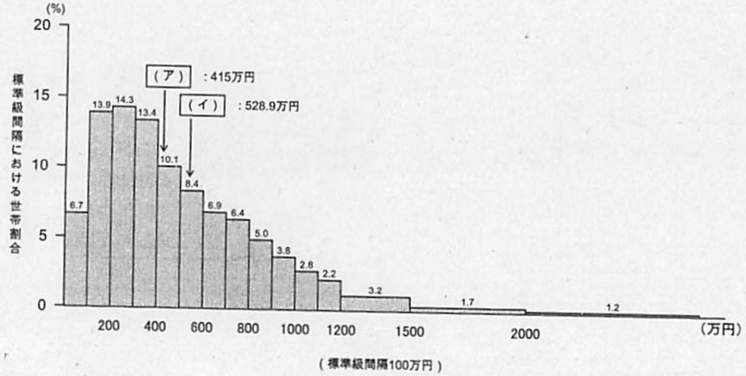
問 3　次のヒストグラムは，平成 26 年国民生活基礎調査（厚生労働省）をもとに作成した所得金額階級別の世帯数の相対度数分布（所得金額の分布）である。ヒストグラムの柱の上の数値は，対応する階級に属する世帯割合（%）を示している。ただし小数点以下の桁表示の関係上，世帯割合の合計が 100 %になるようにした。



資料：厚生労働省「平成 26 年国民生活基礎調査」

〔1〕図中の（ア），（イ）にあてはまるものの組合せとして，次の ① ～ ⑤ のうちから最も適切なものを一つ選べ。 **5**

① （ア）平均　（イ）中央値
② （ア）中央値　（イ）平均
③ （ア）平均　（イ）最頻値
④ （ア）中央値　（イ）最頻値
⑤ （ア）最頻値　（イ）平均

〔2〕次の記述 I ～ III はこのヒストグラムに関するものである。

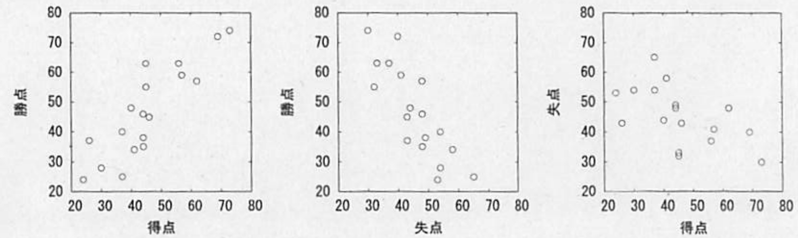
> I. 第 1 四分位数は，100 万円以上 200 万円未満の階級に含まれる。
> II. 四分位範囲の取りうる値は 400 万円以上 600 万円以下である。
> III. この所得金額の分布は右に裾が長い。

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。

6

① I のみ正しい。
② II のみ正しい。
③ III 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。

問 4　次の散布図は，2015 年の J リーグ（J1）18 チームの年間 34 試合の合計の勝点，得点，失点の散布図である。なお，勝点 = 勝ち数 × 3 + 引き分け数，である。



資料：JAPAN PROFESSIONAL FOOTBALL LEAGUE のホームページ

〔1〕得点と勝点の相関係数はいくらか。次の ① ～ ⑤ のうちから最も適切なものを一つ選べ。 **7**

① −0.67　② −0.43　③ 0.12　④ 0.31　⑤ 0.87

〔2〕次の記述 I ～ III はこれらの散布図に関するものである。

> I. 一番得点が多かったチームは，一番失点が少なかったチームである。
> II. 得点と勝点の相関および失点と勝点の相関に比べ，得点と失点の相関は弱い。
> III. 得点の範囲は失点の範囲よりも小さい。

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。

8

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

問 11　$X_1, \ldots, X_n$ は平均 μ，分散 σ^2 の分布に互いに独立に従うものとし，μ と σ^2 はともに未知であるとする。

〔1〕X_1^2 の期待値として，次の ①～⑤ のうちから適切なものを一つ選べ。 **21**

① μ^2　② $\mu^2 - \sigma^2$　③ $\mu^2 + \sigma^2$
④ $(\mu + \sigma)^2$　⑤ $\mu^2 + 3\sigma^2$

〔2〕次の文章は σ^2 と μ^2 の推定量について述べたものである。

「$X_1, \ldots, X_n$ の標本平均を

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

とおく。（ア）は σ^2 の不偏推定量であるので，μ^2 の不偏推定量の 1 つは（イ）で与えられる。」

（ア），（イ）にあてはまるものの組合せとして，次の ①～⑤ のうちから適切なものを一つ選べ。 **22**

① （ア）$\hat{\sigma}^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i - \bar{X})^2$　（イ）$\bar{X}^2 - \hat{\sigma}^2$

② （ア）$\hat{\sigma}^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i - \bar{X})^2$　（イ）$\bar{X}^2 - \dfrac{\hat{\sigma}^2}{n}$

③ （ア）$\hat{\sigma}^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \bar{X})^2$　（イ）$\bar{X}^2 - \hat{\sigma}^2$

④ （ア）$\hat{\sigma}^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \bar{X})^2$　（イ）$\bar{X}^2 - \dfrac{\hat{\sigma}^2}{n}$

⑤ （ア）$\hat{\sigma}^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \bar{X})^2$　（イ）$\bar{X}^2$

問 12　あるコインを投げたとき，表が出る確率を p，裏が出る確率を $1-p$ とし，p は未知であるとする。表が出る確率がある特定の値かどうかを検証するために，n 回コインを投げ，そのうち表が出た回数の割合を使って p を推定する。第 i 回目のコイン投げの結果，表が出たら $X_i = 1$，裏が出たら $X_i = 0$ となる確率変数 $X_i\ (i = 1, \ldots, n)$ を使って，p の推定量 $\hat{p}$ を

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

とする。

〔1〕$\hat{p}$ の分散の取りうる最大値はいくらか。次の ①～⑤ のうちから適切なものを一つ選べ。 **23**

① $\dfrac{1}{n}$　② $\dfrac{1}{2n}$　③ $\dfrac{1}{3n}$　④ $\dfrac{1}{4n}$　⑤ $\dfrac{1}{5n}$

〔2〕次の文章は表が出る確率が p_0 であるという仮説を検定する手続きについて述べたものである。

「帰無仮説 $H_0 : p = p_0$，対立仮説 $H_1 : p \neq p_0$ に対して，検定統計量を

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

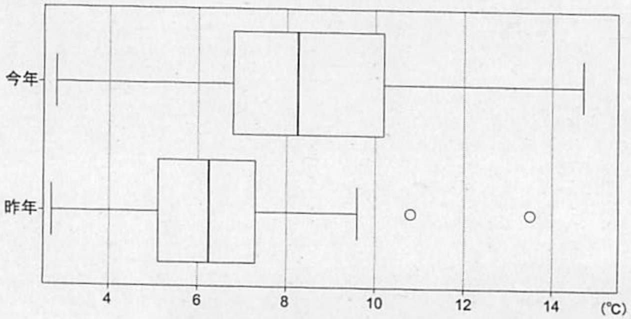
とする。n が十分大きいとき，Z は H_0 の下では標準正規分布で近似できる。この検定は（ア）検定であり，有意水準を 5 %とすると，$|Z| >$（イ）となるとき，（ウ）仮説は有意水準 5 %で棄却される。」

（ア）～（ウ）にあてはまるものの組合せとして，次の ①～⑤ のうちから最も適切なものを一つ選べ。 **24**

① （ア）片側　（イ）1.645　（ウ）対立
② （ア）片側　（イ）1.645　（ウ）帰無
③ （ア）片側　（イ）1.96　（ウ）帰無
④ （ア）両側　（イ）1.96　（ウ）対立
⑤ （ア）両側　（イ）1.96　（ウ）帰無

問1　「今年は暖冬であった」とのニュースを聞き，実際どの程度気温が違うのか興味を持った。そこで，2015年12月1日から50日間の東京の日平均気温（今年の日平均気温と呼ぶ）と，2014年12月1日から50日間の東京の日平均気温（昨年の日平均気温と呼ぶ）を調べた。

〔1〕東京の今年と昨年の日平均気温の箱ひげ図が次のように得られた。なお，この箱ひげ図では，“「第1四分位数」－「四分位範囲」× 1.5”以上の値をとるデータの最小値，および“「第3四分位数」＋「四分位範囲」× 1.5”以下の値をとるデータの最大値までひげを引き，これらよりも遠い値を外れ値として○で示している。



資料：気象庁ホームページ「過去の気象データ検索」

次の記述Ⅰ～Ⅲはこの箱ひげ図に関するものである。

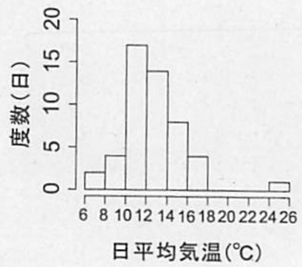
Ⅰ. 今年の日平均気温の標準偏差は昨年の標準偏差より小さい。
Ⅱ. 今年の日平均気温の中央値は昨年の中央値より約2℃高い。
Ⅲ. 今年の日平均気温の範囲は昨年の日平均気温の範囲より約4℃小さい。

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。 1

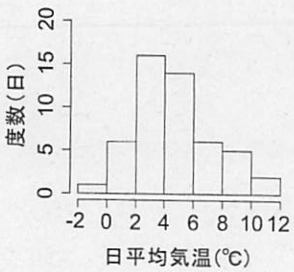
① Ⅰのみ正しい。
② Ⅱのみ正しい。
③ Ⅲのみ正しい。
④ ⅠとⅡのみ正しい。
⑤ ⅠとⅢのみ正しい。

〔2〕今年の日平均気温のヒストグラムとして，次の①～⑤のうちから最も適切なものを一つ選べ。 2

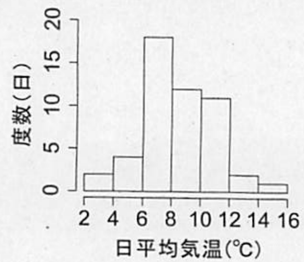
①



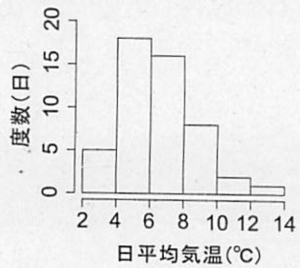
②



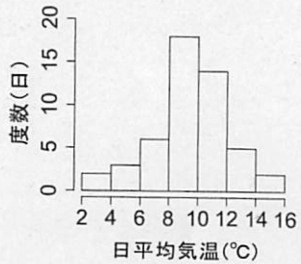
③



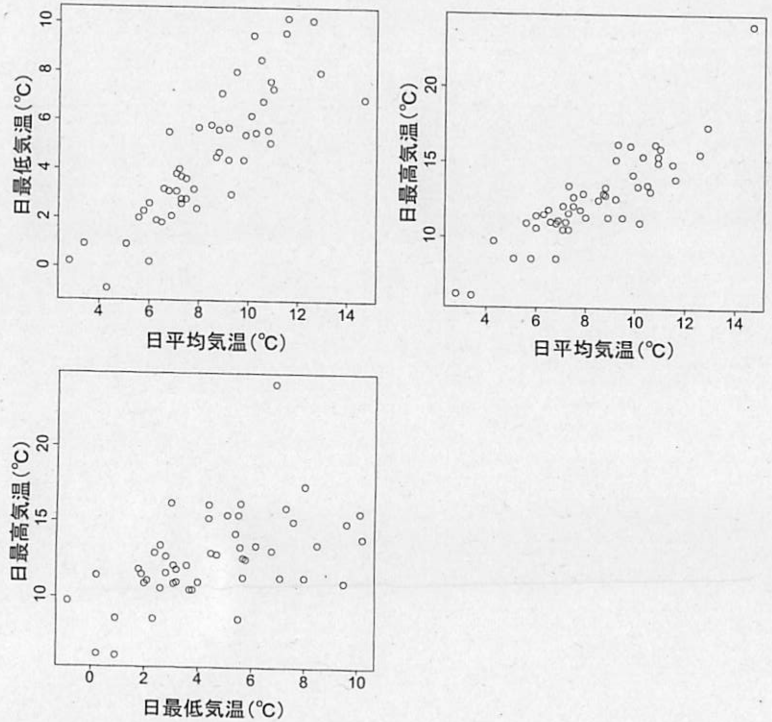
④



⑤



〔3〕2015 年 12 月 1 日から 50 日間の日平均気温，日最高気温，日最低気温についてそれぞれの組合せの散布図を作成した。



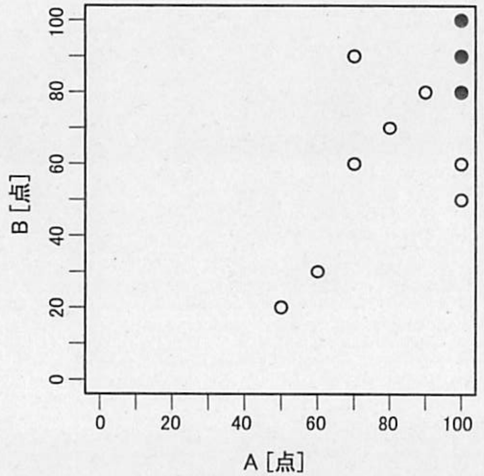
次の記述 I 〜 III はこれらの散布図に関するものである。

I. 日平均気温と日最高気温の間には正の相関がある。
II. 日最低気温は日最高気温より範囲が小さい。
III. 日平均気温と日最低気温の間には負の相関がある。

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。 3

① I のみ正しい。　② II のみ正しい。
③ III のみ正しい。　④ I と II のみ正しい。
⑤ I と III のみ正しい。

問 2　次の図は，あるクラスの 20 人に対して行った 2 つの試験（試験 A と試験 B）の成績を散布図としてプロットしたものである。ただし，成績は 10 点刻みの点数になっており，散布図上の○は 1 人のみ，●は 2 人以上に対応する。また，試験 A の成績の平均は 91 点であり，試験 B の成績の平均は 79.5 点であった。



〔1〕試験 A で 100 点をとった生徒は何人か。次の ① 〜 ⑤ のうちから適切なものを一つ選べ。 4

① 7 人　② 9 人　③ 10 人　④ 12 人　⑤ 14 人

〔2〕試験 A で 100 点をとった生徒に限ったときの試験 B の成績の平均点はいくらか。次の ① 〜 ⑤ のうちから最も適切なものを一つ選べ。 5

① 76.0 点　② 79.5 点　③ 88.6 点　④ 91.4 点　⑤ 100.0 点

問3 100人を対象としたアンケート調査を東京と大阪でそれぞれ行った。回答者は属性AとBのいずれかに分類され，質問に対する回答は「はい」と「いいえ」の二者択一とする。

〔1〕東京でのアンケートで，属性Aに分類された回答者は100人中40人,「はい」と答えた回答者は100人中20人であった。属性ごとに「はい」と答えた割合は等しいとすると，属性Bで「いいえ」と答えた回答者は何人か。次の①～⑤のうちから適切なものを一つ選べ。 **6**

① 8人　② 12人　③ 32人　④ 48人　⑤ 70人

〔2〕大阪でのアンケートの回答者の属性と質問に対する回答を集計したところ，次の2元クロス集計表が得られた。

人数	はい	いいえ
属性A	40	10
属性B	10	40

属性と回答の間の関連の強さを数値で表現するため，属性と回答に数値を割り当てて相関係数を求める。属性Aには0，属性Bには1,「はい」には0,「いいえ」には1を割り当てたときの相関係数をr_1とする。r_1の値はいくらか。次の①～⑤のうちから適切なものを一つ選べ。 **7**

① 0.6　② 0.4　③ 0.36　④ 0.25　⑤ 0.15

〔3〕大阪でのアンケートで，属性Aには0，属性Bには1/2,「はい」には0,「いいえ」には1を割り当てたときの相関係数をr_2とする。また，属性Aには1，属性Bには0,「はい」には0,「いいえ」には1を割り当てたときの相関係数をr_3とする。3つの相関係数r_1，r_2，r_3の関係として，次の①～④のうちから適切なものを一つ選べ。 **8**

① $r_1 = r_2 = -r_3$　② $r_1 = -r_2 = r_3$
③ $r_1 = -r_2 = -r_3$　④ $r_1 = r_2 = r_3$

問4 次の表は，2016年1月29日に公表された日本の消費者物価指数（生鮮食品を除く総合指数。年平均で2010年を100としている）である。

年	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
指数	100.8	100.8	102.3	101.0	100.0	99.8	99.7	100.1	102.7	103.2

資料：総務省「消費者物価指数」

〔1〕上の表にもとづいて，2016年と2017年の物価上昇率（前年比）がそれぞれ1.0％，1.8％であった場合の2017年の消費者物価指数として，次の①～⑤のうちから最も適切なものを一つ選べ。 **9**

① 106.0　② 106.1　③ 106.2　④ 106.3　⑤ 106.4

〔2〕次の記述Ⅰ～Ⅲは表の消費者物価指数に関するものである。

Ⅰ. 2006年と2007年の年平均の消費者物価指数は100.8と同じ値であるので，2007年の月次の消費者物価指数もすべて100.8であることが分かる。

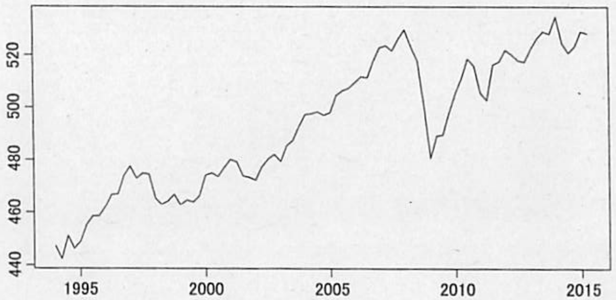
Ⅱ. 消費者物価指数は2006年～2015年にかけて，上昇を続けている。

Ⅲ. 表をもとに2006年の物価指数を100として指数を作成し直した場合でも，各年の物価上昇率は同じである。

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。 **10**

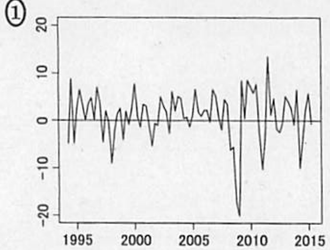
① Ⅰのみ正しい。　② Ⅱのみ正しい。
③ Ⅲのみ正しい。　④ ⅠとⅡのみ正しい。
⑤ ⅠとⅢのみ正しい。

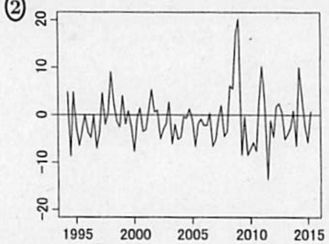
問 5　次の図は，1994 年 1–3 月期から 2015 年 4–6 月期までの 86 四半期分の日本の実質 GDP（兆円，季節調整済み）の系列である。

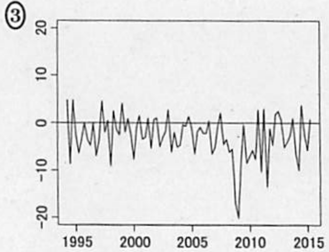


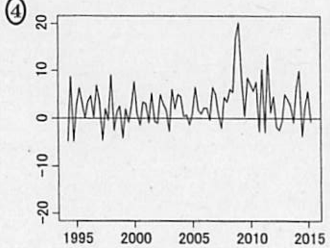
資料：内閣府「国民経済計算」

〔1〕一般に，時系列データ $\cdots, y_{t-1}, y_t, y_{t+1}, \cdots$ に対して一期前との差をとった系列 $\cdots, (y_{t-1} - y_{t-2}), (y_t - y_{t-1}), (y_{t+1} - y_t), \cdots$ を階差系列という。図で示している実質 GDP の階差系列として，次の ① ～ ④ のうちから最も適切なものを一つ選べ。 **11**

①


②


③


④


〔2〕次の表は，被説明変数（従属変数）を実質 GDP y_t，説明変数（独立変数）を時間 t とする回帰モデルを推定した結果である（$t = 1, \ldots, 86$）。表の（ア）の数値として，下の ① ～ ⑤ のうちから最も適切なものを一つ選べ。 **12**

	推定値	標準誤差	t 値	P-値
切片	452.011	2.147	210.56	0.000
時間変数	0.937	（ア）	21.86	0.000

残差の標準誤差	9.867	自由度	84
決定係数 R^2	0.851	自由度修正済み決定係数	0.849
F 値	478.1	自由度（1, 84）	P-値 0.000

① 0.0429　② 0.2071　③ 1.96　④ 21.86　⑤ 23.33

〔3〕2015 年 4–6 月期までのデータによって推定された結果を用いた 2015 年 7–9 月期の実質 GDP の予測値はいくらか。次の ① ～ ⑤ のうちから最も適切なものを一つ選べ。 **13**

① 530.719　② 531.656　③ 532.593　④ 533.530　⑤ 534.593

問 6　標本調査では無作為抽出をはじめ，いくつかの調査方法がある。

〔1〕調査についての説明として，次の ① ～ ⑤ のうちから適切でないものを一つ選べ。 **14**

① 全数調査は標本調査に比べ費用がかかる場合が多い。

② 無作為抽出を行うと誤差の大きさを評価することができない。

③ 標本誤差はどんなに調査員を訓練しても 0 にすることができない。

④ 標本調査は全数調査と比べ速報性の点で優れている。

⑤ 全数調査，標本調査にかかわらず，できるだけ正確な母集団名簿があることが望ましい。

〔2〕ある県の小学生の学習時間の調査を次の方法で実施した。
「最初に県の小学校の名簿から無作為に 100 校を選びだし，その選びだされた小学校に在籍する児童全員について学習時間を調べた。」
この調査で使われた標本抽出法はどれか。次の ① ～ ⑤ のうちから最も適切なものを一つ選べ。 **15**

① 二段抽出法　　② 二相抽出法
③ 単純無作為抽出法　　④ クラスター抽出法（集落抽出法）
⑤ 層別抽出法（層化抽出法）

問 7　2 人で勝負するゲームがある。引き分けはなく，勝つ確率と負ける確率はいずれも 1/2 である。

A，B，C の 3 人が次のルールで対戦し，最初に 2 連勝した人を優勝とする。このルールでは最初に A と B が必ず対戦する。A が勝った場合，A は次に C と対戦し，A が勝てば A の優勝となる。C が勝った場合は C は次に B と対戦し，C が勝てば C の優勝，B が勝った場合は再び A と対戦する。この勝負を 3 人のうちの誰かが 2 連勝するまで繰り返す。最初の対戦で B が勝った場合も同様にする。

このルールの下で A，B，C が優勝する確率をそれぞれ P_A，P_B，P_C と書くことにする。また，ある人が現在の勝負で負けた時点で優勝者がまだ決まらない場合に，その負けた人がその後優勝する条件付き確率を r とする。なお，このゲームでは過去の勝負の結果を条件としていないので，現在の勝負が何回目の勝負であっても条件付き確率 r は同じ値であり，条件付き確率 r は 3 人の誰に対してでも同じ値となる。

〔1〕「A が最初に B に勝ち，次に C に負け，その後 A が優勝する」確率として，次の ① ～ ⑤ のうちから適切なものを一つ選べ。 **16**

① $\frac{1}{8}$　② $\frac{1}{4}+r$　③ $\frac{1}{4}+\frac{1}{2}r$　④ $\frac{1}{2}+\frac{1}{4}r$　⑤ $\frac{1}{4}r$

〔2〕A が優勝するのは，「〔1〕の場合」，「最初から A が 2 連勝する場合」，「最初 B に負け，その後 A が優勝する場合」，の 3 つの場合がある。確率 P_A を表す式として，次の ① ～ ⑤ のうちから適切なものを一つ選べ。 **17**

① $P_A=\frac{1}{4}+\frac{1}{4}r$　② $P_A=\frac{1}{4}+\frac{1}{2}r$
③ $P_A=\frac{1}{4}+\frac{3}{4}r$　④ $P_A=\frac{1}{2}+\frac{3}{4}r$
⑤ $P_A=\frac{3}{4}r$

〔3〕P_A，P_B，P_C の関係を表す式として，次の ① ～ ⑤ のうちから適切なものを一つ選べ。 **18**

① $P_A=P_B=P_C$　② $P_A=P_B>P_C$
③ $P_A=P_B<P_C$　④ $P_A<P_B=P_C$
⑤ $P_A>P_B=P_C$

2016年6月 問題

問 3 B 大学のある学科で国語と英語による入学試験を行った。試験はいずれも 100 点満点である。受験者は 300 人で，得点の平均，分散，中央値は以下のとおりであった。

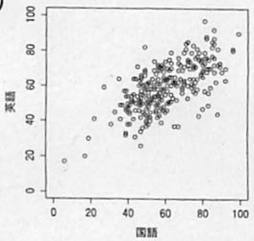
国語：	平均 56.0	分散 236.6	中央値 58.0
英語：	平均 59.1	分散 170.1	中央値 59.0

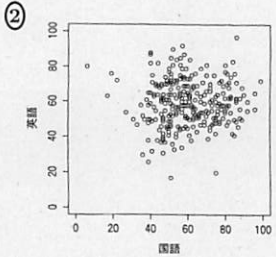
また，国語と英語の共分散は 133.1 であった。

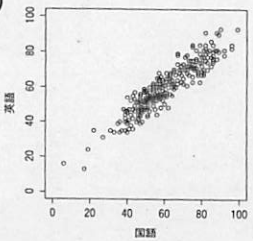
〔1〕国語と英語の相関係数として，次の ①～⑤ のうちから最も近い値を一つ選べ。 $\boxed{7}$

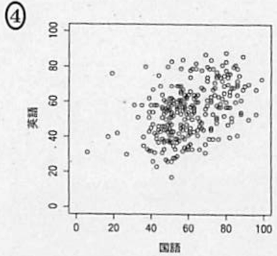
① 0.0　　② 0.3　　③ 0.5　　④ 0.7　　⑤ 0.9

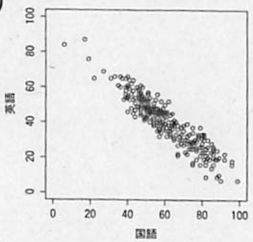
〔2〕国語と英語の散布図として，次の ①～⑤ のうちから最も適切なものを一つ選べ。 $\boxed{8}$

①


②


③


④


⑤


〔3〕得点の合計が 120 点以上の受験者を合格とした。合格者の得点の相関に関する記述として，次の ①～⑤ のうちから最も適切なものを一つ選べ。 $\boxed{9}$

① 合格者は受験者 300 人の一部を取り出したものなので，合格者の得点の相関は受験者全体とほぼ変わらない。

② 合格者はいずれの科目も高い得点を取っているので，合格者は受験者全体よりも強い正の相関となる。

③ 合格者のみの得点の共分散や分散がわからないので，合格者の得点の相関が受験者全体よりも強くなるか弱くなるか，あるいはほぼ変わらないかは見当がつかない。

④ 国語と英語のどちらか片方だけがよい得点の受験者が多かったので，合格者の得点の相関は強い負の相関となる。

⑤ 国語と英語のどちらかの得点が高く，他方は低い合格者もいるので，合格者は受験者全体よりも相関が弱くなる。

問 4 ある飲料会社がこれから売り出す 2 つのドリンク J と K について，ランダムに選ばれた男性 240 人，女性 360 人に両方とも試飲してもらい，どちらが好きかを聞いたところ，

ドリンク J が好きな人：300 人　　ドリンク K が好きな人：300 人

という結果であった。

このとき，男性の中でドリンク J を好む人数が何人である場合に，性別とドリンクの好みの関連が最も弱いといえるか。次の ①～⑤ のうちから最も適切なものを一つ選べ。 $\boxed{10}$

① 0 人　　② 80 人　　③ 100 人　　④ 120 人　　⑤ 240 人

問 14 次の文章は，母比率 p の信頼区間について述べたものである。

標本の大きさを n，標本比率を $\hat{p}$ とする。$\hat{p}$ は確率変数であり，n が十分大きいとき平均 p，標準偏差 $\sqrt{p(1-p)/n}$ の正規分布にほぼ従う。したがって，$\hat{p}$ を標準化した確率変数 $Z=$（ア）は標準正規分布にほぼ従うので，$-1.96 \le$（ア）≤ 1.96 が 95 %の確率で成り立つ。これを変形すると $\hat{p}-$（イ）$\le p \le \hat{p}+$（イ）となり，この区間が p を含む確率は 95%であることがわかる。この（イ）には未知の値 p が含まれるため，p の代わりに標本比率 $\hat{p}$ を用いることで p の近似的な信頼区間が得られる。

一方，標本比率を用いなくても，信頼区間のおおよその幅を見積もることができる。$p(1-p)$ が最大となるのは $p=$（ウ）のときであり，その最大値は（エ）である。そして，1.96 をほぼ 2 とみなすことにより，（イ）の上限はほぼ（オ）となることがわかる。したがって，信頼区間の幅が $2\times$（オ）以下であることがわかる。

〔1〕文中の（ア），（イ）にあてはまるものとして，次の ①～⑤ のうちから最も適切なものを一つ選べ。 **25**

① （ア）$\dfrac{\hat{p}-p}{\sqrt{p(1-p)}}$　（イ）$1.96\sqrt{np(1-p)}$

② （ア）$\dfrac{\hat{p}-p}{\sqrt{p(1-p)/n}}$　（イ）$1.96\sqrt{p(1-p)}$

③ （ア）$\dfrac{\hat{p}-p}{\sqrt{p(1-p)/n}}$　（イ）$1.96\sqrt{p(1-p)/n}$

④ （ア）$\dfrac{\hat{p}-p}{\sqrt{np(1-p)}}$　（イ）$1.96\sqrt{np(1-p)}$

⑤ （ア）$\dfrac{\hat{p}-p}{\sqrt{np(1-p)}}$　（イ）$1.96\sqrt{p(1-p)/n}$

〔2〕文中の（ウ）～（オ）にあてはまるものとして，次の ①～⑤ のうちから最も適切なものを一つ選べ。 **26**

① （ウ）1　（エ）0　（オ）$\sqrt{n}$

② （ウ）1　（エ）$\dfrac{1}{4}$　（オ）$\dfrac{1}{\sqrt{n}}$

③ （ウ）0.5　（エ）$\dfrac{1}{4}$　（オ）$\sqrt{n}$

④ （ウ）0.5　（エ）$\dfrac{1}{4}$　（オ）$\dfrac{1}{\sqrt{n}}$

⑤ （ウ）0　（エ）0　（オ）1

問 15 ある刺激を与えたときの血圧（収縮期血圧）の変化を調べるために，10 人の被験者に対して，刺激を与える前の血圧（mmHg）と刺激を与えた後の血圧（mmHg）を測定した。

No.	1	2	3	4	5	6	7	8	9	10	平均
刺激前	130	118	128	135	126	120	126	140	127	130	128.0
刺激後	135	120	132	135	129	128	135	139	135	132	132.0

〔1〕この刺激を与えた後に血圧が上がる変化があるかを有意水準 5 %で片側検定したい。用いる t 分布の自由度と棄却域について，次の ①～⑤ のうちから最も適切なものを一つ選べ。 **27**

① 自由度 9　棄却域 $t \ge 1.833$

② 自由度 9　棄却域 $|t| \ge 2.262$

③ 自由度 10　棄却域 $|t| \ge 2.228$

④ 自由度 18　棄却域 $t \ge 1.734$

⑤ 自由度 18　棄却域 $|t| \ge 2.101$

〔2〕t 統計量の値と検定の結果について，次の ①～⑤ のうちから最も適切なものを一つ選べ。 **28**

① t 統計量の値は 1.51 であり，帰無仮説は棄却されず，この刺激を与えた後に血圧が上がる変化があるとはいえない，と判断する。

② t 統計量の値は 1.51 であり，帰無仮説は棄却されず，この刺激を与えた後に血圧が上がる変化がないとはいえない，と判断する。

③ t 統計量の値は 3.65 であり，帰無仮説を棄却し，この刺激を与えた後に血圧が上がる変化がないとはいえない，と判断する。

④ t 統計量の値は 3.65 であり，帰無仮説を棄却し，この刺激を与えた後に血圧が上がる変化があるとはいえない，と判断する。

⑤ t 統計量の値は 3.65 であり，帰無仮説を棄却し，この刺激を与えた後に血圧が上がる変化がある，と判断する。

問 18　四国電力が公開している毎時間の電力需要から一日の電力需要の合計値（万 kWh）を求め，2014 年 7 月 1 日から 8 月 31 日までの夏季の電力需要について，気象と関係があるかを調べた。

気象に関する変数としては，高松の一日の平均気温 (℃)，最高気温 (℃)，最低気温 (℃)，降水量の合計 (mm)，日照時間 (時間)，合計全天日射量 (MJ/m^2)，平均蒸気圧 (hPa)，平均雲量 (10 分比) を用いた。

〔1〕まず，電力需要と各変数の相関係数を計算したところ，平均気温との相関が 0.72 で最も強かった。そこで，統計ソフトウェアを用いて平均気温で電力需要を予測する回帰直線を求めたところ，次のような結果が得られた。

モデル 1

```
Residuals:
    Min      1Q  Median      3Q     Max
-1116.1  -522.9   272.5   380.5   901.6

Coefficients:
           Estimate  Std. Error  t value  Pr(>|t|)
切片          467.8       987.3    0.474     0.637
平均気温      289.3        36.1    8.014  4.65e-11

Residual standard error: 571.5 on 60 degrees of freedom
Multiple R-squared:  0.517,  Adjusted R-squared:  0.5089
F-statistic: 64.22 on 1 and 60 DF,  p-value: 4.647e-11
```

この単回帰分析により，平均気温が 1 ℃高くなると，電力需要は何万 kWh 増加すると予測されるか。次の ① ～ ⑤ のうちから最も適切なものを一つ選べ。

32

① 272.5 万 kWh　② 467.8 万 kWh　③ 289.3 万 kWh
④ 987.3 万 kWh　⑤ 36.1 万 kWh

〔2〕次に，説明変数としてすべての変数を用いて重回帰分析を行ったところ，次のような結果が得られた。

モデル 2

```
Residuals:
    Min      1Q  Median      3Q     Max
-1282.2  -390.5   187.2   361.6  1056.1

Coefficients:
             Estimate  Std. Error  t value  Pr(>|t|)
切片          257.089    1667.505    0.154    0.8781
平均気温      602.682     229.375    2.627    0.0112
最高気温      -67.456     117.482   -0.574    0.5683
最低気温     -310.329     151.631   -2.047    0.0457
降水量         -4.504       5.468   -0.824    0.4138
日照時間       51.025      83.871    0.608    0.5455
全天日射量    -44.862      53.694   -0.836    0.4072
平均蒸気圧     51.987      43.386    1.198    0.2362
平均雲量       43.094      74.467    0.579    0.5652

Residual standard error: 577.5 on 53 degrees of freedom
Multiple R-squared:  0.5643,  Adjusted R-squared:  0.4986
F-statistic: 8.582 on 8 and 53 DF,  p-value: 2.065e-07
```

また，影響が小さい変数をモデル 2 から順次取り除いていく変数減少法による変数選択を行ったところ，次のような結果が得られた。

モデル 3

```
Residuals:
    Min      1Q  Median      3Q     Max
-1200.5  -453.6   187.5   378.0  1013.0

Coefficients:
            Estimate  Std. Error  t value  Pr(>|t|)
切片          655.41     1039.12    0.631    0.5307
平均気温      406.19       80.01    5.077  4.25e-06
最低気温     -212.14      116.16   -1.826    0.0730
平均蒸気圧     65.15       35.43    1.839    0.0711

Residual standard error: 562.4 on 58 degrees of freedom
Multiple R-squared:  0.5478,  Adjusted R-squared:  0.5245
F-statistic: 23.42 on 3 and 58 DF,  p-value: 4.653e-10
```

2015年11月 問題