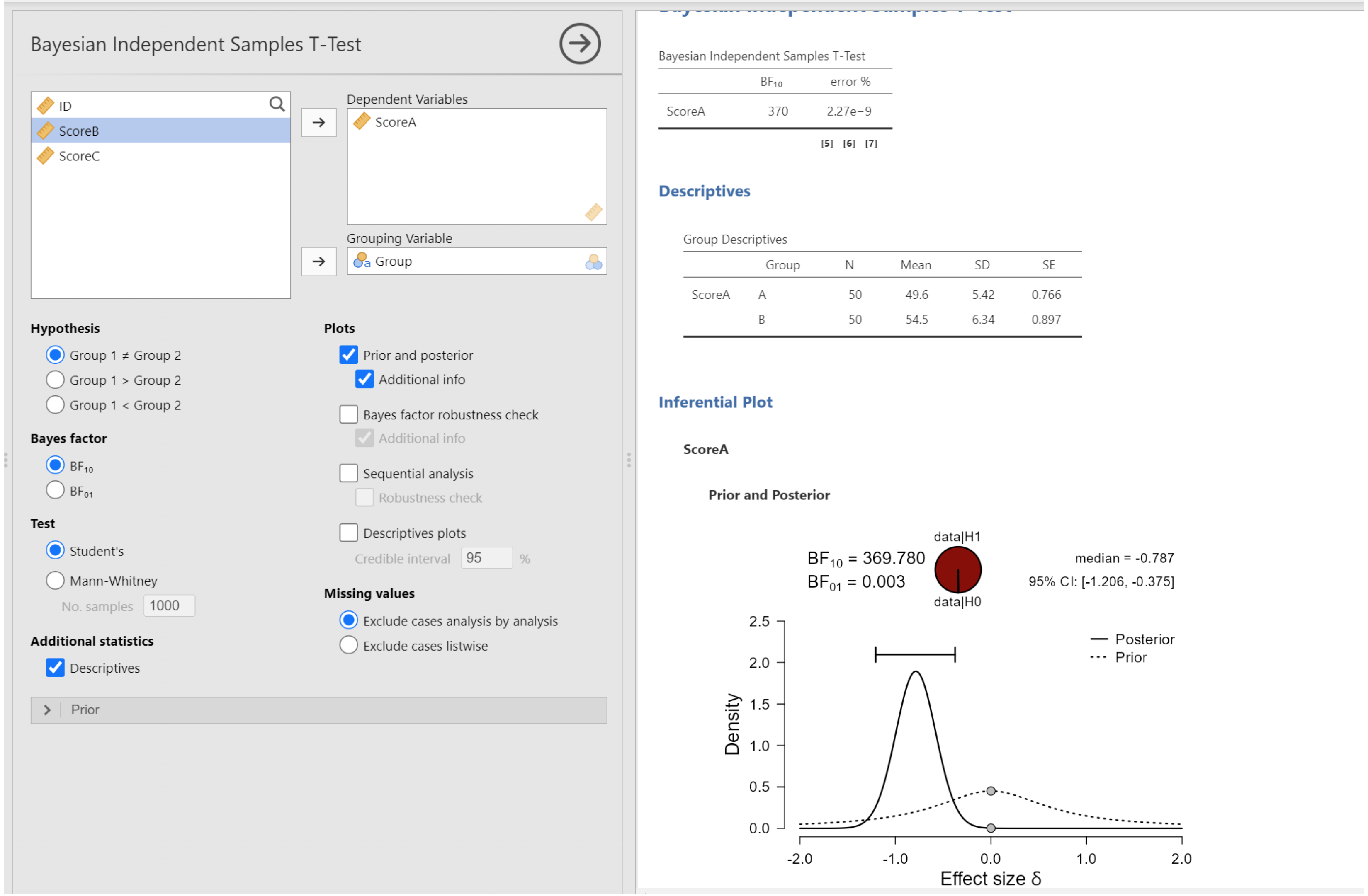
事前分布と事後分布



記述統計量

	群	N	平均値	標準偏差	標準誤差	変動係数	95% 信用区間 下限	95% 信用区間 上限
ScoreA	A	50	49.56	5.420	0.766	0.109	48.02	51.10
	B	50	54.53	6.342	0.897	0.116	52.73	56.33

t検定(ベイズ jamovi)



ベイズ t検定の考え方

- 計算に使う統計量(t値)は頻度流もベイズも同じ

独立したサンプルのt検定 ▼

独立したサンプルのt検定

	t	df	p
ScoreA	−4.214	98	< .001

注 スチューデントのt-検定

記述統計量 ▼

群の記述統計 ▼

	群	N	平均値	標準偏差	標準誤差	変動係数
ScoreA	A	50	49.56	5.420	0.766	0.109
	B	50	54.53	6.342	0.897	0.116

ベイジアン独立サンプルのt検定 ▼

ベイジアン独立サンプルのt検定

	BF_{10}	エラー %
ScoreA	369.8	2.267×10^{-9}

対応なしt検定

対応なしt検定

		統計量	±%	自由度	p
ScoreA	スチューデントのt	−4.21		98.0	< .001
	ベイズ因子$_{10}$	370	2.27e−11		

ノート. H_a $\mu_A \neq \mu_B$

グループ統計量

	グループ	N	平均値	中央値	標準偏差	標準誤差
ScoreA	A	50	49.6	49.8	5.42	0.766
	B	50	54.5	54.4	6.34	0.897

ベイズ t検定の解釈

- 仮説の評価の考え方は異なる
  - 頻度流のt検定はp値で説明する
  - ベイズの場合はベイジアンファクター(BF)で説明する

独立したサンプルのt検定 ▼

独立したサンプルのt検定

	t	df	p
ScoreA	−4.214	98	< .001

注 スチューデントのt-検定

ベイジアン独立サンプルのt検定 ▼

ベイジアン独立サンプルのt検定

	BF_{10}	エラー %
ScoreA	369.8	2.267×10^{-9}

ベイズファクター(BF)

- 対立仮説が起こりそうな尤度(likelifood)と、帰無仮説が起こりそうな尤度の比

$$BF_{10} = 対立仮説 \div 帰無仮説$$

$$BF_{01} = 帰無仮説 \div 対立仮説$$

  - BF_{10}を使うことの方が多い
    - 帰無仮説よりも対立仮説が何倍 という説明にしやすい
    - $BF_{10} > 1$なら対立仮説の方が起こりやすく、$BF_{10} < 1$なら対立仮説の方が起こりにくい

ベイズのt検定の出力

- BF
- モデルの事後確率円グラフ
- δ(効果量)の中央値と95%信頼区間
- δの事前分布と事後分布

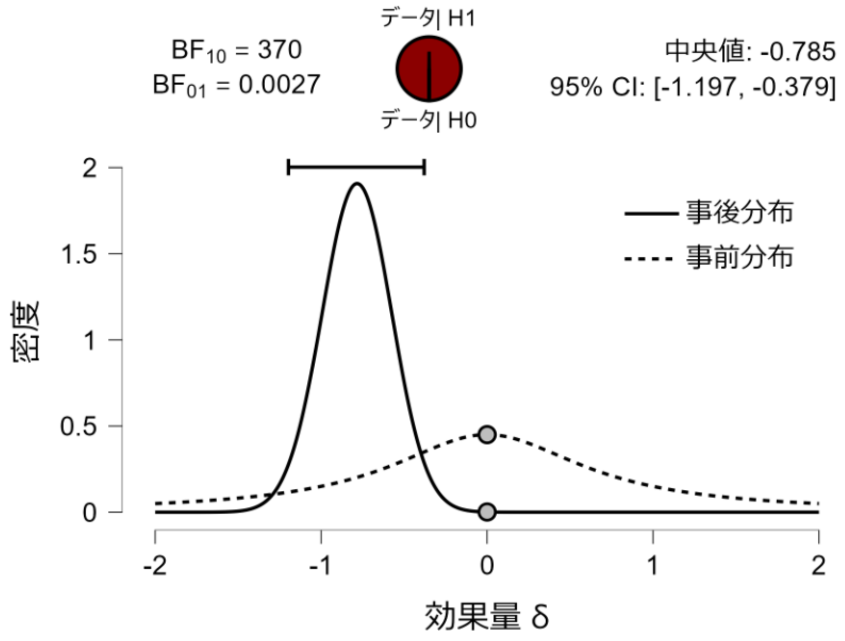
ベイジアン独立サンプルのt検定

	BF_{10}	エラー %
ScoreA	369.8	2.267×10^{-9}

推論プロット

ScoreA

事前分布と事後分布



モデルの事後確率グラフ

- モデル(H_0 / H_1)の事後確率
- H_0 とH_1のどちらが優勢か
  - ベイズファクターそのものではなく、それを事後確率に変換した結果
  - 色が付いている方がH_1 (0にも1にもならない)

$$P(H_1|D) = \frac{BF_{10}}{BF_{10}+1}$$

  - 事後確率は 事前分布の仮定に依存する

事後確率グラフの意味

SUNTORY将棋オールスター
東西対抗戦2025 決勝戦　第5局
藤井聡太竜王名人ー中村太地八段戦
2025年12月14日　ABEMA将棋チャンネル

ベイズの δ の読み方

- δは2群から1人ずつ無作為に選んだときの確率的な優位性を表す効果量

$$\delta = P(X > Y) - P(X < Y)$$

$\delta = 0$：両群は同程度

$\delta > 0$：群Aが大きくなりやすい

$\delta < 0$：群Bが大きくなりやすい

  - 事後分布の中央値：代表的な効果量
  - 信用区間：δが含まれる確率的範囲

δ は標準化された効果量

- 平均との差を標準偏差で割った量
  - 頻度流の Cohen’s d に対応
- ベイズでは「1つの値」ではなく
  分布（中央値＋信用区間）として表される

有意差がなさそうな場合

独立したサンプルのt検定

	t	df	p
ScoreB	0.377	98	.707

注 スチューデントのt-検定

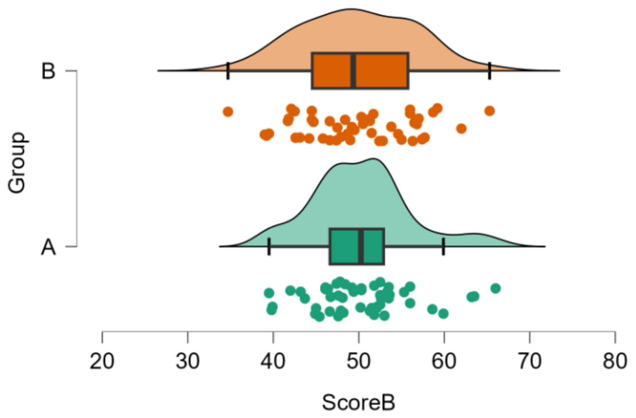
記述統計量

群の記述統計

	群	N	平均値	標準偏差	標準誤差	変動係数
ScoreB	A	50	50.28	5.759	0.814	0.115
	B	50	49.81	6.628	0.937	0.133

雨雲プロット

ScoreB



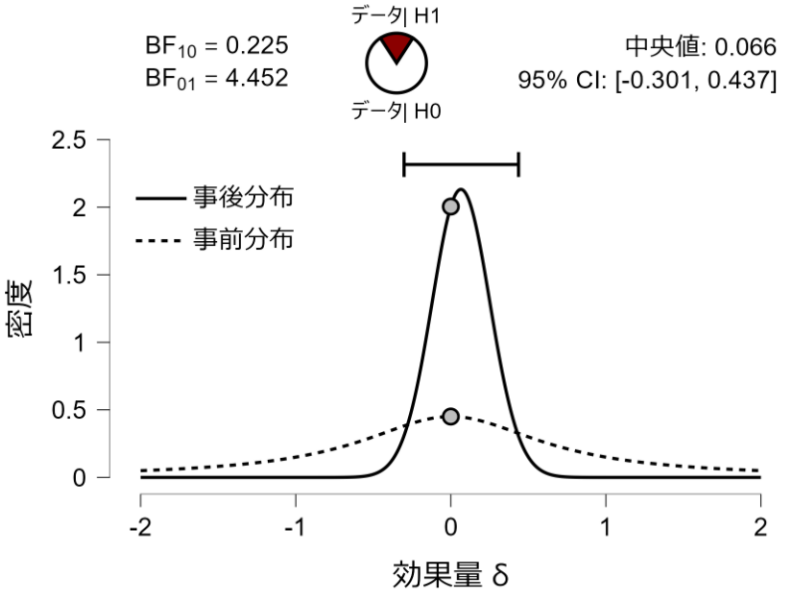
ベイジアン独立サンプルのt検定

	BF_{10}	エラー %
ScoreB	0.225	0.025

推論プロット ▼

ScoreB

事前分布と事後分布



微妙な場合

独立したサンプルのt検定

	t	df	p
ScoreC	−1.983	98	.050

注 スチューデントのt-検定

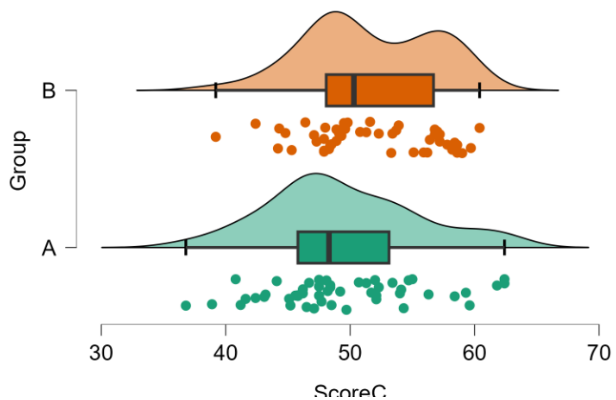
記述統計量

群の記述統計

	群	N	平均値	標準偏差	標準誤差	変動係数
ScoreC	A	50	49.42	6.009	0.850	0.122
	B	50	51.63	5.133	0.726	0.099

雨雲プロット

ScoreC



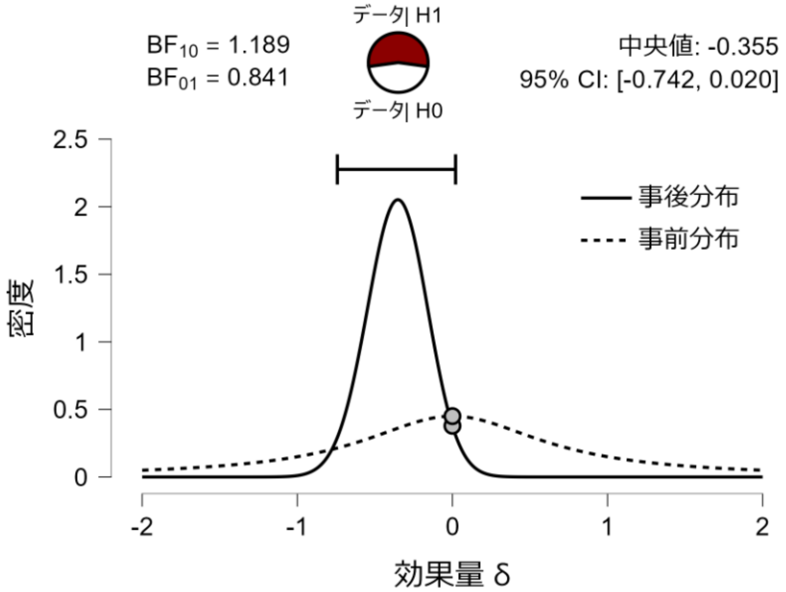
ベイジアン独立サンプルのt検定

	BF_{10}	エラー %
ScoreC	1.189	0.012

推論プロット

ScoreC

事前分布と事後分布



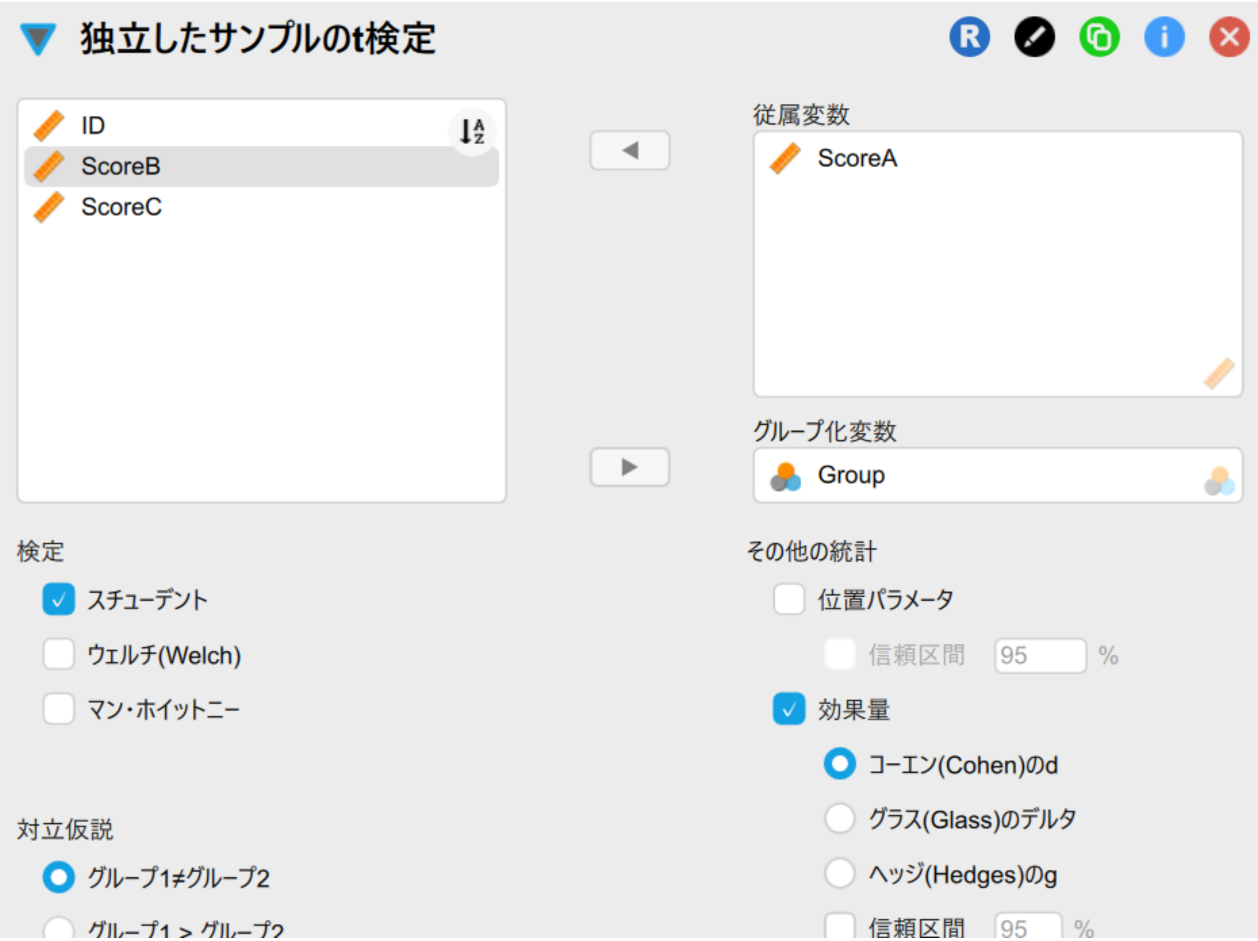
結果に必要なもの

- 頻度流
  t値　p値　df（効果量）

- ベイズ
  （t値）　BF　δ（効果量）の中央値と95%信頼区間

頻度流での効果量

- 頻度流でも効果量はある
- 1つの値として推定



独立したサンプルのt検定

	t	df	p	コーエン(Cohen)のd	SE コーエン(Cohen)のd
ScoreA	−4.214	98	< .001	−0.843	0.217
ScoreB	0.377	98	.707	0.075	0.200

注 スチューデントのt-検定

その後の声明(2019)

- 「$p < 0.05$」を閾値として使う慣習をやめよう
- Statistically significant” という言葉も避けよう
  - 代わりに:
    - 効果量(effect size)
    - 信頼区間(confidence interval)
    - 事前知識や理論的根拠
    - ベイズ推論再現性の重視
  - *Moving to a World Beyond ‘p < 0.05’* Wasserstein, Schirm & Lazar, *The American Statistician*, 2019

まとめ

- 統計は考え方の道具
- 頻度流とベイズは対立するものではなく考え方の違い
  - 珍しさの確率で表現した方がわかりやすいか、対立仮説の起こりやすさを比率で表現した方がいいか
- p値やBFは真理ではない
- ベイズもBFに閾値(例BF>3)を設けたら、頻度流のP値の閾値と一緒
- どちらのやり方でも、閾値ではなくて連続性として考える

まとめ

- 伝統的に頻度流が扱いやすかったので先行研究は多い
  - アプリでベイズを扱いやすくなってきたので流行った
  - 「効果のありなし」「発症するしない」といった場合、頻度流が扱いやすかった
- 確率で話した方がわかりやすいか、起きそうな割合で話した方がわかりやすいかはものによる
- 頻度流が無くなることはなく、状況に応じてどちらで説明するかを選択していけばよい

学習の参考に

新版　統計学のセンス
　デザインする視点・データを見る目
丹後俊郎　著　朝倉書店　2018
ISBN- 978-4-254-12882-6 C3341
(Kindle版あり)