練習問題の詳解

1 章

1.1　練習問題

国勢調査の結果を総務省統計局のホームページから e-Stat を利用すると，次のようなデータが得られる。

	全国人口（千人）			65 歳以上人口（千人）		65 歳以上人口割合	
	男	女	総数	男	女	男	女
平成 2 年	60,697	62,914	123,611	5,988	8,907	9.9%	14.2%
平成 7 年	61,574	63,996	125,570	7,504	10,757	12.2%	16.8%
平成 12 年	62,111	64,815	126,926	9,222	12,783	14.8%	19.7%
平成 17 年	62,349	65,419	127,768	10,875	14,797	17.4%	22.6%
平成 22 年	62,328	65,730	128,057	12,470	16,775	20.0%	25.5%

この結果をみると，これまで人口は男女とも増加傾向にあるが，増加の割合はだんだん下がってきている。一方，65 歳以上人口の割合は，男女とも急激に増えていることがわかる。

2 章

2.1　練習問題

表 2.1 のデータにおいて，男性 10 人の年齢は

47, 46, 50, 42, 55, 50, 48, 46, 48, 52

であり，年齢の合計は 484 となり，これを 10 で割ると平均値は 48.4（歳）となる。一方，中央値を求めるために，10 人のデータを小さい方から順に並べ替えると

42, 46, 46, 47, 48, 48, 50, 50, 52, 55

となる。真ん中にある 2 つの数値はどちらも 48 であるから，中央値は 48（歳）となる。

2.2 練習問題

表 2.1 のデータにおいて，男性 10 人の年齢の平均は 48.4（歳）であるから，10 人の年齢の偏差を小さい方から並べると

－6.4,－2.4, －2.4, －1.4, －0.4, －0.4, 1.6, 1.6, 3.6, 6.6

となる。この偏差の 2 乗の和は 116.4 となり，これを 10 で割ると分散が 11.64 となる。標準偏差は $\sqrt{11.64}$ =3.41 となる。第 1 四分位数 Q_1 は，最初の 5 つの中央値であるから，$Q_1 = 46$ となる。同様に，第 3 四分位数 Q_3=50 となるので，四分位範囲は $Q_3 - Q_1 = 4$ と

なる。

第３章

３．１　練習問題

ウィルス保有者４人に対して、検出率が 50%の検査器を用いて検査を実施したときに陽性反応が出る人数Xの確率分布は二項分布$B\left(4,\frac{1}{2}\right)$である。二項分布$B\left(4,\frac{1}{2}\right)$の実現値は$x=0,1,2,3,4$であり、各実現値$x$に対する確率は次式で計算される。

$$P(X=x)=\ {}_4C_x\left(\frac{1}{2}\right)^x\left(1-\frac{1}{2}\right)^{4-x}=\ {}_4C_x\frac{1}{2^4}$$

X	0	1	2	3	4
P(X = x)	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

第４章

４．１　練習問題

検査器の検出率を$100p\%$（$0<p<1$）とすると、帰無仮説は「検出率は 50%（$p=0.5$）」、対立仮説は「検出率は 50%ではない（$p\neq 0.5$）」の両側検定である。帰無仮説が正しいとしたとき、ウィルス保有者 4 人に対して陽性反応が出る人数Xは二項分布$B(4,0.5)$であることから、p値を計算する。

$$P(X\leq 1)=P(\mathrm{X}=0)+P(X=1)=\ {}_4C_0 0.5^0(1-0.5)^{4-0}+\ {}_4C_1 0.5^1(1-0.5)^{4-1}=0.3125$$

$$P(X\geq 1)=\ 1-P(X=0)=1-{}_4C_0 0.5^0(1-0.5)^{4-0}=0.9375$$

より、p値$=2\times\min(P(X\leq 1),P(X\geq 1))=0.625$。これより、$p$値は有意水準の 0.05 より大きいため、帰無仮説を棄却できない。よって、帰無仮説の「検出率は 50%（$p=0.5$）」を受容する。

４．２　練習問題

検査器の検出率を$100p\%$（$0<p<1$）とすると、帰無仮説は「検出率は 50%（$p=0.5$）」、対立仮説は「検出率は 50%ではない（$p\neq 0.5$）」の両側検定である。帰無仮説が正しいとしたとき、ウィルス保有者 40 人に対して陽性反応が出る人数Xは二項分布$B(40,0.5)$であることからp値を計算できるが、$n=40$と標本サイズが大きいため、Excel を用いて正規近似によりp値を計算する。

$$2*\mathrm{NORMDIST}\big(-\mathrm{ABS}(10-40*0.5),0,\mathrm{SQRT}\big(40*0.5*(1-0.5)\big),1\big)$$

より、p値は約 0.0016 であり、有意水準の 0.05 より小さいので、帰無仮説の「検出率は 50%（$p=0.5$）」を棄却する。よって、対立仮説の「検出率は 50%ではない（$p\neq 0.5$）」が適切であると判断する。

４．３　練習問題

例題４．５より、新薬の有効率p（$0 < p < 1$）の点推定値は$\hat{p} = 0.45$、標準偏差は$S.E. \approx 0.05$である。信頼度が 99%より$K = 2.58$である。これより、有効率pの信頼区間は、

$$\hat{p} - K \times S.E. < p < \hat{p} + K \times S.E.$$
$$0.45 - 2.58 \times 0.05 < p < 0.45 + 2.58 \times 0.05$$
$$0.322 < p < 0.578$$

第５章

５．１　練習問題

⑴ $P(Z \geq 1.65) = 1 - P(Z \leq 1.65)$であり、標準正規分布表から$P(Z \leq 1.65) = 0.9505$である。これより、$P(Z \geq 1.65) = 1 - 0.9505 = 0.0495$

⑵ $P(-1.96 \leq Z \leq 1.96) = P(Z \leq 1.96) - P(Z \leq -1.96)$、標準正規分布表から$P(Z \leq 1.96) = 0.9750$、$P(Z \leq -1.96) = 1 - P(Z \leq 1.96) = 0.0250$である。これより、$P(-1.96 \leq Z \leq 1.96) = 0.9750 - 0.0250 = 0.9500$

５．２　練習問題

正規分布$N(50,10^2)$に従う確率関数Xを標準化した確率変数$Z = \frac{X-50}{10}$は標準正規分布$N(0,1^2)$に従うことを用いる。

⑴ $P(X \geq 66.5) = P\left(\frac{X-10}{10} \geq \frac{66.5-50}{10}\right) = P(Z \geq 1.65)$であり、練習問題５．１の⑴より$P(Z \geq 1.65) = 0.0495$である。したがって、$P(X \geq 66.5) = 0.0495$

⑵ $P(30.4 \leq X \leq 69.6) = P\left(\frac{30.4-50}{10} \leq \frac{X-50}{10} \leq \frac{69.6-50}{10}\right) = P(-1.96 \leq Z \leq 1.96)$であり、練習問題５．１の⑵より、$P(-1.96 \leq Z \leq 1.96) = 0.9500$である。したがって、$P(30.4 \leq X \leq 69.6) = 0.9500$

５．３　練習問題

標準正規分布$N(0,1^2)$に従う確率関数Zに対する確率$P(Z \leq c)$を計算する Excel 関数であるNORMSDIST(c)を用いて計算する。

⑴ $P(Z \geq 1.65) = 1 - P(Z \leq 1.65)$である。$1 - \mathrm{NORMSDIST}(1.65)$の計算結果から、$P(Z \geq 1.65) \approx 0.0495$

⑵ $P(-1.96 \leq Z \leq 1.96) = P(Z \leq 1.96) - P(Z \leq -1.96)$である。$\mathrm{NORMSDIST}(1.96) - \mathrm{NORMSDIST}(-1.96)$の計算結果から、$P(-1.96 \leq Z \leq 1.96) \approx 0.9500$

５．４　練習問題

正規分布$N(50,10^2)$に従う確率関数Xに対する確率$P(X \leq c)$を計算する Excel 関数である$\mathrm{NORMDIST}(c,50,10,1)$を用いて計算する。

(1) $P(X \geq 66.5) = 1 - P(X \leq 66.5)$である。$1 - \mathrm{NORMDIST}(66.5,50,10,1)$の計算結果から、$P(X \geq 66.5) \approx 0.0495$

(2) $P(30.4 \leq X \leq 69.6) = P(X \leq 69.6) - P(X \leq 30.4)$である。$\mathrm{NORMDIST}(69.6,50,10,1) - \mathrm{NORMDIST}(30.4,50,10,1)$ の計算結果から、$P(30.4 \leq X \leq 69.6) \approx 0.9500$

第６章

６．１　練習問題

標本平均の分布の性質から，標本平均の平均は母集団の平均に等しいので 50 である。また，標本平均の分散は 100/400=0.25 となる。よって，標本平均の分布は，平均 50，分散 0.25 の正規分布となる。

６．２　練習問題

$\bar{X} = 50$，$n = 400$，$\sigma^2 = 10^2$ であるから，

$$\bar{X} - 1.96\sigma/\sqrt{n} = 50 - 1.96 \times 10/20 = 49.02$$

$$\bar{X} + 1.96\sigma/\sqrt{n} = 50 + 1.96 \times 10/20 = 50.98$$

となる。よって，95％信頼区間は，49.02 以上 50.98 以下となる。

第７章

７．１　練習問題

介入群の平均は 2，対照群の平均は 0 である。よって，平均の差⊿の推定値は 2 である。それぞれの群において，各値と平均の差の 2 乗和を計算すると，次のようになる。

介入群　$0^2 + (-1)^2 + 1^2 + (-1)^2 + 1^2 + 0^2 = 4$

対照群　$(-3)^2 + 1^2 + 0^2 + 2^2 + 1^2 + (-1)^2 = 16$

となるので，共通する分散の推定は

$$\hat{\sigma}^2 = \frac{4+16}{10} = 2$$

となる。

７．２　練習問題

練習問題 7.1 の結果を用いると，検定統計量の値は

$$T=\frac{2-0}{\sqrt{\left(\frac{1}{6}+\frac{1}{6}\right)\times 2}}=\sqrt{6}=2.45$$

となる。帰無仮説のもとで, T の分布は自由度 10 の t 分布となるので, 上側 2.5%点は 2.228 であるから，有意水準 5%で帰無仮説を棄却できる。よって，このデータから有意水準 5% で，介入の効果があることが示せる。

第 8 章

８．１　練習問題

介入群と対照群の拡張期血圧を合わせて，小さい方から順に並べると，次のようになる。

血圧値	60,	65,	75,	77,	80,	95
順位	1,	2,	3,	4,	5,	6

このことから，介入群の順位和は

2+1+3=6

となる。

８．２　練習問題

帰無仮説を仮定すると，それぞれの順位の組み合わせは同じ確率で起こる。練習問題 8.1 の設定では，全体が 6 人で介入群は 3 人であるから，介入群の順位としてとり得る組み合わせは

（１，２，３），（１，２，４），（１，２，５），（１，２，６），（１，３，４），
（１，３，５），（１，３，６），（１，４，５），（１，４，６），（１，５，６），
（２，３，４），（２，３，５），（２，３，６），（２，４，５），（２，４，６），
（２，５，６），（３，４，５），（３，４，６），（３，５，６），（４，５，６）

の 20 通りある。この中で順位和が 6 以下になるのは 1 つしかない。よって，順位和が 6 以下となる確率は 5%となり，その 2 倍は 10%である。よって，有意水準 5%で介入群と対照群の間に分布の違いがあることを示せない。ただし，この方法ではどんな結果が出ても有意水準 5%で違いを示すことはできない。それぞれの群の対象者をもっと増やす必要がある。

８．３　練習問題

介入群と対照群の測定値を合わせて，小さい方から順に並べ替えると，次のようになる。

－3　－１　０　１　１　１　１　２　２　２　３　３

このとき，それぞれの順位を求めると，

１　２　３　５．５　５．５　５．５　５．５　９　９　９　１１．５　１１．５

となる。このとき，介入群の６人の順位は

　　９　５．５　１１．５　５．５　１１．５　９

となる。このとき，6 人の順位の和は５２となる。

第９章

９．１　練習問題

m=80, n=100，x=50, y=30 として計算すると，カイ 2 乗統計量の値は,

$$\frac{(80+100)(50\times100-30\times80)^2}{80\times100\times(50+30)(80+100-50-30)}=19.01$$

となる。この値は，自由度 1 のカイ 2 乗分布の上側 5%点である 3.84 よりも大きくなるので，有意水準 5%で遺伝子タイプの違いによって薬の効果に違いがあることが示せる。

第 10 章

１０．１　練習問題

プラセボ群と対象薬群の眼圧の差を 2mmHg と仮定することより，Δ=2 とする。両軍の標準偏差が２．５ｍｍHg よりσ=2.5 とし，α＝0.05, β=1-0.8=0.2 とする。標準正規分布を上側%点を調べると，Z(0.025)=1.96, Z(0.2)=0.84 である。これらのことから

$$n=2\left(\frac{1.96+0.84}{2/2.5}\right)^2=24.5$$

であるから，2 つの群の標本サイズを 25 以上とすればよい。

１０．２　練習問題

薬 A および薬 B での改善が認められる割合を 35%と 40%と設定することから，p=0.35, q=0.4 とする。有意水準 5%よりα=0.05, 検出力 80%より，β=1-0.8=0.2 とする。このとき，

$$n=\left(\frac{1.96\sqrt{2\times0.375\times(1-0.375)}+0.84\sqrt{0.35\times(1-0.35)+0.4\times(1-0.4)}}{0.35-0.4}\right)^2=1470.9$$

となるので，プラセボ群と対象薬群それぞれ 1471 人とすればよい。

第１１章

１１．１　練習問題

母親の妊娠前の体重をx、出生時体重をyとすると、$\bar{x}=49.68$、$s_x^2=32.1$、また、例題の解

説から$\bar{y} = 3003$、$s_y^2 = 22195.8$、そして、2 変量の共分散が$s_{xy} = 258.6$と求められることから，相関係数はr_{xy}=258.6/$\sqrt{32.1 * 22195.8}$=0.31 と算出される。この相関係数は、妊娠週数と出生体重との相関係数 0.72 より小さく、その絶対値は 0 により近い。したがって、母親の妊娠前の体重よりも妊娠週数の方が出生体重と直線的な関係がありそうなことが分かる。

１１．２　練習問題

妊娠週数(x)と出生時体重(y)をそれぞれの平均で 2 群に分けたときの分割表は

	$y = 0$	$y = 1$	合計
$x = 0$	7	2	9
$x = 1$	5	11	16
合計	12	13	25

とかける。したがって、求める相関係数は$r_{xy} = (7 * 11 - 2 * 5)/\sqrt{9 * 16 * 12 * 13} = 0.45$と算出できる。これは、平均で分割する前の相関係数 0.72 よりも小さく、絶対値は 0 に近い。このように、一般的には連続変数を 2 値化すると元の相関係数よりも 0 に近くなる傾向がある。

第１２章

１２．１　練習問題

診断後 4 週の重度群における投薬群と反応の２ x ２分割表は

	異常	正常	合計
標準薬	54	46	100
新薬	15	75	90
合計	69	121	190

とかける。したがって、ロジスティック曲線のパラメータは

$$a = \log\frac{46}{54} = -0.16, \qquad b = \log\frac{\frac{75}{15}}{\frac{46}{54}} = 1.77$$

と求められる。よって、標準薬に対する新薬のオッズ比は$\exp(b) = 5.87$となる。

１２．２　練習問題

診断後 2 週の重度群における投薬群と反応の２ x ２分割表は

	異常	正常	合計
標準薬	72	28	100
新薬	45	45	90
合計	117	73	190

とかける。したがって、ロジスティック曲線のパラメータは

$$a = \log\frac{28}{72} = -0.94, \qquad b = \log\frac{\frac{45}{45}}{\frac{28}{72}} = 0.94$$

と求められる。よって、標準薬に対する新薬のオッズ比は$\exp(b) = 2.57$となる。これは後診断後 4 週のオッズ比 5.87 と比べて小さい。逆に言えば、新薬の効果は時間経過とともに大きくなる。

第１３章

１３．１　練習問題

時刻$t = 3$における２群の死亡数に関する表１３．６に対応する２ｘ２分割表は$d_A = 0$、$d_B = 1$、$N_A = 7$、$N_B = 5$、から

	死亡数	生存数	合計
Ａ群	0	7	7
Ｂ群	1	4	5
合計	1.	11	12.

とかける。したがって、$e_A(3) = \frac{7}{12} = 0.583$、$v_A(3) = \frac{7*5}{12^2}$ =0.243。

１３．２　練習問題

表 13.7 の B 群の生存確率は次のように求められる。

死亡時刻	1	2	3	4
生存確率	0.727	0.455	0.364	0.091

第１４章

１４．１　練習問題

あらためて、$U = e_X$、$V = e_Y$とおくと、$\bar{U} = 0$および$\bar{V} = 0$、また、それぞれの分散$s_U^2 = 5.51, s_V^2 = 22.21$と、共分散$s_{UV} = -2.513$から，相関係数は$r_{UV}$=$-\frac{2.513}{\sqrt{5.51*22.21}} = -0.23$と算出される。

第１５章

１５．１　練習問題

図１５．５で推定された回帰係数を用いると、$\theta_1(t) = a_1 + b_1 t = -1.03 + 0.30t$とかける。したがって、年齢が高くなるにつれて（少女を基準とした少年の）性差は正の方向に大きくなる。実際、図１５．５からも性差が大きくなり、２直線が徐々に離れていく様子が読み取れる。