数学２（問題）

［問題１から問題４を通じて、必要であれば（付表）に記載された数値を用いよ。］

問題１．次の各問の $\boxed{}$ に入る答のみを、所定の解答用紙に記入せよ。（３０点）

(1) ある会社では純度がより高い化学製品の購入を望んでいる。そのため、納入されたロットから抽出した10個の製品の純度の平均値がある基準値以下であれば、そのロットは返品することとした。ただし、より多くの製品を購入するために基準値はできるだけ小さくしたいと考えている。今、各ロットは多数の製品で構成され、１つのロットを構成する各製品の純度は正規分布に従うものとする。また、各ロットにおける製品の純度の分散は $\sigma^2 = (0.016)^2$ であることがわかっているものとする。このとき、製品の純度の平均が0.96以下であるロットを返品しない確率を5%以下とするためには基準値を $\boxed{}$ とすればよい。（小数点以下第４位四捨五入）

(2) 正規母集団 $N(\mu, 4^2)$ から大きさ 12 の標本を抽出し、帰無仮説 $H_0 : \mu = 8$ を対立仮説 $H_1 : \mu = 10$ に対し、有意水準 5%で検定したい。統計量 $\bar{X} = \dfrac{X_1 + X_2 + \cdots + X_{12}}{12}$ とするとき検出力が最大になるような棄却域は $\{\bar{x}\,;\ \boxed{}\}$ となる。（小数点以下第２位四捨五入）

(3) ある農作物に関し、A農園の農作物４個とB農園の農作物６個について重さの測定をしたところ、次の結果を得た。

A農園：125.1,　120.9,　107.4,　113.4　（グラム）

B農園：136.8,　122.1,　130.8,　121.5,　129.6,　118.8　（グラム）

A農園の農作物とB農園の農作物の重さはそれぞれ正規分布に従うものとするとき、A農園およびB農園の農作物の重さの母分散比（B農園の母分散／A農園の母分散）の95%信頼区間は（ $\boxed{}$ ， $\boxed{}$ ）である。（小数点以下第３位四捨五入）

(4) ある種の電球の寿命は指数分布に従っているとする。

その電球８個の寿命を調べたところ、次のようなデータを得た。

3200,　3500,　2300,　4800,　3900,　4300,　2800,　3600　（時間）

上のデータを基にして、「電球の使用時間が3600時間を経過した時点で新しい電球と交換する」という方針を立てた場合の信頼度（交換前に電球が切れない確率）は $\boxed{}$ である。なお、電球の寿命の分布の母数の推定には最尤法を用いよ。

(5) 正規母集団$N\left(\mu,\sigma^2\right)$の母平均$\mu$の 90%信頼区間の長さが$\sigma$より小さくなる確率が 95%以上であるための最小の標本数は $\boxed{}$ である。ただし、σ^2は未知とする。

(6) ある計算機には、区間$(0,\ell)$（ℓは十分に大きい）上で発生させた一様乱数Kを入力すると、ある未知の定数$k\ (k>0)$に対して、$K=q\cdot k+x\ \left(0\le x<k,\ \ q;\ \text{整数}\right)$ を満たすxを出力するような演算処理機能がある。

この計算機に 10 個の一様乱数を入力し、上の演算処理を行なったところ、xの値$x_1,x_2,\cdots,x_{10}$を次のとおり得た：

1.2,　1.8,　2.9,　3.6,　2.5,　2.3,　2.1,　1.9,　1.5,　3.2

この 10 個の値からkの値を推定するために、その推定量として、次のT_1,T_2を想定した：

$$T_1=\alpha\cdot\left(\frac{X_1+X_2+\cdots+X_{10}}{10}\right),\quad T_2=\beta\cdot max\left(X_1,X_2,\cdots,X_{10}\right)$$

このとき、T_1,T_2がkの不偏推定量となるように定数α,βを定め、T_1,T_2のうちより有効な方を用いると、kは $\boxed{}$ と推定される。(小数点以下第２位四捨五入)

問題２．次の文中の①から⑩の空欄に当てはまる最も適当な数値、語句あるいは算式を所定の解答用紙に記入せよ。（２０点）

変数xと変量Yの対(x,Y)があって、そのn対の測定値$(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$から$x$と$Y$の間に何らかの線形の関係があると考えた。具体的には１次の直線関係$Y_i=a+bx_i+\varepsilon_i\ (i=1,2,\cdots,n)$という構造模型を想定した。ここで、$\varepsilon_i\ (i=1,2,\cdots,n)$は互いに独立な確率変数で、すべて期待値０、分散$\sigma^2$の正規分布に従うとし、$a,b,\sigma^2$は未知とする。

このとき、以下の手順でa,b,σ^2の最尤推定量を求めることとする。

Y_iの分布関数$F_{Y_i}(y_i)$はε_iの分布関数$F_{\varepsilon_i}(\cdot)$を使って、次のように表すことができる：

$F_{Y_i}(y_i)=P(Y_i\le y_i)=F_{\varepsilon_i}(\boxed{\quad ① \quad})$

よって、Y_iの確率密度関数$f_{Y_i}(y_i)$は$f_{Y_i}(y_i)=f_{\varepsilon_i}(\boxed{\quad ① \quad})=\boxed{\quad ② \quad}$となり（$f_{\varepsilon_i}(\cdot)$は$\varepsilon_i$の確率密度関数）、これより、$(Y_1,Y_2,\cdots,Y_n)$の同時密度関数は$f(y_1,y_2,\cdots,y_n)=\boxed{\quad ③ \quad}\cdot e^{-W}$と得られる。ここに、$W=\boxed{\quad ④ \quad}$である。

そこで、$f(y_1,y_2,\cdots,y_n)$をa,b,σ^2の$\boxed{\quad ⑤ \quad}$関数と考えて、これを$\boxed{\quad ⑥ \quad}$とする推定値a,b,σ^2を求めればよい。

そのために、まずσ^2を固定して考えて、$f(y_1,y_2,\cdots,y_n)$が$\boxed{\quad ⑥ \quad}$となる、すなわち、Wが$\boxed{\quad ⑦ \quad}$となるa,bを求める。次に、$f(y_1,y_2,\cdots,y_n)$を$\boxed{\quad ⑥ \quad}$とするσ^2を求めればよい。

以上よりa,b,σ^2の最尤推定量は、$\hat{a}=\boxed{\quad ⑧ \quad}$，$\hat{b}=\boxed{\quad ⑨ \quad}$，

$\hat{\sigma}^2=\boxed{\quad ⑩ \quad}$と得られる。

なお、解答中には、記号$\bar{x}\left(=\frac{1}{n}\sum_{i=1}^{n}x_i\right)$, $\bar{y}\left(=\frac{1}{n}\sum_{i=1}^{n}y_i\right)$, $\bar{Y}\left(=\frac{1}{n}\sum_{i=1}^{n}Y_i\right)$を必要に応じて用いてよい。また、⑩の解答中には、記号$\hat{a},\hat{b}$を用いてよい。

問題３．ある会社で、過去75日間における１日の苦情受付件数を集計したところ、次の結果を得た。

受付件数	0	1	2	3	4	5	6	7	合計
日数	8	15	21	14	8	6	2	1	75

これについて以下の問に答えよ。（２５点）

(1)　１日の苦情受付件数はポアソン分布に従っているとした場合、母平均λの最尤推定値を求めよ。（結果のみ答えよ。計算過程は不要。）

(2)　１日の苦情受付件数は(1)で求めた最尤推定値を母平均λに持つポアソン分布に従っているといえるか、有意水準5%で検定せよ。

(3)　その後、この会社では苦情対策を実施し、その効果を見るため直近10日間の毎日の苦情受付件数を調べたところ、次のとおりであった。

0,　2,　1,　0,　3,　0,　1,　0,　1,　1（件）

このとき、過去の実績に対し、1日の苦情受付件数は減少しているといえるか、有意水準5%で検定せよ。なお、(2)の結果にかかわらず、１日の苦情受付件数はポアソン分布に従っているものとする。

問題４．母平均がμ、母分散がσ^2の有限母集団π（要素の数はN、各要素の特性値を$\alpha_1, \alpha_2, \cdots, \alpha_N$とする）について、以下の問に答えよ。（２５点）

(1)　πから非復元抽出した、大きさnの標本$X_1, X_2, \cdots, X_n$の１次式

$$T = k_0 + k_1 X_1 + k_2 X_2 + \cdots + k_n X_n$$

が、μ（未知）の値にかかわらず、常に母平均μの不偏推定量であるための$k_i\ (i = 0, 1, \cdots, n)$の条件を求めよ。

(2)　(1)で定義した大きさnの標本$X_1, X_2, \cdots, X_n$に関し、標本共分散$Cov(X_i, X_j)$　$(i \neq j)$を求めよ。

(3)　(1)で求めた条件のもと、Tの分散$V(T)$が最小となるように$k_i\ (i = 0, 1, \cdots, n)$を求めよ。

以　上

（付表）

Ｉ．標準正規分布の上側ε点：$u(\varepsilon)$

ε	0.159	0.050	0.025	0.023	0.010	0.005
$u(\varepsilon)$	1.000	1.645	1.960	2.000	2.326	2.576

Ⅱ．自由度ϕのt分布の上側ε点：$t_\phi(\varepsilon)$

$\phi \diagdown \varepsilon$	0.100	0.050	0.025
15	1.341	1.753	2.131
16	1.337	1.746	2.120
17	1.333	1.740	2.110
18	1.330	1.734	2.101
19	1.328	1.729	2.093
20	1.325	1.725	2.086
21	1.323	1.721	2.080
22	1.321	1.717	2.074
23	1.319	1.714	2.069
24	1.318	1.711	2.064
25	1.316	1.708	2.060

Ⅲ．自由度ϕのχ^2分布の上側ε点：$\chi^2_\phi(\varepsilon)$

$\phi \diagdown \varepsilon$	0.100	0.050	0.025
1	2.706	3.841	5.024
2	4.605	5.991	7.378
3	6.251	7.815	9.348
4	7.779	9.488	11.143
5	9.236	11.070	12.832
6	10.645	12.592	14.449
7	12.017	14.067	16.013
8	13.362	15.507	17.535
9	14.684	16.919	19.023
10	15.987	18.307	20.483

$\phi \diagdown \varepsilon$	0.100	0.050	0.025
15	22.307	24.996	27.488
16	23.542	26.296	28.845
17	24.769	27.587	30.191
18	25.989	28.869	31.526
19	27.204	30.144	32.852
20	28.412	31.410	34.170
21	29.615	32.671	35.479
22	30.813	33.924	36.781
23	32.007	35.172	38.076
24	33.196	36.415	39.364
25	34.382	37.652	40.646

Ⅳ．分母の自由度m、分子の自由度nのF分布の上側ε点：$F_m^n(\varepsilon)$

$\varepsilon = 0.100$

$m \diagdown n$	1	2	3	4	5	6	7	8	9	10
2	8.526	9.000	9.162	9.243	9.293	9.326	9.349	9.367	9.381	9.392
3	5.538	5.462	5.391	5.343	5.309	5.285	5.266	5.252	5.240	5.230
4	4.545	4.325	4.191	4.107	4.051	4.010	3.979	3.955	3.936	3.920
5	4.060	3.780	3.619	3.520	3.453	3.405	3.368	3.339	3.316	3.297
6	3.776	3.463	3.289	3.181	3.108	3.055	3.014	2.983	2.958	2.937
7	3.589	3.257	3.074	2.961	2.883	2.827	2.785	2.752	2.725	2.703
8	3.458	3.113	2.924	2.806	2.726	2.668	2.624	2.589	2.561	2.538
9	3.360	3.006	2.813	2.693	2.611	2.551	2.505	2.469	2.440	2.416
10	3.285	2.924	2.728	2.605	2.522	2.461	2.414	2.377	2.347	2.323

$\varepsilon = 0.050$

$m \diagdown n$	1	2	3	4	5	6	7	8	9	10
2	18.513	19.000	19.164	19.247	19.296	19.329	19.353	19.371	19.385	19.396
3	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.785
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637
8	5.318	4.459	4.066	3.838	3.688	3.581	3.500	3.438	3.388	3.347
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978

$\varepsilon = 0.025$

$m \diagdown n$	1	2	3	4	5	6	7	8	9	10
2	38.506	39.000	39.166	39.248	39.298	39.331	39.356	39.373	39.387	39.398
3	17.443	16.044	15.439	15.101	14.885	14.735	14.624	14.540	14.473	14.419
4	12.218	10.649	9.979	9.604	9.364	9.197	9.074	8.980	8.905	8.844
5	10.007	8.434	7.764	7.388	7.146	6.978	6.853	6.757	6.681	6.619
6	8.813	7.260	6.599	6.227	5.988	5.820	5.695	5.600	5.523	5.461
7	8.073	6.542	5.890	5.523	5.285	5.119	4.995	4.899	4.823	4.761
8	7.571	6.059	5.416	5.053	4.817	4.652	4.529	4.433	4.357	4.295
9	7.209	5.715	5.078	4.718	4.484	4.320	4.197	4.102	4.026	3.964
10	6.937	5.456	4.826	4.468	4.236	4.072	3.950	3.855	3.779	3.717

V．平均値λのポアソン分布表：$P(X=k)=e^{-\lambda}\cdot\dfrac{\lambda^k}{k!}$

k＼λ	2.0	2.1	2.2	2.3	2.4	2.5
0	0.13534	0.12246	0.11080	0.10026	0.09072	0.08208
1	0.27067	0.25716	0.24377	0.23060	0.21772	0.20521
2	0.27067	0.27002	0.26814	0.26518	0.26127	0.25652
3	0.18045	0.18901	0.19664	0.20331	0.20901	0.21376
4	0.09022	0.09923	0.10815	0.11690	0.12541	0.13360
5	0.03609	0.04168	0.04759	0.05378	0.06020	0.06680
6	0.01203	0.01459	0.01745	0.02061	0.02408	0.02783
7	0.00344	0.00438	0.00548	0.00677	0.00826	0.00994
8	0.00086	0.00115	0.00151	0.00195	0.00248	0.00311
9	0.00019	0.00027	0.00037	0.00050	0.00066	0.00086
10	0.00004	0.00006	0.00008	0.00011	0.00016	0.00022

k＼λ	2.6	2.7	2.8	2.9	3.0
0	0.07427	0.06721	0.06081	0.05502	0.04979
1	0.19311	0.18145	0.17027	0.15957	0.14936
2	0.25104	0.24496	0.23838	0.23137	0.22404
3	0.21757	0.22047	0.22248	0.22366	0.22404
4	0.14142	0.14882	0.15574	0.16215	0.16803
5	0.07354	0.08036	0.08721	0.09405	0.10082
6	0.03187	0.03616	0.04070	0.04546	0.05041
7	0.01184	0.01395	0.01628	0.01883	0.02160
8	0.00385	0.00471	0.00570	0.00683	0.00810
9	0.00111	0.00141	0.00177	0.00220	0.00270
10	0.00029	0.00038	0.00050	0.00064	0.00081

数学２ 解答

問題１． *数理統計学のさまざまな分野の基本的事項に関する理解を問うために出題した。計算量が多かったかもしれないが、全体に解法の道筋はつけやすい問題ばかりであるので、計算ミスなどをせずに、できる問題を着実に解答することが望まれる。*

番号	解答
(1)	0.968
(2)	$\bar{x} > 9.90$
(3)	(0.05, 5.97)
(4)	$e^{-\frac{72}{71}}$
(5)	20
(6)	4.0

（１）純度の母平均が$\mu\,(\mu \le 0.96)$のロット（以下、πと呼ぶ）から抽出した製品 10 個の純度の測定値を$X_1, X_2, \cdots, X_{10}$とする。このとき、πの分布は$N(\mu, \sigma^2)$（σは既知）なので、その平均$\overline{X}$は$N\left(\mu, \dfrac{\sigma^2}{n}\right)$に従う。

問題では基準値（以下、kとする）以下であれば返品することから、πを返品しない確率が 5%以下になるとは、$P(\overline{X} \ge k) \le 0.05$であることを意味する。ここで$P(\overline{X} \ge k) = P\left(\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \ge \dfrac{k - \mu}{\sigma/\sqrt{n}}\right)$であり、$\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$は$N(0,1)$に従うから、$\dfrac{k - \mu}{\sigma/\sqrt{n}} \ge u(0.05) = 1.645$を満たすように基準値$k$を定めればよい。

さて、$\mu \le 0.96$となるμについて上の不等式を満たすようにkを定めるには、$\mu = 0.96$について上の不等式を満たすようにkを定めればよい。

よって、$\dfrac{k-\mu}{\sigma/\sqrt{n}}=1.645$ より、$k=1.645\cdot\dfrac{0.016}{\sqrt{10}}+0.96=0.9683$

したがって、求める基準値（純度の平均値が 0.96 以下のロットを返品する確率を 5% とし、かつ、できるだけ小さい基準値）は、0.968 である。

（2） 帰無仮説 $\mathrm{H}_0:\mu=8$、対立仮説 $\mathrm{H}_1:\mu=10$ において、8<10 であるから、検出力を最大にするためには右側検定を行うことになる。

したがって、棄却域を (C,∞)（C は定数）とおくと、有意水準 5% であることより、$P(\overline{X}>C|\mu=8)=0.05$ となる。ここで、$P(\overline{X}>C|\mu=8)=P\left(\dfrac{\overline{X}-8}{4/\sqrt{12}}>\dfrac{C-8}{4/\sqrt{12}}\middle|\mu=8\right)$ であり、

また $\overline{X}$ は $N\left(\mu,\dfrac{4^2}{12}\right)$ に従うので、H_0 が真のとき $\dfrac{\overline{X}-8}{4/\sqrt{12}}$ は $N(0,1)$ に従う。

よって、$P\left(\dfrac{\overline{X}-8}{4/\sqrt{12}}>\dfrac{C-8}{4/\sqrt{12}}\middle|\mu=8\right)=0.05$ を満たすためには、$\dfrac{C-8}{4/\sqrt{12}}=u(0.05)=1.645$ でなければならない。これを解いて、$C=1.645\times\dfrac{4}{\sqrt{12}}+8=9.899\cdots$。

以上より求める棄却域は $\{\bar{x};\bar{x}>9.90\}$ となる。

（3） $S_A{}^2=\dfrac{1}{n_A-1}\sum_{i=1}^{n_A}(X_i-\overline{X})^2, S_B{}^2=\dfrac{1}{n_B-1}\sum_{j=1}^{n_B}(Y_j-\overline{Y})^2$ とすると、$F=\dfrac{S_A{}^2/\sigma_A{}^2}{S_B{}^2/\sigma_B{}^2}$ が分母の自由度 n_B-1、分子の自由度 n_A-1 の F 分布に従うことより、

$P\left\{F_{n_B-1}^{n_A-1}\left(1-\dfrac{\varepsilon}{2}\right)\le\dfrac{S_A{}^2/\sigma_A{}^2}{S_B{}^2/\sigma_B{}^2}\le F_{n_B-1}^{n_A-1}\left(\dfrac{\varepsilon}{2}\right)\right\}=1-\varepsilon$ となるから、

$P\left\{\dfrac{S_B{}^2}{S_A{}^2}\times F_{n_B-1}^{n_A-1}\left(1-\dfrac{\varepsilon}{2}\right)\le\dfrac{\sigma_B{}^2}{\sigma_A{}^2}\le\dfrac{S_B{}^2}{S_A{}^2}\times F_{n_B-1}^{n_A-1}\left(\dfrac{\varepsilon}{2}\right)\right\}=1-\varepsilon$ となる。

したがって、題意の母分散比の信頼区間は、$\left(\dfrac{S_B{}^2}{S_A{}^2}\times F_{n_B-1}^{n_A-1}\left(1-\dfrac{\varepsilon}{2}\right),\dfrac{S_B{}^2}{S_A{}^2}\times F_{n_B-1}^{n_A-1}\left(\dfrac{\varepsilon}{2}\right)\right)$ で与えられる。

これに数値を代入すると、

$n_A=4, n_B=6$

$\overline{X}_A=\dfrac{1}{4}(125.1+120.9+107.4+113.4)=116.7$

$$\overline{X}_B=\frac{1}{6}(136.8+122.1+130.8+121.5+129.6+118.8)=126.6$$

$$(n_A-1)S_A{}^2=(125.1-116.7)^2+(120.9-116.7)^2+(107.4-116.7)^2+(113.4-116.7)^2=185.58$$

$$(n_B-1)S_B{}^2=(136.8-126.6)^2+(122.1-126.6)^2+(130.8-126.6)^2+(121.5-126.6)^2$$
$$+(129.6-126.6)^2+(118.8-126.6)^2=237.78$$

$$F_{n_B-1}^{n_A-1}\left(\frac{\varepsilon}{2}\right)=F_5^3(0.025)=7.764$$

$$F_{n_B-1}^{n_A-1}\left(1-\frac{\varepsilon}{2}\right)=F_5^3(1-0.025)=\frac{1}{F_3^5(0.025)}=\frac{1}{14.885}$$

$$\frac{S_B{}^2}{S_A{}^2}\times F_{n_B-1}^{n_A-1}\left(1-\frac{\varepsilon}{2}\right)=\frac{237.78/5}{185.58/3}\times\frac{1}{14.885}=0.0516\cdots$$

$$\frac{S_B{}^2}{S_A{}^2}\times F_{n_B-1}^{n_A-1}\left(\frac{\varepsilon}{2}\right)=\frac{237.78/5}{185.58/3}\times 7.764=5.9687\cdots$$

であるから、求める信頼区間は（0.05, 5.97）である。

（4） 寿命を表す確率変数をXとしたとき、その確率密度関数は$f(x;\mu)=\begin{cases}\frac{1}{\mu}e^{-\frac{1}{\mu}x} & (x>0)\\ 0 & (x\leq 0)\end{cases}$

（ただし、$\mu(>0)$は定数）で表される。

まず、n個の電球の寿命を測定して$x_1,x_2,\cdots,x_n$が得られたとするとき最尤法によりμの推定を行う。標本変量$(X_1,X_2,\cdots,X_n)$の確率密度関数は

$\prod_{i=1}^{n}f(x_i;\mu)=\left(\frac{1}{\mu}\right)^n e^{-\sum_{i=1}^{n}\frac{x_i}{\mu}}\ (x_i>0)$となる。これを尤度関数と考え、$L(\mu)$と置けば、最尤推定値は$\frac{\partial}{\partial\mu}\log L(\mu)=\frac{\partial}{\partial\mu}\left(-n\log\mu-\sum_{i=1}^{n}\frac{x_i}{\mu}\right)=0$を$\mu$に関して解いたものである。

よって、$\frac{\partial}{\partial\mu}\left(-n\log\mu-\sum_{i=1}^{n}\frac{x_i}{\mu}\right)=-\frac{n}{\mu}+\frac{1}{\mu^2}\sum_{i=1}^{n}x_i=0$より、最尤推定値$\hat{\mu}=\bar{x}$となる。

今、$\bar{x}=\frac{1}{8}(3200+3500+2300+4800+3900+4300+2800+3600)=3550$であるから、

$\hat{\mu}=3550$であるので、求める信頼度は、

$P(X\geq 3600)=\int_{3600}^{\infty}\frac{1}{3550}e^{-\frac{1}{3550}x}dx=\left[-e^{-\frac{x}{3550}}\right]_{3600}^{\infty}=e^{-\frac{3600}{3550}}=e^{-\frac{72}{71}}$となる。

（５）n個の標本値に基づき作成される最短の 90％の信頼区間を求める。

$$\frac{\bar{X}-\mu}{\sqrt{\left(\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2 \Big/ n-1\right)\cdot\frac{1}{n}}}=\frac{\sqrt{n(n-1)}\left(\bar{X}-\mu\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}}$$ は自由度$n-1$のt分布に従うことから、

$$P\left(-t_{n-1}(0.05)\le\frac{\sqrt{n(n-1)}\left(\bar{X}-\mu\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}}\le t_{n-1}(0.05)\right)=0.9$$ が成り立つ。

よって、$\bar{X}-\frac{\sqrt{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}}{\sqrt{n(n-1)}}\cdot t_{n-1}(0.05)\le\mu\le\bar{X}+\frac{\sqrt{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}}{\sqrt{n(n-1)}}\cdot t_{n-1}(0.05)$ がμの 90％信頼区間となる。したがって、題意より次の①が成り立つ最小のnを求めればよい。

$$P\left(2t_{n-1}(0.05)\cdot\frac{\sqrt{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}}{\sqrt{n(n-1)}}\le\sigma\right)\ge 0.95 \quad \cdots①$$

①より、$P\left(\frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}{\sigma^2}\le\frac{n(n-1)}{4t_{n-1}{}^2(0.05)}\right)\ge 0.95$。

$\frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}{\sigma^2}$ は自由度$n-1$のχ^2分布に従うことから、自由度$n-1$のχ^2分布の上側 5%点$\left(\chi^2_{n-1}(0.05)\right)$と$\frac{n(n-1)}{4t_{n-1}{}^2(0.05)}$を比較する。

n	$\chi^2_{n-1}(0.05)$		$\frac{n(n-1)}{4t_{n-1}{}^2(0.05)}$
18	27.587	>	25.267⋯
19	28.869	>	28.435⋯
20	30.144	<	31.778⋯

よって、求める標本数は 20 個となる。

（6）題意より、X_i は確率密度関数 $f(X_i)=\begin{cases}\dfrac{1}{k} & (0\le x_i\le k)\\ 0 & (\text{それ以外})\end{cases}$ を持つ一様分布に従う。

$$\left(\begin{aligned}&E(X_i)=\int_0^k \frac{1}{k}\cdot x dx=\left[\frac{x^2}{2k}\right]_0^k=\frac{k}{2},E(X_i^2)=\int_0^k \frac{1}{k}\cdot x^2 dx=\left[\frac{x^3}{3k}\right]_0^k=\frac{k^2}{3},\\&V(X_i)=E(X_i^2)-(E(X_i))^2=\frac{k^2}{3}-\left(\frac{k}{2}\right)^2=\frac{k^2}{12}\end{aligned}\right)$$

ここで、T_1,T_2 が k の不偏推定量であるためには、$E(T_1)=E(T_2)=k$ であることが必要となる。

$E(T_1)=E\left(\alpha\cdot\sum_{i=1}^{10}X_i\Big/10\right)=\dfrac{\alpha}{10}\sum_{i=1}^{10}E(X_i)=\dfrac{\alpha}{10}\sum_{i=1}^{10}\dfrac{k}{2}=\dfrac{\alpha k}{2}$ より $\dfrac{\alpha k}{2}=k$。よって、$\alpha=2$。

また、$max(X_1,X_2,\cdots,X_{10})=X_{(10)}$ と置くと、$E(T_2)=E(\beta\cdot X_{(10)})=\beta\cdot E(X_{(10)})$ より $\beta\cdot E(X_{(10)})=k$。よって、$\beta=\dfrac{k}{E(X_{(10)})}$。

ここで、$X_{(10)}$ の確率密度関数は $f(x_{(10)})=\left(\int_0^{x_{(10)}}\dfrac{1}{k}dx\right)^9\times\dfrac{1}{k}\times\dfrac{10!}{9!\cdot 1!}=\dfrac{10\cdot x_{(10)}{}^9}{k^{10}}$ より、

$$E(X_{(10)})=\int_0^k x_{(10)}\cdot 10\cdot x_{(10)}{}^9\cdot\frac{1}{k^{10}}dx_{(10)}=\int_0^k 10\cdot\left(\frac{x_{(10)}}{k}\right)^{10}dx_{(10)}=\frac{10}{11}\left[\left(\frac{x_{(10)}}{k}\right)^{11}\cdot k\right]_0^k=\frac{10}{11}k\text{。}$$

よって、$\beta=\dfrac{11}{10}$。

次に、分散について T_1,T_2 を比較する。

$$V(T_1)=V\left(\frac{2}{10}\sum_{i=1}^{10}X_i\right)=\frac{1}{25}\sum_{i=1}^{10}V(X_i)=\frac{1}{25}\times 10\times\frac{k^2}{12}=\frac{k^2}{30}$$

$$\begin{aligned}V(T_2)&=V\left(\frac{11}{10}X_{(10)}\right)=\left(\frac{11}{10}\right)^2 V(X_{(10)})=\left(\frac{11}{10}\right)^2\left\{E(X_{(10)}{}^2)-E(X_{(10)})^2\right\}\\&=\left(\frac{11}{10}\right)^2\left[\int_0^k x_{(10)}{}^2\times\frac{10x_{(10)}{}^9}{k^{10}}dx_{(10)}-\left(\frac{10}{11}k\right)^2\right]=\left(\frac{11}{10}\right)^2\left[\frac{10x_{(10)}{}^{12}}{12k^{10}}\right]_0^k-k^2\\&=\left(\frac{11}{10}\right)^2\cdot\frac{10k^2}{12}-k^2=\frac{k^2}{120}\end{aligned}$$

よって、$V(T_1)>V(T_2)$ となり、T_2 の方がより有効である。

$\therefore T_2=\dfrac{11}{10}x_{(10)}=\dfrac{11}{10}\times 3.6=3.96 \quad\Rightarrow\quad 4.0$

問題２．回帰直線の回帰係数および誤差項の分散の最尤推定量を求める誘導式の穴埋問題である。いわゆる「線形モデル」に対する基本的な理解を問うために出題した。なお、使用する記号が、測定値を表す記号か、変量を表す記号か、等に注意して解答する必要がある。

番号	解答
①	$y_i - a - bx_i$
②	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - a - bx_i)^2}{2\sigma^2}}$
③	$\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n$
④	$\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - a - bx_i)^2$
⑤	尤度
⑥	最大
⑦	最小
⑧	$\bar{Y} - \frac{\sum_{i=1}^{n} x_i Y_i - n\bar{x}\bar{Y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}\bar{x}$
⑨	$\frac{\sum_{i=1}^{n} x_i Y_i - n\bar{x}\bar{Y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$
⑩	$\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{a} - \hat{b}x_i)^2$

⑧の $\bar{x}$ の係数および⑨については次の解答も可。

$$\frac{\frac{1}{n}\sum_{i=1}^{n} x_i Y_i - \bar{x}\bar{Y}}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2} \quad \text{または} \quad \frac{\sum_{i=1}^{n}(x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

① $F_{Y_i}(y_i) = P(Y_i \le y_i) = P(a + bx_i + \varepsilon_i \le y_i) = P(\varepsilon_i \le y_i - a - bx_i) = F_{\varepsilon_i}(y_i - a - bx_i)$

②$f_{Y_i}(y_i)=\dfrac{\partial}{\partial y_i}F_{Y_i}(y_i)=\dfrac{\partial}{\partial y_i}F_{\varepsilon_i}(y_i-a-bx_i)=f_{\varepsilon_i}(y_i-a-bx_i)=\dfrac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(y_i-a-bx_i)^2}{2\sigma^2}}$

③$f(y_1,y_2,\cdots,y_n)=\prod_{i=1}^{n}f_{Y_i}(y_i)=\left(\dfrac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-a-bx_i)^2}=\left(\dfrac{1}{\sqrt{2\pi\sigma^2}}\right)^n\cdot e^{-W}$

④$W=\dfrac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-a-bx_i)^2$

⑧, ⑨, ⑩Wを最小とするa,bを求めるには、a,bの関数Wは下に凸であるから、次の連立方程式を解けばよい。

$$\begin{cases}\dfrac{\partial W}{\partial a}=-\dfrac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-a-bx_i)=0\\ \dfrac{\partial W}{\partial b}=-\dfrac{1}{\sigma^2}\sum_{i=1}^{n}x_i(y_i-a-bx_i)=0\end{cases}$$、すなわち、$$\begin{cases}\sum_{i=1}^{n}y_i-na-b\sum_{i=1}^{n}x_i=0\\ \sum_{i=1}^{n}x_iy_i-a\sum_{i=1}^{n}x_i-b\sum_{i=1}^{n}x_i^2=0\end{cases}$$

これを$\bar{x},\bar{y}$を用いて書き直すと、$$\begin{cases}\bar{y}-a-b\bar{x}=0=0 & \cdots① \\ \sum_{i=1}^{n}x_iy_i-na\bar{x}-b\sum_{i=1}^{n}x_i^2=0 & \cdots②\end{cases}$$

①より　$a=\bar{y}-b\bar{x}$　であるから、これを②に代入してbを求めると、

$b=\dfrac{\sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}}{\sum_{i=1}^{n}x_i^2-n\bar{x}^2}$、したがって、$a=\bar{y}-\dfrac{\sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}}{\sum_{i=1}^{n}x_i^2-n\bar{x}^2}\bar{x}$。

次に、$f(y_1,y_2,\cdots,y_n)$を最大とするσ^2を求めるには、次の式を解けばよい。

$$\frac{\partial}{\partial(\sigma^2)}\log f(y_1,y_2,\cdots,y_n)=0$$

ここで、$\log f(y_1,y_2,\cdots,y_n)=-\dfrac{n}{2}\log(2\pi\sigma^2)-\dfrac{\sum_{i=1}^{n}(y_i-a-bx_i)^2}{2\sigma^2}$であるから、

$$\frac{\partial}{\partial(\sigma^2)}\log f(y_1,y_2,\cdots,y_n)=-\frac{n}{2\sigma^2}+\frac{\sum_{i=1}^{n}(y_i-a-bx_i)^2}{2(\sigma^2)^2}=0$$

$$\therefore\sigma^2=\frac{1}{n}\sum_{i=1}^{n}(y_i-a-bx_i)^2$$

以上より、a,b,σ^2の最尤推定量$\hat{a},\hat{b},\hat{\sigma}^2$は、$\hat{a}=\bar{Y}-\dfrac{\sum_{i=1}^{n}x_iY_i-n\bar{x}\bar{Y}}{\sum_{i=1}^{n}x_i^2-n\bar{x}^2}\bar{x}$, $\hat{b}=\dfrac{\sum_{i=1}^{n}x_iY_i-n\bar{x}\bar{Y}}{\sum_{i=1}^{n}x_i^2-n\bar{x}^2}$,

$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{a} - \hat{b}x_i\right)^2$ と得られた。

問題3．*（2）はポアソン分布の適合度の検定、（3）はポアソン分布の平均値の検定である。（2）では理論的期待値を5以上とするためにデータプーリングを行なうことがポイントである。*

（1）2.40 $\left(\bar{x} = \frac{0\times8+1\times15+2\times21+3\times14+4\times8+5\times6+6\times2+7\times1}{75} = 2.40\right)$

（2）題意より、下記の帰無仮説の適合度検定を行う。

帰無仮説 H_0 ：1日の苦情受付件数は平均値 2.4 のポアソン分布に従う

受付件数 i	日数 f_i	ポアソン確率の値 p_i^0	理論的期待値 np_i^0	$f_i - np_i^0$	$\left(f_i - np_i^0\right)^2$	$\left(f_i - np_i^0\right)^2/np_i^0$
0	8	0.09072	6.80400	1.19600	1.43042	0.21023
1	15	0.21772	16.32900	-1.32900	1.76624	0.10817
2	21	0.26127	19.59525	1.40475	1.97332	0.10070
3	14	0.20901	15.67575	-1.67570	2.80814	0.17914
4	8	0.12541	9.40575	-1.40575	1.97613	0.21010
5	6	0.06020	4.51500	1.48500	2.20523	0.48842
6	2	0.02408	1.80600	0.19400	0.03764	0.02084
7	1	0.00826	0.61950	0.38050	0.14478	0.23371
	75					

ここで、$np_i^0 \geq 5$ とするために n =5〜7 をプールする

5〜7	9		6.94050	2.05950	4.24154	0.61113
計	75					1.41947

自由度が 6-1-1=4 の χ^2 分布の上側 5%点は $\chi_4^2(0.05) = 9.488$ であるから、

$\chi^2 = 1.41947 < 9.488 = \chi_4^2(0.05)$ となり、H_0 は採択される。

よって、1日の苦情受付件数は平均値 2.4 のポアソン分布に従っていないとは言えない。

（3）過去の1日あたりの平均苦情受付件数を λ_0 とすると、（1）より λ_0 の最尤推定値は、

2.40 であるから、帰無仮説 $H_0 : \lambda = \lambda_0 (= 2.40)$ 対立仮説 $H_1 : \lambda < \lambda_0 (= 2.40)$ として片側検定を行う。

1 日の苦情受付件数 X_i は平均 λ のポアソン分布に従うので、総苦情受付件数 $S = \sum_{i=1}^{n} X_i$ は平均 $n\lambda$ のポアソン分布に従うことが言える。

よって、$S < k$ ならば、H_0 を棄却すればよい。

ここで、有意水準 5%より、$P(S < k | \lambda = \lambda_0) = \sum_{i=1}^{n} (n\lambda_0)^i \dfrac{e^{-n\lambda_0}}{i!} < 0.05$ であるように k を定める。

次に、ポアソン分布と χ^2 分布の関係から、$\sum_{i=1}^{n} (n\lambda_0)^i \dfrac{e^{-n\lambda_0}}{i!} = P(\chi^2_{2(k+1)} > 2n\lambda_0) < 0.05$

よって、総苦情受付件数 k を求め、$\chi^2_{2(k+1)}(0.05) < 2n\lambda_0$ ならば H_0 を棄却するとすればよいことになる。

題意より、$n = 10, k = 9$ であるから、$31.410 = \chi^2_{20}(0.05) < 2n\lambda_0 = 48.0$

よって、H_0 は棄却され、苦情対策の効果がなかったとは言えない。

【ポアソン分布の正規分布近似を用いた（3）の別解】

$\sum_{i=1}^{n} x_i = 0+2+1+0+3+0+1+0+1+1 = 9 \geq 5$ であるので、ポアソン分布の部分和と正規分布の関係から、統計量 $U = \dfrac{\sum_{i=1}^{n} X_i - n\lambda}{\sqrt{n\lambda}}$ は $n = 10$ のとき近似的に標準正規分布に従う。

よって、H_0 の棄却域は、$\sum_{i=1}^{10} x_i < 10\lambda_0 - \sqrt{10\lambda_0} \times u(0.05) = 15.941$ となり、$\sum_{i=1}^{n} x_i = 9$ より、H_0 は棄却されるので、苦情対策の効果がなかったとは言えない。

問題４．　*有限母集団から非復元抽出した n 個の標本の 1 次式で表される不偏推定量で分散を最小とするものは、その標本の標本平均であることを導く問題である。母集団が有限であることを十分に意識して解答することが重要である。*

（１）T の期待値 $E(T)$ は、$E(T) = E(k_0 + k_1 X_1 + \cdots + k_n X_n) = k_0 + k_1 E(X_1) + \cdots + k_n E(X_n)$。

ここで、$E(X_1) = \cdots = E(X_n) = \mu$ であるから、$E(T) = k_0 + (k_1 + \cdots + k_n)\mu$

μの値にかかわらず、Tが常にμの不偏推定量であるためには、$E(T)=\mu$がμの恒等式となればよい。したがって、求める条件は$k_0+(k_1+\cdots+k_n)\mu=\mu$が$\mu$の恒等式となる条件であり、$k_0=0$かつ$k_1+\cdots+k_n=1$となる。

（２）$Cov(X_i,X_j)=E[(X_i-\mu)(X_j-\mu)]=\sum_{k\neq l}^{N}(\alpha_k-\mu)(\alpha_l-\mu)P(X_i=\alpha_k,X_j=\alpha_l)\ (i\neq j)$

ここで、$P(X_i=\alpha_k,X_j=\alpha_l)=\dfrac{1}{N(N-1)}$なので、

$$Cov(X_i,X_j)=\frac{1}{N(N-1)}\sum_{k\neq l}^{N}(\alpha_k-\mu)(\alpha_l-\mu)=\frac{1}{N(N-1)}\left[\sum_{k=1}^{N}\sum_{l=1}^{N}(\alpha_k-\mu)(\alpha_l-\mu)-\sum_{k=1}^{N}(\alpha_k-\mu)^2\right]$$

今、$\dfrac{1}{N}\sum_{k=1}^{N}\alpha_k=\mu$であるから、$\sum_{k=1}^{N}(\alpha_k-\mu)=0$となる。

したがって、$\sum_{k=1}^{N}\sum_{l=1}^{N}(\alpha_k-\mu)(\alpha_l-\mu)=\sum_{k=1}^{N}(\alpha_k-\mu)\sum_{l=1}^{N}(\alpha_l-\mu)=0$

よって、$Cov(X_i,X_j)=-\dfrac{1}{N-1}\cdot\dfrac{1}{N}\sum_{k=1}^{N}(\alpha_k-\mu)^2=-\dfrac{\sigma^2}{N-1}$

（３）$V(T)=V(k_0+k_1X_1+\cdots+k_nX_n)=\sum_{i=1}^{n}k_i^2V(X_i)+\sum_{i\neq j}^{n}Cov(k_iX_i,k_jX_j)$

（２）より、$V(T)=\sigma^2\sum_{i=1}^{n}k_i^2-\dfrac{\sigma^2}{N-1}\sum_{i\neq j}^{n}k_ik_j$。

ここで（１）より、$\sum_{i\neq j}^{n}k_ik_j=\left(\sum_{i=1}^{n}k_i\right)^2-\sum_{i=1}^{n}k_i^2=1-\sum_{i=1}^{n}k_i^2$。

したがって、$V(T)=\dfrac{\sigma^2}{N-1}\left(N\sum_{i=1}^{n}k_i^2-1\right)$と表せる。よって、$V(T)$が最小となるのは、$\sum_{i=1}^{n}k_i^2$が最小となるときである。つまり、$\sum_{i=1}^{n}k_i^2$が最小となるように、$k_i$の条件を求めればよい。

ここで、$f(k_1,\cdots,k_n)=\sum_{i=1}^{n}k_i^2,\phi(k_1,\cdots,k_n)=\sum_{i=1}^{n}k_i-1$と定義すると、（１）より、$\phi(k_1,\cdots,k_n)=0$となる。

したがって、$\sum_{i=1}^{n}k_i^2$を最小とするk_iを$k_i^0\ (i=1,\cdots,n)$とすると、ラグランジュの乗数法より、$grad\ f(k_1^0,\cdots,k_n^0)=\lambda grad\ \phi(k_1^0,\cdots,k_n^0)$を満たす定数$\lambda$が存在する。よって、

$2k_i^0 = \lambda\,(i=1,\cdots,n)$となる。

一方、（1）より、$\sum_{i=1}^{n} k_i^0 = 1$となるので、$\lambda = \frac{2}{n}$となる。したがって、$k_i^0 = \frac{1}{n}$である。

よって、求めるk_iの条件は、$k_0 = 0$かつ$k_i = \frac{1}{n}(i=1,\cdots,n)$である。