	A	B	C	D	E	F	G	H	I	J
1	年齢調整死亡率(直接法)及びSMRの計算									
2										
3				入力データ			年齢調整死亡率(直接法)の計算			SMRの計算
4	年齢階級	昭和60年ﾓﾃﾞﾙ人口（千人）	全国人口(平成7年)	全国死亡数(平成7年)	福岡県推計人口(平成7年)	福岡県死亡数(平成7年)	福岡県死亡率(人口千対)	重み	死亡率×重み	期待死亡数
5		①	②	③	④	⑤	⑥	⑦	⑧	⑨
6							⑤÷④×1,000	①÷①の合計	⑥×⑦	③÷②×④
7	0～4歳	8180	5949623	7040	236167	309	=F7/E7*1000	=B7/$B$25	=G7*H7	=D7/C7*E7
8	5～9歳	8338	6493110	1235	264385	46	=F8/E8*1000	=B8/$B$25	=G8*H8	=D8/C8*E8
9	10～14歳	8497	7424703	1184	310225	56	=F9/E9*1000	=B9/$B$25	=G9*H9	=D9/C9*E9
10	15～19歳	8655	8491929	3362	361329	118	=F10/E10*1000	=B10/$B$25	=G10*H10	=D10/C10*E10
11	20～24歳	8814	9765295	5087	395386	202	=F11/E11*1000	=B11/$B$25	=G11*H11	=D11/C11*E11
12	25～29歳	8972	8614403	4596	314347	161	=F12/E12*1000	=B12/$B$25	=G12*H12	=D12/C12*E12
13	30～34歳	9130	7968686	5129	294244	217	=F13/E13*1000	=B13/$B$25	=G13*H13	=D13/C13*E13
14	35～39歳	9289	7709028	6839	306913	263	=F14/E14*1000	=B14/$B$25	=G14*H14	=D14/C14*E14
15	40～44歳	9400	8916937	12814	364869	530	=F15/E15*1000	=B15/$B$25	=G15*H15	=D15/C15*E15
16	45～49歳	8651	10544944	24136	412540	991	=F16/E16*1000	=B16/$B$25	=G16*H16	=D16/C16*E16
17	50～54歳	7616	8867530	32946	325448	1275	=F17/E17*1000	=B17/$B$25	=G17*H17	=D17/C17*E17
18	55～59歳	6581	7912482	44732	296435	1750	=F18/E18*1000	=B18/$B$25	=G18*H18	=D18/C18*E18
19	60～64歳	5546	7445934	68310	288126	2841	=F19/E19*1000	=B19/$B$25	=G19*H19	=D19/C19*E19
20	65～69歳	4511	6373007	89089	249979	3718	=F20/E20*1000	=B20/$B$25	=G20*H20	=D20/C20*E20
21	70～74歳	3476	4674557	102443	187725	4162	=F21/E21*1000	=B21/$B$25	=G21*H21	=D21/C21*E21
22	75～79歳	2441	3276736	125428	128807	4915	=F22/E22*1000	=B22/$B$25	=G22*H22	=D22/C22*E22
23	80～84歳	1406	2293864	157863	92226	6187	=F23/E23*1000	=B23/$B$25	=G23*H23	=D23/C23*E23
24	85歳以上	784	1576179	229269	67300	9416	=F24/E24*1000	=B24/$B$25	=G24*H24	=D24/C24*E24
25	合計	=SUM(B7:B24)	=SUM(C7:C24)	=SUM(D7:D24)	=SUM(E7:E24)	=SUM(F7:F24)	=F25/E25*1000	=SUM(H7:H24)	=SUM(I7:I24)	=SUM(J7:J24)
26							↑		↑	
27							粗死亡率		年齢調整死亡率(直接法)	
28			SMR[⑤の合計÷⑨の合計]＝	=F25/J25						
30			SMRの分散＝	=F25/J25^2						
32			SMRの95%信頼区間＝	=D28-1.96*SQRT(D30)		～	=D28+1.96*SQRT(D30)			
34			正規近似によるZ＝	=ABS(F25-J25)/SQRT(J25)						
36			正規近似による検定結果＝	=IF(D34>2.576,"p<0.01",IF(D34>1.96,"p<0.05","n.s."))						
38			ポアソン分布によるp＝	=IF(J25<5,IF(F25<J25,POISSON(F25,J25,TRUE),1-POISSON(F25-1,J25,TRUE)),"")						
40			ポアソン分布による検定結果＝	=IF(D38<>"",IF(D38<0.005,"p<0.01",IF(D38<0.025,"p<0.05","n.s.")),"")						

【研修における例題回答】

人口動態解析参考資料
医統計テキスト（西村書店）から抜粋

【二項分布の特徴】

★9-A)これまで解説した正規，t，F分布はxが連続変数であったが，二項分布は下図の例（サイコロの目の出る確率）のように，x（0〜n）は整数しかとれないため，離散型の分布の代表である。しかし，nが大きくなると，確率変数$\frac{x}{n}$は連続型のように取り扱われることもある。

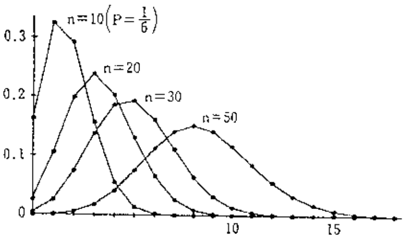


図9-1　サイコロの目の出る確率

表9　n＝10の確率

x	${}_nC_x p^x q^{n-x}$
0	0.161506
1	0.323011
2	0.290710
3	0.155045
4	0.054266
5	0.013024
6	0.002171
7	0.000248
8	0.000019
9	0.000001
10	0.000000

▶9-B)二項確率を$f(x)$とすると，コンピュータでの計算は簡単である。

$$f(0)=(1-P)^n,\ f(i+1)=\frac{n-i}{i+1}\cdot\frac{P}{1-P}\cdot f(i)\,[i=0\sim n-1]$$

ただし，n，Pがともに大きいと，計算精度が問題となる。手計算ではnが大きいと階乗！の計算は無理なので，二項係数${}_nC_x$の数値表が作られているほか，他の分布へのさまざまな変換が試みられてきた。

① 正規分布への近似：二項分布で$P=Q=\frac{1}{2}$でかつ$n\to\infty$の連続型は正規分布となる。そこで，$nP\geqq5$（かつ$nQ\geqq5$）の時，二項分布を正規分布に近似させてよいとされている。

② 逆正弦変換：多くの比率を扱う場合，θ（ラジアン）$=\sin^{-1}\sqrt{P}$と変換すると分散は$\frac{1}{4n}$となり，確率とは無関係になる（詳しくは専門書）。

③ $nP<5$の場合については，詳しく後述する (p. 76〜) 。

① $nP\geqq5(nQ\geqq5)$の場合 ──→ 正規近似式による (1)
② $nP<5$ $\begin{cases} n<50\ \text{の場合} \longrightarrow \text{F検定式による (2)} \\ n\geqq50(P\to0)\ \text{の場合} \longrightarrow \text{ポアソン分布への近似 (3)} \end{cases}$

(1) $nP\geqq5(nQ\geqq5)$の場合 ──→ 正規近似式による

(IX-3)式を用いてもよいが，離散型である二項分布を連続である正規分布に近似させると，多少無理が生じてしまう。したがって，以下のようにイェーツ(Yates)の修正項$-\frac{1}{2n}$によって連続修正する。

H_0：$p_s=P$のもとで，

$$z_0=\frac{|p_s-P|-\frac{1}{2n}}{\sqrt{\frac{P(1-P)}{n}}}\sim N(0,\ 1^2) \qquad \cdots\cdots(\text{IX-6})$$

※厳密にいえば，(IX-4)式にもイェーツの修正項$\pm\frac{1}{2n}$が必要であるが，nが十分に大きいという仮定のもとでは，0に近いため修正しなくてもよい。

ところで，一般的に標本では陽性者x人の数は知られており，$x=np_s$であるから，(IX-6)式の分母・分子にnを乗じて，

$$z_0=\frac{|x-nP|-0.5}{\sqrt{nP(1-P)}}\sim N(0,\ 1^2) \qquad \cdots\cdots(\text{IX-7})$$

となる。期待値$nP\geqq5$は計算済みなので，(IX-7)式の方が計算しやすい。

【例9-2】昭和58年の「学校保健統計」によると，小学校6年生の永久歯罹患者率は，X県全体で87.4%であった。A小学校では6年生54人中，罹患者は30人(55.6%)であったという。$\alpha=0.05$として検定しなさい。

① $nP=54\cdot0.874=47.196>5$であるから，正規近似式を用いてよい。

② H_0：標本比率＝母比率のもとで，　　$[n(1-P)=6.804>5]$

$$z_0=\frac{|30-47.196|-0.5}{\sqrt{47.196\cdot0.126}}=6.847>z\left(\frac{0.001}{2}\right)=3.291$$

したがってA小学校の罹患者率は，X県より有意に($p<0.001$)低い。

(2) $nP<5$かつ$n<50$の場合 ──→ F検定式による

2．区間推定で述べたように標本数nが小さい場合，信頼区間が広くなりすぎて意味がない。しかし，検定では実例が多いにもかかわらず，F検定式による場合を明記してあるテキストは少ない。H_0：標本比率＝母比率として，図9-2，3のように，陽性者が$\left\{\begin{matrix}\text{x人以下}\\ \text{x人以上}\end{matrix}\right\}$の片側確率を2倍して評価する。

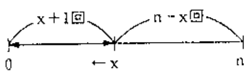


図9-2　陽性者がx人以下の場合

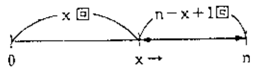


図9-3　陽性者がx人以上の場合

① 陽性者がx人以下の場合，$n_1=2(x+1)$，$n_2=2(n-x)$であるから，

$$F_L=\frac{n_2P}{n_1Q}>F^{n_1}_{n_2}\left(\frac{\alpha}{2}\right)\ \text{なら}H_0\text{を棄却する。} \qquad \cdots\cdots(\text{IX-8})$$

② 陽性者がx人以上の場合，$n_1=2(n-x+1)$，$n_2=2x$であるから，

$$F_U=\frac{n_2Q}{n_1P}>F^{n_1}_{n_2}\left(\frac{\alpha}{2}\right)\ \text{なら}H_0\text{を棄却する。} \qquad \cdots\cdots(\text{IX-9})$$

【例9-3】非流行年の風疹抗体陽性率は，「伝染病流行予測調査事業」によって6歳女児で15.0%と予想された。B小学校1年女子25人中抗体陽性者が，

a) 1人
b) 8人 } であった場合について，それぞれ$\alpha=0.05$として検定しなさい。

① $nP=25\cdot0.15=3.75<5$なので，正規近似式を使うことができない。

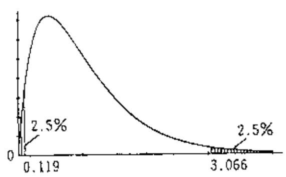


図9-4　F^{4}_{48}の分布

② H_0：$p_s=P$，$Q=1-P=0.85$

a)の場合，$n_1=2(1+1)=4$，$n_2=2(25-1)=48$から，

$$F_L=\frac{48\cdot0.15}{4\cdot0.85}=2.118<F^{4}_{48}\left(\frac{0.05}{2}\right)=3.066$$

したがって，H_0を棄却できず，陽性者1人は有意に少ないとはいえない。

b)の場合，$n_1=2(25-8+1)=36$，$n_2=2\cdot8=16$

$$F_U=\frac{16\cdot0.85}{36\cdot0.15}=2.519<F^{36}_{16}\left(\frac{0.05}{2}\right)=2.528$$

したがって，H_0を棄却できず，陽性者8人は有意に多いとはいえない。

③ 念のため正規近似式でも検討してみる。

a)の場合，$z_0=1.260<1.96$で有意差なしと判定は変わらない。

b)の場合，$z_0=\frac{|8-3.75|-0.5}{\sqrt{3.75\cdot0.85}}=2.100>z\left(\frac{0.05}{2}\right)=1.96$

つまり，うっかり正規近似式を用いると，誤った判定を下すことになる。

(3) $nP<5$かつ$n\geqq50$の場合 ──→ ポアソン分布に近似させる

たとえば，市町村における死因別死亡率を考えると，nが大きいのに対してPは0とみなせるくらい，非常に小さくなっている。そこでまれな事象の確率分布である，ポアソン分布(Poisson distribution)に近似させる。

★補足9-C

ポアソン分布の確率$P(x)$は，平均$m=nP$として，

$$P(x)=\frac{e^{-m}\cdot m^x}{x!} \qquad \cdots\cdots(\text{IX-10})$$

($x=0,\ 1,\ 2,\ \cdots\cdots,\ n$)［eは自然対数の底＝2.71828……］

として求められる。ただし，両側でなく片側の確率である。

【例9-4】昭和54〜63年の10年間において，全国男性の白血病死亡率は人口10万対49.7であった。間接法を用いて年齢調整を行うと，C町の男性（1万人）の期待死亡数は4.97人であったとする。C町の10年間の観察死亡数が，

a) 1人である場合
b) 10人である場合 } について，それぞれ$\alpha=0.05$として検定しなさい。

① 標準化死亡比(SMR)$=\frac{観察値}{期待値}\times 100$ であるから，▶補足9-D

a)では $SMR=\frac{1}{4.97}\times 100=20.1$，b)では $SMR=201.2$ となる。

② $m=nP=4.97<5$ かつ $n\geqq 50$ なので，(IX-10)式を用いる。a)の場合，$P(1)$ を求めただけでは下側確率にならないことに注意する。

$$P(0)=\frac{e^{-4.97}\cdot 4.97^0}{0!}=0.006943 \qquad P(1)=\frac{P(0)\cdot 4.97}{1}=0.034507$$

③ 両側確率に直すため2倍する。つまり求める確率 p は，

$p=2\cdot\{P(0)+P(1)\}=2\cdot 0.04145=0.0829$

したがって H_0 は棄却できず，全国に比べて SMR は低いとはいえない。

b)の場合，図9-5のように9人以下の確率 $\sum_{x=0}^{9}P(x)$ を求め，1から引く。

つまり，$p=2\cdot\{1-\sum_{x=0}^{9}P(x)\}=2\cdot 0.030753=0.061506$

したがって H_0 は棄却できず，全国に比べて SMR は高いとはいえない。

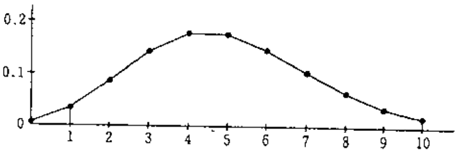


図9-5　ポアソン分布の確率(m=4.97の場合)

④ ちなみに正規近似式で検定すると，

a)の場合，$z_0=1.557<1.96$ で有意差なしと判定は変わらない。

b)の場合，$z_0=\frac{|10-4.97|-0.5}{\sqrt{4.97(1-0.000497)}}=2.032>z\left(\frac{0.05}{2}\right)=1.96$

となり，正規近似式を用いた場合，誤った判断を下すことになる。

※ただし，SMR の判定に関しては，片側検定でよいとする考え方もある。

【ポアソン分布の特徴】

★9-C)ポアソン分布は，非常にまれな事象の確率で，離散型分布のひとつである。「馬にケラレテ死ぬ確率」が有名であるが，医学的データにも例が多い。二項分布で平均 $m=nP$ を一定にし，$n\to\infty$，$P\to 0$ の極限型がポアソン分布であるが，$nP<5$ かつ $n\geqq 50$の場合，正規分布に近似できる。

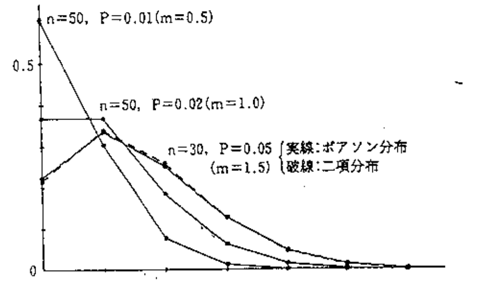


図9-6　二項分布とポアソン分布との関係

医学的データで，ポアソン分布のよくあてはまる例として，工場や学校での負傷発生数，市町村での1年間のがん死亡数，先天異常の発生数，飲料水検査での大腸菌数を予測する最確数(MPN)等が知られている。

▶9-D)$nP\geqq 5$ の場合，SMR(Standardized Mortality Ratio)の検定については，観察値$=O$，期待値$=E=nP$ とおいて，

① 正規近似式 $z_0=\frac{|O-E|-0.5}{\sqrt{E(1-P)}} \sim N(0,\ 1^2)$ ……(IX-11)

② $P\to 0$，$1-P\to 1$ から，(IX-11)式を2乗する。

$$\chi_0^2=\frac{(|O-E|-0.5)^2}{E} \sim \chi^2 \quad \cdots\cdots(\text{IX-12})$$

自由度1のカイ2乗分布(p.84)に従うとみなしてよい。

つまり，ポアソン分布でもカイ2乗検定でも母比率Pは必要がない。

③ F検定式において，片方の自由度→∞と考えて検定してもよい。

【例9-4】b)は $n_1=19982\to\infty$，$n_2=20$ で(IX-9)式から，

$F_u=\frac{20\cdot(1-P)}{19982\cdot P}=2.013<F_{\infty}^{20}\left(\frac{0.05}{2}\right)=2.085$ と有意性は変わらない。

ただし，期待値 $F_j<5$ の f_j をまとめるためには，(XI-2)式の方がよい。

【例11-3】職業別に肥満度を調べた結果は，表11-7のとおりであった。肥満者の割合は職業によって差が認められるか，$\alpha=0.05$ として検定しなさい。

表11-7　職業別肥満度の分布

	事務職	商　業	農林業	鉱工業	計
肥満度　20%以上	150	70	60	100	380
20%未満	550	230	340	500	1620
計	700	300	400	600	2000
肥満者の割合(%)	21.4	23.3	15.0	16.7	19.0

H_0：肥満者の割合は職業別に差はない。自由度は $l-1=3$ である。

$$\chi_0^2=\frac{2000^2}{380\cdot 1620}\left(\frac{150^2}{700}+\frac{70^2}{300}+\frac{60^2}{400}+\frac{100^2}{600}-\frac{380^2}{2000}\right)=12.624$$

$$>\chi_3^2(0.01)=11.345$$

したがって H_0 は棄却され，肥満者の割合は，職業別に有意差($p<0.01$)が認められる。しかし，このままではどこに差があるかは明らかではない。図11-2に示すように2×2表の組み合わせの結果をまとめておくとよい。

▶補足11-E

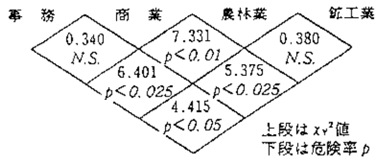


図11-2　2×2表の組合せ結果

(5) カイ2乗-傾向性の検定 ←―― l が順序尺度の場合

解説しているテキストは少ない。例をあげて手順を説明する。

【例11-4】年齢階級別の陽性率を調査したら，表11-8のような結果を得たとする。年齢階級が進むにつれて，陽性率が上昇していることを証明したい。

表11-8　年齢階級別陽性率(%)

	20～29歳	30～39歳	40～49歳	計
陽性　A_j	38	50	30	118　(A)
陰性　m_j-A_j	162	150	70	382　(N−A)
計　m_j	200	200	100	500　(N)
陽性率 $p_j=m_j/A_j$	19.0	25.0	30.0	23.6(p=A/N)
得　点　x_j	−1	0	1	――

この場合，横の項目 l は年齢階級順に並んでいるので，自由に入れ替えることができない。そこで，(XI-4)式だけではなく，特別なカイ2乗検定式を用いる。

① 全体の χ_0^2 値を(XI-4)式から求める。

$$\chi_0^2=\frac{500^2}{118\cdot 382}\left(\frac{38^2}{200}+\frac{50^2}{200}+\frac{30^2}{100}-\frac{118^2}{500}\right)=4.836$$ df: R-1

※ここで $\chi_0^2<\chi_2^2(0.05)=5.991$ から，直ちに NS と判定しないこと。

② 回帰による χ_1^2 値を次式から求める。(df=1)表11-8から，

$$\left.\begin{array}{l}\sum A_j=A,\\ \sum m_j=N,\\ x_j=-1,\ 0,\ 1\end{array}\right\}\text{とおくと，}\begin{array}{l}\sum A_jx_j=38\cdot(-1)+30\cdot 1=-8\\ \sum m_jx_j=200\cdot(-1)+100\cdot 1=-100\\ \sum m_jx_j^2=200\cdot 1+100\cdot 1=300\end{array}$$ df: 1

$$\chi_1^2=\frac{N(N\sum A_jx_j-A\sum m_jx_j)^2}{A(N-A)\{N\sum m_jx_j^2-(\sum m_jx_j)^2\}}=\frac{500\{500\cdot(-8)-118\cdot(-100)\}^2}{118\cdot 382\{500\cdot 300-(-100)^2\}}$$

$=4.820>\chi_1^2(0.05)=3.841$($H_0$：回帰による傾向が認められない，と考えればこのままで有意差ありと判定してよいが，③の結果も考慮する。)

③ 図11-3のように，仮想的な回帰直線からの偏りを χ_{11}^2 とする。H_0：回帰直線からの偏り(へだたり)はない，という仮定のもとで， df: R-2

$$\chi_{11}^2=\chi_0^2-\chi_1^2 \sim \chi_{l-2}^2 \quad \cdots\cdots(\text{XI-5})$$

$\chi_{11}^2=4.836-4.820=0.016<\chi_1^2(0.05)=3.841$

したがって H_0 は棄却できず，② から回帰による傾向が認められるので，年齢階級が進むにつれて，陽性率は有意に($p<0.05$)上昇する。▶補足11-F

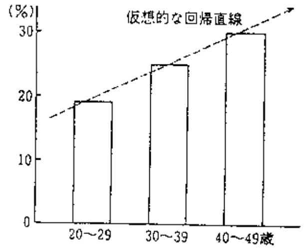


図11-3　年齢階級別陽性率

(6) リジット解析(Ridit Analysis)

順序尺度(p. 3)である *I* について対照群と比較する場合である。たとえば，食事調査で，① 毎日，② 週4〜6回，③ 週1〜3回，④ あまり食べない等，順序をもつ例は医学的データには多い。しかし，解説しているテキストは少ない。

【例11-5】治療薬A，Bについて薬効試験を行った結果は，表11-9のとおりであった。両薬の効果に差が認められるか，$\alpha=5\%$として判定しなさい。

表11-9　薬効試験の例

	a)著効	b)有効	c)やや有効	d)不変	e)悪化	計	a)〜c)	d)+e)
対照薬A	5	10	15	15	5	50	30	20
実験薬B	10	15	12	10	3	50	37	13
計	15	25	27	25	8	100	67	33

※(4)，(5)の場合は表を↓方向に見たが，ここでは計が→方向にある。

A，Bの入れ替えは可能であるが，対照群が基準となるので表11-9のように書いた方がよい。(XI-4)式によると，$\chi_0^2=4.500<\chi_4^2(0.05)=9.488$ である。

また，e)悪化のB薬の期待値は $50\cdot8/100=4<5$ なので，表11-9のようにa)〜c)とd)+e)とに分割する。2×2表として(X-4)式を用いても，$\chi^2_Y=1.628<\chi_1^2(0.05)=3.841$ と，有意差が認められないことに注意されたい。

表11-10　リジット解析の計算例

	1) 対照群	2) 1)/2	3) 累積度数	4) 2)+3)	5) 4)/n_1	6) 実験群	7) 5)×6)
著　効	5	2.5	0	2.5	0.05	10	0.50
有　効	10	5.0	5	10.0	0.20	15	3.00
やや有効	15	7.5	15	22.5	0.45	12	5.40
不　変	15	7.5	30	37.5	0.75	10	7.50
悪　化	5	2.5	45	47.5	0.95	3	2.85
	$n_1=50$	——	50	50.0	1.00	$n_2=50$	19.25

そこでリジット解析を適応するが，手順は表11-10を参照のこと。

▶補足11-G

① 対照群の度数*1)*の半分を*2)*とする。

② *3)*に累積度数(p. 10)をとるが，はじめは0として通常より一段下げる。

③ *2)*+*3)*を求め*4)*とする。累積度数について(各区分+次区分)/2でもよい。

④ *4)*を対照群の n_1 で割った値を*5)*とし，対照群の期待的な分布と考える。

⑤ *5)*に実験群の各度数*6)*をかけて，*7)*の計を求める。

$$\begin{cases} \text{実験群の平均リジット } \overline{R}_B \text{ は，} \overline{R}_B=\dfrac{7)\text{の計}}{n_2}=0.385 \\ \text{対照群の平均リジット } \overline{R}_A \text{ は常に，} \overline{R}_A=0.5\left(\because \dfrac{\sum 1)\cdot 5)}{n_1}=\dfrac{25}{50}\right) \end{cases}$$

⑥ 近似的に平均リジットの標準誤差(p. 28) $s=\dfrac{1}{2\sqrt{3n_2}}$ である。

したがって，$H_0:\overline{R}_A=\overline{R}_B$ のもとで，

$$z_0=\frac{|\overline{R}_A-\overline{R}_B|}{s}=\sqrt{12n_2}\,|\overline{R}_A-\overline{R}_B| \sim N(0,\ 1^2) \qquad \cdots\cdots(\text{XI-6})$$

を用いて検定する。〔例11-5〕では，

$$z_0=\sqrt{12\cdot50}\,|0.5-0.385|=2.817>z\left(\frac{0.01}{2}\right)=2.576$$

したがって $p<0.01$ で H_0 を棄却でき，両群の分布に差が認められる。

⑦ ところで，両群の平均リジット $\pm s$ を表示してみると，図11-4のように実験群の方が左にシフトしている。表11-10では，「著効」を先にとったの

統計学の基礎１，２

目　次　　　　　　　　　　ページ

統計学の基礎１，２
～統計分布による推定・検定法を中心として～

福岡県保健環境研究所　篠原志郎

A．統計的方法の概要

１．統計的な考え方

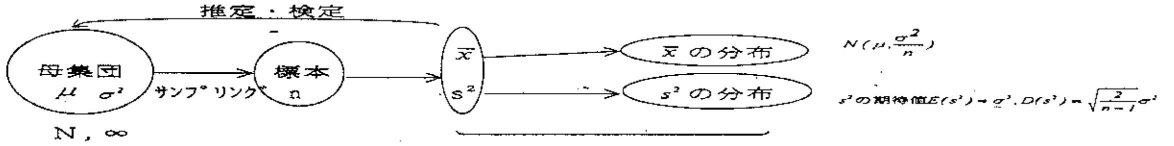
１）測定値にはバラツキ（*Dispersion*）がつきもの（誤差）

２）そのバラツキには統計的な規則性がみられる（分布）

３）大数の法則：　ｘが平均μ，分散σ^2の分布に従うとき，xがどんな分布であっても，標本数nの無作為標本（*Random sampling*）による$\bar{x}$は正規分布$N(\mu, \sigma^2/n)$にnが大きくなるにつれて近づく（中心極限定理：一般に，$n>25$で適用）

４）統計的推論に基づいて行動する（検定）

Fisher の３原則＋１（*1935*）：反復（*Replication*），局所管理（*Local control*），無作為化（*Randomization*），因子の組み合わせ（*Factorial combination*）



２．特性値

母集団を知る手がかりとして、母集団から取られた標本の度数分布表を作り、ヒストグラム等を描く．それによって分布の形や特徴がつかめるが、更に、数量化できれば、なお、的確な情報を得ることができる．この数量的に表現された指標を特性値とよぶ．

１）分布の中心を代表する指標

①平均値・・・a．算術平均値　$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{k=1}^{n} x_k = \sum x / n,$

b．幾何平均値　$\bar{y} = \sqrt[n]{x_1 \cdot x_2 \cdots x_n} = (x_1 \cdot x_2 \cdots x_n)^{\frac{1}{n}},$

c．調和平均値　$\bar{z} = \frac{1}{n}(\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}), x_1, x_2, \cdots, x_n \neq 0$

②中央値・・・Ｍｅ＝（n+1）／2番目の測定値　（ｎ：奇数odd number）
＝[n/2番目＋（n/2+1）番目]／2　（ｎ：偶数even number）

③モード（最頻値）Ｍｏ＝最も度数の多い測定値

④最大値、最小値

２）分布の広がりの程度を代表する指標

①分散（不偏分散）　$s^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \bar{x})^2 = \frac{1}{n-1}\sum(x_k^2 - n\bar{x}^2)$

②標準偏差　$s = \sqrt{s^2}$

③範囲　$R = x_{max} - x_{min}$

④変動係数　$CV = \frac{s}{\bar{x}} \times 100(\%)$

３）正規性からのズレと異常

①ゆがみ具合（歪度：Skewness）　$\alpha_3 = \frac{\mu_3}{\sigma^3}, \mu_k = E((X-\mu)^k), \alpha_3 > 0$（分布が右に広がる）

②尖り具合（尖度：Kurtosis）　$\alpha_4 = \frac{\mu_4}{\sigma^4} - 3, \alpha_4 > 0$（トガリ型），$\mu_4$の推定$= \frac{1}{n}\sum(x_i - \bar{x})^4, \sigma^3 = (\frac{1}{n}\sum(x_i - \bar{x})^2)^{\frac{3}{2}} = (\sigma^2)^{\frac{3}{2}}$

４）データのスクリーニング

①異常データの検出　*Grubbs,F.E.* 法

B．標本抽出法（*Sampling*）

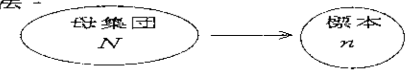
1．有意抽出法・・・意図的に標本を抽出する方法

1）典型法

母集団の特性値から母集団の相似形を意識して求めた抽出法

2）割当法

母集団の特性値が同じ比率となるように選ぶ方法



2．確率抽出法

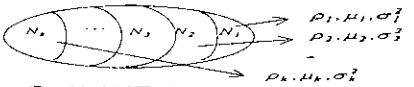
1）単純無作為抽出法（*Simple random sampling*）

N個の要素の中から大きさ nの標本$x_1, x_2, \cdots, x_n$を非復元抽出して標本平均$\bar{x}$を求める．

$\bar{x}=\frac{1}{n}\sum x,\quad E(\bar{x})=\mu,\quad V(\bar{x})=\frac{N-n}{N-1}\cdot\frac{\sigma^2}{n}$ で与えられる．

2）層別抽出法（*Stratified sampling*）

母集団をk個の層に分け、各層から標本を抽出する方法



全平均$\mu=\sum_{i=1}^{k}\rho_i\mu_i,\ (\rho_1+\rho_2+\cdots+\rho_k=1,\ \rho_i=\frac{N_i}{N})$

全分散$\sigma^2=\sum\rho_i\sigma_i^2+\sum\rho_i(\mu_i-\mu)^2=\sigma_w^2+\sigma_b^2$

$\sigma_w^2=\bar{\sigma}^2+\sum\rho_i(\sigma_i-\bar{\sigma})^2,\quad \bar{\sigma}=\sum\rho_i\sigma_i$

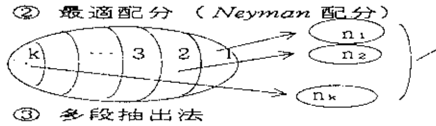
① 比例配分

k個の層に分けられている母集団から、各層の大きさに比例した大きさの標本を抽出する方法、即ち、$\rho_i=\frac{N_i}{N}=\frac{n_i}{n}$、標本平均 $\bar{x}_s=\sum\rho_i\bar{x}_i$ （$\bar{x}_i$は第i層からの標本平均）

∴ 分散 $V(\bar{x}_s)=\sum\rho_i\frac{N_i-n_i}{N_i-1}\cdot\frac{\sigma_i^2}{n}\fallingdotseq\sum\rho_i\cdot\frac{\sigma_i^2}{n}$　　実は、$V(\bar{x})\geqq V(\bar{x}_s)$

が成り立つ．即ち、単純ランダムサンプリングより層別抽出法が分散を小さくできる．

② 最適配分（*Neyman* 配分）



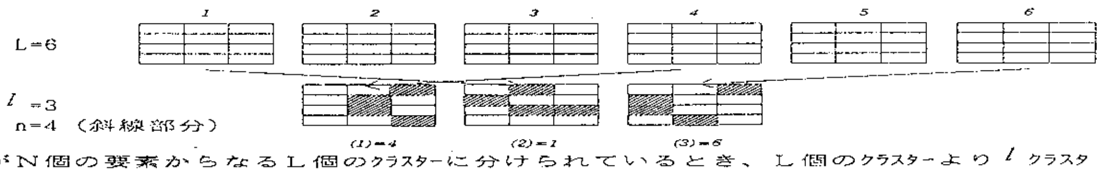
$n_1+n_2+\cdots+n_k=n$

$n_i=\frac{\rho_i\sigma_i}{\rho_1\sigma_1+\rho_2\sigma_2+\cdots+\rho_k\sigma_k}\cdot n,\ (i=1,2,\cdots,k)$とおくと，

$V(\bar{x}_s)=\sum\rho_i{}^2\frac{N_i-n_i}{N_i-1}\cdot\frac{\sigma_i^2}{n_i}\fallingdotseq\sum\rho_i{}^2\cdot\frac{\sigma_i^2}{n_i}$　最小$\rightarrow\frac{(\sum\rho_i\sigma_i)^2}{n}$

③ 多段抽出法

a．二段抽出法：第1段階として全母集団から適当な大きさの中間的抽出集落（クラスター）をいくつか抽出し、次に本来の調査個体を抽出する方法



母集団がN個の要素からなるL個のクラスターに分けられているとき、L個のクラスターよりlクラスター抽出し、抽出された クラスターからそれぞれn個の標本を選び、標本平均 $\bar{x}_{(1)}, \bar{x}_{(2)}, \cdots, \bar{x}_{(l)}$を求め、

$\bar{\bar{x}}=\frac{\bar{x}_{(1)}+\bar{x}_{(2)}+\cdots+\bar{x}_{(l)}}{l},\ E(\bar{\bar{x}})=\mu$　（母平均）

$V(\bar{\bar{x}})=\frac{N-n}{N-1}\cdot\frac{\sigma_w^2}{ln}+\frac{\sigma_b^2}{l}.$

$\sigma_w^2=\frac{1}{L}\sum_{j}^{L}\sigma_j^2,\ \sigma_b^2=\frac{1}{L}\sum_{1}^{L}(\mu_j-\mu)^2$

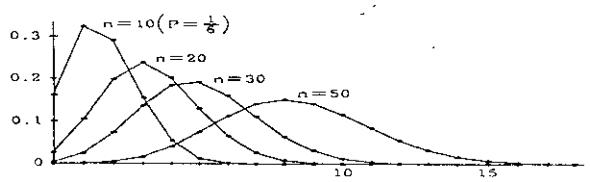
C．確率分布

1．二項分布（*Binomial distribution*）　B（np, npq），$q=1-p$

二項定理によって、$(p+q)^n=q^n+npq^{n-1}+\frac{n(n-1)}{2!}p^2q^{n-2}+\cdots+\frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}p^kq^{n-k}+\cdots+np^{n-1}q+p^n$

二項分布とは離散型変数の分布で、表と裏、1と0のように2つのうちどちらかを選択す場合の確率分布である。この場合、p＋q＝1，p，q≧0である。例えば、n回のうちk回は事象Aが起き（確率p）、（n－k）回は事象A以外が起きる（確率：1－p）場合の確率をP（k）とすると、$P(k)=\binom{n}{k}p^kq^{n-k}=\frac{n!}{k!(n-k)!}p^kq^{n-k}=\frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}p^kq^{n-k}$

で示される。このとき、二項分布の平均値は　np，分散は　np（1－p）＝npq．



サイコロの目の出る確率

n＝10の確率

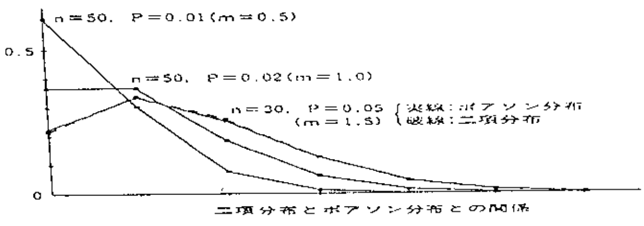
x	${}_nC_xp^xq^{n-x}$
0	0.161506
1	0.323011
2	0.290710
3	0.155045
4	0.054266
5	0.013024
6	0.002171
7	0.000248
8	0.000019
9	0.000001
10	0.000000

2．ポアソン分布（*Poisson distributiom*）P（λ）

ポアソン分布も離散型変数の分布で、例えば、交通事故や疾病による死亡数のような希に起こる事象の分布である。二項分布の平均値npにおいて、pが非常に小さく0に近く、nが大きいがnp＜5の場合、このポアッソン分布によく当てはまる。ポアソン分布の確率をP（x）とすると、$P(x)=e^{-\lambda}\frac{\lambda^x}{x!},\ x=0,1,2,\cdots$　　ここに、e（自然対数の底）＝2.71828・・・

ポアソン分布の平均値λ，分散λである。

二項分布で平均 m＝nP を一定にし，n→∞，P→0 の極限型がポアソン分布であるが，nP＜5かつ n≧50の場合，正規分布に近似できる。



二項分布とポアソン分布との関係

3．正規分布（*Normal distribution*）N（μ，σ^2）

正規分布は連続型変数の分布である。最もよく使われる分布である。自然界の現象でこの正規分布が適用されるケースが非常に多い。正規分布の確率（この場合、確率密度関数とい

う） f（X）で示すと、 $f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty<x<\infty$

また、正規分布をN（μ，σ^2）で表す。正規分布の平均値はμ、分散はσ^2である。正規分布のグラフはx＝μで左右対称である。正規分布関数は

$$F(x)=\int_{-\infty}^{x}f(x)dx=P(-\infty<X<x)=1-P(x\leq X)$$

分布形の面積が確率を表している。全面積は１である。z＝（x－μ）／σの変換を施せば、N（μ，σ^2）はN（０，１）に変わる。即ち、平均０、分散１の正規分布である。これを規準（標準）正規分布という。

○正規確率紙

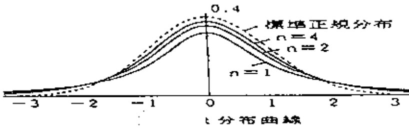
作成した分布が正規分布しているかどうかを調べるためのもので、正規分布に近ければ、それぞれの確率点を正規確率紙にプロットしたものが、ほぼ直線になる。より厳密に確かめるにはχ^2－分布による検定を行う。

自由度がnのt分布の基本的な性質は以下のとおりである。

①分布の全面積は１である
②t＝０で極大値をとる
③分散は $\frac{n}{n-2}$（n＞2）

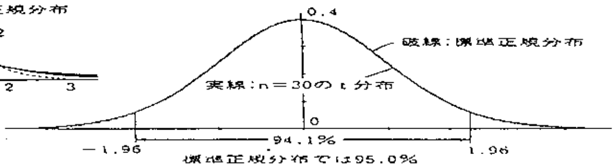
確率密度関数は、ガンマΓ関数(p. 36)を用いて、

$$y=\frac{\Gamma\left(\frac{n+1}{2}\right)}{n\pi\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}$$



t分布曲線

（※nは自由度）



t分布と標準正規分布との比較

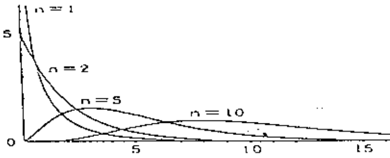
4．t－分布

標本平均$\bar{x}$の分布はnが大きいときはN（μ，σ^2／n）に従う。σが未知のとき、σの代わりに標本分散s^2を使うことがよくある。しかし、標本数n≦25の場合、$t=\frac{\bar{x}-\mu}{\sqrt{s^2/n}}$は、自由度$\nu$＝n－１のt－分布に従う。t－分布はn→大にすると、正規分布に近づいていく。t－分布のグラフは左右対称の分布である。

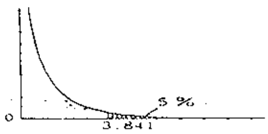
5．カイ２乗分布（χ^2－分布）（*Chi－square distribution*）

２つ以上の選択しがあるケース、あるいは正規分布に従う変数Xの２乗X^2はχ^2－分布に従う。χ^2－分布は自由度νをもち、ν（*0*、*1*、*2*、‥）の大きさで分布曲線が異なってくる。大きさnの標本が正規母集団（正規分布に従う母集団）から抽出されたものであるなら、そこから求められる変数、$\frac{nS^2}{\sigma}=[(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2]/\sigma=\frac{\sum x^2-n\bar{x}^2}{\sigma}$は自由度$\nu$＝n－１の$\chi^2$分布に従う。

表の縦・横の項目に関連がない時、互いに「独立である」という。平均m，分散S^2の母集団から大きさnの標本を無作為抽出する時，$\chi^2=\frac{1}{S^2}\Sigma(x_i-m)^2$［≧0］は自由度nのカイ２乗分布に従う。①定義により正側確率しか存在しない。確率密度はガンマ関数を用いるが省略する。②nによって形が大きく異なり，nが大きくなると左右対称に近くなる。



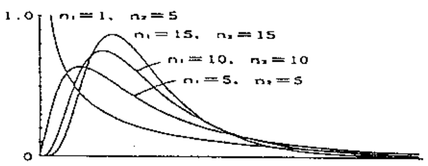
カイ２乗分布曲線



自由度１の棄却域

6．F－分布

標本分散s^2の検定によく用いられる分布である。例えば、分散がσ_1^2，σ_2^2の２つの正規母集団から、大きさmとn'の標本を取ったとき、その標本分散を$s_1^2 s_2^2$とすると、$F=\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$は自由度$\nu_1$＝m－１、$\nu_2$＝n－１のF－分布に従うことが分かっている。これを使えば、分散の大きさの違いを比較検定することができる。



F分布の自由度と形状

F分布は左右対称でないので，正規分布やt分布と異なり，棄却域も(図)に示すように，上側と下側で左右対称となっていない。

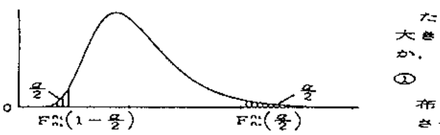
しかし，$F_{n_2}^{n_1}\left(1-\frac{\alpha}{2}\right)=\frac{1}{F_{n_1}^{n_2}\left(\frac{\alpha}{2}\right)}$

の関係が成立するので，等分散の検定の棄却域は，

$$\frac{u_1^2}{u_2^2}\geq F_{n_2}^{n_1}\left(\frac{\alpha}{2}\right),\quad \frac{u_1^2}{u_2^2}\leq F_{n_2}^{n_1}\left(1-\frac{\alpha}{2}\right)=\frac{1}{F_{n_1}^{n_2}\left(\frac{\alpha}{2}\right)}$$

つまり，$\frac{u_1^2}{u_2^2}\geq F_{n_2}^{n_1}\left(\frac{\alpha}{2}\right)$，または$\frac{u_2^2}{u_1^2}\geq F_{n_1}^{n_2}\left(\frac{\alpha}{2}\right)$の判定となるため，大きい方の不偏分散を分子として，上側確率だけで判定してよい。

また，t分布との関係は，$\left\{t_n\left(\frac{\alpha}{2}\right)\right\}^2=F_n^1(\alpha)$である。



$F_{n_2}^{n_1}\left(1-\frac{\alpha}{2}\right)$　$F_{n_2}^{n_1}\left(\frac{\alpha}{2}\right)$

F検定の棄却域

二項確率をf(x)とすると，コンピュータでの計算は簡単である。

$f(0)=(1-P)^n$，$f(i+1)=\frac{n-i}{i+1}\cdot\frac{P}{1-P}\cdot f(i)$［$i$＝0～n－1］

ただし，n，Pがともに大きいと，計算精度が問題となる。手計算ではnが大きいと階乗！の計算は無理なので，二項係数$_nC_x$の数値表が作られているほか，他の分布へのさまざまな変換が試みられてきた。

① 正規分布への近似：二項分布でP＝Q＝$\frac{1}{2}$でかつn→∞の極限型は正規分布となる。そこで，nP≧５（かつnQ≧５）の時，二項分布を正規分布に近似させてよいとされている。

② 逆正弦変換：多くの比率を扱う場合，θ（ラジアン）＝$\sin^{-1}\sqrt{P}$と変換すると分散は$\frac{1}{4n}$となり，確率とは無関係になる

7．二項分布の正規近似

○二項分布において、p≦*1/2*のとき、np＞５ならば、正規分布に近似する。また、p＞*1/2*のときはnq＞５ならば、正規分布に近似すると考えてよい。二項分布の平均値はnp、分散はnpqであるから、規準化して、 $z=(x-np)/\sqrt{npq}$ にすると、

zは正規分布*N(0,1)*に従う。即ち、平均０、分散１の正規分布に従う。

○zを更に、nで割って、$z=(\frac{x}{n}-p)/\sqrt{\frac{pq}{n}}$ は、割合x／nが平均p、分散pq／nの正規分布が使えることになる。

○xがN（μ，σ^2）に従うなら、標本数nのランダムサンプリングによる$\bar{x}$は$\bar{x}\sim N(\mu,\frac{\sigma^2}{n})$であることがわかっている。これを$\bar{x}$の標本分布という。更に、nがかなり大きい数、例えば、n≧*50*、実用的にはn≧*25*なら、xが正規分布していなくても平均μ、分散σ^2の分布でありさえすれば、$\bar{x}\sim N(\mu,\frac{\sigma^2}{n})$であることもわかっている。【中心極限定理といって統計学では重要な法則の一つである。この法則があるため、あまり厳密な検証なしに$\bar{x}$の推定・検定ができるのである。】

D．推定と検定

1．母数（*Parameter*）

母集団を特徴づける変量をさす。例えば、母集団が正規分布N（μ，σ^2）に従うとき、この正規分布を決定づける変量μ、σを母数（ﾊﾟﾗﾒｰﾀ）という。μ、σが分かれば、その母集団は既知と考える。母集団が二項分布B（n，p）に従うなら母数はpである。pが決まれば、試行nにより二項分布は決まる。

2．統計量（*Statistic*）

母数を推定するために標本から得られる任意の量のことである。例えば、母集団から抽出された$x_1, x_2, \cdots, x_n$は標本であるが、平均$\bar{x}=(x_1+x_2+\cdots+x_n)/n$、分散s^2，標準偏差sなどは、標本から計算された統計量である。そして、この統計量の分布が考えられる。

3．点推定（*Point estimation*）

一点で与えられる母数の推定を点推定という。例えば、母平均μの推定値として$\bar{x}$、Me、Mo等を用いることができるが、どれがμの良い推定量であるかは、多数の繰り返し実験のなかで、どの統計量の値がμの代用値として適切であるかということである。良い推定量の基準として不偏推定量があげられる。不偏推定量はE（x）＝μと記す。E（x）は平均$\bar{x}$の標本分布の平均のことで、期待値とよび、これがμに等しいとき$\bar{x}$はμの不偏推定量であるという。そのような不偏推定量がいくつかある場合、最小の分散をもつものが最も有効であると考える。例えば、$\bar{x}$とMeは共に同じ母平均をもっている。しかし、$\bar{x}$の標本分布の分散はσ^2／n、Meの標本分布の分散は$\pi\sigma^2/(2n)$で、σ^2／nの方が小さい。従って、$\bar{x}$の方がMeより有効な推定量といえる。また、分散$s^2=(\sum x_i^2-n\bar{x}^2)/(n-1)$は$E(s^2)=\sigma^2$であることがわかっている。s^{2}は$\sigma^2$の不偏推定量である。それゆえ、このs^{2}を不偏分散とよんでいる。

4．区間推定（*Interval estimation*）

母数が与えられた2点の間に存在すると考える推定を区間推定という。一般に、推定の信頼度を95％あるいは99％に置く。その信頼度における推定区間を信頼区間という。この区間の両端を信頼限界値という。信頼度を高めようとすれば、推定区間は広くなる。

$\bar{x}\sim N(\mu,\frac{\sigma^2}{n})$のとき、母平均$\mu$の95％信頼限界は$\bar{x}\pm 1.96\frac{\sigma}{\sqrt{n}}$、真の平均$\mu$は、95％の信頼確率で、$\bar{x}-1.96\frac{\sigma}{\sqrt{n}}<\mu<\bar{x}+1.96\frac{\sigma}{\sqrt{n}}$　の間に存在するといえる。

1）正規分布によるμの区間推定

95％の信頼区間は　$\bar{x}-1.96\frac{\sigma}{\sqrt{n}}<\mu<\bar{x}+1.96\frac{\sigma}{\sqrt{n}}$

99％の信頼区間は　$\bar{x}-2.576\frac{\sigma}{\sqrt{n}}<\mu<\bar{x}+2.576\frac{\sigma}{\sqrt{n}}$

2）t－分布によるμの区間推定

$\bar{x}-t_{n-1}(\alpha)\frac{s}{\sqrt{n}}<\mu<\bar{x}+t_{n-1}(\alpha)\frac{s}{\sqrt{n}}$，　$s=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}$，$t_{n-1}(\alpha)$は自由度$\phi=n-1$，危険率$\alpha$の$t$－分布値

例えば、n＝21のとき、$n-1=20$、t－分布95％値は2.086、99％値は2.845

3）割合pの区間推定

n回の試行で事象Aがx回起これば、割合の実現値　$\hat{p}=\frac{x}{n}$　、n→大のとき、$\hat{p}$はN（p，pq／n）に従うので、95％信頼区間は

$$\hat{p}-1.96\sqrt{\frac{pq}{n}}<p<\hat{p}+1.96\sqrt{\frac{pq}{n}}$$　、q＝1－p

99％信頼区間の場合は、1．96の代わりに2．576とすればよい。

5．標本数の決定

1）比率の場合

母集団の大きさが有限Nであるとき、比率pの標準偏差$\sigma_p=\sqrt{\frac{N-n}{N-1}\cdot\frac{p(1-p)}{n}}$より、$p$の95％信頼限界は$p$の代わりに$\hat{p}$を用いて、$\hat{p}\pm 1.96\sqrt{\frac{N-n}{N-1}\cdot\frac{\hat{p}(1-\hat{p})}{n}}$、従って、母数の推定誤差を$d$とすれば、$\hat{p}\pm d$と書けるので、$n=\frac{N}{(\frac{d}{1.96})^2\cdot\frac{N-1}{\hat{p}(1-\hat{p})}+1}<\frac{N}{(\frac{d}{1.96})^2\cdot\frac{N-1}{0.3^2}+1}$

例えば、$d=\hat{p}/10$、即ち、母比率の10％とおくと、$\hat{p}=0.3$なら$d=0.03$
信頼度99％の場合は、1．96の代わりに2．576とする。dは精度であるので適当に設定すればよい。

2）平均値の場合

大きさNの有限母集団の平均値$\bar{x}$の標準偏差は$\sigma_{\bar{x}}=\sqrt{\frac{N-n}{N-1}\cdot\frac{\sigma^2}{n}}$であるから、同様に$d=1.96\sqrt{\frac{N-n}{N-1}\cdot\frac{\sigma^2}{n}}$とおいて、$n=\frac{N}{(\frac{d}{1.96})^2\cdot\frac{N-1}{\sigma^2}+1}$

何らかの形でσの値を推定して、nを求める。精度dを小さくすれば、nは大きくなり、σが大きければ、nも大きくなることがわかる。

6．検定

1）仮説検定（*Test of hypothesis*）

仮説検定とは、ある仮説の下で母集団の母数、例えば、母平均μ、母分散σ^2、母比率p等同士の比較あるいは母集団の母数と標本統計量、即ち、標本平均$\bar{x}$、標本分散s^2、標本比率$\hat{P}$等を比較し、確率分布による判定を行うことをさす。検定においては、まず、仮説を立て、設定したある確率の下でこの仮説が棄てられるかどうかを判定する。この設定した確率を有意水準または危険率という。

2）有意水準（*Significant level*）

仮説は棄てられることに有意な（偶然に起こったものとは認められない）意味を積極的に含むものである。その意味で、帰無仮説といい、判定する確率を有意水準とよぶ。危険率という言葉は、この仮説が実は正しいにも拘わらず、誤ってそれを棄てる危険を犯す場合があり、その確率をさす。この誤りを第1種の過誤（α）、生産者危険率、あわてものの誤りなどという。また、この仮説が実は正しくないにも拘わらず、誤って採用する危険も含んでいる。この誤りを第2種の過誤（β）、消費者危険率、ぼんやりものの誤りなどという。この2つの誤りα、βは共に小さいに越したことはないが、同時に最小にする方法はない。特に、βに対する検定法がない。そこで、一般に、α＝0．05（5％）あるいは0．01（1％）に設定した上で、βを小さくするには標本の大きさnを大きくする。

3）両側検定と片側検定（*Two－tailed test, One－tailed test*）

この仮説検定において、仮説を否定した場合は何を採用するかが問題になる。これが対立仮説である。例えば、薬効試験の場合、仮説は薬効なしとすれば対立仮説は薬効ありとなる。血圧降下剤試験であれば、薬効ありなら投与後、明らかに血圧は下がっていることになるから、投与前と投与後の母数には大小関係、即ち、母数に関する事前情報があることになる。母数に関する事前情報がない場合は仮説検定が両側検定となり、事前情報がある場合は片側検定が考えられる。

4）検定手順

仮説検定の手順は次のように行う。

① 仮説H$_0$、対立仮説H$_1$を立てる。
② 仮説の下での標本統計量zを計算する。
③ 有意水準（または危険率）5％に対応する分布値z (0.05)、有意水準1％に対応する分布値z (0.01) を表より決定する。
④ z＜z (0.05) なら危険率5％で有意差は認められない。即ち、仮説はすてられない。
⑤ z (0.05) ≦z＜z (0.01) なら危険率5％で有意差ありとする。即ち、仮説は棄却される。
⑥ z (0.01) ≦zなら危険率1％で有意差ありとする。即ち、仮説は棄却される。

7．分布による仮説検定

1）平均値の検定（分散σ^2未知、n≧25のとき）

① 仮説H_0：$\mu=\mu_0$、H_1：$\mu\neq\mu_0$

② 仮説の下では、$z=\dfrac{|\bar{x}-\mu_0|}{\sqrt{\dfrac{s^2}{n}}}$は規準正規分布$N(0,1)$に従う。

③ 正規分布N(0,1)の危険率5%点は
$z(0.05)=1.96$、1%点$z(0.01)=2.576$

④ $z<1.96$なら危険率5%で有意差は認められない。

⑤ $1.96\leqq z<2.576$なら危険率5%で有意差あり。

⑥ $2.576\leqq z$なら危険率1%で有意差あり。

2）t－検定（分散σ^2未知、n＜25のとき）

① 仮説H_0：$\mu=\mu_0$、H_1：$\mu\neq\mu_0$

② 仮説の下では、$t_0=\dfrac{|\bar{x}-\mu_0|}{\sqrt{\dfrac{s^2}{n}}}$は自由度$(n-1)$のt－分布に従う。

③ 自由度（n－1）のt－分布表より、
$t_{n-1}(0.05)$、$t_{n-1}(0.01)$を求める。

④ $t_0<t_{n-1}(0.05)$なら危険率5%で有意差は認められない。

⑤ $t_{n-1}(0.05)\leqq t_0<t_{n-1}(0.01)$なら危険率5%で有意差あり。

⑥ $t_{n-1}(0.01)\leqq t_0$なら危険率1%で有意差あり。

3）対応のある場合の検定

2つの標本群が同一人の投薬前と投薬後のようにデータx_1とy_1、x_2とy_2のように互い対応関係があるとき、それぞれの差$z_1=x_1-y_1$、$z_2=x_2-y_2$、・・・、$z_n=x_n-y_n$をつくる。そして、z_1、z_2、・・・、z_nについて$\bar{z}=\frac{1}{n}\sum_{i=1}^{n}z_i$、$s^2=\frac{1}{n-1}(\sum_{i=1}^{n}z_i^2-n\bar{z}^2)$を求める。以下は2）t－検定を適用する。（n≦25のとき）

4）2つの割合の差の検定（$m\hat{p}_1\geq5, n\hat{p}_1\geq5$のとき）

母集団A：母比率p_1・・・・標本数m、割合　$\hat{p}_1$

母集団B：母比率p_2・・・・標本数n、割合　$\hat{p}_2$

A, B全体の母比率$p=\dfrac{mp_1+np_2}{m+n}$で推定する。このとき、

分散$s^2=pq(\frac{1}{m}+\frac{1}{n})$, $p+q=1$, $p,q\geq0$である。

① 仮説H_0：$p_1=p_2$、対立仮説H_1：$p_1\neq p_2$

② 仮説の下では、$z=\dfrac{|p_1-p_2|}{\sqrt{pq(\frac{1}{m}+\frac{1}{n})}}$は規準正規分布$N(0,1)$に従う。

③ $z<1.96$なら危険率5%で有意差は認められない。

④ $1.96\leqq z<2.576$なら危険率5%で有意差あり。

⑤ $2.576\leqq z$なら危険率1%で有意差あり。

5）2つの平均値の差の検定（n＜25のとき）

母集団A：母平均μ_1、母分散σ^2_1・・・標本数m、平均$\bar{x}_1$、分散s_1^2

母集団B：母平均μ_1、母分散σ^2_2・・・標本数n、平均$\bar{x}_2$、分散s_2^2

① 仮説H_0：$\mu_1=\mu_2$、対立仮説H_1：$\mu_1\neq\mu_2$

② 仮説の下では、$\sigma^2_1=\sigma^2_2$のとき、

$$t=\frac{|\bar{x}_1-\bar{x}_2|}{\sqrt{\dfrac{(m-1)s_1^2+(n-1)s_2^2}{m+n-2}(\dfrac{1}{m}+\dfrac{1}{n})}}$$
は自由度$\phi=m+n-2$のt－分布に従う。

③ 自由度m＋n－2のt－分布表より$t(0.05)$、$t(0.01)$を求める。

④ $t<t(0.05)$なら、危険率5%で有意差は認められない。

⑤ $t(0.05)\leqq t<t(0.01)$なら、危険率5%で有意差あり。

⑥ $t(0.01)\leqq t$なら、危険率1%で有意差あり。

6）独立性の検定（H_0：A, B間に関連がない）適合度の検定

①$k>2$の場合

	B_1	B_2	B_3	…	B_k	Total
A_1	n_{11}	n_{12}	n_{13}	…	n_{1k}	$n_{1.}$
A_2	n_{21}	n_{22}	n_{23}	…	n_{2k}	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	$n_{.3}$	…	$n_{.k}$	n

$$\chi_0^2=\frac{n^2}{n_{1.}n_{2.}}\left(\frac{n_{11}^2}{n_{.1}}+\frac{n_{12}^2}{n_{.2}}+\frac{n_{13}^2}{n_{.3}}+\cdots+\frac{n_{1k}^2}{n_{.k}}-\frac{n_{1.}^2}{n}\right)$$
$>\chi^2_{k-1}(\alpha)$のとき　A, Bは独立ではない。

②k＝2の場合

	B_1	B_2	Total
A_1	a	b	g
A_2	c	d	h
Total	e	f	n

$$\chi_0^2=\frac{(|ad-bc|-n/2)^2n}{efgh}$$
≧3.84なら危険率5%で有意差あり。
≧6.63なら危険率1%で有意差あり。

E．相関分析

1．相関係数 r の検定

1）標本が１組（H₀ : ρ＝0 の場合）

$$r=\frac{S_{xy}}{\sqrt{S_{xx}\cdot S_{yy}}}=\frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2\cdot\sum(y-\bar{y})^2}}=\frac{\sum xy-n\bar{x}\bar{y}}{\sqrt{(\sum x^2-n\bar{x}^2)\cdot(\sum y^2-n\bar{y}^2)}}$$

$F_0=\frac{(n-2)r^2}{1-r^2}$ は自由度 $\nu_1=1,\nu_2=n-2$ のＦ(α)分布に従う。ここに、αは有意水準100α%である。

2）標本が１組の場合（H₀ : ρ＝ρ₀ の場合）

２次元正規母集団からの標本とみなす。

① Ｚ変換 $r^*=\frac{1}{2}ln\frac{1+r}{1-r},\quad \rho^*=\frac{1}{2}ln\frac{1+\rho}{1-\rho}$ （ln はeを底とする自然対数）

② $Z=\sqrt{n-3}(r^*-\rho^*)$ のとき $|Z|\sim N(0,1)$

即ち、$|Z|\geq 1.96$ のとき有意水準5%で仮説を棄却する。

③ r の 95 %信頼限界は

下限値 $r_1^*=r^*-\frac{1.96}{\sqrt{n-3}}$，上限値 $r_2^*=r^*+\frac{1.96}{\sqrt{n-3}}$ を求め、

$r_1^*=\frac{1}{2}ln\frac{1+r_1}{1-r_1},r_2^*=\frac{1}{2}ln\frac{1+r_2}{1-r_2}$ を満たすr_1,r_2を信頼下限、信頼上限とする。

即ち、 $r_1=\frac{exp(2r_1^*)-1}{exp(2r_1^*)+1},\quad r_2=\frac{exp(2r_2^*)-1}{exp(2r_2^*)+1}$

3）標本が２組（H₀ : ρ₁＝ρ₂の場合）

① $r_1^*=\frac{1}{2}ln\frac{1+r_1}{1-r_1},\quad r_2^*=\frac{1}{2}ln\frac{1+r_2}{1-r_2}$ を求める。

② $Z=\frac{r_1^*-r_2^*}{\sqrt{\frac{1}{n_1-3}+\frac{1}{n_2-3}}}$ （n_iは第i標本のデータ数,$i=1,2$）

$|Z|\geq 1.96$ なら有意水準５%で仮説を棄却する。即ち、母相関係数 ρ₁≠ρ₂とする。

あるいは、

$Z^2=\frac{(r_1^*-r_2^*)^2}{\frac{1}{n_1-3}+\frac{1}{n_2-3}}\sim\chi_1^2(\alpha)$ 即ち、Z^2は自由度１のχ^2分布に従う。

この $Z^2\geqq 3.84$ （$=1.96^2$）なら仮説を棄却する。

4）標本がｋ組（H₀ : ρ₁＝ρ₂＝・・・＝ρₖの場合）

r_1 標本１	r_2 標本２	・・・・	r_k 標本ｋ
n_1個	n_2		n_k

対立仮説H₁ : $\rho_i\neq\rho_j$ $(i\neq j)$

① $r_i^*=\frac{1}{2}ln\frac{1+r_i}{1-r_i}(i=1,2,\cdots,k)$

② $\chi_0^2=\sum_i^k(n_i-3)r_i^{*2}-\frac{(\sum_i^k(n_i-3)r_i^*)^2}{\sum_i^k(n_i-3)}$

このχ_0^2は$\chi_{k-1}^2(\alpha)$分布に従う。従って、自由度$(k-1)$のχ^2分布の5%点を$\chi_{k-1}^2(0.05)$とすると、$\chi_0^2\geq\chi_{k-1}^2(0.05)$のとき有意水準5%で仮説を棄却する。

2．偏相関係数の検定（ｐ変量の場合）

1）偏相関係数 $r_{12\cdot 3}$ は変数３の値がすべて同一である個体の切断面内の変数 1,2 の間の相関である。

$r_{12\cdot 3}=\frac{r_{12}-r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}}=r$ とおく。

①仮説H_0 : $\rho_{12\cdot 3}=0$

$t=\frac{r\sqrt{n-3}}{\sqrt{1-r^2}}$ とおくと、$|t|\geq t(n-3,0.05)$ なら、５%で仮説を棄却する。

t(n−3, 0, 05)は自由度n−3の両側5%点である。

注）変数x,y,zがあって、yのxに対する回帰式$y'=a_1+b_1x$を求める。残差$d_{yi}=y_i-y'_i$ $(i=1,2,..,n)$が得られる。同様に、zのxに対する回帰式$z'=a_2+b_2x$を求める。残差$d_{zj}=z_j-z'_j$ $(j=1,2,...,n)$が求められる。残差の組$(d_{y1},d_{z1}),(d_{y2},d_{z2}),$.(d_{yn},d_{zn})について、単相関を求める。それを$r_{yz\cdot x}$で表すのである。

2）偏相関係数 $r_{12\cdot 34}$（４変数の場合）

$r_{12\cdot 34}=\frac{r_{12\cdot 4}-r_{13\cdot 4}r_{23\cdot 4}}{\sqrt{(1-r_{13\cdot 4}^2)(1-r_{23\cdot 4}^2)}}=\frac{r_{12\cdot 3}-r_{14\cdot 3}r_{24\cdot 3}}{\sqrt{(1-r_{14\cdot 3}^2)(1-r_{24\cdot 3}^2)}}=r$

$t=\frac{r\sqrt{n-4}}{\sqrt{1-r^2}}$ と置くと、$|t|\geq t(n-4,0.05)$ なら有意水準５%で仮説を棄却する。

即ち、仮説H₀ : ρ＝$\rho_{12\cdot 34}$＝0 を捨て、$\rho_{12\cdot 34}\neq 0$ とするのである。

3．平行性検定法

２組のデータ列（$x_{s1},x_{s2},\cdots,x_{sn}$）に対する（$y_{s1},y_{s2},\cdots,y_{sn}$）と（$x_{t1},x_{t2},\cdots,x_{tm}$）に対する（$y_{t1},y_{t2},\cdots,y_{tm}$）を考える。２つの回帰直線

$$Ys=\bar{y}_s+b_s(x_s-\bar{x}_s)$$
$$Yt=\bar{y}_t+b_t(x_t-\bar{x}_t)$$

結論から述べよう。次の分散分析表を作成する。

要　因	SS	df	Ms
R（共有直線性）	S_R	1	V_R
Dp（非平行性）	S_{Dp}	1	V_{Dp}
RΣ（全直線性）	$S_{R\Sigma}$	2	$V_{R\Sigma}$
e　（残　差）	Se	n＋m−4	Ve
Y　（全　体）	Syy	n＋m−2	−

仮説Ｈ o ： $b_s=b_t$ （平行である） に対して、

$$F_{cal}=\frac{S_{Dp}}{V_e}\sim F^{1}_{n+m-4}(\alpha)$$ 即ち、F_{cal}は自由度$(1,n+m-4)$のＦ分布に従う。

これにより、Ｆ $cal \geqq F(1,n+m-4:\ \alpha\ )$ なら $100\ \alpha$ %で仮説を棄却する。
即ち、 $b_s \neq b_t$
とする。２つの直線は平行ではないとする。Ｆ $cal<F(1,n+m-4:\ \alpha\ )$ なら平行とはいえない。

$Ys=\overline{y}_s+b_s(x_s-\overline{x}_s)$ において、$Sxxs=\sum x_{si}^2-\frac{(\sum x_{si})^2}{n}$

$$Syys=\sum y_{si}^2-\frac{(\sum y_{si})^2}{n},\quad Sxys=\sum x_{si}y_{si}-\frac{(\sum x_{si})(\sum y_{sj})}{n},$$
$$b_s=\frac{Sxys}{Sxxs},\quad S_{Rs}=\frac{(Sxys)^2}{Sxxs}=b_s\cdot Sxys,\quad Ses=Syys-S_{Rs}$$ （sの残差）

$Yt=\overline{y}_t+b_t(x_t-\overline{x}_t)$ についても同様に、$Sxxt=\sum x_{ti}^2-\frac{(\sum x_{tj})^2}{m}$

$$Syyt=\sum y_{ti}^2-\frac{(\sum y_{ti})^2}{m},\quad Sxyt=\sum x_{ti}y_{ti}-\frac{(\sum x_{ti})(\sum y_{tj})}{m},$$
$$b_t=\frac{Sxyt}{Sxxt},\quad S_{Rt}=\frac{(Sxyt)^2}{Sxxt}=b_t\cdot Sxyt,\quad Set=Syyt-S_{Rt}$$ （tの残差）

ｓ、ｔの全体のデータで考える。即ち、ｓとｔをプールする。

$$Sxx=Sxxs+Sxxt,\quad Syy=Syys+Syyt,\quad Sxy=Sxys+Sxyt$$
$$b=\frac{Sxy}{Sxx},\quad S_R=\frac{(Sxy)^2}{Sxx}=b\cdot Sxy$$
$$S_{R\Sigma}=S_{Rs}+S_{Rt},\quad S_{Dp}=S_{R\Sigma}-S_R$$ （非平行性）
$$Se=Syy-S_{R\Sigma}=Ses+Set$$
$$Ve=\frac{Se}{n+m-4},\quad V_{Dp}=\frac{S_{Dp}}{1},\quad V_{R\Sigma}=\frac{S_{R\Sigma}}{2}$$

F_{cal}で検定したものは、また、

$$t_{cal}=\frac{|b_s-b_t|}{\sqrt{V_e\cdot(\frac{1}{Sxxs}+\frac{1}{Sxxt})}}$$ が$t_{n+m-4}(\alpha)$に従うことを利用することと同じになる。
$t_{n+m-4}(\alpha)$は自由度$n+m-4$のt分布のα点である。

以上

1998.11.20 & 12.9

統計学演習問題

問題１　小学生 5000人の都市で虫歯保有率 Pを標本比率 Pで推定するとして、その誤差 dを95%信頼度で５%以下にしたい。何人調査すればよいか？
＜ヒント＞　pが不明であれば、p=0.5として標本の大きさ nを求めればよい。

1) ある地方で小学生 200人を無作為に選び、虫歯調査をしたら 137人が虫歯であった。この地方の小学生の虫歯保有率 pの95%信頼区間を求めよ。

問題２　次表は1985年、F県在住女性について栄養調査を実施し、その時得られた脂質栄養素摂取量(g)のデータである。

年　齢	人口	調査数	平均摂取量	標準偏差	比例配分	最適配分
10代	70,000	8	56.9	11.5		
20	167,000	255	51.2	13.5		
30	140,000	1156	49.4	13.4		
40	139,000	1424	47.6	12.4		
50	110,000	1367	44.9	12.8		
60	85,000	532	43.6	12.2		
70	67,000	96	40.5	11.0		
Total	778,000	4838	46.9	13.0	4838	4838

1) 実際の調査では、母平均μの95%信頼区間はいくらか？

2) n=4838について、比例配分(Propotional allocation)で各年齢層の調査数を求めよ。

3) n=4838を最適配分(Neyman allocation)で各年齢層の調査数を求めよ。更に、実際の調査数の問題点を批評せよ。

問題３　ある電気メーカーは蛍光灯の新製法を開発した。新製品から 25本を無作為に抽出し測定したら、平均寿命 1160時間であった。従来の製品仕様書によれば、平均寿命は 1130時間、標準偏差 80時間であった。新製法によって寿命が延びたといえるか 有意水準５%で検定せよ？

問題４　ある製品の融点がＡ、Ｂ両社で異なるかどうかを調べるため、両社からそれぞれ９個づつ製品をとって実験したところ次のようであった。この実験結果からＡ社とＢ社とでは融点が違うといえるか？　ただし、Ａ、Ｂ両社の融点の標準偏差はそれぞれ３，４であるとする。

Ａ社	149	147	152	148	149	145	154	156	154
Ｂ社	148	149	144	140	139	145	137	151	148

問題５　２種類の薬剤の効果を比較するため、それぞれ１０回ずつ実験したところ次のような得点があった。Ａ、Ｂ薬剤の効果は同じと考えてよいか有意水準 0.05で検定せよ。

A	5	8	7	6	6	7	7	8	5	6
B	7	6	5	4	8	5	5	6	3	4

問題６　次の数値はある製薬工場で錠剤をつくる工程から10分間隔で１粒宛１００個の錠剤を抜き出して、その重さを測定して得られたものである。（単位ｍｇ）
下の表を完成させ、正規分布の当てはめ、その適合性を検定せよ。

256	254	262	267	261	247	244	245	250	253
237	251	249	250	244	239	260	240	231	251
255	249	253	260	250	259	245	265	254	263
266	256	252	237	250	249	260	247	259	242
259	244	275	251	248	249	258	229	263	255
235	255	259	253	255	239	246	247	243	262
257	241	254	248	241	254	257	270	239	259
267	253	244	242	234	251	270	239	237	245
234	234	260	255	242	266	247	260	256	243
253	242	249	257	251	239	245	236	259	252

階級値x_i	観測度数f_i	期待度数np_i	$(f_i-np_i)^2/np_i$
229.5			
233.5			
237.5			
241.5			
245.5			
249.5			
253.5			
257.5			
261.5			
265.5			
269.5			
273.5			
計	100	100.0	

問題７　次の資料は、スチール缶再資源化率（ｙ）とアルミ缶再資源化率（ｚ）の年次推移データである。
直線回帰式　　$y=\bar{y}+b_1(x-\bar{x}),\quad z=\bar{z}+b_2(x-\bar{x})$

を求め、２つの資源率は同じ伸び率にあるとみて見てよいか？即ち、傾き b_1、b_2が同じと見なして良いか検定せよ。

年（19―）ｘ	89	90	91	92	93	94	95	96
ｙ	―	44.8	50.1	56.8	61.0	69.8	73.8	77.3
ｚ	42.5	42.6	43.1	53.8	57.8	61.1	65.7	70.2

問題８　ある地域の婦人５４人について栄養調査を実施し、２種の摂取栄養素と血中コレステロール濃度を測定した。年齢（Ａ）、蛋白質（Ｐ）、脂肪（Ｆ）、血中コレステロール（Ｃ）の相互の相関係数は次のとおりである。
蛋白質と脂肪の摂取量の影響を除いた年齢、血中コレセテロールの相関はどの程度か？　即ち、$r_{AC\cdot PF}$ はいくらか。

	A	P	F
P	−0.4865		
F	−0.5296	0.5784	
C	0.4737	−0.4249	−0.3135

問題９　ある種の降圧剤の一定濃度溶液を等分し、一方をそのまま（Ｓ）、他方に酸化剤を加え（Ｔ）１か月保存してから、それぞれの降圧作用 *ymmHg* を調べた。用量―反応（*D―R*）直線は平行しているとみていいか？

用量 (1/10ml)	x=log 用量	ys	yt
2	0.301	12	20
3	0.477	17	25
5	0.699	29	38
7	0.845	40	49
10	1.000	44	55
15	1.176	56	66
20	1.301	59	75
30	1.477	72	82
合計	7.276	329	410
平均	0.9095	41.125	51.250

統計学 Example 1　　　　　　　　　　　　　　　　　　　　出題:1998. 11. 11

問題1. 次のデータ1～5(n=100)について下の階級値で度数分布表を作成し,
棒グラフ(あるいはヒストグラム)をかきなさい.

データ1	データ2	データ3	データ4	データ5
15.26	15.43	15.56	15.44	15.48
15.12	15.29	15.31	15.73	15.41
15.23	15.39	15.51	15.37	15.21
15.34	15.54	15.32	15.43	15.21
15.39	15.81	15.16	15.41	15.70
15.16	15.54	15.48	15.21	15.32
15.25	15.67	15.44	15.46	15.27
15.37	15.21	15.53	15.19	15.30
15.44	15.27	15.28	15.27	15.26
15.39	15.32	15.55	15.34	15.64
15.33	15.39	15.50	15.40	15.57
15.18	15.27	15.31	15.65	15.33
15.47	15.24	15.50	15.20	15.23
15.73	15.60	15.29	15.27	15.33
15.43	15.20	15.40	15.57	15.35
15.46	15.48	15.28	15.61	15.51
15.26	15.14	15.38	15.34	15.29
15.36	15.38	15.31	15.22	15.37
15.52	15.26	15.25	15.23	15.40
15.42	15.32	15.38	15.45	15.36

度数分布表

no	階級下限値	階級上限値	度数
1	15.105	15.155	
2	15.155	15.205	
3	15.205	15.255	
4	15.255	15.305	
5	15.305	15.355	
6	15.355	15.405	
7	15.405	15.455	
8	15.455	15.505	
9	15.505	15.555	
10	15.555	15.605	
11	15.605	15.655	
12	15.655	15.705	
13	15.705	15.755	
14	15.755	15.805	
15	15.805	15.855	
		Total	100

統計演習.xls

統計学演習問題解答例　　　　　　　　　　　　　　　　　　　　1998.11.20 / 12.9

問題 1

1）　母集団(N)の虫歯保有率（母比率）p の 95 %信頼区間は 2 項分布の正規分布近似により、

$$\hat{p} - 1.96\sqrt{\frac{N-n}{N-1}\cdot\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + 1.96\sqrt{\frac{N-n}{N-1}\cdot\frac{\hat{p}\hat{q}}{n}}, \qquad \hat{q} = 1 - \hat{p}$$

(無限母集団に対しては，$\hat{p} - 1.96\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + 1.96\sqrt{\frac{\hat{p}\hat{q}}{n}}$)

誤差 $d \leqq 0.05$ であるから、

$d = 1.96\sqrt{\frac{N-n}{N-1}\cdot\frac{\hat{p}\hat{q}}{n}} \leq 0.05,$　　仮に $\hat{p} = \hat{q} = 0.5$ として，$\frac{N-n}{N-1}\cdot\frac{0.5^2}{n} \leq (\frac{0.05}{1.96})^2$

$$n \geq \frac{N}{(\frac{0.05}{1.96\times 0.5})^2(N-1)+1} = \frac{1}{(\frac{0.05}{1.96\times 0.5})^2(1-\frac{1}{N})+\frac{1}{N}} = 356.8, \quad (\frac{1}{N} \to 0 \text{とすれば}, \quad n \geq 384.2)$$

故に、N=5000 のとき，$n \geqq 357$ 例以上調査すればよい．

2）p の 95 %信頼区間は　$\hat{p} - 1.96\sqrt{\frac{\hat{p}\hat{q}}{n}} \leq p \leq +1.96\sqrt{\frac{\hat{p}\hat{q}}{n}}$　で与えられる．

$\hat{p} = \frac{137}{200} = 0.685$　であるから，

$$0.685 - 1.96\sqrt{\frac{0.685\times 0.315}{200}} \leq p \leq 0.685 + 1.96\sqrt{\frac{0.685\times 0.315}{200}}, \qquad \therefore 0.621 \leq p \leq 0.749$$

即ち，この地方の小学生の虫歯保有率 p の 95 %信頼区間は 62.1 %から 74.9 %の間といえる．

問題 2

年齢	人口	調査数	平均摂取量	標準偏差	比例配分	最適配分
10 代	70,000	8	56.9	11.5	435	396
20	167,000	255	51.2	13.5	1,038	1,109
30	140,000	1,156	49.4	13.4	871	922
40	139,000	1,424	47.6	12.4	864	847
50	110,000	1,367	44.9	12.8	684	692
60	85,000	532	43.6	12.2	529	510
70	67,000	96	40.5	11.0	417	362
Total	778,000	4,838	46.9	13.0	4,838	4,838

1）母平均 μ の 95 %信頼区間は 平均摂取量（標本平均値）を $\bar{x}$, 標準偏差を σ とすれば

$$\bar{x} - 1.96\cdot\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 1.96\cdot\frac{\sigma}{\sqrt{n}}, \quad 46.9 - 1.96\cdot\frac{13.0}{\sqrt{4838}} \leq \mu \leq 46.9 + 1.96\cdot\frac{13.0}{\sqrt{4838}}, \quad 46.5 \leq \mu \leq 47.3$$

2）　$\rho_i = \frac{N_i}{N} = \frac{n_i}{n}$, $N = 778{,}000, N_1 = 70{,}000, N_2 = 167{,}000, \cdots, N_7 \fallingdotseq 67{,}000$ より

$n_i = \frac{N_i \times n}{N}$ で求めればよい

３） 最適配分の場合は，

$\rho_i = \frac{N_i}{N}$より，$n_i = \frac{\rho_i \sigma_i}{\rho_1\sigma_1 + \rho_2\sigma_2 + \cdots + \rho_k\sigma_k} \cdot n$として求めればよい． $\sigma_1 = 11.5, \sigma_2 = 13.5, \cdots$

問題３

標準偏差σが既知であるので、正規分布 による検定法（８ページ,7.の 1)参考）を使用する

仮説は$H_0 : \mu = 1130$である $n = 25, \bar{x} = 1160, \sigma = 80$より $z = \frac{|\bar{x} - \mu_0|}{\frac{\sigma}{\sqrt{n}}} = \frac{1160 - 1130}{\frac{80}{\sqrt{25}}} = \frac{30}{\frac{80}{5}} = 1.875 < 1.96$

故に,仮説は有意水準 5 %で棄却されない 即ち,新製法によって蛍光灯の平均寿命が延びたとは言えない．

問題４

２つの平均値の差の検定である． 9 ページの 5)を参考にするが,この問題は標準偏差がともに既知であるので,正規分布による検定法を用いる．

一般に．

変数xが正規分布$N(\mu, \sigma_1{}^2)$に従い,変数yが同じく$N(\mu_2, \sigma_2^2)$に従うとき，

合成変数$x \pm y$は正規分布$N(\mu_1 \pm \mu_2, \sigma_1^2 + \sigma_2^2)$に従うという性質が分かっている．

従って,平均値についても変数$\bar{x} \pm \bar{y}$は正規分布$N(\mu_1 \pm \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$に従う

注） 変数$x + y$も$x - y$も分散は$\sigma_1^2 + \sigma_2^2$となることに注意する

$H_0 : \mu_1 = \mu_2$, $z = \frac{|\bar{x}_1 - \bar{y}_1|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{150.4 - 144.6}{\sqrt{\frac{4^2}{9} + \frac{3^2}{9}}} = 3.48 > 1.96.$

故に,有意水準（危険率)5%で有意差が認められる． 即ち,A社とB社とでは製品の融点に差があるといえる．

問題５

この問題では標準偏差が未知であるので, t 一分布による平均値の差の検定 （9 ページ 5)）を適用する．

$m = 10, \bar{x}_1 = 6.5, s_1^2 = \frac{1}{m-1}(\sum x_{1i}^2 - m\bar{x}_1^2) = \frac{1}{9}(433 - 422.5) = \frac{10.5}{9}$,

$n = 10, \bar{x}_2 = 5.3, s_2^2 = \frac{1}{n-1}(\sum x_{2i}^2 - n\bar{x}_2^2) = \frac{1}{9}(301 - 280.9) = \frac{20.1}{9}$,

$t = \frac{|6.5 - 5.3|}{\sqrt{\frac{9 \times s_1^2 + 9 \times s_2^2}{m + n - 2}(\frac{1}{m} + \frac{1}{n})}} = \frac{1.2}{\sqrt{0.34}} = 2.06 < 2.101 = t_{18}(0.05)$,

ここに,$t_\phi(\alpha)$は自由度ϕ,有意水準αのt分布値である．

故に,有意差は認められない．

即ち,A,B の２種類の薬剤の効果は同じと考えられる．

問題６ *Excel* の関数 *NORMDIST*(*x*,平均値,標準偏差,*TRUE*) を用いて正規分布関数の値 （累積確率） を求める．

階級値	観測度数 f_i	累積度数	累積確率	累積期待値	期待度数 np_i	$(f_i-np_i)^2/np_i$
229.5	2	2	0.02458	2.5	2.5	
233.5	4	6	0.05874	5.9	3.4	0.002
237.5	9	15	0.12225	12.2	6.3	1.157
241.5	9	24	0.22298	22.3	10.1	0.120
245.5	13	37	0.35924	35.9	13.6	0.026
249.5	16	53	0.51642	51.6	15.7	0.006
253.5	16	69	0.67106	67.1	15.5	0.016
257.5	13	82	0.80080	80.1	13.0	0.000
261.5	10	92	0.89365	89.4	9.3	0.053
265.5	5	97	0.95031	95.0	5.6	0.365
269.5	2	99	0.97979	98.0	3.0	
273.5	1	100	1.00000	100.0	1.3	
計	100	100			100.0	1.745

全データの平均値$\bar{x} = \frac{1}{n}\sum x_i = 251.09$,標準偏差$s = \sqrt{\frac{1}{n-1}\sum(x_i^2 - n\bar{x}^2)} = 9.9586$

*Excel*の分析ツールの基本統計量で求められる．

検定H_0：理論分布 （この場合,正規分布） に当てはまる （適合度の検定）

$\chi_0^2 = \sum \frac{(f_i - np_i)^2}{np_i}$,自由度$\phi = m - 1 - 2 = 6$ のχ^2分布に従う

$\chi_0^2 = 1.745 < \chi_6^2(0.05) = 12.592$, 故に正規分布に適合するという仮説を否定できない．

即ち,正規分布すると考えられる．

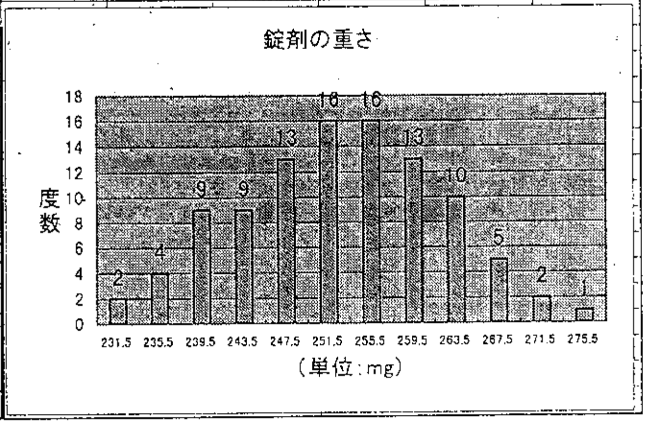
自由度のうち,3を差し引いたのは,条件が3つあるからである.一つは期待値の合計値を100に合わせたこと,平均値,標準偏差を使用したことによる．

適合度は $P(\chi_6^2 \geq 1.745) = 0.94159$ (94.1%)である．

注） *Excel*の関数*CHIDIST*(*x*,自由度)= *p* （確率） を使用する．

問題6	Excelで ツール(T)→分析ツールのヒストグラムを使用								
256	254	262	267	261	247	244	245	250	253
237	251	249	250	244	239	260	240	231	251
255	249	253	260	250	259	245	265	254	263
266	256	252	237	250	249	260	247	259	242
259	244	275	251	248	249	258	229	263	255
235	255	259	253	255	239	246	247	243	262
257	241	254	248	241	254	257	270	239	259
267	253	244	242	234	251	270	239	237	245
234	234	260	255	242	266	247	260	256	243
253	242	249	257	251	239	245	236	259	252

軸	データ区間	頻度	累積 %
	227.5	0	.00%
1	231.5	2	2.00%
2	235.5	4	6.00%
3	239.5	9	15.00%
4	243.5	9	24.00%
5	247.5	13	37.00%
6	251.5	16	53.00%
7	255.5	16	69.00%
8	259.5	13	82.00%
9	263.5	10	92.00%
10	267.5	5	97.00%
11	271.5	2	99.00%
12	275.5	1	100.00%
	合計	100	



列1	
平均	251.09
標準誤差	0.99585961
中央値(ﾒｼﾞｱﾝ)	251.5
最頻値(ﾓｰﾄﾞ)	259
標準偏差	9.958596104
分散	99.17363636
尖度	−0.599801532
歪度	−0.06320175
範囲	46
最小	229
最大	275
合計	25109
標本数	100
信頼区間(95.0%)	1.976001873

階級下限	累積度数	逆推定値	累積確率	理論累積数	理論度数
231.5	2	230.6376	0.02458	2.5	
235.5	6	235.6067	0.05874	5.9	3.4
239.5	15	240.7686	0.12225	12.2	6.3
243.5	24	244.0562	0.22298	22.3	10.1
247.5	37	247.7852	0.35924	35.9	13.6
251.5	53	251.8396	0.51642	51.6	15.7
255.5	69	256.028	0.67106	67.1	15.5
259.5	82	260.2058	0.8008	80.1	13
263.5	92	265.0826	0.89365	89.4	9.3
267.5	97	269.82	0.95031	95	5.6
271.5	99	274.2571	0.97979	98	3
275.5	100	#NUM!	0.99288	99.3	1.3

Sheet1の作図の手順(マウス指示:カーソルを左クリック)

1. 作図

E13

ツールバーからｸﾞﾗﾌｳｲｻﾞｰﾄﾞ

縦棒→次へ

データ範囲(D):　C15からC26までドラッグ(左ｸﾘｯｸを押したまま移動)→系列

名前(N):　"錠剤の重さ"と入力

項目軸ラベルに使用(T):　B15からB26までドラッグする

凡例　凡例を表示する(S)を空白にする

タイトルとラベル　x/項目軸(C):　"(単位:mg)"と入力

y/数値軸(V):　"(度数)"と入力

データラベル　値を表示する(V):　ｸﾘｯｸした後完了をクリック

2. 編集作業　図が表示されるので図内にカーソルをおいてドラッグして配置を移動

図中のX軸の下の数値にカーソルを合わせる:右クリック

軸書式設定:　フォント:　サイズ(S."6"としOKをクリック

表示形式 数値小数点以下の桁数(D):　"1"

図中のY軸の左の"度数"にカーソルを合わせる:右クリック

軸書式設定:　配置:　方向:　縦文字に合わせてOKをクリック

Sheet1

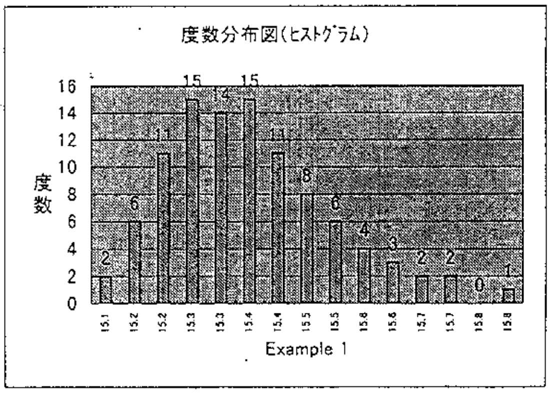
問題7の解答:		xに対するy(y=ax+b), xに対するz(z=a'x+b') の関係から回帰分析を行う								
dis-paralle	1	2	3	4	5	6	7	8		
x	89	90	91	92	93	94	95	96		
y	0	44.8	50.1	56.8	61.0	69.8	73.8	77.3		
z	42.5	42.6	43.1	53.8	57.8	61.1	65.7	70.2		
xx	7921	8100	8281	8464	8649	8836	9025	9216		
yy		2007.04	2510.01	3226.24	3721	4872.04	5446.44	5975.29		
zz	1806.25	1814.76	1857.61	2894.44	3340.84	3733.21	4316.49	4928.04		
xy	0	4032	4559.1	5225.6	5673	6561.2	7011	7420.8		
xz	3782.5	3834	3922.1	4949.6	5375.4	5743.4	6241.5	6739.2		
yz	0	1908.48	2159.31	3055.84	3525.8	4264.78	4848.66	5426.46		
1)	Ns	7		Nt	8			15		
	mean x	93		mean x	92.5		mean x+x	92.73		
	mean y	61.94286		mean z	54.6		mean y+z	58.03		
	SSx	60571.0		SSx	68492.0		SSx+x	129063.0		
	SSy	27758.06		SSz	24691.64		SSy+z	52449.7		
	SSxy	40482.7		SSxz	40587.7		SSxy+xz	81070.4		
	sxx: Sxxs	28	①	sxx: Sxxt	42.0	②	Sx+x	80.2065	⑳	
	syy: Syys	899.6371	③	szz: Syyt	842.4	④	Syy+Szz	1937.487	⑨	
	sxy: Sxys	157.9	⑤	szx: Sxyt	183.7	⑥	Sxy+xz	353.5715	⑲	
	bs	5.64	⑤／①	bt	4.37	⑥／②	b'	4.41	⑲／⑳	
	Srs	890.44	⑦	Srt	803.47	⑧	SR	1558.64	⑩=⑲＊⑲/⑳	
	Ses	9.197	③−⑦	Set	38.89	④−⑧	SE	378.846	⑪	
			分散分析表							
2).	要因		SS	df	Ms	F	Sxx	70.0	⑫=①+②	
	R(共有直線性)	⑱	1667.01	1			Syy	1742.0	⑬=③+④	
	DP(非平衡性)	⑮−⑱	26.89	1	26.89	8.15	Sxy	341.6	⑭=⑤+⑥	
	RΣ(全直線性)	⑮	1693.90	2			b	4.88	⑭／⑫	
	e(残差)	⑯	48.10	11	4.37		Srσ	1693.9	⑮=⑦+⑧	
	Y(全体)	⑬	1742.00	13			Se	48.087	⑯=⑬−⑮	
							Se2	75.0	⑰=⑬−⑱	
							⑱R共線性	1667.008	⑭＊⑭/⑫	

問題8の解答:					
	A	P	F	r(AC・F)	0.381969
P	−0.4865			r(AP・F)	−0.26039
F	−0.5296	0.5784		r(CR・F)	−0.31444
C	0.4737	−0.4249	−0.3135	r(AC・PF)	0.327422
	r(AC)	0.4737	r(AP)	−0.4865	r(CP)	−0.4249
	r(AF)	−0.5296	r(AF)	−0.5296	r(CF)	−0.3135
	r(CF)	−0.3135	r(PF)	0.5784	r(PF)	0.5784
	r(AC・F)	0.381969	r(AP・F)	−0.26039	r(CP・F)	−0.31444

統計学 Example 1(1998.11.20)

ﾃﾞｰﾀ 1	ﾃﾞｰﾀ 2	ﾃﾞｰﾀ 3	ﾃﾞｰﾀ 4	ﾃﾞｰﾀ 5
15.26 D	15.43 G	15.56 J	15.44 G	15.48 H
15.12 A	15.29 D	15.31 E	15.73 M	15.41 G
15.23 C	15.39 F	15.51 I	15.37 F	15.21 C
15.34 E	15.54 I	15.32 E	15.43 G	15.21 C
15.39 F	15.81 P	15.16 B	15.41 G	15.70 L
15.16 B	15.54 I	15.48 H	15.21 C	15.32 E
15.25 C	15.67 L	15.44 G	15.46 H	15.27 D
15.37 F	15.21 C	15.53 I	15.19 B	15.30 D
15.44 G	15.27 D	15.28 D	15.27 D	15.26 D
15.39 F	15.32 E	15.55 I	15.34 E	15.64 K
15.33 E	15.39 F	15.50 H	15.40 F	15.57 J
15.18 B	15.27 D	15.31 E	15.65 K	15.33 E
15.47 H	15.24 C	15.50 H	15.20 B	15.23 C
15.73 M	15.60 J	15.29 D	15.27 D	15.33 E
15.43 G	15.20 B	15.40 F	15.57 J	15.35 E
15.46 H	15.48 H	15.28 D	15.61 K	15.51 I
15.26 D	15.14 A	15.38 F	15.34 E	15.29 D
15.36 F	15.38 F	15.31 E	15.22 C	15.37 F
15.52 G	15.26 D	15.25 C	15.23 C	15.40 F
15.42 G	15.32 E	15.38 F	15.45 G	15.36 F

階級中央値	階級上限値	度数	
15.13	15.155	2	A
15.18	15.205	6	B
15.23	15.255	11	C
15.28	15.305	15	D
15.33	15.355	14	E
15.38	15.405	15	F
15.43	15.455	11	G
15.48	15.505	8	H
15.53	15.555	6	I
15.58	15.605	4	J
15.63	15.655	3	K
15.68	15.705	2	L
15.73	15.755	2	M
15.78	15.805	0	N
15.83	15.855	1	P
		100	TOTAL



<資料１> F検定

研究例１　障害児の発達評価に及ぼす療育経験の効果
・・・・２要因分散分析における主効果と交互作用の分析・・・・

２×３ 要因

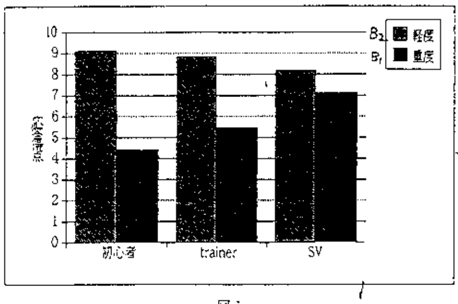


図1

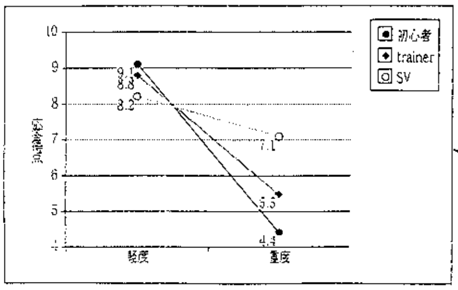


図2

データファイル作成 ――――→ 二要因分散分析の計算

実際はＳＰＳＳ，ＳＡＳ等で処理する

変動因	平方和(SS)	自由度(df)	平均平方(MS)	E(MS)
要因A	$n\sum(\bar{X}_{\cdot j}-\bar{X}_{\cdot\cdot})^2$	$k-1$	$n\sum(\bar{X}_{\cdot j}-\bar{X}_{\cdot\cdot})^2/(k-1)$	$\sigma^2+\frac{n\sum(\mu_{\cdot j}-\mu)^2}{k-1}$
要因B	$k\sum(\bar{X}_{i\cdot}-\bar{X}_{\cdot\cdot})^2$	$n-1$	$k\sum(\bar{X}_{i\cdot}-\bar{X}_{\cdot\cdot})^2/(n-1)$	$\sigma^2+\frac{k\sum(\mu_{i\cdot}-\mu)^2}{n-1}$
誤差	$\sum\sum(X_{ij}-\bar{X}_{\cdot j}-\bar{X}_{i\cdot}+\bar{X}_{\cdot\cdot})^2$	$(k-1)\times(n-1)$	$\sum\sum(X_{ij}-\bar{X}_{\cdot j}-\bar{X}_{i\cdot}+\bar{X}_{\cdot\cdot})^2\div(k-1)(n-1)$	σ^2
全体		$kn-1$		

両要因の効果を検定するには誤差平均平方を分母としてF比を計算すればよい。

分散分析表

S.V.	S.S.	d.f.	M.S.	F	
A 療育者の水準（3）	12.5920	2	6.2960	1.36	ns
B 障害の程度（2）	217.9117	1	217.9117	47.21	＊＊
A×B 交互作用	49.4274	2	24.7137	5.35	＊＊
SUB 被験者	470.7770	102	4.6155		
計	750.7081	107	nsP＜.10	＊p＜.05	＊＊P＜.01

統計的有意性の判定表

pの値	マーク	文章中の表現
p＞.10	なし	有意でない
.05＜p＜.10	†	有意傾向である
p＜.05	＊	（5％水準で）有意である
p＜.01	＊＊	（1％水準で）有意である

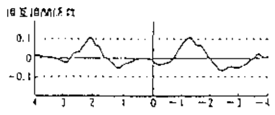
（注）マークを論文に記載するときは肩付きとする。

相互作用の分析表

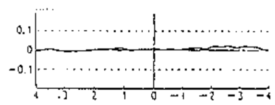
S.V.	S.S.	d.f.	M.S.	F	
A at B1	55.9260	2	27.9630	6.06	＊＊
A at B2	6.0934	2	3.0467	0.66	ns
B at A1	10.4832	1	10.4832	2.27	ns
B at A2	85.0956	1	85.0956	18.44	＊＊
B at A3	171.7603	1	171.7603	37.21	＊＊
SUB	470.7770	102	4.6155		
Total	750.7081	107	nsP＜.10	＊p＜.05	＊＊P＜.01

<資料２>

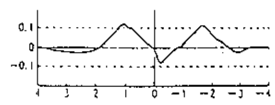
研究例２　新生児期における非言語的な母子相互交渉（エントレインメント）
・・・・相互相関関数による分析・・・・・



(a) 母親が新生児に自由に語りかけた場合



(b) コンピュータ合成音を新生児に聞かせた場合

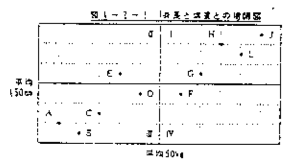


(c) 成人同士の会話の場合

図3

相関分析の基礎

◆ ２つの変数ＸとＹについて、Ｘの値が決まればＹの値が決まるというわけには行かないが、両者の間に直線的な関連性が認められるとき「ＸとＹの間には相関関係がある」といい、相関関係の程度を示す数値を単相関係数（ピアソンの偏差積率相関係数ともいう）という。単相関係数は－１から＋１の間の値を取る。以下に、体重と身長の関係について、単相関係数を求める具体例を示す。



学生	x_i	y_i	$x_i-\bar{x}$	$y_i-\bar{y}$	$(x_i-\bar{x})(y_i-\bar{y})$	$(x_i-\bar{x})^2$	$(y_i-\bar{y})^2$
A	45	146	−5	−4	20	25	16
B	46	145	−4	−5	20	16	25
C	47	147	−3	−3	9	9	9
D	49	149	−1	−1	1	1	1
E	48	151	−2	1	−2	4	1
F	51	149	1	−1	−1	1	1
G	52	151	2	1	2	4	1
H	53	154	3	4	12	9	16
I	54	153	4	3	12	16	9
J	55	155	5	5	25	25	25
計	Σx_i 500	Σy_i 1500	$\Sigma(x_i-\bar{x})$ 0	$\Sigma(y_i-\bar{y})$ 0	$\Sigma(x_i-\bar{x})(y_i-\bar{y})$ 98	$\Sigma(x_i-\bar{x})^2$ 110	$\Sigma(y_i-\bar{y})^2$ 104

$n=10$，$\bar{x}=\Sigma x_i/n=500\div10=50$

$\bar{y}=\Sigma y_i/n=1500\div10=150$

$$r=\frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\Sigma(x_i-\bar{x})^2\times\Sigma(y_i-\bar{y})^2}}=\frac{98}{\sqrt{110\times104}}=0.9162$$

◆ 相関係数には、単相関係数、偏相関係数、ノンパラメトリックの相関（ファイ係数、独立係数、相関比、順位相関など、いろいろある。

◆ ある集団における変数相互の相関関係を推定、または相関関係があるかどうかの検定を行うとき、その集団の一部を調べ結論を導くことがある。これが統計的推定であるが、以下の点に留意すること。

・rの有意性検定はr＝0でないことを検定する。
・rの有意性は"相関が強い"ということを意味するわけではない。
・相関の強さは下の表により経験的に判定する。

負の相関	相関の強さの判定	正の相関
−1		＋1
	強い相関がある	
−.7		＋.7
	中程度の相関がある	
−.4		＋.4
	弱い相関がある	
−.2		＋.2
	ほとんど相関がない	
0		0

研究例3　　共同注意の行動的指標の構造
・・・・因子分析による検討・・・・

		Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
1	指さし表出（マンド）	$\boxed{.72088}$	.49878	.16296	-.01956	.02134
	指さし表出（タクト）	$\boxed{.71253}$	.50841	.17314	-.00473	-.02420
	指さし理解（視野外）	$\boxed{.67195}$	.29104	.27132	.05555	.26265
	指さし理解（視野内対象）	$\boxed{.66312}$	.13230	.26746	.10487	.26020
	referencial looking（マンド）	$\boxed{.64853}$	.37037	.23029	.28173	-.06464
	指さし理解（視野内方向）	$\boxed{.64441}$	.17548	.19503	.10189	.45089
	referencial looking（タクト）	$\boxed{.62539}$	.40390	.20513	.27640	-.13820
	指さし表出（応答）	$\boxed{.62481}$	.32590	.25882	.03512	-.01070
	referencial looking（応答）	$\boxed{.53101}$	.21286	.37116	.13783	.05514
	視線追従（視野外）	.51140	.16929	.50866	.08444	.11928
	追視（視野外）	.30061	.23608	.27803	.21792	.29508
	showing（応答）	.36555	$\boxed{.71812}$	.11052	.16192	.22861
	showing（自発）	.31180	$\boxed{.70465}$	.16132	.20355	.17354
	giving（応答）	.60511	$\boxed{.60736}$	.23184	.06852	.10480
	giving（自発）	.54478	$\boxed{.57761}$	.24337	.08397	.12424
	アイ・コンタクト	-.02746	-.06806	-.02257	-.01469	.04775
3	視線追従（視野内対象）	.35113	.15912	$\boxed{.74130}$	.13032	.10343
	視線追従（視野内方向）	.28932	.21521	$\boxed{.70876}$	.14510	.20361
4	視線指示（タクト）	.24197	.17567	.08206	$\boxed{.70725}$	-.04973
	視線指示（マンド）	.11159	.08790	.00961	$\boxed{.69419}$	.09373
	指さし理解（手を見る）	-.09089	.00569	.11434	.35094	.12466
	追視（視野内）	.06942	.03974	.13294	.12232	.30594
	寄与率	50.0	7.6	6.4	4.6	4.3
	累積 〃	50.0	57.7	64.1	68.1	73.0

全被験者の素点データをもとに、バリマックス回転、主因子法による項目間の因子分析を行った結果、以下のように分類された。

第1因子：視覚対象を介した3者関係
（指差しの理解・指差し表出・交互凝視）
・・・11～19カ月の間に出現

第2因子：手に握った触覚的な対象を介した2者関係
（showing・giving）
・・・11カ月頃出現

第3因子：母子の2者関係に見られる視野内の視線追従
（視野内の視線追従）
・・・9カ月以降に出現

第4因子：発達初期に現れ、後に消えていく行動
（視線による指示・指さしたとき手を見る）
・・・5～11カ月の間に出現

◆ 因子分析の基本的な考え方

因子分析は、多変量データを集約的に記述し、その相関的構造を明らかにすることを目的とした多変量解析法の一つである。

その考え方は「相関」（correlation）に基づいている。すなわち、相関のある変数どうしはいっしょに動く。一方の値が大きくなれば規則的に他方の値も大きくなる（または小さくなる）という関係にある。ここで、どうしてそれらの変数がそうかんするかを考えると、同じ「因子」に支配されているからと考えることが出来る。こうして変数感の相関関係に基づいて因子を探しにゆく。

◆ 一般的な手続き

A．質問調査紙の作成と調査の実施

1. 質問項目の開発（内容の収集、内容の選別、表現のチェック、逆転・ダミー）
2. 評定尺度の作成（単極尺度か双極尺度か、ポイント数、文か数字）
3. 質問紙の編集（フェースシート項目、項目の掲載順序、関連質問、謝辞など）
4. 対象者の確保（項目数の2～5倍、実施期間、回答者の選定）
5. 調査の実施（インドームドコンセント、プライバシー保護,回収方法など）

B．因子分析実行マニュアル

1. データファイルの作成（回答のチェック、回答の得点化、入力、保存）
2. 基本統計量の計算と不良項目のチェック（ソフト起動、データ読み込み、実行）
3. 1回目の因子分析（標準得点化、共通性の初期値設定、因子の抽出法、実行）
4. 2回目の因子分析（因子抽出数と因子軸の回転法設定、実行）
5. 因子の解釈と命名（所定の因子負荷量以上の項目をマーク、共通性の推理・命名
6. 因子得点の利用（因子得点を用いた分散分析、回帰分析、他のデータとの相関）

研究例4　　共同注意の成立時期
・・・・クラスター分析による検討・・・・

予備調査1　外部基準との相関

次に、これらの共同注意行動と標準化された外部基準（遠城寺式発達検査）との相関(table.2)を見ると、出現時期の遅い行動ほど、外部基準との相関が高いことが分かる。これは、出現時期の遅い行動が下降することなく伸びていくいくのに比して、発達初期に現れる行動は、はじめから通過率が高かったり、後に衰退していくからである。このことから、table.2 の中で外部基準との相関が高い行動が、共同注意行動の発達の指標として妥当であると言うことができるが、一方で、外部基準との相関の低い初期の行動は、共同注意行動の前提となる行動と見ることができる。

table.2　外部基準との相関

	移動運動	手の運動	対人関係	発語	言語理解
指さし理解(手を見る)	-.01	-.01	-.01	-.03	-.06
アイ・コンタクト	-.04	.04	.02	.02	-.03
視線指示(マンド)	.12**	.12**	.16**	.0800	.13**
視野内追視	.12**	.12**	.14**	.15**	.13**
視線指示(タクト)	.28**	.27**	.32**	.24**	.30**
視野外追視	.42**	.38**	.43**	.35**	.40**
視線追従(視野内方向)	.50**	.48**	.51**	.44**	.51**
視線追従(視野内対象)	.54**	.51**	.54**	.47**	.54**
showing(自発)	.53**	.51**	.56**	.45**	.55**
showing(応答)	.57**	.52**	.59**	.45**	.57**
交互凝視(応答)	.55**	.52**	.58**	.50**	.57**
指さし理解(視野内方向)	.57**	.53**	.57**	.48**	.58**
視線追従(視野外)	.58**	.55**	.59**	.50**	.59**
指さし理解(視野内対象)	.59**	.56**	.58**	.51**	.61**
交互凝視(マンド)	.65**	.62**	.67**	.57**	.65**
交互凝視(タクト)	.65**	.64**	.68**	.56**	.67**
指さし理解(視野外)	.69**	.58**	.71**	.62**	.70**
giving(自発)	.70**	.68**	.71**	.60**	.71**
指さし表出(応答)	.74**	.73**	.73**	.69**	.75**
giving(応答)	.76**	.72**	.76**	.65**	.76**
指さし表出(タクト)	.77**	.73**	.75**	.68**	.76**
指さし表出(マンド)	.78**	.75**	.77**	.70**	.75**

** : $p < .01$

予備調査2　指さし理解の発達過程

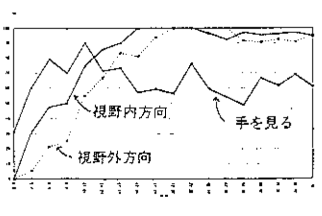
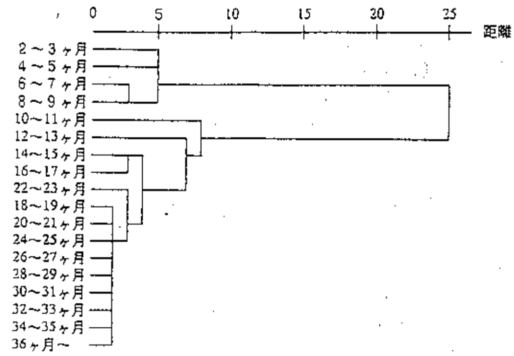


Fig.3　指さしの理解

指さし理解の発達メカニズム

1　生態学的メカニズム
（指差した指を見る）
↓
2　幾何学的メカニズム
（視野内の方向を見る）
↓
3　表象的メカニズム
（視野外の方向をみる）

クラスター　結果；完全連結法を使用した系統図（デンドログラム）



◆ クラスター分析とは、対象間の距離にもとづき、それらをいくつかの群（クラスター）に分類する手法。対象の分類基準（外部基準）がない場合の多変量解析法の一つとみなされる。
生物学では、クラスター分析は動植物の分類に、医学では、病名とその進行状態を識別する時などに使われる。
市場調査でも、類似した購買傾向を持つ人々を識別して、将来の販売戦略の目標をより効率的に定めるために使われる。

◆ 基礎知識および分析の実際は、ＳＰＳＳ（ProfessionalStatistics）等を参照のこと

ご記入されている方について

（１）障害を持たれているお子さんと貴方とはどのような間柄にありますか？下の項目から、当てはまるものに○をつけてください。

（　　）父親　（　　）母親　（　　）兄弟・姉妹
（　　）同居している親族　（　　）同居していない親族　（　　）その他

（２）あなたの年齢をお書きください。　（　　　　）歳

お子さんについて

（３）お子さんは現在何歳ですか？　（　　　　）歳

（４）お子さんの性別をお答えください。　（　　）男　（　　）女

（５）言語的な理解だけではなく、ジェスチャーや手話、サポート機器を用いても結構ですが、お子さんは対人的なコミュニケーションがどの程度可能ですか。当てはまると思う程度の（　）に○を付けてお答え下さい。

（　　）自分の意志を相手に伝えたり、相手の言葉を理解することに難はない。
（　　）自分の意志を伝えることはできるが相手の言葉を理解することが難しい。
（　　）自分の意志を伝えることは難しいが、相手の言葉は理解できる。
（　　）自分の意志を伝えることも、相手の言葉を理解することもまだ難しい。

（６）あなたの目から見て、お子さんにはどのくらい注意や世話が必要であると思われますか？

（　　）常時必要とする。
（　　）常時ではないが、頻繁に必要。
（　　）短時間なら目を離せるが、かなり必要。
（　　）危険な時や気づいた時以外は、あまり必要としない。
（　　）全然必要としない。

（７）お子さんの障害や問題について、療育手帳や身障者手帳をお持ちの方や、医学的な診断名が出されている場合、差しつかえなければ、その診断名をお答えください。（例：視覚障害、聴覚障害、ダウン症など。特に診断名のない方やご存知ない方は空白のままで結構です。）

（　　　　　　　　　　　　　　　　　　　　）

遊び場所について

（８）下に挙げる場所のうち、お子さんの主たる遊び場所はどこですか？第１位から第３位まで順位をつけて、数字をご記入ください。

１．家の中
２．家の周囲
３．子どもの友達の家
４．親戚・知人宅
５．公園・広場
６．その他（具体的にお書き下さい。　　　　　　　　）

第１位　場所　（　　　　）
第２位　場所　（　　　　）
第３位　場所　（　　　　）

（９）上記のような場所でお子さんが遊んでいる時（遊びに行かせる場合）、あなたはどのようになさっていますか？また、なぜそのようにされるのか、理由がある場合にはどのような理由があるのか、差しつかえがなければ教えてください。

（　　）いつも一緒につきそう（遊ぶ）。
（それは何故ですか？：　　　　　　　　）

（　　）場所によって付き添う。
（どのような場所ですか？：　　　　　　　　）
（それは何故ですか？：　　　　　　　　）

（　　）いつでも一人で遊びに行かせられる（遊ばせられる）。