やさしい t 検定

Excel を利用した t 検定①

データを入力する
 < 例>ビタミンCの含有量の熱変化

К	3 🔻	÷	×	$\checkmark f_x$	2.98								
1	A	1	з	С	D	E	F	G	Н	1	J	K.	L
1			1	2	3	4	5	6	7	8	9	10	
2	加熱なし		3.21	3.34	3.26	3.22	3.15	3.21	3.08	3.32	3.15	3.07	
3	加熱あり		2.84	2.97	2.91	3.02	3.05	2.91	2.92	3.04	2.96	2.98	
4											8		

(単位は×10⁻³mol/l)

上の結果から,「加熱するとビタミンCの含有量が減少する」と判断してよいでしょうか。平均は加熱 ありの方が小さいですが,本当は差がほとんどないのに,たまたまこのような結果が出たということはな いでしょうか。

こうした問題を解くために編み出されたのが「仮説検定」という方法です。仮説検定には様々な種類が あり、その実験内容によって適切な検定方法が異なります。

今回の場合は2つの母集団の平均の比較をしたいのでt検定を行います。t検定にも様々な種類がありますが、今回は2つの母集団の分散が等しいとは限らないと考えて検定を行うことにします。この方法をWelch(ウェルチ)のt検定といいます。

2標本平均,不偏分散を計算する。

・標本平均・・・標本(得られたデータ)の平均

・不偏分散・・・母集団の分散の不偏推定量(標本の属する母集団の分散を推定したもの)

X:加熱なしの場合の標本平均 Y:加熱ありの場合の標本平均

 μ_X :加熱なしの場合の母平均 μ_Y :加熱ありの場合の母平均

 $\widehat{\sigma_{X}^{2}}$:加熱なしの場合の不偏分散 $\widehat{\sigma_{Y}^{2}}$:加熱ありの場合の不偏分散

※不偏分散の計算式 $\widehat{\sigma_X^2} = \frac{1}{m-1} \sum_{i=1}^m (x_{i-\bar{x}})^2$, $\widehat{\sigma_Y^2} = \frac{1}{n-1} \sum_{i=1}^n (y_{i-\bar{y}})^2$ 【データ数-1で割る】

標本分散の計算式 $\widehat{\sigma_X}^2 = \frac{1}{m} \sum_{i=1}^m (x_{i-\bar{X}})^2$, $\widehat{\sigma_Y}^2 = \frac{1}{n} \sum_{i=1}^n (y_{i-\bar{Y}})^2$ 【データ数で割る】

つまり, (不偏分散) = m/m-1 (標本分散)

※ Excel 上での関数 標本平均 『= AVERAGE(B2:K2)』 不偏分散 『= VAR(B2:K2)』

Ν	3 *	1 ×	$\checkmark f_x$	=VAR(B3	:K3)								
1	A	В	С	D	E	F	G	Ĥ	- E -	j į	К	L	Μ
1		1	2	3	4	5	6	7	8	9	10	平均	不偏分散
2	加熱なし	3.21	3.34	3.26	3.22	3.15	3.21	3.08	3.32	3.15	3.07	3.201	0.008277
3	加熱あり	2.84	2.97	2.91	3.02	3.05	2.91	2.92	3.04	2.96	2.98	2.96	0.0044

3 仮説検定の考え方

帰無仮説 H_0 :主張したいことの逆の仮説 対立仮説 H_1 :もともと主張したい仮説 fightarrow Factorian Fact

仮説H₁「加熱するとビタミンCの含有量が減少する」←*証明したい仮説* ひっくり返す↓

仮説 H_0 「加熱してもビタミンCの含有量は変わらない」←*その反対の仮説* 仮説を検証↓

仮説H₀は実験で得られたデータと矛盾する

 \downarrow

よって、仮説 H_0 は正しくない。つまり、仮説 H_1 は正しい。

4 帰無仮説H₀,対立仮説H₁を立てる。

帰無仮説 H_0 :含有量は加熱あり/なしで変わらない。 $\mu_X = \mu_Y$ 対立仮説 H_1 :加熱すると含有量が減少する。 $\mu_X > \mu_Y$ →帰無仮説 H_0 が成り立つとして、計算を進めます。 片側検定で行い、有意水準は5%とします。 ※有意水準α・・・仮説を検証するための判断基準(帰無仮説 H_0 を棄却する基準)

5 統計検定量T を求める。

2 で求めた値, m = n = 10(データ数)を代入する。帰無仮説 H_0 より, $\mu_X - \mu_Y = 0$ とします。

$$T = \frac{\left(\overline{X} - \overline{Y}\right) - (\mu_X - \mu_Y)}{\sqrt{\frac{\widehat{\sigma_X}^2}{m} + \frac{\widehat{\sigma_Y}^2}{n}}} =$$
(1)

(小数第3位四捨五入)

J	K	L	М	N
9	10	平均	不偏分散	
3.15	3.07	3,201	0.008277	
2.96	2.98	2.96	0.0044	
	T =	=(L2-L3)/	SORT(M2/	10+M3/

6 棄却域を設定する。

統計検定量T は帰無仮説Hoが正しいとすれば、自由度df のt分布にしたがいます。 ただし、自由度df の値は次のように計算します。



|7| 帰無仮説が正しいかどうかを検証する。|重要|

1 横軸上に(4)の値をプロットし,横軸に垂直に線分を引く。

①の線分、グラフ、横軸で囲まれた部分を塗る。これが棄却域である(斜線部分の確率は5%)。 (2)



横軸に上(1)の値をプロットする。 (3)

であるので, 棄却域に含まれます。したがって, 帰無仮説H₀は棄却され, 対立仮説H₁が採択されま す。すなわち、加熱なしの場合のビタミンCの含有量と、加熱ありの場合のビタミンCの含有量の平 均には、有意な差があることがわかります。

Excel を利用した t 検定②

1分析ツールを読み込む

8 5.0	e ÷ ≑							Book1 - Exc	æl
ファイル ホー	ム挿入	ページ レイアウト	数式	データ 校園	1 表示	へげ	JUST PDF 4	♀ 実行したい	作業
* *	游ゴシック	* 11	* A* A*	===	89/ +	む 折り返し	に全体を表示する	標準	
貼り付け	В <u>I</u> <u>U</u> ∗	🖽 • 💩 • 🛕	- 2 -		€≣ ∋≣	目もルを結	洽して中央揃え *	5 - % ,	•.0
クリップボード ちょ		フォント	. IS			配置	្តត	数値	

① [ファイル]タブをクリック



H 5 . 12 . 3	feed OxToria	9084	. Excel	2	< 0000054			
Bound 2 . B 1 B - E -	都本設定 新式 オーロ	Microsoft Office 07/1/204-204817	-44°		ųм	27 8×1022 2839-	P HRT- HRT-	
990/25-4 3 222 A1 A B C 1 1 1 2 198/A/6/L 3.221 3 198/A/6/L 3.221 3 198/A/6/L 3.221 5 6 5	スペロジ 何符 日記 日本時行 日本時行 日本時行 日本時行 日本時行 日本時行 日本時行 日本時行	2017 2017 (ロアルウーラロンドトク) 2017 (ロアロンドトク) 2017 (ロアロンドトク)	Earl Call, Africe and an Africa and Africa and Africa Call, 2007 Control and Africa and Africa Africa Call, 2007 Control and Africa Africa Africa Call, 2004 Africa Africa Africa Africa Africa Africa Africa Africa Africa Africa Africa Africa Africa Call, 2004 Africa Africa Africa Africa Africa Call, 2004 Africa Africo Africa Africa Africa Africa Af	1000 2F42 1000 2F42 1000 2F40 000 2F40 000 2F40 000 2F40 000 2F40 000 2F40 000 2F40 200 2F40 307 2F40 307 2F40 307 2F40		0	3	
8 9 1 2 3 3 44 5 5 5		9272-1-1-148 P13-32514887P10280324 P14320424 P14320424 P14320424 P14320424 P143204 P143	CALLEROUNLERY/MAKENALINEALISE	4				
7 8 Sheet: Shar		RC(A): Lood (Fr) Ris(g)		× ====================================				1
as7 • A B 🛐 🗴	1 🖪 🛷			10	14 IV	~ ≈ 6	• • • ₂₀	100% 16:42 20:01/10

③[アドイン]→[設定]をクリック



④[分析ツール]チェックボックスをオンにして[OK]をクリック

2 読み込んだ分析ツールを利用して t 検定を行う



▲▲▲▲▲▲ ①[データ]タブ→[分析ツール]をクリック

H																	
				19 200	3-9 0	ND 75		MAT FOR 4	۶F	RICCORRENDUTE							
1070 M	0 7≉31 0 web: 0 11 7−71	FREERCRY S 195 FREERCRY 17-90/RHS	ne 7, and 1) and 1) 10 and 10 and 10 and	50.7-7. 1655	147 Days 147 Days 147 Days 148 Days	19 9933 777 18 2240	112 11411 7 11411 7	¥ € 169- € €	ut o		01 8-8 6 52- 5863 - 12 5 -9 7-6	18 	921 Wint-27-3987	第のループル 第回から一の副 シート 目やは アウトライン		ान 7-998 न अस	
91			$= \sqrt{-\beta_i}$														
4	A	в	c	0	ε	F	5	H	li -	JK	L	M	N	-	-	R	
L		. 1	2		3 4	9	6	71			10.干灯	不偏分散					
Į.	目的なし	3.21	3.34	3.20	5 3.22	3.15	3.21	3.08	3.3	7-9分析				1 ×			
	的机石川	2.84	2.97	2.91	1 3.02	3.05	2.91	2.92	3.04	2 30 10 10 - 3(5)			OK			
										ヒストグラム 教会会 ほう				*			
8										礼祝完建				- ALACH		/	
										100	12			65.750	~		
										1975/5	100001-174	ugoest.					
2										1 11 21 2	分記を約定した	2.0012.2012	-				
1										2 後老: 2	846387498	横支		5			
1													/				
1												1					
1											_						
3																	
8																	
5																	
		Change	Chevera	Charles I.	1 0						1 D.L					-	
	1	Diffeerr	Olicera	oneers	1 1									Hard on the		- C	1000
	1		-	-			-			_			_	10 D 10	-		100
\pm		121	1 12	-												\$× © .mp	/or/m

②[t 検定:分散が等しくないと仮定した2標本による検定]を選択し、[OK]をクリック

入力元	12		OK
変数1の入力範囲(<u>1</u>):	\$B\$2:\$K\$2	1	OK
変数 2 の入力範囲(<u>2</u>):	\$B\$3:\$K\$3	Ť	キャンセル
- 檀太の平均値の差(日)			ヘルプ(<u>H</u>)
α(<u>A</u>): 0.05			
ロ ラベル(<u>L)</u> a(<u>A</u>): 0.05 出力オプション			
□ ラベル(L) a(<u>A</u>): 0.05 出カオプション ○ 出力先(<u>0</u>):		<u> </u>	
 □ ¬√\(L) a(A): 0.05 出力オプション ○ 出力先(Q): ⑨ 新規ワークシート(P): 	t検定	Ì	

③変数1の入力範囲:加熱なしのデータ範囲
 ④変数1の入力範囲:加熱ありのデータ範囲
 ⑤α:有意水準5%の検定
 ⑥出力オプション:t検定という名前のシート
 に結果を出力する

→[OK]をクリック

検定結果

3

	A	В	С	D	
1	t-検定: 分散が等し	くないと仮	定した2標	本による検	定
2					
3		変数 1	変数 2		
4	平均	3.201	2.96		
5	分散	0.008277	0.0044	•	
6	観測数	10	10		
7	仮説平均との差異	0			
8	自由度	16			
9	t	6.768844			
10	P(T<=t) 片側	2.25E-06			
11	t 境界値 片側	1.745884			
12	P(T<=t) 両側	4.51E-06			
13	t 境界值 両側	2.119905			

太枠の数値を記入し, Excel を利用し た t 検定①で求めた数値と一致するこ とを確かめよう。

- 分散:標本を元に計算された、母集団の分散の推定値(不偏分散)
- ② 自由度:分散が等しくないと仮定した2標本による検定(Welch 検定)のため、正確には小数点以 下も数値があるが、Excel上では四捨五入された整数値が表示される。
- ③ P(T<=t)片側:帰無仮説が真であるとしたときに、このデータが得られる確率
 2.25E-06 = 2.25×10⁻⁶ = 0.00000225(= 0.000225%)
 有意水準 0.05(5%)より小さくなっています。つまり、2つのデータの平均が等しいという仮説の
 域に入らない(棄却域に入る)ため、帰無仮説は棄却され、2つのデータの平均には有意な差があることがわかります。
- ④ t境界値片側:片側検定で有意差が認められる境界値
 1.745884 に対して、与えられたデータから計算したt値が 6.7888844 となり、境界値よりも大きくなっています。つまり、2つのデータの平均値が等しいという仮説の域に入らない(棄却域に入
 - る)ため、帰無仮説は棄却され、2つのデータの平均には有意な差があることがわかります。

