



Copyright © SAS Institute Inc. All rights reserved.



1 データの代表値 平均・バラツキ、中央値 2 データの視覚化 箱ひげ図、ヒストグラム

3 正規分布、t 分布

4 t 検定

1標本、2標本



平均(算術平均)

(例) 英語の研修前後の成績(20人)

	A君の得点	全体の平均点	得点-平均点
研修前	70	58.3	+11.7
研修後	75	58.3	+13.7

A君の成績の評価は?

平均の信頼性

(例)暑い日に暑い場所で待ち合わせをした。 いつも遅れてくる人が何分後に来るのか予測。

<過去10回の遅刻データ(分)>
平均
① 21 46 8 28 19 34 13 33 19 31 25.2分
② 25 26 23 24 26 27 26 25 24 26 25.2分
③ 21 22 21 23 22 23 22 21 23 54 25.2分

<u>①のデータは「バラツキ」が大きい。</u>

「バラツキ」の大きいデータの平均は信頼できない。

③のデータは「外れ値」を含む。

「外れ値」を含むデータの平均は信頼できない。





大きさの順番に並び替えて、中央に位置する値を代表値

5

データの視覚化

データを分析するには全体を把握することが重要。

- (1) 各データ群の把握
 ①外れ値の有無の確認
 ②分布の偏り
 箱ひげ図、ヒストグラム
- (2) 2群以上のデータ群の関係の把握 散布図







Q1 第1四分位数(25%) : 252 Q2 第2四分位数(50%) : 292 Q3 第3四分位数(75%) : 331

<u>IQR 四分位範囲 Q3-Q1=331-252=79</u>

Q1-1.5×IQR以下 Q3+1.5×IQR以上

 $252 - 1.5 \times 79 = 133.5$ $331 + 1.5 \times 79 = 449.5$

外れ値: 133.5以下 449.5以上





(外れ値がない場合)



Ssas.



SAS Studio

1.SAS Studio にログインする。 2.SAS®Studio をクリックする。

	N SAS® OnDema Sig	^{otices} nd for Academics gn In			
گ	SAS Profile email add	ress or user ID			
f	Password				
	Accept the terms of use and conditions	Applications	Enrollments	Courses	
		SAS® Studio Write and run SA Actions: Clear	AS code with a Web-b my saved tabs.	based SAS development env	ironment.

Ssas

3. 「SASプログラマ」をクリックし、「ビジュアルプログラマ」を選択する。





4. 新規フォルダを作成する。

「ファイル(ホーム)」を右クリックー「新規作成」ー「フォルダ」をクリック、 「新規フォルダ名(データ分析の基礎1)」を入力し、「保存」をクリックする。

AS [®] Studio	
・サーバーファイルとフォルダ	・ 実行 □ □ □ + Kの生成 № ・ ドの生成 № ・ ブロパティ ・ 十 - □ 色 - 前 ■ チ № ・
▲ 🔽 ファイル (木一/コ) 開く	
新規作成 テキスト形式でファイルを表述	 ▶ SAS プログラム (F4) 示する データのインポート
データのインボート	クエリ
マイタスクに追加 マイス二ペットに追加	YML
作成	フォルダ フォルタショートカット
名前の変更削除	
移動	新規フォルダ
ファイルのアップロード ファイルのダウンロード	場所: /home/u61364448名前: データ分析の基礎1
プロパティ	保存 キャンセル



5. ファイルをアップロードする。 「データ分析の基礎1」を右クリック、「ファイルのアップロード」をクリックし、 「ファイルの選択」から、ファイルを選択し「アップロード」をクリックする。

SAS [®] Studio			
 サーバーファイルとフォルダ ・	<mark>1</mark> プロセス: ▶ 実行 6 + ▼ 6	〕 — 1 × ■ 幽: 結果 ● ▼ 音	
■ テータ分析の基礎1	開<		
	新規作成 テキスト形式でファイルを表示す	ファイルのアップロード	×
	データのインポート	ファイルのマップロード先、 (heme////12/////2/データイン	七个甘味1
	マイタスクに追加	ファイルのアックロード元: //nome/u81364448/テータ分	竹の基礎1
	マイスニペットに追加	ファイルの選択	
	作成	▶ 選択済みファイル:	
	名前の変更	1 XLSX EXCELデータ.xlsx	15.1 kb
	削除		
	を割 コピー		
	コマイルのアップロード		
	ファイルのダウンロード		
	プロパティ		
l			

6.「データ分析の基礎1」フォルダを開きファイル(EXCELデータ)を 右側のプロセスフロー画面にドラッグし、右クリックして「開く」を選択する。



7.「ワークシート名(Data1)」を入力し、実行ボタンをクリックする。

SAS [®] Studio	
▼ サーバーファイルとフォルダ	ta *プロセスフロー 1 ×
は 一曲 出 平 目 55	<u>プロセスフロー1</u> > "EXCELデースxlsx" のインポート
⊿ 🛱 odaws01-apse1	
🔁 フォルダショートカット	オプション ノード
▲ □ ファイル (ホーム)	▼ ファイル情報
▲ - テータ分析の基礎1	ソース ファイル
X EXCELT-Y.XISX	ファイル名: EXCELデータ.xlsx
	 ソースの場所・/home/u£1364448/データ分析の基礎1 ワークシート名: Data1 出力データ SAS Server: SASApp データセット名: IMPORT ライブラリ: WORK 変更 オブション ファイルの種類: デフォルト(ファイル拡張子に基づく) ▼ コード ログ 結果 こード ログ 結果 図 配 配 里 ● 行番号 ② ● 純 編集 1 /* 生成コード (IMPORT) */ 2 /* ソースファイル: EXCELデータ.xlsx */

8. 「出力データ」をクリックし、「テーブル(データセット名)」、データの内容を 確認する。

AS [®] Studio				
サーバーファイルとフォルダ └ ・ 面 上 〒 目 い ・ 型 odaws01-apse1 □ フォルグショートカット ・ □ ファイル(ホーム) ・ ■ データ分析の基礎1 □ EXCELデータ.xlsx	*ブロセスフロー1 ^プ ロセスフロー1 ^プ ロセスフロー1 ^プ ロセスフロー1 ^プ ^プ ^プ ^ブ	× "EXCELデータ.xlsx"のインオ 分割 ・ え Q ジ - に EXCELデータ.xlsx home/u61364448/データ分析 SASApp IMPORT1 WORK ブ 活果 ビュー: ビュー:	-タ -タ -タ ト 	フィルタ: (なし) 1
	2 ② 報酬A		1 2 3	
タスクとユーティリティ			4	
	プロパティ	値	4	

9. 「プロセスフロー」をクリックしてフロー画面に戻り、「タスクとユーティリティ」を開く。





10. 「タスク」 – 「統計量」をクリックし、「要約統計量」をフロー画面にドラッグし 「EXCELデータ・・」と結合、右クリック – 「開く」をクリックする。



18

11.「データ」の「テーブル(データセット名)」を確認する。 変更する場合は「テーブルの選択」から「データセット」を選択する。 12.「分析変数」をセットする。(右の+マークをクリック)

SAS [®] Studio	
▶ サーバーファイルとフォルダ	ti *プロセスフロー1 ×
▼ タスクとユーティリティ	<u>プロセスフロー1</u> > 要約統計量
☆ ★ 龠 民 ■ (5)	設定 コード/結果 分割 🖌 😡 🐹
	● ① データ オプション 出力 情報 ノード
	WORK.IMPORTI
	→役割
	*分析変数: 🚹 🕈 🖬 🕇
タ探索	22 列
国要約統計量	
M 分布分析	
而一元度数表	
↓ 相関分析	
☑ 分割表分析	
H t 検定	
▶ ■ 線形モデル	
▶ 🖪 生存時間分析	分類変数: ▲ ▲ 皿 +
▶ 🗳 多変量分析	
▶ ■ クラスター分析	
 ▶ <mark>□</mark> 検定力とサンプルサイズ	
▶ 🛄 統計的工程管理	
▶ 🛄 組み合わせと確率	
▶ 🖪 データマイニング	
▶ 🖪 計量経済	
▶ 🛄 予測	● 追加役割
▶ 📑 ネットワーク最適化	
▲ ಾ ユーティリティ	

13. 分析変数を確認し、OKボタンをクリックする。



14. 「オプション」をクリックし、「パーセント点」をクリックし、出力項目に チェックを入れ、実行する。





15. 要約統計量が表示される。

	分析変数:報酬A 報酬A							
平	平均 標準偏差 最小值 最大值 中央值 N 下側四分位点 上側四分位点 四分位範囲					四分位範囲		
401.11111	1 250.2056376	229.0000000	851.0000000	292.0000000	9	252.0000000	331.0000000	79.0000000





箱ひげ図の作成 (SAS Studio)

- 1. 「タスクとユーティリティ」→「タスク」→「グラフ」の「箱ひげ図」を選択、
- 2. フロー画面にドラッグし、「EXCELデータ・・」と結合、右クリック-「開く」 をクリックする。



3.「データ」の「テーブル(データセット名)」を確認する。 変更する場合は「テーブルの選択」から「データセット」を選択する。 4.「分析変数」を確認し、実行ボタンをクリックする。



5. 箱ひげ図が表示される。





1 2 3 4 5 6 7 8 9 10 11 12 13 3 11 11 12 13 13 14 20 25 28 50 75 92

Q1 四分位点(25%) : 12 Q2 中央値 : 14 Q3 四分位点(75%) : 28 四分位範囲 Q3-Q1=28-12=16 12-1.5×16=-12 28+1.5×16=52 外れ値 : -12以下 52以上

	分析変数:報酬B 報酬B							
平均 標準偏差 最小值 最大值 中央值 N 下側四分位点 上側四分位点 四分位範					四分位範囲			
28.2307692	27.33481.86	3,0000000	92,0000000	14.000000	13	12.0000000	28.0000000	16.000000









成績

88	88	60	86	86
48	86	60	46	40
64	84	70	34	62
78	36	77	72	76
65	68	90	90	56
72	54	54	52	66
68	68	60	74	72
78	56	62	96	90
36	86	48	48	90
48	68	84	66	46



度数分布表



ヒストグラムの作成(SAS Studio) 1.「データ分析の基礎1」フォルダを開きファイル(EXCELデータ)を 右側のプロセスフロー画面にドラッグし、右クリックして「開く」を選択する。



Sas

2.「ワークシート名(Data5)」を入力し、実行ボタンをクリックする。

SAS [®] Studio	
 SAS[®] Studio サーバーファイルとフォルダ ● 面 査 〒 目 い ● 切 odaws01-apse1 ● フォルダショートカット ● ファイル(ホーム) ● データ分析の基礎1 ■ データ.xlsx 	*プロセスフロー1 × プロセスフロー1 > "EXCELT ち xlsx" のインボート 設定 コード/結果 分割 メブション ノード * ファイル情報 ソースファイル ファイル名: EXCELデータ.xlsx ソースの場所: /home/u61364448/データ分析の基礎1 ワークシート名: Data5 出力データ SAS Server: SASApp データセット名: IMPORT1 ライブラリ: WORK アフ
	変更 オプション ファイルの種類: デフォルト(ファイル拡張子に基づく) ▼

3. 「出力データ」をクリックし、「テーブル(データセット名)」、データの内容を 確認する。

SAS [®] Studio		
・ サーバーファイルとフォルダ	t *プロセスフロー1 ×	
[✿マ 奤 匙 〒 围 \$5 ▲ ⑮ odaws01-apse1	<u>プロセスフロー1</u> > "EXCELデータ.xlsx"のイ 設定 コード/結果 分割 ↓ ★ Q ↓ \$	ンボート
■ フォルダショートカット ■ ファイル (ホーム)	オプションノード → ファイル情報	
▲ ■ データ分析の基礎1	ソースファイル	
	ファイル名: EXCELテータ.XISX ソースの場所: /home/u61364448/データ	分析の基礎1
	ワークシート名: Data5	
	出力データ	
	SAS Server: SASApp データセット名: IMPORT1	
	ライブラリ: WORK 変更	
	• オプション	
	ファイルの種類: デフォルト(ファイル拡張子に基づく) 🗸	
	JU: WORKIMPORTI ▼	: 例名 ▼ 戦 量 5 目 ▼ フィルタ: (なし ④ 合計行数:50 合計列数:1
	✓ すべて選択	
	☑ @ 成績	1
		3
タスクとユーティリティ		4



SAS Studio	
▼ サーバーファイルとフォルダ	t *プロセスフロー1 ×
は~ 竜 玉 平 目 い	▶ 実行 🔒 😡 😫 コードの生成 🚼
⊿ 🖫 odaws01-apse1	フロー 結果 プロパティ
🔁 フォルダショートカット	+- 늘 色- 命 晶 / + い 論 能
🖌 🗖 ファイル (ホーム)	
▲ データ分析の基礎1	
R EXCELデータ.xlsx	
▶ タスクとユーティリティ	
▶ タスクとユーティリティ ▶ ユニペット	

5. 「タスク」 – 「グラフ」をクリックし、「ヒストグラム」をフロー画面にドラッグし 「EXCELデータ・・」と結合、右クリック-「開く」をクリックする。



6.「データ」の「テーブル(データセット名)」を確認する。 変更する場合は「テーブルの選択」から「データセット」を選択する。 7.「分析変数」をセットし、実行ボタンをクリックする。

SAS [®] Studio		
▶ サーバーファイルとフォルダ	t *プロセスフロー1 ×	
▼ タスクとユーティリティ	<u>プロセスフロー1</u> > ヒストグラム	
雍→ 亩 民 目 \$5		20
▶ 🚾 マイタスク	テーダ 表示 情報	7-1
▲ ■ タスク	 データ 	
▶ 🖴 データ	WORK.IMPORT1	
⊿ Щ グラフ	▼フィルタ:(なし)	
山棒グラフ	▼ 役割	
┙ 棒-折れ線グラフ	*分析変数: (1項目)	命 +
■● 箱ひげ図	2 成績	-
▶ バブルプロット		
調ヒートマップ	尺度: パーセント(デフォルト) 🗸	
山. ヒストグラム	▶ 追加役割	
▶ 折れ線グラフ		
■ モザイクプロット		
● 円クラフ		
▶ ※列フロット		
▶ ■ マッフ		
▶ 📲 統計量		
8. ヒストグラムが表示される。





英語の研修前後の成績(20人)

	A君の得点	全体の平均点	得点-平均点
研修前	70	58.3	+11.7
研修後	75	58.3	+16.7

A君の成績の評価は?

平均を評価基準とする合理性?



• <研修前>

70、56、89、27、69、57、69 50、33、67、37、49、98、69 68、25、65、67、33、68

• <研修後>

75、31、95、36、89、88、89 76、28、47、23、28、96、48 51、20、30、91、27、98

全員の成績

• <研修前>

70、56、89、27、69、57、69 50、33、67、37、49、98、69 68、25、65、67、33、68

• <研修後>

75、31、95、36、89、88、89 76、28、47、23、28、96、48 51、20、30、91、27、98

研修後の成績順位は下がった!

Sas

要約統計量(SAS Studio)

1.「データ分析の基礎1」フォルダを開きファイル(EXCELデータ)を 右側のプロセスフロー画面にドラッグし、右クリックして「開く」を選択する。



41

2.「ワークシート名(Data8)」を入力し、実行ボタンをクリックする。

SAS [®] Studio	
 サーバーファイルとフォルダ ● 面 ▲ 〒 目 い ● 切aws01-apse1 ● フォルダショートカット ● ファイル(ホーム) ● データ分析の基礎1 ● EXCELデータ.xlsx 	ご *ブロセスフロー1 × プロセスフロー1 > *EXCEL × k/sx* のインボート 該定 コード/箱里 分割 オブション ノード * ファイル/雨線 ファイル ファイル名: アスの場所: /home/u61364448/データ分析の基礎1 ワークシート名: Data8 エリアータ SAS Server: SASApp データセット名: IMPORT2 ライブラリ: WORK 変更 ・ ・ オブション ファイル加速子に基づく) ▼ ニード ログ 福集 ① ● ● ● ● ● ① ● ● ● ● ● ● ● ● ● ● ● ● ●

3. 「タスク」ー「統計量」をクリックし、「要約統計量」をフロー画面にドラッグし 「EXCELデータ・・」と結合、右クリックー「開く」をクリックする。



43 **Sas**

4.「データ」の「テーブル(データセット名)」を確認する。 変更する場合は「テーブルの選択」から「データセット」を選択する。 5.「分析変数」をセットする。



「オプション」をクリックし、「統計量」→「基本統計量」の出力項目に チェックを入れ選択する。 「詳細」をクリックし、「標準偏差と分散の除数」から「オブザベーションの数」を 選択、実行する。



7. 要約統計量が表示される。





研修前:19.19 研修後:28.84





	成績	平均	成績-平均	標準偏差
研修前	70	58.3	11.7	19.19
研修後	75	58.3	13.7	28.84

研修前		研修後
-----	--	-----

$\frac{70-58.3}{19.19} = 0.610 >$	75-5 28.	<u>8.3</u> = 0.579 <mark>84</mark>
成績と平均値の差		フ/店
標準偏差	$\overline{}$	ム言語

47

Z値の比較によりデータの評価・比較が可能





研修前の偏差値= $0.610 \times 10 + 50 = 56.09$ 研修後の偏差値= $0.579 \times 10 + 50 = 55.79$



Z値は、平均値=0 平均値より大のときプラスの値、小のときマイナスの値

偏差値は、平均値 = 50 平均値より大のとき50より大きな値、平均値より小の とき50より小さな値

49

Sas

Ζ値の特性

	データ	Z值	データ	Z值
	1	-1.4142	3	-1.1844
	2	-0.7071	8	0.55739
	3	0	6	-0.1393
	4	0.70711	4	-0.8361
	5	1.41421	11	1.60249
平均值	3	0	6.4	0
標準偏差	1.41421	1	2.87054	1



標準正規分布(standard normal distribution)



Z値=1(偏差値60)以上の割合は約15.87% Z値=2(偏差値70)以上の割合は全体の約2.28%



■日本の成年男子の身長が平均170cm 標準偏差6cm、 182cm以上の人は、全体の何%を占めるか?







Ssas.

◇競合メーカー 有効成分含有量 平均=60mg、標準偏差=5mg

市場から4つ購入した平均:65mg 60mgより多いのか?



P値(有意確率)を求める! 違うと判定したときの危険率



$$Z = \frac{65-60}{5} \times \sqrt{4} = 2$$

P値 (有意確率) = 2.28%

サンプルサイズが大きくなると、Z 値も大きくなる



Z値による検定は標本数が少ないときに精度が良くない。



t 分布とは

1908年、W.S.ゴセット(ギネスビールの研究員)が考案。 データが少ないときに精度を上げるために t 値を使う。



■全て(データの大小に関係なく) t 分布を使う。 ■この検定方法を t 検定という。

Sas

t 分布



データの大きさによって形が変わる。 大きくなると正規分布と同じ形。

注) 自由度: データ数-1



t 値が大 = P値(有意確率)は小 t 値が小 = P値(有意確率)は大

(A) ランダムに選んだ製品10個の有効成分含有量 60mg (母平均) と違いはみられるか?

65 62 57 61 60 58 63 67 62 63

10個の平均値を求めると (65+62+・・・+62+63) ÷10 = 61.8

標本平均(61.8)と母平均(60.0)の比較

1標本 t 検定(SAS Stutio)

1.「データ分析の基礎1」フォルダを開きファイル(EXCELデータ)を 右側のプロセスフロー画面にドラッグし、右クリックして「開く」を選択する。



2.「ワークシート名(Data10)」を入力し、実行ボタンをクリックする。

SAS [®] Studio	
 サーバーファイルとフォルダ ● daws02-apse1 ● フォルダショートカット ● ファイル(ホーム) ● sasuser.v94 ● データ分析の基礎1 ● EXCELデータ.xlsx 	こ

61

3. 「タスクとユーティリティ」→「タスク」→「統計量」の「t検定」をフォロー画面に ドラッグし、「EXCELデータ・・」と結合、右クリック-「開く」をクリックする。



4. 「t検定」を右クリック、開き、「データ」、「t検定」、「分析変数」をセットする。

SAS [®] Studio	
サーバーファイルとフォルダ	t *プロセスフロー1 ×
 マタスクとユーティリティ 確マ 亩 民 目 い ■ マイタスク 	<u>プロセスフロー1</u> > t 検定 設定 コード/結果 分割 火
 ▶ № マイタスク ▲ ● タスク ▶ ● データ ▶ ● グラフ ▶ ● マップ ▲ ● 統計量 職 データ探索 錣 要約統計量 M 分布分析 Ⅲ 一元度数表 ☑ 相関分析 亞 分割表分析 ☑ ビ検定 	 データ WORK.IMPORT プフィルタ: (なし) 役割 t検定: 1 標本検定 *分析変数: (1項目) ★
▶ ■ 線形モデル ▶ ■ 生存時間分析	



5. 「オプション」をクリックし、mu^=60を入力し、実行ボタンをクリックする。



6.1標本t検定の結果が表示される。

Ν	4	均	價準	相差	價	隼誤差	最	小値	最大値
10 61.8000		000		3.0111		0.9522		.0000	67.0000
平均 95			平均	りの 頼限界	. 4	票準偏差	ę i	標準(95% 信	偏差の 「頼限界
61.8000		59.6	460	63.954	10	3.011	1 :	2.0711	5.4971
自由度 t值 Pr>[t] 8 1.89 0.0913									

t值 = 1.89 有意確率 = 0.0913

サンプルサイズが拡大した場合

(B) ランダムに選んだ製品20個の有効成分含有量 母平均:60mg 違いはみられるか?

6562576160586367626365625761605863676263

20個の平均値を求めると (65+62+・・・+62+63) ÷20 = 61.8

標本平均(61.8)と母平均(60.0)の比較

「ワークシート名(Data11)」を入力し、実行ボタンをクリックする。

SAS [®] Studio	
▶ サーバーファイルとフォルダ	t *プロセスフロー1 ×
▼ タスクとユーティリティ	プロセスフロー1 〉 "EXCELデーシャlsx" のインボート
確→ 曲 民 目 65	
▶ 🔽 マイタスク	オプションノード
▲ 📫 タスク	▼ ファイル情報
▶ 📫 データ	ソース ファイル
▶ 🛄 グラフ	ファイル名: EXCELデータ.xlsx
▶ 🗳 マップ	ソースの場所: /home/u61364448/データ分析の基礎1
▲ 🛄 統計量	ワークシート名:
👯 データ探索	Data11
醫 要約統計量	
<u>M</u> 分布分析	出力データ
Ⅲ 一元度数表	SAS Server: SASApp
∠相関分析	
五 分割表分析	
ビt検定	変更
▶ 🔩 線形モデル	 オプション
▶ 💶 生存時間分析	ファイルの種類:
▶ 💶 多変量分析	デフォルト(ファイル拡張子に基づく) 👻
▶ 🖪 クラスター分析	

67

N	平均	標準偏差	標準誤差	最小值	最大值	
20	61.8000	2,9308	0.6553	57.0000	67.0000	

平均	平均の 95% 信頼限界		標準偏差	標準(95% 信	扁差の 頼限界
61.8000	60.4284	63.1716	2,9308	2.2288	4.2806



t值 = 2.75 > 1.89 P値(有意確率) = 0.0128 < 0.0913



t値(P値)はサンプルサイズの影響を受ける





1標本 t 検定 効果量の比較

	P値	t值	n	効果量
(A)	0.0913	1.89	10	0.60
(B)	0.0128	2.75	20	0.61

◇効果量(Cohen's d)の大きさの評価







(A)新製品の好感度について、男女別各10人に 10点満点にて調査した。 男女間の評価に違いは見られるか?



男性平均(6.1)と女性平均(5.3)の比較

◇1標本t検定
2標本 t 検定(SAS Stutio)

2.「ワークシート名(Data15)」を入力し、実行ボタンをクリックする。

SAS [®] Studio	
サーバーファイルとフォルダ	t *プロセスフロー1 ×
▼ タスクとユーティリティ	<u>プロセスフロー1</u> > "EXCEL ^T スxlsx" のインポート
雍→ 亩 民 目 \$5	
▶ 🔽 マイタスク	オプションノード
▲ № タスク	▼ ファイル情報
▶ 📲 データ	ソース ファイル
▶ 📫 グラフ	ファイル名: EXCELデータ.xlsx
▶ 🗳 マップ	ソースの場所: /home/u61364448/データ分析の基礎1
▲ 🖴 統計量	ワークシート名:
15 データ探索	Data15
器 要約統計量	
M 分布分析	出力データ
Ⅲ 一元度数表	SAS Server: SASApp
∠相関分析	
国 分割表分析	51759: WORK
<u>H</u> t 検定	変更
▶ 📲 線形モデル	 オプション
▶ 🔩 生存時間分析	ファイルの種類:
▶ 🔩 多変量分析	デフォルト(ファイル拡張子に基づく) ▼
▶ 🔩 クラスター分析	

3. 「タスクとユーティリティ」→「タスク」→「統計量」の「t検定」をフォロー画面に ドラッグし、「EXCELデータ・・」と結合、右クリック-「開く」をクリックする。



4. 「t検定」を右クリック、開き、「データ」、「t 検定(2標本検定)」、 「分析変数(評価)」、「グループ変数(性別)」をセットする。

SAS [®] Studio	
▶ サーバーファイルとフォルダ	ta *プロセスフロー1 ×
▼ タスクとユーティリティ	
確→ 曲 民 目 \$5	
▶ 🚾 マイタスク	
▲ ■ タスク	 データ
▶ 🛄 データ	WORK.IMPORT 👻 🎛
▶ 🛄 グラフ	♥フィルタ:(なし)
▶ 📫 マップ	投帯
▲ 🛄 統計量	t 検定:
15 データ探索	2 標本検定
殿 要約統計量	*公托亦为:// [百日]
121 分布分析	
III 一元度数表	● 第十1四
▶ 相関分析	*グループ変数: (1 項目)
狙 分割表分析	▲ 性別
ビ t 検定	
▶ 🛄 線形モデル	
▶ 💶 生存時間分析	



5. 結果が表示される。

	888	10	Pr > [t]
Equal	18	-1.25	0.2281
Unequal	17.036	-1.25	0.2290
	Equal Unequal	Equal 18 Unequal 17.036	Equal 18 -1.25 Unequal 17.036 -1.25

等分散性				
手法	分子の自由度	分母の自由度	F值	₽r≻F
Folded F	9	9	1.62	0.4813

◇分散が違うとは言えないとき P値(有意確率) = 0.2281

◇分散が違うときP値(有意確率) = 0.2290

サンプルサイズが拡大した場合

(B) 新製品の好感度について、男女別各30人に 10点満点にて調査した。 男女間の評価に違いは見られるか? 平均



男女各10人から男女各30人に増加

「ワークシート名(Data16)」を入力し、実行ボタンをクリックする。

SAS [®] Studio	
 サーバーファイルとフォルダ 	t: *プロセスフロー1 ×
 ▼ タスクとユーティリティ ・	<u>プロセスフロー1</u> > "EXCELデータ.xlsx" のインポート 設定 コード/結果 分割 ★ 😡 🔀
▶ № マイタスク ▲ № タスク	オプション ノード ・ファイル情報
▶ ■ データ ▶ ■ グラフ	ソースファイル ファイル名: EXCELデー タ.xisx
▶ ee マップ ▲ ee 統計量 ■ データ探索	ソースの場所: /home/u61364448/データ分析の基礎1 ワークシート名: Data16
器 要約統計量 <u> ()()</u> 分布分析	出力データ 000000000000000000000000000000000000
Ⅲ 一元度数表	SAS Server: SASApp データセット名: IMPORT ライブラリ: WORK
	変更
▶ ■ 生存時間分析 ▶ ■ 多変量分析	ファイルの種類:
、 ● カニフカー公析	

手法	分散	自由度	t値	Pr > [t]
Pooled	Equal	58	-2.24	0.0290
Satterthwaite	Unequal	54,895	-2.24	0.0292

等分散性				
手法	分母の自由度	F値	Pr > F	
Folded F	29	29	1.62	0.1976

P値(有意確率) = 0.0292 < 0.2290



◇2標本 t 検定

= 効果量 ×
$$\sqrt{\frac{n}{2}}$$

2標本 t 検定 効果量の比較

	P値	t值	n	効果量
(A)	0.2290	1.25	10	0.56
(B)	0.0292	2.24	30	0.58

◇効果量(Cohen's d)の大きさの評価





「違いの大きさ」はサンプルサイズの影響を受ける。 ビッグデータにおいては 「違いがある!」という分析結果が出やすい。



<u>サンプルサイズと効果量を検討することが重要!</u>



まとめ



アンケートのお願い・ご質問 8月3日 データ分析の基礎-1

今後の参考にさせていただくため、ぜひともアンケートにご協力を お願いします。

・無記名
 ・所要時間目安: 1~3分

<u>アンケートURL</u>

https://sas.qualtrics.com/jfe/form/SV_eeRL4HhHpOxpaKi



- ・お客様講演会のアーカイブは、2022年8月8日~2023年3月31日迄 視聴できます。
- 本日の内容に関するご質問は、以下宛にご連絡ください。
 que@datascience.co.jp

ご視聴ありがとうございました。

Sas