



総務省統計局

社会人のためのデータサイエンス演習

第4週：課題の補講 Excel分析演習（各検定の実行方法）

講師名：菅 由紀子

講座内容

1

- 課題の補講 Excel分析演習 回帰分析による予測モデル作成手順

2

- 課題の補講 Excel分析演習 各検定の実行方法

※Microsoft® , Microsoft® Office Excel® は、米国 Microsoft Corporation の、米国およびその他の国における登録商標または商標です。

※本資料は、Microsoft Corporation と提携しているものではなく、また、Microsoft Corporationが許諾、後援、その他の承認をするものではありません。

※本資料の本文では、©、®、™などの表記は割愛いたします。

※本資料ではデータ分析ツールとして Microsoft® Office Excel®2019を利用しています。

Microsoft® Office の他のバージョンや他の分析ツールを利用している場合は、ヘルプやインターネットなどで各自で調査し、該当機能に置き換えて参照してください。

1. 仮説検定とは（4-6 おさらい）

- 仮説に対して、正しいか否かを統計学的に検証する手法
- 統計的な判断として、差に意味があるかを明らかにする

有意差とは

立てた仮説と結果の差について、統計的に意味があるものを「有意差」という。統計調査などによって得られた2つの値の差が、統計的に信頼できるものか、偶然のものかどうかを判定する方法を有意差検定という。

ミュージカル鑑賞の1年間の平均支出金額

男性・女性それぞれ100人に対するアンケート



男女の結果に差があるといえる



男女の結果差はある・・・？

2.t検定による母平均の差の検証

ミュージカル鑑賞への1年間の支出の男女の平均差が統計的に有意なものなのか検定します。

	A	B	C	D
1	男女別ミュージカル鑑賞への1年間の支出（円）			
2				
3	男	女		
4	15,000	10,000		
5	0	20,000		
6	20,000	9,000		
7	9,000	0		
8	12,000	12,000		
9	25,000	24,000		
10	19,000	20,000		
11	12,000	12,000		
12	0	0		
13	13,000	10,000		

帰無仮説

男女の支出の平均は等しい

対立仮説

男女の支出の平均は異なる

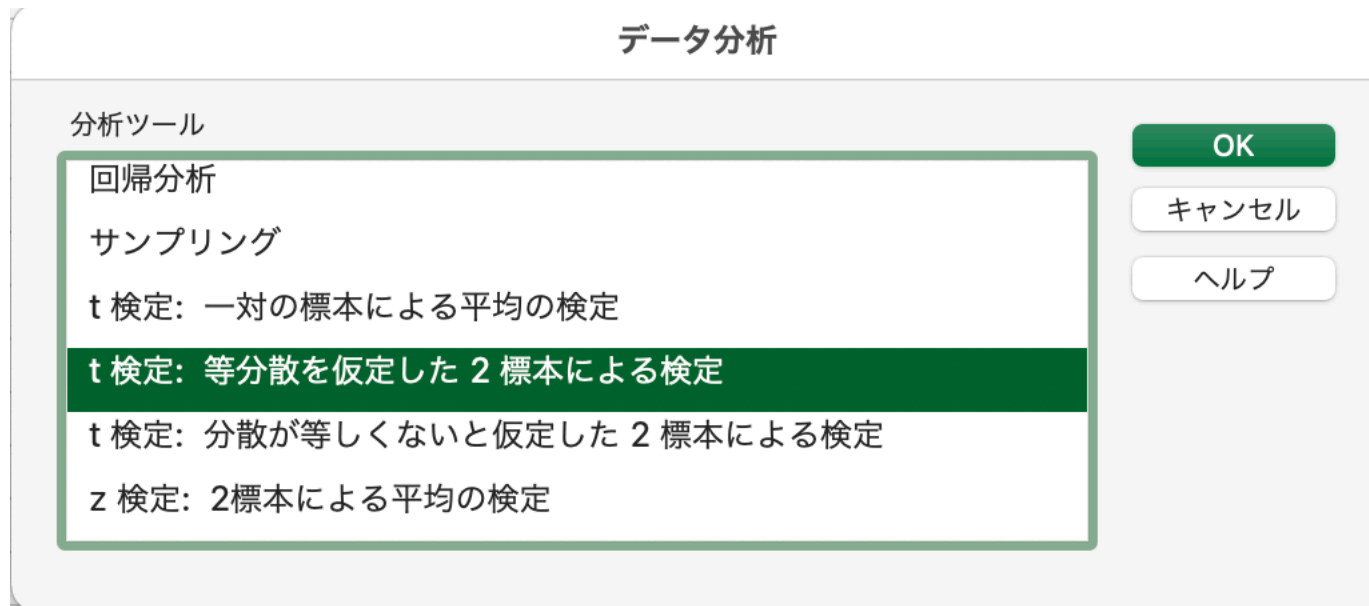
■ 演習ファイル名
hokou_data_W42.xlsx
■ ワークシート名
t検定

2. t検定による母平均の差の検証

1) ツールバーからデータを選択。一番右にある「データ分析」をクリック



2) 分析ツールから「t検定：等分散を仮定した2標本による検定」を選択



2. t検定による母平均の差の検証

3) データの範囲を指定

The screenshot shows the 't-Test: Two-Sample Assuming Equal Variances' dialog box. Three callouts provide instructions:

- ① データの範囲を指定: Points to the input range fields for Variable 1 (\$A\$4:\$A\$103) and Variable 2 (\$B\$4:\$B\$103).
- ② 仮説平均との差異には"0"を入力: Points to the 'Hypothesized Mean Difference' field, which contains the value 0.
- ③ 有意水準は0.05を使用: Points to the 'Alpha' field, which contains the value 0.05.

Other visible fields include 'Labels' (unchecked), 'Output Options' (New Worksheet selected), and buttons for OK, Cancel, and Help.

4) t検定の結果が出力

t-検定: 等分散を仮定した 2 標本による検定		
	変数 1	変数 2
平均	1600	1670
分散	19570707.1	17541515.2
観測数	100	100
プールされた分散	18556111.1	
仮説平均との差異	0	
自由度	198	
t	-0.11490517	
P(T<=t) 片側	0.45431837	
t 境界値 片側	1.65258578	
P(T<=t) 両側	0.90863673	
t 境界値 両側	1.97201748	

2. t検定による母平均の差の検証

5) 結果の考察

t-検定: 等分散を仮定した 2 標本による検定		
	変数 1	変数 2
平均	1600	1670
分散	19570707.1	17541515.2
観測数	100	100
プールされた分散	18556111.1	
仮説平均との差異	0	
自由度	198	
t	-0.11490517	
P(T<=t) 片側	0.45431837	
t 境界値 片側	1.65258578	
P(T<=t) 両側	0.90863673	
t 境界値 両側	1.97201748	

- 平均の差は70円
- $P(T \leq t)$ 両側 = 0.908... > 有意水準 (0.05)
- 「男女の支出の平均は等しい」という帰無仮説を採択

3. カイ 2 乗検定:適合度の検定

サイコロの出目の観測値が理論値に当てはまるのなのか検定する

サイコロの理論値と観測値						
	1	2	3	4	5	6
理論値	500	500	500	500	500	500
観測値	499	485	497	543	489	487

帰無仮説

観測値は理論値に当てはまる

対立仮説

観測値は理論値に当てはまらない

■ 演習ファイル名
hokou_data_W42.xlsx
■ ワークシート名
適合度検定

3. カイ 2 乗検定:適合度の検定

サイコロの理論値と観測値						
	1	2	3	4	5	6
理論値	500	500	500	500	500	500
観測値	499	485	497	543	489	487
=CHISQ.TEST(B5:G5,B4:G4)						

=CHISQ.TEST(観測値のデータ範囲,理論値のデータ範囲)

サイコロの理論値と観測値						
	1	2	3	4	5	6
理論値	500	500	500	500	500	500
観測値	499	485	497	543	489	487
0.44740205						

p値=0.447...>0.05
→帰無仮説を採択
(サイコロは一樣である)

4. カイ 2 乗検定:独立性の検定

性別と好きな教科は独立であるか検定します。

性別と最も好きな教科のクロス集計表							
	国語	数学	理科	社会	英語	合計	
男	8	11	14	8	9	50	
女	10	9	16	10	5	50	
合計	18	20	30	18	14	100	

帰無仮説

性別と好きな教科は独立である

対立仮説

性別と好きな教科は独立でない

■ 演習ファイル名
hokou_data_W42.xlsx
■ ワークシート名
独立性検定

4. カイ 2 乗検定:独立性の検定

期待度数の計算

$$\text{期待度数} = \frac{\text{横計} \times \text{縦計}}{\text{全合計}}$$

$$\frac{50 \times 18}{100} = 9$$

期待度数						
	国語	数学	理科	社会	英語	合計
男	9	10	15	9	7	50
女	9	10	15	9	7	50
合計	18	20	30	18	14	100

縦計

横計

全合計

4. カイ 2 乗検定:独立性の検定

性別と最も好きな教科のクロス集計表						
	国語	数学	理科	社会	英語	合計
男	8	11	14	8	9	50
女	10	9	16	10	5	50
合計	18	20	30	18	14	100
期待度数						
	国語	数学	理科	社会	英語	合計
男	9	10	15	9	7	50
女	9	10	15	9	7	50
合計	18	20	30	18	14	100
=CHISQ.TEST(B3:F4,B10:F11)						

=CHISQ.TEST(観測値のデータ範囲,理論値のデータ範囲)

男	9	10	15	9	7	50
女	9	10	15	9	7	50
合計	18	20	30	18	14	100
0.75035336						

p値=0.75...>0.05
→帰無仮説を採択
(性別と好きな教科は独立である)