大学教育のガバナンスと成績評価基準(中)質保証とGPA制度

HAYASHI, Naotsugu / 林, 直嗣

(出版者 / Publisher)
法政大学経営学会
(雑誌名 / Journal or Publication Title)
The Hosei journal of business / 経営志林
(巻 / Volume)
47
(号 / Number)
2
(開始ページ / Start Page)
39
(終了ページ / End Page)
47
(発行年 / Year)
2010-07-31
(URL)
https://doi.org/10.15002/00008912

〔研究ノート〕

大学教育のガバナンスと成績評価基準（中）

＝質保証とGPA制度＝

林　　直　　嗣

目　次

1．はじめに
2．大学教育のガバナンス
3．大学設置基準で定める授業，試験，及び成績評価基準
（以上前号）
4．試験等の得点分布の正規性と中心極限定理，正規分布検定
5．正規分布検定の実証分析
（以上本号）
6．現行成績評価基準の問題点
7．現行GPA制度の問題点
8．適正な成績評価基準とGPA制度
9．おわりに
参考文献

4．試験等の得点分布の正規性と中心極限定理，正規分布検定

4.1．試験等の得点分布の正規性と中心極限定理

各授業科目の試験等による各履修者の総合得点 X_i（i＝1～n）は，学生側の要因と教員側の教え方や出題・採点方法などに依存して決まる。学生側の要因としては，各人の授業出席，理解度，予習，復習，テスト勉強の仕方など多数の独立した要因により X_i が決まり，試験前の共同答案練習，複数によるカンニング行為などがなければ，X_i が相関をもつことはないといえる。また同じ大学の同じ学部の学生は，元々入学試験で同じような学力レベルの学生が合格しているので，教員側の要因としては，特定の集団・グループには理解しやすいが，別の集団・グループには理解しにくいような，偏りのある教え方や出題・採点方法などをしなければ，X_i は相関をもたないといえる。

これらの条件が満たされるとき，X_i は独立で同一な分布をする（independently identically distributed: i.i.d.）確率変数と見なすことができる。

いま試験の得点 X_i が，平均 μ，標準偏差 σ（分散は σ^2）の任意の独立同分布（i.i.d.）に従うならば，大きさ n の無作為標本に基づく標本平均 $m_n=(X_1+X_2+\cdots\cdots+X_n)/n$ は，n が大きくなるとき，平均 μ，標準偏差 $\sigma/\sqrt{n}$（分散は σ^2/n），確率密度関数

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(x-\mu)^2}{2\sigma^2})$$

の正規分布（normal distribution）に弱収束する。すなわち

$$\lim_{n\to\infty}P\left[\frac{m_n-\mu}{\sigma/\sqrt{n}}\leqq y\right]=\int_{-\infty}^{y}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$$

よって標本平均と母平均の誤差 $(m_n-\mu)/(\sigma/\sqrt{n})$ は，n が大きくなるとき，平均0，標準偏差1，確率密度関数

$$f(x)=\frac{1}{\sqrt{2\pi}}\exp(-\frac{x^2}{2})$$

の標準正規分布（standard normal distribution）に弱収束する。すなわち

$$\lim_{n\to\infty}P\left[\frac{m_n-\mu}{\sigma/\sqrt{n}}\leqq y\right]=\int_{-\infty}^{y}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx$$

この定理は，中心極限定理（the central limit theorem）として，証明されている。

確率変数 X が 標準正規分布N（0，1）に従う時，この確率密度関数に基づいて y が取る確率は標準正規分布表として計算されている。それによると，N(μ，σ^2）に従う時は，平均 μ

からのずれが $\pm 1\sigma$ 以下の範囲に X が含まれる確率は 68.26％，$\pm 2\sigma$ 以下だと 95.44％，さらに $\pm 3\sigma$ だと 99.74％ と計算されている。

なお試験の得点，身長，体重のような社会現象や気体分子の速度のような自然現象は，標本数 n が十分に大きくなると，経験的に正規分布に近づくことがしばしば確認されてきた。正規分布は，1733年にド・モアブルにより最初に研究されたが，ガウスは実験における偶然の誤差の分布が正規分布であることを発見し，この貢献により正規分布は，ガウス分布（Gaussian distribution）とも呼ばれている。ケトレーは成年男子の身長の分布が正規分布に近似できることを確認した。またマックスウェルは気体分子の速度が正規分布に従うことを検証した。

4.2. 正規分布検定とコルモゴロフ＝スミルノフ検定

個々の事例の経験分布が実際に正規分布に従うかどうかは，アプリオリに仮定されるべきことではなく，実証的に検証されるべきことである。

ある標本分布が正規分布に従うか否かを検定する統計的方法には幾つかあるが，その一つにコルモゴロフ＝スミルノフ検定（Kolmogorov-Smirnov test; KS test）[注9] がある。KS検定は，有限個の標本に基づいて，標本の経験分布が帰無仮説で仮定された累積分布と有意に異なるか否かを検証したり，あるいは2つの標本の経験分布が有意に異なるか否かをノンパラメトリックに検証する仮説検定である。前者の1標本 KS 検定（one-sample KS test）は，標本の経験分布関数と帰無仮説で仮定された累積分布関数との最短距離をコルモゴロフ＝スミルノフ統計量（KS test statistic）として数量化し，所定の有意水準にしたがって有意性を判定する。正規分布や一様分布，ポアソン分布などの適合度検定に用いることができる。帰無仮説は，標本が正規分布など仮定する分布から抽出されたこと，両者が有意に異ならないことである。後者の2標本 KS 検定（two-sample KS test）は，2つの標本の経験分布関数の最短距離を KS 統計量として数量化し，所定の有意水準にしたがって有意性を判定する。帰無仮説は，2つの標本が同じ分布から抽出されたこと，すなわち両者が有意に異ならないことである。

いま n 個の独立で同一な分布をする（independently identically distributed: i.i.d.）確率変数 X_i（$=X_1, X_2, \cdots\cdots, X_n$）の経験分布を次の分布関数で定義する。

$$F_n(x)=\frac{1}{n}\sum_{i=1}^{n} I \qquad (I=1 \text{ if } X_i \leqq x,\ I=0 \text{ if } X_i > x)$$

このとき所与の累積分布関数 $F_n(x)$ に対する KS 検定統計量は，次式で与えられる。

$$D_n=\sup_x |F_n(x)-F(x)|$$

あるいは2つの片側 KS 検定統計量として表せば，次のようになる。

$$D_n^{+}=\max(F_n(x)-F(x)),$$

$$D_n^{-}=\max(F(x)-F_n(x))$$

コルモゴロフ分布は，$B(t)$ を Brown ブリッジとして，次式で定義される。

$$K=\sup_{t\in[0,1]} |B(t)|$$

この K の累積分布関数は，次のようになる[注10]。

$$\Pr(K\le x)=1-2\sum_{i=1}^{\infty}(-1)^{i-1}e^{-2i^2x^2}$$

標本分布が仮定された分布関数 $F(x)$ から得られる分布と有意に異ならない，という帰無仮説の下で，n が大きくなるとき，$\sqrt{n}\,D_n$ はコルモゴロフ分布に収束する。

$$\sqrt{n}D_n = \xrightarrow{n\to\infty} \sup_t |B(F(t))|$$

有意水準 α のもとでの有意確率 K_α は，

$$\Pr(K\le K_\alpha)=1-\alpha$$

と求められ，これに対して，

$$\sqrt{n}D_n > K_\alpha$$

であれば，帰無仮説は棄却される。

正規分布の検定に関しては，Shapiro-Wilk 検定や Anderson-Darling 検定に比べて，KS 検定は分布の裾よりも中央値付近に感応的であることが知られている[注11]。

５．正規分布検定の実証分析

山本（2008）によれば，上智大学では2007年度から各開講科目の成績評価の分布を，学生と教職員に対して公開している。これにより，同一科目を担当している複数の教員の間で，成績評価に大きなバラツキがないように調整したり，極端に甘い成績評価や極端に辛い成績評価の教員が他の成績評価分布を知ることにより再考をするなど，成績評価の在り方を考え直すよう試みたという。成績評価を大学設置基準が要請する「客観的で厳格な」ものに近づけていく上で，画期的な試みと評価できよう。

個々の学生の成績評価は当該個人の個人情報であり，本人の同意なく第三者に開示することは個人情報保護法に抵触する恐れがあるが，各開講科目の成績評価分布あるいは学部全体の平均的な成績評価分布は，個人を特定する個人情報をまったく含んでいないので，本人の同意を得ることなく開示しても個人情報保護法には抵触しない。

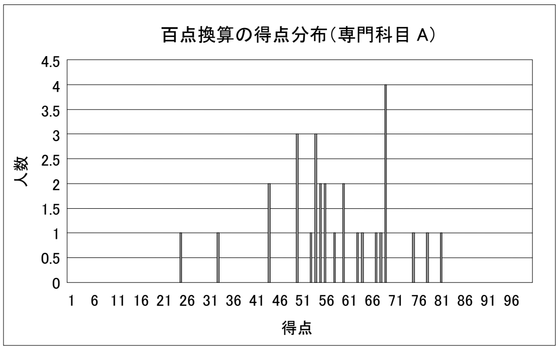
本節では，各開講科目及び学部全体の成績評価の分布が正規分布に従う事例が多いという事実を，統計的に検証するが，そのことは実は成績評価の理論を実証的に裏付ける上で非常に重要な意義を持つ。仮想的な成績数値シミュレーションだけでは，説得力に乏しい。ただし成績評価の情報は，特定の個人情報を含まないように加工した形態ではあれ，従来から余り公表されてこなかった微妙な実情があるので，それを統計分析用に加工した上で，科目名や学部名や大学名など固有名詞を一切特定しない形態で以下に正規分布検定を行う。

５.１．標本数約30の科目Ａのケース

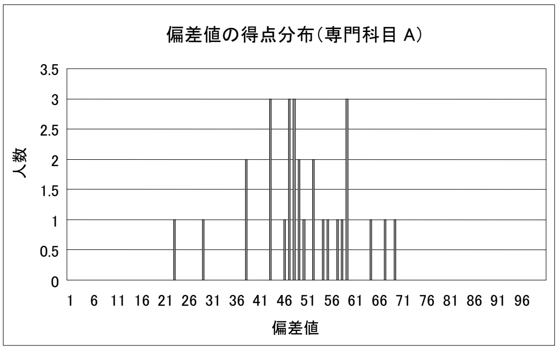
Ｚ大学α学部の通期の専門科目Ａでは，試験100点，平常の課題40点，自由研究課題20点，ノート筆記20点，出席点30点からなる合計得点は210点満点であり，それを100点満点に換算した得点を，履修者合計30名（履修登録はしても未受験は除く）について図示したのが（5-1図）である。最高点は80.95点，最低点は25.24点，平均点は57.87点，標準偏差は12.05である。

百点換算の得点分布は，難易度があまり易しくもなく難しくもなく，平均点は57.84点と分布の位置（location）が50点よりやや上方にある。その難易度の僅かな偏りを是正するため，得点を偏差値（standard score = 50 + 10 × 偏差／標準偏差）に変換して図示したのが，（5-2図）である。最高点は69.16点，最低点は22.91点，平均点は50点，標準偏差は10である。得点から偏差値への変換は１次式による線型変換であるので，位置は平均が50点の位置に是正されるものの，標準偏差が16.34から10へ縮小された分だけ横幅が縮小するだけで，釣り鐘型（bell-shaped）の山の形状はそのまま維持される。

（5-1 図）

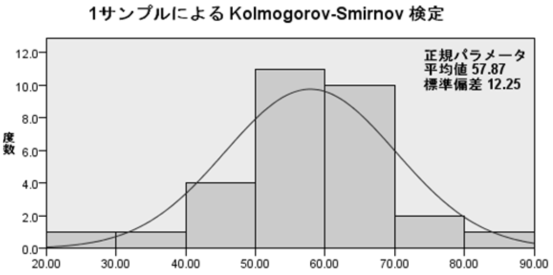


（5-2 図）



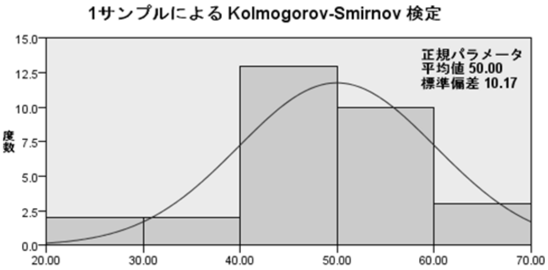
そこで有意水準α＝0.05（５％），信頼区間0.95で，上記の KS 正規分布検定を SPSS により実施したところ，百点換算点でも偏差値でも検定結果はまったく同じであり，同じく正規分布に従うことが検証された。このように標本数が30ほどの小規模サンプルでも，正規分布であることが検証できる。

（5-3 図）百点換算の得点分布の KS 検定



KS 統計量=0.627,
漸近有意確率=0.827, 有意水準=0.05
最遠距離差（正=0.088, 負=-0.114）

（5-4 図）偏差値の得点分布の KS 検定



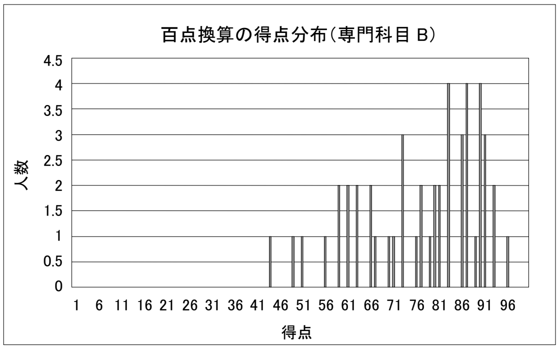
KS 統計量=0.627,
漸近有意確率=0.827, 有意水準=0.05
最遠距離差（正=0.088, 負=-0.114）

5.2. 標本数50の科目Ｂのケース

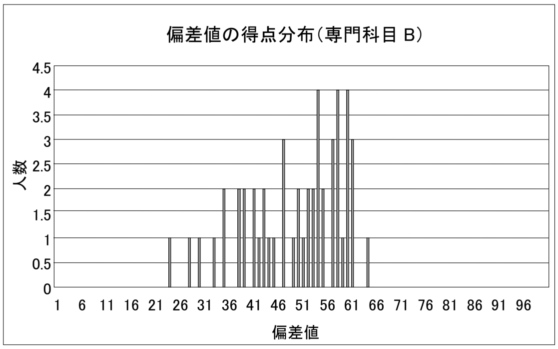
Ｚ大学β学部の専門科目Ｂでは，定期試験50点，レポート20点からなる合計得点は70点満点であり，それを100点満点に換算した得点を，履修者合計50名（履修登録はしても未受験は除く）について図示したのが（5-5 図）である。最高点は95.71点，最低点は44.29点，平均点は77.31点，標準偏差は12.85である。

百点換算の得点分布は，かなり難易度が易しいか履修者の理解度が高いかで，平均点が77.31点と分布の位置（location）が50点より27点も上方にある。難易度の偏りを是正するため偏差値を計算して，その分布を（5-6 図）に示してある。最高点は64.32点，最低点は24.30点，平均点は50，標準偏差は10である。

（5-5 図）

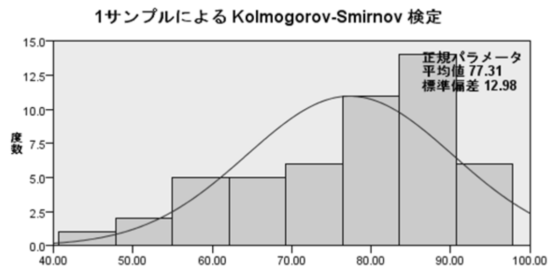


（5-6 図）



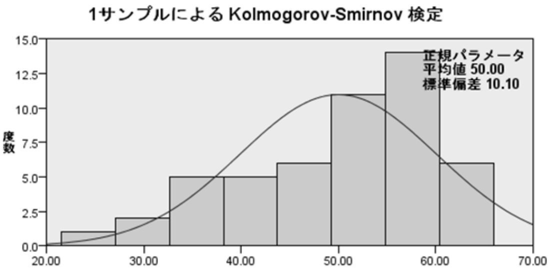
そこで有意水準α=0.05（5％），信頼区間0.95で，上記の KS 正規分布検定を実施したところ，百点換算点でも偏差値でも検定結果はまったく同じであり，同じく正規分布に従うことが検証された。科目Ａのケースと比較すると，上方へ(できる方へ) 分布が偏っているため, KS 統計量が大きく，漸近有意確率が小さくなっているので，分布の正規性の程度は低くなるが，正規分布であることは検証できる。

（5-7図）百点換算点の KS 正規分布検定



KS 統計量=1.027,
漸近有意確率=0.242，有意水準=0.05
最遠距離差（正=0.096，負=-0.145）

（5-8図）偏差値の KS 正規分布検定

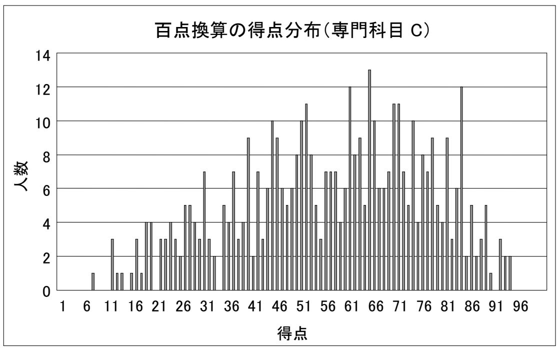


KS 統計量=1.027,
漸近有意確率=0.242，有意水準=0.05
最遠距離差（正=0.096，負=-0.145）

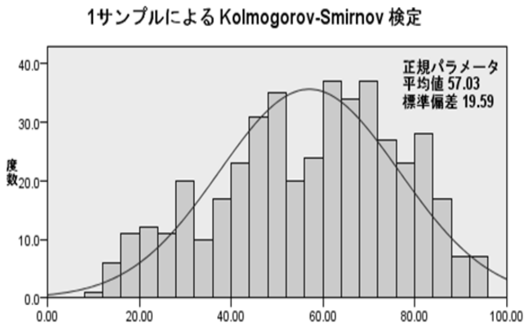
5.3．標本数約400の科目Ｃのケース

Ｚ大学α学部の通期の専門科目Ｃでは，定期試験50点×２回=100点，前期レポート10点，後期授業内小テスト10点，出席点12点（学生に告知せずにランダムに４回出欠調査×３点）からなる合計得点は132点満点であり，それを100点満点に換算した得点を，履修者合計438名（履修登録はしても未受験は除く）について図示したのが下図である。百点換算の得点分布は，難易度があまり易しくもなく難しくもなく，平均点は57.03点と50点よりやや上方にある。最高点は94.08点，最低点は8.45点，平均点は57.03点，標準偏差は19.57である。

（5-9図）



（5-10図）百点換算点の KS 正規分布検定



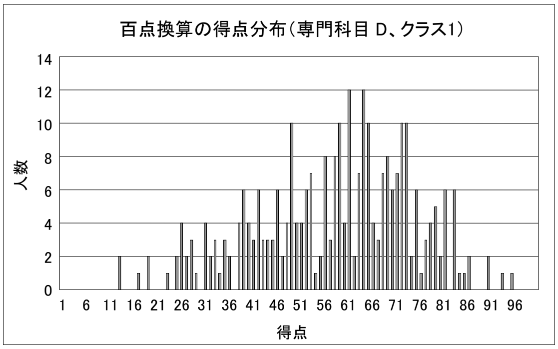
KS 統計量=1.235,
漸近有意確率=0.095，有意水準=0.05
最遠距離差（正=0.039，負=-0.059）

そこで有意水準α=0.05（５%），信頼区間0.95で，上記の KS 正規分布検定を実施したところ，同じく正規分布に従うことが検証された。平均点が７点ほど50点より高く，僅かに分布が上方へ偏ってはいるものの，サンプル数が大きい分だけ滑らかな分布に近づいており，やはり正規性が検証できる。

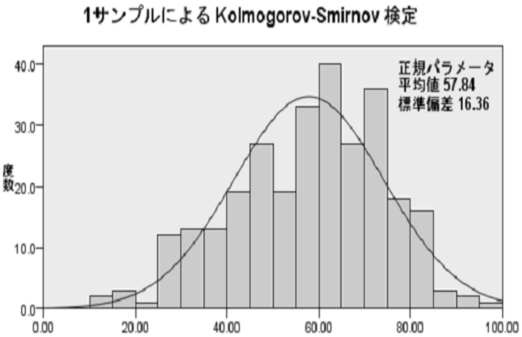
5.4．標本数約300の科目Ｄ，クラス１のケース

Ｚ大学α学部の専門科目Ｄ，クラス１では，定期試験50点，授業内小テスト10点，出席点12点（学生に告知せずにランダムに４回出欠調査×３点）からなる合計得点は72点満点であり，それを100点満点に換算した得点を，履修者合計286名（履修登録はしても未受験は除く）について図示したのが下図である。最高点は95.28点，最低点は12.5点，平均点は57.84点，標準偏差は16.34である。

（5-11図）



（5-12図）百点換算点の KS 検定



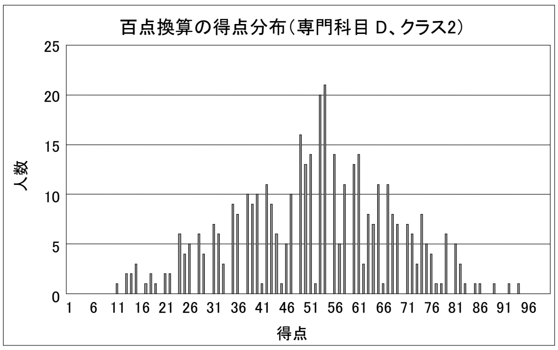
KS 統計量＝1.122,
漸近有意確率＝0.161，有意水準＝0.05
最遠距離差（正＝0.033，負＝-0.066）

そこで有意水準α＝0.05（5％），信頼区間0.95で，上記の KS 正規分布検定を実施したところ，同じく正規分布に従うことが検証された。科目Ｃと比較して，平均点はほぼ同じであるが，分布の形状がスムーズな釣り鐘型に近いので，KS 統計量は小さく，漸近有意確率は大きくなっており，正規性の程度はやや高いと判定できる。

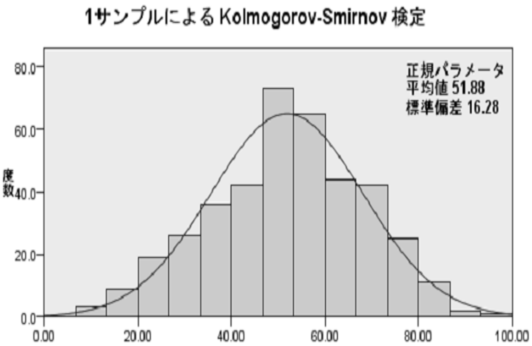
5.5．標本数約400の科目Ｄ，クラス２のケース

Ｚ大学α学部の上記と同じ専門科目Ｄ，クラス２では，定期試験50点，授業内小テスト10点，出席点12点（学生に告知せずにランダムに４回出欠調査×３点）からなる合計得点は72点満点であり，それを100点満点に換算した得点を，履修者合計398名（履修登録はしても未受験は除く）について図示したのが下図である。最高点は94.44点，最低点は11.11点，平均点は51.88点，標準偏差は16.26である

（5-13図）



（5-14図）百点換算点の KS 検定



KS 統計量＝1.118,
漸近有意確率＝0.164，有意水準＝0.05
最遠距離差（正＝0.026，負＝-0.056）

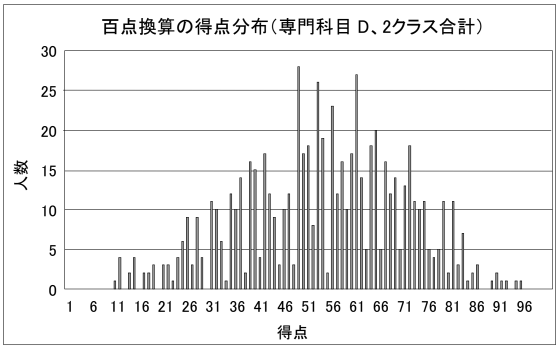
そこで有意水準α＝0.05（5％），信頼区間0.95で，上記の KS 正規分布検定を実施したところ，同じく正規分布に従うことが検証された。科目Ｄのクラス１とは授業内容は同じで，試験は出題方法や難易度はまったく同じであるが，平均点が約52点と前者の約58点より６点低いのは，元々クラス編成の時に学力レベルがやや異なる編成をしていたと推察できる。しかし標本数が

多いクラスの方が KS 統計量は小さく，漸近有意確率は高いので，i.i.d.の確率変数は標本数が大きくなると正規分布に近づくという中心極限定理を例証するといえる。

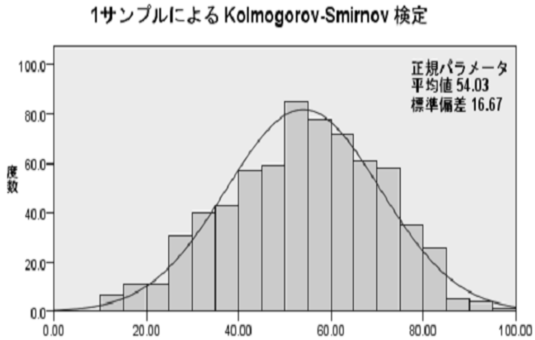
5.6． 標本数約700の科目Ｄ，２クラス合計のケース

Ｚ大学α学部の半期の専門科目Ｄの２クラス合計では，定期試験50点，授業内小テスト10点，出席点12点（学生に告知せずにランダムに４回出欠調査×３点）からなる合計得点は72点満点であり，それを100点満点に換算した得点を，２クラスの履修者合計684名（履修登録はしても未受験は除く）について図示したのが下図である。最高点は95.28点，最低点は11.11点，平均点は54.03点，標準偏差は16.65である。

（5-15図）



（5-16図）百点換算点の KS 正規分布検定



KS 統計量＝1.059，
漸近有意確率＝0.212，有意水準＝0.05
最遠距離差（正＝0.025，負＝−0.040）

そこで有意水準α＝0.05（５％），信頼区間0.95で，上記の KS 正規分布検定を実施したところ，同じく正規分布に従うことが検証された。クラス毎のケースより合計した場合は KS 統計量が小さく，漸近有意確率が大きくなるので，正規性は高くなると判定できる。よって i.i.d.の確率変数は標本数が大きくなると正規分布に近づくという中心極限定理を例証する。

5.7． 正規分布からずれる事例

中心極限定理では，母集団分布は任意であっても，nを大きくするにつれて，無作為抽出した標本平均の分布が正規分布に近づくことをいうのであって，母集団分布が正規分布であることは前提していない。標本数が大きくなるときに標本分布が正規分布ないしそれに近い分布に近づくためには，幾つか条件があって，科目によっては必ずしもそれが十分に満たされない場合があることは事実である。

授業や試験が難解で難易度が難しい場合，あるいは履修者の理解度が低い場合は，平均点が50点より低くなり，平均点以下の度数（人数）が大きくなるので，低い方（下方）に分布の歪みが生じる。逆に難易度が易しい場合，あるいは履修者の理解度が高い場合は，平均点が50点より高くなり，平均点以上の度数（人数）が大きくなるので，高い方（上方）に分布の歪みが生じる。歪みの原因が教員側にあるのか学生側にあるのか，先験的には特定できないので，個々のケースにより分析する必要がある。平均点が50点から大きくずれるケースで教員側に原因がある場合には，偏差値換算などにより難易度の歪みを是正する必要がある。

ある程度以上の勉強をする学生にはほぼ興味を持って理解できる教え方であっても，それ以下の学生にはほとんど興味も感じずに理解できないような場合には，できるグループとできないグループとに２極分化して，得点分布の山が２つできるケースもあり得る。同じ大学の同じ学部の学生は，入学時にできるグループとでき

ないグループを２極分化して合格させているわけではないので，そこから抽出した標本分布に２極分化したコブができるか否かは，主として入学後の学生側か教員側かいずれかの側に何らかの原因があると考えられる。たとえば公認会計士試験や司法試験などの国家試験講座を併設している場合に，それを受講する学生は興味を持ってその関連科目を勉強するので成績も良くなる傾向があるが，この講座を受講しない学生はその関連科目では相対的に成績は低くなる。教員ができるグループに合わせて授業レベルを設定すると，できない集団がますます理解できなくなる。逆にできないグループに合わせて授業レベルを設定すると，できるグループは授業がつまらなく感じる。つまり一つの授業科目に２つの性質の違うグループが混在することになり，そのため得点分布が２極分化しうる。そうしたケースでは，プレースメントテストなどによりクラス分けをして，それぞれのクラスの学力レベルに適合した授業や試験を行う方が，受講者全体の学力向上に役立ちうる。

5.8.　成績評価分布の正規分布検定

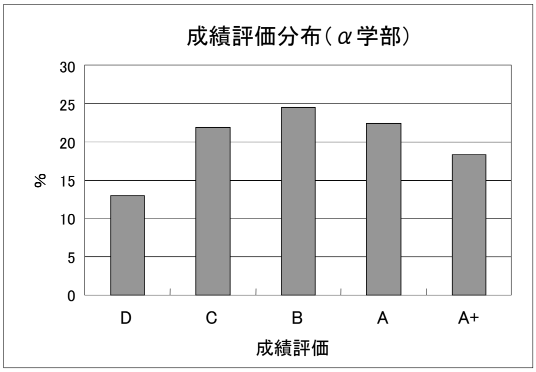
次にＺ大学α学部の専門科目全体（専任教員が担当）の成績評価分布を見ると，（5-1表）と（5-17図）のようになる。有意水準α=0.05（５％），信頼区間0.95で，KS 正規分布検定を実施したところ，比較的に KS 統計量は小さくて漸近有意確率は高いので，正規性が高い正規分布に従うことが検証された。

不合格率＝Ｄの比率は12.94％であり，絶対評価基準 AE5の理論確率72.57％より圧倒的に低く，相対評価基準 RE5の10％にかなり近く，次節で検討する卒業不可不合格率よりかなり低いので，平均して見ると専任教員は卒業所要単位数に対する履修可能単位数の割合などを十分に配慮した上で，成績評価をしているといえる。また各成績評価区分のバランスもよくスムーズな正規分布に従っているので，平均して言えばほぼ適正な成績評価をしているといえよう。これは絶対評価基準AE5よりも相対評価基準RE5に非常に近い基準である。この RE5基準は，実はヨーロッパの大学で一般的に採用されている ECTS（European Credit Transfer System）の基準そのものである。

（5-1表）　α 学部の成績評価分布

成績評価	実際の成績評価分布	相対評価基準 RE5
S, A+	100～81.7％(上位18.3％)	100～90％(上位10％)
A	81.6～59.4％(次の22.3％)	89～65％(次の25％)
B	59.3～35％(次の24.4％)	64～35％(次の30％)
C	34.9～13.0％(次の21.9％)	34～10％(次の25％)
D	12.9～0％(下位12.9％)	9～0％(下位10％)

（5-17図）



KS 統計量=0.723，
漸近有意確率=0.672，有意水準=0.05
最遠距離差（正=0.192，負=-0.723）

したがって建前上・形式上はアメリカ型の絶対評価基準 AR5を謳っていても，実際上・実質的にはヨーロッパ型の相対評価基準RE5を採用しているといえる。その背景には，多くの教員は難易度が通常で得点分布が正規分布に近い場合には絶対評価基準 AE5が適用できないことを意識的にあるいは暗黙裏に認識していると見られる。ただし各評価区分のパーセンテージは一律のルールに従っているのではなく，各教員がそれぞれに設定しているので，その総体的結果として上記の比率になったといえる。

こうした状況は同大学の他のすべての学部及び大学全体でも確認することができる。他のすべての学部及び大学全体で専任教員が担当する科目全体の成績評価分布について，有意水準α=0.05（５％），信頼区間0.95で，KS 正規分布検定を実施したところ，すべてが例外なく正規分布に従うことが検証された。よって建前上・形式

上は絶対評価基準AR5を謳っていても，多くの教員が実際上・実質的には相対評価基準RE5を採用しているといえる。

ただしこのことは，教員全員がそうしているということを意味しない。非常に辛い基準で成績評価をする科目が一部にある一方で，非常に甘い基準で成績評価をする科目が一部にあるので，両者の効果が相殺されて，しかも正規分布の成績評価をする科目が相対的に多いために，総体として正規分布になっているといえる。

また日本経済学教育協会では，ミクロ経済学とマクロ経済学の2科目からなる検定試験と，それに財政学，金融論，国際経済，統計学を加えた6科目からなる検定試験を年2回実施しているが，（5-2表）のような7段階の成績評価のランク判定は得点分布が正規分布することを前提としているという。ただしそれが正規分布から大きくずれる場合は試験委員会で調整するとしている。

（5-2表）

ランク	偏差値	範囲
S	73以上	上位1%以上
A+	66～73未満	1.1～5.0%
A	60～66未満	5.1～15.0%
B+	55～60未満	15.1～30%
B	47～55未満	30.1～60.0%
C	37～47未満	60.1～90.0%
D	37未満	90.1～100%

日本英語検定協会やG-TELP日本事務局にヒヤリングしたところ，英検や国際英検の得点分布は受験者層が広いので一部に例外はあるものの，ほぼ正規分布に従うという。

また大手の予備校である河合塾にヒヤリングをしたところ，全国模擬試験における各科目及び全科目の得点分布は，厳密な正規分布検定をしてはいないけれども，例外なく釣り鐘型をしているという。

このように試験の得点分布が正規分布に従う事例は非常に多く，一部に例外はあるとはいえ，普遍的に観察されるといえる。

〔注〕

注 9：KS 検定の着想は最初に Kolmogorov（1933）により与えられ，後にSmirnovの数学的研究により発展させられたため，両者の名前を冠して呼ばれている。

注10：この数学的な導出過程については，Marsaglia, Tsang and Wang (2003) を参照。

注11：これらの検定力の統計学的な比較については，Stephens (1974) を参照。