経済統計概論模擬試験問題解答例

１．以下の問に答えよ．

(ア) 標本数が９０，標本平均が６，標本分散は９であった．このとき，母集団平均の９５％信頼区間を求めよ．

（解答例）

標本数が十分大きいので，$\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ は正規分布で近似できる．また，σ は $\hat{\sigma}=\sqrt{\dfrac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}$ で近似できる．従って，$P\left(-1.96\le\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}\le 1.96\right)=0.95$ の性質を使って９５％信頼区間を構成する．（）の中を変形すると $P\left(\overline{X}-1.96\dfrac{\hat{\sigma}}{\sqrt{n}}\le\mu\le\overline{X}+1.96\dfrac{\hat{\sigma}}{\sqrt{n}}\right)=0.95$ となる．標本分散は $S^2=\dfrac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$ であるから，$\dfrac{1}{90-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2=9$ より，$\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2=9\times 89$．$\hat{\sigma}=\sqrt{\dfrac{9\times 89}{90}}=2.98$．$\overline{X}=6$．これらを代入して，$P\left(6-1.96\dfrac{2.98}{\sqrt{90}}\le\mu\le 6+1.96\dfrac{2.98}{\sqrt{90}}\right)=0.95$．ここから，$P(5.38\le\mu\le 6.62)=0.95$ となり，９５％信頼区間は，$5.38\le\mu\le 6.62$ である．

（注意）

σ の推定値としてＳを使用しても良い．この場合は，$6-1.96\dfrac{3}{\sqrt{90}}\le\mu\le 6+1.96\dfrac{3}{\sqrt{90}}$，ここからも，$5.38\le\mu\le 6.62$ である．（初稿は計算間違えがありましたね．質問してくれた人，どうもです）

(イ) ６００世帯を調査する視聴率調査で視聴率（「北条時宗」初回視聴率）は１９．６％＜関東地区＞であった，ここから母集団に対する真の視聴率の９５％信頼区間を求めよ．

（解答例）

比率の推定を使用する．スライドシリーズ６の３１枚目によると，ｎを標本数，ｍを視聴している世帯数とすると，９５%信頼区間は,

$$\frac{m}{n}-1.96\sqrt{\frac{m}{n}\left(1-\frac{m}{n}\right)}\Big/\sqrt{n}\le\mu\le\frac{m}{n}+1.96\sqrt{\frac{m}{n}\left(1-\frac{m}{n}\right)}\Big/\sqrt{n}$$

であるから，$m/n=0.196$，$n=600$を代入して,

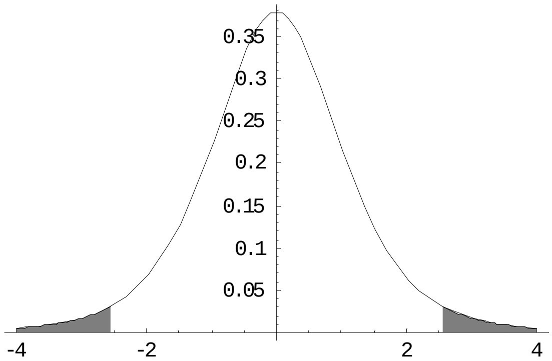
$$0.196-1.96\sqrt{0.196\times0.804}\Big/\sqrt{600}\le\mu\le0.196+1.96\sqrt{0.196\times0.804}\Big/\sqrt{600}$$

となる. 計算すると$0.164\le\mu\le0.228$であるから, ９５%信頼区間は16.4%から22.8%ということになる.

(ウ) 標本のデータ値は，１，３，２，６，１０，１３であった，母集団の分布は正規分布であることがわかっている．このとき，母平均の９５%信頼区間を求めよ.

（解答例）

$\dfrac{\overline{X}-\mu}{S\big/\sqrt{n}}$が自由度$n-1$のｔ分布に従う．ただし，$S^2=\dfrac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$である．自由度（６－１）のｔ分布の密度関数は下の通りであり，黒く塗ってある部分の確率合計は 0.05 である．この黒塗りの部分との境目の値は，教科書のｔ分布表の自由度（ｄｆ）5 の行の.0.025 の列を参照すればよい．参照すると±2.57 であるから，９５%信頼区間は，$P\left(-2.57\le\dfrac{\overline{X}-\mu}{S\big/\sqrt{n}}\le2.57\right)=0.95$を元に構成すればよい.



従って，$P\left(\overline{X}-2.57\dfrac{S}{\sqrt{n}}\le\mu\le\overline{X}+2.57\dfrac{S}{\sqrt{n}}\right)=0.95$である． $\overline{X}=5.83$，

$$S^2=\frac{1}{5-1}\left\{(1-5.83)^2+(3-5.83)^2+(2-5.83)^2+(6-5.83)^2+(10-5.83)^2+(13-5.83)^2\right\}$$
$$=22.97$$

，$S=\sqrt{22.97}=4.79$， $n=6$を代入して， $5.83-2.57\dfrac{4.79}{\sqrt{6}}\le\mu\le5.83+2.57\dfrac{4.79}{\sqrt{6}}$．

計算すると，$0.80\le\mu\le10.85$である．

(エ) 集団 A については標本数が１０１，平均が１０，標本分散は９であった．集団 B については標本数が５１，平均が１２，標本分散は１６であった．集団 A，B の母分散は等しいと仮定して，集団 A の母平均と集団 B の母平均が等しいか，等しくないか検定を行え．

（解答例）

＜理屈＞

帰無仮説は「両集団の母平均が等しい」である．標本数が十分大きいので，

$\dfrac{\overline{X}_{A集団}-\mu_{A集団}}{\sigma/\sqrt{n_{A集団}}}$， $\dfrac{\overline{X}_{B集団}-\mu_{B集団}}{\sigma/\sqrt{n_{B集団}}}$はともに標準正規分布で近似でき無相関である．ここで，添え字の「A集団」はA集団内で計算した標本平均，母平均，標本数を示し，添え字「Ｂ集団」はＢ集団内で計算した標本平均，母平均，標本数を示す．

$Z_1=\dfrac{\overline{X}_{A集団}-\mu_{A集団}}{\sigma/\sqrt{n_{A集団}}}$，$Z_2=\dfrac{\overline{X}_{B集団}-\mu_{B集団}}{\sigma/\sqrt{n_{B集団}}}$とおく．$\overline{X}_{A集団}-\mu_{A集団}=\left(\sigma/\sqrt{n_{A集団}}\right)Z_1$，$\overline{X}_{B集団}-\mu_{B集団}=\left(\sigma/\sqrt{n_{B集団}}\right)Z_2$と書き換え，差をとると，$\left(\overline{X}_{A集団}-\overline{X}_{B集団}\right)-\left(\mu_{A集団}-\mu_{B集団}\right)=\left(\sigma/\sqrt{n_{A集団}}\right)Z_1-\left(\sigma/\sqrt{n_{B集団}}\right)Z_2$．さらに帰無仮説が正しいとすると，$\mu_{A集団}=\mu_{B集団}$であるから，代入して，

$\overline{X}_{A集団}-\overline{X}_{B集団}=\left(\sigma/\sqrt{n_{A集団}}\right)Z_1-\left(\sigma/\sqrt{n_{B集団}}\right)Z_2$．

$E\left[\overline{X}_{A集団}-\overline{X}_{B集団}\right]=\left(\sigma/\sqrt{n_{A集団}}\right)E[Z_1]-\left(\sigma/\sqrt{n_{B集団}}\right)E[Z_2]=0$． Z_1,Z_2は無相関だから，

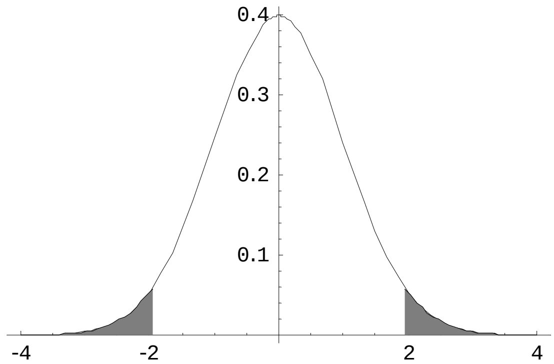
$V\left[\overline{X}_{A集団}-\overline{X}_{B集団}\right]=\dfrac{\sigma^2}{n_{A集団}}V[Z_1]+\dfrac{\sigma^2}{n_{B集団}}V[Z_1]=\left(\dfrac{1}{n_{A集団}}+\dfrac{1}{n_{B集団}}\right)\sigma^2$．さらに，$\overline{X}_{A集団}-\overline{X}_{B集団}=\left(\sigma/\sqrt{n_{A集団}}\right)Z_1-\left(\sigma/\sqrt{n_{B集団}}\right)Z_2$は無相関な正規分布の線形和で近似できるから，それ自身正規分布で近似できる．以上を総合すると，$\overline{X}_{A集団}-\overline{X}_{B集団}$は$N\left[0,\left(\dfrac{1}{n_{A集団}}+\dfrac{1}{n_{B集団}}\right)\sigma^2\right]$で近似できる．よって，$\dfrac{\overline{X}_{集団A}-\overline{X}_{集団B}}{\sigma\sqrt{\dfrac{1}{n_{A集団}}+\dfrac{1}{n_{B集団}}}}$は標準正規分布で近似できる．σはその推定値$\hat{\sigma}^2=\dfrac{1}{n_{A集団}+n_{B集団}}\left\{\sum_{集団A}\left(X_i-\overline{X}_{A集団}\right)^2+\sum_{集団B}\left(X_i-\overline{X}_{B集団}\right)^2\right\}$で推定して，統計量は

$\dfrac{\overline{X}_{集団A}-\overline{X}_{集団B}}{\hat{\sigma}\sqrt{\dfrac{1}{n_{A集団}}+\dfrac{1}{n_{B集団}}}}$で，境界値は標準正規分布表に基づいて決めることになる．

＜実際の解法＞

帰無仮説は「両集団の母平均が等しい」，対立仮説は「両集団の母平均が異なる」であるので, 両側検定となる. 統計量は $\dfrac{\overline{X}_{集団A}-\overline{X}_{集団B}}{\hat{\sigma}\sqrt{\dfrac{1}{n_{A集団}}+\dfrac{1}{n_{B集団}}}}$ で, 有意水準５%で検定すると,

境界値は帰無仮説の場合の分布の両側合計５%のはずれを考慮して決める.



上図の網掛け部が合計確率５%になるように決める. よって, ±1.96 が境界値である.

$S^2_{A集団}=\dfrac{1}{n-1}\sum_{集団A}\left(X_i-\overline{X}_{A集団}\right)^2$， $S^2_{B集団}=\dfrac{1}{n-1}\sum_{集団B}\left(X_i-\overline{X}_{B集団}\right)^2$ であるから,

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n_{A集団}+n_{B集団}}\left\{\sum_{集団A}\left(X_i-\overline{X}_{A集団}\right)^2+\sum_{集団B}\left(X_i-\overline{X}_{B集団}\right)^2\right\}\\ &= \frac{1}{n_{A集団}+n_{B集団}}\left\{\left(n_{A集団}-1\right)S^2_{A集団}+\left(n_{B集団}-1\right)S^2_{B集団}\right\}\\ &= \frac{1}{101+51}\{100\times 9+50\times 16\}=11.18\end{aligned}$$

よって，統計量の値は, $\dfrac{\overline{X}_{集団A}-\overline{X}_{集団B}}{\hat{\sigma}\sqrt{\dfrac{1}{n_{A集団}}+\dfrac{1}{n_{B集団}}}}=\dfrac{10-12}{3.34\sqrt{\dfrac{1}{101}+\dfrac{1}{51}}}=-3.49$． したがって,

統計量値＜－１．９６なので，統計量値は棄却域に入り，有意水準５%で帰無仮説は棄却され，対立仮説が採択される．したがって，二つの集団の母平均は異なるといえる.

(オ) （エ）と同じ状況で，集団 A，B の母分散は等しいと仮定して，集団 A の母平均より集団 B の母平均が大きいかどうか検定を行え.

（解答例）

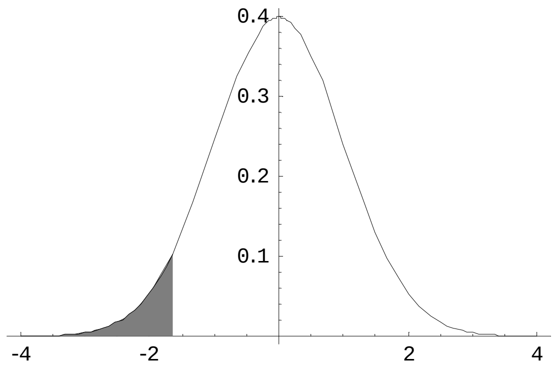
$\left(\overline{X}_{A集団}-\overline{X}_{B集団}\right)-\left(\mu_{A集団}-\mu_{B集団}\right)=\left(\sigma/\sqrt{n_{A集団}}\right)Z_1-\left(\sigma/\sqrt{n_{B集団}}\right)Z_2$　より，

$\left(\overline{X}_{A集団}-\overline{X}_{B集団}\right)=\left(\mu_{A集団}-\mu_{B集団}\right)+\left(\sigma/\sqrt{n_{A集団}}\right)Z_1-\left(\sigma/\sqrt{n_{B集団}}\right)Z_2$．従って，

$E\left(\overline{X}_{A集団}-\overline{X}_{B集団}\right)=\left(\mu_{A集団}-\mu_{B集団}\right)_2$，$V\left(\overline{X}_{A集団}-\overline{X}_{B集団}\right)=\sigma^2\left(1/n_{A集団}+1/n_{B集団}\right)$

である．対立仮説は，$\mu_{A集団}<\mu_{B集団}$ であるから，

$E\left(\overline{X}_{A集団}-\overline{X}_{B集団}\right)=\left(\mu_{A集団}-\mu_{B集団}\right)<0$ となり，$\overline{X}_{A集団}-\overline{X}_{B集団}<0$ の場合に棄却するのが妥当となる（片側検定)．したがって，統計量値が負の場合に棄却することになる．数で言うと，網掛けの部分が棄却域であるから，境界値は－1.64 である．統計量値がこの値より小さい場合に帰無仮説を棄却することになる．ところが，統計量値は（エ）によると－3.49 なので棄却域に入り，帰無仮説は有意水準５％で棄却される．



(カ)（エ）と同じ状況で，集団 A，B の母分散は等しくないと仮定して，集団 A の母平均より集団 B の母平均が等しいかどうか検定を行え．

（解答例）

＜理屈＞

$\dfrac{\overline{X}_{A集団}-\mu_{A集団}}{\sigma_{A集団}/\sqrt{n_{A集団}}}$，$\dfrac{\overline{X}_{B集団}-\mu_{B集団}}{\sigma_{B集団}/\sqrt{n_{B集団}}}$ はともに無相関の標準正規分布で近似できる．ただ

し，母分散が異なるので，それぞれの集団の母分散を $\sigma^2_{A集団}, \sigma^2_{B集団}$ としてある．

$Z_1 = \dfrac{\overline{X}_{A集団} - \mu_{A集団}}{\sigma_{A集団} / \sqrt{n_{A集団}}}$ ， $Z_2 = \dfrac{\overline{X}_{B集団} - \mu_{B集団}}{\sigma_{\mathrm{B}集団} / \sqrt{n_{B集団}}}$ とおく．

$\left(\overline{X}_{A集団} - \overline{X}_{\mathrm{B}集団}\right) - \left(\mu_{A集団} - \mu_{\mathrm{B}集団}\right) = \dfrac{\sigma_{A集団}}{\sqrt{n_{A集団}}} Z_1 - \dfrac{\sigma_{B集団}}{\sqrt{n_{B集団}}} Z_2$ ．帰無仮説

$\mu_{A集団} = \mu_{\mathrm{B}集団}$ の下では， $\overline{X}_{A集団} - \overline{X}_{\mathrm{B}集団} = \dfrac{\sigma_{A集団}}{\sqrt{n_{A集団}}} Z_1 - \dfrac{\sigma_{B集団}}{\sqrt{n_{B集団}}} Z_2$ である．

$E\left[\overline{X}_{A集団} - \overline{X}_{\mathrm{B}集団}\right] = \dfrac{\sigma_{A集団}}{\sqrt{n_{A集団}}} E[Z_1] - \dfrac{\sigma_{B集団}}{\sqrt{n_{B集団}}} E[Z_2] = 0$， Z_1, Z_2 は無相関だから，

$V\left[\overline{X}_{A集団} - \overline{X}_{\mathrm{B}集団}\right] = \dfrac{\sigma^2_{A集団}}{n_{A集団}} V[Z_1] + \dfrac{\sigma^2_{B集団}}{n_{B集団}} V[Z_2] = \dfrac{\sigma^2_{A集団}}{n_{A集団}} + \dfrac{\sigma^2_{B集団}}{n_{B集団}}$ ，また，

$\overline{X}_{A集団} - \overline{X}_{\mathrm{B}集団} = \dfrac{\sigma_{A集団}}{\sqrt{n_{A集団}}} Z_1 - \dfrac{\sigma_{B集団}}{\sqrt{n_{B集団}}} Z_2$ は無相関な正規分布の線形和で近似できる

から，それ自身正規分布で近似できる．以上を総合すると， $\overline{X}_{A集団} - \overline{X}_{B集団}$ の帰無仮説

の下での分布は $N\left[0, \dfrac{\sigma^2_{A集団}}{n_{A集団}} + \dfrac{\sigma^2_{B集団}}{n_{B集団}}\right]$ で近似できる．したがって，検定統計量は，

$\dfrac{\overline{X}_{A集団} - \overline{X}_{B集団}}{\sqrt{\dfrac{\hat{\sigma}^2_{A集団}}{n_{A集団}} + \dfrac{\hat{\sigma}^2_{B集団}}{n_{B集団}}}}$ でその分布は標準正規分布ということになる．ただし，

$\hat{\sigma}^2_{A集団} = \dfrac{1}{n} \sum_{A集団} \left(X_i - \overline{X}_A\right)^2$ ， $\hat{\sigma}^2_{B集団} = \dfrac{1}{n} \sum_{B集団} \left(X_i - \overline{X}_B\right)^2$ である．

<実際の解法>

帰無仮説は「両集団の母平均が等しい」，対立仮説は「両集団の母平均が異なる」であ

るので，両側検定となる．統計量は$\dfrac{\overline{X}_{A集団}-\overline{X}_{B集団}}{\sqrt{\dfrac{\hat{\sigma}^2_{A集団}}{n_{A集団}}+\dfrac{\hat{\sigma}^2_{B集団}}{n_{B集団}}}}$で，有意水準５％で検定すると，

境界値は帰無仮説の場合の分布の両側合計５％のはずれを考慮して決める．この時の境界値は ± 1.96 である．統計量の値を計算すると，

$$\frac{\overline{X}_{A集団}-\overline{X}_{B集団}}{\sqrt{\dfrac{\hat{\sigma}^2_{A集団}}{n_{A集団}}+\dfrac{\hat{\sigma}^2_{B集団}}{n_{B集団}}}}=\frac{10-12}{\sqrt{\dfrac{9\frac{100}{101}}{101}+\dfrac{16\frac{50}{51}}{51}}}=-3.18$$

で，境界値の外側なので，帰無仮説を棄却する．

(キ) （エ）と同じ状況で母集団分散が集団 A と集団 B で等しいかどうか検定を行え．

（解答例）

母集団が正規分布に従うと仮定してＦ検定を行う．帰無仮説：$\sigma_A^2=\sigma_B^2$，対立仮説：$\sigma_A^2\neq\sigma_B^2$である．統計量は，

$$\frac{S^2_{A集団}}{S^2_{B集団}}=\frac{\dfrac{1}{n_A-1}\displaystyle\sum_{A集団}^{n}\left(X_i-\overline{X}_A\right)^2}{\dfrac{1}{n_B-1}\displaystyle\sum_{B集団}^{n}\left(X_i-\overline{X}_B\right)^2}\cong\frac{\chi^2(n_A-1)/(n_A-1)}{\chi^2(n_B-1)/(n_B-1)}=F(n_A-1,n_B-1)$$

である．境界値は$F(100,50)$の 2.5%点と 97.5%点である．教科書の表は片側検定用の表なので，両側検定用の表を作成した．表には$F(n1,n2)$の 97.5%点を示した．

分子自由度 n1

分母自由度 n2	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	50	100
1	647.79	799.48	864.15	899.60	921.83	937.11	948.20	956.64	963.28	968.63	976.72	982.55	986.91	990.35	993.08	1008.10	1013.16
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43	39.44	39.44	39.45	39.48	39.49
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.28	14.23	14.20	14.17	14.01	13.96
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.68	8.63	8.59	8.56	8.38	8.32
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.46	6.40	6.36	6.33	6.14	6.08
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.30	5.24	5.20	5.17	4.98	4.92
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.60	4.54	4.50	4.47	4.28	4.21
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.13	4.08	4.03	4.00	3.81	3.74
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.80	3.74	3.70	3.67	3.47	3.40
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.55	3.50	3.45	3.42	3.22	3.15
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.21	3.15	3.11	3.07	2.87	2.80
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.05	2.98	2.92	2.88	2.84	2.64	2.56
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.89	2.82	2.76	2.72	2.68	2.47	2.40
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.77	2.70	2.64	2.60	2.56	2.35	2.27
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.60	2.55	2.50	2.46	2.25	2.17
50	5.34	3.97	3.39	3.05	2.83	2.67	2.55	2.46	2.38	2.32	2.22	2.14	2.08	2.03	1.99	1.75	1.66
100	5.18	3.83	3.25	2.92	2.70	2.54	2.42	2.32	2.24	2.18	2.08	2.00	1.94	1.89	1.85	1.59	1.48

ここから，$F(100,50)$の 97.5%点は 1.66，2.5%点は$F(50,100)$の 97.5%点の逆数なの

で，$1/1.59 = 0.63$である．一方統計量の値は，$\dfrac{S_A^2}{S_B^2} = \dfrac{9}{16} = 0.56$で，棄却域にはいるので有意水準５％で帰無仮説は棄却され，対立仮説が採択される．従って，集団Ａの分散とＢの分散は異なる．したがって, 異分散を仮定した平均差の検定を使う方がよい.

(ク) 集団 A は標本数が４，標本平均が６，標本分散は４，集団 B は標本数が８，標本平均が３，標本分散は１である．集団 A,B ともに母集団の分布は正規分布で分散は等しいと仮定して，集団 A と集団 B の母平均が等しいか検定せよ.

（解答例）

帰無仮説 $\mu_{A集団} = \mu_{B集団}$，対立仮説 $\mu_{A集団} \neq \mu_{B集団}$ として検定を行う．統計量 $\dfrac{\overline{X}_{A集団} - \overline{X}_{B集団}}{S\sqrt{\dfrac{1}{n_{A集団}} + \dfrac{1}{n_{B集団}}}}$ は帰無仮説の下で自由度 $n_{A集団} + n_{\mathrm{B}集団} - 2$ の t 分布に従う. この場合は，自由度は４＋８－２＝１０である．したがって，有意水準５％の場合の境界値は±2.23 である.

$$
\begin{aligned}
S^2 &= \frac{1}{n_A + n_B - 2}\left\{ \sum_{\mathrm{A}集団}\left(X_i - \overline{X}_A\right)^2 + \sum_{B集団}\left(X_i - \overline{X}_B\right)^2 \right\} \\
&= \frac{1}{n_A + n_B - 2}\left\{ (n_A - 1)S_{A集団}^2 + (n_{\mathrm{B}} - 1)S_{\mathrm{B}集団}^2 \right\} \\
&= \frac{1}{10}\{3 \times 4 + 7 \times 1\} = \frac{19}{10} = 1.9
\end{aligned}
$$

$S = \sqrt{1.9} = 1.38$ より，統計量の値は，$\dfrac{6-3}{1.38\sqrt{\dfrac{1}{4} + \dfrac{1}{8}}} = 3.55$ となり，棄却域にはいるので，帰無仮説は棄却される.

(ケ) ６００世帯に対する視聴率調査の結果，今年の NHK 大河ドラマ「新撰組！」の初回視聴率は２６．３％，最近の幕末物である「徳川慶喜」の初回視聴率は２４．

４％であった（いずれも関東地区）．母集団で考えた視聴率で「新撰組！」と「徳川慶喜」の視聴率が変わらないかどうかを検定せよ．

（解答例）

サンプル調査数をｎ，「新選組！」の視聴世帯数 m_1，「徳川慶喜」の視聴世帯数を m_2，「新選組！」の真の視聴率を p_1，「徳川慶喜」の真の視聴率を p_2 とする．帰無仮説は「二つの視聴率が等しい， $p_1 = p_2$」，対立仮説は「等しくない， $p_1 \neq p_2$」ととる．

$\dfrac{\sqrt{n}\left(m_1/n - p_1\right)}{\sqrt{p_1\left(1-p_1\right)}}$，$\dfrac{\sqrt{n}\left(m_2/n - p_2\right)}{\sqrt{p_2\left(1-p_2\right)}}$ はともに標準正規分布する． $Z_1 = \dfrac{\sqrt{n}\left(m_1/n - p_1\right)}{\sqrt{p_1\left(1-p_1\right)}}$，

$Z_2 = \dfrac{\sqrt{n}\left(m_2/n - p_2\right)}{\sqrt{p_2\left(1-p_2\right)}}$ とすると，

$$\frac{m_1}{n} - \frac{m_2}{n} = \left(p_1 - p_2\right) + \left(\frac{\sqrt{p_1\left(1-p_1\right)}}{\sqrt{n}} Z_1 - \frac{\sqrt{p_2\left(1-p_2\right)}}{\sqrt{n}} Z_2\right)$$

であるが，

$$E\left[\frac{m_1}{n} - \frac{m_2}{n}\right] = \left(p_1 - p_2\right) + \left(\frac{\sqrt{p_1\left(1-p_1\right)}}{\sqrt{n}} E\left[Z_1\right] - \frac{\sqrt{p_2\left(1-p_2\right)}}{\sqrt{n}} E\left[Z_2\right]\right)$$
$$= p_1 - p_2$$
$$V\left[\frac{m_1}{n} - \frac{m_2}{n}\right] = \left(\frac{p_1\left(1-p_1\right)}{n} V\left[Z_1\right] + \frac{p_2\left(1-p_2\right)}{n} V\left[Z_2\right]\right)$$
$$= \frac{p_1\left(1-p_1\right) + p_2\left(1-p_2\right)}{n}$$

であるから，

$$\frac{m_1}{n} - \frac{m_2}{n} \cong N\left(p_1 - p_2, \frac{p_1\left(1-p_1\right) + p_2\left(1-p_2\right)}{n}\right)$$

である．帰無仮説 $p_1 = p_2$ の下では，$\dfrac{m_1}{n} - \dfrac{m_2}{n} \cong N\left(0, \dfrac{p_1\left(1-p_1\right) + p_2\left(1-p_2\right)}{n}\right)$ となる．

ここから，$\left(\dfrac{m_1}{n} - \dfrac{m_2}{n}\right) \Big/ \sqrt{\dfrac{p_1\left(1-p_1\right) + p_2\left(1-p_2\right)}{n}} \cong N(0,1)$ となる． p_1, p_2 の推定値は，

それぞれ $m_1/n, m_2/n$ とすればよいから，検定統計量は，

$$\left(\frac{m_1}{n}-\frac{m_2}{n}\right)\Bigg/\sqrt{\frac{m_1/n(1-m_1/n)+m_2/n(1-m_2/n)}{n}}$$

$$=(0.263-0.244)\Bigg/\sqrt{\frac{0.263\times0.737+0.244\times0.756}{600}}=0.76$$

である，有意水準５％とすれば，これが±1.96 の中にあるかどうかである．この場合，棄却域に入っていないので，帰無仮説は棄却できず，二つの番組の真の視聴率に差がないということになる．

(コ) 今度は昔の幕末を扱った大河ドラマと比較する．６００世帯に対する視聴率調査の結果，今年の NHK 大河ドラマ「新撰組！」の初回視聴率は２６．３％，１９６８年の「竜馬がゆく」の初回視聴率は４３０世帯に対する視聴率調査の結果，２２．６％であった．母集団で考えた視聴率で「新撰組！」の第１回目が「竜馬がゆく」の第１回目を上回っているとまでいえるかどうか検定せよ．

（解答例）

帰無仮説は「二つの視聴率が等しい， $p_1=p_2$」，対立仮説は「$p_1>p_2$」である．

（ケ）とほぼ同様であるが，調査世帯数の変化を考慮する必要がある．現在の調査世帯数を n_1，６８年当時の調査世帯数を n_2 とすると，

$$\frac{m_1}{n_1}-\frac{m_2}{n_2}=(p_1-p_2)+\left(\frac{\sqrt{p_1(1-p_1)}}{\sqrt{n_1}}Z_1-\frac{\sqrt{p_2(1-p_2)}}{\sqrt{n_2}}Z_2\right)$$

となり，帰無仮説が正しければ，$\left(\frac{m_1}{n_1}-\frac{m_2}{n_2}\right)\Bigg/\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}\cong N(0,1)$ となる． p_1, p_2 の推定値は，それぞれ $m_1/n, m_2/n$ とすればよいので，統計量は $\left(\frac{m_1}{n_1}-\frac{m_2}{n_2}\right)\Bigg/\sqrt{\frac{m_1/n_1(1-m_1/n_1)}{n_1}+\frac{m_2/n_2(1-m_2/n_2)}{n_2}}$ である．境界値は片側検定と

なり，対立仮説の下では検定統計量の分子は正になりがちであるから，棄却域は正の範囲となる．具体的には，有意水準を５％とすれば，標準正規分布の 95%点が境界値となる．したがって，１．６４である．統計量の値は,

$$\left(\frac{m_1}{n_1}-\frac{m_2}{n_2}\right)\Bigg/\sqrt{\frac{m_1/n_1\left(1-m_1/n_1\right)}{n_1}+\frac{m_2/n_2\left(1-m_2/n_2\right)}{n_2}}$$

$$=\left(0.263-0.226\right)\Bigg/\sqrt{\frac{0.263\times 0.737}{600}+\frac{0.226\times 0.774}{430}}=1.37$$

となり,「新選組！」の初回視聴率が「竜馬がゆく」の初回視聴率を上回ったとまではいえない.

(サ) ある正規分布に従う集団の母分散は９であった．母平均に関する９５％信頼区間の幅を０．０１以下にしたい．標本をいくつ以上採ればよいか.

（解答例）

９５％信頼区間は，$P\left(\overline{X}-1.96\frac{\sigma}{\sqrt{n}}\le\mu\le\overline{X}+1.96\frac{\sigma}{\sqrt{n}}\right)=0.95$であるから，信頼区間の幅は$2\times 1.96\times\frac{\sigma}{\sqrt{n}}=\frac{2\times 1.96\times 3}{\sqrt{n}}\le 0.01$となる．これを解いて，$n\ge 1.38298\times 10^{6}$を得る.

２. 以下の表は NHK 大河ドラマの関東地区視聴率表である.（ビデオリサーチ社 HP より）

放送年度	タイトル	初回視聴率(%)	順位	最高視聴率(%)	順位	平均視聴率(%)	順位
2004年度	新選組！	26.3	23				
2003年度	武蔵　MUSASHI	21.7	37	24.6	36	16.7	40
2002年度	利家とまつ･加賀百万石物語	26.1	25	27.6	28	22.1	24
2001年度	北条時宗	19.6	39	21.2	41	18.5	35
2000年度	葵徳川三代	22.6	35	22.6	39	18.5	35
1999年度	元禄繚乱	25.0	29	28.5	26	20.2	31
1998年度	徳川慶喜	24.4	30	29.7	24	21.1	27
1997年度	毛利元就	25.3	28	28.5	26	23.4	21
1996年度	秀吉	26.6	21	37.4	7	30.5	8
1995年度	八代将軍吉宗	22.1	36	31.4	19	26.4	10
1994年度	花の乱(94年4～12月)	17.9	41	18.3	42	14.1	42
1993年度	琉球の風(93年1～6月)	24.1	31	24.1	37	17.3	39
	炎立つ(93年7月～94年3月)	20.8	38	21.6	40	17.7	38
1992年度	信　長	25.4	27	33.0	15	24.6	16
1991年度	太　平　記	34.6	4	34.6	12	26.0	12
1990年度	翔ぶが如く	26.9	20	29.3	25	23.2	22
1989年度	春　日　局	33.1	6	39.2	6	33.1	3
1988年度	武田信玄	42.5	1	49.2	2	39.2	2
1987年度	独眼竜政宗	28.7	14	47.8	3	39.7	1
1986年度	い　の　ち	26.6	21	36.7	10	29.3	9
1985年度	春の波涛	23.9	32	24.7	35	18.2	37
1984年度	山河燃ゆ	30.5	10	30.5	21	21.1	27
1983年度	徳川家康	34.9	3	37.4	7	31.2	6
1982年度	峠の群像	31.3	9	33.8	14	23.7	19
1981年度	おんな太閤記	32.2	8	36.8	9	31.8	5
1980年度	獅子の時代	26.2	24	26.7	33	21.0	29
1979年度	草燃える	27.9	16	34.7	11	26.3	11
1978年度	黄金の日日	29.8	13	34.4	13	25.9	13
1977年度	花　神	16.5	43	25.9	34	19.0	34
1976年度	風と雲と虹と	28.0	15	30.1	22	24.0	18
1975年度	元禄太平記	29.9	12	41.8	4	24.7	15
1974年度	勝　海舟	30.5	10	30.9	20	24.2	17
1973年度	国盗り物語	27.5	18	29.9	23	22.4	23
1972年度	新･平家物語	17.3	42	27.2	31	21.4	26
1971年度	春の坂道	19.1	40	27.5	30	21.7	25
1970年度	樅の木は残った	27.6	17	27.6	28	21.0	29
1969年度	天と地と	23.5	33	32.4	17	25.0	14
1968年度	竜馬がゆく	22.9	34	22.9	38	14.5	41
1967年度	三　姉　妹	27.0	19	27.0	32	19.1	33
1966年度	源　義経	32.5	7	32.5	16	23.5	20
1965年度	太　閤　記	35.2	2	39.7	5	31.2	6
1964年度	赤穂浪士	34.3	5	53.0	1	31.9	4
1963年度	花の生涯	25.6	26	32.3	18	20.2	31

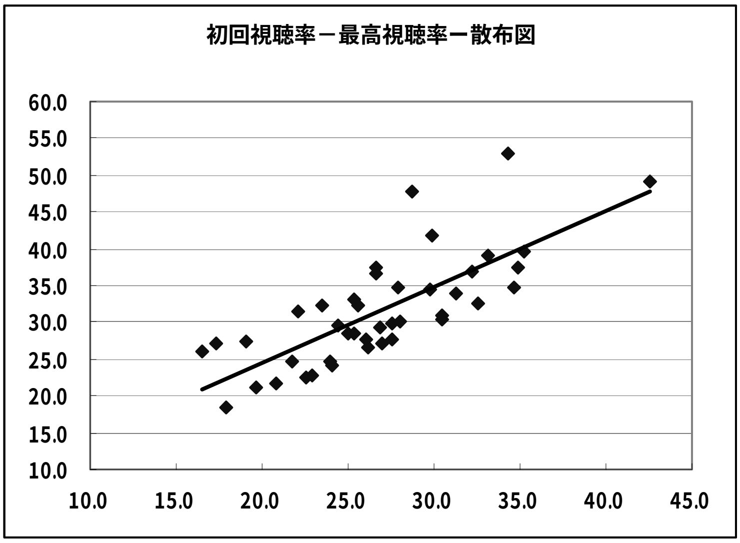
ここから，（推定）初回視聴率と（推定）最高視聴率の標本相関係数，（推定）初回視聴率と（推定）平均視聴率の標本相関係数を求めよ．そこから，初回視聴率の重要性を述べよ．

（答え）（推定）初回視聴率と（推定）最高視聴率の標本相関係数＝０．７６

（推定）初回視聴率と（推定）平均視聴率＝０．７２

とかなり相関があることが解る．従って，初回視聴率が悪いとその後の１年間の視聴率も芳しくないという可能性が高い．したがって，初回視聴率は大河ドラマの成績を占う上でかなり信頼できる指標といえる．ちなみに，初回にふるわず盛り返したのを，最高視聴率順位が一桁で初回視聴率順位が２桁という指標で取り出してみると「秀吉」（竹中直人主演），「独眼竜政宗」（渡辺謙主演），「元禄太平記」（石坂浩二主演，江守徹が大石内蔵助役）しかない．逆に最高視聴率順位が２桁で初回視聴率順位が１桁という例も「太平記」，「峠の群像」，「源義経」だけである．また，散布図を書き，近似直線を書き入れると以下の図になるが，初回視

聴率と最高視聴率の関係が強いことが解ると同時に，初回視聴率の割に最高視聴率が驚くほど悪い例はほとんどなく，初回のわりに最高視聴率が驚くほど良かった例も，長谷川一夫主演「赤穂浪士」と渡辺謙主演の「独眼竜政宗」しかない．（さすが名優長谷川一夫と渡辺謙である）したがって，今回の三谷幸喜原作の「新選組！」はそこそこの成功を収めると予想できる．ちなみに，線形回帰モデルで予測する場合の式は，最大視聴率＝１．０４×初回視聴率＋３．５であり，「新選組！」の予想最高視聴率は３１．０％となる．ほんとうは誤差評価が必要だが，最大視聴率も推定値，初回視聴率も推定値なので，測定誤差モデルというのを使ってやらなければいけない．それは説明が難しいので省略．~~散布図でみると下側５％以上ずれているケースはまれだから，おそらく最高視聴率は２６％を下回らないであろう．ということは，最近大河ドラマの視聴率低迷が続く中で「元禄繚乱」の水準に最高視聴率を回復する可能性が高いといえる．~~（なんかトリビアルなことを言っていました．初回が２６．３％だから最高視聴率が２６．３％を下回ることはない．すでに「元禄繚乱」のレベルに回復していますね．それに，この回帰モデルは説明変数である初回視聴率と近似直線と実際の値の差＝誤差とが相関を持っています．なぜかというと，被説明変数の最高視聴率は説明変数の初回視聴率を決して下回らないから．だから，回帰モデルとしては変則です．さらに，散布図の近似曲線下のデータに関してのはずれが小さく見えるのも最高視聴率が初回視聴率を超えないというトリビアルな事実の結果です．皆さんは下の図に最高視聴率＝初回視聴率の線を書き入れてみてください．いわば，統計をつかって嘘をつく方法の一種ですね．だから，はずれについては信用しないでください．中心線として，最大視聴率＝１．０４×初回視聴率＋３．５の式がでたというだけです．このように統計学をよく知っていないと自分で自分にだまされることにもなりかねません．ですからちゃんと勉強しましょう．来年の回帰分析の授業では統計学をちゃんと知っていないとこんな判断ミスをするという例でこの失敗を取り上げることになりそうです．）つまり，時計の針を５年戻すことが出来るというわけで，さすが三谷幸喜といってもよいであろう．



３．以下の問いに答えなさい.

左の列と右の列の正しい組み合わせを線で結びなさい.

（１）正規分布　　　ａ．独立な標準正規分布の２乗の和

（２）カイ２乗分布　ｂ．分布のグラフは左右対称でしかも釣り鐘型でガウス分布とも呼ばれる.

（３）ｔ分布　　　　ｃ．二項分布に極めて似ている分布

４．下の表はある病院を訪れた 100 人の患者の収縮期血圧(mmHg)の値である．以下の問いに答えよ.

階級	階級値	度数
100～110	１０５	3
110～120	１１５	8
120～130	１２５	11
130～140	１３５	12
140～150	１４５	15
150～160	１５５	27
160～170	１６５	10
170～180	１７５	7
180～190	１８５	6
190～200	１９５	1
計		100

（イ）　階級値を埋めよ.

表参照

（ロ） 収縮期血圧の平均を求めよ.

階級値×度数の合計を標本数１００で割る．147.7mmHg である.

（ハ） 分散を求めよ

（階級値－平均）の二乗×度数の合計を１００で割る．419.71mmHg である.

（ニ） 標本分散を求めよ

（階級値－平均）の二乗×度数の合計を９９で割る．423.95mmHg.

（ホ）変動係数を標本分散から求めよ

$\sqrt{標本分散}/平均$を計算する．0.139

５．全国の大学生 10000 人の体重測定を実施した．平均 48kg, 標準偏差 4kg の正規分布をしていた．計算過程を書くこと.

Ｘを大学生の体重値とする． $Z=\dfrac{X-\mu}{\sigma}\sim N(0,1)$ を使う．数値を当てはめると $Z=\dfrac{X-48}{4}\sim N(0,1)$ である.

（イ）体重 58kg 以上の人は何人いるか

$P(X\geq 58)=P\left(Z=\dfrac{X-48}{4}\geq\dfrac{58-48}{4}=2.5\right)$ を使う．標準正規分布表より，

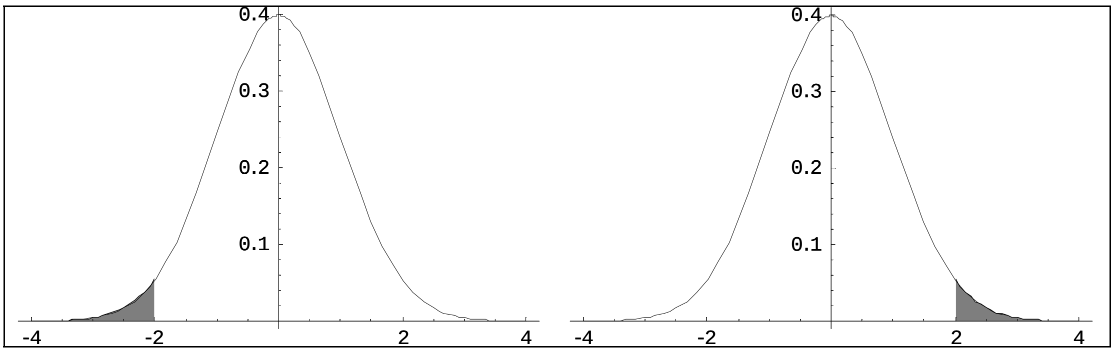
$P(Z\geq 2.5)=1-P(Z<2.5)=1-0.9938=0.0062$ である．全員は 10000 人であるから，その 0.0062 は６２人ということになる.

（ロ）体重 40kg 以上の人は何人いるか

$P(X\geq 40)=P\left(Z=\dfrac{X-48}{4}\geq\dfrac{40-48}{4}=-2.0\right)$ より，

$P(Z\geq -2.0)=1-P(Z<-2.0)=1-P(Z>2.0)=P(Z\leq 2.0)=0.9773$ ということで，９７７３人.

$\boxed{P(Z<-2)=P(Z>2)}$ の説明



（ハ）体重 42kg 以上 52kg の人は何人いるか

$$P(42 \le X \le 52) = P\left(\frac{42-48}{4} \le \frac{X-48}{4} \le \frac{52-48}{4}\right) = P(-1.5 \le Z \le 1.0)$$
$$= P(Z \le 1.0) - P(Z \le -1.5) = P(Z \le 1.0) - P(Z \ge 1.5)$$
$$= P(Z \le 1.0) - \{1 - P(Z \le 1.5)\}$$
$$= 0.8413 + 0.9332 - 1 = 7745$$

より７７４５人である.