TA報告（5）
実証会計学
「2グループ間の差異の検定」

藤井ゼミ サブゼミ
5/24(Fri) @106演習室
京都大学大学院 経済学研究科
博士後期課程 1回生
渡邉 誠士

報告内容

2グループ間の差異の検定

- 分散の検定
  - 2つの（標本）グループの属する母集団の分散に差異があるかどうかの検定。
  - ⇒ F検定
  - （注意）3（以上の）グループ間の平均差の検定もF検定というので注意。
- 平均差の検定
  - 2つの（標本）グループが属する母集団の平均値に差異があるかどうかの検定。
  - 2つのグループの分散が等しい ⇒ t検定
  - 2つのグループの分散が異なる ⇒ Welchのt検定
- 独立性の検定
  - 2つの事象が独立して起こっているのかどうかの検定。
  - ⇒ χ^2検定

グループ間比較の例

- 近代企業は所有（株主）と経営（経営専門家）が分離しているものであるとされているが，東証1部上場企業においても多くの非分離企業が存在している。これは非分離企業の優れたパフォーマンスゆえに許されるものである。
- 仮説（理論モデル）
  - 日本の製造業において，創業者一族が株式を所有し経営に関与している企業群と，所有と経営の分離が行われている企業群とでは，異なった性質がある。
- 操作可能化
  - 社長・会長・代表権のある相談役に創業者一族がついており，創業者一族によって最大株主グループを形成している企業を「ファミリー企業」と定義する。
  - パフォーマンスは産業調整済みROAによって測定する。
- 作業仮説（実証モデル）
  - ファミリー企業のROAの分散は非ファミリー企業のROAの分散よりも小さい。
  - ファミリー企業のROAの平均は非ファミリー企業のROAの平均よりも大きい。
  - ファミリー企業はROAが高い傾向にある。

分散の検定

- A,B2つのグループ間に分散の差異があるかどうかの検定
  - 直感的には，Aの分散を$s_A{}^2$，Bの分散を$s_B{}^2$とするとき，$\frac{s_A{}^2}{s_B{}^2}$が1であるといっていいかどうかを検定する。
  - この$\frac{s_A{}^2}{s_B{}^2}$は自由度(n_A-1, n_B-1)のF分布に従う。よって，前回の平均値の検定（t検定）と同様の検定をF分布で行う。
  - 帰無仮説H0：$\sigma_A{}^2 = \sigma_B{}^2$　実際には$\frac{\sigma_A{}^2}{\sigma_B{}^2} = 1$
  - 対立仮説H1：$\sigma_A{}^2 \neq \sigma_B{}^2$　実際には$\frac{\sigma_A{}^2}{\sigma_B{}^2} \neq 1$

分散の検定の例

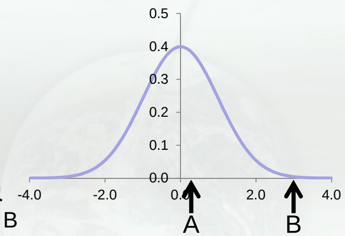
- 「ファミリー企業のROAの分散は非ファミリー企業のROAの分散よりも小さい。」
  - ファミリー企業のROA（128社）の分散は0.001699
  - 非ファミリー企業のROA（751社）の分散は0.002482
- Excelのデータ→分析→データ分析から「F検定：2標本を使った分散の検定」を選択し，比較したい2グループのデータを指定。
  - 有意。

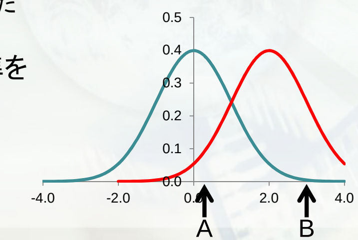
F-検定: 2 標本を使った分散の検定

	変数 1	変数 2
平均	0.005054	-0.000846
分散	0.001699	0.002482
観測数	128	751
自由度	127	750
観測された分散比	0.684467	
P(F<=f) 片側	0.004195	
F 境界値 片側	0.791208	

平均差の検定のイメージ

- 例えば，あることがらが，右のような確率分布で起こるという仮定を置く。
- 実際に起こったのはAとBであったとき，
  - この分布（仮定）のもとにこの事象（標本）が起こったと考えてもよいのか？
  - それとも下の図のように2つの分布を持つ事象があり，Aは青（左）の分布を持つ母集団から，Bは赤（右）の分布を持つ母集団から抽出された事象（標本）という可能性があるのか？
- この問題に統計的に合理的な判断基準を与えるものが差の検定である。





平均差の検定（等分散）

- それぞれのグループの（標本）平均を計算する ⇒ $\overline{X_A}$，$\overline{X_B}$
- 平均差の検定を行う前に，等分散の検定を行い，帰無仮説が棄却できなかった場合。

注）等分散という仮定を棄却できなかっただけであり，等分散だと言っているわけではないことに留意。

- 以下の帰無仮説に基づいて，t検定を行う。
  - 帰無仮説H0：$\mu_A = \mu_B$　実際には$\mu_A - \mu_B = 0$
  - 対立仮説H1：$\mu_A \neq \mu_B$　実際には$\mu_A - \mu_B \neq 0$
  - 統計量：$t_0 = \frac{\overline{X_A} - \overline{X_B}}{\sqrt{U_e\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}}$
  - $U_e = \frac{(n_A-1)s_A{}^2 + (n_B-1)s_B{}^2}{n_A+n_B-2}$
  - t_0が自由度$n_A + n_B - 2$のt分布に従う。

平均差の検定(等分散)の例

- 「ファミリー企業のROAの平均は非ファミリー企業のROAの平均よりも大きい。」
  – ファミリー企業のROA(128社)の平均は0.00505
  – 非ファミリー企業のROA(751社)の平均は-0.00085
- Excelのデータ→分析→データ分析から「t-検定: 等分散を仮定した2標本による検定」を選択し，比較したい2グループのデータを指定。
  – **有意とはならず。**

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	0.00505	-0.00085
分散	0.00170	0.00248
観測数	128	751
プールされた分散	0.00237	
仮説平均との差異	0	
自由度	877	
t	1.26777	
P(T<=t) 片側	0.10261	
t 境界値 片側	1.64659	
P(T<=t) 両側	0.20522	
t 境界値 両側	1.96267	

平均差の検定(異分散)

- それぞれのグループの(標本)平均を計算する ⇒ $\overline{X_A}$, $\overline{X_B}$
- 平均差の検定を行う前に，等分散の検定を行い，帰無仮説が棄却された場合。
- 以下の帰無仮説に基づいて，t検定を行う。
  – 帰無仮説H0: $\mu_A = \mu_B$ 実際には$\mu_A - \mu_B = 0$
  – 対立仮説H1: $\mu_A \neq \mu_B$ 実際には$\mu_A - \mu_B \neq 0$
  – ただし，通常のt検定ではなく，Welchのt検定を行う必要がある。
  – 統計量: $t_0 = \frac{\overline{X_A} - \overline{X_B}}{\sqrt{\frac{s_A{}^2}{n_A} + \frac{s_B{}^2}{n_B}}}$
  – t_0が自由度νのt分布に従う。※ νについてはここで詳細は述べません。

平均差の検定(異分散)の例

- 「ファミリー企業のROAの平均は非ファミリー企業のROAの平均よりも大きい。」
  – ファミリー企業のROA(128社)の平均は0.00505
  – 非ファミリー企業のROA(751社)の平均は-0.00085
- Excelのデータ→分析→データ分析から「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，比較したい2グループのデータを指定。
  – **片側検定で有意となった。**

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	0.00505	-0.00085
分散	0.00170	0.00248
観測数	128	751
仮説平均との差異	0	
自由度	196	
t	1.44910	
P(T<=t) 片側	0.07445	
t 境界値 片側	1.65267	
P(T<=t) 両側	0.14891	
t 境界値 両側	1.97214	

χ^2検定

- 2つの変数間の独立性を検定するときに使える検定。
- 簡単化のため，2×2のマトリクスに表現できる場合を考える。
  – 帰無仮説H0:ファミリー企業であるか否かはROAが業界平均を超えるか否かと独立
  – 対立仮説H1:ファミリー企業であるか否かはROAが業界平均を超えるか否かと独立でない

実測値 サンプル数T	ROA 業界平均以上	ROA 業界平均以下	周辺確率
ファミリー	x	y	γ=(x+y)/T
非ファミリー	z	w	δ=(z+w)/T
周辺確率	α=(x+z)/T	β=(y+w)/T	

期待値 サンプル数T	ROA 業界平均以上	ROA 業界平均以下
ファミリー	αγT	βγT
非ファミリー	αδT	βδT

– $\chi^2 = \frac{(x-\alpha\gamma T)^2}{\alpha\gamma T} + \frac{(y-\beta\gamma T)^2}{\beta\gamma T} + \frac{(z-\alpha\delta T)^2}{\alpha\delta T} + \frac{(w-\beta\delta T)^2}{\beta\delta T}$
– χ^2が自由度1のχ^2分布に従う。 自由度=(列の数−1)×(行の数−1)

χ^2検定の例

- 「ファミリー企業であるか否かはROAが業界平均を超えるか否かに影響を及ぼす」

実測値 サンプル数879	ROA 業界平均以上	ROA 業界平均以下	周辺確率
ファミリー	65	63	0.146
非ファミリー	332	419	0.854
周辺確率	0.452	0.548	

期待値 サンプル数879	ROA 業界平均以上	ROA 業界平均以下
ファミリー	57.8	70.2
非ファミリー	339.2	411.8

  - Excelの分析ツールではできないが，上記の表を作り，CHISQ.TEST(実測値領域,期待値領域)を利用することで検定（p値が求まる。）が可能。
  - χ^2 =1.908
  - 自由度1の5%有意水準はχ^2 =3.841なので，帰無仮説（ファミリー企業であるか否かはROAが業界平均を超えるか否かと独立である）は棄却できない。

結果の頑健性

- 本日紹介したように，同じような2グループ間の変数の特性の差を検定する方法が多数あります。
- これらを組み合わせることによってより頑健な（robust）結果となります。
- 今回紹介した「ファミリー企業のROA」については，頑健性（robustness）に問題があると言えるでしょう。
- Eviews等では，さらに，中央値の差の検定もあり，研究の際には時間が許す限りさまざまな方法を試してみるということが必要となります。

次回までに…

- いよいよ来週はみなさんに何か仮説検定をしてきてもらいたいと思います。
- 用意してもらいたいもの。
  - 理論仮説（直感的なもので結構です。）
  - 作業仮説
  - サンプルの説明
  - 記述統計量
  - 検定の結果（ExcelやEviews等で出てきたものをそのままで結構です。）
  - 結論
- では，楽しみにしています。