

<母平均の差の検定、分散分析>

1. 準備

今回も、実習にはあらかじめ河田が作成したファイルを用いる。

【課題 7】 統計学の講義用 HP から、業種間賃金格差のデータとエアコン価格差のデータのファイルをダウンロードしてみよう。

📖 手順

- ① 検索エンジンで、「河田研究室」と入力し検索すると、「河田研究室」のページにジャンプする。(ここまでの手順は、<http://www2.tokuyama-u.ac.jp/kawada> とアドレスを直接入力してもよい)
- ② 「統計学」をクリックし、「第25回 7月11日(木)」の配布資料にある、「業種間賃金価格差」と「エアコン価格差」をクリックし、自分の使いやすい場所に保存する。

2. 母分散が等しいとみなした場合の母平均の差の検定

【課題 8】 業種間賃金価格差のデータについて、分析ツールを用いて、母分散が等しいとみなした場合の母平均の差の検定をおこなってみよう。

データタブの中に「データ分析」のボタンが出てくる¹ので、下のほうにある、「t 検定: 等分散を仮定した 2 標本による検定」を選び、ウィザードの要求に従ってデータ範囲を指定すれば、下のような表が計算される。

ここで求められた分散は、標本不偏分散(nで割る代わりにn-1で割ったもの) \hat{s}_1^2 , \hat{s}_2^2 である。

プールされた分散 (母分散の不偏推定量) は、

$$\hat{\sigma}^2 = \frac{(n_1 - 1)\hat{s}_1^2 + (n_2 - 1)\hat{s}_2^2}{n_1 + n_2 - 2}$$

となることを確認してみる。

D7セルに $=((B6-1)*B5+(C6-1)*C5)/(B6+C6-2)$ と入力すると、B7セルの値と同じ値が計算される。

D10セルに $=(B4-C4)/SQRT(B7*(1/B6+1/C6))$ と入力すると、B10セルと等しい値、検定仮説が正しいとみなした t 統計量が計算される。これを t 境界値と比較することによって、2業種間の賃金に差があるかどうかの検定がおこなわれる。(p 値を用いる方法もある)

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	6	5.7
分散	0.071111	0.026667
観測数	10	10
プールされた分散	0.048889	
仮説平均との差異	0	
自由度	18	
t	3.033899	
P(T<=t) 片側	0.003569	
t 境界値 片側	1.734064	
P(T<=t) 両側	0.007139	
t 境界値 両側	2.100922	

¹分析ツールを最初に使用する場合には、アドイン(有効にすること)しなくてはならない。分析ツールのアドインは次のようにおこなう。

- ① 「ファイル」のタブをクリックし、下にある「オプション」のボタンをクリックする。
 - ② 「Excel のオプション」のウィンドウが開くので、左側の「アドイン」をクリックする。
 - ③ 一番下に表示される「Excel アドイン」の右の設定ボタンを押す。
 - ④ 「分析ツール」にチェックをつけ、OK ボタンをクリックする。
- すると、データタブの中に「データ分析」のボタンが出てくるので、回帰分析を選べばよい。

3. 等分散の検定

【課題 9】 業種間賃金価格差のデータについて、分析ツールを用いて、母分散が等しいかどうかの検定をおこなってみよう。

「データ分析」のボタンを押し、中ほどにある、「F 検定: 2 標本を使った分散の検定」を選び、ウィザードの要求に従ってデータ範囲を指定すれば、下のような表が計算される。

観測された分散比は、

$$F = \frac{\frac{n_1}{n_1 - 1} s_1^2}{\frac{n_2}{n_2 - 1} s_2^2} = \frac{\hat{s}_1^2}{\hat{s}_2^2}$$

となることを確認してみる。

D8セルに **=B5/C5** と入力すると、B8セルの値と同じ値が計算される。これをF境界値と比較することによって、2業種間の賃金の分散が等しいとみなしてよいかどうかの検定がおこなわれる。(p値を用いる方法もある)

F-検定: 2 標本を使った分散の検定		
	変数 1	変数 2
平均	6	5.7
分散	0.0711111	0.0266667
観測数	10	10
自由度	9	9
観測された分散比	2.666667	
P(F<=f) 片側	0.080073	
F 境界値 片側	3.178893	

4. 母分散について全く分からない場合の母平均の差の検定

【課題 10】 業種間賃金価格差のデータについて、分析ツールを用いて、母分散について全く分からない場合の母平均の差の検定をおこなってみよう。

「データ分析」のボタンを押し、下のほうにある、「t 検定: 等分散を仮定した 2 標本による検定」を選び、ウィザードの要求に従ってデータ範囲を指定すれば、下のような表が計算される。

ここで求められるt統計量は、分散が等しいと仮定した場合と同じであるが、自由度が異なる。

自由度が、

$$v = \frac{\left(\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}\right)^2}{\frac{(s_1^2 / (n_1 - 1))^2}{n_1 - 1} + \frac{(s_2^2 / (n_2 - 1))^2}{n_2 - 1}}$$

$$= \frac{\left(\frac{\hat{s}_1^2}{n_1} + \frac{\hat{s}_2^2}{n_2}\right)^2}{\frac{(\hat{s}_1^2 / n_1)^2}{n_1 - 1} + \frac{(\hat{s}_2^2 / n_2)^2}{n_2 - 1}}$$

となることを確認してみる。

D8セルに **=(B5/B6+C5/C6)^2/((B5/B6)^2/(B6-1)+(C5/C6)^2/(C6-1))** と入力すると、B8セルの値と近い値が計算される。(自由度としては一番近い整数をとった)

この自由度に対応したt境界値とt統計量とを比較することによって、2業種間の賃金に差があるかどうかの検定がおこなわれる。(p値を用いる方法もある)

t-検定: 分散が等しくないと仮定した2標本による検定		
	変数 1	変数 2
平均	6	5.7
分散	0.0711111	0.0266667
観測数	10	10
仮説平均との差異	0	
自由度	15	
t	3.033899	
P(T<=t) 片側	0.004187	
t 境界値 片側	1.75305	
P(T<=t) 両側	0.008374	
t 境界値 両側	2.13145	

5. 分散分析法

【課題 11】 エアコン価格差のデータについて、分析ツールを用いて、分散分析法をおこなってみよう。

「データ分析」のボタンを押し、一番上にある、「分散分析: 一元配置」を選び、ウィザードの要求に従ってデータ範囲を指定(入力範囲を A1:D10 とし、「先頭行をラベルとして使用」にチェックをつける)すれば、下のような表が計算される。

分散分析: 一元配置						
概要						
グループ	標本数	合計	平均	分散		
A	7	52	7.428571	0.085714		
B	9	67.7	7.522222	0.064444		
C	7	53.5	7.642857	0.05619		
D	9	65.3	7.255556	0.052778		
分散分析表						
変動要因	変動	自由度	分散	観測された分散	P-値	F 境界値
グループ間	0.650481	3	0.216827	3.393213	0.031617	2.946685
グループ内	1.789206	28	0.0639			
合計	2.439688	31				

観測された分散比が、グループ間分散をグループ内分散で割ったものであり、これを F 境界値と比較することによって、グループ間で差がないかどうかを検定する。