SAS (SAS Enterprise Guide) 使用方法 (2種類のデータの関係を分析する)

目次

(1)	散布図の表示	2
(2)	2種類のデータのヒストグラムおよび、累積度数分布の確率プロットの表示	3
(3)	相関係数を求める	6
(4)	データを関数変換して分析を行う	8
(4)	Spearman の順位相関係数を求める	11
(5)	記号の順位(順序)尺度データを数字に変換する	13
(6)	クラメールのC係数、φ(ファイ)係数を求める	17

- (1) 散布図の表示
 - 1) SAS Enterprise Guide を起動し、使用するデータの表が表示されている状態まで進める。
 - 2) 上部メニュー内の [グラフ]をクリックし、その下に表示されるメニュー内の [散布図] をクリックする。

🕵 SAS Enterprise Guide - Sample-Load			
ファイル(E) 編集(E) 表示(V) コード(C) データ(D) 記述統計(S)	グラ:	フ(G) 分析(<u>A</u>) フ	パイン(1)
` - ☆ - 囁 昌 →		ウィザード(<u>W</u>)	2
SASUSERIMPW_0006 (プロセスフロー) 🔹 🔉 🕞 🖉		棒グラフ(<u>B</u>)	
	Ø	円グラフ(<u>P</u>)	L.
Sample-Load	\sim	折れ線グラフ(<u>L</u>)	
E Seg プロセスフロー	<u>ka</u>	散布図(S)	
□ 語 Sample xls (Sample\$) □ 語 データのインポート 2 2	M	領域プロット(<u>A</u>)	

3)新しいウィンドウが開くので、散布図に使用する変数の一方を「変数リスト」の下のリストから選んでクリックする。そして、その右隣りにある右向き矢印のボタンをクリックし、新たに表示されたメニューから「X軸」をクリックする。

同様に、散布図に使用する変数のもう一方を「変数リスト」の下のリストから選んでク リックする。そして、その右隣りにある右向き矢印のボタンをクリックし、新たに表示さ れたメニューから「Y軸」をクリックする。

(注; この分析の場合、X軸とY軸に設定する変数は、2種類のデータのどちらを使 用しても問題はない。なお、回帰分析ではX軸とY軸に設定する変数を間違える と大きな問題になるので注意が必要である)



4) 右の「タスクの役割」の下の「X軸」と「Y軸」の下に、選択した変数名が表示されて いることを確認した後、下にある「実行」ボタンをクリックする。 5) 画面に下図のような「散布図」が表示されるので、2種類のデータ間にどのような関係 がありそうなのかを推定する。



散布図

- (2) 2種類のデータのヒストグラムおよび、累積度数分布の確率プロットの表示
 - 1) SAS Enterprise Guide を起動し、使用するデータの表が表示されている状態まで進める。
 - 2) 上部メニュー内の [記述統計]をクリックし、その下に表示されるメニュー内の [分布] をクリックする。



3)新しいウィンドウが開くので、関係を分析する2種類のデータの一方を「変数リスト」の下のリストから選んでクリックする。そして、その右隣りにある右向き矢印のボタンを クリックし、新たに表示されたメニューから「分析変数」をクリックする。

更に、関係を分析する2種類のデータのもう一方を「変数リスト」の下のリストから選 んでクリックする。そして、その右隣りにある右向き矢印のボタンをクリックし、新たに 表示されたメニューから「分析変数」をクリックする。

山 IMPW_0006 の 分布 タスクの役割 分布	タスクの役割	>
概要 正規分布 対数正規分布 指数分布 地域の分布 ベータ分布 ガンマ分布 核 グラフ 表示 インセット テーブル タイトル	変数リスト(<u>A</u>) 名前 ③ F1 ④ 漢字氏名 ④ 性別 ③ 年齢(才) ④ 年齢(才) ④ 年齢(才) ④ 年令階級 ④ 呼煙 ③ 年数(年/日) ④ 本数(本/日) ④ 身長(cm) ④ 体重(Ke) ④ 最高血圧(mmHe) ④ 最低血圧(mmHe) ④ 尿運白 ④ 尿灌血 ③ 赤血球(万個/mm3) ④ 血色素(e/d)	
		実行(13) 上書き(保存(3) キャンセル ヘルラ
分析変数 役割には	少ななも「変数を割り当てる	必要かめります。

- 4) 右の「タスクの役割」の下の「分析変数」の下に、分析する2種類のデータの名前が表示されていることを確認する。
- 5) 左端のメニューから「分布」の下の「正規分布」をクリックし、新たに表示された右側 の画面で、「正規分布」の前のボックスにチェックを入れる。

概要	□ 正規分布(N)	
<u>正規分布</u> 対数正規分布	☑ 分布をすべての変数	(に適用する(<u>A</u>)
指数分布 Weibull 分布	分析変数(⊻):	平均 (ミュー)(M) (シュー)(M) (シュー)(M) (シュー)(M)
ベータ分布 ガンマ分布		標準偏差(シグマ)(<u>G</u>)
核 パニコ		種類()):
表示		
インゼット テーブル		幅(<u>D</u>):
タイトル		□ 分布テーブルを表示しない(P) □ □ □ □ □
		推定値の使用(山)
		(指定しない場合、パラメータは推定されます。)
	1	
]	

6) 左端のメニューから「グラフ」の下の「表示」をクリックし、新たに表示された右側の 画面で、「ヒストグラム」と「確率プロット」の前のボックスにチェックを入れ、下にあ る「実行」ボタンをクリックする。

分布 概要 正規分布	注: インセット プロットのみす	は、ヒストグラム、確率プロット、QQ す効です。	軸の色:	背景色:	軸の幅:	
N 叙止規方布 指数分布 Weibull 分布 Airーの公本		ヒストグラム(出)	•	•	1	•
バーダガボー ガンマ分布 核 ガニコ	<u></u>	▶ 確率ブロット(P)		•	1	•
ッファ 表示 インセット テーゴリ		 עם ליםיאל <u>(</u>	-	v	1	¥
テーフル タイトル	!	 □ 箱ひげ図(<u>B</u>)	-	-	1	Y
	**************************************	🗖 テキストべースのプロッド(L)	オブザベーシ かを、また箱 By 変数がの	/ョン数に応じて 取げ図、正規で ある場合、プロッ	幹葉図、棒グ 確率ブロットを トは並列に作	ラフのいずれ 作成します。 :成されます。
	連続型の理論 タを示す参照線	分布の確率プロットを作成し、理論分 泉を重ね合わせます。	布に対して指	定または推定さ	れた位置パラッ	リータと尺度パラメー
^{∞∞})⊐. ២. ठ.– 91.18		# %2/	n) L#	±±/₽/≂/c)	الاصل المداحل	1 011-2

7) 画面に「分布」の分析結果が表示されるが、それぞれのデータの分析結果の最下部に下 図のような累積度数分布の確率プロットがあるので、どちらのデータも青い正方形のプ ロットが、左下から右上に伸びている黄色い直線上に乗っているかどうかを確認する。

青いプロットが黄色い直線上にほぼ乗っている場合は、その変数(データ)は正規分布していると判断してよい。もし、青いプロットが黄色い直線から大きくずれている場合は、この変数は正規分布していないと判断できる。





- (3) 相関係数を求める
 - 1) SAS Enterprise Guide を起動し、使用するデータの表が表示されている状態まで進める。
 - 2) 上部メニュー内の [分析]をクリックし、その下に表示されるメニュー内の [相関関係] をクリックする。

🌀 SAS Enterprise Guide - Sample-Load			
□ ファイル(E) 編集(E) 表示(V) コード(C)	データ(<u>D</u>) 記述統計(<u>S</u>) グラフ(<u>G</u>) [:	分析(<u>A)</u> アドイン(<u>1</u>) OLAP(<u>O</u>)	、ツール(I) ウィンドウ(W) ヘルプ
14-3-614/BBX/2) 🔍 🔍 🚬 🔤 プロジェクトデザイナ	分散分析(<u>A</u>)	🔄 タスクステータス(<u>K</u>)
		回帰分析(<u>R</u>)	
3A303EK1MFW_0000 (70 EX70 -) + +			☑ 相関関係(C)
70919F19770-5 + *	🗞 👸 プロジェクトデザイナ 🛛 🔀 SASI	生存時間分析(S)	☑ 正準相関分析(<u>A</u>)
Sample-Load 中品のプロセスフロー	😡 F1 🔬 漢字		🗵 主成分分析(P)
⊟ ∰ Samplexis (Sample\$)		管理図(0)	L。因子分析(<u>F</u>)
📕 👘 🖅 データのインポート	2 110 利 1月		1.* カニフカニノン#6(の)

3)新しいウィンドウが開くので、関係を分析する2種類のデータの一方を「変数リスト」の下のリストから選んでクリックする。そして、その右隣りにある右向き矢印のボタンを クリックし、新たに表示されたメニューから「分析変数」をクリックする。

更に、関係を分析する2種類のデータのもう一方を「変数リスト」の下のリストから選 んでクリックする。そして、その右隣りにある右向き矢印のボタンをクリックし、新たに 表示されたメニューから「分析変数」をクリックする。



- 4) 右の「タスクの役割」の下の「分析変数」の下に、分析する2種類のデータの名前が表示されていることを確認する。
- 5) 左端のメニューから「オプション」をクリックし、新たに表示された右側の画面で、「相関係数」の下の「Pearson」の前のボックスにチェックが入っていることを確認し、入っていない場合はクリックしてチェックを入れ、最下部の「実行」ボタンをクリックする。

IMPW_0006 の 3	相関関係	<u>×</u>
タスクの役割	オプション	
オフション 結果	- 相関係数	Pearson 相関のオプション
出力データ 🤇	Pearson(P)	🥅 Cronbach のアルファ係数(B)
\$1 hu	Hoemaing(H)	□ 共分散(N)
	☐ Kendall(K)	
	j opearman(<u>c</u>)) 19正/東干 カ 和2/頃和山) 「 Pearson の相関係数を表示しない(11)
	7 畝(0) 万 珥(型)	
		まえ ハロトロノゴ 売買服金 まえいへい
	次損値を含む行至1年を相関係数の計算から味外	98 (UXPU1 X CHID#98)(U)
		<u> </u>
		_
 	(の) 実行	·(B) トまき保存(S) をおうけれ ヘルプ

6)分析結果が表示されるので、まず、分析しているデータが、母集団全体を調べたものか、 母集団から抽出したサンプルを調べたものかを確認し、母集団から抽出したサンプルを調 べたものの場合は、「Pearsonの相関係数.....」という表で、関係を分析している2種類 のデータの組み合わせのセルに記載された上から2番目の数値が危険率(p)なので、これを 確認する。もし、分析しているデータが、母集団全体を調べたものの場合は、危険率(p)は 無視して、直ぐに7)の相関係数の確認を行う。

危険率(p)が 0.05(5%)未満の場合、統計的に有意な関係がこの2種類のデータの間に存在すると考えられる(ただし、この時点では結論はまだ出さない)ので、7)の操作に進む。もし、0.05(5%)以上の場合は、2種類のデータの間に関係があるかどうかの判定は「保留」となり、「2種類のデータの間に有意な関係があるとはいえない」あるいは「2種類のデータの間に有意な関係があるとはいえない」あるいは「2種類のデータの間に有意な関係は見られなかった」という結論になり、ここで分析は終了となり、相関係数を確認する必要は無い。なお、決して「2種類のデータの間に関係が無い」あるいは「無関係」という結論は出せない。これは、サンプル数を増加させた場合や他の検定方法を使用した場合、有意な関係が検出される可能性があるためである。



7)分析結果が表示された中で、「Pearson の相関係数.....」という表で、関係を分析して いる2種類のデータの組み合わせのセルに記載された一番上の数値が相関係数なので、こ れを確認する。

相関係数の確認では、求められた相関係数が1に近いほど正の相関関係が強いと考えられる。(正の相関関係とは、一方のデータが増加すると他方のデータも増加する関係)逆に、相関係数が-1に近いほど負の相関関係が強いと考えられる。(負の相関関係とは、一方のデータが増加すると他方のデータは減少する関係)そして、相関係数が0に近いほど 2つのデータの関係は弱くなる。



(4) データを関数変換して分析を行う

1) SAS Enterprise Guide を起動し、使用するデータの表が表示されている状態まで進める。

2) 上部メニュー内の [データ]をクリックし、その下に表示されるメニュー内の [フィルタ とクエリ]をクリックする。

2	12/ 7 7 7 9 00				
	🌠 SAS Enterprise Guide - Sample-Load				
	ファイル(E) 編集(E) 表示(V) コード(<u>C</u>)	デー	-タ(<u>D</u>) 記述統計(S)	グラフ(<u>G</u>)	分析(<u>A</u>)
	1-6-5 (4 % 6 8 × 4		読み取り専用(E)] [
	SASUSERIMPW_0006(プロセスフロー) 🔩		フィルタとクエリ(<u>F</u>)		
	プロジェクトエクスプローラ 🛛 🖡 🗙] 📰	行(<u>R</u>)		PV
	🕵 Sample-Load		列(上)		
			移動(<u>G</u>)	Ctrl	+G
	⊡ 🗃 Samplexis (Sample\$)	RTE.	ニーゴル の注意物(の)		

3)新しいウィンドウが開くので、左側の変数のリストから、関係を分析する2種類のデー タ(変数)を選び、ダブルクリックし、右側の枠内にその2種類の変数が表示されているこ とを確認する。

🕷 SASUSERIMPW_0006 のクエリ - クエリビルダ				X
クエリ名(Q): SASUSERIMPW_0006のク.	エリ 出力名(0):	SASUSER.QUERY_FO	R_SASUSER_IMPW_0	変更(<u>C</u>)
🏢 計算列(M) 🛛 🥫 パラメータ(P) 💋 検証(V)	▼ 📆 ブレビュー(E) 🔮 オブ	∋∋) (<u>N</u>) ▼		
	データの選択 フィルタデータ	データの並べ替え		
	列名	入力	概要	
6 F1 (F1)	() (年齢(才)(年齢(才))	IMPW_0006.年齢(才)		
·····································	(i) 'GPT(IU/I)' (GPT(IU/	IMPW_0006.GPT(IU/I)		
(1) 年令階級(年令階級)				
() 身長(cm)(身長(cm))				
→ → → → → → → → → → → → → → → → → → →				
(mms,())) (m色素(g/d))(me)				
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	- 要約グループ			
	🗌 🗍 グループを自動選択する	U	グルー	ブの編集(G)…
	グループが選択されていません	U		
International log(GPT) (log(GPT) )		-		

4) 左上にある「計算列」をクリックする。



5)「計算列」というダイアログが開くので、右上の「新規作成」ボタンをクリックし、その 下に新たに表示されたメニュー内の「式の作成」をクリックする。

鬱 計算列	× 新規作成(N)▼
	デジルサゴ Thaty 式の作成(B) 買(B金(D)
	2前の変更(日)
	閉じる
✓ クエリに新しい計算列を追加する(A)	

6)「拡張式エディタ」というダイアログが開くので、左中央の「関数」タグをクリックし、 右下の「関数」の下のリストから、変換に用いる関数を選び、ダブルクリックする。

■写拡張式エディタ 式テキスト( <u>T</u> ): + - × / × AND OR NOT =	<> < <= > >= = ^x	× × × ×
	式のクリア( <u>C</u> )	式へ追加(0)
カテゴリ: <b>すべての講教</b> 集計 算術 ビットごとの論理演算 文字 文字列の合致 条件 Data Quality (SAS/DQUALITY が必要です) 日付と時間 DBCS 記述統計量 外部ファイル	■ ■ ■ ■ ■ ■ ■ ■ ABS ARSIN ATAN ATAN ATTRC ATTRC ATTRN AVG {AII} AVG {Distinct} BAND BETAINV BLSHIFT	
ABS(引数) ABS 関数は、引数と同じ大きさの正の数を返します。	OK	キャンセル ヘルフ

(例;対数変換を行う場合は、「LOG」の関数を選択する)

7)「式テキスト」の下のボックスに選択した関数が表示されるので、その中の「<numValue>」 をマウスで選択した後、「Back Space」キーか「Delete」キーで削除する。

Σ覧拡張式エディタ
式テキスト(工):
LOG(KnumValue>)
+ - * / ** AND (

8) 左中央の「データ」タグをクリックし、「変数値」の下のリストから変換する変数を選んでダブルクリックし、「式テキスト」の下の関数の中にその変数が表示されたことを確認した後、最下部の「0K」ボタンをクリックする。

Σ₩拡張式エディタ	×
式テキスト(工):	
LOG(IMPW_0006!GPT(IU/I)'n )	×
+ - * / ** AND OR NOT =	↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓
$\frown$	式のクリア(C) 式へ追加(D)
データし戦	
使用できる変数( <u>A</u> ):	変数値( <u>R</u> ):
□       第         □       ④         □       ④         ○       ○         □       ④         □       ○         □       ○         □       ○         □       ○         □       ○         □       □         □       □         □       □         □       □         □       □         □       □         □       □         □       □         □       □         □       □         □       □         □       □         □       □         □       □         □       □         □       □         □       □         □       □         □       □         □       □         □       □         □       □         □       □         □       □         □       □         □       □         □       □         □       □ <td< th=""><th></th></td<>	
	<u>OK</u> キャンセル ヘルプ

9)「拡張式エディタ」が閉じて、「計算式」ダイアログに戻るので、左のボックス内に表示 されている「Calculation1」をクリックした後、「名前の変更」ボタンをクリックする。

鬱 計算列	×
Calculation1	新規作成( <u>N</u> )▼
	編集( <u>E</u> )
	削除( <u>D</u> )
	名前の変更( <u>R</u> )
	閉じる
☑ クエリに新しい計算列を追加する(A)	,

10)変換後の変数の名前を適当に入力した後、「閉じる」ボタンをクリックする。

鬱 計算列	×
Log(GPT)	新規作成( <u>N</u> )▼
	編集( <u>E</u> )
	削除( <u>D</u> )
	名前の変更(R)
	閉じる
▼ クエリに新しい計算列を追加する(A)	

11) もし、もう一方のデータも変換が必要な場合は、4)~10)の操作を繰り返して、もう一方のデータも関数変換する。

- 12) 関数変換した変数名が、元の変数の下に表示されていることを確認した後、「実行」ボ タンをクリックする。
- 13) 関数変換前の変数と関数変換後の変数の表が表示されるので、これらを用いて、この資料の「(2) 2種類のデータのヒストグラムおよび、累積度数分布の確率プロットの表示」の操作を行い、関数変換後の比較する変数が、2つのグループのどちらでも正規分布しているかどうかを判定する。

もし、正規分布していた場合は、14)の操作に進むが、もし、関数変換後も正規分布していない場合は、配布資料「統計学基礎(4)2種類のデータの関係を分析する」の(9) を参照する。

- 14) もし、関数変換後の比較する変数が、2つのグループのどちらでも正規分布している場合は、この資料の「(3) 相関係数を求める」の操作を行い、2種類のデータ間の関係を分析する。
- (4) Spearman の順位相関係数を求める
  - 1) SAS Enterprise Guide を起動し、使用するデータの表が表示されている状態まで進める。
  - 2) 上部メニュー内の [分析]をクリックし、その下に表示されるメニュー内の [相関関係] をクリックする。

🕼 SAS Enterprise Guide - Sample-Load	
̄ ファイル(E) 編集(E) 表示(V) コード(C) データ(D) 記述統計(S) グラフ(G) 分析(A) アドイン(I) OLAP(O)	_ツール( <u>T</u> ) ウィンドウ( <u>W)</u> ヘルプ
日・G・ 幅 昌 ※ 時 電 × □ ○ ○ ○ ○ ↓ 品語 プロジェクトデザイナ 分散分析(A)     分散分析(A)	🔯 タスクステータス( <u>K</u> ) 💡
SASUSERIMPW 0006 (プロセスフロー) - 回帰分析(R)	
●	☑ 相関関係(C)
フロジェクトエクスフローラ 本本 Beg フロジェクトデザイナ III SASI 生存時間分析(S)	
Sample-Cool     Dia Cool     Dia Cool	🗵 主成分分析(P)
日間 Samplexis (Sample\$)	⊾ 因子分析(E)
	L* 5575_45(m)

3)新しいウィンドウが開くので、関係を分析する2種類のデータの一方を「変数リスト」 の下のリストから選んでクリックする。そして、その右隣りにある右向き矢印のボタンを クリックし、新たに表示されたメニューから「分析変数」をクリックする。

更に、関係を分析する2種類のデータのもう一方を「変数リスト」の下のリストから選 んでクリックする。そして、その右隣りにある右向き矢印のボタンをクリックし、新たに 表示されたメニューから「分析変数」をクリックする。

I	/ IMPW_0006 の ネ	相関関係		
:	タスクの役割	タスクの役割		
•	オノンヨン   結果	変数リスト( <u>A</u> )	タスクの役割(E):	_
	出力データ   タイトル	<u>名前</u> 词 F1	▲ ● ● 分析変数 	
:		▲ 漢字氏名	<ul> <li>         → 相関変数         <ul> <li></li></ul></li></ul>	
		(1) (1) (1) (1) (1) (1) (1) (1) (1) (1)		
		<ul> <li>(2) 年令階級</li> <li>▲ 喫煙</li> </ul>	● 部方変数 ● 重み (選択の上限:1)	
•		(1) (1) (1) (1) (1) (1) (1) (1) (1) (1)		
		(1)/4 € (x (4/日)) (1)/2 ● 身長(cm)	◎ 相関変数	
1		123 体重(Kg) 133 最高血圧(mmHg)	🔄 🔄 🞯 グループ分析	

4) 右の「タスクの役割」の下の「分析変数」の下に、分析する2種類のデータの名前が表 示されていることを確認する。 5) 左端のメニューから「オプション」をクリックし、新たに表示された右側の画面で、「相関係数」の下の「Pearson」の前のボックスのチェックを外し、更に、その下の「Spearman」の前のボックスをクリックしてチェックを入れた後、最下部の「実行」ボタンをクリックする。

IMPW_0006 の	相関関係	×
タスクの役割	- オプション	
オプション    結果		- Pearson 相関のオブション
出力データ	Pearson(P)	▶ Cronbach のアルファ(系数(B)
97.FJU	Hoeffding( <u>H</u> )	□ 共分散(N)
	☐ Kendall(K)	□ 平方和と積和( <u>M</u> )
	Spearman( <u>E)</u>	○ 修正済半方和と積和(1)
•		Pearson ()作目詞(新教)を設め示い(3に)(U)
1		
•	分散の分母(⊻):	
	目田度(N-1) ▼	
	□ 欠損値を含む行全体を相関係数の計算から除外	する (リストワイズで削除する)(0)
	相関分析の種類を1つ以上選択します。	*
	Spearman の順位相関係数を計算します。データ値(	D順位に基づき関連性のノンパラメトリックな指標が計算されます。
<b>」</b> ドのプレビ	<u>ニュー(C)</u>	テ(B) 上書き保存(S) キャンセル ヘルプ

6)分析結果が表示されるので、まず、分析しているデータが、母集団全体を調べたものか、 母集団から抽出したサンプルを調べたものかを確認し、母集団から抽出したサンプルを調 べたものの場合は、「Spearmanの相関係数.....」という表で、関係を分析している2種類 のデータの組み合わせのセルに記載された上から2番目の数値が危険率(p)なので、これを 確認する。もし、分析しているデータが、母集団全体を調べたものの場合は、危険率(p)は 無視して、直ぐに7)の Spearmanの順位相関係数の確認を行う。

危険率(p)が 0.05(5%)未満の場合、統計的に有意な関係がこの2種類のデータの間に存 在すると考えられる(ただし、この時点では結論はまだ出さない)ので、7)の操作に進む。 もし、0.05(5%)以上の場合は、2種類のデータの間に関係があるかどうかの判定は「保 留」となり、「2種類のデータの間に有意な関係があるとはいえない」あるいは「2種類 のデータの間に有意な関係は見られなかった」という結論になり、ここで分析は終了とな り、相関係数を確認する必要は無い。なお、決して「2種類のデータの間に関係が無い」 あるいは「無関係」という結論は出せない。これは、サンプル数を増加させた場合や他の 検定方法を使用した場合、有意な関係が検出される可能性があるためである。

Spearman の相関係数, N = 1006 H0: Rho=0 に対する Prob >  r			
	GOT(IU/I)	GPT(IU/I)	
GOT(IU/I) GOT(IU/1)	1.00000	0.71243	
GPT(IU/I) GPT(IU/1)	0.71343	1.00000	

7)分析結果が表示された中で、「Spearman の相関係数.....」という表で、関係を分析して いる2種類のデータの組み合わせのセルに記載された一番上の数値が Spearman の順位相 関係数なので、これを確認する。

Spearman の順位相関係数の確認では、求められた Spearman の順位相関係数が1に近い ほど正の相関関係が強いと考えられる。(正の相関関係とは、一方のデータが増加すると 他方のデータも増加する関係)逆に、相関係数が-1 に近いほど負の相関関係が強いと考 えられる。(負の相関関係とは、一方のデータが増加すると他方のデータは減少する関係) そして、Spearman の順位相関係数が0に近いほど2つのデータの関係は弱くなる。



- (5) 記号の順位(順序)尺度データを数字に変換する
  - 1) SAS Enterprise Guide を起動し、使用するデータの表が表示されている状態まで進める。
  - 2) 上部メニュー内の [データ]をクリックし、その下に表示されるメニュー内の [フィルタ とクエリ]をクリックする。

🕵 SAS Enterprise Guide - Sample-Load		
ファイル(E) 編集(E) 表示(⊻) コード(C)	データ(D) 記述統計(S) グラフ(G) 分析(	<u>A</u> )
1 1 • G • G   4 • B B ×   #	▶ / 読み取り専用(E)	1
SASUSERIMPW_0006 (プロセスフロー) -	フィルタとクエリ(E)…	
プロジェクトエクスプローラ 🔹 🗶	] 🚍 行(R)	P
💽 Sample-Load	■ 列(L)	L.
	移動( <u>G</u> ) Ctrl+G	K
□ 🔐 Sample xls (Sample\$)		

3)新しいウィンドウが開くので、左側の変数のリストから関係を分析する2種類のデータ (変数)を選び、ダブルクリックし、右側の枠内にその2つの変数が表示されていることを 確認する。

斷 Samplexis (Sample\$) のクエリー クエリビルダ				x
クエリ名(Q): Samplexis (Sample\$)の	カエリ 出力名(0):	SASUSER.QUERY_FOR	SAMPLE_XLS_SAI	変更( <u>C</u> )
🎰 計算列(M) 🛛 🤯 パラメータ(P) 💋 検証(V	) 🔹 🗟 ブレビュー(E) 📓 オブ	ション( <u>N</u> ) ・		
	データの選択 フィルタデータ	データの並べ替え		1
🖃 🎘 Sample xis. Sample 🔺	列名	入力	概要	
6 F1 ( F1 )	▲ '喫煙' (喫煙 )	Sample_xls_Sample_喫		
漢字氏名(漢字氏名)	▲ 尿_糖 (尿_糖)	Sample_xls_Sample_尿		
(1) 身長(cm)(身長(cm))				
→ 14里(NE)(14里(NE) ) → 日本市区(mmUr)(日本市区(				

4) 左上にある「計算列」をクリックする。

	📓 SASUSER IMP	PW_0006 のクエリ	- クエリビルダ	
	クエリ <u>名(Q</u> ):	SASUSER.	IMPW_0006 のク	IU
(	⊞計算列(M)	🔜 パラメータ( <u>F</u>	) 🔊 検証(V)	• 🗟 🤊
	🔠 追加(T)	¥ 削除(D)	≞ 結合(」)	データの選び
		006		和么

5)「計算列」というダイアログが開くので、右上の「新規作成」ボタンをクリックし、その 下に新たに表示されたメニュー内の「列の再コード化」をクリックする。

鬱 計算列	×
	新規作成(№)▼
	列の再コード化(R) 式の作成(B)
	肖耶余( <u>D</u> )
	名前の変更(民)
	閉じる
✓ クエリに新しい計算列を追加する(A)	11

6)「項目の選択」というダイアログが開くので、「項目の選択」の下のリストから数字に変換する変数をクリックし、最下部の「継続」ボタンをクリックする。

E	■ 項目の選択	×
	項目の選択	
	<ul> <li>□ Sample_xls_Sample_</li> <li>□ □ F1</li> <li>□ □ F1</li> <li>□ □ P5</li>     &lt;</ul>	
, (	維続(C) キャンセル	

7)「列の再コード化」というダイアログが開くので、もし、新しい変数名を付ける場合は、 「新しい列名」の右横のボックス内の名前を変更する。次に、その下の「追加」ボタンを クリックする。

鬱 列の再コード化- Sample_xk	s_Sample_尿塘	×
新しい列名( <u>N</u> ):	Recode_尿	
置換( <u>E</u> ):	<b>追加(<u>A</u>)</b> 前條( <u>R</u> )	
置換	置換後	Ι
, 一他の値		
すべての他の値を次の値で置	き換える:	
④ 現在の値(C)		
C 欠損値(M)		
		-
	l.	
「新しい列の種類」		
● 文字(H)	○ 数値(山)	
08	1 50° 11711 1 1 1.1.7	
UK.		
		///

8)「置換の指定」というダイアログが開くので、中央右にある、「値の取得」ボタンをクリッ クする。

	直探の指定
	値の置換
-	
	値の取得いた。
	この値で置き換える(W):
	この値で置き換える(W):
	この値で置き換える(W):
i ; ]	この値で置き換える(W):
i ; ;	この値で置き換える(W): とント: 欠損値は、空白 で表示されます。 OK キャンセル ヘルプ

9) 新しいダイアログが開くので、データ値のリストからどれかの値をクリックし、「OK」 ボタンをクリックする。

		<u>&gt;</u>
データ値		
値	出力形式が適用された値	
-	_	
+	+	
+-	+-	
++++	+++	
		-
値の追加取得(⊻	) C OK キャンセル	

10)「置換の指定」ダイアログに戻るので、「この値で置き換える」の下のボックスに変換 する数字の値を入力し、最下部の「OK」ボタンをクリックする。

鬱 置換の指定	x
値の置換単範囲の置換	1
-	
	-1
この値で置き換える( <u>W</u> ):	
ヒント: 欠損値は、空白 で表示されます。	
	7
	///

11) 7)~10)を繰り返して、再分類前の値と再分類後の値を全て設定する。
例)例えば、尿糖の「-」「+-」「+」「++」を数字に変換する場合は、 最初に9)で「-」を選んだ後、10)で「0」を入力、
次に9)で「+-」を選んだ後、10)で「1」を入力、
更に9)で「+」を選んだ後、10)で「2」を入力、
その次に9)で「++」を選んだ後、10)で「3」を入力、
最後に9)で「+++」を選んだ後、10)で「4」を入力する 12) 全ての設定が終了したら、「列の再コード化」ダイアログの下部にある「新しい列 の種類」の所で、「数値」の前の丸をクリックしてマークを付けた後、最下部の「OK」 ボタンをクリックする。

	。列の再コード化- Sample_xis	Sample尿塘
1	新しい列名( <u>N</u> ):	Recode_尿糖
1	置換( <u>E</u> ):	追加( <u>A</u> ) 削除(B)
:	置換	置換後
i	- + +-	0 2 1
1	++ +++	3
	- 他の値	
	すべての他の値を次の値で置	き換える:
	C 現在の値(⊆)	
	⊙ 欠損値( <u>M</u> )	
	○ 値(1):	
1	新しい列の種類	
-	○ 文字(出)	○ 數值( <u>U)</u>
	ОК	キャンセル ヘルプ
		lh

- 13)「計算列」ダイアログに戻るので、最下部の「閉じる」ボタンをクリックする。
- 14) 最初のダイアログに戻るので、最下部の「実行」ボタンをクリックする。
- 15) 数字変換前のデータと数字変換後のデータの表が表示されるので、正しく数字変換されていることを確認する。
- 16) 上記の表を用いて、適切な統計手法で分析を行う。
- (6) クラメールの C 係数、φ (ファイ)係数を求める

SAS Enterprise Guide を起動し、使用するデータの表が表示されている状態まで進める。
 上部メニュー内の[記述統計]をクリックし、その下に表示されるメニュー内の[分割表

分析」をクリ	リックする。		
	🕵 SAS Enterprise Guide – Sample-Load		
	ファイル(E) 編集(E) 表示(V) コード(C)	データ( <u>D</u> ) 📑	記述統計(S) グラフ(G) 分析(A)
	1 1 • 6 • 6 4 5 6 X 0	0 10	ウィザード( <u>W)</u>
	QUERY_FOR_SASUSER_IMPW_0006		<u>п</u> улк <u>и</u>
	プロジェクトエクスプローラ	Σ	∑ 要約統計量(S)
		II 9 →nd	b
	nple-Load	800 7 D1	h. 分布( <u>D</u> )
	nple-Load プロセスフロー		▶ 分布( <u>D</u> ) ■ データの特性分析( <u>H</u> )
	nple-Load プロセスフロー 細 Samplexis (Sample\$)		▶ 分布(D) データの特性分析(H) 集計表(T)
	nple-Load プロセスフロー 語 Samplexis (Sample\$) 日 語 データのインボート		<ul> <li>▶ 分布(D)</li> <li>□ データの特性分析(H)</li> <li>□ 集計表(T)</li> <li>□ 一元度数表(Q)</li> </ul>
	nple-Load プロセスフロー 語 Samplexls (Sample\$) □ 記 データのインボート □ □ 最後にサブミットしたコード □ □ ゴ		<ul> <li>分布(D)</li> <li>データの特性分析(H)</li> <li>集計表(T)</li> <li>一元度数表(Q)</li> <li>分割表分析(A)</li> </ul>

3)分割表分析のダイアログが開くので、「変数リスト」から、関係を分析する2種類のデータ(変数)をそれぞれクリックした後、それぞれで、その右の右向き矢印のボタンをクリックし、新たに表示されたメニューから「表変数」をクリックし、「タスクの役割」の下の「表変数」の下に、関係を分析する2種類のデータ(変数)が表示されたことを確認する。



4)ダイアログの左端のリストから「テーブル」をクリックし、新たに右に表示された「テー ブルで許可されている変数」の下のリストから、どちらかの変数をその右の「プレビュー」 の下の領域にドラッグアンドドロップし、続いて、もう一方の変数を同じ領域にドラッグ アンドドロップする。その結果、「プレビュー」の下の表に、関係を分析する2種類のデー タ(変数)が表示されていることを確認する。

QUERY_FOR_SASUSE	:R_IMPW_0006_0007 の 分割表分析	×
タスクの役割	テーブル	
セル統計量	テーブルで注言されている変動(\/)・	プレドュー
表統計量		「ノニニに変数をトラックトマイナさいい」
関連	(词)年令階級	
一一致 順度付けなれた美		
傾向検定		
計算オプション		
結果		
表統計量の結果		
タイトル		
		·
	$\checkmark$	
QUERY_FOR_SASUSEF	R_IMPW_0006_0007 の 分割表分析	×
タスクの役割	テーブル	
ーナーノル ヤル統計量	テーブルで計算すれている感謝化いか	สำนักระห
表統計量	) ///Cat-9/Civico/@@gx( <u>v</u> ).	
関連		平节階級
一致		
傾向検定		
計算オプション		
「結果」		
表統計量の結果		
タイトル	1	
	生成されるテーブル(工)・	
	- wept by "年谷階級"	
	기학에 가는 그렇게 찌르기꾼은 며머니, 나가 내 스카는 정생이다.	71#71X

5) ダイアログの左端のリストから「セル統計量」をクリックし、新たに右に表示された「使 用できる統計値」の下で、既にチェックの入っているものはそのままにして、「期待セル 度数」の前のボックスにチェックを入れる。



6) ダイアログの左端のリストから「表統計量」の下の「関連」をクリックし、新たに右に 表示された「独立性の検定」の下で、「χ2 乗検定(Pearson、....」の前のボックスと、 「r×c 表に対するフィッシャーの直接法」の前のボックスにチェックを入れ、最下部の「実 行」ボタンをクリックする。

QUERY_FOR_SASUSE	R_IMPW_0006_0007 の 分割表分析	×
タスクの役割	表統計量>関連	
<ul> <li>ハック(スロ) <ul> <li>テーブル</li> <li>セル統計量</li> <li>表統計量</li> <li>関連</li> <li>一致</li> <li>順序付けされた差</li> <li>傾向検定</li> <li>計算オプション</li> </ul> </li> <li>結果</li> <li>セル統計量の結果</li> <li>表統計量の結果</li> <li>タイトル</li> </ul>	独立性の検定 ×2 乗検定() ✓ Cochran-Mantel-Haenszel の ×2 乗 検定、2×2 表に対する Fisher の直接注() ✓ r x c 表に対するフィッシャーの直接注() ✓ r x c 表に対するフィッシャーの直接注() Cochran-Mantel-Haenszel 統計量 CMH 統計量( <u>H</u> ) スコアの種類:表(B)	<ul> <li>独立性の統計量</li> <li>統計量(M)</li> <li>(2 × 2 表に対する相対危険度およびオッズ比を含む)</li> <li>オッズ比に対する正確な p 値および信頼区間(P)</li> <li>統計量が 0 に等しいという検定(D)</li> <li>2 × 2 表に対するリスクの差(K)</li> <li>2 × 2 表に対する相対危険度(L)</li> <li>次の場合は [計算オプション] ページを使用してください。</li> <li>正確な p 値の計算時に制限時間を設ける。</li> <li>CMH 統計量のスコアを選択する。</li> </ul>
<u>ראשלעבר(6)</u>	2 X 2 より大きい表に対する、Fisher の正確な p 値 も呼ばれます。 	検定を計算します。この検定は、Freeman-Halton 検定と ・ ・ 上書き保存(S) キャンセル へルブ

7)分析結果が表示されるので、まず、分析しているデータが、母集団全体を調べたものか、 母集団から抽出したサンプルを調べたものかを確認し、母集団から抽出したサンプルを調べたものの場合は、9)以降の処理を行う。もし、分析しているデータが、母集団全体を調 べたものの場合は、危険率(p)は無視して、直ぐに12)の処理に進み、必要な係数の確認を 行う。

- 8) 分析結果の中で、まず、下図のような「FREQ プロシジャ」と書かれた下の表を見る。その表の各セルで、真ん中の「期待度数」を全て確認し、下記の条件に合うものがあれば、その指示に従う。
  - (a) **2 × 2表(分割表の行も列も2個のもの)で、どれかのセルの期待度数が5未満の 場合、**10)に進み、そこに記載された方法で結果を判定する。

# (b) <u>2 × 2 表以外の場合で、セルの期待度数が以下のどちらかの条件に当てはまる場合</u>

#### (1) セルのうちで 20%の期待度数が 5 未満

(2) セルの中に、1つでも期待度数が0のものがある

どちらかの条件に当てはまる場合は、11)に進み、そこに記載された方法で分析 する。

上記の(a)(b)どちらの条件にも当てはまらない場合は、9)に進み、そこに記載された 方法で危険率(p)を判定する。

度数	表:喫煙*年令階級			
朝存反数列のバーセント		年令階級(		
p	喫煙(喫煙)	30	50	슴탉
	А	259	192	451
		82.75	65.75	
	С	54	100	154
		79.673	74.327	
	슴計	313	292	605

FREQ プロシジャ

9) 8)の分割表のすぐ下にある下図のような表を見て、「カイ2乗値」の行の「p値」を確認 し、以下の方法で判定する。

統計量	自由度	値	p値
カイ2乗値	1	22.9925	<.0001
尤度比カイ2乗値	1	23.2135	<.0001
連続性補正カイ2乗値	1	22.1056	<.0001
Mantel-Haenszelのカイ 2 乗値	1	22.9545	<.0001
ファイ係数		0.1949	
一致係数		0.1913	
Cramerの V 統計量		0.1949	

この値は、本当は関係が無いのに「関係がある」と間違う確率の値である。従って、 この値が<u>5%(0.05)未満</u>の場合は、帰無仮説を棄却できる(否定できる)ため、5%の有 意水準で「<u>2種類のデータの間で有意な関係がある</u>」とみなすことができる(ただし、 この時点では結論はまだ出さない)ので、12)に進み、必要な係数の確認を行う。 もし、この値が<u>5%(0.05)以上</u>の場合は、2種類のデータの間に関係があるかどうか の判定は「保留」となり、「2種類のデータの間に有意な関係があるとはいえない」あ るいは「<u>2種類のデータの間に有意な関係は見られなかった</u>」という結論になり、<u>この</u> 後、係数を確認する必要は無い。なお、決して「2種類のデータの間に<u>関係が無い」あ</u> るいは「無関係」という結論は出せない。これは、サンプル数を増加させた場合や他の 検定方法を使用した場合、有意な関係が検出される可能性があるためである。

 2×2表の分析で、4つのセルのうち1つでも期待度数が5未満のものがある場合は、 Fisher の直接確率法(正確確率法, Fisher's exact probability test)</u>を使用する必要が ある。このため、9)のカイ2乗検定のすぐ下にある下図のような「Fisher の正確検定」の 表の「両側 Pr<=P」右横の値を確認する。</li>

この値は、本当は関係が無いのに「関係がある」と間違う確率の値である。従って、 この値が<u>5%(0.05)未満</u>の場合は、帰無仮説を棄却できる(否定できる)ため、5%の有 意水準で「<u>2種類のデータの間で有意な関係がある</u>」とみなすことができる(ただし、 この時点では結論はまだ出さない)ので、12)に進み、必要な係数の確認を行う。

もし、この値が<u>5%(0.05)以上</u>の場合は、2種類のデータの間に関係があるかどうか の判定は「<u>保留</u>」となり、「<u>2種類のデータの間に有意な関係があるとはいえない</u>」あ るいは「<u>2種類のデータの間に有意な関係は見られなかった</u>」という結論になり、<u>この</u> 後、係数を確認する必要は無い。なお、決して「2種類のデータの間に<u>関係が無い」あ</u> <u>るいは「無関係」という結論は出せない</u>。これは、サンプル数を増加させた場合や他の 検定方法を使用した場合、有意な関係が検出される可能性があるためである。

Fisherの正確検定			
セル (1,1) 度数 (F)	259		
左側 Pr <= F	1.0000		
右側 Pr >= F	1.167E-06		
表の確率 (P)	7.165E-07		
両側 Pr <= P	1.720E-06		

- (注;上記の表の危険率のように「1.720E-06」という数値は、「1.720×10⁻⁶」 を意味している。すなわち、「○○E-X」という数値は、「○○×10^{-x}」 のことである。
- 2×2表以外の分析では、下記のような場合、χ²検定(カイ二乗検定)はできない可能性 がある。

#### (1) セルのうち 20%の期待度数が 5 未満

#### (2) セルの中に、1つでも期待度数が0のものがある

- このような場合は、下記のような操作を行う。
  - 1) データの再分類を行い、期待度数の小さなセルを無いようにして、再度、χ²検 定(カイ二乗検定)を行う。
  - 2) データの再分類ができない場合は、χ²検定(カイ二乗検定)の結果の解釈や採用 するかどうかを慎重に検討(統計の専門家に相談するのも良い)し、場合によって は分析を中止する。

12) 8)の分割表のすぐ下にある下図のような表を見て、以下の方法で判定する。

関係を分析した2種類のデータのどちらかが2分類の分類尺度(名義尺度)ではない 場合(分割表の行も列も2個の場合[2×2表])、下記の「クラメールのC係数 [クラメー ルの連関係数]」で判定する。

関係を分析した2種類のデータのどちらも2分類の分類尺度(名義尺度)である場合 (分割表の行か列が3以上ある場合)、下記の「 $\phi$ (ファイ)係数」で判定する。

(クラメールのC係数 [クラメールの連関係数])

下図の表で「Cramer の V 統計量」と書かれたところの値が「クラメールの C 係数 [ク ラメールの連関係数]」である。

クラメールのC係数は、0から1の間の数値を取り、マイナスの数値になることは ない。これが1に近いほど2種類のデータ間の関係は強く、0に近いほど関係は弱い と考えられる。しかし、名義尺度(分類尺度)のデータを含む分析では、相関係数のよ うに2つの変数の関係の方向性を示すことはできないので、結論を出す際には注意が 必要である。(可能ならば、名義尺度のデータを再分類して2×2表にし、(13)の (ファイ)係数やユールの連関関数を求めれば、解釈が容易となる)

統計量	自由度	値	p値
カイ2乗値	1	22.9925	<.0001
尤度比カイ2乗値	1	23.2135	<.0001
連続性補正カイ2乗値	1	22.1056	<.0001
Mantel-Haenszelのカイ2乗値	1	22.9545	<.0001
ファイ係数		0.1949	
一致係数		0.1913	
Cramerの V 統計量		0.1949	

(φ(ファイ)係数)

下図の表で「ファイ係数」と書かれたところの値が「φ(ファイ)係数]」である。 φ(ファイ)係数は、-1から1の間の数値を取るが、その絶対値が1に近いほど二 種類のデータの間の関係は強く、0に近いほど関係は弱いと考えられる。

統計量	自由度	値	p値
カイ2乗値	1	22.9925	<.0001
尤度比カイ2乗値	1	23.2135	<.0001
連続性補正カイ2乗値	1	22.1056	<.0001
Mantel-Haenszelのカイ2乗値	1	22.9545	<.0001
ファイ係数	(	0.1949	
一致係数		0.1913	
Cramerの V 統計量		0.1949	

Copyright (C) 2011 渡辺博且, All Rights Reserved.