

TA報告(6) 実証会計学 「回帰分析」

藤井ゼミ サブゼミ
6/7(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
渡邊 誠士

1

報告内容

1. 回帰分析
2. 単回帰分析
 1. 例
 2. Excellによる単回帰分析
 3. 結果
3. 重回帰分析
 1. 例
 2. 結果
 3. ダミー変数
4. 時系列回帰分析
 1. 例
 2. 結果
5. 回帰分析における留意点
 1. 不均一分散
 2. 多重共線性
 3. 系列相関
6. 次回予告

2

回帰分析

- 回帰分析とは、2変数X, Yのデータがあるとき、YをXで定量的に説明する回帰式というものを推定することである。
 - Y: 従属変数, 被説明変数, 内生変数
 - X: 独立変数, 説明変数, 外生変数
- アーカイバル分析上の分類
 - **クロスセクション分析**
 - 同時点の各企業のデータを用いて回帰分析を行う。
 - 時系列分析
 - 同一企業の各時期のデータを用いて回帰分析を行う。
- 統計手法上の分類
 - **単回帰分析**
 - 説明変数が1つの場合の分析。
 - **重回帰分析**
 - 説明変数が2つ以上の場合の分析。
 - 非線形回帰分析

3

単回帰分析

- 1つの説明変数によって、被説明変数を定量的に説明する以下の回帰式を推定する。
 - 推定した式を、 $y = \alpha + \beta x$ とする
- しかしながら、実際のデータをそれぞれ代入した場合にはこの等号は成り立たない。よって、各データ(i)によって $y_i = \alpha + \beta x_i + \varepsilon_i$ となる。
 - ここで、 ε_i は誤差項であり、 ε_i^2 を最小にするような α, β を推定する。
- ただし、以下のような仮定がおかれていることに留意。
 1. 説明変数 x_i は確率変数ではない。
 2. 誤差項 ε_i は確率変数で期待値は0。
 3. 誤差項どうしは無相関。
 4. 誤差項の分散が一定。(分散の均一性の仮定)

4

単回帰分析の例

- 理論仮説
 - 株価は、将来キャッシュフローの現在割引価値である。また、利益は長期的にみればキャッシュフローと一致する。よって、恒常的な利益によって、株価は決定される。
- 作業仮説
 - 株価は、過去5年間の平均利益によって、説明される。
- サンプル
 - 電気機器業界2008年3月期～2012年3月期
- 回帰式

$$P_i = \alpha + \beta \overline{EPS}_i + \varepsilon_i$$
 - P_i : 企業*i*の株価
 - \overline{EPS}_i : 企業*i*の過去5年間の平均1株当たり利益
- 外れ値の除去, 利益額が負の企業の消去を行い, サンプルサイズ59社

Excelによる単回帰分析

- データ⇒データ分析⇒回帰分析

Excelによる単回帰結果

回帰統計	
重相関 R	0.772117
重決定 R2	0.596165
補正 R2	0.58908
標準誤差	720.3483
観測数	59

モデルの説明力のチェック

分散分析表					
	自由度	変動	分散	観測された分散比	有意 F
回帰	1	43663965	43663965	84.14690142	8.06E-13
残差	57	29577394	518901.6		
合計	58	73241359			

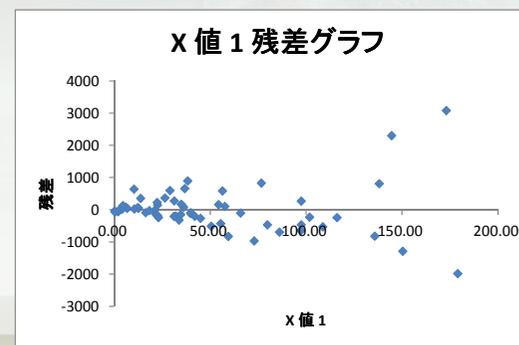
	係数	標準誤差	t	P-値
切片	129.507	141.8023	0.91329	0.36493594
X 値 1	18.9155	2.062047	9.17316	8.0604E-13

係数の妥当性のチェック
帰無仮説: 係数=0

回帰分析における留意点①

不均一分散問題

- 回帰分析において、分散は一定であるという仮定がおかれていたが、これが満たされていない場合がある。⇒ 不均一分散問題
- チェックの方法
 - 残差グラフの利用。
 - Whiteの検定の利用 (Excelにはおそらく無い)
- 不均一分散なら？
 - Whiteの修正の利用 (Excelには無い)



重回帰分析

- 複数の説明変数によって、被説明変数を定量的に説明する以下の回帰式を推定する。
 - 推定した式を、 $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots$ とする
- しかしながら、実際のデータをそれぞれ代入した場合にはこの等号は成り立たない。よって、各データ(i)によって $y_i = \alpha + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \dots + \varepsilon_i$ となる。
 - ここで、 ε_i は誤差項であり、 ε_i^2 を最小にするような $\alpha, \beta_1, \beta_2, \dots$ を推定する。
- ただし、以下のような仮定がおかれていることに留意。
 1. 説明変数 x_i は確率変数ではない。
 2. 誤差項 ε_i は確率変数で期待値は0。
 3. 誤差項どうしは無相関。
 4. 誤差項の分散が一定。(分散の均一性の仮定)
 5. **説明変数行列の階数(ランク)が説明変数の数+1である。**

9

重回帰分析の例

- 理論仮説
 - 株価は、将来の配当の流列の現在割引価値である。クリーンサープラス関係を所与とすると、現在の純資産簿価と将来の超過利益の流列の現在割引価値となる。
- 作業仮説
 - 株価は、現在の1株当たり純資産簿価と、過去5年間の1株当たり純利益の平均で説明される。
- サンプル
 - 電気機器業界2008年3月期～2012年3月期(外れ値等の処理後59社)
- 回帰式

$$P_i = \alpha + \beta_1 \overline{BVPS}_i + \beta_2 \overline{EPS}_i + \varepsilon_i$$
 - P_i : 企業iの株価
 - \overline{BVPS}_i : 企業iの純資産簿価
 - \overline{EPS}_i : 企業iの過去5年間の平均1株当たり利益

10

Excelによる重回帰結果

回帰統計	
重相関 R	0.788924
重決定 R2	0.622401
補正 R2	0.60891
標準誤差	702.7484
観測数	59

モデルの説明力のチェック

分散分析表					
	自由度	変動	分散	観測された分散比	有意 F
回帰	2	45585463	22792732	46.15265299	1.4351E-12
残差	56	27655896	493855.3		
合計	58	73241359			

モデルの妥当性のチェック
帰無仮説: $\beta_1 = \beta_2 = 0$

	係数	標準誤差	t	P-値
切片	-30.086	160.2608	-0.1877	0.8517648
X 値 1	0.35664	0.180802	1.97251	0.05349615
X 値 2	15.6148	2.616662	5.96744	1.7172E-07

係数の妥当性のチェック
帰無仮説: 係数=0

11

回帰分析における留意点②

- **多重共線性問題**
 - 説明変数間に高い相関があると、結果が不安定となる。
 - チェックの方法
 - 相関係数を見る。(0.4が1つの境界)
- | | BVPS | EPS |
|------|---------------|-----|
| BVPS | 1 | |
| EPS | 0.6395 | 1 |
- VIFの利用
 - VIFを調べる係数を被説明変数に、残りの説明変数を説明変数にして回帰を行う。その重決定R2を1から引いたものの逆数がVIFとなる。

12

回帰分析における留意点②

- VIFの利用

- VIFを調べる係数を被説明変数に、残りの説明変数を説明変数にして回帰を行う。その重決定R2を1から引いたものの逆数がVIFとなる。
- VIFのカットオフ基準については10以上といわれる場合もあれば5以上といわれる場合もある。

回帰統計	
重相関 R	0.6395
重決定 R2	0.40896 ⇒ $1/VIF=1-0.40896=0.59104$
補正 R2	0.398591 VIF=1.692
標準誤差	514.8231
観測数	59

	係数	標準誤差	t	P-値
切片	447.4966	101.3442	4.415611	4.55E-05
X 値 1	9.255157	1.473717	6.280147	4.98E-08

13

ダミー変数

- 説明変数は常に量的情報であるとは限らず、質的情報である場合はどのように扱えばよいか。
- 例えば、「企業が東証に上場しているか大証に上場しているか」などといった量的ではなく、質的な要因が被説明変数に影響を与えていると考えられるとき、「0」or「1」となる2項選択の変数(ダミー変数)を入れることがある。
- また、複数年度のデータを扱う際には、年度ごとにダミー変数を入れることによってマクロ要因をコントロールすることがある。

14

時系列回帰分析

- クロスセクション分析(これまで扱った分析)とは、同時点のA社、B社、...といった企業のデータを用いて分析を行う手法。
- 時系列分析とは、長期に渡る1企業のデータを用いて、ある期の数値が、別の期(主にそれ以前の期)の要因によって説明できるかを分析するもの。
 - 推定した式を、 $y_t = \alpha + \beta x_t$ とする(もちろん重回帰の場合もある)
- どちらの分析も今までと同様、回帰分析については以下のような仮定がおかれていることに留意。
 1. 説明変数 x_t は確率変数ではない。
 2. 誤差項 ε_t は確率変数で期待値は0。
 3. 誤差項どうしは無相関。
 4. 誤差項の分散が一定。(分散の均一性の仮定)
 5. 説明変数行列の階数(ランク)が説明変数の数+1である。

15

時系列回帰の例

- 理論仮説
 - 売上高に対しては広告宣伝費が重要な役割を持っており、売上高が広告宣伝費の影響を受けている。
- 作業仮説
 - 当期の売上高は、前期の広告宣伝費の影響を受けている。
- サンプル
 - 花王の1989年3月期～2012年3月期のデータ
- 回帰式
$$SALES_t = \alpha + \beta ADV_t + \varepsilon_t$$
 - $SALES_t$: t期の売上高
 - ADV_t : t期の広告宣伝費

16

Excelによる時系列回帰結果

回帰統計	
重相関 R	0.936084
重決定 R2	0.876254
補正 R2	0.87036
標準誤差	72528.41
観測数	23

分散分析表					
	自由度	変動	分散	観測された分散比	有意 F
回帰	1	7.82E+11	7.82E+11	148.7019791	5.41136E-11
残差	21	1.1E+11	5.26E+09		
合計	22	8.93E+11			

	係数	標準誤差	t	P-値
切片	267395	56850.72	4.70345	0.00012107
X 値 1	1.62181	0.132997	12.1943	5.4114E-11

係数の妥当性のチェック
帰無仮説: 係数=0

17

回帰分析における留意点②

● 系列相関問題

- 時系列分析を行う上で、注意しなければならない問題で、t-1期の誤差項とt期の誤差項の間に相関がある場合には通常最小二乗法が使用できない。
- チェックの方法
 - ダービー・ワトソン検定 (Excelではコマンドは無い)
 - $$DW = \frac{\sum(\varepsilon_t - \varepsilon_{t-1})^2}{\sum \varepsilon_t^2}$$
 - このDWが2から遠ざかると系列相関の恐れあり。(サンプルサイズ、棄却域の設定にもよるが、1.5を下回る(2.5を上回る)と黄色信号、1を下回る(3を上回る)と赤信号となる。)
- 系列相関が認められる場合には、推定方法を変える必要がある。

18

次回までに...

- 来週はみなさんに何か回帰分析(できれば重回帰を希望)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, 相関係数表
 - 検定の結果 (ExcelやEviews等で出てきたものをそのまま結構です。)
 - できれば多重共線性や不均一分散のチェックをお願いします。
 - 結論
- では、楽しみにしています。

19