

2000年度教養科目「統計学」(社会) 期末試験解説／担当：星野(経済学部講師)

- このハンドアウトは <http://www.ec.kanazawa-u.ac.jp/hoshino/stats/list.html> から取得可

この講義は今年が初めてだったが、ある程度手応えは有った。諸君に書いてもらった授業に関する提案は、出来るだけ活かしたいと考えている。なお統計のテストの統計は、付録に与えている。

採点の大方針を説明しよう。主観を書かせる問題は、できるだけ広く解釈して得点を与えるようにした。ただ客観問題は、講義で与えた定義式以外の計算結果を正答として認めなかった。本によっては同じ統計量に異なる定義をしている場合がある。しかし授業に出ない事で不利益を被るのは自己責任である。

任意提出のレポートが二回出ているが、二回とも出した学生には最大で10点(= α) 足した。テストの素点 + α が60点を下回ると「不可」、69点までが「可」、79点までが「良」で、それ以上が「優」である。レポートを二回出して「可」がついた諸君は、テストだけでは「不可」だったという事である。講義理解が不十分だという事を自覚して、努力を重ねて欲しい。

「不可」の諸君は、自分に何が足りなかったのか良く考えて欲しい。「数学が苦手だから」で片づけてはならない。努力や勉学のあり方について、考え違いをしているのではないか。授業中に「特性値が計算できれば単位を出す」と言った。問いの1,2が完全に出来れば45点。それ以外で5点取れば、レポートの10点と合わせて「可」が付くようになっている。算数の知識が有れば、単位はとれるのである。確かに努力は尊い。しかし授業に出席しただけでは、努力したとは言えない。あらかじめ「特性値が計算できる」という目標が示されていたにも関わらず、必要な準備=努力を怠ったのではないだろうか。

授業中にはほとんど問題演習をしない私の責任を問いたいかもかもしれない。しかし問題演習は、全員一律の授業にはなじまない。諸君が高校までで受けてきた授業は、黒板で問題の解き方を示すというものが多かったであろう。これが大部分の学生にとって退屈なのは、自分のレベルに合っていないからである。退屈でも必要な事はやらねばならぬが、問題演習は各自が必要に応じてやるのが良い。自分が理解しているかどうかは、自分が一番良く分かるはずだ。練習問題のついている本を紹介しているのだから、自分で選び取ることだ。それに自分で選び取る事に意義がある。本当の勉学は、自分に何が足りないか分析し、自分で課題を立て、それを解決するプロセスである。これが上手くできれば、結果として成績は良くなる。成績が良いからといってこのプロセスが上手いとは限らないが、これが出来る人が社会的に「有能」と評価されるのである。口を開けて(教師の)課題を待っているだけの人間は、社会的に求められていない。勉学の意義は、専門知識獲得にも有るが、目標を明らかにしてその為は何をしたら良いか自分で考える習慣を身につける中に有る。これが必要とされる人材になるための、最大の訓練なのである。

本講義は、来年度以降も私が持つ機会が有るはずだ。その場合講義の内容は大きく変えない。出席は取らない。再履修して本年度のノートを使って試験を受け、堂々と単位を取得してみよ。

1. (配点：30) 度数分布表が読めていない人は居なかったようだ。行和、列和を「周辺度数」と呼び、一変数の度数となる。これくらいの計算が出来ないと「金沢大の学生は分数が出来ない」と言われる。それでは困るので、部分点などは与えていない。
 - (a) 2.89 ぐらい。
 - (b) 1.58 ぐらい。この講義では総変動(=平均からの偏差の二乗和)をデータ数 n で割った物が「分散」である。 $n-1$ で割った物を分散としている本も有る。
 - (c) 3 である。総度数が9で奇数な事に注意。
 - (d) 2 である。2点を取った人がもっとも多い。
 - (e) $4=5-1$ 。この講義では、最大値-最小値をレンジと定義した。

(f) 平均の桁丸めをせずに計算すると 0.938 になる。平均を下一桁まで合わせて計算すると、1.1 くらいらしい。共変動 (= 平均からの偏差を二変数について掛け合わせた総和) を n で割った物が、この講義での「共分散」の定義。 $n - 1$ で割る定義を採用している本も有る。この問いはかなり出来が悪いが、相関係数を理解するには共分散の挙動を押さえる事は必須である。必ず復習の事。

2. (配点:15)

(a) 次のように書けばよい。葉は大きさ順に並び替える事。

0	1 5
1	0 0 3 6 7
2	2 5
3	1

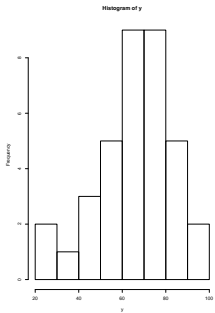
(b) データ数は 10 である。20%分位点は $7.5 = (x_{(2)} + x_{(3)})/2$ 。58%分位点は $16 = x_{(6)}$ 。本によっては分位点を違うやり方で定義している。

3. (配点:10) $(e) < (c) < (b) < (a) < (d) < (f)$ となる。それぞれ相関係数は、-0.98,-0.75,0.23,0.75,0.87,0.98 である。これらが -1 から 1 までの範囲をとる事に注意。比例なら正で、反比例なら負である。a,d,f は明らかに比例、c,e は反比例。そして線形関係が強ければ、相関係数の絶対値は 1 に近くなる。例えば e と f は「ひっくり返した」関係に有るが、「線形度」は等しいので相関係数の絶対値は同じ。
4. (配点:10) 例えば平均 3.5 ぐらい、メディアン 4 ぐらい、モード 4.8 ぐらい。それぞれ 3,4,5 などでも良い。グラフの読み方を理解していない人も散見された。縦軸を相対度数と考えるという事で、モードが一番度数が多いところの $x \doteq 4.8$ ぐらいか。この周辺なら正解にしている。メディアンは 50%分位点だから、度数を半分にわけるとあたり。つまり曲線で囲まれた部分の面積を二分するような $x \doteq 4$ ぐらい? 3.5 より下は明らかに半分ではないように見えるので減点。それからこれは左に歪んだ分布であるから、平均 < メディアン < モードと予想される。この順序が崩れた解答は大幅減点。しかし大した歪みではないので、平均は 3.5 程度だろう。これは良く分からないので、許容範囲が広い。
5. (配点:10) このような M を「切り落とし平均」という。一部の教科書には紹介されていると思う。しかし授業だけでも理解していれば、(a) 中心の特性値で有る事 (b) ロバスト (外れ/異常値に強い) で有る事くらいは考えられるはずだ。なお計算量が減るという指摘 (減点せず) もあったが、データを並び替える作業を含めると計算は減らず、むしろ増える。
6. (配点:10) 何でも合理的と思われる解答には得点を与えた。当然のことながら日常用語は曖昧なので、厳密な議論に耐えない。各種専門分野で用語が溢れるのは、用語こそが議論の基礎だからである。しかし、中心の特性値がいくつも有る理由を記した答案が目立った。問いは中心の特性値に焦点が有るのではない。一般的な傾向をきいている。個別的事実から普遍的命題を推論する事が求められているのだ。ノート・参考書から関係すると思われる箇所を探して抜き書きした諸君、そのような処世術は即座に捨てなさい。学校のテストで部分点は取れるかもしれないが、生きる力にはならないよ。日本の大学は優しいので、目先の成績にこだわらなくても大丈夫。慣れない事をするのは怖いかもしれないけど、自分で考える訓練のラストチャンスなのだから。
7. (配点:15) 気温・売り上げという二変数データの分析手順を説明すればよい。散布図を書き、相関関係を調べ、度数分布表を作る。一変数の分析も加え、外れ値のチェックなど有効だろう。何より大切なのは、グラフを書くという事である。データは予見以上のことを語ってくれる可能性を持つ。授業の主題は、このようなパターンの発見方法である。— 他にも季節・曜日や商品別の分析など、色々工夫できるだろう。なおこれは経営学の試験ではないので、商店がウナギ屋か灯油屋かというような具体性は問われない。そのかわりデータ分析方法説明に具体性が無いものは不可。

付録：統計のテストの統計

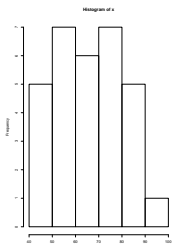
一変数の分析の復習を兼ねて、素点について結果を示す。受験者は 67 (文系 31、理系 36) 名であった。以下の計算は R というソフトで行ったが、四分位点の計算法が講義で採用したやり方と異なる事に注意。まず五数要約すると、{ 最小値 24, 第一四分位点 55.5, メディアン 68, 第三四分位点 76.5, 最大値 99 } である。標準偏差は 15.46、平均点は 66.25 であり、左に歪んだ分布で有ることが予想される。実際、度数分布のヒストグラムを見ると以下のようなになる。

1. 全体の得点分布

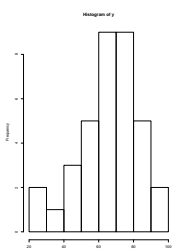


文系 (文・法・経・教) だけ見ると、五数要約は {41,55,68,76.5,93}、標準偏差 13.85、平均 66.39 点である。理系 (理・医・工) では、五数要約が {24,57.25,67.5,76.25,99}、標準偏差 16.92、平均 66.14 点だ。平均、メディアンは文系の方が良い。実は全く準備せずに受験した理系の学生が、全体を下に引っ張っているのである。ヒストグラムを見てみよう。

2. 文系の得点分布



3. 理系の得点分布



文系のモードは 50-60 と 70-80 に有る。理系のモードは 60-80 である。これを見ると、理系の諸君のプライドも保たれるだろうか。しかしこの講義は数学力というより、論理性を要求しているつもりである。以上。