分析系検査管理論演習　「統計解析の基礎」

5月 15日（水）

担当：加藤一夫（愛知学院大学歯学部）

＜統計的仮説の検定＞

1．分布の中心を表すもの（代表値）

①意味が明瞭　②計算が簡単　③数学的発展性　④統計的推論に有用

(1)平均（mean）

μ（母平均）

$\bar{x}$（標本平均）

相加(算術)平均(arithmetical mean)ともよばれ、最も多用される代表値。

データの数値(変量の値)の合計($\Sigma x = x_1 + x_2 + \cdots\cdots + x_n$)をデータの個数($n$)で割ることにより求められる。すべてのデータの数値を利用するため、極端な数値の影響を受けやすい。平均値と呼ばれるものには、これ以外に相乗(幾何)平均、調和平均などがあるがいずれもあまり一般的ではない。

(2)メジアン(median)Md

データを大きさの順に並べたときその中央に位置する数値で中央値ともいう。

算出方法はデータの個数により次の二通りがある。

①データ数が奇数($2n+1$)とき、$n+1$番目のデータ。

②データ数が偶数($2n$)とき、n番目と$n+1$番目のデータの平均値。

(3)モード(mode)Mo

データの中に最もよく出現する値。最頻値ともよばれる。度数分布表では最大度数の階級値を示す。この場合、モードは階級の取り方で変化する。また二峰性の分布の場合、高い方の頂点からそれぞれ第1モード、第2モードという。データの1部を利用する。

(4)代表値の選択と利用

分布の形が左右対称に近い釣り鐘型(正規分布)では、平均値、メジアン、モードは互いに近似した値をとる。この場合の代表値には、計算処理が容易で、統計的推論に優れた平均値が使用される。一方、分布の形が非対称型やL字型であったり、データのばらつきが大きい（平均値より標準偏差が大きい）と、平均値の代表性は低下し意味もわかりにくくなる。こうした場合には、変量の順序のみに関係するメジアンの方が分布から外れた数値の影響を受けにくいため、平均値より代表性に優れている。

モードもメジアン同様、極端な数値の影響を受けにくい。規格品の導入を考えるといった場合には、モードの利用が適している。

2．データの広がりを示すもの（散布度）

(1)範囲(range)

(2)分散(variance)

$$\mu^2 = \frac{1}{n}\Sigma(x_i - \bar{x})^2 \quad \text{（母分散 population v.）}$$

母集団の分散には σ^2 が使われるが、標本集団の分散を母分散の推定値に用いるとき、そのままでは標本分散は母分散の($n-1$)／n 倍になるため、標本集団には分母の n を $n-1$ に代えた S^2 を用いる。

$$S^2 = \frac{1}{n-1}\Sigma(x_i - \bar{x})^2 \quad \text{（不偏分散 sample v. ／ 標本分散 unbiased v.）}$$

(3)標準偏差(S.D.: standard deviation)

$$\sigma = \sqrt{\frac{1}{n}\Sigma(x_i - \bar{x})^2}$$

分散と同様、標本集団には分母を $n-1$ に代えた標本標準偏差 S を用いる。

(4)標準誤差(S.E.M.: standard error of mean)

「標本調査」における平均値の信頼度を判断する指標で、標本数が多くなると標準誤差は小さくなる(信頼度は増す)。σが不明であれば代わりにSを用いる。

$$\mathrm{S.E.} = \frac{\sigma}{\sqrt{n}}$$

平均μ、標準偏差σの母集団を仮定し、そこから n 個の標本を抜き出し平均を求める。このとき、標本平均の分布は、どのような母集団についても、平均μ、標準偏差 $\sigma / \sqrt{n}$ の正規分布になることが証明されている(中心極限定理)。

標準誤差は、この標本平均のバラツキの程度を示す。

(5)変動係数(C.V.: coefficient of variation)

平均値に対する標準偏差の割合を示す係数で、平均値の異なるデータ間の散布度を比較するのに用いる。

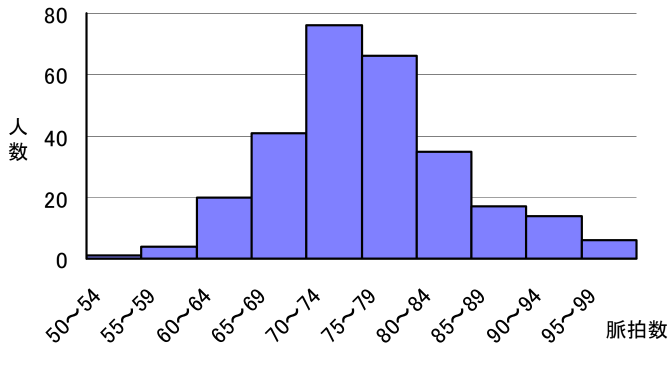
例えば、平成7年度の男子の身長(平均±標準偏差)は、幼稚園児(5歳)111.0 ± 4.70cm、高校生(17歳)170.8 ± 5.64cmである。両者の変動係数4.2 %と3.3 %から、幼稚園児の方が身長に関して個体差が大きい。

$$\mathrm{C.V.} = \frac{\sigma}{\bar{x}} \times 100\ (\%)$$

3. 測定データの確率分布

(1)正規分布の性質

次の図はある学校に通う学生280名の心拍数の度数分布を表したものである。これを相対度数分布に変換し、全体の面積が1となるように階級の幅を使って縦軸のスケールを変える。そして、データ数を増やしながら階級の幅を小さくしていくと、次第にスムーズで釣り鐘状の「確率密度関数」に近づいていく。度数分布が標本集団である実測値の分布を示すのに対して、確率密度関数は予想される母集団の分布の確率を理論的に示したもので、その曲線下の面積は1となる。正規分布(normal distribution)は連続変量の分布を表す代表的な確率密度関数であり、心拍数以外にも身長や血圧など生体の測定値や測定誤差など多くの変量が正規分布に適合するといわれている。



正規分布は両側に無限に広がる一峰性の分布で、左右対称なため平均値、メジアン、モードはいずれも一致する。正規分布の形は平均値μと分散 σ^2(標準偏差σ)により完全に決まることから、記号を用いて $N(\mu, \sigma^2)$と表記している。正規分布と標準偏差の関係には、μを中心にした±1σの範囲にデータ全体の68.3%、±2σの範囲では95.5%を含むというたいへん重要な性質がある。よく臨床検査などで集団の平均値を中心に標準偏差の2倍以内を正常域として設定することがあるが、それもこの性質を利用しているといえる。

データが正規分布するかどうかは、ヒストグラムでも見当はつくが、累積相対度数を求め正規確率紙にプロットすることで容易に判定することができる。

(2)正規分布の利用

一般の正規分布 $N(\mu, \sigma^2)$は平均値と標準偏差の値の違いによってさまざまな形になる。そのうちで平均値0、標準偏差1の正規分布を標準正規分布 $N(0,1^2)$とよんでいる。これは一般の正規分布の平均値を－μだけ移動し、標準偏差が1になるように横軸のスケールを変えたものであるから、一般の正規分布の値 x は次の式で標準正規分布の値 Z に変換、すなわちデータを標準化することができる。

$$Z = \frac{x - \mu}{\sigma}$$

標準正規分布の横軸上の値0と Z で区切られた曲線下の面積について、いろいろな Z の値について計算した結果をまとめたものを正規分布表という。正規分布は左右対称なので、正規分布表では Z=∞のときの面積は 0.5 というように片側の数値が示されている。データの標準化により、この正規分布表は一般の正規分布に応用することができる。

(3)その他の理論分布

保健医療領域で扱うほとんどの場合では母集団の母平均μと母標準偏差σは未知である。しかし母集団が正規分布しなくても、標本集団が十分大きい（データ数 30 以上）ときには、標本平均 $\bar{x}$ とその標準偏差（前述の標準誤差）を母集団の推定値として代用することができる。このとき値 Z は標準正規分布をする。

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

これを利用すれば、標本平均からある確率のもとで母平均の存在する範囲を推測することができる。これを区間推定とよび、確率としては 95%がよく用いられる。この場合、正規分布は左右対称なので、面積が 0.45 になるときの Z(1.96)を用いる。母標準偏差σが不明であっても、データ数が多ければ標本標準偏差Sが代用できる。

$$\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}$$

これを平均値の 95%信頼区間という。

しかし、データ数が少ない場合はどうだろうか。小数データを正規分布に当てはめると誤差が大きいことから、母標準偏差σに代わって標本標準偏差Sを用いて標準化した値t が利用される。

$$t = \frac{\chi - \mu}{S / \sqrt{n}}$$

この値tは自由度 $n-1$ のt 分布(t-distribution)に従う。

t 分布は左右対称で山型の分布であるが、標本の大きさによって分布の形が変わり、自由度が大きくなるにつれて標準正規分布に近づいていく。

t 分布のほかに標準正規分布から派生した分布に χ^2 分布とF分布がある。

【演習1】血液データからヒストグラムを作成しなさい。（データはエクセルファイルを参照 ）

4．仮説検定の key word

(1)帰無仮説(null hypothesis : H_0)

仮説検定の原理は背理法である。棄てられる運命にある仮説を帰無仮説という。一方、A ＝ B という帰無仮説に対して A ≠ B(H_1)、A ＞ B(H_2)、A ＜ B(H_3)という 3 種類の仮説を対立仮説という。統計的仮説には帰無仮説を用いる。

(2)有意水準(level of significance : α)

判断の基準となる確率(5%または 1%)を有意水準といい、α で表わす。

帰無仮説の成立する可能性が α 以下のとき、帰無仮説を棄却する。つまり 5%の有意水準という表現は、本当は正しい H_0(A ＝ B)を否定して、誤った H_1 、H_2 、H_3 を採用する可能性が 20 回に1回あることを意味している。換言すれば、$P < 0.05$ という表現は、帰無仮説の成立する確率が 0.05 より小さいことを示す。

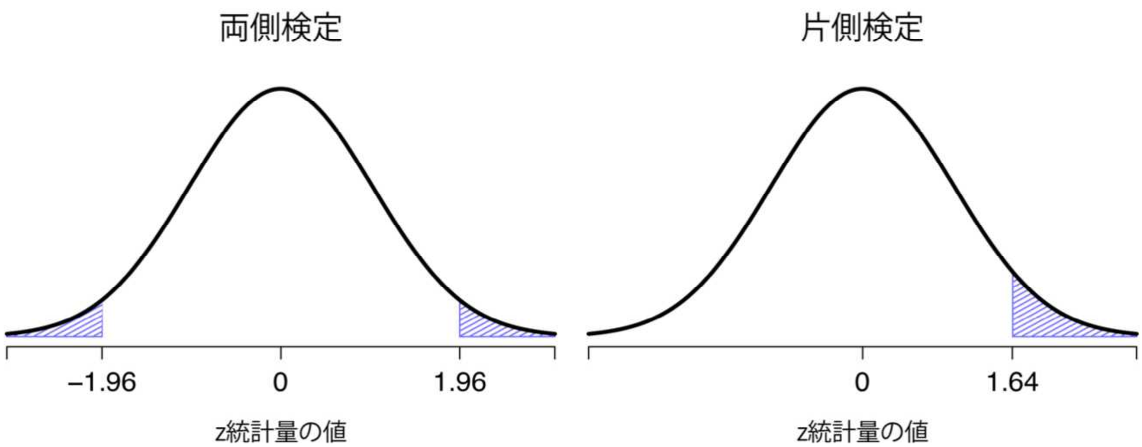
(3)第 1 種のエラー(type 1 error : α)と第 2 種のエラー(type 2 error : β)

検定に際して、正しい H_0 を誤って棄てる誤りを第 1 種のエラーといい、H_0 が誤っているにもかかわらず、棄却しない誤りを第 2 種のエラーという。この両者は一方を小さくすると他方は大きくなる関係にあり、バランスを取る必要から、α は 0.01 ～ 0.05 に設定される。

(4)両側検定(two tailed test)と片側検定(one tailed test)

母平均 μ_1, μ_2 をもつ 2 つの母集団があり、μ_1 と μ_2 の差を検定する時の H_0 を $\mu_1=\mu_2$ とする。このときの対立仮説は $\mu_1>\mu_2$ または $\mu_1<\mu_2$ である。仮説が正しいにもかかわらず、$\mu_1>\mu_2$ と結論を出す確率を α_1 、$\mu_1<\mu_2$ と結論を出す確率を α_2 とすると、$\mu_1=\mu_2$ 以外の結論を出す確率は($\alpha_1+\alpha_2$)である。 このように($\alpha_1+\alpha_2$)のような確率水準を採用する場合を両側検定という。正規分布や t 分布を応用して検定する場合、$\alpha=\alpha_1+\alpha_2$ であるから、$\alpha=0.025$ に相当する Z 値や t 値を用いる。

一方、明らかに $\mu_1 \geqq \mu_2$ である場合には、誤って $\mu_1<\mu_2$ となる確率をゼロとして有意水準 $\alpha=\alpha_1$ を採用する。これが片側検定である。



＜データの統計処理＞

1．平均値の差の検定

正規分布母分散が既知のとき（大標本の場合 σ を S で代用）

t 分布　母分散が未知のとき（小標本の場合）

(1)独立 2 標本の平均の差の検定（母分散既知か大標本）

(2)独立（対応のない）2 標本の平均の差の検定（母分散未知）

母分散が等しいとき（ Student's t-test ）

母分散が等しくないとき（ Welch の方法）

(3)対応のある 2 群の小標本平均の差の検定（ paired t-test ）

【例題1】次の表は、昨年の同時期にある歯科医院を初めて受診し、その後 1 年ぶりに来院にした幼児 8 名（ A ）と、年間を通じて定期的に指導を受けてきた幼児 8 名（ B ）の年間う蝕発生歯面数を示している。両者の平均はそれぞれ 5.64 と 3.14 であった。

この結果から、この間の定期的な受診によるう蝕抑制効果があったと判断してよいか。

幼児　A	5	3	8	6	0	5	4	3
幼児　B	4	2	0	0	3	2	5	3

（有効数字 3 桁、危険率 5 %）

（解説）

等質の母集団から抽出された独立 2 標本（ n_A ， n_B ）において、標本平均の差（ $\bar{x}_A - \bar{x}_B$ ）は自由度（ $n_A + n_B - 2$ ）の t 分布となることを利用する

	幼児　A	幼児　B
標本数	n_A	n_B
標本平均	$\bar{x}_A$	$\bar{x}_B$
不偏分散	S_A^2	S_B^2

$n_A = 8$

$\bar{x}_A = 1 ／ 8\ (5 + 3 + \cdots + 3) = 4.25$

$S_A^2 = 1 ／ 8 - 1\ \{(5 - 4.25)^2 + (3 - 4.25)^2 + \cdots + (3 - 4.25)^2\} = 5.64$

$n_B = 8$

$\bar{x}_B = 1 ／ 8\ (4 + 2 + \cdots + 3) = 2.38$

$S_B^2 = 1 ／ 8 - 1\ \{(4 - 2.38)^2 + (2 - 2.38)^2 + \cdots + (3 - 2.38)^2\} = 3.13$

(1)H_0 の設定（両者の平均年間う蝕発生歯面数に差はない）

(2)母分散が等質かどうかの確認（ F 検定）

①両者の母分散に差はない（ H_0 ’）

② F_0 値の計算（ $F_0 > 1$ ）

$S_A^2 > S_B^2$ より $F_0 = S_A^2 ／ S_B^2 = 5.64 ／ 3.13 = 1.80$

③ F 値を求める

$F\,(n_A - 1,\ n_B - 1,\ \alpha) = F\,(7,\ 7,\ 0.05) = 3.79$　→ 自由度:分子(横)分母(縦)

④ $F_0 < F$ より H_0 ’は棄却できない　∴ 母分散は等しい

(3)共分散 ω^2 を求める

$$\omega^2 = \frac{\Sigma\,(x_{Ai} - \bar{x}_A)^2 + \Sigma\,(x_{Bi} - \bar{x}_B)^2}{n_A + n_B - 2}$$

$$= \frac{(n_A - 1)\,S_A^2 + (n_B - 1)\,S_B^2}{n_A + n_B - 2}$$

$$= \frac{(8 - 1) \times 5.64 + (8 - 1) \times 3.13}{n_A + n_B - 2} = 4.39$$

(4)t_0値を求め t 値と比較する

$$t_0 = \frac{|\ \bar{x}_A - \bar{x}_B\ |}{\sqrt{\omega^2\,(1 ／ n_A + 1 ／ n_B)}} = \frac{|\ 4.25 - 2.38\ |}{\sqrt{4.39\,(1 ／ 8 + 1 ／ 8)}} = 1.79$$

$t\,(n_A + n_B - 2,\ \alpha) = t\,(14,\ 0.05) = 2.145$

$t_0 < t$ より　H_0 は棄却できない

(5)結論（危険率 5%で両者の年間う蝕発生歯面数に差はあるとはいえない）

【演習２】を行いなさい。（データはエクセルファイルを参照 ）

２．３群以上のデータの取り扱い（分散分析：ANOVA analysis of variance ）

標本分散の違いを検定する

（１）一元配置分散分析（ one way ANOVA ） １因子により分類される多群

（２）二元配置分散分析（ two way ANOVA ） ２因子により分類される多群

繰り返しのある二元配置

繰り返しのない二元配置

一元配置分散分析について例題で示す

【例題２】抜歯後疼痛を押さえるため患者に 3 種類の鎮痛剤を経口投与した。作用の発現までの時間（分）は次のようであった。薬剤により差があるといえるか。

（解説）

H_0：薬剤による作用発現時間（分）に差はない。

薬剤 A 投与群	薬剤 B 投与群	薬剤 C 投与群
185	200	265
225	186	214
215	178	226
245	196	224
232		248

A 群の平均	B 群の平均	C 群の平均	全体の平均
220.4	190	235.4	217.0714286

１）群間の変動を求める

群間の偏差平方和 4666.528571

２）群内の変動を求める

A の偏差平方和　2043.2

B の偏差平方和　296

C の偏差平方和　1711.2

群内の偏差平方和 4050.4

全体（群間＋群内）の偏差平方和　8716.928571

３）分散分析表にまとめる

変動要因	偏差平方和	自由度	分　散	分散比	P
群間変動	4666.528571	2	2333.264286	6.33663518	0.0148
群内変動	4050.4	11	368.2181818		
総 変 動	8716.928571	13			

４）判定

危険率 5%で薬剤による作用発現時間（分）に差がある。

（分散分析ではどの薬剤間に差があるかは分からない。薬剤間の差は多重比較で行うが省略する。）

【演習３】を行いなさい。（データはエクセルファイルを参照 ）

＜対になったデータの取り扱い（相関と回帰）＞

1．相関係数（ r : correlation coefficient ）

相関図（散布図）に分布する点が、回帰直線の近辺に集中して分布していれば、x の任意の値に対する y の中心的な値を推定することができる。相関係数はデータの分布が回帰直線の回りに集中する程度を表す。

すべてのデータは（ x_i ， y_i ）という対のデータとして書き表される。そこで、X ， Y 成分の変動として各々の偏差平方和（ T ）を考える。

X 成分の変動　$T_x = \Sigma (x_i - \bar{x})^2 = \Sigma x_i^2 - \dfrac{(\Sigma x_i)^2}{n}$

Y 成分の変動　$T_y = \Sigma (y_i - \bar{y})^2 = \Sigma y_i^2 - \dfrac{(\Sigma y_i)^2}{n}$

X ， Y 成分間に関連する変動を両者の偏差の積で考える。それを共変動という。

X と Y の共変動　$T_{x.y} = \Sigma (x_i - \bar{x})^2 (y_i - \bar{y})^2 = \Sigma x_i y_i - \dfrac{\Sigma x_i \Sigma y_i}{n}$

この共変動は、2 変数の関係が直線的であるほど正または負の大きな値になるため、X ， Y 成分の関連の強さを示す指標となる。

X ， Y 成分の共変動を各々の変動の平方根の積で除したものを相関係数（ r ）という。

$$r = \frac{T_{x.y}}{\sqrt{T_x T_y}}$$

相関係数の有意性の検定

$$t = | \, r \, | \sqrt{\frac{n-2}{1-r^2}}$$

t 値が、自由度 $n - 2$ の t 分布することを利用して、t 分布表から判定する

2．直線回帰（ linear regression ）

変数 x, y に相関関係があるとき、その分布の中心的傾向を予測する直線を回帰直線という。

求める直線の方程式を $y = ax + b$ と仮定し、P（ x_i ， y_i ）から垂線をおろし、直線との交点を Q（ x_i ， y' ）とすれば、$\overline{PQ}$ 間の差は（ y_i ， y' ）となる。

変数 X から Y を予測する場合の誤差を最小とするのは $\Sigma (y_i , y') = 0$ であり、$\Sigma (y_i , y')^2$ が最小となる a ， b を求めればよい。

$\overline{PQ}^2 = \Sigma (y_i , y')^2 = \Sigma \{ y_i - (ax_i + b) \}^2$ を最小とする a ， b は

$$a = \frac{T_{x.y}}{T_x}$$

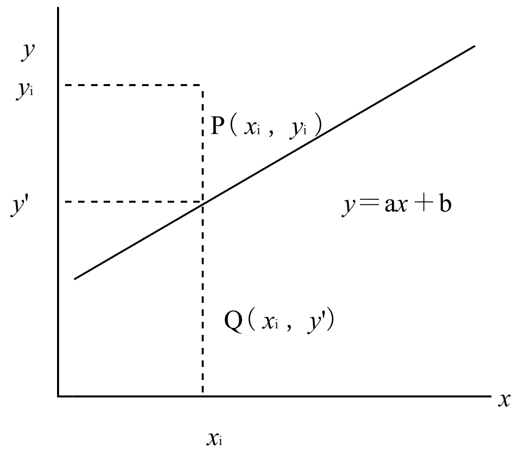
$$b = \bar{y} - a \bar{x}$$

となる。

$y = ax + b$ に a ， b を代入すると

$$y - \bar{y} = \frac{T_{x.y}}{T_x} (x_i - \bar{x})$$

これを X から Y への回帰直線という。



【演習4】を行いなさい。（データはエクセルファイルを参照 ）

＜離散量（discrete variate）データの取り扱い（χ^2 検定）＞

構成集団中である特性を持ったものの割合が問題となることが多い。

分割表を利用する。

1．適合度の検定（1要因：単純集計の場合）

1つの母集団の性質を検討する。

観測値O、期待値Eとするとき、χ_0^2 値より一致度を算出する。

$$\chi_0^2 = \Sigma \frac{(O - E)^2}{E}$$

一般には連続性の補正（Yate's の補正）を行う。

$$\chi_0^2 = \Sigma \frac{(|O - E| - 0.5)^2}{E}$$

この χ_0^2 値を自由度 $n-1$ の χ^2 値と比較し、$\chi_0^2 \geqq \chi^2(\alpha)$ であれば、危険率 α で H_0 を棄却する。

2．独立性の検定（2要因：$l \times m$ 分割表の場合）

2つの母集団の性質を比較する。

次の式は2×2分割表を利用した場合のみ利用できる。

$$\chi_0^2 = \frac{n(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)} \qquad (n = a+b+c+d、a, b, c, d \neq 0)$$

一般には連続性の補正（Yate's の補正）を行う。

$$\chi_0^2 = \frac{n(|ad - bc| - n/2)^2}{(a+b)(c+d)(a+c)(b+d)}$$

この χ_0^2 値を自由度 $(l-1)\times(m-1)$ の χ^2 値と比較し、$\chi_0^2 \geqq \chi^2(\alpha)$ であれば、危険率 α で H_0 を棄却する。

3．McNemar の検定（対応のある2標本の比率の差の検定）

同一被験者を対象に2度同じ検査をするといった、ある事象前後での変化が有意かどうか（H_0：前後の変化は有意ではない。）を検定するもの。

次の表で、1回目と2回目で反応の異なるbとcに注目する。

		1回目の反応		計
		+	−	
2回目の反応	+	a	b	a + b
	−	c	d	c + d
計		a + c	b + d	

a（1回目、2回目ともに＋）とd（1回目、3回目ともに－）は情報にならない。

変化のなかった場合の期待値 $E = \frac{b+c}{2}$ と観察値b，cの差を利用する。

$$\chi_0^2 = \frac{\left(b - \frac{b+c}{2}\right)^2}{\frac{b+c}{2}} + \frac{\left(c - \frac{b+c}{2}\right)^2}{\frac{b+c}{2}} = \frac{(b-c)^2}{b+c}$$

連続性の補正をする場合は

$$\chi_0^2 = \frac{(|b - c| - 1)^2}{b+c}$$

この χ_0^2 値を自由度1の χ^2 値と比較し、$\chi_0^2 \geqq \chi^2(\alpha)$ であれば、危険率 α で H_0 を棄却する。

【演習5】と【演習6】を行いなさい。（データはエクセルファイルを参照 ）

＜順位で示されるデータの取り扱い（ノンパラメトリック検定）＞

ここでは、平均値の差の検定を扱ったものを取り上げる。

1．Mann-Whitney の U 検定　独立した 2 群の比較

例題で示す

【例題３】妊娠中の女性患者 7 名と同世代の女性患者（非妊娠）13 名の唾液のグルコースクリアランス値を測定しデータを得た。両群には差があるか。

H_0 :グルコースクリアランス値に差はない。（2 群の順序関係に偏りはない）

	1	2	3	4	5	6	7	8	9	10	11	12	13	
女性（非妊娠）	80	110	130	160	170	170	170	180	190	190	220	270	400	
妊娠女性	170	220	280	290	360	380	470							(mg/dl)
	①	②	③	④	⑤	⑥	⑦							

1）U 値を求める（ここでは妊娠女性に注目しグループ1、対照群をグループ2とする）。

個々の点に注目して、その点より大きい他方の群のデータ数を調べ合計する。

	①	②	③	④	⑤	⑥	⑦
順位	6.5	12.5	15	16	17	18	20

注 ①は 5, 6, 7 とおよび 8 の平均
②は 12 と 13 の平均

$$\text{U値}(U_1) = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

但し　n_1 :グループ1の標本数　R_1 :グループ1の順位和(①～⑦の順位の和: 105)

2）この U 値を Mann-Whitney の U 表で $n_1 = 7$ のときの値と比較し、それよりも小さければ仮説を棄却する。（U 値が小さいほど群間の差は大きくなる）

（確率の求め方）

$n_1 \leqq 20$ かつ $n_2 \leqq 20$ のとき　Mann-Whitney 検定表を利用。

n_1, n_2 の一方が 20 以上のとき

$$\mu_U = \frac{n_1 n_2}{2} \quad と \quad \sigma_U = \sqrt{\frac{n_1 n_2(n_1 + n_2 + 1)}{12}} \quad から \quad Z = \frac{U - \mu_U}{\sigma_U} \quad を算出。$$

正規分布表で Z を求める。

Mann-Whitney 検定表 : U値の有意点(両側確率〉

$P < 0.05$

n2=	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	=n2
n1=1	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	n1=1
2	−	−	−	−	0	0	0	0	1	1	1	1	1	2	2	2	2	2
3	−	0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	3
4	0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	13	4
5		2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20	5
6			5	6	8	10	11	13	14	16	17	19	21	22	24	25	27	6
7				8	10	12	14	16	18	20	22	24	26	28	30	32	34	7
8					13	15	17	19	22	24	26	29	31	34	36	38	41	8
9						17	20	23	26	28	31	34	37	39	42	45	48	9
10							23	26	29	33	36	39	42	45	48	52	55	10
11								30	33	37	40	44	47	51	55	58	62	11
12									37	41	45	49	53	57	61	65	69	12
13										45	50	54	59	63	67	72	76	13
14											55	59	64	67	74	78	83	14
15												64	70	75	80	85	90	15
16													75	81	86	92	98	16
17														87	93	99	105	17
18															99	106	112	18
19																113	119	19
20																	127	20

2．Wilcoxon の T 検定　対応のある 2 群の比較

例題で示す

【例題4】患者 10 名について右側と左側で臼歯部の咬合圧に差があるか。

H_0：左右の咬合圧に差はない。　（kg）

患者	1	2	3	4	5	6	7	8	9	10
左側	49	35	40	49	34	42	52	41	32	40
右側	53	42	46	46	39	42	49	44	36	49
差	－4	－7	－6	3	－5	0	3	－3	－4	－9

1）符合を無視して並び替える。（同順位 0 は除外）

3　3　－3　－4　－4　－5　－6　－7　－9

2）符合を無視して順位をつける。

2　2　2　4.5　4.5　6　7　8　9

3）符合つき順位を求める。

2　2　－2　－4.5　－4.5　－6　－7　－8　－9

左側 $T_{+} = 2 + 2 = 4$

右側 $T_{-} = 2 + 4.5 + 4.5 + 6 + 7 + 8 + 9 = 41$

4）小さい方の T_{+} 値を T 表（ $n = 9$ のとき）の T 値と比較し、
T_{+} 値がそれよりも小さければ仮説を棄却する。

（確率の求め方）

$n \leqq 25$ のとき　Wilcoxon 検定表を利用。

$n > 25$ のとき

$$\mu_T = \frac{n(n+1)}{4}$$ と

$$\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{4}}$$ から

$$Z = \frac{T - \mu_T}{\sigma_T}$$ を算出する。

正規分布表で Z を求める。

Wilcoxon 検定表：*T*の有意点

片側確率	P<0.025	P<0.005
両側確率	P<0.05	P<0.01
n = 6	0	
7	2	
8	3	0
9	5	1
10	8	3
11	10	5
12	13	7
13	17	9
14	21	12
15	25	15
16	29	19
17	34	23
18	40	27
19	46	32
20	52	37
21	58	42
22	65	48
23	73	54
24	81	61
25	89	68
26	98	75
27	107	83
28	116	91
29	126	100
30	137	109
31	147	118
32	159	128
33	170	138
34	182	148
35	195	159
36	208	171
37	221	182
38	235	194
39	249	207
40	264	220
41	279	233
42	294	247
43	310	261
44	327	276
45	343	291
46	361	307
47	378	322
48	396	339
49	415	355
50	434	373

参考資料

ノンパラメトリック法　日本文化科学社　1964

生物統計学入門　培風館　1975

医学保健学の例題による統計学　現代数学社　1982

臨床医のための医学統計　メディカルコア　1983

バイオサイエンスの統計学　南江堂　1990　（使用した検定表は、ここから引用した。）