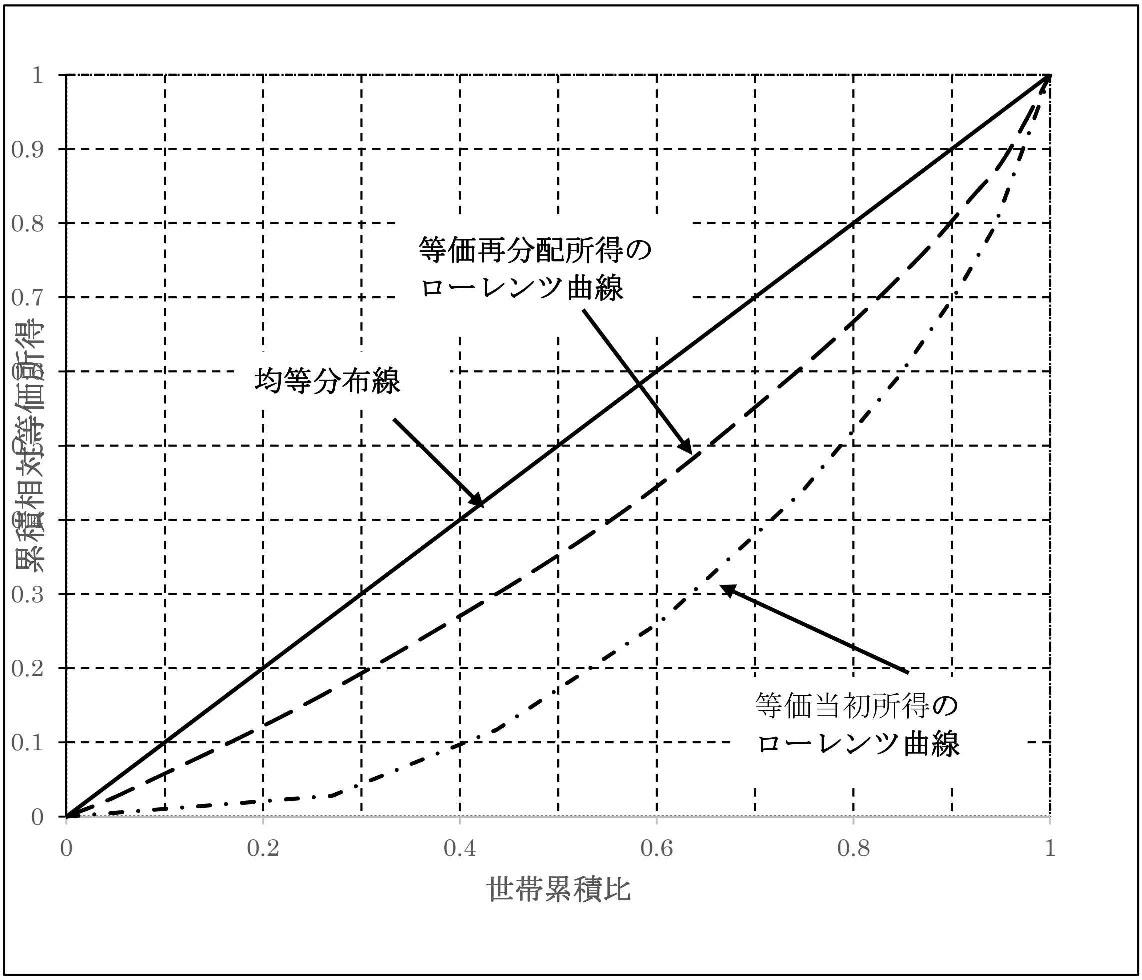
平成 27 年公認会計士論文式試験　統計学　第 7 問、第 8 問　解答
クレアール　無断複製・流布を禁じます

第 7 問

問題 1

(1)



(2)　ジニ係数　＝　0.21　(小数点以下 2 ケタ)

(3)

① 当初所得のローレンツ曲線よりも、再分配所得のローレンツ曲線の方が、均等分布線により近い
② ジニ係数も、再分配所得のジニ係数の方が当初所得の場合よりも小さい
などの理由から、所得再分配によって世帯間の所得格差が縮小しているといえるであろう。また両者の曲線が交差することがないことから、所得格差の是正は、特定の所得層に偏ることなく、すべての所得層にくまなく実施されているといえよう。

クレアール　無断複製・流布を禁じます

問題 2

(1) $E(X) =$ $\boxed{0}$　　$E(Y) =$ $\boxed{0.2}$

(2) $V(X) =$ $\boxed{0.6}$　　$V(Y) =$ $\boxed{0.56}$

(3) $Cov(X, Y) =$ $\boxed{0}$

$Corr(X, Y) =$ $\boxed{0}$

(4) $E(X+Y) =$ $\boxed{0.2}$

$E(X-Y) =$ $\boxed{-0.2}$

(5) $V(X+Y) =$ $\boxed{1.16}$

$V(X-Y) =$ $\boxed{1.16}$

(6) $Cov\,(X+Y\,, X\text{-}Y) =$ $\boxed{0.04}$

$Corr\,(X+Y\,, X\text{-}Y) =$ $\boxed{\dfrac{1}{29}}$

平成 27 年公認会計士論文式試験　統計学　第 7 問、第 8 問　解答

クレアール　無断複製・流布を禁じます

問題 3

(1) $\Pr(X=1) =$ 36/125　　$\Pr(X=2) =$ 54/125

$\Pr(X=3) =$ 27/125

(2) $\Pr(Y=1) =$ 3/10　　$\Pr(Y=2) =$ 3/5

$\Pr(Y=3) =$ 1/10

(3) $E(X) =$ 9/5　　$V(X) =$ 18/25

(4) $E(Y) =$ 9/5　　$V(Y) =$ 9/25

(5)

$\Pr(X=x) =$ ${}_{\mathrm{n}}C_x (M/N)^x (1-\mathrm{M/N})^{n-x}$

$\Pr(Y=y) =$ ${}_{\mathrm{M}}C_y\ {}_{\mathrm{N-M}}C_{n-y}\ /\ {}_{\mathrm{N}}C_n$

クレアール　無断複製・流布を禁じます

第 8 問

問題 1

(1) 平均μ、分散σ^2/n の正規分布

(2) 自由度 n の χ^2 分布

(3) 自由度 n の t 分布

(4) 自由度 n の χ^2 分布

(5) 自由度 $n-1$ の χ^2 分布

(6) 自由度 $n-1$ の t 分布

平成 27 年公認会計士論文式試験　統計学　第 7 問、第 8 問　解答

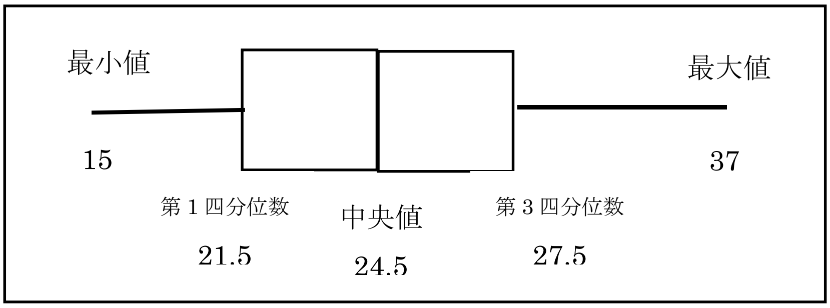
クレアール　無断複製・流布を禁じます

問題 2

(1)　1)　中央値(中位値)　＝　24.5

2)　四分位範囲　＝　6

3)　箱ひげ図



(2)　1)

値	10	14	15	18	20	20	21	21	22	22
順位	1	2	3	4	5.5	5.5	7.5	7.5	9.5	9.5

値	23	24	25	26	27	27	28	28	29	37
順位	11	12	13	14	15.5	15.5	17.5	17.5	19	20

2)　順位の合計(サッカー部)　＝　61

3)

(仮説検定の詳細と検定結果)

帰無仮説 H_0：　「野球部とサッカー部の懸垂の回数の分布に差がない」
対立仮説 H_1：　「野球部とサッカー部の懸垂の回数の分布に差がある」
サッカー部の順位和は 61 であるが、帰無仮説 H_0 が正しいときは、検定統計量の下位 2.5%は、問題の表より 58、また上位 2.5%は 8×(12＋8＋1)－58＝110 となることから、帰無仮説 H_0 は棄却されない。よって野球部とサッカー部の懸垂の回数の分布に差がないといえる。

クレアール　無断複製・流布を禁じます

問題 3

(1)　予測式

$$C_t = 5.416\times10^4 + 1.682\times10^{-1}Yd_t + 1.387\times10^2 Trend_t$$

(2)　予測値　＝　74864　(整数)

(3)　差の絶対値　＝　4361　(整数)

(4)

(理由)

　ダミー変数を 3 つにするのは、この線形回帰分析では、定数項 α が第 1 四半期の時の基準値にし、ダミー変数は、それ以外の時期の値が第 1 四半期の値との差を表しているからである。

　もしこの推計式に、あらたな第 1 四半期のダミー変数を加えて、ダミー変数を 4 つにすれば、定数項と第 1 四半期のダミー変数との違いが区別つかなくなり、結果、多重共線性の問題を引き起こすことになる。

クレアール　無断複製・流布を禁じます

平成 27 年公認会計士論文式試験　統計学　第 7 問、第 8 問　解答解説

第 7 問

問題 1

(1)　ジニ係数は、世帯割合の累積比とそれに対応する所得の累積比を結んだ直線である。そして均等分布線(対角線、すなわち角度が 45 度の直線)は、所得分布が完全に均等である時に実現される、世帯割合と所得割合の累積比の組み合わせである。これは解答の図のように描かれる。

(2)　ジニ—係数は、均等分布線より右下の 3 角形の面積から、ローレンツ直線より右下の面積を引いたものを 2 倍したものに等しくなる。これはローレンツ曲線が均等分布線と等しいときには 0、また完全不平等なケース(ローレンツ曲線が L 字型になる)では、1 になる。等価再分配の時のジニ—係数は、

$$2\times\left\{\begin{array}{c}0.5\times(0.033+0)(0.06-0)+\ 0.5\times(0.174+0.033)(0.274-0.06)\\+\,0.5\times(0.391+0.174)(0.544-0.274)+\ 0.5\times(0.597+0.391)(0.74-0.544)\\0.5\times(0.742+0.597)(0.858-0.74)+\ 0.5\times(0.836+0.742)(0.921-0.858)\\+0.5\times(0.889+0.836)(0.954-0.921)+\ 0.5\times(1+0.889)(1-0.954)\end{array}\right\}$$

となり、これを計算すると 0.21(小数点以下 2 桁)となる、

(3) ローレンツ曲線とジニ係数の結果から、①当初所得のローレンツ曲線よりも、再分配所得のローレンツ曲線の方が、均等分布線により近い、②ジニ係数も、再分配所得のジニ係数の方が当初所得の場合よりも小さい、などの理由から、所得再分配によって世帯間の所得格差が縮小していると結論付けられる。また両者の曲線が交差することがないことから、所得格差の是正は、特定の所得層に偏ることなく、すべての所得層にくまなく実施されているといえよう。仮に中間所得層の家計から、低所得層および高所得層に有利な所得分配がなされているならば、高所得層に対応する等価再分配所得のローレンツ曲線が、等価当初所得のローレンツ曲線を横切る結果となる可能性がある。

クレアール　無断複製・流布を禁じます

第 7 問

問題 2

(1)同時確率分布の定義より、

Xが－1 をとる確率は、0.1＋0＋0.2＝0.3
Xが 0 をとる確率は、0＋0.4＋0＝0.4
Xが 1 をとる確率は、0.1＋0＋0.2＝0.3

Yが－1 をとる確率は、0.1＋0＋0.1＝0.2
Yが 0 をとる確率は、0＋0.4＋0＝0.4
Yが 1 をとる確率は、0.2＋0＋0.2＝0.4

となる。よって期待値の定義より

$$\mathrm{E(X)} = -1\times(0.1+0+0.2)+ 0\times(0+0.4+0)+1\times(0.1+0+0.2)=0$$
$$\mathrm{E}(Y) = -1\times(0.1+0+0.1)+ 0\times(0+0.4+0)+1\times(0.2+0+0.2)=0.2$$

となる。

(2)分散の定義より、$\mathrm{V(X)} = E(X^2)-\left(E(X)\right)^2$ となる。

$$\mathrm{E}(X^2) = (-1)^2\times(0.1+0+0.2)+ (0)^2\times(0+0.4+0)+(1)^2\times(0.1+0+0.2)=0.6$$

なので、$\mathrm{V(X)} = 0.6-0^2$ ＝0.6となる。同様に

$$\mathrm{E}(Y^2) = (-1)^2\times(0.1+0+0.1)+ (0)^2\times(0+0.4+0)+(1)^2\times(0.2+0+0.2)=0.6$$

なので、$\mathrm{V(Y)} = 0.6-0.2^2$ ＝0.56となる。

(3)共分散の定義より、$\mathrm{Cov(X,Y)} = E((X-E(X))(Y-E(Y)))=E(XY)-E(X)E(Y)$ となる。

$$E(XY)=(-1)\times(-1)\times0.1+(-1)\times(0)\times0+(-1)\times(1)\times0.2$$
$$+(0)\times(-1)\times0+(0)\times(0)\times0.4+(0)\times(1)\times0$$
$$+(1)\times(-1)\times0.1+(1)\times(0)\times0+(1)\times(1)\times0.2=0$$

また$E(X)E(Y)$＝0　であることから、$\mathrm{Cov(X,Y)}=0$となる。また相関係数の定義より、

$$\mathrm{Corr(X,Y)}=\frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}}$$

となることから、$\mathrm{Corr(X,Y)}=0$　となる。

クレアール　無断複製・流布を禁じます

⑷$\mathrm{E(X+Y)=E(X)+E(Y)=0+0.2=0.2}$、$\mathrm{E(X－Y)=E(X)－E(Y)=0－0.2=－0.2}$となる。

(5)　　　分散の定義より、

$$\mathrm{V(X+Y)=V(X)+V(Y)+2Cov(X,Y)=0.6+0.56+2\times 0=1.16}$$

$$\mathrm{V(X－Y)=V(X)+V(Y)－2Cov(X,Y)=0.6+0.56－2\times 0=1.16}$$

⑹共分散の定義より、

$$\mathrm{Cov(X+Y,X-Y)}=E\Big(\big(X+Y-E(X+Y)\big)\big(X-Y-E(X-Y)\big)\Big)$$

$$=E\Big(\big(X-E(X)\big)^2\Big)-E\Big(\big(Y-E(Y)\big)^2\Big)=V(X)-V(Y)=0.6-0.56=0.04$$

となる。また相関係数の定義より、

$$\mathrm{Corr(X＋Y\,,X-Y)}=\frac{Cov(X+Y,X-Y)}{\sqrt{V(X+Y)V(X-Y)}}=\frac{0.04}{\sqrt{1.16^2}}=\frac{1}{29}$$

となる。

平成 27 年公認会計士論文式試験　統計学　第 7 問、第 8 問　解答

クレアール　無断複製・流布を禁じます

第 7 問

問題 3

この問題は、試行 A が二項分布、試行 B が超幾何分布に従う。そして N 個の玉のうち、M 個が白で、それを n 試行繰り返したときに白が得られる確率は、

試行 A：　$\Pr(X{=}x) = {}_{\mathrm{n}}C_x \left(\frac{M}{N}\right)^x \left(1-\frac{\mathrm{M}}{\mathrm{N}}\right)^{n-x}$

試行 B：　$\Pr(Y = y) = {}_{\mathrm{M}}C_y \ {}_{\mathrm{N-M}}C_{n-y} \ / \ {}_{\mathrm{N}}C_n$

なお ${}_{\mathrm{n}}C_x = n!/(x!\,(n-x)!)$ である。試行 A は、1 回目の試行で白の玉が得られる確率 p が、M/N であることがわかれば、理解できるであろう。また試行 B は非復元抽出の際に用いられる確率分布である。これをもとに、以下解説を行う。

(1)　N=5、M=3、n=3 であることから、試行 A については、

$\Pr(X{=}x) = {}_3C_x(3/5)^x(1-3/5)^{3-x} = {}_3C_x(3/5)^x(2/5)^{3-x}$

となる。よって、

$\Pr(X{=}1) = {}_3C_1(3/5)^1(1-3/5)^{3-1} = {}_3C_1(3/5)^1(2/5)^2 = 3\times(3/5)^1\times(2/5)^2 = 36/125$

$\Pr(X{=}2) = {}_3C_2(3/5)^2(1-3/5)^{3-2} = {}_3C_2(3/5)^2(2/5)^1 = 3\times(3/5)^2\times(2/5)^1 = 54/125$

$\Pr(X{=}3) = {}_3C_3(3/5)^3(1-3/5)^{3-3} = {}_3C_3(3/5)^3(2/5)^0 = 1\times(3/5)^3\times(2/5)^0 = 27/125$

(2)　N=5、M=3、n=3 であることから、試行Bについては、

$\Pr(Y{=}y) = {}_{\mathrm{M}}C_y \ {}_{\mathrm{N-M}}C_{n-y} \ / \ {}_{\mathrm{N}}C_n$

となる。よって、

$\Pr(Y{=}1) = {}_3C_1 \ {}_{5-3}C_{3-1} - {}_5C_3 = \frac{3!2!}{1!2!2!0!}\frac{3!2!}{5!} = 3/10$

$\Pr(Y{=}2) = {}_3C_2 \ {}_{5-3}C_{3-2} - {}_5C_3 = \frac{3!2!}{2!1!1!1!}\frac{3!2!}{5!} = 3/5$

$\Pr(Y{=}1) = {}_3C_3 \ {}_{5-3}C_{3-3} - {}_5C_3 = \frac{3!2!}{3!0!0!2!}\frac{3!2!}{5!} = 1/10$

(3)　二項分布の期待値 E(X)は、$E(X) = p{=}M/N$ 、また分散 V(X)は、$V(X) = np(1\text{-}p) = n(M/N)(1-M/N)$ である。よって

$E(X) = 3\times(3/5) = 9/5$、　$V(X) = 3\times(3/5)\times(2/5) = 18/25$

(4)　超幾何分布の期待値 E(Y)は、$E(X) = n(M/N)$ 、また分散 V(X)は、

クレアール　無断複製・流布を禁じます

$V(X) = n(M/N)(1-M/N)(N-n)/(N-1)$である。よって、

$E(Y) = 3 \times (3/5) = 9/5$、　$V(X) = 3 \times (3/5) \times (2/5) \times ((5-3)/(5-1)) = 9/25$

(5)　二項分布に従うので、$\Pr(X=x) = {}_{n}C_{x}\left(\frac{M}{N}\right)^{x}\left(1-\frac{M}{N}\right)^{n-x}$

(6)　超幾何分布に従うので、$\Pr(Y = y) = {}_{M}C_{y}\ {}_{N-M}C_{n-y}\ /\ {}_{N}C_{n}$

クレアール　無断複製・流布を禁じます

第8問

問題1

(1) $\bar{X}=\frac{X_1+\cdots+X_n}{n}$ であり、また各X_i $(\mathrm{i}=1,2,\cdots,\mathrm{n})$は正規分布で、各々独立であることから、その和である$\bar{X}$もまた、正規分布に従う。また$\mathrm{E}(\overline{X})=\frac{E(X_1)+\cdots+E(X_n)}{n}=\frac{n\times\mu}{n}=\mu$、

$\mathrm{V}(\overline{X})=\frac{V(X_1)+\cdots+V(X_n)}{n^2}=\frac{n\times\sigma^2}{n^2}=\frac{\sigma^2}{n}$ となり、$\bar{X}\sim N(\mu,\sigma^2/n)$ となる。

(2)$\sum_{i=1}^{n}Z_i^2$=$Z_1^2+\cdots+Z_n^2$ は、χ^2分布の定義より自由度nのχ^2分布に従う。

(3)$\frac{Z}{\sqrt{Y/n}}$ は、分子が標準正規分布、また分母が自由度nのχ^2分布に従う変数Yを自由度nで割って平方値をとっていることから、t分布の定義より自由度nのt分布に従う。

(4)　$\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}=\sum_{i=1}^{n}(\frac{X_i-\mu}{\sigma})^2$ となり、$\frac{X_i-\mu}{\sigma}$が平均0、分散1の標準席分布に従うことから、χ^2分布の定義より自由度nのχ^2分布に従う

(5)$\sum_{i=1}^{n}(\frac{X_i-\mu}{\sigma})^2=\sum_{i=1}^{n}(\frac{X_i-\bar{X}}{\sigma})^2+(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}})^2$ となり、$\sum_{i=1}^{n}(\frac{X_i-\mu}{\sigma})^2$が自由度$n$の$\chi^2$分布に、また$(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}})^2$が自由度1の$\chi^2$分布に従うことから(平均0、分散1の標準正規分布になっている)、$\sum_{i=1}^{n}(\frac{X_i-\bar{X}}{\sigma})^2$は自由度$n$−1の$\chi^2$分布に従う。

(6)$\frac{\sqrt{n}(\bar{X}-\mu)}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2/(n-1)}}=\frac{(\bar{X}-\mu)/(\sigma/\sqrt{n})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X}/\sigma)^2/(n-1)}}$ となる。分母は自由度n−1のχ^2分布に従うχ^2分布に従う変数を、自由度n−1で割った後で、平方値をとっていること、また分子が平均0、分散1の標準正規分布になることから、t分布の定義より、自由度n−1のt分布に従う

クレアール　無断複製・流布を禁じます

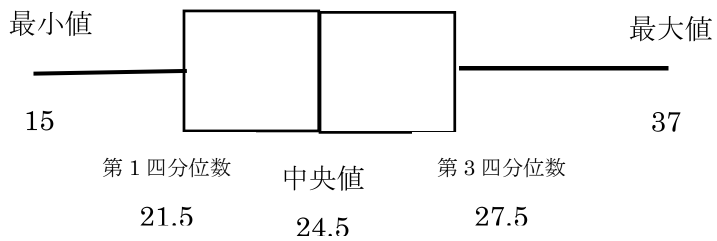
第 8 問

問題 2

(1)　1)　中央値は、データ数が 12 と偶数であることから、定義より、(25+24) / 2 = 24.5

2)　第 3 四分位点が(28+27) / 2 = 27.5、第 1 四分位点が(22+21) / 2 = 21.5　である。そして四分位範囲は、第 3 四分位点から第 1 四分位点を引いた値であり、27.5－21.5 = 6 となる。

3)　最小値は 15、最大値は 37、第 3 四分位点が 27.5、第 1 四分位点が 21.5、そして中央値が 24.5 であることから、箱ひげ図は以下のようにかける。



（2）　1)

値	10	14	15	18	20	20	21	21	22	22
順位	1	2	3	4	5.5	5.5	7.5	7.5	9.5	9.5

値	23	24	25	26	27	27	28	28	29	37
順位	11	12	13	14	15.5	15.5	17.5	17.5	19	20

2)　サッカー部の値に対する順位は、それぞれ 1、2、4、5.5、7.5、9.5、14、17.5 であることから、それを合計すると、1+2+4+5.5+7.5+9.5+14+17.5=61 となる。

3)　順位和の検定を行うと、以下のようになる。まず以下の仮説を設定する。

帰無仮説 H_0:　「野球部とサッカー部の懸垂の回数の分布に差がない」

対立仮説 H_1:　「野球部とサッカー部の懸垂の回数の分布に差がある」

クレアール　無断複製・流布を禁じます

そのうえで、サッカー部の順位和は 61 であるが、帰無仮説 H_0 が正しいときは、検定統計量の下位 2.5%は、問題の表より 58、また上位 2.5%は 8×(12＋8＋1)－58＝110 となることから、帰無仮説 H_0 は棄却されない。よって野球部とサッカー部の懸垂の回数の分布に差がないといえる。

クレアール　無断複製・流布を禁じます

第8問

問題3

(1)第1四半期の場合、各四半期ダミーは、$Q_{IIt} = Q_{IIIt} = Q_{IVt} = 0$となるため、予測式は、

$C_t = \alpha + \beta Yd_t + \gamma Trend_t$ となる。そして各パラメータの推定値を代入すれば、

$C_t = 5.416\times10^4 + 1.682\times10^{-1} Yd_t + 1.387\times10^2 Trend_t$

となる。

(2)2014年第II四半期のC_tの予想値は、Q_{IIt}=1　またそれ以外のダミー変数は0になることから、$C_t = \alpha + \beta Yd_t + \gamma Trend_t + \delta_2$となり、各変数と、それに対応するパラメータの推定値を代入すると、

$$C_t = 5.416 \times 10^4 + 1.682 \times 10^{-1} \times 85000 + 1.387 \times 10^2 \times 81 - 4.828 \times 10^3 \times 1$$

となり、これを計算すると、$C_t = 74863.7 \cong 74864$ となる。

(3)第II四半期のC_tの予想値は、$C_{IIt} = \alpha + \beta Yd_t + \gamma Trend_t + \delta_2$、

第III四半期のC_tの予想値は、$C_{IIIt} = \alpha + \beta Yd_t + \gamma Trend_t + \delta_3$、

となることから、この差の絶対値は、

$$|C_{IIt} - C_{IIIt}| = |\alpha + \beta Yd_t + \gamma Trend_t + \delta_2 - (\alpha + \beta Yd_t + \gamma Trend_t + \delta_3)| = |\delta_2 - \delta_3|$$

となる。これを計算すると、$|\delta_2 - \delta_3| = |-4.828 \times 10^3 - (-4.673 \times 10^2)| = 4360.7 \cong 4361$ となる。

(4) ダミー変数を3つにしているのは、この線形回帰分析では、定数項αが第1四半期の時の基準値にし、ダミー変数は、それ以外の時期の値が第1四半期の値との差を表しているからである。たとえば、第2四半期の場合は、定数項αに加えてδ_2の値の効果が加わることで、第2四半期の定数項の値が決定する形式になっている。

そのため、この式に新たなダミー変数、第1四半期のダミー変数を加えて推定して回帰分析をしてしまうと、定数項と第1四半期のダミー変数との間に違いが存在しないことになり、結果、多重共線性の問題が発生することになる。よって第1四半期でもどの木のものでもいいのであるが、定数項を加えて回帰分析を行う際は、どれかのダミー変数を最低1つは除いて回帰分析を行う必要がある。

もし第1四半期のダミーを加えたいのであれば、定数項部分を除去して、定数項なしのダミー変数をすべて用いた回帰分析を行えばよい。

クレアール　無断複製・流布を禁じます