平均値の条件差の t 検定の問題

シミュレーションによる若干の検討

岡本安晴
日本女子大学

データの平均値の比較は、最もよく行われる
統計分析の1つ（Derrick et al., 2016）

独立な２つのサンプル

Welch’s pはどこまで使えるか？

対応のある２つのサンプル

t 検定での取り残し情報

Derrick, B., Toher, D. and White, P. (2016) Why Welch’s test is type I error robust. *The Quantitative Methods for Psychology*, 12, 30-38.

２つのデータの比較

Studentの t 検定がデフォルト（Delacre et al., 2017）

Studentの t 検定は p 値に集約される。
p 値は今後も使われ続けると予想されている（Benjamin et al., 2019）

p 値にはよく知られた批判がある（小杉、2019）
p 値は約100年にわたる問題である（Ruberg, 2021）

p 値は有用な統計的指標でありうる（ASA Statement, 2016）

p 値の算出方法の問題

Delacre, M., Lakens, D. and Leys, C. (2017) *International Review of Social Psychology*, 30, 92-101
Benjamin, D. J. and Berger, J. O. (2019) The American Statistician, 73, 186-191.
小杉孝司（2019）社会と調査、23、92－98。
Ruberg, S. J. (2021). Clinical Pharmacology & Therapeutics, 109, 1489-1498.
Wasserstein, R. L., and Lazar, N. A. (2016) The ASA statement on p-values.
The American Statistician, 70, 129-133.

２つの独立なサンプルの平均値

Studentの p 値　　仮定：正規分布、等分散

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{s^2_{\bar{X}-\bar{Y}}}} \qquad s^2_{\bar{X}-\bar{Y}} = \frac{1}{M}s^2 + \frac{1}{N}s^2 \qquad s^2 = \frac{\sum(X_i - \bar{X})^2 + \sum\left(Y_j - \bar{Y}\right)^2}{M + N - 2}$$

$$df = M + N - 2$$

Welchの p 値　　仮定：正規分布

$$t' = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s^2_X}{M} + \frac{s^2_Y}{N}}} \qquad s^2_x = \frac{\sum(X_i - \bar{X})^2}{M - 1} \qquad s^2_Y = \frac{\sum\left(Y_j - \bar{Y}\right)^2}{N - 1}$$

$$df' = \left(\frac{s^2_X}{M} + \frac{s^2_Y}{N}\right)^2 \Big/ \left[\frac{s^4_X}{M^2(M-1)} + \frac{s^4_Y}{N^2(N-1)}\right]$$

p 値（Wasserstein & Lazar, 2016）

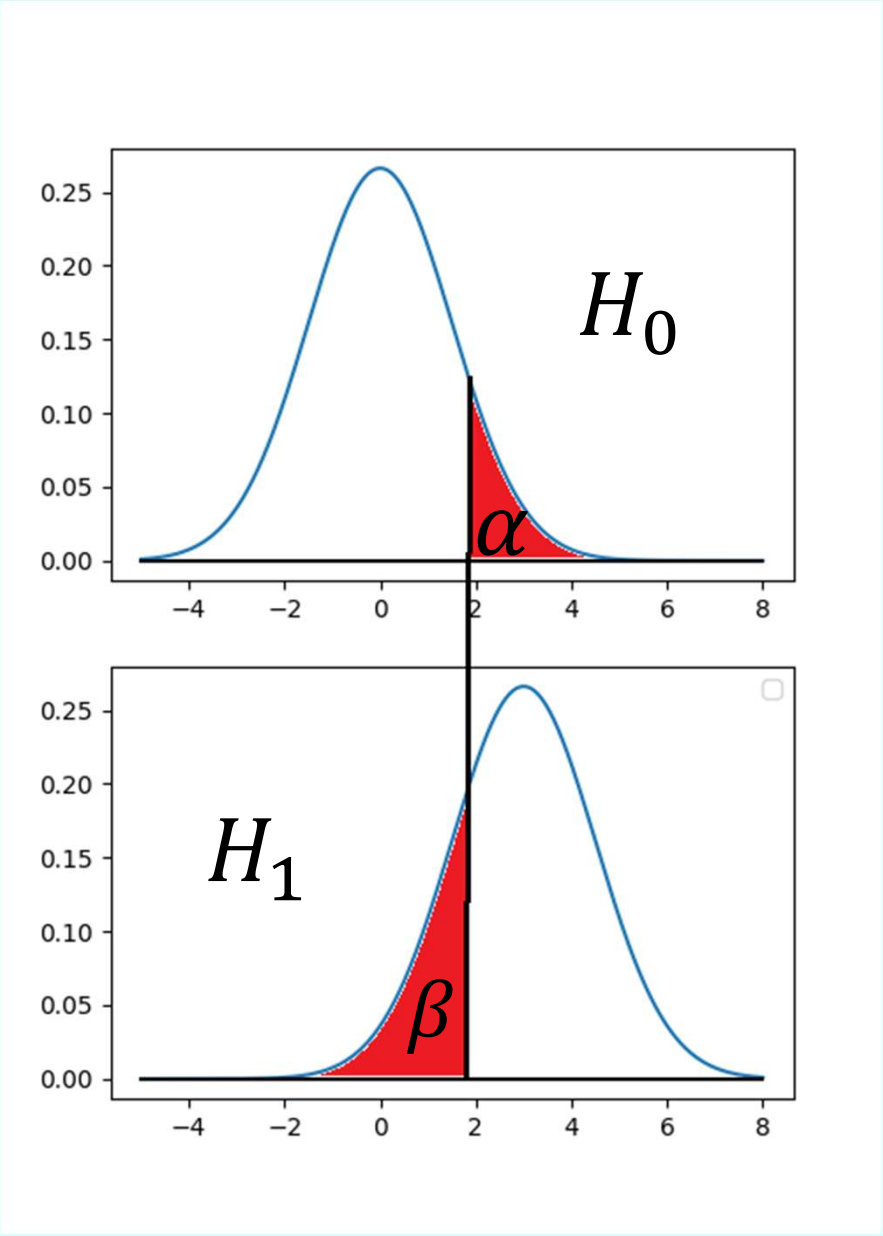
統計量がデータによって与えられる値以上の極端な値を取る確率

$$(1) \qquad P(\text{p値} \leq u) = u$$

算出された p 値の特性を式（１）により検討する。

Wasserstein, R. L., and Lazar, N. A. (2016) The ASA statement on p-values. *The American Statistician*, 70, 129-133.

有意水準の期待効用



$P(H_0)$：　帰無仮説が正しい確率

α：　第１種の誤りの確率（有意水準）

$U_{0,Sig}$：　帰無仮説が正しいときに、効果有りと判断することの効用

$U_{0,NSig}$：　帰無仮説が正しいときに、効果無しと判断することの効用

$P(H_1)$：　対立仮設が正しい確率

β：　第２種の誤りの確率　　$1-\beta$：　検定力

$U_{1,Sig}$：　対立仮説が正しいときに、効果有りと判断することの効用

$U_{1,NSig}$：　対立仮説が正しいときに、効果無しと判断することの効用

期待効用

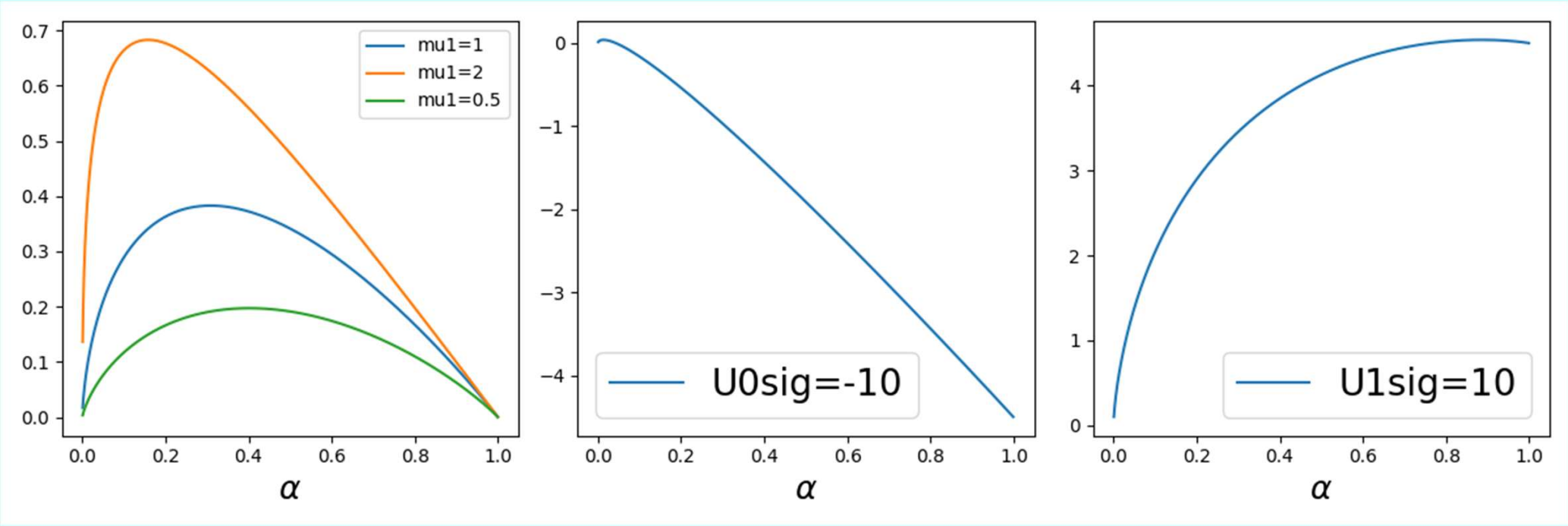
$$EU(\alpha) = P(H_0)\left[\alpha U_{0,Sig} + (1-\alpha)U_{0,NSig}\right] + P(H_1)\left[(1-\beta)U_{1,Sig} + \beta U_{1,NSig}\right]$$

p値 $\leq \alpha$：　有意差あり　　p値 $> \alpha$：　有意差なし

$$P(H_0) = P(H_1) = 0.5$$

$U_{0.Sig} = -1$　　$U_{0.NSig} = 1$

$U_{1.Sig} = 1$　　$U_{1.Nig} = -1$


mu1=1
mu1=2
mu1=0.5
α
U0sig=-10
α
U1sig=10
α

p 値（Wasserstein & Lazar, 2016）

統計量がデータによって与えられる値以上の極端な値を取る確率

$$(1) \qquad P(\text{p値} \leq u) = u$$

算出された p 値の妥当性を式（１）により検討する。

Delacreら（2017）は、正規分布、非対称正規分布、二重指数分布、カイ二乗分布、一様分布の場合のp値を調べている。

本報告では、正規分布、 t 分布、ポアッソン分布、対数正規分布、順序カテゴリ値の分布（多項分布）の場合を調べる。

Wasserstein, R. L., and Lazar, N. A. (2016) The ASA statement on p-values. *The American Statistician*, 70, 129-133.

Delacre, M., Lakens, D. and Leys, C. (2017) *International Review of Social Psychology*, 30, 92-101

正規分布の場合

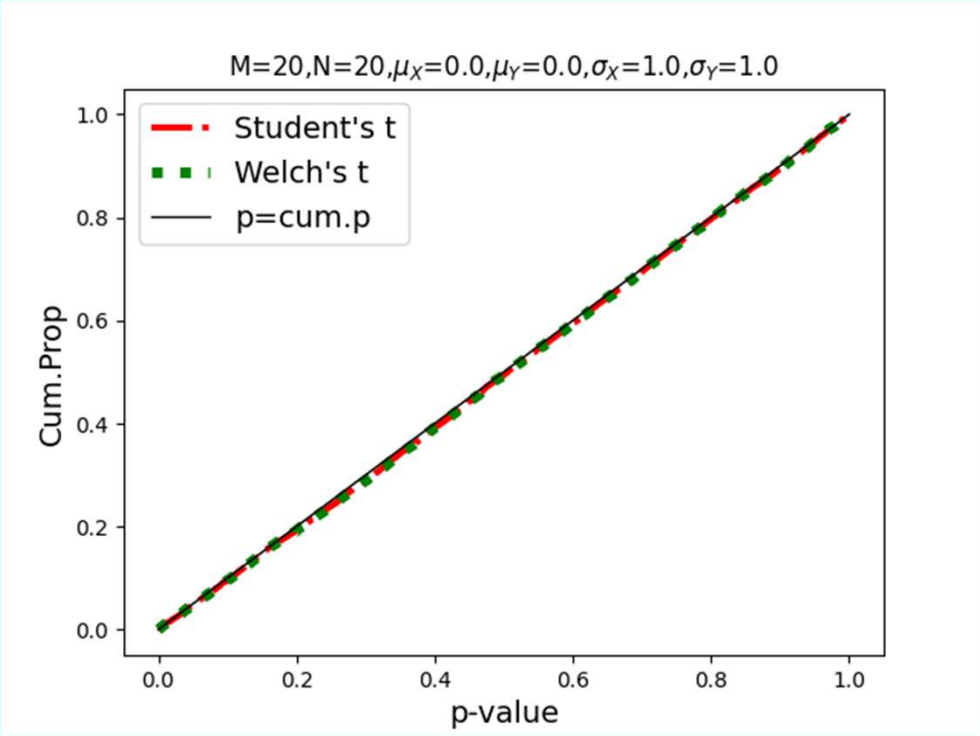
正規分布は心理学においてデフォルトとして用いられている分布（椎名、2013）

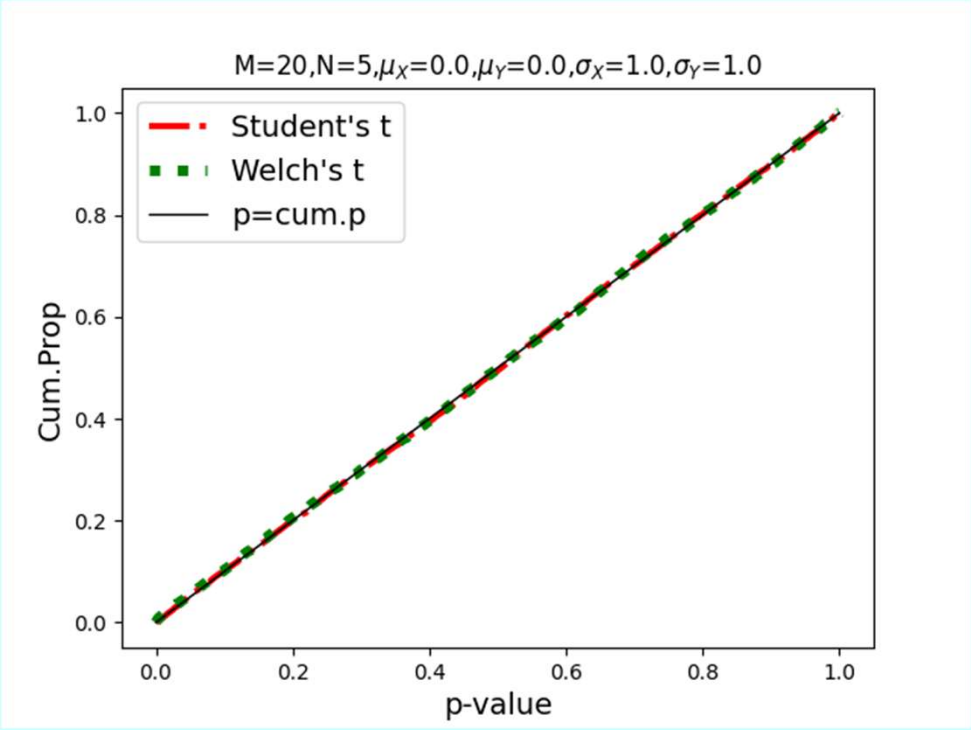
$$p(\theta|\mu,\sigma^2)=\frac{1}{\sqrt{2\pi}\cdot\sigma}exp\left(-\frac{1}{2\sigma^2}(\theta-\mu)^2\right),\qquad E(\theta)=\mu,\qquad var(\theta)=\sigma^2$$

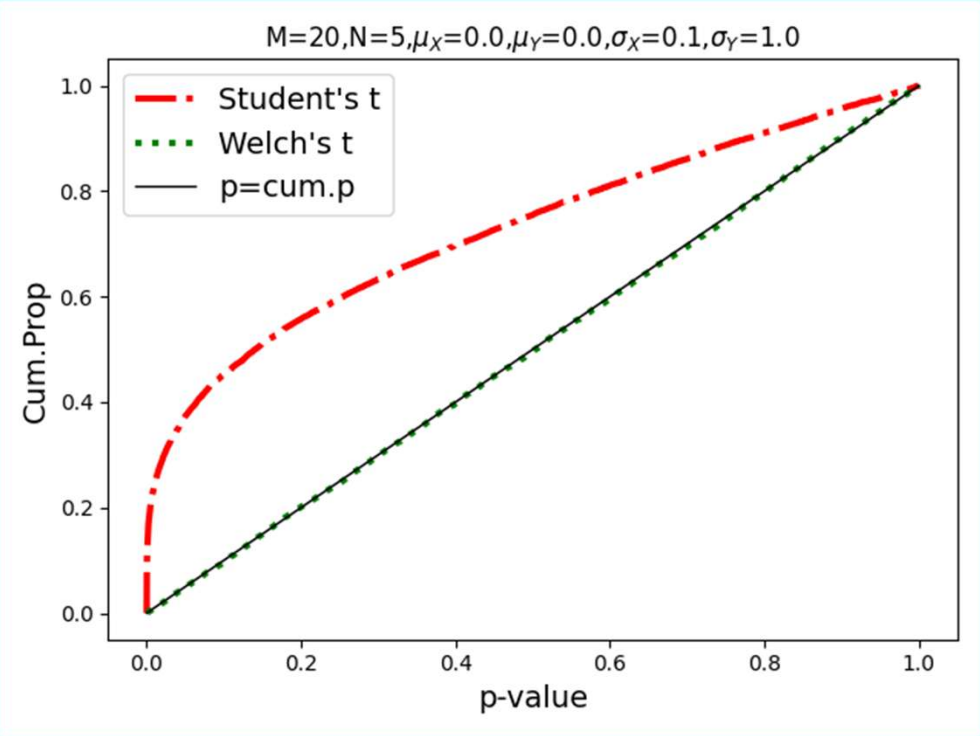
Studentのt値とdf、およびWelchのt’値とdf’は、
データの原点および単位の影響を受けない。

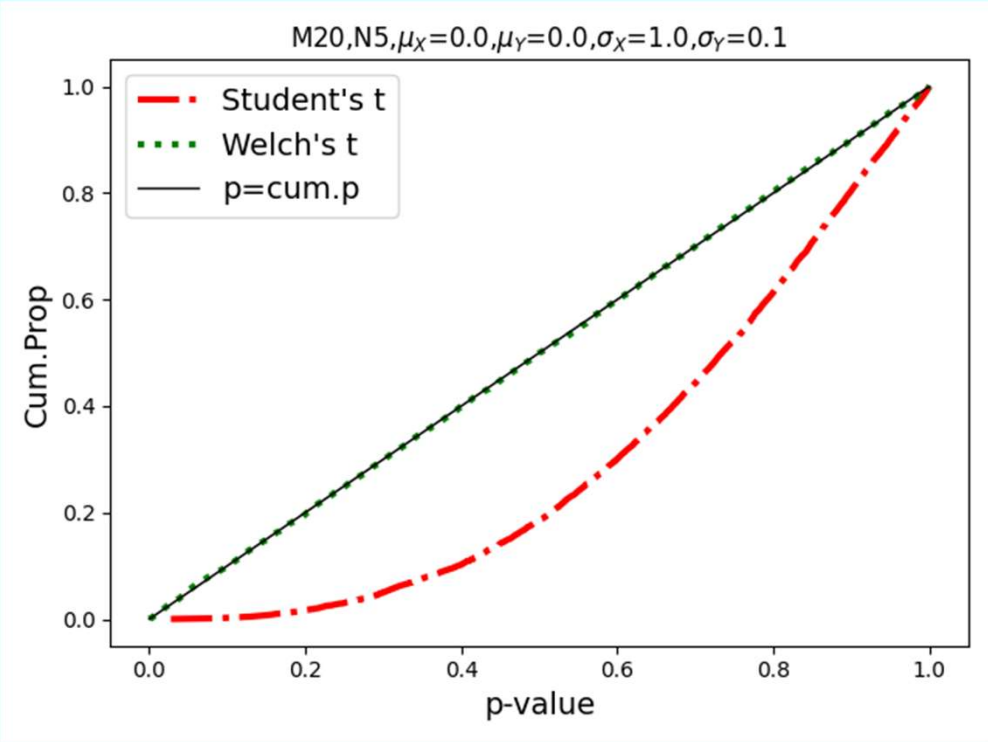
$\mu=0,\sigma=1$として考える

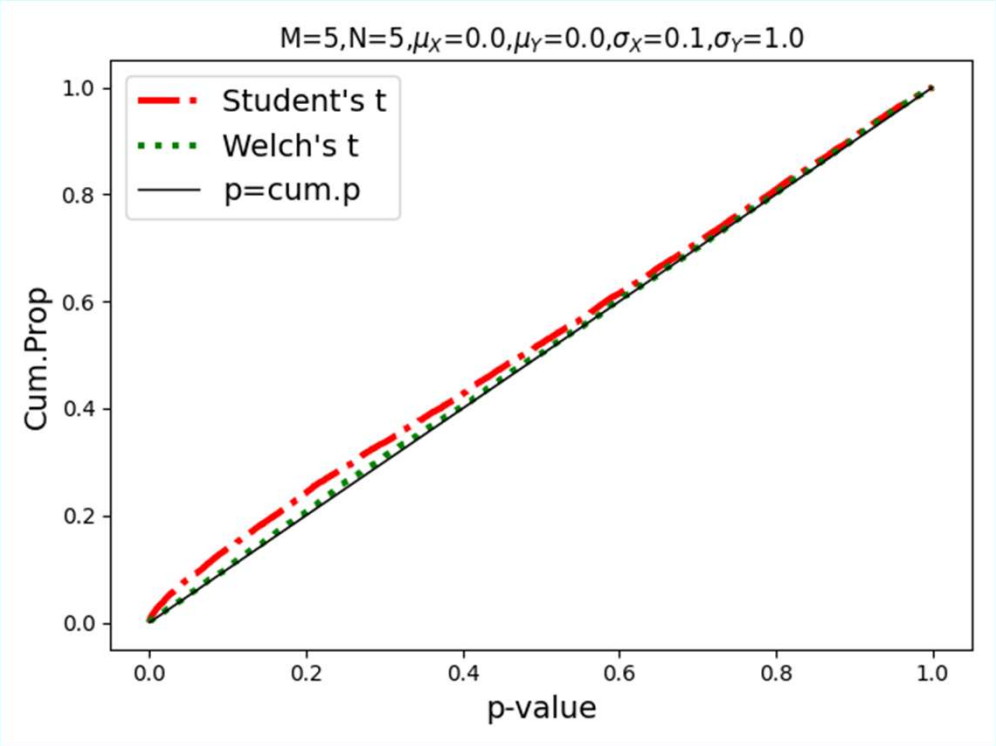
椎名乾平（2013）心理学評論、56、7－34.

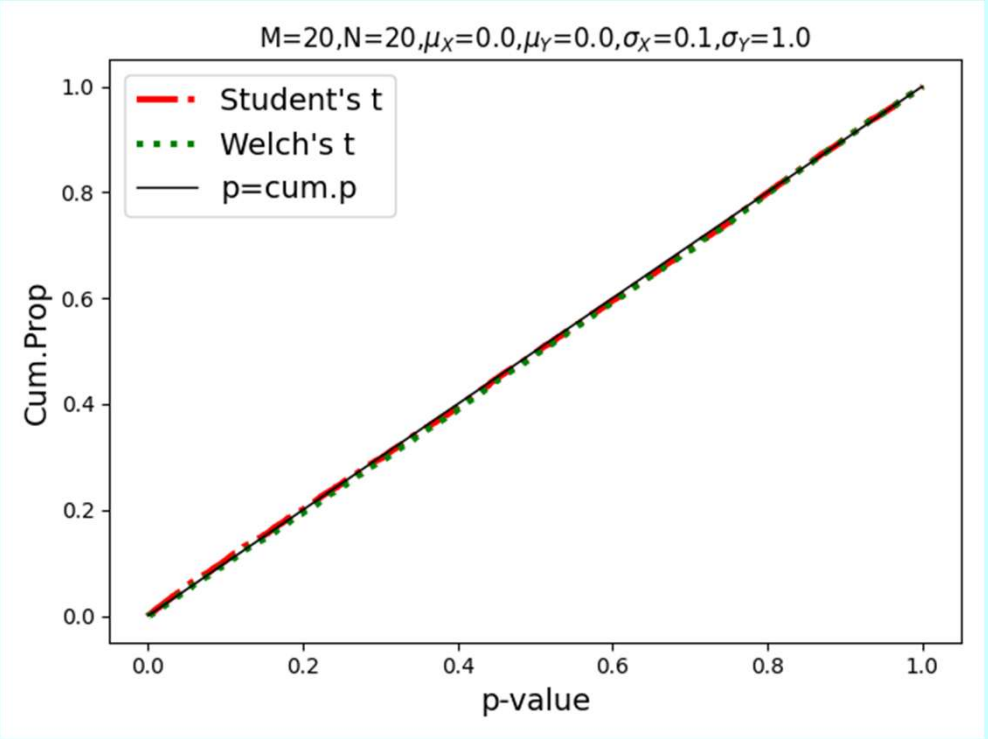

M=20,N=20,μ_X=0.0,μ_Y=0.0,σ_X=1.0,σ_Y=1.0
Student's t
Welch's t
p=cum.p
Cum.Prop
p-value
0.0
0.2
0.4
0.6
0.8
1.0


M=20,N=5,μ_X=0.0,μ_Y=0.0,σ_X=1.0,σ_Y=1.0
Student's t
Welch's t
p=cum.p
Cum.Prop
p-value
0.0
0.2
0.4
0.6
0.8
1.0


M=20,N=5,μ_X=0.0,μ_Y=0.0,σ_X=0.1,σ_Y=1.0
Student's t
Welch's t
p=cum.p
Cum.Prop
p-value
0.0
0.2
0.4
0.6
0.8
1.0


M20,N5,μ_X=0.0,μ_Y=0.0,σ_X=1.0,σ_Y=0.1
Student's t
Welch's t
p=cum.p
Cum.Prop
p-value
0.0
0.2
0.4
0.6
0.8
1.0


M=5,N=5,μ_X=0.0,μ_Y=0.0,σ_X=0.1,σ_Y=1.0
Student's t
Welch's t
p=cum.p
Cum.Prop
p-value
0.0
0.2
0.4
0.6
0.8
1.0


M=20,N=20,μ_X=0.0,μ_Y=0.0,σ_X=0.1,σ_Y=1.0
Student's t
Welch's t
p=cum.p
Cum.Prop
p-value
0.0
0.2
0.4
0.6
0.8
1.0

t 分布

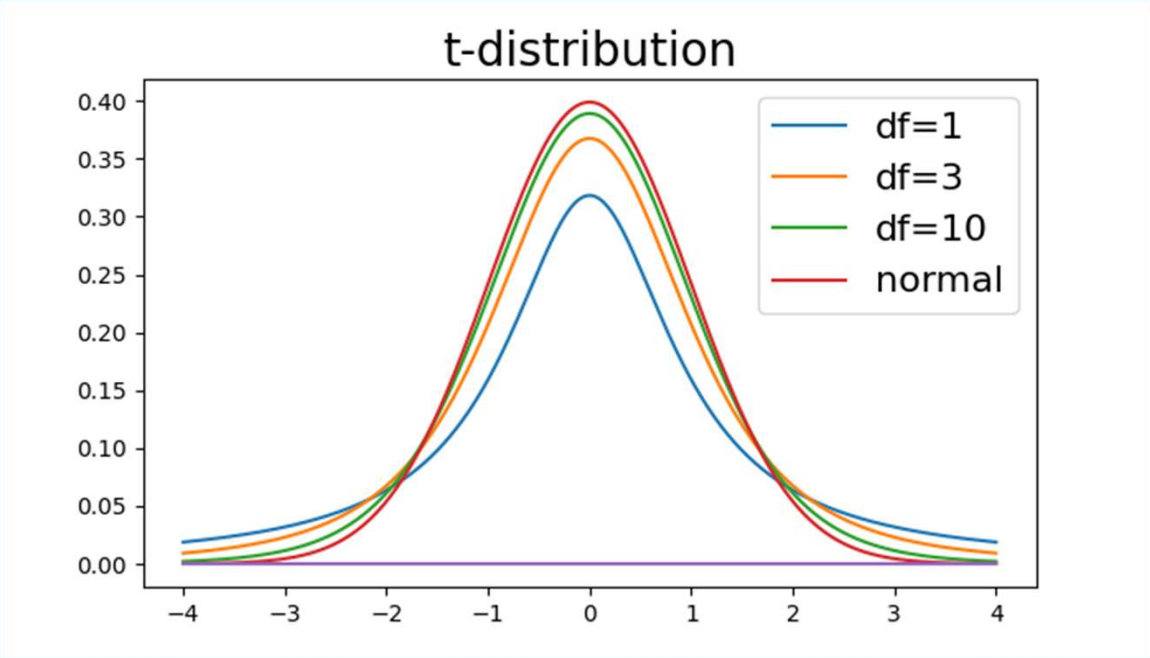
正規分布の観点からは外れ値と見做される値も、他の裾の広い分布による分析を行うときは外れ値として扱う必要がない場合がある（Gelman et al., 2014; Martin, 2018）。

$$p(\theta|\nu,\mu,\theta) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\nu\pi}\sigma}\left(1+\frac{1}{\nu}\left(\frac{\theta-\mu}{\sigma}\right)^2\right)^{-(\nu+1)/2}$$

$$\nu > 0$$

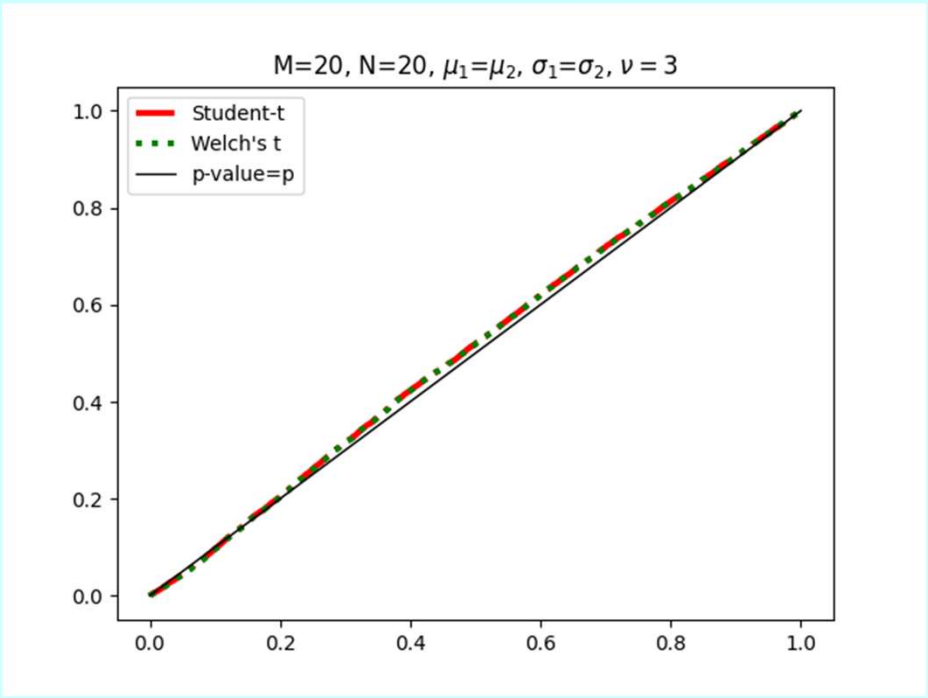
$$E(\theta) = \mu, \quad for\ \nu > 1$$

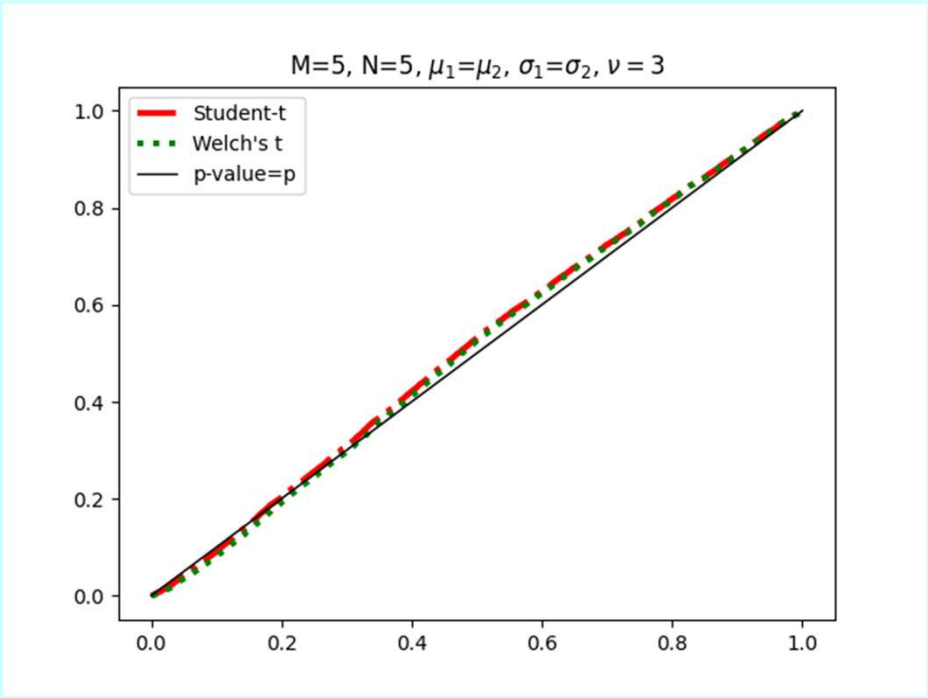
$$var(\theta) = \frac{\nu}{\nu-2}\sigma^2, \quad for\ \nu > 2$$

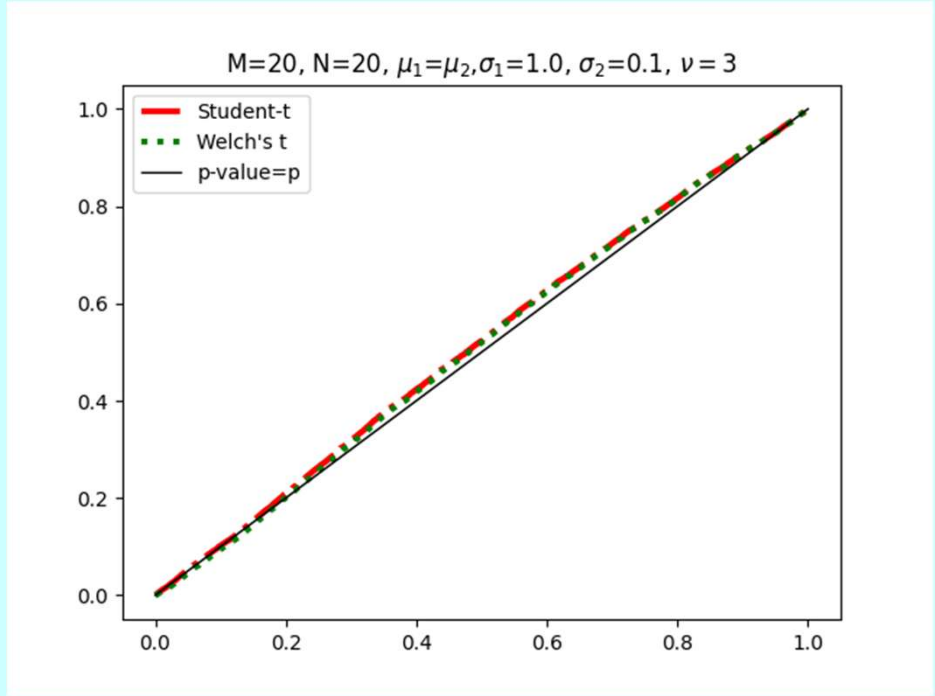


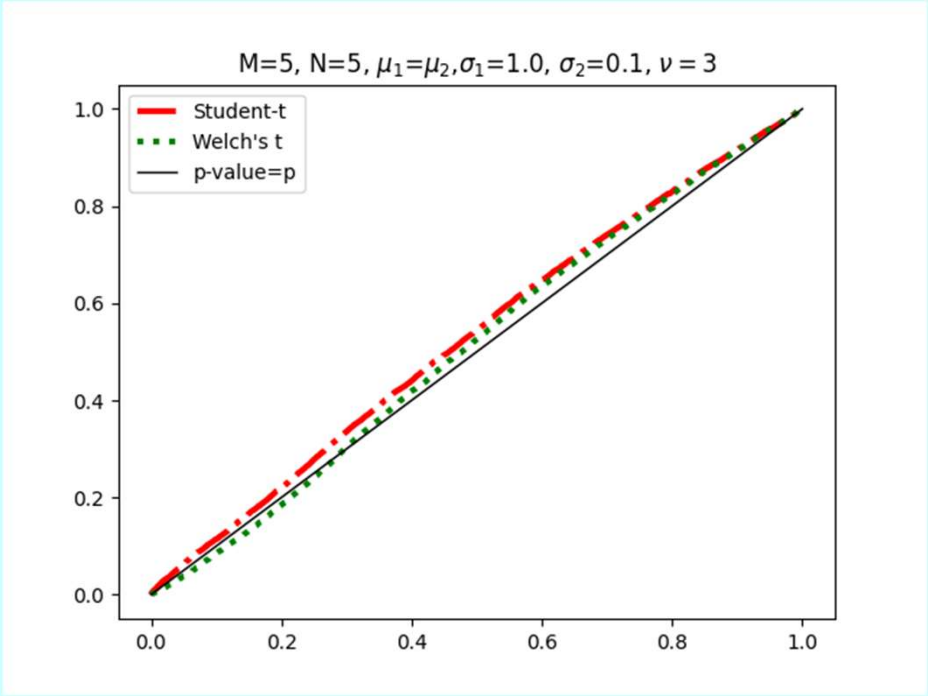
Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D> B., Vehtari, A, & Rubin, D. B. (2014). *Bayesian data analysis, 3rd ed*. CRC Press.

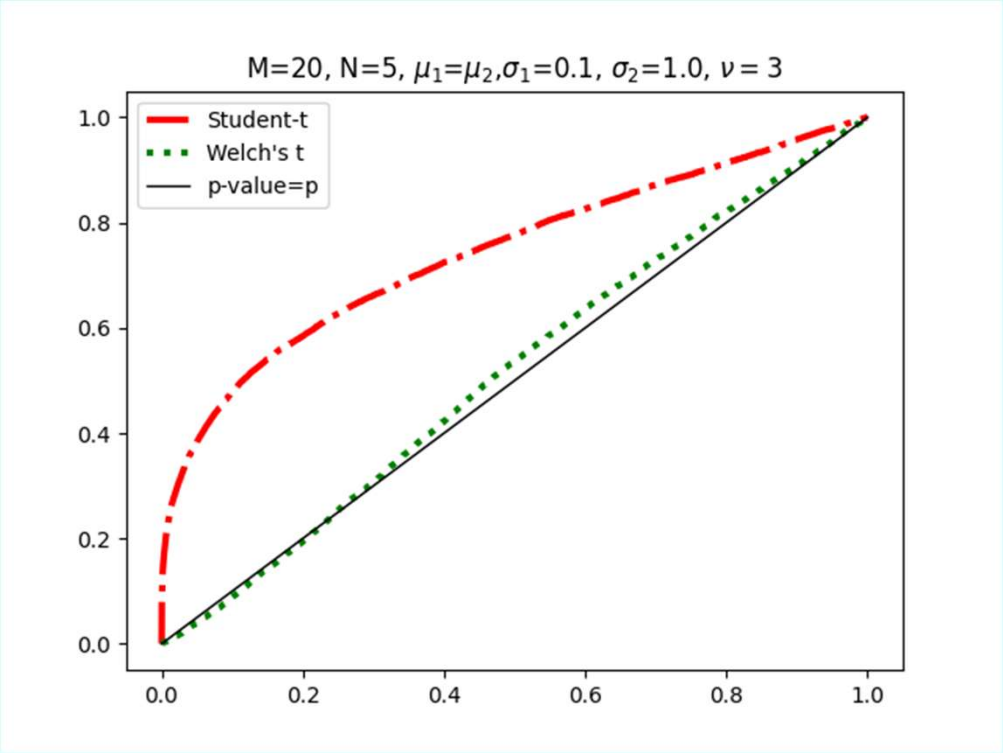
Martin, O. (2018) *Bayesian analysis with Python.* Packt.

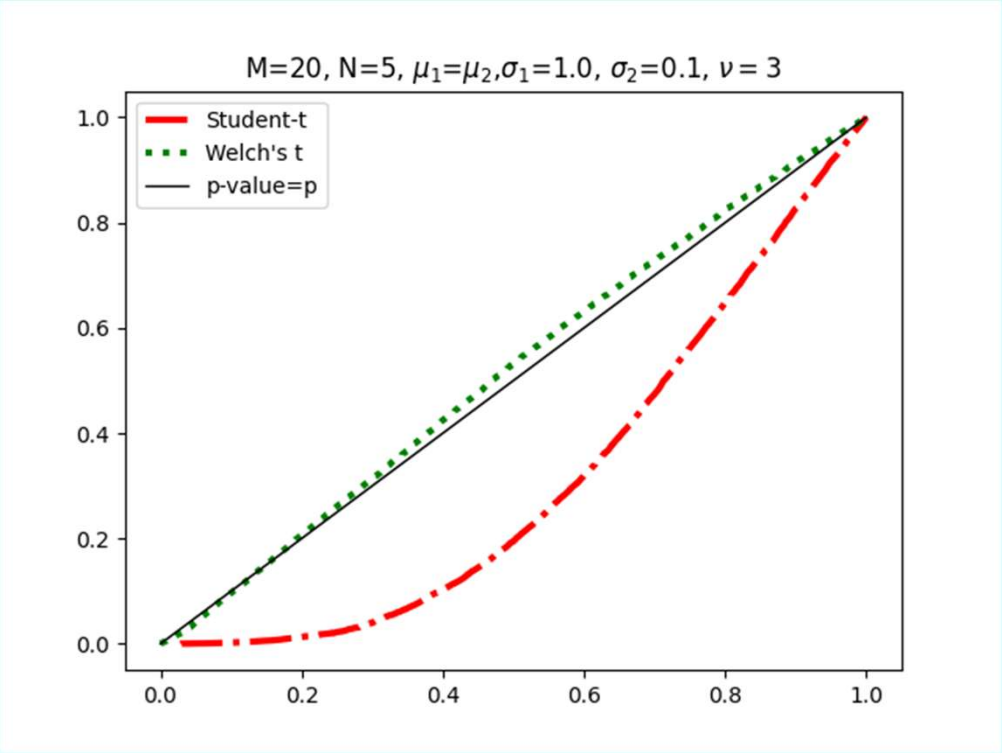

M=20, N=20, $\mu_1=\mu_2$, $\sigma_1=\sigma_2$, $\nu=3$
Student-t
Welch's t
p-value=p
0.0
0.2
0.4
0.6
0.8
1.0

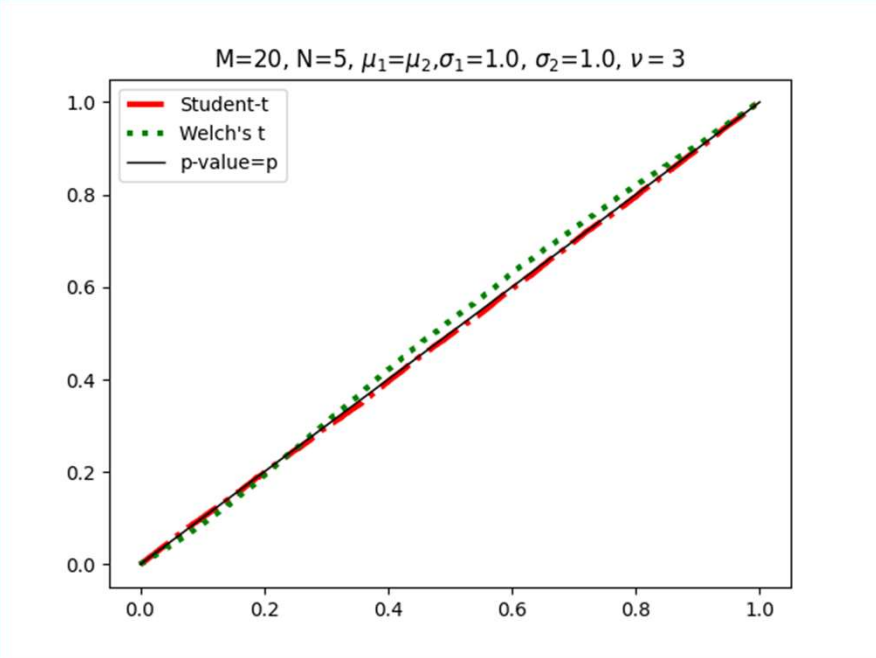

M=5, N=5, $\mu_1=\mu_2$, $\sigma_1=\sigma_2$, $\nu=3$
Student-t
Welch's t
p-value=p
0.0
0.2
0.4
0.6
0.8
1.0


M=20, N=20, $\mu_1=\mu_2$,$\sigma_1=1.0$, $\sigma_2=0.1$, $\nu=3$
Student-t
Welch's t
p-value=p
0.0
0.2
0.4
0.6
0.8
1.0


M=5, N=5, $\mu_1=\mu_2$,$\sigma_1=1.0$, $\sigma_2=0.1$, $\nu=3$
Student-t
Welch's t
p-value=p
0.0
0.2
0.4
0.6
0.8
1.0


M=20, N=5, μ_1=μ_2,σ_1=0.1, σ_2=1.0, ν = 3
Student-t
Welch's t
p-value=p
0.0
0.2
0.4
0.6
0.8
1.0

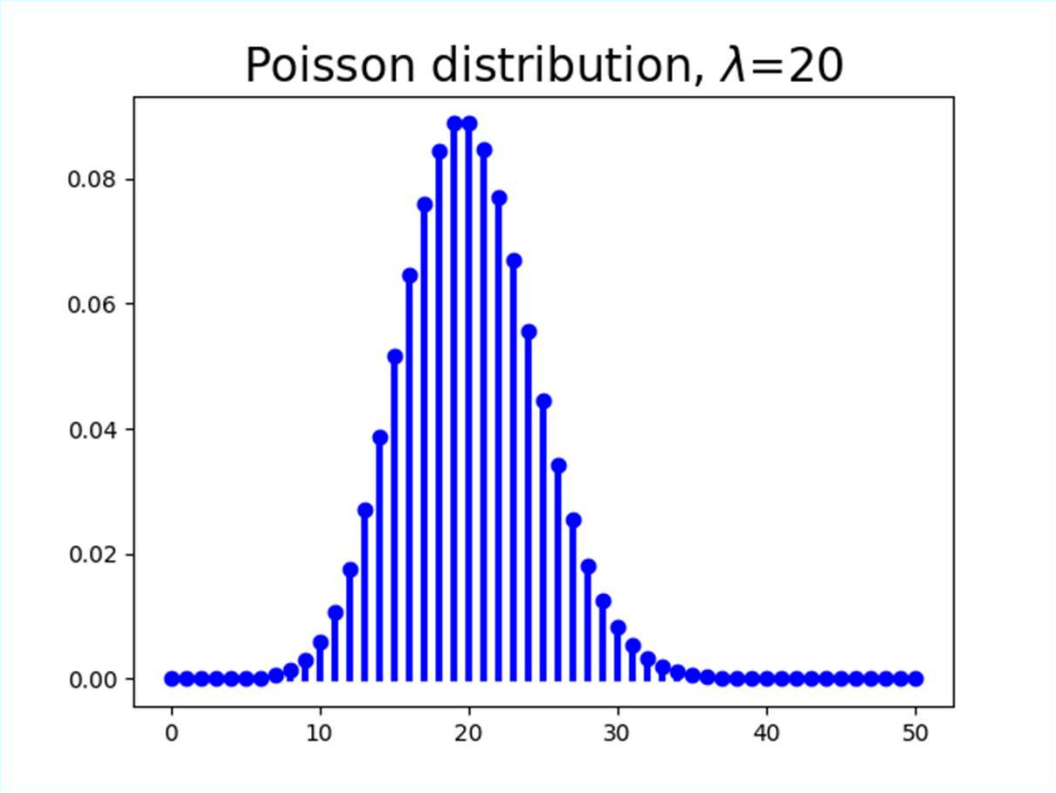

M=20, N=5, μ_1=μ_2,σ_1=1.0, σ_2=0.1, ν = 3
Student-t
Welch's t
p-value=p
0.0
0.2
0.4
0.6
0.8
1.0


M=20, N=5, μ_1=μ_2,σ_1=1.0, σ_2=1.0, ν = 3
Student-t
Welch's t
p-value=p
0.0
0.2
0.4
0.6
0.8
1.0

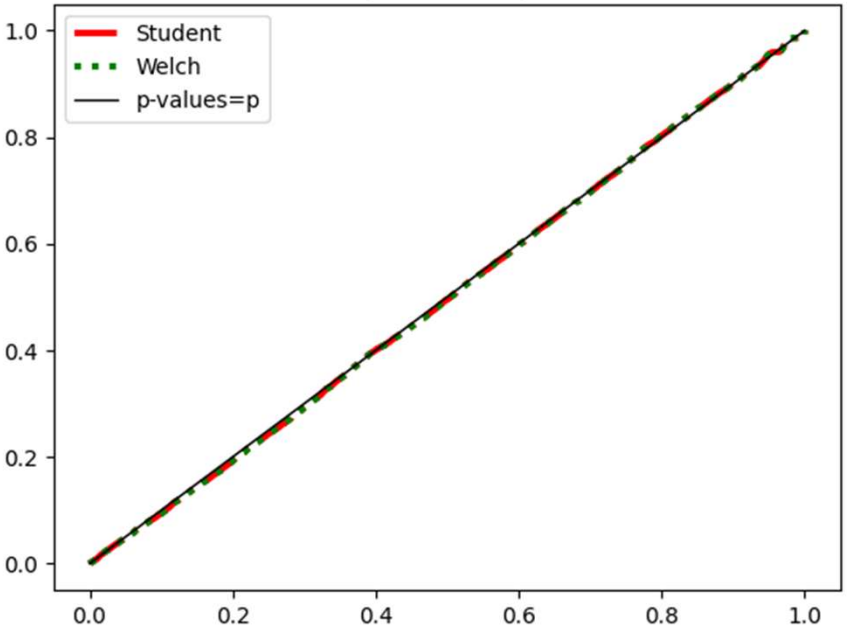
ポアッソン分布

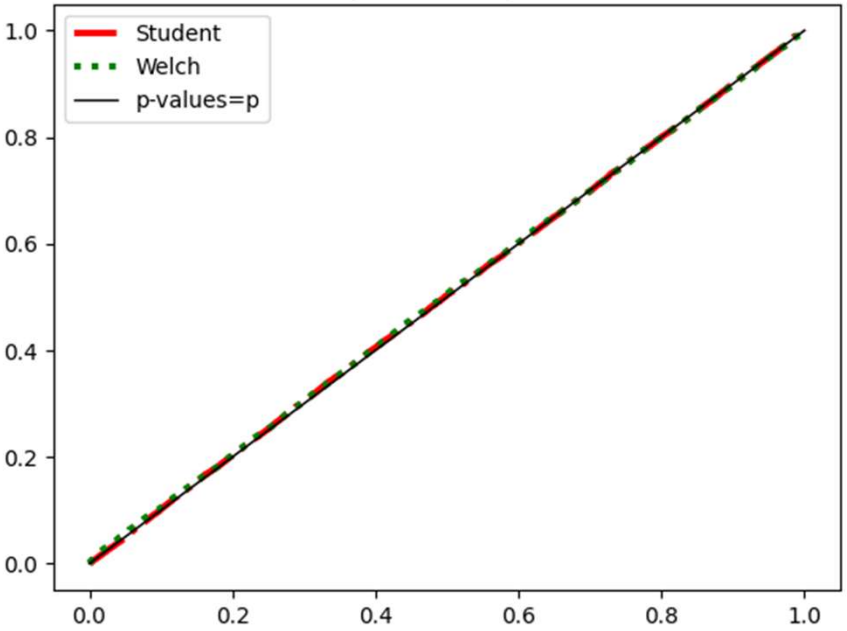
度数データの分析にポアッソン分布が使われることがある（Spieglhalter, 2019）

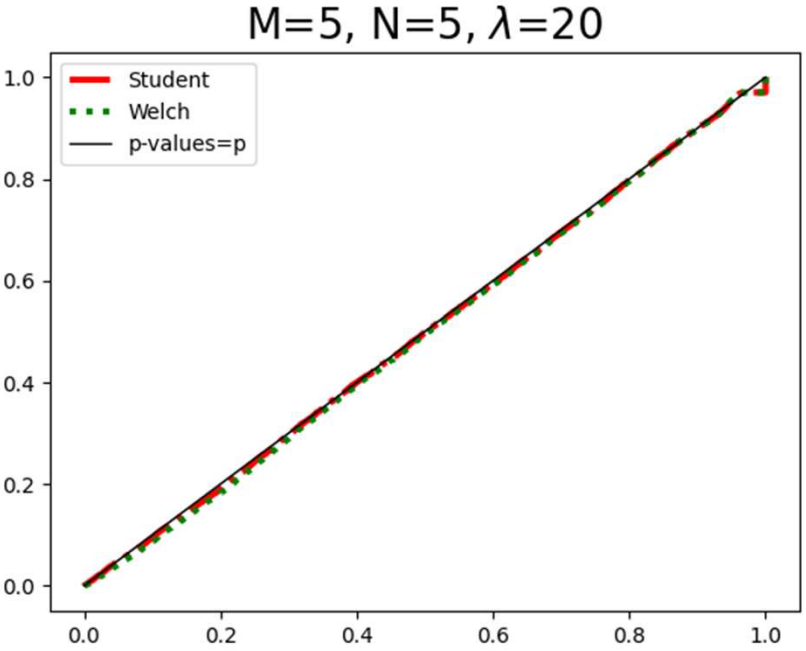
$$p(\theta|\lambda) = \frac{1}{\theta!}\lambda^{\theta} exp(-\lambda), \qquad \theta = 0,1,2,\ldots \qquad E(\theta) = \lambda, \qquad var(\theta) = \lambda$$



Spiegelhalter, D. (2019). The art of statistics: Learning from data. Pelican


M=20, N=20, λ=20
Student
Welch
p-values=p
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0


M=5, N=20, λ=20
Student
Welch
p-values=p
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0


M=5, N=5, λ=20
Student
Welch
p-values=p
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

対数正規分布

対数正規分布は、恒星の密度、鉱物の塊粒の寸法・重量、鉱床の規模、音の強さ、材料の寿命、生物群の個体数、生物個体の寸法・重量、市町村の人口、個人所得額など、天文、地質、物理、生物、社会の各現象に広く見られる統計分布であり、石油・天然ガス鉱床においても、層厚、孔隙率｛こうげきりつ｝、埋蔵量の値が対数正規分布をすることが認められている。天文、地質、物理、生物、社会の各現象に広く見られる統計分布であると説明されている（エネルギー・金属鉱物資源機構ホームページ、2025.04.21）。

対数正規分布は，これまで紹介した自然現象のみならず，社会現象においてもしばしば観測される．例えば，研究所内で発表された論文の発表数が対数正規分布に従うこと）や，いくつかの国の比例代表選挙において，候補者の得票数を，政党が名簿に記載した候補者数や各政党の得票数などでスケールした量の分布が，ユニバーサルな対数正規分布に従う36｝など，人間の生産活動や社会行動にも対数正規性がよく現れることが報告されている（國仲・小林・松下、日本物理学会誌、2011、658－665）。

$X_1 + X_2 + \cdots + X_n \rightarrow$ 正規分布

$Y_1 Y_2 \cdots Y_n \rightarrow$ 対数正規分布　　$logY_1 + logY_2 + \cdots + logY_n \rightarrow$ 正規分布

$$p(\theta|\mu,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma\theta}exp\left(-\frac{1}{2\sigma^2}(log\theta-\mu)^2\right)$$

すなわち、$log\theta$は平均μ、分散σ^2の正規分布に従う。ここで、

$$E(\theta)=exp\left(\mu+\frac{1}{2}\sigma^2\right),\qquad var(\theta)=exp(2\mu+\sigma^2)(exp(\sigma^2)-1)$$

いま

$$mn=exp\left(\mu+\frac{1}{2}\sigma^2\right),\qquad sd^2=exp(2\mu+\sigma^2)(exp(\sigma^2)-1)$$

とおけば、

$$p(\theta|mn,sd)=p(\theta|\mu,\sigma^2)$$

$$\mu=log(mn)-\frac{1}{2}log\left((\mathrm{sd}^2mn^{-2})+1\right),\qquad \sigma^2=log\left((\mathrm{sd}^2mn^{-2})+1\right)$$

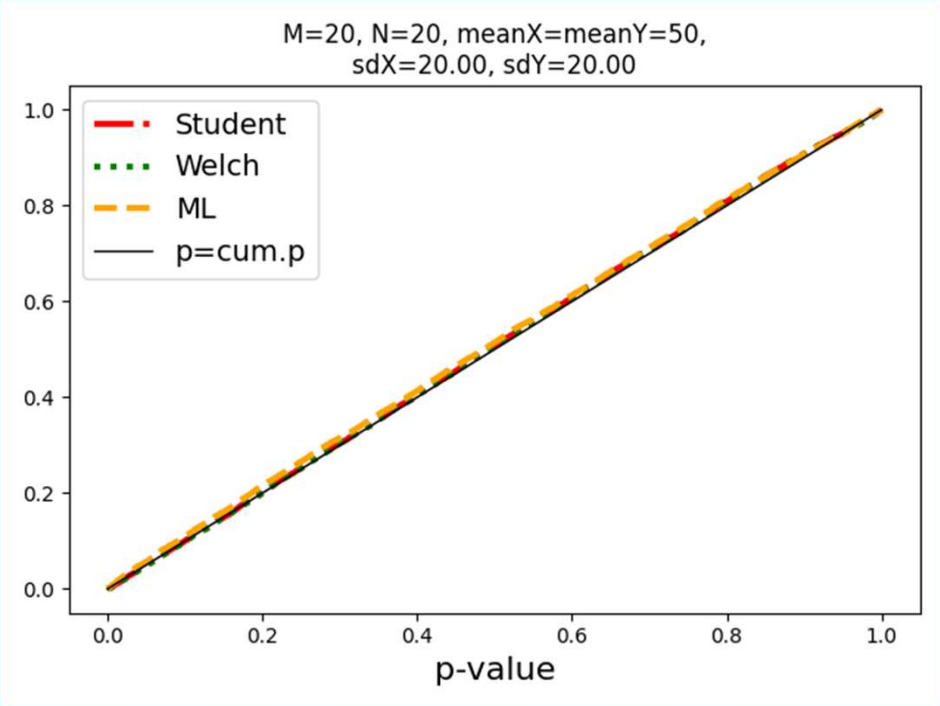
である。

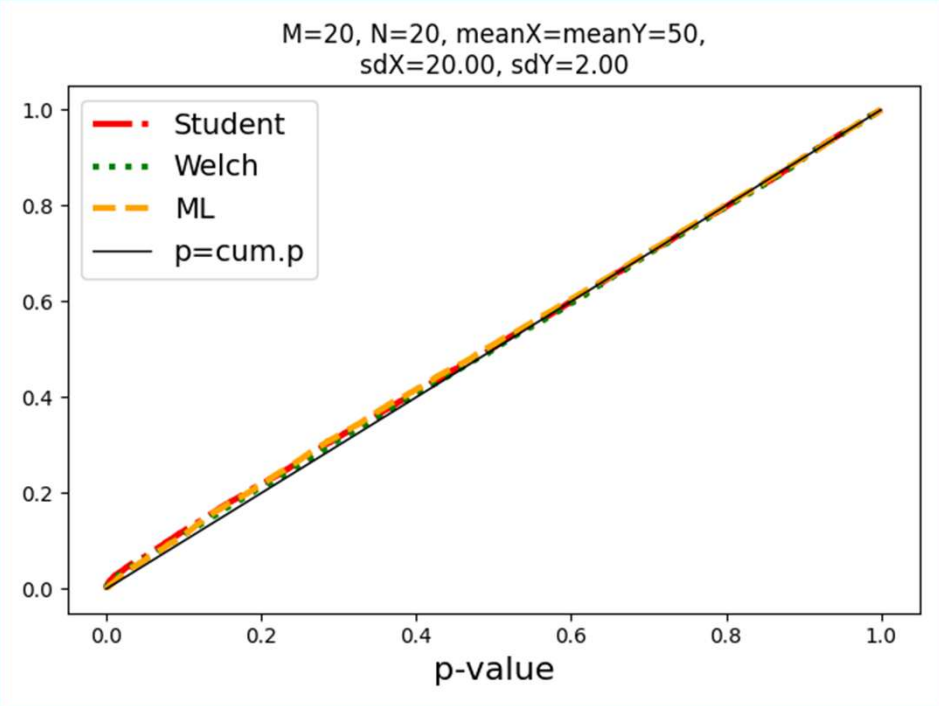
最尤法

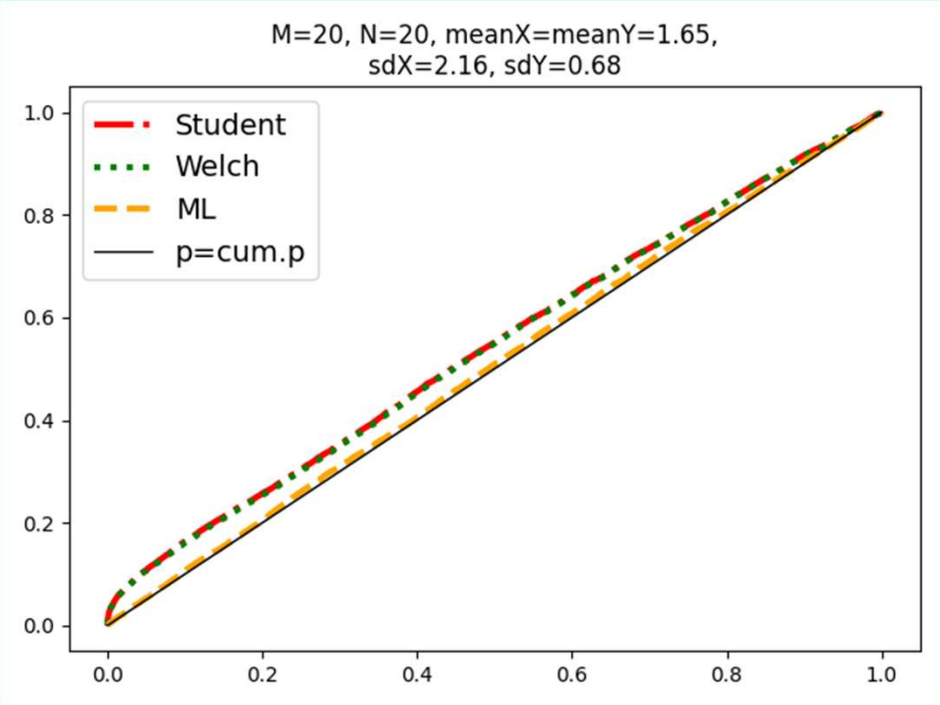
L_0：平均値が等しいという制約条件での最大尤度

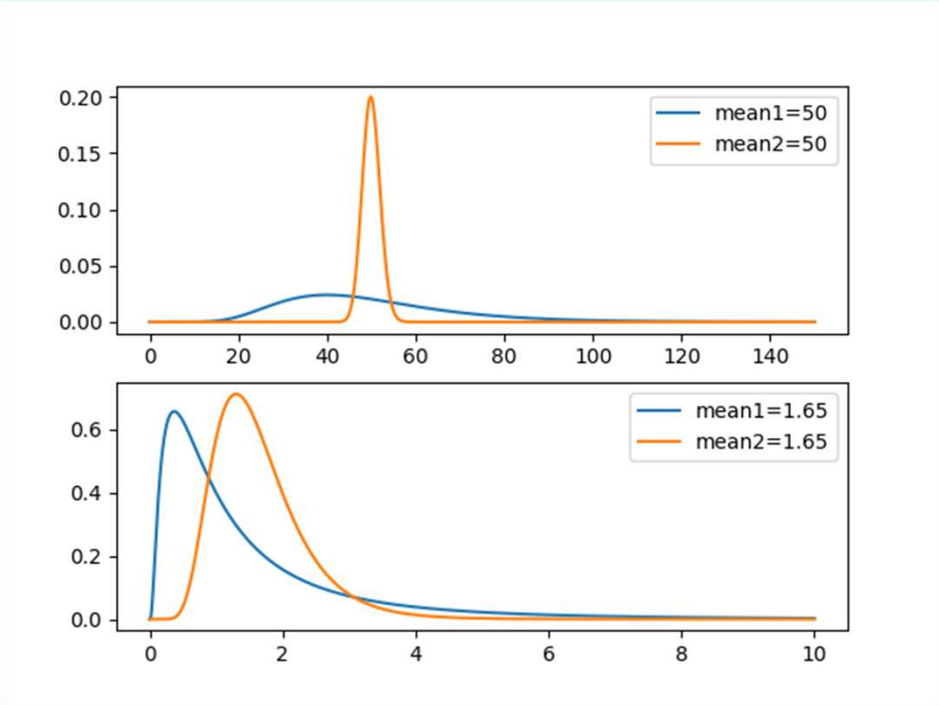
L_1：平均値に制約がない条件での最大尤度

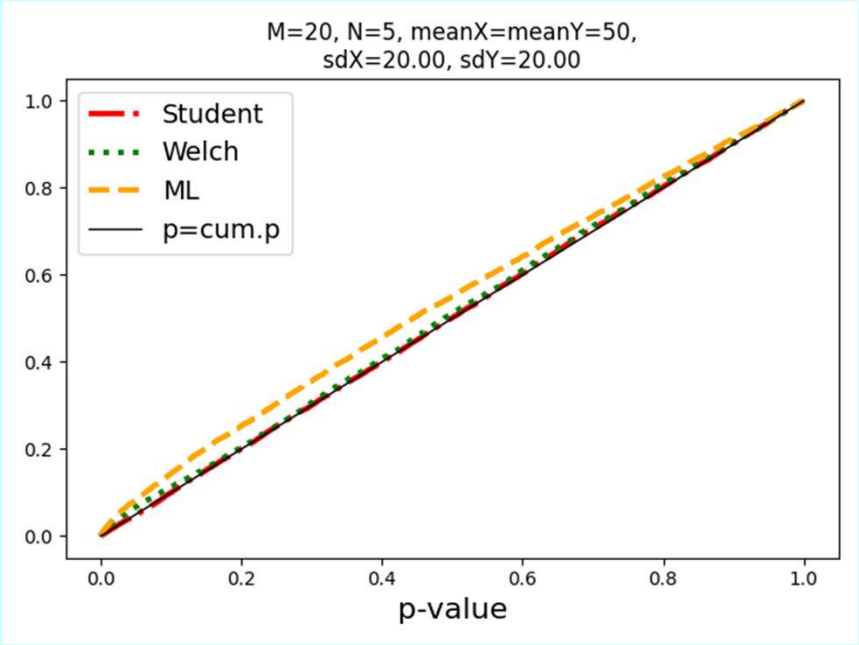
$$-2log\frac{L_0}{L_1} \quad \rightarrow \quad \chi^2_{df} \qquad\qquad df = dim_1 - dim_0$$

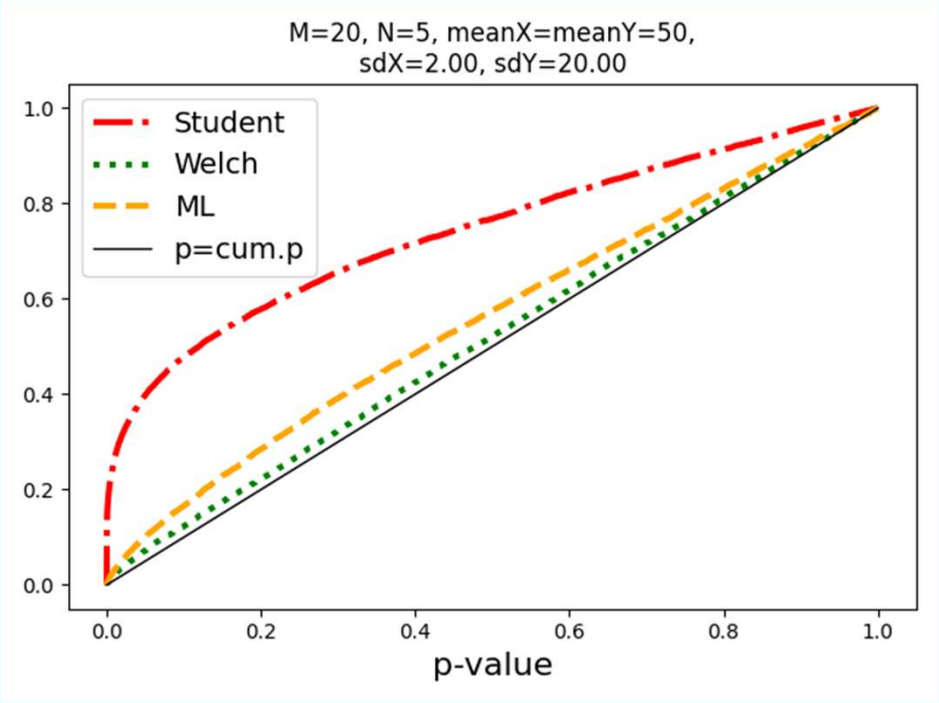

M=20, N=20, meanX=meanY=50,
sdX=20.00, sdY=20.00
Student
Welch
ML
p=cum.p
p-value

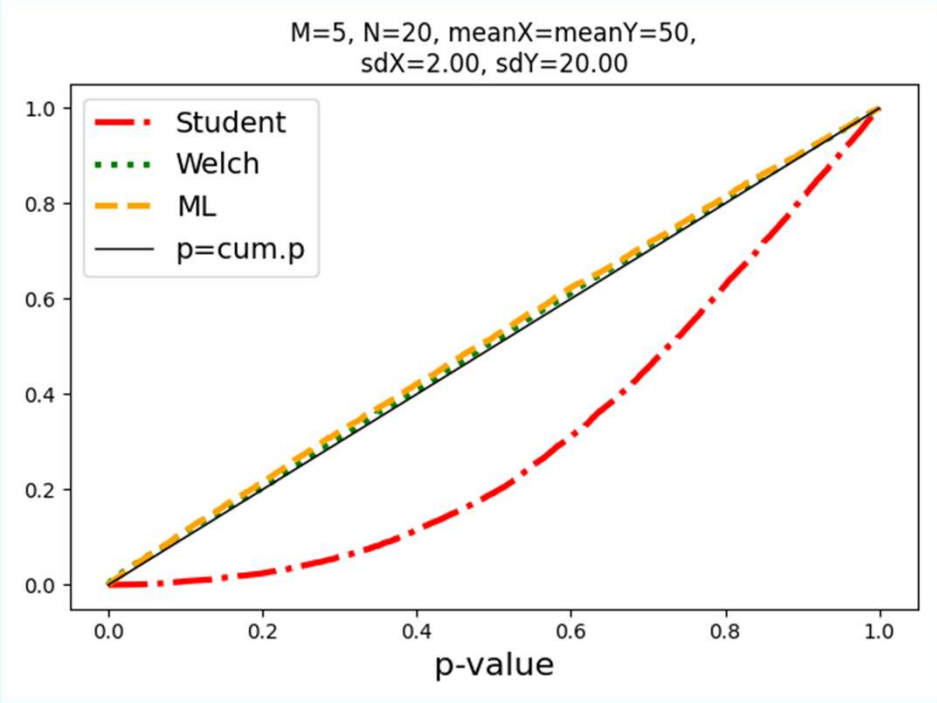

M=20, N=20, meanX=meanY=50,
sdX=20.00, sdY=2.00
Student
Welch
ML
p=cum.p
p-value

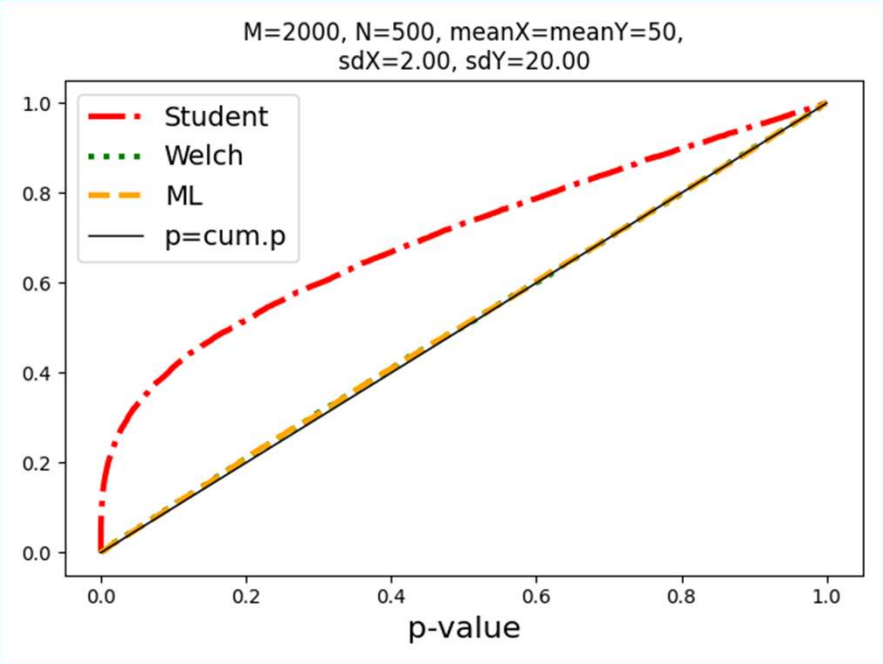

M=20, N=20, meanX=meanY=1.65,
sdX=2.16, sdY=0.68
Student
Welch
ML
p=cum.p

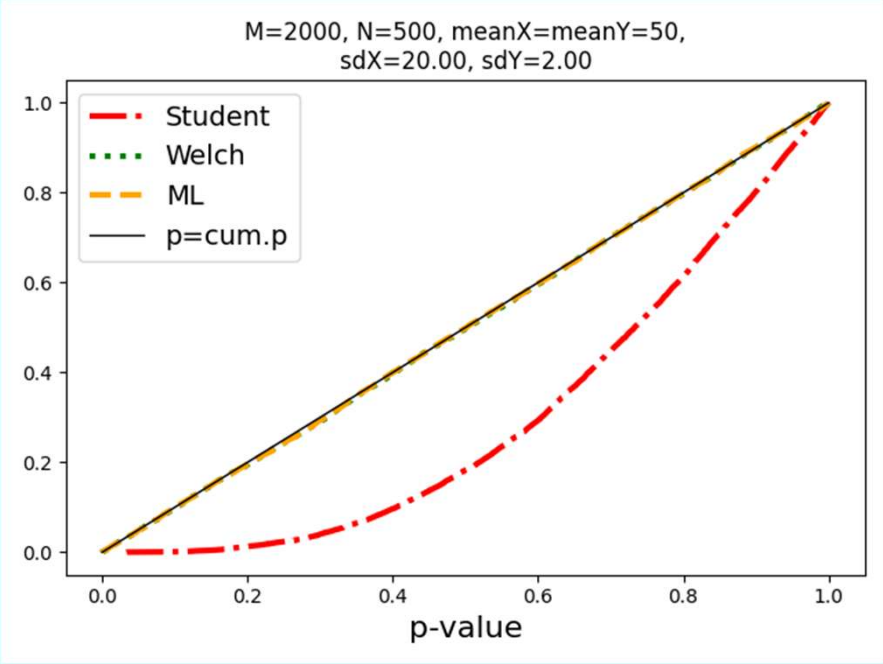

mean1=50
mean2=50
mean1=1.65
mean2=1.65

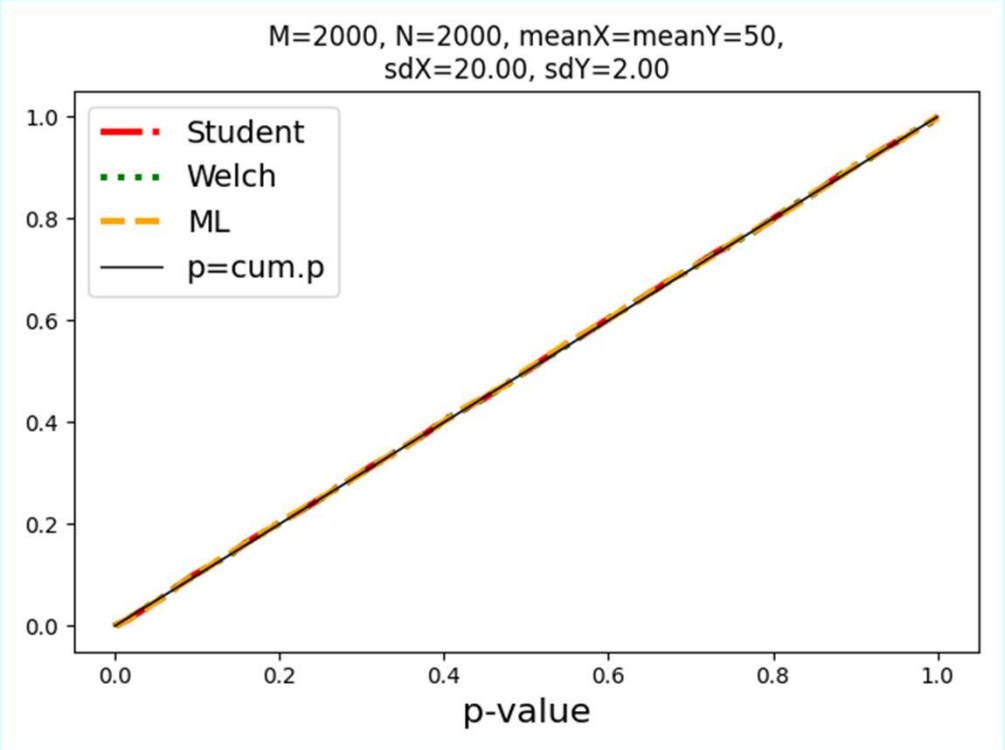

M=20, N=5, meanX=meanY=50,
sdX=20.00, sdY=20.00
Student
Welch
ML
p=cum.p
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
p-value


M=20, N=5, meanX=meanY=50,
sdX=2.00, sdY=20.00
Student
Welch
ML
p=cum.p
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
p-value


M=5, N=20, meanX=meanY=50,
sdX=2.00, sdY=20.00
Student
Welch
ML
p=cum.p
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
p-value


M=2000, N=500, meanX=meanY=50,
sdX=2.00, sdY=20.00
Student
Welch
ML
p=cum.p
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
p-value


M=2000, N=500, meanX=meanY=50,
sdX=20.00, sdY=2.00
Student
Welch
ML
p=cum.p
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
p-value


M=2000, N=2000, meanX=meanY=50,
sdX=20.00, sdY=2.00
Student
Welch
ML
p=cum.p
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
p-value

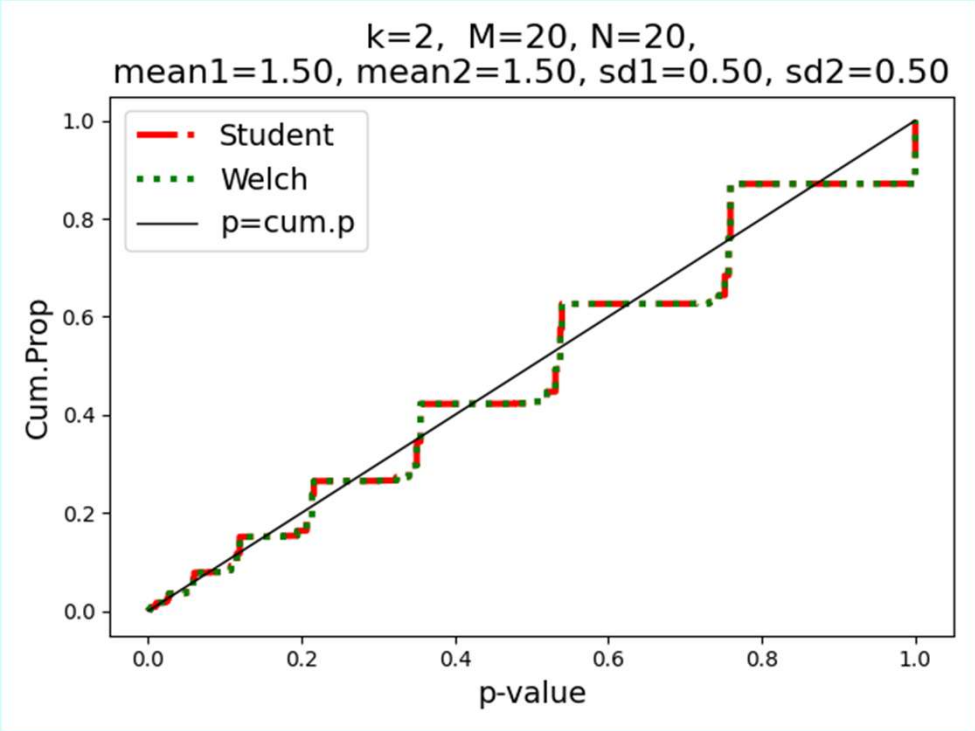
順序カテゴリ値

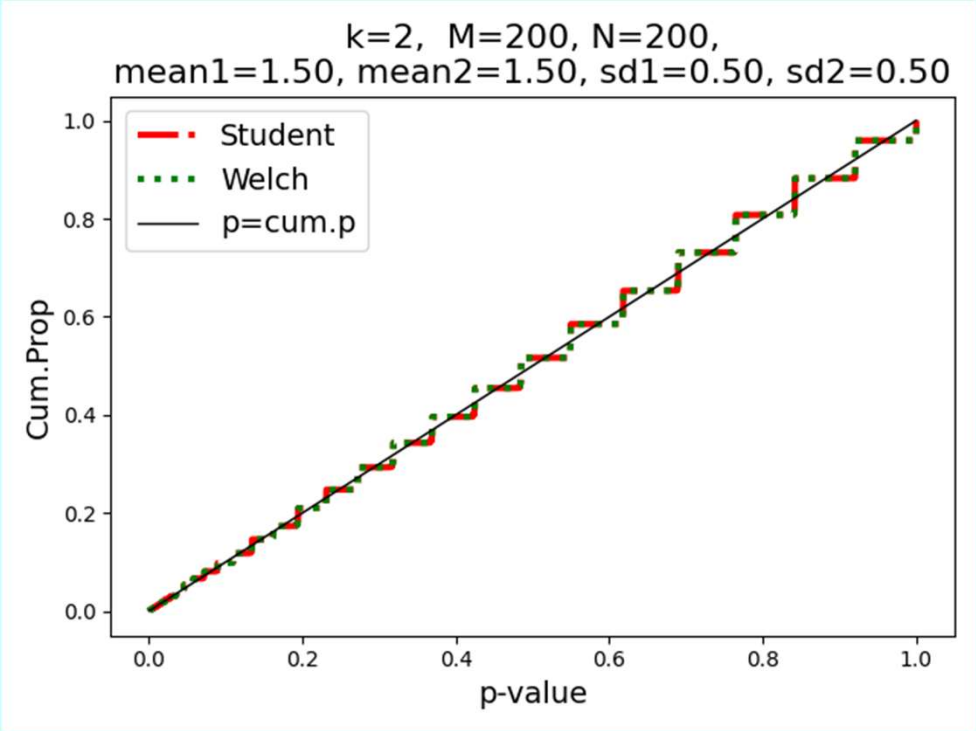
カテゴリ数がk、
カテゴリー値が１からkまでの整数で与えられているとする。

カテゴリー値がiである確率を$p[i]$で表す。

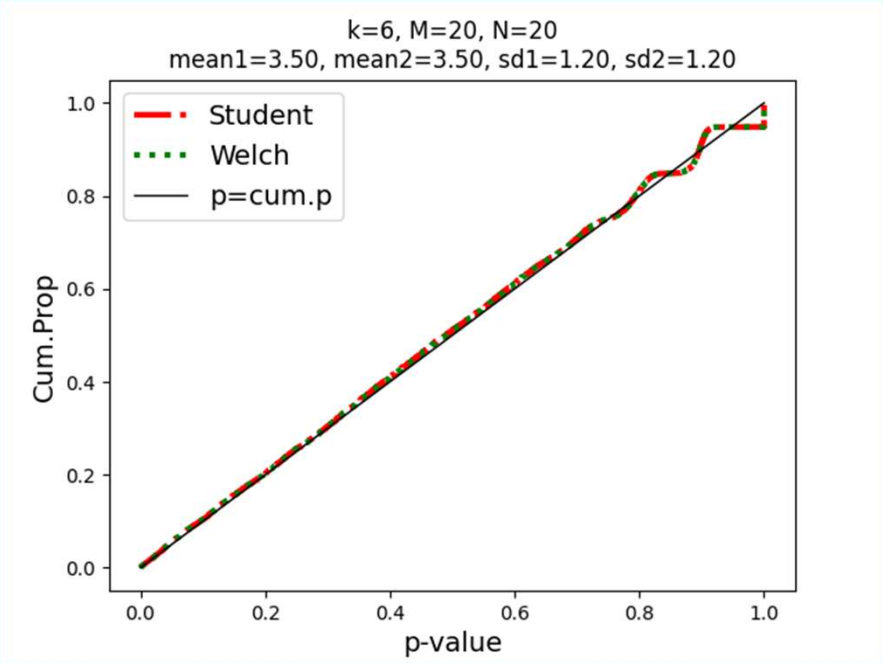
平均値　　$E[i] = \sum_{i=1}^{k} i \times p[i]$

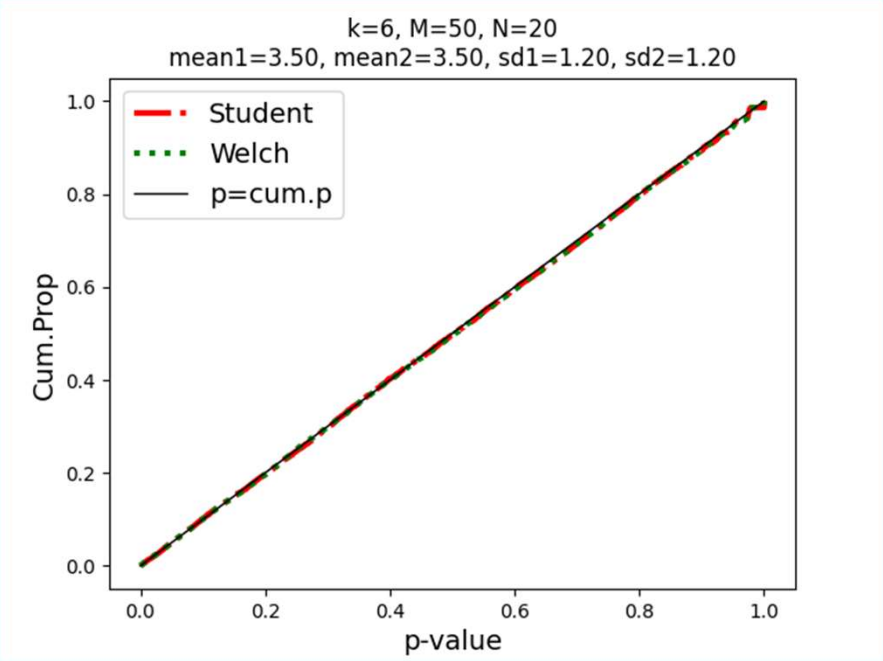
$$p_1 = [0.5 \quad 0.5]$$


k=2, M=20, N=20,
mean1=1.50, mean2=1.50, sd1=0.50, sd2=0.50
Student
Welch
p=cum.p
Cum.Prop
p-value
0.0
0.2
0.4
0.6
0.8
1.0


k=2, M=200, N=200,
mean1=1.50, mean2=1.50, sd1=0.50, sd2=0.50
Student
Welch
p=cum.p
Cum.Prop
p-value
0.0
0.2
0.4
0.6
0.8
1.0

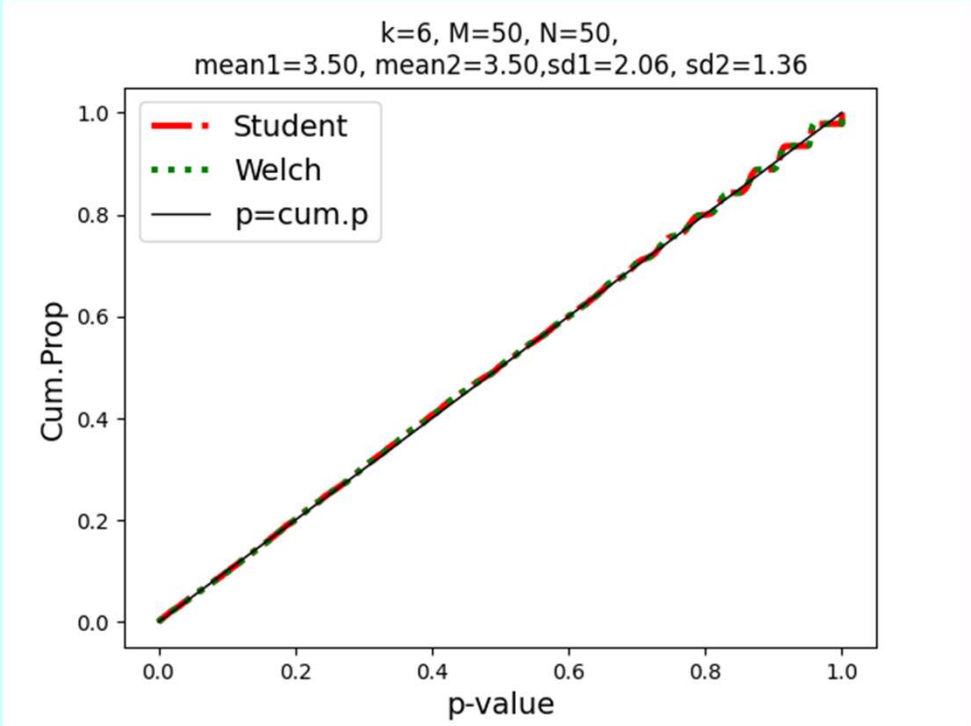
p1=p2 = [0.05, 0.15, 0.3, 0.3, 0.15, 0.05]



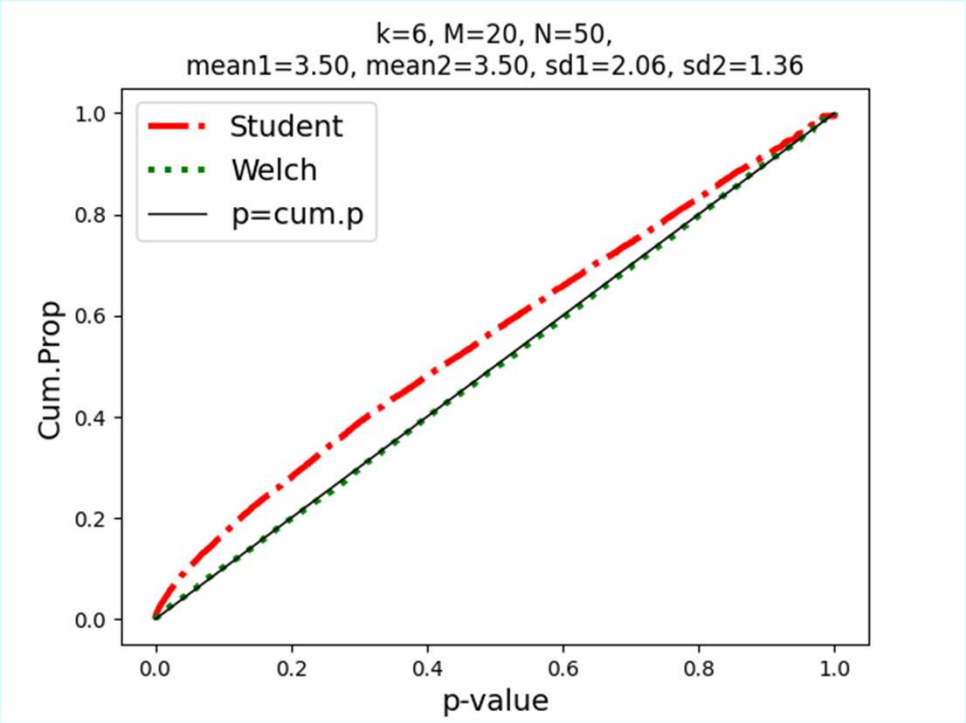


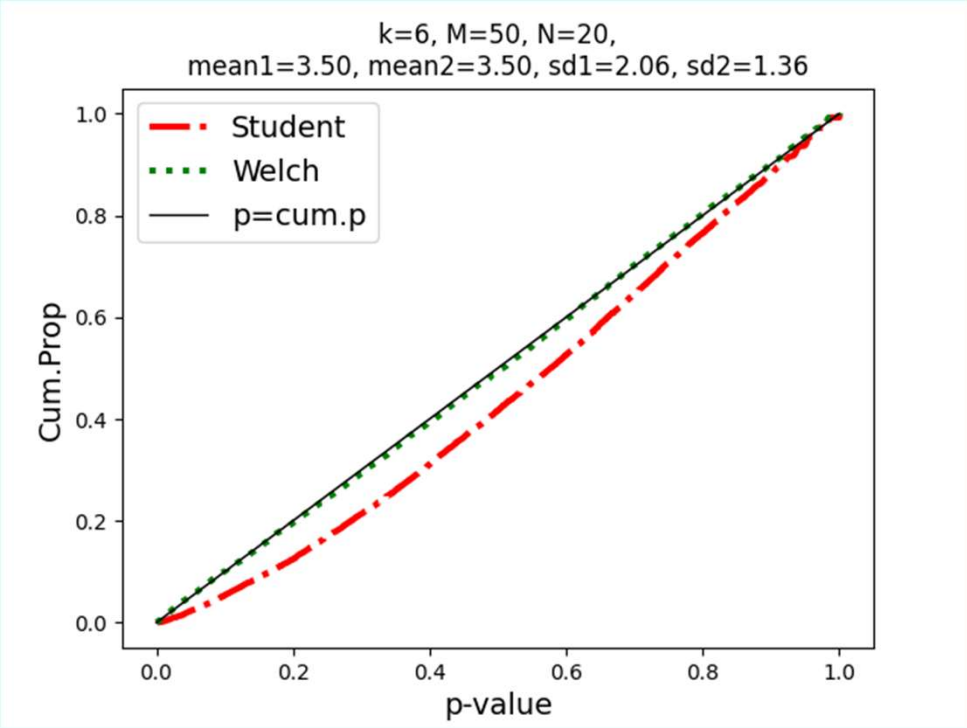
p1 = [0.3, 0.1, 0.1, 0.1, 0.1, 0.3]

p2 = [0.1, 0.1, 0.3, 0.3, 0.1, 0.1]



$$p1 = [0.3, 0.1, 0.1, 0.1, 0.1, 0.3]$$
$$p2 = [0.1, 0.1, 0.3, 0.3, 0.1, 0.1]$$





まとめ１

正規分布、 t 分布、ポアッソン分布、対数正規分布、順序カテゴリ値

サンプルサイズが等しい（M=N）とき

　対数正規分布の場合を除いて、Studentの p 値、Welchの p 値ともに、等分散あるいは不等分散の場合に（ポアッソン分布は、不等分散の場合はない）、一様分布に従っていたといえる。
　対数正規分布の場合は、等分散のとき、Studentのp値、Welchのp値、尤度比のP値とも一様分布に従っていたが、不等分散のときは、分布の歪みが大きいときにStudentのp値とWelchのp値は、p値の小さい領域において一様分布からずれる傾向がみられた。サンプルサイズをM=N=2000と大きくとると、３つのp値はいずれも一様分布に一致した。

サンプルサイズが異なる（M≠N）とき

正規分布	等分散	Studentのp：〇	Welchのp：〇
	不等分散	Studentのp：×	Welchの p ：〇
t 分布	等分散	Studentのp：〇	Welchのp：〇
	不等分散	Studentのp：×	Welchの p ：〇
ポアッソン分布			
	等分散のみ	Studentのp：〇	Welchのp：〇

対数正規分布

等分散　　Studentの p ：〇　　Welchの p ：　〇　　尤度比の p ：△

不等分散　　Studentの p ：×　　Welchの p ：　〇

尤度比の p ：分散の大きい方がサイズ大〇
分散の大きい方がサイズ小△

サンプル数 M=2000, N=500
Studentのp：×　　Welchの p ：〇　　尤度比の p ：〇

順序カテゴリ値

等分散　　Studentの p ：〇　　Welchの p ：〇

不等分散　　Studentの p ：×　　Welchの p ：〇

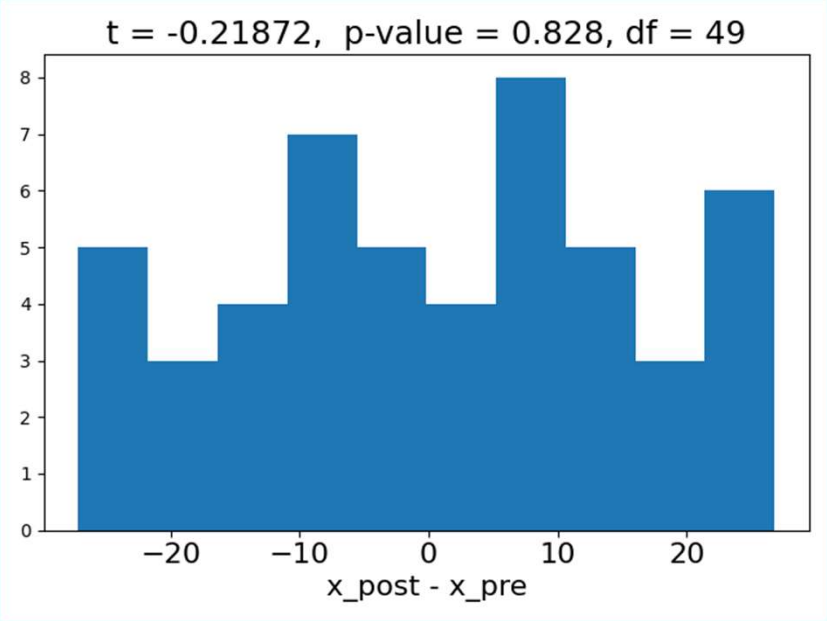
対応のあるデータの分析

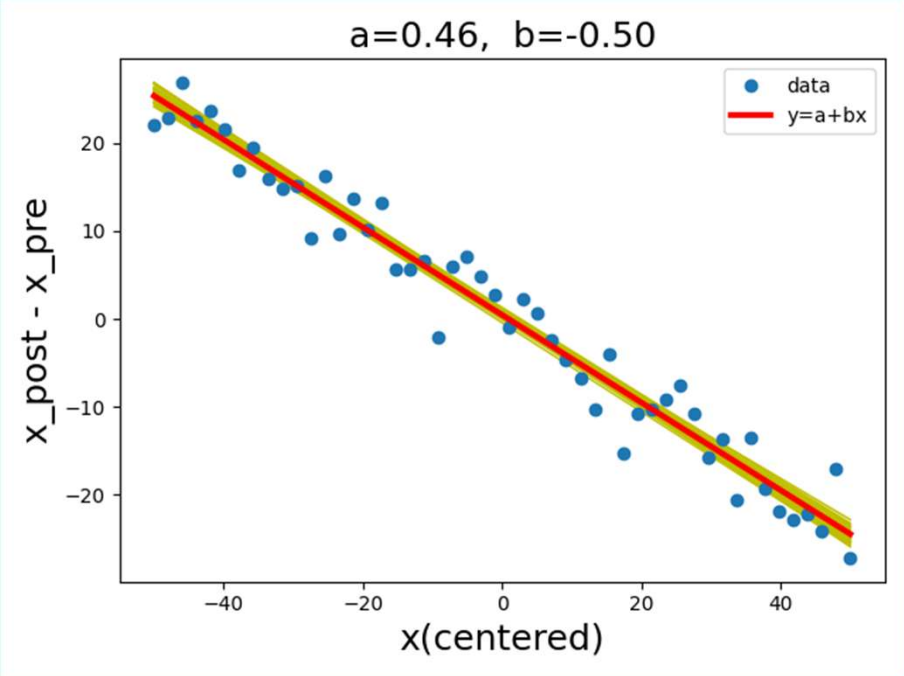
$$(X_{pre,i}, X_{post,i}) \qquad d_i = X_{post,i} - X_{pre,i}$$

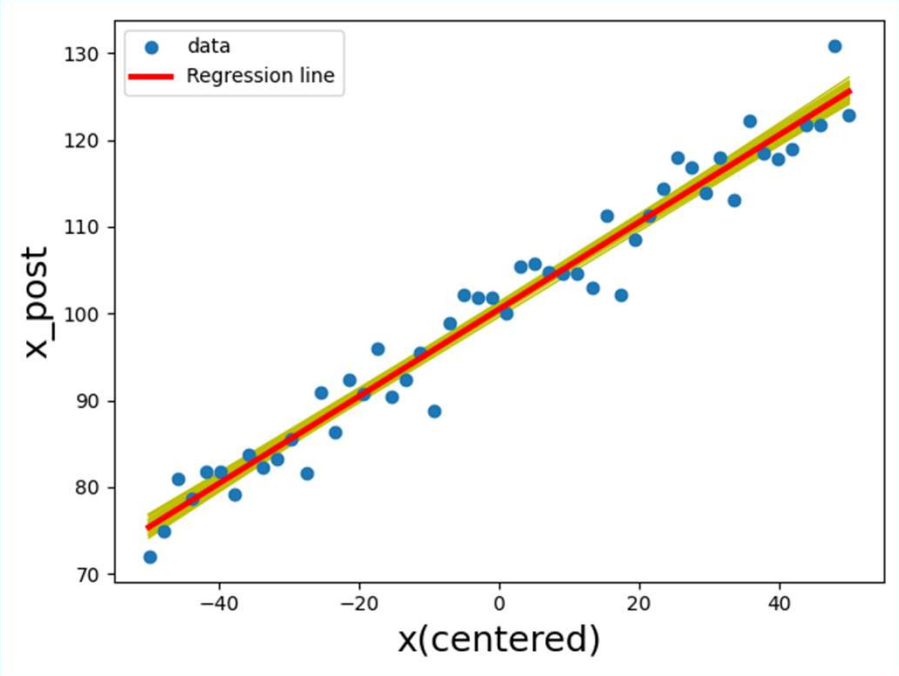
$$t = \frac{\bar{d}}{\sqrt{s^2/N}} \qquad df = N - 1$$

$$\bar{d} = \frac{1}{N}\sum d_i \qquad s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(d_i - \bar{d})^2$$

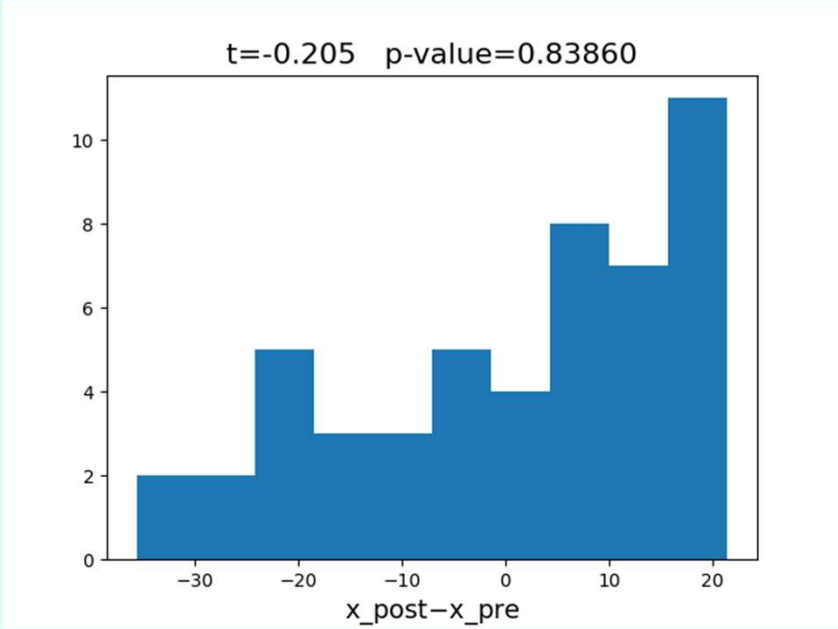
線型関係の場合

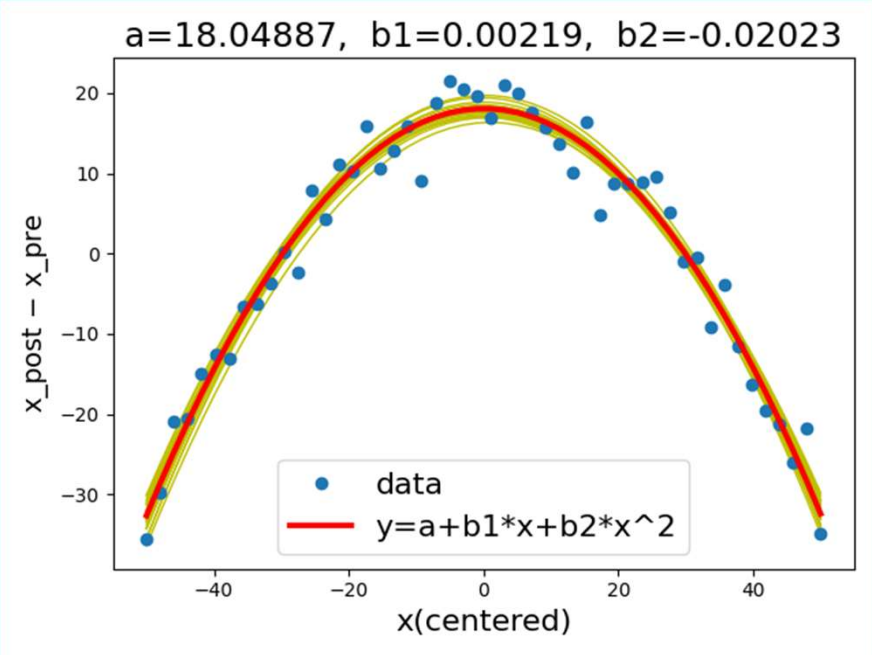

t = -0.21872, p-value = 0.828, df = 49
x_post - x_pre

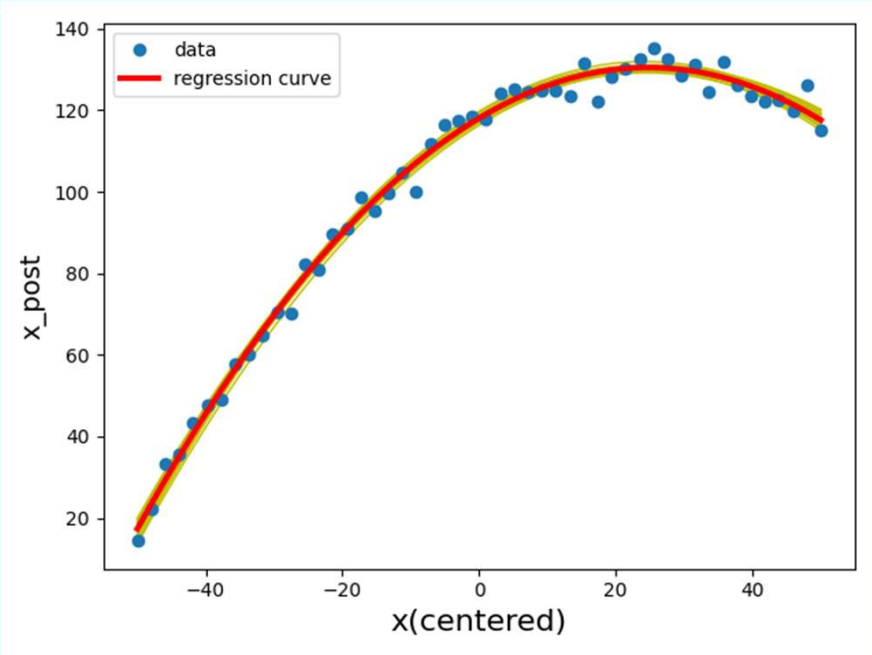

a=0.46, b=-0.50
data
y=a+bx
x_post - x_pre
x(centered)


data
Regression line
x_post
x(centered)

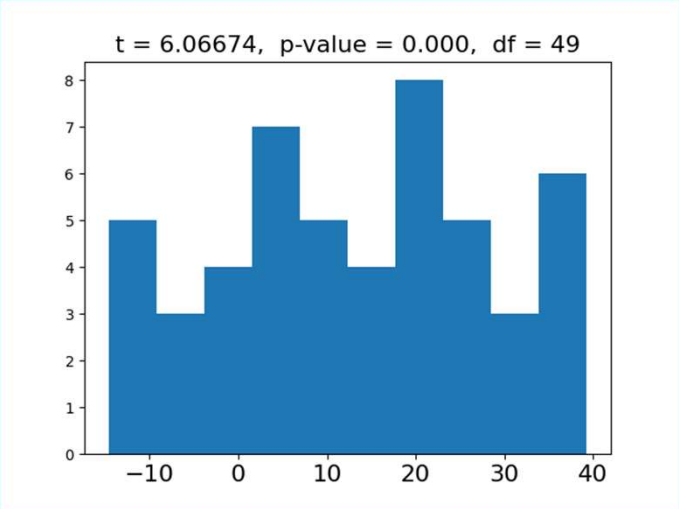
逆U字関係の場合

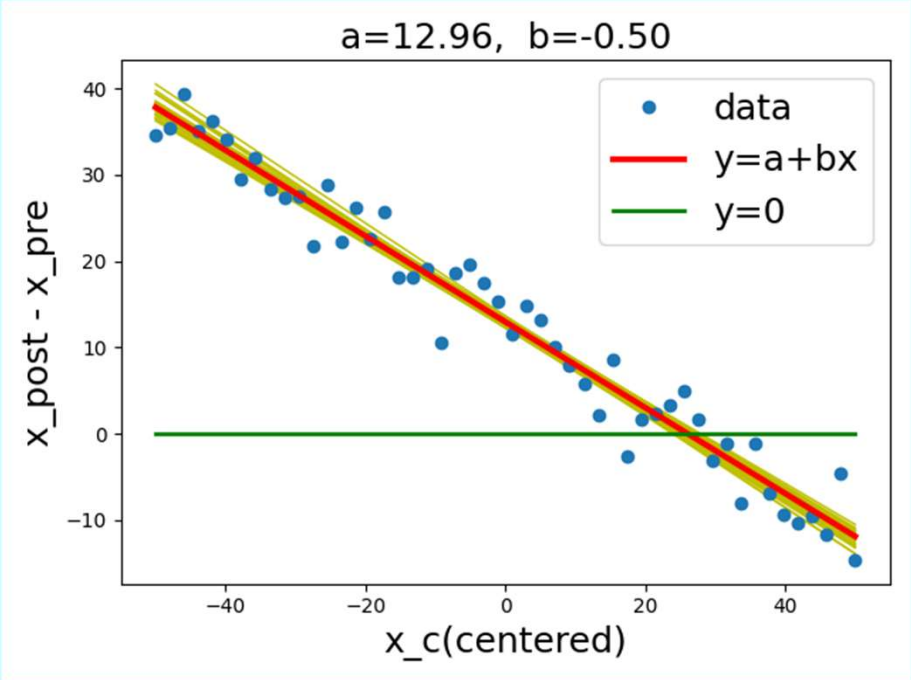

t=-0.205　p-value=0.83860
x_post−x_pre

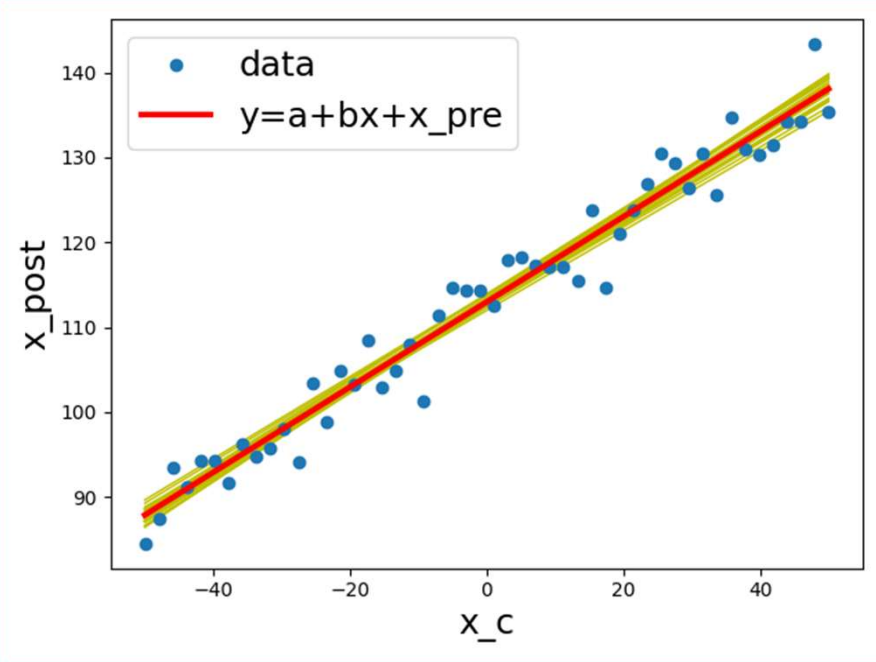

a=18.04887, b1=0.00219, b2=-0.02023
x_post − x_pre
x(centered)
data
y=a+b1*x+b2*x^2


data
regression curve
x_post
x(centered)

t 検定有意、1次式回帰関係あり







まとめ２

１．ｔ検定が非有意でも、回帰モデルで分析すると関係が見られることがある。

２．ｔ検定が有意であるとき、回帰分析を行うことによりより詳しい関係が見られることがある。