

Ⅲ 研究ノート Ⅲ

高校数学新課程に対応した 実習形式の統計学の授業内容について

渋谷 綾子

1. はじめに

大学での統計学に関する授業は、対象学生、授業回数、授業形式（実習形式か講義形式か）によってその内容がさまざまに変化する。本研究ノートの著者は文系の学生への半期の実習授業¹⁾、

講義授業²⁾、理系の学生を含むクォーターの講義授業³⁾、大学院での半期の実習授業⁴⁾とさまざまな形態での統計に関する授業を経験してきている。そのうち、本研究

対象学生	授業回数	授業形式
文系	16回	実習
理系	8回	講義

ノートでは、「文系の学部学生を対象とし、**図1 本研究ノートで想定した授業**授業回数は16回、ExcelとSPSSが使用できる実習形式」の授業内容について考察する。この形態は「経営統計」という科目名で、本研究ノートの著者が最も多く経験してきている形態であり、年々その内容を変更してきており、本年度の授業ではほぼ取り入れたいと考えていた内容を取り入れることができたので、その授業内容を本研究ノートでまとめてみたい。

1) 本学経済学部専門科目「経営統計」

2) 本学経済学部「社会科学のためのやさしい統計学」(2単位) および他大学の「統計学」(4単位)

3) 本学共通教育科目「経済データと統計学」(医学部医学科、医学部保健学科、人文学部、教育学部対象)(1単位)

4) 本学経済学研究科の科目「経営数理計画研究」(2単位) でのExcelやSPSSを用いた授業

その際、この授業がVBAによる正誤判定つきの実習シートを使用していることに留意してもらいたい。内容が多いと感じられることもあるかと思うが、実習シートでは、学生が個々のペースで課題を解いていくことができ、また、インターネット上でシートの受け渡しをしているので授業時間内に完成しなかった学生は自宅学習で続きを行い、その場で提出もできる仕組みになっている⁵⁾。実習シートは教室内の8割程度の学生が授業時間内で完成させられるような分量を目指した。これまで、早く仕上げてしまう学生と時間のかかる学生の差が大きく、早く仕上げてしまう学生には物足りない授業かもしれないが、近年は内容の高度化により、早く仕上がってしまう学生は少なくなってきている。なかでも、「t検定」と「分散分析」はようやく本年度から本格的に取り入れることができた⁶⁾。また、平成27年度の新入生より高校の数学で統計に関する内容を学んできているので⁷⁾、高校で学んだことと重複せずに新しい内容を大学で学べるように工夫することにも配慮する。

2. 新課程の高等学校のカリキュラムと大学で初めて学ぶ事項のリストアップ

平成27年度入学生から高校数学で新課程に移行している。新課程の高校数学では「記述統計」を「数学I」、「集合」を「数学A」、「推測統計」を「数学B」に振り分けていると理解してよいと思う。従来の大

	旧課程	新課程
集合	数学A	数学A
記述統計	数学B	数学I
推測統計	数学C	数学B

図2 高等学校数学カリキュラム

図2 高等学校数学カリキュラム
大学の統計学の授業では「記述統計」と「推測統計」に関する内容が主であり、集合については、既知のこととして授業をしていたので留意すべき点としては、「記述統計」が「数学I」に移行したため、大学では、推測統計や

5) 提出期限は授業の1週間後

6) 正規分布の関数は3年ほど前から部分的に取り入れていた

7) 特に記述統計については「数学I」ではほぼ全員が既習

回帰分析などのより高度な内容を扱うことができるようになったということである。また、推測統計に分類される項目のうち、「検定」は高等学校数学Bでも扱っていない。

高校数学新課程での変化を詳しく分析してみよう。統計学に関連する内容は、旧課程では「数学A」,「数学B」,「数学C」で教えられていた。新課程では「記述統計」が「数学I」に移行したことが最も大きな変化といえる。具体的にいえば、新課程の学生に、度数分布、平均、分散、標準偏差、相関係数等の記述統計に分類される項目の説明に時間をかける必要はないということである。特に理系の学生には確認程度でよい。クォーターの場合はさらにこの部分に時間をかけないことが重要になる。

なお、高等学校では、これらの内容をすべて講義で行っており、PCの使用は想定されていないようである。(表計算ソフトウェアを用いた計算方法に触れている部分があるが、あくまで参考程度である。) ちなみに、旧課程と新課程の数学の教科書全体に対する、統計学に関わる内容は14%程度で旧課程と新課程で変化していない。

また、高校では扱われず、大学の統計学の授業で比較的良好に取り扱われる項目は「時系列データの分析(移動平均, 指数平滑, 期別平均法等)」,「質的データ(連関係数)」,「回帰分析」,「平均値の差の検定(t検定)」,「分散分析」,「P値と多重比較」などがあげられる。

表1 高校数学新課程の統計に関わる内容一覧⁸⁾と大学で新しく学ぶべき内容

数学 I

度数分布表とヒストグラム, 中央値, 四分位数, 四分位範囲, 四分位偏差, 箱ひげ図, 分散(標本分散)と標準偏差, データの相関, 散布図, 共分散, 相関係数
--

数学 A

場合の数と確率, 和の法則, 積の法則, 順列・組合せ, 独立な試行, 反復試行, 条件つき確率
--

数学 B

確率分布と統計的な推測, 確率変数と確率分布, 確率変数の平均, 確率変数の分散と標準偏差, 和の平均と分散, 二項分布〔発展〕ポアソン分布 正規分布 統計的な推測, 母集団と標本, 推定
--

大学で新しく学ぶ内容

時系列データの分析(移動平均, 指数平滑, 期別平均法等), 質的データ(連関係数), 回帰分析(単回帰, 重回帰), 平均値の差の検定(t検定), 分散分析, P値と多重比較
--

3. 文系・セメスタ(全16回)・実習の授業スケジュール

前項までで述べてきたことと矛盾するが, 対象が文系の学生の場合は, たとえ, 新課程であっても, 高等学校での数学のカリキュラムが様々であること, 履修してきたはずの内容でも理解が十分でない, などの理由により, 既習の事項でも全く説明しないではすまされないと思われる。パーソナル・コンピュータを使用できることでより理解しやすくなる内容などはしっかり確認するべきである。平成27年度の学部の「経営統計」の授業では(現2年生が旧課程の最後の年代なので, 全員旧課程)5回程度(実習授業回数15回のほぼ3分の1)は記述統計に費やされた。今後はこの授業を講義とする可能性があるが, もし, 実習であるなら, 記述統計に関する授業回数は減らすことが望ましい。その分をt検定や分散分析に時間をかけるべきである。大学で新しく学ぶ内容としては, 時系列データの取り扱い(移動平均, 指数平滑, 期

8) 啓林館高等学校数学科用教科書「数学I」「数学A」「数学B」より

別平均法)を2回の授業で学び、単回帰分析、重回帰分析を各1回ずつとそれらのまとめ1回、計5回をあてた。最後の5回で、正規分布、仮説検定の基礎知識、t検定、分散分析、信頼区間の推定の推測統計、および、卒論等で実施される可能性があると考えられるアンケート集計のためのピボットテーブルを学ぶ。

まとめると、本年度では学部学生を対象としたセメスタの実習では、3分の1が記述統計、3分の1が大学で初めて学ぶ分野、3分の1が推測統計となる。表2に本研究ノートが提案する授業スケジュールを示す。

表2 文系・セメスタ・実習の授業スケジュール

	授業の主目的	備考
1	記述統計 平均, 分散, 標準偏差	表形式で分散や標準偏差の公式どおりに計算する。次に, Excelの関数を学ぶ。標本分散(=var.p)と不偏分散(=var), および対応する標準偏差(=stdev.pと=stdev)を学ぶ。データ分析の基本統計量の実行。
2	記述統計 代表値(幾何平均, 調和平均, 中央値, 四分位数) 二変量(共分散, 相関係数)	幾何平均と調和平均。 SPSSでは箱ひげ図が作れる。[分析]→[記述統計]→[度数分布表]→[統計量] 共分散も分散の説明のように表形式→=covar関数
3	記述統計 度数分布表とヒストグラム	四分位数がExcelとSPSSで異なることを確認。 度数分布表とヒストグラムをExcelとSPSSどちらでも作れるようにする。その際, 階級の分割方法も学ぶ。 偏差値の公式。
4	記述統計 共分散と相関係数	Excelで関数(=covar, =correl)と共に, 「分散共分散行列」と「相関行列」の出力方法と見方を学ぶ。散布図にも言及。
5	記述統計のまとめ インターネットと移動平均	記述統計についてこれまで学んだことをまとめる。 時系列データについては, インターネットから株価データを検索して, 13週と26週の移動平均を求める。
6	時系列データ 指数平滑	[データ]→[データ分析]→[指数平滑]
7	時系列データ 期別平均法 変動係数	季節調整の実際を経験する。

8	単回帰分析	3つの方法で単回帰式を求める（ Σ を使った式の変形と偏微分の知識）。決定係数について。
9	重回帰分析	Excelの「データ分析」とSPSSでの重回帰式の求め方。
10	これまでの復習	できるだけ問題を多く解く
11	正規分布とz検定（2標本による平均の検定）	=norm.dist関数, =norm.inv関数, =norm.s.dist関数, =norm.s.inv関数
12	仮説検定の考え方	仮説検定の考え方（母平均）、帰無仮説と対立仮説、片側検定、両側検定、棄却値、棄却域、有意水準、P値などの基礎概念を学ぶ。 Excelの「データ分析」による仮説検定としてZ検定を実習する
13	平均値の差の検定（t検定）	F検定で等分散か分散が等しくないかを調べ、Excelのデータ分析で検定する ①等分散を仮定した2標本のt検定 ②分散が等しくないと仮定した2標本のt検定 ③一対の標本による平均の検定（F検定不要） 各検定をSPSSでも実施し、結果が一致することを確認する。SPSSでは「分析」→「平均の比較」 ⁹⁾
14	分散分析	①一元配置分散分析 ②繰り返しのある二元配置の分散分析 ③繰り返しのない二元配置の分散分析 をExcelの「データ分析」とSPSSの両方で学ぶ
15	アンケートの集計とピボットテーブル	アンケート集計を例にクロス集計などはアドインを用いずにピボットテーブルとピボットグラフで簡易に分析することを学ぶ
16	筆記試験	定期試験期間内に実施

4. 各授業の内容と留意点

1回目の授業

分類；記述統計

内容：「平均，分散，標準偏差」

統計学を学ぶにあたって、よく用いられる数式に親しむために、あえて最初の授業で Σ を用いた数式を提示する。

$$\text{平均の公式} \quad \bar{x} = \sum x_i / n$$

9) 複数の方法があり、今後結果の検証も合わせてメニュー構造を探索する予定

分散の公式 $\sigma^2 = \sum (x_i - \bar{x})^2 / n$

標準偏差の公式 $\sigma = \sqrt{\sum (x_i - \bar{x})^2 / n}$

のそれぞれの計算過程を Excel シート上で実習する。また、標本分散と不偏分散の違いとそれぞれに対応する関数について、標本分散が =var.p¹⁰⁾、その平方根である標準偏差が =stdev.p、不偏分散が =var、それに対応する標準偏差が =stdev として、表で求めた値と合致するかを確認する。また、Excel の [データ分析]¹¹⁾ の [基本統計量] で「統計情報」にチェックして実行すると、図3の表が出力され、ここの分散と標準偏差は不偏分散とそれにもとづく標準偏差であることを確認する。

日経平均株価	
平均	19789.71
標準誤差	200.39
中央値 (メジアン)	20174.24
最頻値 (モード)	#N/A
標準偏差	961.0367
分散	923591.6
尖度	-0.11498
歪度	-1.05598
範囲	2932.4
最小	17792.16
最大	20724.56
合計	455163.3
標本数	23

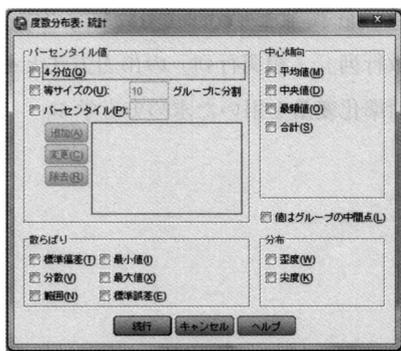
図3 基本統計量の出力

2回目の授業

分類：記述統計

内容：「代表値としての中央値、幾何平均、調和平均、また、それぞれの Excel と SPSS での求め方」

幾何平均 (=geomean) と調和平均 (=harmean) について、使用例を含めて説明。中央値は四分位数と合わせて説明。



また、SPSS の [分析] → [記述 図4 SPSS の [統計量] オプション画面

10) Excel2007以前のバージョンと互換性を持たせるなら、var と p の間のドットは削除して =varp とする

11) 分析ツールのアドイン後、[データ] メニューに [データ分析] が表示される

統計] → [度数分布表] → [統計量] でさまざまな統計量を求める。ここで、SPSS でも分散や標準偏差は不偏であることに注意する。

3回目の授業

分類：記述統計

内容：「中央値，四分位数，度数分布表，ヒストグラム」

四分位数は，データが少ないときは Excel の =quartile 関数と SPSS では異なることを確認¹²⁾。また，Excel も SPSS も階級幅を設定しないと自動的に階級に分割して度数分布表やヒストグラムを出力する。望ましい階級幅を設定する方法も Excel，SPSS それぞれで学ぶ。偏差値の公式も学ぶ。

次回の授業の準備として共分散について，分散の計算方法と関連付けての表形式での求め方と関数での求め方の両方を学ぶ。

4回目授業

分類：記述統計

内容：「共分散と相関係数」

インターネットから株価データを検索して Excel シートに貼り付け，=covar 関数と =correl 関数を適用する。また，[データ分析] で「分散共分散行列」，「相関行列」の出力方法とその見方を学ぶ。相関係数については，標準化変量を用いた求め方も学ぶ。

12) 第一四分位数は SPSS<Excel，第三四分位数は Excel<SPSS

	九州電力	中国電力	東京電力	トヨタ自動車	日産自動車	日経平均株価	TOPIX
九州電力	0.022549						
中国電力	0.016198	0.016531					
東京電力	0.023166	0.010136	0.054219				
トヨタ自動車	0.007587	0.00568	0.009735	0.005302			
日産自動車	0.004422	0.003574	0.007794	0.004706	0.006466		
日経平均株価	0.005458	0.00394	0.008012	0.003418	0.002651	0.002898	
TOPIX	0.00562	0.003943	0.007904	0.003532	0.002723	0.003029	0.003265

	九州電力	中国電力	東京電力	トヨタ自動車	日産自動車	日経平均株価	TOPIX
九州電力	1						
中国電力	0.838955	1					
東京電力	0.662545	0.33855	1				
トヨタ自動車	0.693868	0.606657	0.574194	1			
日産自動車	0.366194	0.345702	0.416269	0.803774	1		
日経平均株価	0.675269	0.56928	0.639219	0.872003	0.612518	1	
TOPIX	0.655006	0.536687	0.594081	0.848875	0.592655	0.98468	1

図5 Excel [データ分析] の分散共分散行列 (上) と相関行列 (下) の出力

5回目授業

分類：記述統計のまとめと (大学で新しく学ぶ) 時系列データの扱い (移動平均法)

内容：「記述統計のまとめと時系列データ (移動平均)」

移動平均法自体は非常に平易なので、インターネットからデータを取得する方法を合わせて学ぶ。データをコピー&ペーストするだけでなく、並べ替える等の処理を施し、13週移動平均と26週移動平均のグラフからゴールデンクロスやデッドクロスを読み取る。

6回目授業

分類：(大学で新しく学ぶ) 時系列データの扱い (指数平滑法)

内容：「指数平滑法」

直近の予測値と実測値から次期の予測値を求める。 x が実測値、 y が予測値である。

$$\begin{aligned}
 y_{t+1} &= \alpha \times x_t + (1 - \alpha) \times y_t \\
 &= y_t + \alpha(x_t - y_t) \quad \text{ただし, } 0 < \alpha < 1
 \end{aligned}$$

Excel の [データ分析] の [指数平滑] を選択。減衰率を指定する。為替

相場のデータを用いて減衰率が大きくなるほど、予測が当たりにくくなることをグラフで確認。

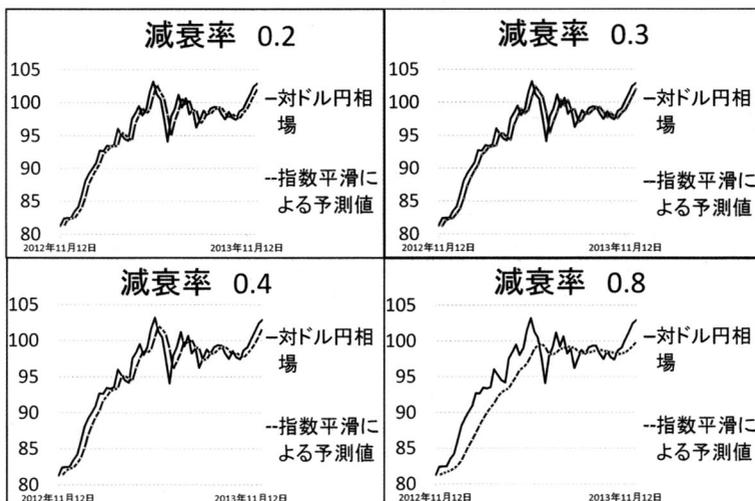


図6 減衰率による予測精度の比較

7回目授業

分類：(大学で新しく学ぶ) 時系列の扱い（期別平均法）と記述統計の復習
（変動係数）

内容：期別平均法と変動係数

季節調整の例として期別平均法を学ぶ。また、ばらつきの比較には変動係数を用いることを学ぶ。

8回目授業

分類：(大学で新しく学ぶ) 単回帰分析

内容：「単回帰分析」

①分散や共分散を用いた公式から単回帰式を求める、②散布図の近似オプションとして単回帰式を求める、③正規方程式から単回帰式を導き出す、の

3種類の方法を学ぶ。③の方法は式の変形や偏微分を用いるところが文系の学生には難しいかと予想したが、よく理解したようであった。

$y = a_0 + a_1x$ として、 a_0 と a_1 を求める問題に帰着させる。3とおりの方法は以下のとおり。

① $a_1 = \frac{x \text{ と } y \text{ の共分散}}{x \text{ の分散}}, a_0 = y \text{ の平均} - x \text{ の平均} \times a_1$

② 散布図を作成して、近似曲線を描くオプションで「グラフに数式を表示する」をチェックする。

③ 残差 (e) の2乗和を最小化する問題とする。

$$\begin{aligned} \phi &= \sum e_i^2 \\ &= \sum (y_i - a_0 - a_1x_i)^2 \\ &= \sum (y_i^2 + a_0^2 + a_1^2x_i^2 - 2a_0y_i + 2a_0a_1x_i - 2a_1x_iy_i) \\ &= \sum y_i^2 + na_0^2 + a_1^2 \sum x_i^2 - 2a_0 \sum y_i + 2a_0a_1 \sum x_i - 2a_1 \sum x_iy_i \\ &= 0 \end{aligned}$$

とおき、この式を a_0 と a_1 で偏微分する。

$$\frac{\partial \phi}{\partial a_0} = 2na_0 - 2 \sum y_i + 2a_1 \sum x_i = 0 \quad \text{より,} \quad \sum y_i = na_0 + a_1 \sum x_i \quad \text{①}$$

$$\frac{\partial \phi}{\partial a_1} = 2a_1 \sum x_i^2 + 2a_0 \sum x_i - 2 \sum x_iy_i = 0 \quad \text{より,} \quad \sum x_iy_i = a_0 \sum x_i + a_1 \sum x_i^2 \quad \text{②}$$

①と②の式を「正規方程式」とよび、この連立方程式を解いて a_0 と a_1 の値を求める。

また、決定係数の意味とそれが相関係数の2乗であることを確認する。

9回目授業

分類：(大学で新しく学ぶ) 重回帰分析

内容：「重回帰分析」

Excel と SPSS で重回帰式の求め、結果が一致することを確認。Excel で

は [データ] → [データ分析] → [回帰分析], SPSS では [分析] → [回帰] → [線型]。例題は「バス」「徒歩」「広さ」「築年数」から家賃を求める問題。次頁の図7と図8で示すように Excel も SPSS も大量のデータが出力される。「係数」と「t 値」の欄を読み取る。また、求められた重回帰式の係数の符号と相関行列での符号が矛盾する場合は説明変数（独立変数）から除外するなどの注意点を伝える。t 値の絶対値が2未満の説明変数は削除した方が良いことなどを説明する。

概要

回帰統計	
重相関 R	0.843085
重決定 R2	0.710792
補正 R2	0.63367
標準誤差	5.15447
観測数	20

分散分析表

	自由度	変動	分散	測された分散	有意 F
回帰	4	979.4716	244.8679	9.21645307	0.00057616
残差	15	398.5284	26.56856		
合計	19	1378			

	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	52.60442	10.38623	5.064822	0.00013971	30.4666868	74.74215	30.46669	74.74215
バス(分)	-1.01826	0.355764	-2.86219	0.0118701	-1.7765569	-0.25997	-1.77656	-0.25997
徒歩(分)	-0.60057	0.314466	-1.9098	0.07547178	-1.2708365	0.069701	-1.27084	0.069701
広さ(m ²)	0.739422	0.180119	4.105193	0.00093635	0.35550816	1.123336	0.355508	1.123336
築年数(年)	-1.01711	0.338255	-3.00694	0.00884688	-1.7380873	-0.29614	-1.73809	-0.29614

図7 Excel の重回帰分析の出力例 (Excel)

係数^a

モデル		標準化されていない係数		標準化係数	t 値	有意確率
		B	標準誤差	ベータ		
1	(定数)	52.604	10.386		5.065	.000
	バス	-1.018	.356	-.577	-2.862	.012
	徒歩	-.601	.314	-.366	-1.910	.075
	広さ	.739	.180	.658	4.105	.001
	築年数	-1.017	.338	-.426	-3.007	.009

a. 従属変数 家賃

図8 SPSS の重回帰分析の出力例

10回目授業 これまでの復習

11回目授業

分類：推測統計（正規分布）（理論的には高等学校の「数学B」に含まれる）

内容：「確率密度関数と =norm.dist, =norm.inv, =norm.s.dist, =norm.s.inv」

確率密度関数 $(f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}})$ はグラフ描画を通して学ぶ¹³⁾、各関数

については入力方法と出力の解釈について学ぶ。

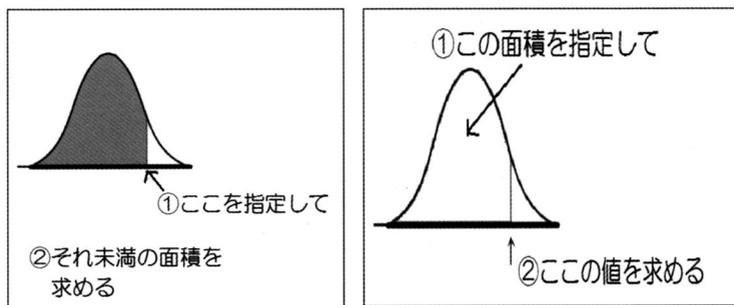


図9 norm.dist 関数（左）と norm.inv 関数（右）のイメージ

12回目授業

分野：推測統計（大学で新しく学ぶ）

内容：「仮説検定の考え方（母平均の仮説検定）」

帰無仮説と対立仮説について、両側検定と片側検定について、有意水準、P値などの基礎概念を学ぶ。新しく =tinv, =tdist 関数も学ぶ。また、Excelの「データ分析」のうち、Z検定を実習する。図10のような出力を得た場合、片側でも両側でもP値が0.05より小さいので母集団の平均に「差がある」と結論する。

13) 本質的な理解にはおよばないが、正規分布の形状については具体的なイメージが形成されやすくなると思われる

z-検定: 2 標本による平均の検定

	A社	B社
平均	12.0028571	12.5914286
既知の分散	0.81852101	0.49551261
観測数	35	35
仮説平均との差異	0	
z	-3.0375956	
P(Z<=z) 片側	0.00119237	
z 境界値 片側	1.64485363	
P(Z<=z) 両側	0.00238474	
z 境界値 両側	1.95996398	

図10 Excel の z 検定の出力

13回目授業

分類：推測統計（大学で新しく学ぶ）

内容：「t 検定」

Excel の [データ] → [データ分析] では

- ① t 検定：一対の標本による平均の検定
- ② t 検定：等分散を仮定した2標本による検定
- ③ t 検定：分散が等しくないと仮定した2標本による検定

の3種が用意されている。実習では②→③→①の順で学ぶ。②と③では t 検定に先立って、分散が等しいか等しくないかを調べるために F 検定を行う。F 検定も [データ分析] のメニューにある。

②等分散を仮定した2標本による検定の例題

「二つの養鶏場の卵の重さに差があるか」を有意水準5%で検定¹⁴⁾。

Excel の [データ分析] では、まず、F 検定を実施。「P(F<=f)片側」の欄が0.05より大きいため (0.366818) 等分散であるとし、[データ分析] のメニューから「t 検定：等分散を仮定した2標本による検定」を実行する。出力の「P(T<=t)片側」「P(T<=t)両側」がともに0.05より小さいので (0.017149

14) ひとつめの養鶏場の10個のサンプルの平均は55.84g, もう一つの養鶏場の8個のサンプルの平均は57.7625g

と0.034298) 片側検定でも両側検定でも二つの養鶏場の卵の重さには「差がある」。

F-検定: 2 標本を使った分散の検定			t-検定: 等分散を仮定した2標本による検定		
	重さ(g)	重さ(g)	重さ(g)	重さ(g)	
平均	55.84	58.7625	平均	55.84	58.7625
分散	6.378222	8.005536	分散	6.3782222	8.0055357
観測数	10	8	観測数	10	8
自由度	9	7	プールされた分散	7.0801719	
観測された分散比	0.796726		仮説平均との差	0	
P(F<=f) 片側	0.366818		自由度	16	
F 境界値 片側	0.303698		t	-2.313848	
			P(T<=t) 片側	0.017149	
			t 境界値 片側	1.7458837	
			P(T<=t) 両側	0.034298	
			t 境界値 両側	2.1199053	

図11 Excel [分析ツール] の F 検定と t 検定の出力例 (等分散)

SPSS では [分析] → [平均の比較] → [独立したサンプルの t 検定]¹⁵⁾。等分散性は Levene 検定の有意確率で示され、有意水準が0.05より大きい(0.553) ので等分散である。両側検定の「有意確率 (両側)」0.34¹⁶⁾ より「差がある」と検定される。この出力には片側の有意確率は出力されない。

独立サンプルの検定										
		等分散性のための Levene の検定		2 つの母平均の差の検定						
		F 値	有意確率	t 値	自由度	有意確率 (両側)	平均値の差	差の標準誤差	差の 95% 信頼区間	
重さ	等分散を仮定する。 等分散を仮定しない。	.367	.553	-2.314	16	.034	-2.9225	1.2630	-5.6000	-.2450
				-2.283	14.261	.038	-2.9225	1.2800	-5.6632	-.1818

図12 SPSS による検定 (養鶏場の問題)

③分散が等しくないと仮定した2標本による検定の例題

「サバの養殖場で従来のエサと新しいエサを与えたとき、サバの体重に差があるか」を有意水準5%で検定¹⁷⁾。

15) 等分散か分散が等しくないか、はユーザが検定する必要はない

16) Excel の等分散を仮定した t 検定の両側検定の値と一致する

17) 従来のエサを与えられたサバ10匹のサンプルの平均は1.359Kg, 新しいエサを与えられたサバ10匹のサンプルの平均は1.94Kg

Excelの「データ分析」では、分散が等しいか等しくないかをF検定で検定する。「P(F<=f)片側」の欄が0.05より小さいため(0.016135955)分散が等しくないとし、「データ分析」のメニューから「t検定：分散が等しくないと仮定した2標本による検定」を実行する。出力の「P(T<=t)片側」は0.048171001で0.05より小さいので「差がある」。「P(T<=t)両側」は0.096342001で0.05より大きいので「差がない」。片側検定では差があり、両側検定では差がないという結果を得る。この結果について討論すると理解が深まる。

F-検定: 2 標本を使った分散の検定

	従来のエサ	新しいエサ
平均	1.359	1.94
分散	0.186698889	0.863888889
観測数	10	10
自由度	9	9
観測された分散比	0.216114469	
P(F<=f) 片側	0.016135955	
F 境界値 片側	0.314574906	

t-検定: 分散が等しくないと仮定した2標本による検定

	従来のエサ	新しいエサ
平均	1.359	1.94
分散	0.186698889	0.863888889
観測数	10	10
仮説平均との差異	0	
自由度	13	
t	-1.79250329	
P(T<=t) 片側	0.048171001	
t 境界値 片側	1.770933396	
P(T<=t) 両側	0.096342001	
t 境界値 両側	2.160368656	

図13 Excel [分析ツール] の F 検定と t 検定の出力例 (分散が等しくない)

SPSSでは「分析」→「平均の比較」→「独立したサンプルのt検定」¹⁸⁾。Levene 検定の有意確率が0.012で、分散は等しくない。また、両側検定の有意確率が0.097なので「差はない」。(Excelの両側検定の結果とほぼ一致)

独立サンプルの検定

		等分散性のための Levene の検定		2つの母平均の差の検定						
		F 値	有意確率	t 値	自由度	有意確率 (両側)	平均値の差	差の標準誤差	下の 95% 信頼区間	上限
重さ	等分散を仮定する。	7.724	.012	-1.793	18	.090	-.58100	.32413	-1.26197	.09997
	等分散を仮定しない。			-1.793	12.716	.097	-.58100	.32413	-1.28283	.12083

図14 SPSS による検定 (サバの問題)

18) 等分散が分散が等しくないか、はユーザが検定する必要はない

19) 1回目の平均点は48.73点、2回目の平均点は61.2点

① t 検定：一対の標本による平均の検定の例題

「30人が試験を受け、その1カ月後に同じ30人が試験をうけたとき、1回目の試験と2回目の試験の平均点に差があるか」を有意水準5%で検定¹⁹⁾。

Excelの「データ分析」で「t 検定：一対の標本による平均の検定」を実施。

片側検定でも両側検定でも「差がある」。

t-検定: 一対の標本による平均の検定ツール

	1回目試験	2回目試験
平均	48.73333333	61.2
分散	73.71954023	58.7862069
観測数	30	30
ピアソン相関	0.653504135	
仮説平均との差!	0	
自由度	29	
t	-10.0173064	
P(T<=t) 片側	3.17294E-11	
t 境界値 片側	1.699127027	
P(T<=t) 両側	6.34587E-11	
t 境界値 両側	2.045229642	

図15 Excel [分析ツール] 一対の標本による平均の検定の出力例

SPSSでは「分析」→「平均の比較」→「対応のあるサンプルの t 検定を実施」。両側検定の有意確率の欄から「差がある」。

対応サンプルの検定								
	対応サンプルの差					t 値	自由度	有意確率 (両側)
	平均値	標準偏差	平均値の標準誤差	差の 95% 信頼区間				
				下限	上限			
ペア 1 試験 - 得点	-53.467	9.928	1.282	-56.031	-50.902	-41.717	59	.000

図16 SPSS による検定 (1回目の試験と2回目の試験)

14回目の授業

分類：推測統計 (大学で新しく学ぶ)

内容：「分散分析」

2つの標本の平均の比較を行うのが「t検定」で13回目の授業で学んだが、14回目の授業では3つの標本の平均の比較を行う「分散分析」を学ぶ。

Excelの[データ] → [データ分析]では

- ①分散分析：一元配置
- ②分散分析：繰り返しのある二元配置
- ③分散分析：繰り返しのない二元配置

の3種が用意されている。

一元配置の分散分析の例題「3つの支店で年収に差があるか」を有意水準5%で分散分析する。

Excelの[データ分析]で[分散分析：一元配置]を実行する。分散分析表のP-値が0.00473なので「差がある」。

分散分析: 一元配置

概要

グループ	標本数	合計	平均	分散
東京支店	6	45775	7629.167	54588.57
ニューヨーク支店	6	43723	7287.167	66747.37
ロンドン支店	6	41694	6949	145106.4

分散分析表

変動要因	変動	自由度	分散	観測された分散比	P-値	F 境界値
グループ間	1387895	2	693947.4	7.813481	0.00473	3.68232
グループ内	1332212	15	88814.11			
合計	2720106	17				

図17 Excel [分析ツール] 分散分析：一元配置の出力例

SPSSでは[分析] → [平均の比較] → [一元配置分散分析]を実行。出力のなかの分散分析表の有意確率0.005により、「差がある」。

分散分析

年取	平方和	自由度	平均平方	F 値	有意確率
グループ間	1387894.778	2	693947.389	7.813	.005
グループ内	1332211.667	15	88814.111		
合計	2720106.444	17			

図18 SPSS による一元配置の分散分析

繰り返しのある二元配置の分散分析の例題「3つの支店のそれぞれ営業職3名と技術職3名の年取に差があるか」を有意水準5%で分散分析する。

Excel の [データ分析] で [分散分析：繰り返しのある二元配置] を実行する。二元配置とは、「支店」と「職」の二つの要因で分析することを表し、「繰り返しのある」とは、支店と職で分類された各項目に3人ずつ（複数）の年取のデータがあることを表す。二元配置の分散分析では要因どうしが影響を及ぼすこと²⁰⁾を「交互作用」とよぶ。

出力結果の分散分析表の P- 値では標本の P- 値が0.05より小さいと職（営業職か技術職か）で差があり，列の P- 値が0.05より小さいと支店間に差があり，交互作用の P- 値が0.05より小さいと交互作用があるということになる。この場合は，職間，支店間に有意差があり，職と支店の交互作用はないという結果になる。次頁図19に Excel の [データ分析] による繰り返しのある二元配置の分散分析の結果を示す。

SPSS では [分析] → [一般線型モデル] → [1変量] を実行し，[作図] オプションにて，グラフも表示させる。2本の線が並行であれば交互作用は「ない」。ほぼ平行であるため，交互作用はない。「被験者間効果の検定」の有意確率の欄から職間，支店間に有意差があり，職と支店の交互作用はないという Excel と一致する結果が得られる。図20に SPSS による二元配置の分散分析の結果を示す。

20) たとえば「相乗効果」

繰り返しのない二元配置の分散分析は二つの要因の組合せにデータが1つずつしかない場合に実行する。つまり、東京支店の営業職、東京支店の技術職、ニューヨーク支店の営業職、ニューヨーク支店の技術職、ロンドン支店の営業職、ロンドン支店の技術職の6データの分析である。

Excelの「データ分析」では「分散分析：繰り返しのない二元配置」を選択し、SPSSでは繰り返しのある場合と同様のメニュー選択を行う。それぞれの結果は図21と図22に示す。

分散分析: 繰り返しのある二元配置

概要	東京支店	ニューヨーク支店	ロンドン支店	合計
営業職				
標本数	3	3	3	9
合計	23353	22445	21563	67361
平均	7784.333	7481.667	7187.667	7484.556
分散	8920.333	18569.33	87830.33	95586.78
技術職				
標本数	3	3	3	9
合計	22422	21278	20131	63831
平均	7474	7092.667	6710.333	7092.333
分散	55321	34808.33	104050.3	157892.5
合計				
標本数	6	6	6	
合計	45775	43723	41694	
平均	7629.167	7287.167	6949	
分散	54588.57	66747.37	145106.4	

分散分析表

変動要因	変動	自由度	分散	観測された分散比	P-値	F 境界値
標本	692272.2	1	692272.2	13.42048	0.003246	4.747225
列	1387895	2	693947.4	13.45295	0.000861	3.885294
交互作用	20940.11	2	10470.06	0.202974	0.819045	3.885294
繰り返し誤	618999.3	12	51583.28			
合計	2720106	17				

図19 Excel [分析ツール] 分散分析：繰り返しのある二元配置の出力例

被験者間効果の検定

従属変数: 年収

ソース	タイプ III 平方和	df	平均平方	F	有意確率
修正モデル	2101107.11 ^a	5	420221.422	8.146	.001
切片	956185603.6	1	956185603.6	18536.736	.000
支店	1387894.778	2	693947.389	13.453	.001
職	692272.222	1	692272.222	13.420	.003
支店 * 職	20940.111	2	10470.056	.203	.819
エラー	618999.333	12	51583.278		
合計	958905710.0	18			
修正総和	2720106.444	17			

a. R2 乗 = .772 (調整済み R2 乗 = .678)

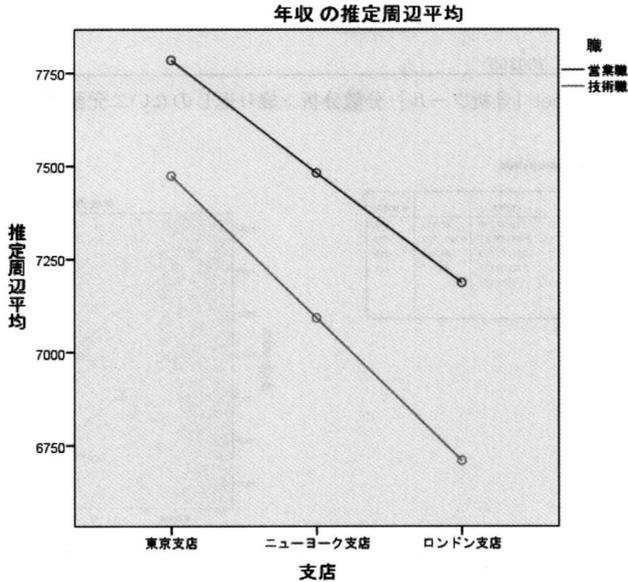


図20 SPSS による二元配置の分散分析 (18データ)

分散分析: 繰り返しのない二元配置

概要	標本数	合計	平均	分散
営業職	3	22453.67	7484.556	89009.03704
技術職	3	21277	7092.333	145796.7778
東京支店	2	15258.33	7629.167	48153.38889
ニューヨーク支店	2	14574.33	7287.167	75660.5
ロンドン支店	2	13898	6949	113923.5556

分散分析表

変動要因	変動	自由度	分散	観測された分散比	P-値	F 境界値
行	230757.4	1	230757.4	66.11925014	0.014789503	18.51282051
列	462631.6	2	231315.8	66.2792461	0.014863425	19
誤差	6980.037	2	3490.019			
合計	700369	5				

図21 Excel [分析ツール] 分散分析: 繰り返しのない二元配置の出力例

被験者間効果の検定

ソース	タイプ III 平方和	df	平均平方	F	有意確率
修正モデル	693291.167 ^a	3	231097.056	65.426	.015
切片	31873393.5	1	31873393.5	90237.360	.000
支店	462403.000	2	231201.500	65.456	.015
職	230888.167	1	230888.167	65.367	.015
エラー	7094.333	2	3547.167		
合計	319433749.0	6			
修正総和	700355.500	5			

a. R2 値 = .990 (調整済み R2 値 = .975)

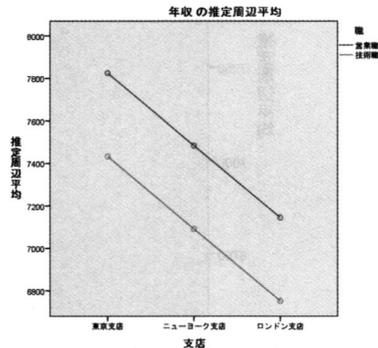


図22 SPSS による二元配置の分散分析 (6データ)

15回目授業

分類: アンケート集計 (大学で新しく学ぶ)

内容: 「ピボットテーブルを用いたアンケート集計」

Excel のアドインや SPSS といった特定用途の機能を用いず, メニューバーの [挿入] から [ピボットテーブル] を選択することにより簡単に度数分布分析やクロス集計を行えることを学ぶ。合わせてピボットグラフも学ぶ。

4. 課題

統計に関する知識、およびソフトウェアの設計には独特の敷居の高さがあり、“Excelによる統計分析”の書籍もその内容に統一がない。講義として授業するのか、実習とするのかによってもまったく異なった内容になる。また、SPSSではメニュー構造が深く展開しており、同じことを複数の方法で実行できる。このことは、どの方法を採用すべきかという新たな悩みの原因ともなっている。「3.文系・セメスタ(全16回)・実習の授業スケジュール」で述べた授業内容に間違いがないかはいまだに確信がない。この研究ノートをお読みになった方から間違いやより正しい知識や操作方法についてご教示いただければ幸いである。

[参考文献]

1. 啓林館高等学校数学科用教科書「数学I」「数学A」「数学B」
2. 上田太一郎監修,「Excelでかんたん統計分析—分析ツールをつかいこなそう!—」, オーム社, 2009年
3. 岸学著,「SPSSによるやさしい統計学」, オーム社, 2012年
4. 木下宗七「入門統計学 [新版]」, 有斐閣ブックス, 2011年
5. 近藤宏, 末吉正成, 村田真樹著,「Excelでかんたん統計分析—「分析ツール」を使いこなそう!」, オーム社, 2007年
6. 浅井隆著,「いまさら誰にも聞けない医学統計の基礎のキソ 第1巻まずは統計アレレギーを克服しよう!」, アトムス, 2010年
7. 石村貞夫・石村資郎著,「SPSSによる分散分析と多重比較の手順」, 東京図書, 2011年
8. 菅民郎著,「Excelで学ぶ統計解析入門」, オーム社, 2011年
9. 涌井良幸, 涌井貞美著,「Excelで学ぶ統計解析」, ナツメ社, 2013年
10. 数研版教科書ガイド「数学I」「数学A」「数学B」