

2014 年 RSS/JSS 試験 (Higher Certificate)



HIGHER CERTIFICATE IN STATISTICS, 2014

モジュール 1 : データの収集と解釈

制限時間: 90 分

4 問中 3 問を選択の上解答のこと.

各問は合計 20 点である. 小問の配点は括弧の中に記されている.

グラフ用紙と統計数値表は配布する.

解答にあたっては電卓を使用してよい.
ただし, 一般財団法人統計質保証推進協会による「受験要領」に記された範囲で使用する.

数学記号 \log は e を底とする自然対数を表す.
その他の底をもつ対数は, 例えば \log_{10} のように底を明示する.

また, $\binom{n}{r}$ は ${}_nC_r$ と同じ意味とする.

問題用紙は 8 頁からなり, それぞれの頁は片面にのみ印刷されている.

この表紙が 1 頁目である.
第 1 問は 2 頁目から始まる.

問題は全部で 4 問である.

1. ある難関大学では、毎年数千人の学生が入学している。この入学生の全体的な能力を調べるために、毎年新入生の一部に対し、1時間の学力全般の試験が行われる。この試験は志願した最初の200人の学生に対して課される。学生は監督のもとで試験を受け、その試験に費やした時間に対して幾らかの謝礼が与えられる。この試験の得点は、仮にその国の全成人に対して行われたならば、平均100、標準偏差15の正規分布になるように設計されている。

表1は過去5年間の点数の要約である。

年	2009	2010	2011	2012	2013
平均	132.8	132.6	130.1	127.9	125.7
標準偏差	8.6	8.1	8.8	9.4	9.5
最低点	108.9	114.2	105.5	94.5	97.6
5パーセンタイル値	118.9	119.1	115.9	113.0	109.6
25パーセンタイル値	126.6	126.6	124.7	121.4	120.1
50パーセンタイル値	132.6	132.7	130.4	128.8	126.0
75パーセンタイル値	139.8	138.6	135.9	134.9	131.6
95パーセンタイル値	146.9	145.6	144.4	141.8	141.4
最高点	153.6	151.7	150.0	150.0	148.0

表1

表2はその国の全成人でのこの試験のいくつかの点数に対する累積割合である。

得点	100	105	110	115	120	125	130	135	140	145	150
全成人での累積割合	50.0	63.1	74.8	84.1	90.9	95.2	97.7	99.0	99.6	99.9	100.0

表2

- (i) 過去5年間の新入生の能力について、年毎に比較し、また適切な方法で全成人と比較を行い、このデータが表していることを要約したレポートを書け。レポートにはデータの要約と統計図表を適切な場所に含め、特別な統計の知識を持っていない人に分かるように書け。(10)
- (ii) (i)のレポートの中で示したデータをどのように選んだのかを簡潔に述べよ。もし用いなかった特定のデータ項目があるならば、その理由を説明せよ。(4)
- (iii) この情報を集める手順に関して大学にコメントし、相応しいと思われる変更があれば勧告せよ。(6)

2. 客船タイタニック号は 1912 年に沈没し, 3 つ全ての等級の乗客と乗員の多くの命が奪われた. 事件後, 一等客室の乗客は三等客室の乗客よりも救命ボートに乗るチャンスが高かったと言われている. また男性客は女性と子供を優先して救命ボートに乗せたとの数々の物語がある.

以下の表は様々なカテゴリーの乗客, 乗員の生存者数と死亡者数である.

等級	年齢区分	性別	死亡者数	生存者数	総数
一等	大人	女性	4	140	144
		男性	118	57	175
	子供	女性		1	1
		男性		5	5
二等	大人	女性	13	80	93
		男性	154	14	168
	子供	女性		13	13
		男性		11	11
三等	大人	女性	89	76	165
		男性	387	75	462
	子供	女性	17	14	31
		男性	35	13	48
乗員	大人	女性	3	20	23
		男性	670	192	862
総数			1490	711	2201

以下の間に答えるためにこのデータを解析せよ. 各々の問において, 必要ならば, 適切な統計図表を用いて結論を説明せよ.

- (i) 等級間で生存率はどのように異なるか? (注:この問と(ii)において, 乗員は一つの等級として扱うこと) (6)
- (ii) 等級間で共通する特徴, また等級における特質について述べよ. (9)
- (iii) このデータに基づき, 救命ボートに乗る際に「女性と子供を優先すべき」という方針を裏付ける, あるいは反証するための証拠を示せ. (5)

- 3 スペインのバリャドリッド地域において、65歳以上男性の体重の標本調査が行われた。この地域にはこの年齢層の男性が42,000人いて、その中から508人の標本を抽出した。年齢と居住形態（自宅、公的介護施設、民間介護施設）によってカテゴリー化された。

表1はカテゴリー毎の人数を示している。

年齢 (歳)	自宅	公的 介護施設	民間 介護施設	総数
65-69	41	16	18	75
70-74	41	21	21	83
75-79	36	28	27	91
80-84	34	20	26	80
85-89	37	17	23	77
90-94	25	14	16	55
95歳以上	23	10	14	47
総数	237	126	145	508

表1

表2は各カテゴリーにおける男性の体重（キログラム）の平均を示している。

年齢 (歳)	自宅	公的 介護施設	民間 介護施設	全体
65-69	66.1	68.0	67.4	66.8
70-74	65.0	67.5	68.1	66.4
75-79	63.8	66.2	66.5	65.3
80-84	61.7	65.3	64.0	63.3
85-89	60.7	63.4	62.8	a
90-94	58.0	62.0	60.1	59.6
95歳以上	58.8	61.4	58.9	59.4
全体	62.5	65.3	b	63.7

表2

- (i) 表2における a と b の値を求めよ。 (2)

- (ii) 調査報告書は、調査された508人全員の体重の標準偏差は9.7 kgだと報告している。508人の体重の標本平均の標準誤差を推定せよ。

各カテゴリーにおける標準偏差は全て9.7 kgであると仮定して、次のカテゴリーにおける体重の標本平均の標準誤差を推定せよ。

- (a) 65歳以上69歳以下の自宅居住者
(b) 公的介護施設居住者

標準誤差はデータの解釈に際してどのように使われるか説明せよ。

(6)

問題は次のページに続く

- (iii) この 508 人を選ぶために層別抽出で選んだならば、どのようにすべきか説明せよ。層別抽出の長所を 1 つ述べよ。

集落抽出について説明し、その長所を 1 つ述べよ。この事例において集落抽出を用いることが難しいのは何故か？

(8)

- (iv) データの部分集合間の比較、対比のために適切な図を描き、このデータにおいて明らかな主要なパターンを説明せよ。

(4)

- 4 ボランティアグループが地域ラジオ局を立ち上げようと思っている。ラジオ局は非営利企業となり、極めて限られた資金しかないが、ウェブサイトを持っている。このラジオは地元で放送され、インターネット上でも聴取できるようにする。

ボランティアグループは、放送予定地域において、聴取者となり得る人たちが最も重要だと思ふ話題を確認するために市場調査を行うこととした。またこの市場調査を実施することによって、ラジオ局の知名度を高め、放送開始後の聴取者を増やし、ラジオ局を手伝う新たなボランティアを増やすことも期待されている。

- (i) ボランティアによって街の中心部で通行人に尋ねるためのアンケートを設計せよ。アンケートでは、次のトピックに関し、要約と処理に適した形式で情報を集める必要がある。

- 回答者の年齢と性別
- 現在のラジオ聴取形態一例：時間帯、場所、放送局
- 好きな番組内容一例：ニュース、スポーツ、天気、交通状況（各々に関して地元のものど全国のもの）、トークショー、電話による聴取者参加型、好きな音楽と嫌いな音楽の種類
- 地域ラジオ局を立ち上げることに對する一般的な反応
- 宗教的な番組に対する意見（ラジオ局は宗教番組を放送することで地元の宗教団体から資金を得られるかもしれない）

(12)

- (ii) ラジオ局のウェブサイト上で回答できるように、オンライン版のアンケートも作られる。オンライン版を作る際に、加えて考慮すべき設計上の問題を簡潔に述べなさい。

(4)

- (iii) 聞き取りやオンラインで必要な情報を集める際に、バイアスや誤差が生じ得る要因を簡潔に説明しなさい。

(4)

BLANK PAGE

BLANK PAGE

BLANK PAGE