第7部 統計的思考によって大学入試・統計検定を乗り切ろう！

～大学入試・統計検定問題からみる統計的思考力～

2020年度（平成32年度）から、現在の「大学入試センター試験」に代わって「大学入学希望者学力評価テスト（仮称）」が実施される。そこでは、一体どのような問題が出題されるのだろうか？

2016年３月に公表された文部科学省の報告書では、例えば、国語のイメージ問題例として下記の問題が取り上げられている。

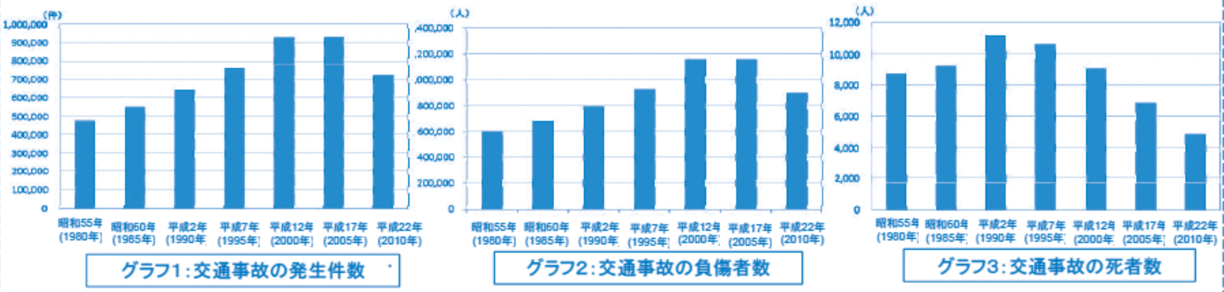
図１　高大接続システム改革会議「最終報告」参考資料イメージ問題例＜国語＞の抜粋

問題イメージ＜例１＞

〔国立教育政策研究所「特定の課題に関する調査（論理的な思考）」（平成24年2月実施）より一部改題〕

次の文章とグラフを読み，後の問いに答えよ。

次に示すのは，警察庁事故統計資料に基づいて作成された交通事故の発生件数，負傷者数，死者数のグラフと，この3つのグラフを見て，交通事故の死者数が他よりも早く，平成2年（1990年）以降減少傾向になっていることについて，4人の高校生が行った話し合いの一部である。

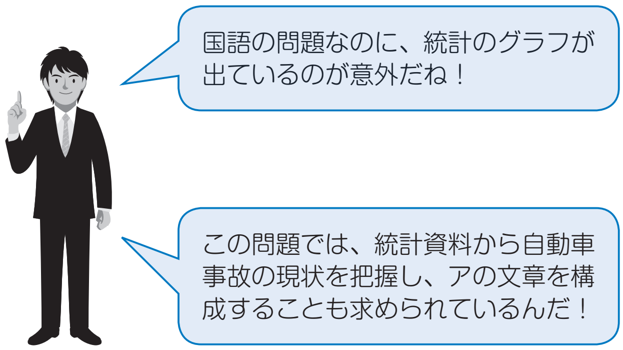


Aさん：交通事故の死者数が他よりも早く，平成2年（1990年）以降減少傾向になっているのは，交通安全に関する国民の意識の変化が関係しているのではないかと思います。
　その裏付けとなる資料として，「交通違反で検挙された人数の推移が分かる資料」があると思います。その資料を見れば，飲酒運転やスピード違反など，死亡事故につながるような重大な違反の割合が少なくなっていることが分かるはずです。

Bさん：私は，この30年間で販売されてきた自動車の台数と安全性に関係があると思います。
(a)つまり，自動車の台数は年々増加し続けているので事故件数と負傷者数はなかなか減らなかったけれども，［ア］［ア］ということです。
　例えば，最近30年間における，「車の総販売台数の推移が分かる資料」と，「車の安全に関する装置の装備率の推移が分かる資料」があれば，このことを裏付けることができると思います。

136

資料：高大接続システム改革会議「最終報告」



この問題では、棒グラフと会話文から情報を解釈し、問題場面に適した考えを構成し、表現することが求められている。新テストでは、教科の枠を超えて統計を活用する能力が求められており、これから統計的思考力を養うことはますます重要になってくるといえる。

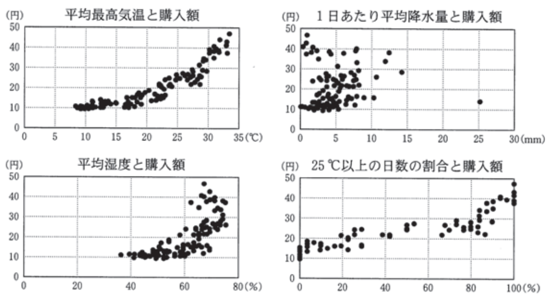
以降では、大学入試や統計検定の過去問題を参考に、現在、どのような統計的思考力が求められているかについて考えていこう。

1　「数学Ⅰ」：データの分析の問題から見る統計的思考力

現行の「数学Ⅰ」データの分析の単元では、箱ひげ図や散布図、相関係数といった記述統計に関わる事柄を学ぶ。2016年度の大学入試センター試験では、アイスクリームの購入金額と気象データとの散布図から、「アイスクリームが売れる条件」を分析する問題が出題された。

〔例題１〕2016年度　大学入試センター試験「数学Ⅰ・A」第２問

〔2〕　次の４つの散布図は，2003年から2012年までの120か月の東京の月別データをまとめたものである。それぞれ，１日の最高気温の月平均(以下，平均最高気温)，１日あたり平均降水量，平均湿度，最高気温25℃以上の日数の割合を横軸にとり，各世帯の１日あたりアイスクリーム平均購入額(以下，購入額)を縦軸としてある。



出典：総務省統計局(2013)『家計調査年報』，『過去の気象データ』(気象庁 Webページ)などにより作成

次の $\boxed{ス}$，$\boxed{セ}$ に当てはまるものを，下の⓪～④のうちから一つずつ選べ。ただし，解答の順序は問わない。

これらの散布図から読み取れることとして正しいものは，$\boxed{ス}$ と $\boxed{セ}$ である。

⓪　平均最高気温が高くなるほど購入額は増加する傾向がある。
①　１日あたり平均降水量が多くなるほど購入額は増加する傾向がある。
②　平均湿度が高くなるほど購入額の散らばりは小さくなる傾向がある。
③　25℃以上の日数の割合が80％未満の月は，購入額が30円を超えていない。
④　この中で正の相関があるのは，平均湿度と購入額の間のみである。

〔解説〕

⓪…「平均最高気温と購入額」の散布図では、正の相関を持つので ○
①…「１日あたり平均降水量と購入額」の散布図では、負の相関を持つので ×
②…「平均湿度と購入額」の散布図では、平均湿度が高くなるほど購入額の散らばりは大きくなるので ×
③…「25℃以上の日数の割合と購入額の散布図」では、横軸が80％未満の範囲で、縦軸の購入金額が30円を超えていないので ○
④…「平均湿度と購入額」のほか「平均最高気温と購入額」にも正の相関があるので ×

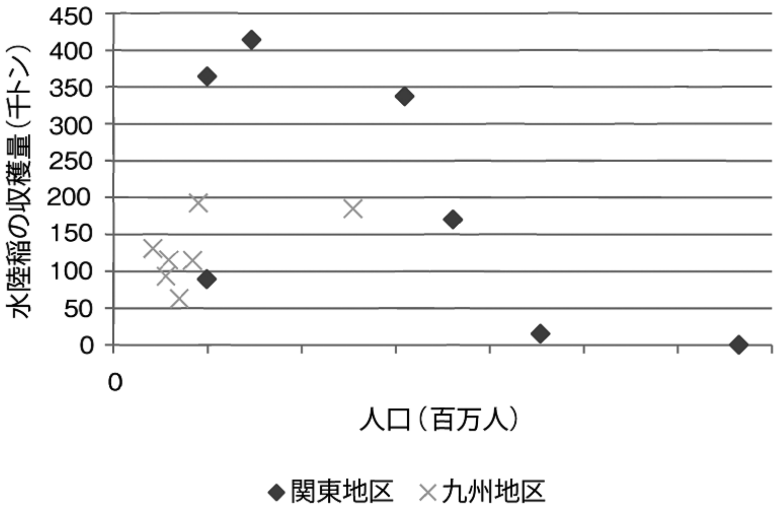
答：⓪と③

やはり、気温が高い日や湿度が高い日にアイスクリームはよく売れる傾向にあるのですね。

〔練習１〕社会データの分析

（１） 2014年11月実施　統計検定３級問題「水陸稲の収穫量と人口の相関」

〔2〕各都道府県の人口と水陸稲の収穫量の相関係数を関東地区の 7 都県（群馬，栃木，茨城，埼玉，千葉，東京，神奈川）と九州地区の７県（福岡，佐賀，長崎，熊本，大分，宮崎，鹿児島）について調べたところ，散布図は次のようになり，それぞれの相関係数は関東地区で −0.67，九州地区で 0.60 であった。



資料：総務省「人口推計」および農林水産省「作物統計」

これについて，次の I〜III の記述を考えた。

I. 人口と水陸稲の収穫量について，関東地区には負の相関がある。

II. 人口と水陸稲の収穫量について，九州地区には正の相関がある。

III. 関東地区と九州地区を合わせた 14 都県での相関係数は $(-0.67 + 0.60) \div 2 = -0.035$ と計算できる。

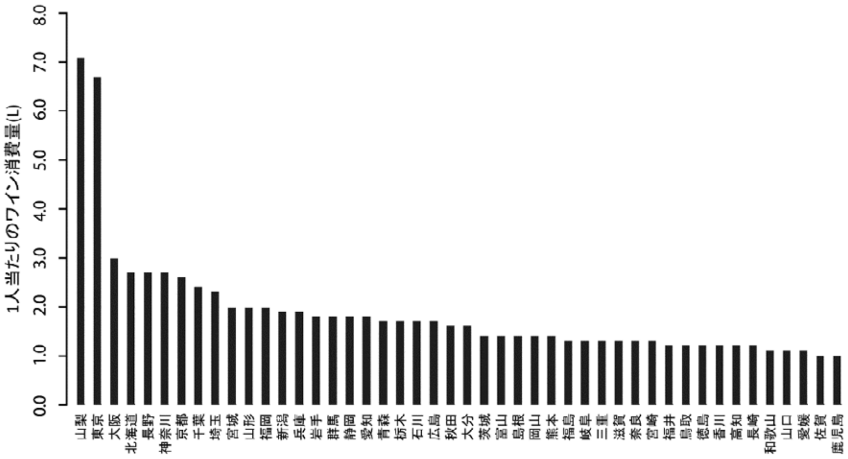
この記述 I〜III に関して，次の ①〜⑤ のうちから最も適切なものを一つ選べ。

$\boxed{16}$

① I のみ正しい。
② II のみ正しい。
③ I と II のみ正しい。
④ II と III のみ正しい。
⑤ I と II と III はすべて正しい。

（２）　2014年６月実施　統計検定３級問題「ワインの消費量を表す箱ひげ図の選択」

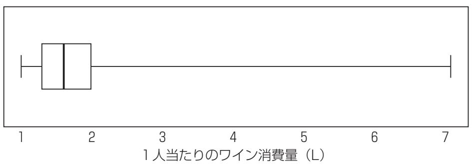
問17　次の棒グラフは、平成22年度の沖縄県を除く46都道府県の成人１人当たりのワイン消費量（L）を表している。



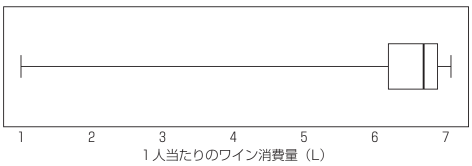
資料：国税庁課税部酒税課「平成22年度酒類販売（消費）数量表」

この結果の中で，7.1Ｌの山梨県が最も多く，次いで東京都の6.7Ｌであり，最も少ないのは鹿児島県，佐賀県で消費量は1.0Lであった。この棒グラフから作成した46の都道府県のワイン消費量の箱ひげ図として①～⑤のうちから最も適切な者を一つ選べ。 $\boxed{24}$

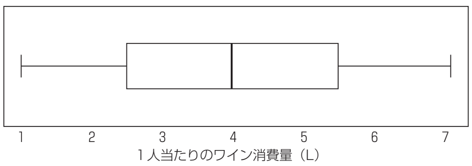
①



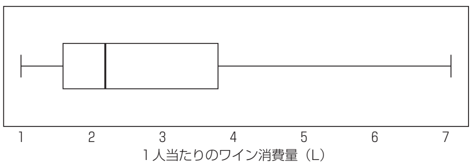
②



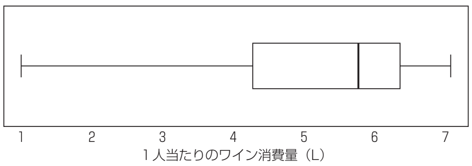
③



④



⑤



2 「数学 B」：確率分布と統計的な推測の問題から見る統計的思考力

現行の「数学 B」確率分布と統計的な推測の単元では、二項分布や正規分布、区間推定といった推測統計に関わる事柄を学ぶ。2015年度の大学入試センター試験では、標準正規分布の性質を利用した区間推定の公式そのものの意味理解を問う問題が出題された。

〔例題 2〕2015年度　大学入試センター試験「数学Ⅱ・B」第 5 問

第 5 問 （選択問題）（配点　20）

以下の問題を解答するにあたっては，必要に応じて 29 ページの正規分布表を用いてもよい。

また，小数の形で解答する場合，指定された桁(けた)数の一つ下の桁を四捨五入し，解答せよ。途中で割り切れた場合，指定された桁まで⓪にマークすること。

(2)　確率変数 Z が標準正規分布に従うとき

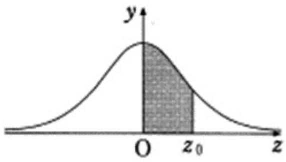
$$P\left(-\boxed{\text{タ}} \leqq Z \leqq \boxed{\text{タ}}\right) = 0.99$$

が成り立つ。$\boxed{\text{タ}}$ に当てはまる最も適切なものを，次の⓪～③のうちから一つ選べ。

⓪ 1.64　　① 1.96　　② 2.33　　③ 2.58

正 規 分 布 表

次の表は，標準正規分布の分布曲線における右図の灰色部分の面積の値をまとめたものである。



z_0	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

(3)　母標準偏差 σ の母集団から，大きさ n の無作為標本を抽出する。ただし，n は十分に大きいとする。この標本から得られる母平均 m の信頼度(信頼係数) 95 % の信頼区間を $A \leqq m \leqq B$ とし，この信頼区間の幅 L_1 を $L_1 = B - A$ で定める。

この標本から得られる信頼度 99 % の信頼区間を $C \leqq m \leqq D$ とし，この信頼区間の幅 L_2 を $L_2 = D - C$ で定めると

$$\frac{L_2}{L_1} = \boxed{チ}.\boxed{ツ}$$

が成り立つ。また，同じ母集団から，大きさ $4n$ の無作為標本を抽出して得られる母平均 m の信頼度 95 % の信頼区間を $E \leqq m \leqq F$ とし，この信頼区間の幅 L_3 を $L_3 = F - E$ で定める。このとき

$$\frac{L_3}{L_1} = \boxed{テ}.\boxed{ト}$$

が成り立つ。

〔解説〕

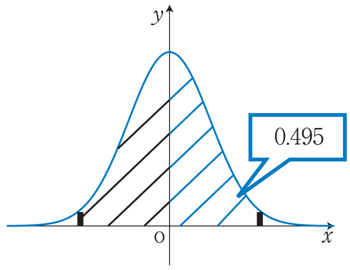
(2)　$P(0 \leqq Z \leqq z_0) = 0.99 \div 2 = 0.495$

したがって、正規分布表より右の図の状況を満たす

z_0は2.57と2.58の間であるから

$$z_0 = \frac{(2.57 + 2.58)}{2} = 2.575 \fallingdotseq 2.58$$

答：③



(3)　(2)と同様の考え方を用いて、$P(-1.96 \leqq Z \leqq 1.96) = 0.95$

母平均 m、母標準偏差 σ、大きさ n の標本の標本平均を $\overline{X}$ とすると、$\overline{X}$ は平均 m、分散 σ^2/n の標準正規

分布に従うので　$P\left(-1.96 \leqq \dfrac{\overline{X} - m}{\dfrac{\sigma}{\sqrt{n}}} \leqq 1.96\right) = 0.95$ と表せる。

母平均 m の信頼度95％信頼区間は、上のかっこ内の式を変形して、

$$\overline{X} - 1.96\frac{\sigma}{\sqrt{n}} \leqq m \leqq \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}$$

したがって、　$L_1 = \overline{X} + 1.96\dfrac{\sigma}{\sqrt{n}} - \left(\overline{X} - 1.96\dfrac{\sigma}{\sqrt{n}}\right) = 2 \times 1.96\dfrac{\sigma}{\sqrt{n}}$　…①

同様に、$P(-2.58 \leqq Z \leqq 2.58) = 0.99$であるから、母平均 m の信頼度99％信頼区間は

$$\overline{X} - 2.58\frac{\sigma}{\sqrt{n}} \leqq m \leqq \overline{X} + 2.58\frac{\sigma}{\sqrt{n}}$$

したがって、　$L_2 = \overline{X} + 2.58\dfrac{\sigma}{\sqrt{n}} - \left(\overline{X} - 2.58\dfrac{\sigma}{\sqrt{n}}\right) = 2 \times 2.58\dfrac{\sigma}{\sqrt{n}}$…②

①、②より

$\dfrac{L_2}{L_1} = \dfrac{2.58}{1.96} = 1.316... \fallingdotseq 1.3$　　　　答：$\dfrac{L_2}{L_1} \fallingdotseq 1.3$

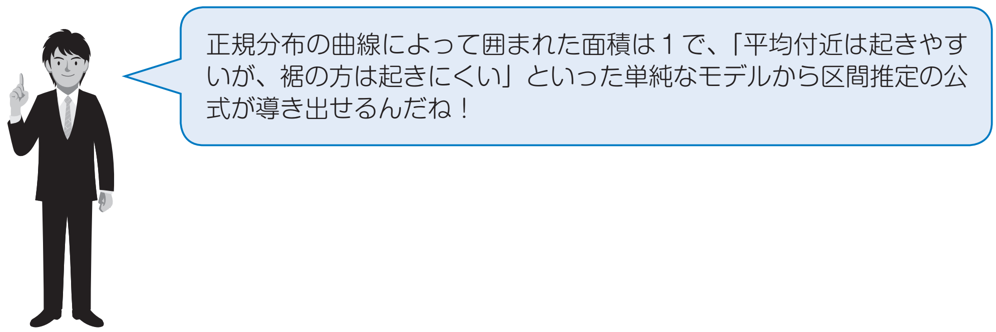
また、大きさ 4 n の標本の標本平均を $\overline{X}'$ とすると、$\overline{X}'$ は平均 m、分散 $\sigma^2/4n$ の標準正規分布に従うので、母平均 m の信頼度95％信頼区間は

$$\overline{X}' - 1.96\frac{\sigma}{\sqrt{4n}} \leqq m \leqq \overline{X}' + 1.96\frac{\sigma}{\sqrt{4n}}$$

したがって、　$L_3 = \overline{X}' + 1.96\dfrac{\sigma}{\sqrt{4n}} - \left(\overline{X}' - 1.96\dfrac{\sigma}{\sqrt{4n}}\right) = 2 \times 1.96\dfrac{\sigma}{\sqrt{4n}} = 1.96 \times \dfrac{\sigma}{\sqrt{n}}$…③

①、③より

$\dfrac{L_3}{L_1} = \dfrac{1}{2} = 0.5$　　　　答：$\dfrac{L_3}{L_1} = 0.5$



〔練習２〕正規分布の利用

2014年度　鹿児島大学２次試験の問題「合格最低点の推定」

ある企業の入社試験は採用枠300名のところ500名の応募があった。試験の結果は500点満点の試験に対し、平均点245点、標準偏差50点であった。得点の分布が正規分布であるとみなされるとき、合格最低点はおよそ何点であるか。小数点以下を切り上げて答えよ。ただし、確率変数Ｚが標準正規分布に従うとき、$P(Z>0.25)=0.4$、$P(Z>0.5)=0.3$、$P(Z>0.54)=0.2$とする。

〔鹿児島大学　理・工・医（医）・歯〕

３　「大学」：統計学の問題から見る統計的思考力

大学の統計学の授業では、正規分布以外の確率分布や検定といった、さらに専門的な推測統計学に関わる内容を学ぶ。高等学校では、「正規分布の裾の部分が起きにくい」といった程度しか言及しないが、実は、正規分布の裾の見方は、検定で重要となる“棄却域”の考え方につながる。たとえば、2014年６月に実施された統計検定２級では、私たちも馴染み深い「二項分布」を題材に検定の考え方を問う問題が出題された。

〔例題３〕2014年６月実施　統計検定２級問題「公正なサイコロか？」

問 13　サイコロを 7 回投げたときに 3 の目が 4 回出たことから，公正なサイコロ（どの目も出る確率は 6 分の 1）であるかどうか疑問となった。

〔1〕サイコロが公正であるとして，サイコロを 7 回投げたときに 3 の目が 4 回出る確率を求める式として正しいものはどれか。次の ① ～ ⑤ のうちから適切なものを一つ選べ。 $\boxed{20}$

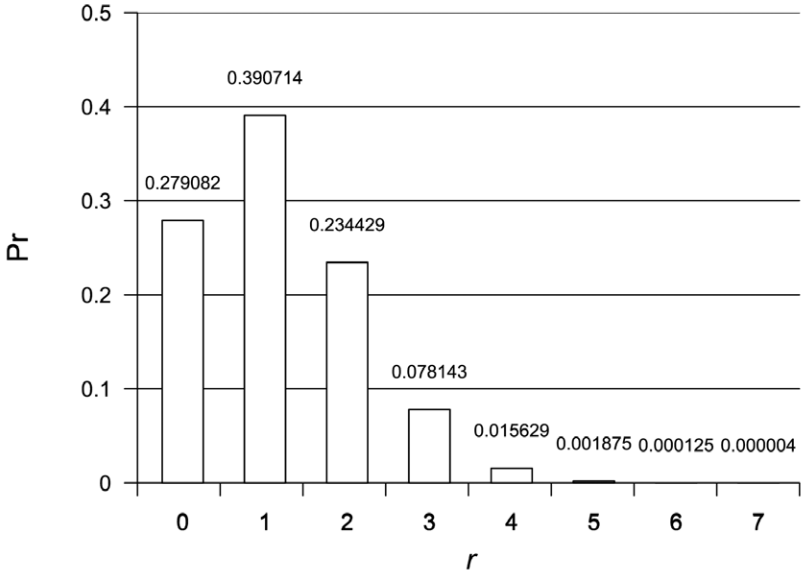
① $840\times\left(\frac{1}{6}\right)^4\times\left(\frac{5}{6}\right)^3$　　② $35\times\left(\frac{1}{6}\right)^4$

③ $\frac{1}{35}\times\left(\frac{1}{6}\right)^4$　　④ $35\times\left(\frac{1}{6}\right)^4\times\left(\frac{5}{6}\right)^3$

⑤ $\frac{1}{35}\times\left(\frac{1}{6}\right)^4\times\left(\frac{5}{6}\right)^3$

〔2〕このサイコロが公正なサイコロであるかどうかを調べることにする。そのために，このサイコロが公正なサイコロであった場合に，サイコロを 7 回投げたときに 3 の目が r 回出る確率を求め，判断することにした。

その確率 Pr を r = 0, 1, ⋯ , 7 の場合についてグラフにすると，次の図のようになる。



r = 0, 1, ⋯ , 7 のときの確率

帰無仮説として「このサイコロの 3 の目の出る確率は 6 分の 1 である」を用い，有意水準 5% で仮説検定を行ったとき，結論として正しい判断をしているものはどれか。次の ① ～ ⑤ のうちから適切なものを一つ選べ。 $\boxed{21}$

① 3 の目が 4 回以上出る確率は 5% より小さい。したがって，帰無仮説を棄却して，公正なサイコロでないと結論する。

② 3 の目がちょうど 4 回出る確率は 0.015629 となり，5% より小さい。したがって，帰無仮説を棄却して，公正なサイコロでないと結論する。

③ 3 の目が何回出るかは事前にわからず，今回はたまたま 4 回出ただけである。したがって，公正なサイコロと結論する。

④ 3 の目が出る回数が 4 回未満である場合で最も確率が小さいのは 3 回の 0.078143 で，5% より大きい。したがって，公正なサイコロと結論する。

⑤ 3 の目が 4 回出る場合の前後を考慮し，3 回，4 回，5 回出る確率を足すと 0.095647 となり 5% より大きい。したがって，帰無仮説は棄却できず，公正なサイコロと結論する。

〔解説〕

⑴　サイコロを７回投げたときに３の目が出た回数を確率変数Ｘとおく。

３の目が４回出る確率は、

$$\mathrm{P}(\mathrm{X}=4)={}_7\mathrm{C}_4\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)^3=35\times\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)^3$$

答：④

⑵　①…３の目が出る確率は１/６であると仮定し、右片側検定で考えると、

３の目が４回以上出る確率は、

$$\mathrm{P}(X\geqq 4)=P(X=4)+P(X=5)+P(X=6)+P(X=7)$$
$$=0.015629+0.001875+0.000125+0.000004=0.017633$$

有意水準５％で、$\mathrm{P}(\mathrm{X}\geqq 4)<0.05$であるから、サイコロを７回投げて３の目が４回以上出ることはめったにないといえる。したがって、帰無仮説は棄却できるので ○

②…４回ちょうどの場合を考えても、片側検定・両側検定の考え方に適さないので ×

③…仮説検定の枠組みになっていないので ×

④…３の目が４回も多く出たことが疑問であるので、４回未満の左片側検定をそもそも考える必要がないので ×

⑤…３の目が４回出る前後を考えても、②と同様に仮説検定の考え方に適さないので ×

答：①

二項分布で、右裾の３の目が４回以上出る確率はすごく小さい。だから、それが「有意水準５％未満の確率ならば、めったに起きないことが起きたと判断する」といった、仮説検定ならではの考え方が問われているんだね。

〔練習 3〕仮説検定の考え方

2014年 6 月実施　統計検定 2 級問題「仮説検定における用語の確認」

問 14　次の文章中の（ア）～（オ）にあてはまる用語の正しい組合せとして，下の ① ～ ⑤ のうちから最も適切なものを一つ選べ。 $\boxed{23}$

統計的推測では，母集団からの標本として得られたデータを用いて，測定値を得た対象だけに限定することなく，母集団について何らかの判断を下す方法論を扱っている。統計的推測は大きく 2 つに分けられる。（ア）と（イ）である。（イ）は，はじめに母集団に対して（ウ）と呼ばれる特定の仮説を設定し，観測したデータがこの仮説を否定するかどうかを調べる手法である。このとき，仮説を支持するか否かは，確率を伴う判断が必要となる。（ウ）が正しいにも関わらず，（ウ）を否定してしまう確率をある値以下にする必要がある。この値を（エ）と呼ぶ。（エ）を決めることで，（オ）と呼ばれる判定のための領域を決めることができる。

① （ア）推定，（イ）検定，（ウ）帰無仮説，（エ）第 1 種過誤の確率，（オ）検出力
② （ア）検定，（イ）推定，（ウ）対立仮説，（エ）P-値，（オ）有意水準
③ （ア）推定，（イ）検定，（ウ）対立仮説，（エ）有意水準，（オ）棄却域
④ （ア）検定，（イ）推定，（ウ）帰無仮説，（エ）P-値，（オ）有意水準
⑤ （ア）推定，（イ）検定，（ウ）帰無仮説，（エ）有意水準，（オ）棄却域

〔本節の解答〕

〔練習 1〕 (1) ③ (2) ①

〔練習 2〕 試験の得点を X 点とし、500名中300名が合格することから、

$$\mathrm{P}\left(\frac{X-245}{50} \geqq z_0\right)=\frac{300}{500}(=0.6)$$

$\mathrm{P}\left(\frac{X-245}{50} \geqq z_0\right)=1-\mathrm{P}\left(\frac{X-245}{50} \leqq z_0\right)$ であるので、$\mathrm{P}\left(\frac{X-245}{50} \leqq z_0\right)=\mathrm{P}(Z \leqq z_0)=0.4$ を満たす X は不合格となる最高点を与える。

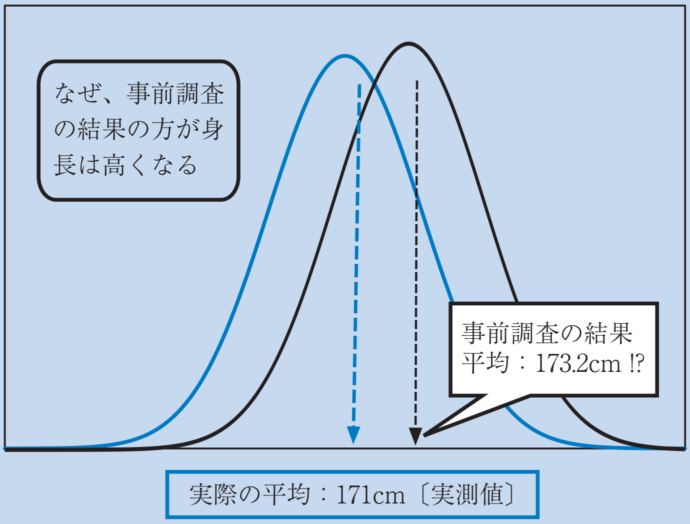
正規分布は左右対称であるから、$\mathrm{P}(Z \leqq z_0)=\mathrm{P}(Z \geqq -z_0)=0.4$

$-z_0=0.25$ より、$X \leqq 245-50\times 0.25$ であるので、232.5点以下の得点は不合格となる。したがって、合格最低点は約233点。

〔練習 3〕 ⑤

正規分布の歪みから見る若者たちの身長の真実

ある地域の17歳の男子生徒の身長の分布（cm）



文部科学省の「平成27年度学校保健調査」によると、日本の17歳の男子生徒の平均身長は、170.7cm です。ある地域では、どの年も、17歳の男子生徒の身長の測定結果は、全国結果とほぼ変わらない傾向であることが知られていました。しかし、事前にアンケート調査を実施すると、その結果（平均身長173.2cm）は、実際の測定結果（平均身長171cm）より高い数値を示しました。このことから、正規分布を用いることで､「生徒たちは、事前調査では、実際の身長より高めの身長を回答した」と推測できます。

一方、この話題とは逆に、統計学者のアドルフ・ケトレー（Adolphe Quételet）は、1844年に男性の身長が正規分布に従うことを利用して、フランス軍の徴兵検査の際に測定された若者たちの身長について、次のウソを見抜きました。

当時の記録では、157cm よりやや背が高い者が少なく、逆に157cm よりやや背が低い者の数が極端に多かったようです。そのため、身長の分布が一部凹んだ正規分布になりました。この結果を受け、ケトレーは、当時のフランス軍は身長157cm 以上の若者を徴兵していたことを踏まえ､「157cm よりわずかに身長が高い若者たちの何人かが、身長を低くごまかし、徴兵から逃れた」と推測しました。

したがって、この正規分布の歪みは､「徴兵を逃れたい、157cm よりやや高い若者たちのごまかしによって、現れた歪みである」と結論付けることができます。

フランスの徴兵検査の際には、モデルとして正規分布を利用したが、グラフの一部が歪んだことで、ケトレーにより記録の偽りが見抜かれました。このように、現実の事象を分析する際には、状況に合わせた確率分布をモデルとして用いることで、さまざまな統計分析を行うことができます。