トレーニング コース

IBM SPSS Statistics データ加工

演習解答例

IBM SPSS Statistics V20 2012/09 IBM、IBMロゴおよびibm.comは、世界の多くの国で登録されたInternational Business Machines Corporationの商標です。

他の製品名およびサービス名等は、それぞれIBMまたは各社の商標である場合があります。

現時点でのIBMの商標リストについては、www.ibm.com/legal/copytrade.shtmlをご欄ください。

この資料は研修用教材として作成したものです。

この資料は2012年9月1日現在のものであり、将来この資料の全体または一部につき変更する場合があります。

本書の内容についてお気づきの点がございましたら、下記までお知らせください。

〒103-8510 東京都中央区日本橋箱崎町19-21

TEL. 0120-105553

日本アイ・ビー・エム株式会社

演習解答例

第2章 数値データの計算

ここでは、ある金融機関の従業員のデータ(BANK.sav)を使います。このデータ には、従業員コード(id)、性別(gender)、初任給(salbeg)、現在の給与(salnow)、 などの情報が記録されています。

数値データの変換

- 1. IBM SPSS Statisticsを起動してBANK.savファイルを開いてください。
- WindowsのスタートメニューからIBM SPSS Statisticsを起動します。
 ファイルメニューの開くのデータを選択します。
 C:¥train¥DataMani¥BANK.savを選択し、開くボタンをクリックします。
- 2. predictという名前の新しい変数を作成してください。この変数は、回答者の年齢 (age)と教育年数(edlevel)に基づく現在の収入の予測値です。次の式を使っ てください。

1592 * edlevel + 25 * age - 8600

 変換メニューの変数の計算を選択します。
 目標変数ボックスにpredict、数式に1529 * edlevel + 25 * age – 8600と入力し
 ます。

 OKボタンをクリックします。

🏭 変数の計算		×
目標変数(①: predict = 型とラベル(山) ✓ age ✓ edlevel ✓ gender ✓ id ✓ jobcat ■ minority ✓ salbeg ✓ salnow ✓ time ✓ work ↓ (任意のケースの選択条件)	数式(E): 1592*edlevel + 25*age - 8600 + < 2 7 8 9 - <= ≥= 4 5 6 * = ~= 1 2 3 / & 1 0 . ** ~ () 前順命 Function () 有利命 Function () 有利命 <td></td>	
OK	貼り付け(ビ) [戻す(ビ) [キャンセル] ヘルブ	

図E.1 新しい変数を計算するダイアログボックス

- 初任給(salbeg)と現在の給与(salnow)の算術平均をあらわす変数を作成して ください。
- 変換メニューの変数の計算を選択します。
 目標変数ボックスにavesal(任意の変数名)、数式に(salbeg + salnow)/2と
 入力します。
 OKボタンをクリックします。
- 4. MEAN関数を使って、3.と同様の平均をあらわす変数を作成してください。2つの結果に違いはありますか?
- 変換メニューの変数の計算を選択します。
 目標変数ボックスにmeansal(任意の変数名)、数式にMEAN(salbeg,salnow)
 を入力します。
 OKボタンをクリックします。

分析メニューの記述統計の記述統計を選択します。 avesalとmeansalを変数リストボックスに移動し、OKボタンをクリックします。

<mark>翻</mark> 記述統計量	×
	オブション(の) ブートストラップ(B)
標準化された値を変数として保存(Z)	
OK「貼り付け(P) / 戻す(R) (キャンセ	ル ヘルプ

記述統計量

	度数	最小値	最大値	平均値	標準偏差
avesal	474	5190.00	42996.00	10287.1308	4858.37139
meansal	474	5190.00	42996.00	10287.1308	4858.37139
有効なケースの数 (リストごと)	474				

図E.2 2つの平均を比較する

統計量の値がすべて等しいので、2つの変数には違いがないと考えられます。

5. 性別と人種をあらわすgenderraceという質的データから、ダミー変数を作成してく ださい。コード化の例は以下を参考にしてください。

genderrace	genderrace1	genderrace2	genderrace3
1(白人男性)	1	0	0
2(白人以外男性)	0	1	0
3(白人女性)	0	0	1
4(白人以外女性)	0	0	0

OKボタンをクリックします。

図E.3 genderrace1を作成

⑦ 同じ要領で、genderrace2(数式:genderrace = 2)、genderrace3(数式: genderrace = 3)

を作成します。シンタックスの貼り付けを使って作成してもよいでしょう。

🔚 *シンタックス1 - PASW Statis	etics シンタックス エディタ
ファイ) 編集(表示(データ 変換(分析(ダイレクト マーケミ グラフ ユーティリミ実行(ツール ウィンド! ヘルプ
۱ 🛄 🖨 🔚 🔁	
/* 🔵 🍼 🐚	アクティブ: データセット1 🔻
DATASET ACTIVATE COMPUTE COMPUTE COMPUTE EXECUTE.	1 2 DATASET ACTIVATE データセット1. 3 COMPUTE genderrace1=genderrace = 1. 4 COMPUTE genderrace2=genderrace = 2. 5 COMPUTE genderrace3=genderrace = 3. 6 EXECUTE. 7 ►
	PASW Statistics プロセッサは使用可能です In 7 Col 0 In 7 Col 0

図E.4 シンタックスを使ったダミー変数の作成

📰 *BANK.sav [データセット1] – PASW Statistics データエディタ							
ファイル 編集(生表示(データ) 変換(分析(ダイレクト マーケテ グラフ(ユーティリテ ウィンドウ ヘルプ)							
🔄 🗄 🖨 🛄 🗠 🦘 📓 🏪 🗐 🏦 🚟 🚟 📻 (
1 : id	6	28			表示	〒: 17 個 (17 変数中)	
	avesal	meansal	genderrace	genderrace1	genderrace2	genderrace3	
1	24480.00	12240.00	1	1.00	0.	.0 📥	
2	65400.00	32700.00	1	1.00	.0	.0	
3	32160.00	16080.00	1	1.00	.0	.0	
4	27900.00	13950.00	1	1.00	.0	.0	
5	45750.00	22875.00	1	1.00	.0	.0	
6	40246.00	20123.00	1	1.00	.0	.0	
7	22980.00	11490.00	1	1.00	.0	.0	
8	19500.00	9750.00	1	1.00	.0	.0	
·	4						
データ ビュー(D) 変数 ビュー(V)							
			PASW Stat	tistics プロセッサ	は使用可能です		

図E.5 作成したダミー変数を表示

第3章 日付データと文字データの操作

文字型変数の変換

1. POST&TEL.savファイルを開いてください。この郵便番号コードを示すpostcodeの 値は正しいものではありません。最初の文字Dの後にEを追加しなければなりませ ん。変数の計算手続きと文字型関数を組み合わせて、イギリスの正しい郵便番号 コードをあらわす新しい変数newpostを作成してください。

 ファイルを開くボタンをクリックし、POST&TEL.savを選択し、開くボタン をクリックします。
 変換メニューの変数の計算をクリックします
 目標変数ボックスにnewpostと入力します。
 型とラベルボタンをクリックして型の文字型を選択し、続行ボタンをクリックします。
 文字式ボックスに
 CONCAT(CHAR.SUBSTR(postcode,1,1),'E',CHAR.SUBSTR(postcode,2))と 入力します。

OKボタンをクリックします。

🔢 変動の計算		×
<mark>課 変動の計算</mark> 目標変数①: newpost 型とラベル(<u>し</u>) ♪a postcode ♪a tel#	文字式(G) = CONCAT(CHAR.SUBSTR(postcode,1,1),'E',CHA	R.SUBSTR(postcode,2)) 関数グループ(G): 乱数 検索 有意確率 統計 文字列 時間の長さの作成 時間の長さの抽出
「F…」(任意のケースの選択 の)	*** ~ () 削除 CHAR.SUBSTR(strexpr,pos[,length])。文字 列。strexprの位置 pos で始まる分の文字列を戻します。任意の3つ目の引入す字列の文字数を表します。任意の引数は部分文字列の文字数を表します。任意の引数 length を省略した場合 は、strexprの pos 文字目から最後までの部分文字 列を返します。たとえば、CHAR.SUBSTR('abcd', 2.1) 2011 「Ibcd」 本ivit OHAP SUBSTR('abcd', 2.1) 条件)	Length Lower Ltrim(1) Ltrim(2) Mblen.Byte Normalize Ntrim Replace(3)

図E.6 newpostを作成する

 tel#として保存されている電話番号のうち、01で始まるロンドンの市外局番の値は 正しいものではありません。変数の計算手続きで、条件を指定し、文字型関数を 使ってロンドンの電話番号の最初の0の直後に8を挿入してください(01を081へ変 更)。

 変換メニューの変数の計算を選択します。
 目標変数ボックスにtel#と入力します。
 文字式ボックスに、CONCAT(CHAR.SUBSTR(tel#,1,1),'8',
 CHAR.SUBSTR(tel#,2))と入力します。
 IFボタンをクリックし、IF条件を満たしたケースを含むを選択します。
 条件式CHAR.SUBSTR(tel#,1,2) = '01'と入力し、続行ボタンをクリックします。
 OKボタンをクリックします。



POST&	「EL.sav【データセ	zット2] – PASW Statist	ics データ エディタ							
ファイル(E)	編集(E) 表示(V) データ(D) 変換(T)	分析(A) ダイレク	トマーケテ	ィングM	グラフ(3	り ユーティ	リティ(U)	ウィンドウ@	り ヘルゴ(円)
		r 🤉 📱		11 🎽			sta 📰	(ABC
1 : postcode	D27A	F							表示	示: 3 個 (3 変数中)
	postcode	tel#	newpost	var	1	/ar	var	var	var	var
1	D27AF	081 284 3333	DE27AF							
2	D136TH	081 234 5678	DE136TH							
3	D305TY	0283 45672	DE305TY							
4	D16YH	081 234 6794	DE16YH							
5	D195RT	081 435 7843	DE195RT							
6	D26DF	021 3456 678	DE26DF							i i i
7	D73GS	081 728 9830	DE73GS							
8	D407YU	071 234 5678	DE407YU							
9	D124FG	091 234 5683	DE124FG							
10	D34FV	081 234 5678	DE34FV							
4.4	1									•
						PASW Sta	tistics プロセ	ッサは使用す	J能です	

図E.7 計算ダイアログと実行結果

日付型データの変換

- 3. HOSPITAL.savファイルを開いてください。このファイルには4人の患者の情報が 記録されています。
- ファイルメニューの開くのデータを選択します。HOSPITAL.savを選択し、開くボタンをクリックします。
- 4. 入院期間(日数単位)を計算してください。
- 変換メニューの日付と時刻ウィザードを選択します。
 日付と時刻ウィザードで日付と時刻で計算を選択し、次へボタンをクリックします。
 日付と時刻ウィザード(ステップ1/3)で、2つの日付間の時間単位数の計算を選択し、次へボタンをクリックします。
 日付と時刻ウィザード(ステップ2/3)で、日付1に退院日、引く日付2に入院日を移動し、単位を日に変更します。
 次へボタンをクリックします。
 日付と時刻ウィザード(ステップ3/3)で、変数ボックスに入院期間と入力し、完了ボタンをクリックします。
- 5. 各患者の入院時の年齢(満年齢)を計算してください。

 ・
 ・
 変換メニューの日付と時刻ウィザードを選択します。
 日付と時刻ウィザードで日付と時刻で計算を選択し、次へボタンをクリックします。
 日付と時刻ウィザード(ステップ1/3)で、2つの日付間の時間単位数の計算を選択し、次へボタンをクリックします。
 日付と時刻ウィザード(ステップ2/3)で、日付1に入院日、引く日付2に
 誕生日日を移動します(単位はデフォルトの年を使用します)。
 次へボタンをクリックします。
 日付と時刻ウィザード(ステップ3/3)で、変数ボックスに入院時の年齢と
 入力し、完了ボタンをクリックします。

HOSPIT	📰 *HOSPITAL.sav [データセット3] – PASW Statistics データエディタ								
ファイル(E 編	ファイル匠 編集(E) 表示(M) データ(D) 変換(D) 分析(A) ダイレクトマーケティング グラフ(G) ユーティリティ(ウィンドウ(M) ヘルプ(H)								
	😑 🗄 🌐 💷 🗠 🥣 🎬 🏪 重 👫 🚟 📰 🐴 🖓 🙆								
1:患者氏名	Sm	ith				表示: 6	個 (6 変数中)		
	患者氏名	誕生日	入院日	退院日	入院期間	入院時の年齢	var		
1	Smith	21-May-1946	22-Sep-1993	07-Oct-1993	15	47			
2	Rogers	26-Sep-1956	30-Jun-1993	27-Aug-1993	58	36			
3	Grey	27-Sep-1964	24-May-1993	19-Sep-1993	118	28	3		
4	Harris	08-Jul-1912	19-Sep-1992	13-Jul-1993	297	80			
5							-		
データ ビ	データ ビュー(D) 変数 ビュー(V)								
				PASW Statisti	cs プロセッサは使	用可能です			

図E.8 入院時の年齢

- 6. 患者の氏名、入院日数、入院時の年齢を記述する出力を作成してください。
- ⑦ 分析メニューの報告書のケースの要約を選択します。

患者氏名、入院期間、入院時の年齢を変数リストボックスに移動します。 OKボタンをクリックします。

🏢 ケースの要約			X
 ▲ 入院日 ▲ 誕生日 ▲ 退院日 	*	変数(V): ▲ 患者氏名 ◆ 入院期間 ◆ 入院時の年齢 グループ変数(G):	統計量(S) オブション(O)
	•		
☑ ケースの表示(L)			
👿 ケースの制限は最初の(E) 🛛 1	00		
📝 有効なケースだけを示す(H)			
📃 ケース番号を示す(C)			
ок [貼り付け(<u>ی</u> (چ	す® キャンセル ·	ヘルプ

		患者氏名	入院期間	入院時の年齢
1		Smith	15	47
2		Rogers	58	36
3		Grey	118	28
4		Harris	297	80
合計	度数	4	4	4

ケースの集計®

a. 最初の 100 のケースlこ制限されています。

第4章 ケースの選択とファイルの分割

CUSTOMER.savは金融機関の顧客のデータです。このデータには、顧客番号(id)、 収入(income)、性別(gender)、地域(region)、婚姻状況(marital)、子供の数(child)、 車の所有(car)、担保の有無(mortgage)、口座の種類(acct)が記録されており、358 のケースがあります。

ケースの選択

- 1. CUSTOMER.savを開いてください。
- ファイルメニューの開くのデータを選択します。 CUSTOMER.savを選択し、開くボタンをクリックします。
- 2. 収入 (income) の平均値を求めてください。
- 分析メニューの記述統計の記述統計を選択します。 incomeを変数リストボックスに移動し、OKボタンをクリックします。

図E.9 ケースの集計手続きと結果

	度数	最小値	最大値	平均値	標準偏差
収入	358	5960	61476	27124.01	12522.262
有効なケースの数 (リストごと)	358				

記述統計量

図E.10 記述統計量

- 3. ケースの選択ダイアログボックスを使用して、収入が平均値(2.で求めた値)以 上のケースを選択してください。
- データメニューのケースの選択を選択します。
 IF条件が満たされるケースを選択してIFボタンをクリックします。
 条件式income >= 27124.01を入力して続行ボタンをクリックします。
 OKボタンをクリックします。

🏭 ケースの選択	
 ✓ 顧客番号 [id] ✓ 年齢 [age] ◆ 性別 [gender] ◆ 地域 [region] ✓ 収入 [income] ◆ 婚姻状況 [marital] ✓ 子供の数 [child] ◆ 車の所有 [car] ◆ 担保の有無 [mortga ◆ 口座の種類 [acct] 	 選択状況 ● すべてのケース(A) ● IF 条件が満たされるケース(C) IF income >= 27124.01 ○ ケースの無作為抽出(D) サンブル(S) ● 日付かケース番号の範囲(B) 範囲(N) ○ フィルタ変数の使用(U):
現在の状況: ケースの選択な OK 話し	 ● 選択されなかったケースを分析から除外(F) ● 選択されたケースを新しいデータセットにコピー(0) データセット名(8): ● 選択されなかったケースを削除(L) し J付け(P) 戻す(R) キャンセル ヘルプ

図E.11 ケースの選択

- 4. 選択されたケースには、男性と女性のどちらが多いでしょうか? 婚姻状況ではど のカテゴリが多いでしょうか?
- 分析メニューの記述統計の度数分布表を選択します。 gender、maritalを変数リストボックスに移動し、OKボタンをクリックします。

… 度数分布表		×		
 acct ✓ age car ✓ child ✓ filter_\$ ✓ id ✓ income ✓ mortgage ✓ region 	変数(\):	統計量(S) 図表(C) 書式(F) ブートストラップ(B)		

性別

		度数	バーセント	有効バーセン ト	累積バーセン ト
有効	女性	80	54.1	54.1	54.1
	男性	68	45.9	45.9	100.0
	合計	148	100.0	100.0	

婚姻状況

		度数	バーセント	有効バーセン ト	累積バーセン ト
有効	既婚	85	57.4	57.4	57.4
	未婚	63	42.6	42.6	100.0
	合計	148	100.0	100.0	

図E.12 度数分布表手続きと実行結果

男性より女性の方が多いようです。また、未婚者より既婚者のほうが多いようです。

- 5. ケースの選択を解除してください。
- データメニューのケースの選択を選択します。 すべてのケースを選択してOKボタンをクリックします。

ファイルの分割

- 6. ファイルの分割ダイアログボックスを使用して、地域(region)でグループを作 成し比較するように指定をしてください。
- データメニューのファイルの分割を選択します。
 グループの比較を選択します。
 regionをグループ化変数リストボックスに移動します。
 OKボタンをクリックします。

🏭 ファイルの分割	
 acct age car child filter_\$ gender id income marital mortgage 現在の状態: グループ毎に分 	 ② 全てのケースを分析(A) ◎ グルーブの比較(C) ◎ グルーブごとの分析(Q) グルーブ化変数(G): グルーブ化変数(G): ● グルーブ変数によるファイルの並び替え(S) ◎ ファイルはすでに並び替え済み(F) 析しない。 ● 戻す(R) キャンセル ヘルプ

図E.13 ファイルの分割ダイアログボックス

7. 地域による収入の違いを比較できるような手続きを実行してください。

分析メニューの記述統計の記述統計を選択します。 incomeを変数リストボックスに移動し、OKボタンをクリックします。

		6C.)24	Male			
地域		度数	最小値	最大値	平均値	標準偏差
郊外	収入	63	9785	51468	26211.25	12366.113
	有効なケースの数 (ソストごと)	63				
市街地	収入	110	5960	61476	24762.93	10575.899
	有効なケースの数 (ソストごと)	110				
都市	収入	145	6531	60635	28343.70	13218.709
	有効なケースの数 (リストごと)	145				
農村	収入	40	8005	60176	30633.20	14116.449
	有効なケースの数 (ソストごと)	40				

記述統計量

図E.14 実行結果

収入の平均値を比較すると、収入が最も多いのは農村部、最も少ないのは市街地で あることがわかります。本当に違いがあるのかを調べるためには、さらに分析が必要 です。

第5章 グループ集計とケースの重み付け

CUSTOMER.savのデータには、同じid(顧客番号)を持つケースが複数入力されて います。このデータをグループ集計し、1ケースが1個人を表すように集計の単位を 指定します。

1. データのグループ集計機能を使用して、id(顧客番号)ごとに次の集計を行っ てケースを作成します。

変数名	集計方法
age	age(年齢)の平均値
income	income(収入)の平均値
child	child(子供の数)の最大値
marital	marital(婚姻状況)の最後の値
region	region(地域)の最後の値

ファイルメニューの開くのデータを選択します。
 CUSTOMER.savを選択し、開くボタンをクリックします。
 データメニューのグループ集計を選択します。
 idをブレーク変数リストボックスに、age、income、child、marital、region
 を変数の集計リストボックスに移動します。
 関数ボタンを使用して、集計方法を上記の指定内容に変更します。
 変数名とラベルボタンを使用して、変数名を上記名に変更します。

 集計ファイルをAGCUSTOMER.savとして保存します。保存先は C:\train\DataManiフォルダを指定します。

第 集計変数のみを含む新しいデータファイルを作成するを選択します。 ファイルボタンをクリックし、C:\train\DataManiフォルダで、ファイル名テキ ストボックスにAGCUSTOMER.savと入力し、保存ボタンをクリックします。 OKボタンをクリックします。 データのグループ集計:一致する名前の警告ダイアログボックスで、上書きボ タンをクリックします。

🏭 データのグループ集計						
 acct age car child filter_\$ gender marital mortgage region 	ブレーク変数(B): ✓ id 集計変数 変数の集計(S): age = MEAN(age) income = MEAN(income) child = MAX(child) marital = LAST(marital) region = LAST(region) 関数(F) 変数名とラベル(N)					
保存 ② アクティブなデータセッ ③ 集計変数のみを含む新し データセット名(D): [③ 集計変数のみを含む新し ファイル(し) C 「非常に大きなデータセットの ③ ファイルは既にブレーク ⑤ グループ集計前にファイ	トに集計変数を追加(D) いデータセットを作成する(E) いデータ ファイルを作成する(W) ::train\DataMani\AGCUSTOMER.sav マオプション - *変数で並べ替えられている(A) 「ルを並べ替え(G)					
 ブルーブ集計前にファイ OK< 貼り 	 □ グループ集計前にファイルを並べ替え(G) OK 貼り付け(P) 戻す(R) キャンセル ヘルプ 					

	JSTO	MER.sav [データ	!ቲット9] – PASW S	tatistics データ エデ	伊				
ファイル	(E) 編	[集(E) 表示() :	データ(D) 変換(D) 分	淅(A) ダイレクトマー	-ケティング() グラ	ラフ(の) ユーティレ	ノティ(ビ ウィン	・ドウ 🕐 へル	,ゴ(H)
					#1 🍇	🖬 🔚 4	ا 🖽 🕻		9
1 : id		12701					144	表示: 6 個 (6 変	(中)
		id	age	income	child	marital	region	var	
1		12701	23.00	18766.00	0	1	3		
2		12702	30.00	9915.00	1	2	4		
3		12703	45.00	21881.00	0	2	4		
4		12704	50.00	46794.00	2	1	2		
5		12705	41.00	20721.00	0	1	3		
6		12706	20.00	16688.00	1	2	3		
7		12707	46.00	39068.00	0	1	4		
8		12708	50.00	27740.00	1	1	3		
9		12709	42.00	33584.00	3	2	3		
10		12710	57.00	19621.00	1	1	2		
11		12711	63.00	47630.00	0	1	3		-
データ	Ľ٦	<mark>.ー(D)</mark> 変数 ビ	⊐ – (V)						
					PASW Statis	tics プロセッサは	使用可能です		

図E.15 データをidで集計

3. 次に、CUSTOMER.savを、region(居住地域)の単位で集計します。次の集計を ダイアログボックスで指定してください。

変数名(変数ラベル)	集計方法
age_mean(平均年齡)	最適な方法を選択
income_mean (平均収入)	最適な方法を選択
child_mean (子供の数の平均)	最適な方法を選択
break (地域の人数)	最適な方法を選択

データメニューのグループ集計を選択します(必要に応じて戻すボタンをクリックします)。

regionをグループ変数リストボックスに、age、income、childを変数の集計 リストボックスに移動します。 変数名とラベルボタンをクリックして上記の変数ラベルを入力します。

ケースの数を選択して名前ボックスにBREAKと入力します。

- ファイル名をAGRESION.savとして保存します。保存先は、 C:\train\DataManiフォルダを指定します。
- 全 集計変数のみを含む新しいデータファイルを作成するを選択します。 ファイルボタンをクリックし、C:\train\DataManiフォルダで、ファイル名テキ ストボックスにAGRESION.savと入力し、保存ボタンをクリックします。

OKボタンをクリックします。

🏭 データのグループ集計					
データのグループ集計 またt ふ acct ふ age car ぐ filter_\$ gender が income marital デ mortgage	× ブレーク変数(B):				
	▼ ケースの数(C) 名前(M): BREAK				
 ◎ アクティブなデータセットに集計変数を追加(D) ◎ 集計変数のみを含む新しいデータセットを作成する(E) データセット名(D): 					
 ● 集計変数のみを含む新しいデータ ファイルを作成する(W) □アイル(L) C:\Train\DataMani\AGRESION.sav 					
- 非常に大きなデータセットのオプション □ ファイルは既にブレーク変数で並べ替えられている(A) □ グループ集計前にファイルを並べ替え(G)					
ок 貼り作	オナ® 戻す® キャンセル ヘルプ				

AGREGIO	🗄 AGREGION.sav [データセット6] – PASW Statistics データ エディタ							
ファイル(E 編	譙(E) 表示() き	データ(D) 変換(D) 分析	f(A) ダイレクトマーケティン	ノグ グラフ(⑤ ユーティリ	ティリウィンドウ(⊻ ヘルプ(⊞		
😑 🖩 🖨 🛄 🖛 🛥 🎬 🏪 🚅 👫 🚟 🚟 🐴 🖓 🤆								
1 : region	1:region 1 表示:5個(5変数中)							
	region	age_mean	income_mean	child_mean	BREAK	var		
1	1	42.60	26211.25	1.13	63			
2	2	40.36	24762.93	1.31	110			
3	3	42.15	28343.70	1.06	145	3		
4	4	47.70	30633.20	.80	40			
5								
	4					▼ 		
データ ビュー(D) 変数 ビュー(V) PASW Statistics ブロセッサは使用可能です								

図E.16 データのグループ集計ダイアログと集計結果

第6章 ファイルの結合 – ケースの追加

銀行の顧客に関する97年度(97DATA.sav)と98年度(98DATA.sav)のデータがあ ります。データに入力されている情報は、顧客番号(id)、口座の種類(acct)、口 座開設時の残高(openbal)、現在の残高(currbal)です。この2つのデータの結合を 行い、1つのデータファイルを作成します。さらに、そのデータを顧客番号で集計し たファイルを作成します。

- 97年度のデータを開きます。ファイルの結合ダイアログボックスを使用して、 98年度のデータファイルを追加します。ケースソースを示すための新しい変数 sourceを作成します。
- ファイルメニューの開くのデータを選択します。
 97DATA.savを選択し、開くボタンをクリックします。
 データメニューのファイルの結合のケースの追加を選択します。
 外部SPSSデータファイルを選択し、参照ボタンをクリックします。
 98DATA.savを選択し、開くボタンをクリックします。
 続行ボタンをクリックします。
 変数としてソースケースを示すを選択し、変数名をSOURCEに変更します。
 OKボタンをクリックします。

 グループ集計ダイアログボックスを使って集計ファイルを作成します。 ブレーク変数には、顧客番号(id)を投入し、ブレークグループのケース数を N BREAKの変数名で保存します。

変数名	集計方法	(変数ラベル)
~~ ~		\XXX / '/V/

openbal	口座開設時残高の平均
currbal	現在の残高の平均

- データメニューのグループ集計を選択します。
 idをグループ変数リストボックスに、openbal、currbalを変数の集計リストボックスに移動します。
 変数名とラベルボタンをクリックして上記の変数名、変数ラベルを入力します。
 ケースの数を選択します。
- 新しいデータファイルを作成し、保存します。C:\train\DataManiフォルダに、
 97 98DATA.savのファイル名で保存します。
- 第計変数のみを含む新しいデータファイルを作成するを選択します。 ファイルボタンをクリックし、C:\train\DataManiフォルダで、ファイル名テキ ストボックスに97_98DATA.savと入力し、保存ボタンをクリックします。 OKボタンをクリックします。

🎫 *97_98DATA.sav [データセット8] - PASW Statistics データ エディタ							
ファイル(編集(E 表示(V データ(E 変換(E 分析(A ダイレクト マーケティングラフ(C ユーティリティ ウィンドウ(ヘルプ(E							
😑 🗄 🖨 🛄 🗠 🤉 🦉 🕌 🗯 👫 🚟 🗮 🖄 🚟							
1 : id	12701				表示:4個	(4 変数中)	
	id	openbal	currbal	N_BREAK	var	Va	
1	12701	1000.00	1005.32	1			
2	12702	100.00	144.51	1			
3	12703	225.00	58.35	2			
4	12704	2000.00	2022.02	1			
5	12705	1027.33	1080.63	3			
6	12706	614.67	691.48	3			
7	12707	10.00	55.03	1			
8	12708	5.00	10.55	1			
9	12709	50.00	57.21	1		-	
データービュ	<mark>ュー(D)</mark> 変数 ビ	- − (V)					
			PASW Statistics	プロセッサは使用可能	です		

図E.17 97_98Data.savファイル

第7章 ファイルの結合 – 変数の追加

1対1の結合

第5章の演習問題で作成したAGCUSTOMER.savと第6章の演習問題で作成した 97_98DATA.savの2つのデータの結合を行います。

- 1. AGCUSTOMER.sav ファイルと97_98DATA.savファイルを検討し、キー変数を決定します。
- ファイルメニューの開くのデータを選択します。
 AGCUSTOMER.savを選択し、開くボタンをクリックします。
 データビュー、変数ビューを表示して内容を確認します。
 同じように、97_98DATA.savも確認します。

どちらのデータもid(顧客番号)ごとにケースが集計されています。これをキー変数として使うことにします。

- 2. 2つのファイルをキー変数でソートします。
- G AGCUSTOMER.savが表示されているデータエディタで、データメニューのケースの並び替えを選択します。
 idを並べ替えボックスに移動し、OKボタンをクリックします。
 同じように、97_98DATA.savもidで並び替えます。
- 変数の追加ダイアログボックスを使用して、AGCUSTOMER.savと 97_98DATA.savを結合します。
- G AGCUSTOMER.savが表示されているデータエディタで、データメニューのファイルの結合の変数の追加を選択します。
 開いているデータセットを選択し、97_98DATA.savを選択して続行ボタンをクリックします。
 キー変数によるケースの結合を選択します。
 除外された変数リストボックスのidをキー変数リストボックスに移動します。
 OKボタンをクリックします。

■ 変数の追加 データセット8		X
愛教の追加 データセット8 除外された変数(E): 名前の変更(A) マー変数によるケースの結合 の 両方のコッく しがたっつ を提供(P)	*	¥しいアクティブデータセット(N): age(*) income(*) child(*) marital(*) region(*) openbal(+) currbal(+) N_BREAK(+)
 ● 両方のファイルかケースを提供(B) ○ アクティブでないデータセットが検索テーブル(K) ○ アクティブなデータセットが検索テーブル(A) 	*	id
📃 変数としてケース ソースを示す(): ソース01		
(*)=アクティブデータセット (+)=データセット8		
OK 貼り付けの 戻すの	++	ンセル

AGCUS	TOMER.sav [データセ	2ット9] – PASW	Statistics データ	፤ ፲ምሳያ				×
ファイル(E)	編集(E) 表示(V) デー	タ(D) 変換(T) 分	ì析(A) ダイレクト	マーケティング	() グラフ(の) ユー	・ティリティ(ぜ ウ	ィンドウ(\) ヘルプ	Э
2	🖨 🛄 🛯	ດ 🦳 🚦	š 🄚 🗐	#1 🕺	ş 🔛 🚍			•
1 : id	12701						表示: 9 個 (9 変数	(中)
	income	child	marital	region	openbal	currbal	N_BREAK	
1	18766.00	0	1	3	1000.00	1005.32	1	
2	9915.00	1	2	4	100.00	144.51	1	
3	21881.00	0	2	4	225.00	58.35	2	
4	46794.00	2	1	2	2000.00	2022.02	1	
5	20721.00	0	1	3	1027.33	1080.63	3	
6	16688.00	1	2	3	614.67	691.48	3	
7	39068.00	0	1	4	10.00	55.03	1	
8	27740.00	1	1	3	5.00	10.55	1	
9	33584.00	3	2	3	50.00	57.21	1	
10	19621.00	1	1	2	511.50	586.84	2	
11	47630.00	0	1	3	1000.00	1082.27	1	•
	1							
データ ビ	<mark>ュー(D)</mark> 変数 ビュー	- (V)						
				PASW	Statistics プロセッ	サは使用可能です	-	

図E.18 変数の追加ダイアログと実行結果

- 4. 結合後のファイルを、FULLDATA.savのファイル名で保存します。
- ファイルメニューの名前を付けて保存を選択します。
 ファイル名にFULLDATA.savと入力して保存ボタンをクリックします。

検索テーブルによる結合

1対1の結合で作成したFULLDATA.savに、検索テーブルによる結合を使用して、 第5章の演習問題で作成したAGRESION.savを結合します。結合したデータで、各顧 客のデータと地域ごとの平均値とを比較します。 1対1の結合で作成したFULLDATA.savを開きます。キー変数を決定し、ソートしておきます。

AGRESION.savはregion(居住地域)で集計されています。FULLDATA.savにも regionがあります。regionをキー変数として、AGRESION.savを検索テーブルとし たファイルの結合を行います。

- ファイルメニューの開くのデータを選択します。
 FILLDATA.savを選択し、開くボタンをクリックします。
 データメニューのケースの並び替えを選択します。
 regionを並び替えリストボックスに移動し、OKボタンをクリックします。
- 変数の追加ダイアログボックスを開きます。結合するファイル (AGRESION.sav)を選択します。外部ファイルが検索テーブルであることを 指定します。
- FULLDATA.savが表示されているデータエディタで、データメニューのファイルの結合の変数の追加を選択します。
 外部SPSSデータファイルを選択し、参照ボタンをクリックして
 AGRESION.savを選択し、開くボタンをクリックします。
 続行ボタンをクリックします。
 キー変数によるケースの結合を選択し、アクティブでないデータセットが検索テーブルを選択します。
 除外された変数リストボックスのregionをキー変数リストボックスに移動します。
 OKボタンをクリックします。



🔠 *FULLDA	TA.sav [データ	22911] - PA	ASW Statist	ics データ エディタ	<u>,</u>		
ファイル(E 編	集(E) 表示() ÷	データ(D 変換(<u>円</u> 分析(A) ダー	イレクトマーケテ	ィン: グラフ() ユ	ーティリティ(ウ	マンドウ(2 ヘルゴ(日
			~		# 👬	2	
26 : currbal	11	1.34				表	辰示: 13 個 (13 変数中)
	child	marital	region	openbal	currbal	N_BREAK	age_mean
5	0	1	1	683.00	737.13	3	42.60 📥
6	0	2	1	57.00	64.36	2	42.60
7	0	2	1	61.00	114.05	1	42.60
8	3	2	1	485.50	529.33	2	42.60
9	2	2	1	487.50	549.72	2	42.60
10	1	1	1	48.00	111.83	1	42.60
11	0	1	1	1149.50	1168.72	2	42.60
12	1	1	1	265.00	317.08	2	42.60 🚽
データ ビー							
				PASW S	itatistics プロセッち	力は使用可能です	

図E.19 変数の追加ダイアログと実行結果

結合したファイルを使用して、各顧客の収入と地域ごとの収入の平均値とを比較します。

収入が地域の平均より低いことをあらわす変数を作成し、その度数分布表を作成し て、収入が地域の平均より低い人がどれくらいいるのか調べます。

 ・変換メニューの変数の計算を選択します。

 目標変数ボックスにlow_income (任意の名前)、数式にincome <
 income_meanと入力し、OKボタンをクリックします。

📰 変数の計算			×
目標変数①: low_income 型とラベル(L)	= *	数式(E): income < income_mean + < > 7 8 9 - <= >= 4 5 6 * = ~= 1 2 3 / & 1 0 . / & 1 0 . #* ~ () 削除 J 認知グルーブ(G): オペて 算術 CDF と非心度 CDF 変換 現在の日付と時刻 算術日 日付作成 関数と特殊変数(F):	
0	ĸ	貼り付け(P) 戻す(R) キャンセル ヘルプ	

図E.20 変数の計算ダイアログボックス

分析メニューの記述統計の度数分布表を選択します。 low_incomeを変数リストボックスに移動し、OKボタンをクリックします。

		度数	バーセント	有効バーセント	累積バーセント
有効	.00	77	38.9	38.9	38.9
	1.00	121	61.1	61.1	100.0
	合計	198	100.0	100.0	

low_income

図E.21 low_incomeの度数分布表

全体の約6割の人が、収入が地域の平均より低いことがわかります。

- 8. 収入が地域の平均より高いグループと低いグループを比較し、現在の残高 (currbal) に違いがあるか調べます。
 - 7. で求めた変数を使って、2つのグループのcurrbalの平均値を比較します。
- 分析メニューの平均値の比較のグループの平均を選択します。 currbalを従属変数ボックスに、low_income(7.で求めた変数)を独立変数ボックスに移動し、OKボタンをクリックします。

🏭 グループの平均		×
 ゴルーナの平均 ダ age ▲ age_mean ▲ BREAK ④ child_mean ④ child_mean ④ income ✓ income_mean ▲ marital ✓ N_BREAK 	従属変数(D): ✓ currbal //1 /// 次(N) 独立変数: ✓ low_income	▼ オプション() ブートストラップ(B)
OK	貼り付け(の) 戻す(の) キャンセル	ヘルプ

116音音	郣	浩	書
-------	---	---	---

low_income	平均値	度数	標準偏差
.00	730.8029	77	927.00134
1.00	833.3249	121	1052.86227
合計	793.4552	198	1004.62397

図E.22 currbalの平均値の比較