

**中南米における
ファイナンスデータの
ジェンダー差別化をした
傾向・特徴の分析**

東海大学
朝日研究室
中村勇太

目次

1. 経済状況
2. 中南米での二輪自動車
3. 先行研究・研究目的
4. データ概要
5. 分析使用ソフトウェア
6. 分析手順
7. 基礎集計
8. 相関分析
9. クラスタ分析
10. まとめ・今後の課題

参考文献

Appendix

1. 経済状況

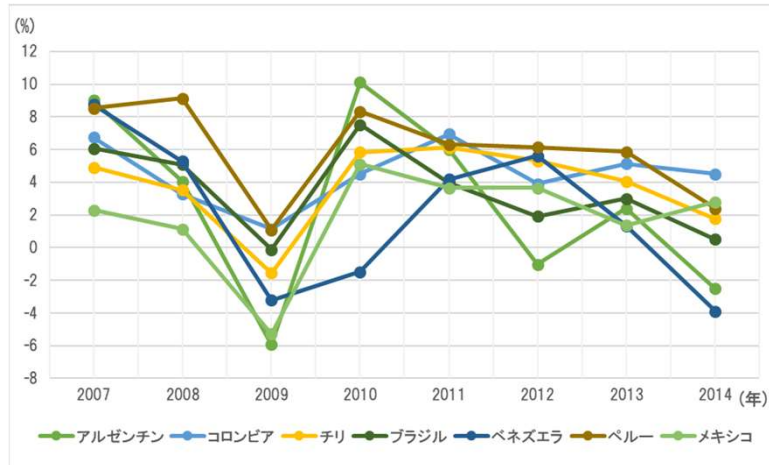


図1.中南米諸国の実質GDP成長率[1]

中南米の経済成長率は2007年の夏に金融不安を引き起こした米国のサプライローン問題を引き金に発生したリーマン・ショックの影響を受け、景気後退へ陥ったが2010年には10%近い高成長を達成した。しかし、2011年には成長は鈍化し、その後は成長率はマイナスまでに陥った。

中南米は物価上昇を抑えるため、政策金利による利下げを行った。2012年10月には年率7.25%と過去最低水準を更新した。これにより、**貧困層でも以前より簡単にローンを組むことが可能**となり、個人でローンを組み商品を購入するのが一般的となった。しかし、**ローン経験が少ない人から債務不履行が多発**するようになった。



図2.中南米の政策金利推移[2]

2. 中南米での二輪自動車

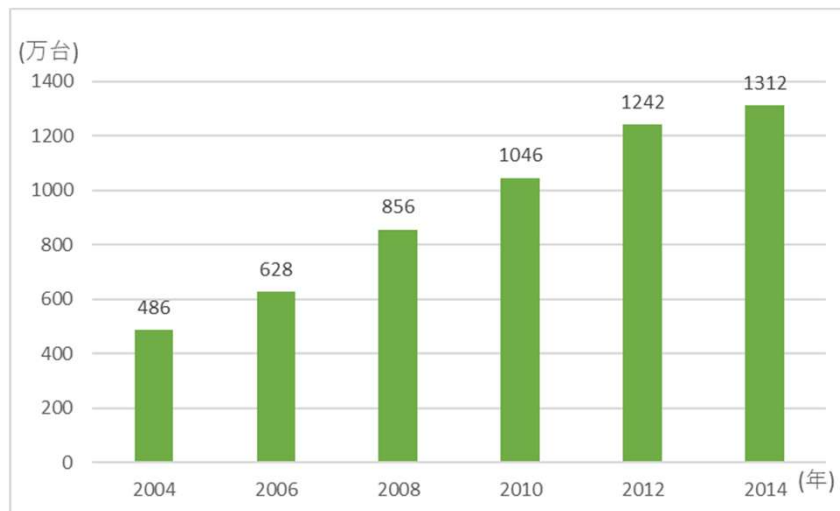


図3.中南米のバイク保有台数推移^[3]

中南米においては主な移動手段が自動車ではなく二輪自動車である。その為、保有台数は増加傾向にあり10年間で2.7倍まで増加している。

表1.2008年度中南米のバイク新車販売台数^[4]

順位	車名	年間販売数
1	HONDA・CG150	442,569台
2	HONDA・CG125	383,002台
3	HONDA・BIZ	196,171台
4	YAMAHA・YBR125	127,223台
5	HONDA・NXR150	111,416台
6	SUZUKI・EN125	84,760台
7	HONDA・CBX250 TWISTER	76,131台
8	HONDA・POP100	67,769台
9	YAMAHA・FAZER	37,816台
10	DAFRA・SPEED150	37,105台

日本メーカーの二輪自動車は中南米でも人気だ。新車販売台数においては上位9つを日本メーカーが独占している。

中南米では経済状況に関わらず
日本メーカーの自動二輪車が人気である。

3. 先行研究・研究目的

ローン未払い者の経歴の特徴についての分析,東海大学,厚木麻里^[5]

中南米における地域と顧客の性質・関係性に焦点を当て、債務者の特徴を研究した。ローン未払い者の特徴としては、地域ごとの経済格差が影響しており、平均主収入額が低い地域ほど未払い者になる傾向が強いことが判明した。しかし、収入額が低いからローン未払いになるとは言い切れず、学歴に関しては平均主収入額が多い第二学位卒業の方が平均主収入額の低い上級学位中退者よりも未払い者になりやすいということも判明した。



本研究では、学歴に焦点を当てつつ、債務者の経済状況だけでなく文化的背景を考慮し、分析を行い、ローン未払い者の傾向・特徴を探ることによって債務不履行に陥る債務者の早期発見に努める。

4. データ概要

表2. データ詳細

提供	二輪自動車販売会社
期間	2010年9月から2012年6月
対象エリア	中南米(A国)

- ◆ 全データ件数 :13,214件
- ◆ 欠損値・異常値削除後 :11,203件
- ◆ データ項目 :45変数
- ◆ データ内容
 - ◆ 年齢、性別、学歴、職業などの顧客情報
 - ◆ 利子率、借入金、サイズなどの商品情報
 - ◆ 未払い月、クレジットスコなどの信用情報

5. 分析使用ソフトウェア



図4.VMSパンフレット表紙^[6]

データマイニングソフト Visual Mining Studio(Ver 8.7)

データマイニングは、人間が機械学習を含む様々な手法を駆使し、膨大かつ複雑など様々な特性があるデータから、有用な知識、規則性を見つけ出し、現実に応用することを目的としています。Visual Mining Studioはデータマイニングに欠かせないデータ読み込み、集計、可視化、前処理、分析といったプロセスをマウスのみのシンプルな操作で実行できます。Visual Mining Studioを使えば、データから価値ある情報を得て、研究や戦略に生かします。定番から先進まで網羅した充実の分析機能を搭載しており、Visual Mining Studioを導入することで深い分析が可能になります^[7]。

6. 分析手順

基礎集計

中南米における債務者の状況把握

相関分析(VMS)

男女別で変数どうしの相関の違いがあるか確認

クラスター分析(VMS)

男女別に学歴毎の性質を鑑みるために分類

7. 基礎集計①

表2.学歴別性別と年代の割合

	データ数	性別		年代					
		男性	女性	10代	20代	30代	40代	50代	60代以上
教育歴なし	253	68.0%	32.0%	0.8%	34.8%	29.2%	20.9%	9.9%	4.3%
第一学位中退	121	76.0%	24.0%	1.7%	15.7%	28.1%	24.0%	27.3%	3.3%
第一学位卒業	437	77.1%	22.9%	0.7%	22.9%	26.8%	27.5%	18.3%	3.9%
第二学位中退	564	75.2%	24.8%	1.2%	25.4%	29.8%	25.2%	16.8%	1.6%
第二学位卒業	8114	69.8%	30.2%	0.9%	33.9%	31.7%	20.5%	11.3%	1.8%
上級中退	657	65.3%	34.7%	1.4%	44.6%	29.5%	15.4%	7.5%	1.7%
上級卒業	981	56.5%	43.5%	0.4%	26.0%	33.3%	24.8%	13.0%	2.4%
POS卒業	76	44.7%	55.3%	0.0%	22.4%	19.7%	34.2%	17.1%	6.6%
総計	11203	68.8%	31.2%	0.9%	32.7%	31.3%	21.2%	12.0%	2.0%

表3.中南米と日本の学歴関係

中南米	日本
教育歴なし	教育歴なし
第一学位	小学校
第二学位	中学校
上級	高校
POS	大学

データ数としては第二学位卒業者が約70%を占めている。
 しかし、**家庭環境や経済的影響から義務教育を受けられていない債務者が約12%存在しており教育格差が確認できる。**
 男女比は男性の方が圧倒的に多いが学歴が上がるにつれ女性の割合が高くなり、POS卒業者に関しては女性が55.3%と男性よりも多くなっている。**債務者の年齢は主に20代から40代が中心だ。**

7. 基礎集計②

表4.学歴別男女での在住地域率

	男性					女性				
	北部	北東部	中西部	南東部	南部	北部	北東部	中西部	南東部	南部
教育歴なし	28.5%	16.9%	11.6%	33.1%	9.9%	23.5%	14.8%	17.3%	34.6%	9.9%
第一学位中退	39.1%	42.4%	10.9%	5.4%	2.2%	27.6%	51.7%	13.8%	6.9%	0.0%
第一学位卒業	19.6%	34.7%	15.7%	21.1%	8.9%	29.0%	30.0%	12.0%	27.0%	2.0%
第二学位中退	24.5%	26.9%	13.0%	22.4%	13.2%	19.3%	26.4%	13.6%	22.9%	17.9%
第二学位卒業	18.6%	30.4%	14.1%	25.6%	11.3%	19.7%	32.2%	15.9%	20.4%	11.8%
上級中退	21.4%	17.0%	12.8%	38.9%	9.8%	20.2%	23.7%	14.9%	26.8%	14.5%
上級卒業	13.2%	23.8%	10.8%	43.0%	9.2%	20.1%	26.9%	10.8%	35.1%	7.0%
POS卒業	20.6%	17.6%	14.7%	38.2%	8.8%	19.0%	33.3%	11.9%	28.6%	7.1%
総計	19.2%	29.0%	13.7%	27.2%	10.9%	20.1%	30.5%	15.0%	23.2%	11.2%

学歴別に男女の在住地域を見たところPOS卒業以外は大きな差を見ることが出来なかった。POS卒業については、男性は南東部在住が半数近いのに対して、女性は北東部、南東部共に約30%ずつ異なっている。南東部は経済の中心的地域である為、高学歴の割合が高くなっている。しかし、教育歴なしの割合も高いことから5つ地域の中で最も貧富の差が大きいと思われる。また、北東部に低学歴の債務者が集中しているのは経済発展が遅れている為だと推測される。

7. 基礎集計③

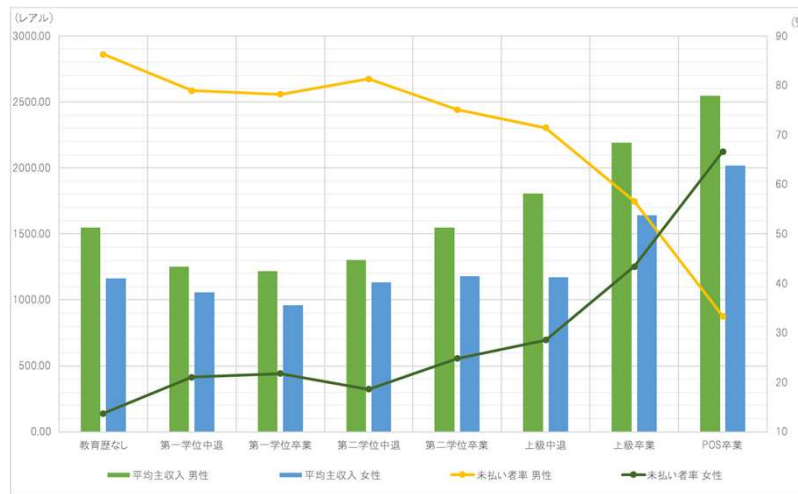


図5.学歴別男女での平均主収入と未払い者率

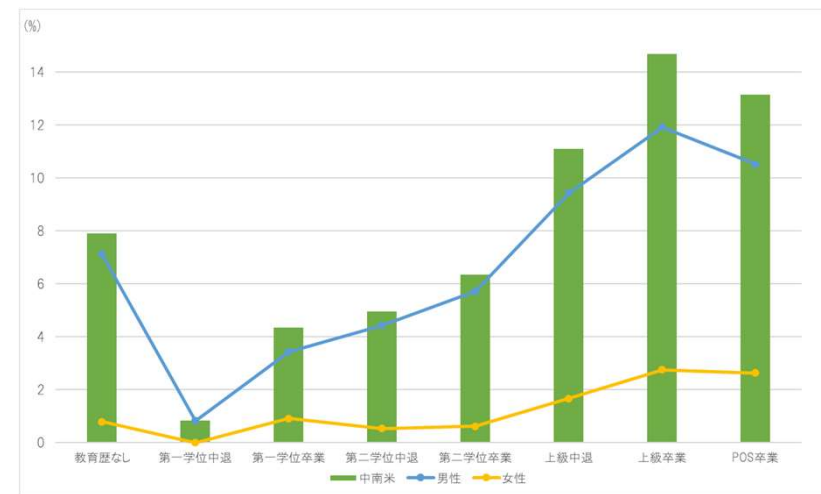


図6.学歴別購入二輪自動車サイズ割合平均

学歴が上がるにつれて男女ともに平均主収入が増加傾向にある。教育歴なしが第二学位卒業以下の主収入よりも高いのは就労年数の差と思われる。収入・学歴が高くなるにつれて女性の未払い者率が増加し、POS卒業に関しては割合が逆転している。また、購入サイズも学歴が上がるにつれ中型は増加傾向にあり、10%以上増加している。以上のことから債務者は性別によって特徴・傾向が異なっている可能性があると推測される。

8. 相関分析

債務者を性別で分けた際、データ同士の関係性が異なるかを確かめるために「Visual Mining Studio」を用いて相関分析を行った。使用変数は全て数値変数のため相関係数を求めた。

表5.相関分析男性結果

男性	不動産価値	主収入	分割数	利子%	頭金	借入金	未払い
在住年	0.41	0.29	-0.30	0.11	0.70	-0.02	-0.44
不動産価値		-0.26	0.36	0.29	-0.21	-0.29	-0.08
主収入			-0.92	-0.74	0.31	0.92	-0.87
分割数				0.50	-0.44	-0.82	0.72
利子%					0.07	-0.74	0.78
頭金						-0.03	-0.21
借入金							-0.73

表6.相関分析女性結果

女性	不動産価値	主収入	分割数	利子%	頭金	借入金	未払い
在住年	-0.34	0.69	0.13	-0.51	0.18	-0.03	-0.25
不動産価値		-0.11	0.51	-0.16	-0.70	0.19	0.02
主収入			0.23	-0.80	-0.06	0.47	-0.56
分割数				-0.53	-0.81	0.30	0.00
利子%					0.21	-0.77	0.76
頭金						-0.12	-0.31
借入金							-0.74

結果別に相関が強い上位4つを色付けしたところ一致しているのは1項目のみとなった。また、男性においては、主収入、分割数、利子、借入金の4項目が未払いとの強い相関なのに対して、女性においては、利子と借入金の2項目のみが未払いと強い相関である。更に分割数については女性側では全く相関がみられなかった。その他の項目でも相関の強弱が異なることから債務者は男性と女性で異なった傾向・特徴があるということが判明した。

9. クラスタ分析

表7.設定内容

距離計算法	ユークリッド距離
分析方法	Ward法
使用データ	年齢、不動産価値、主収入、分割回数、利子%、商品価格、頭金、借入金の各平均未払い者、在住地域の各割合

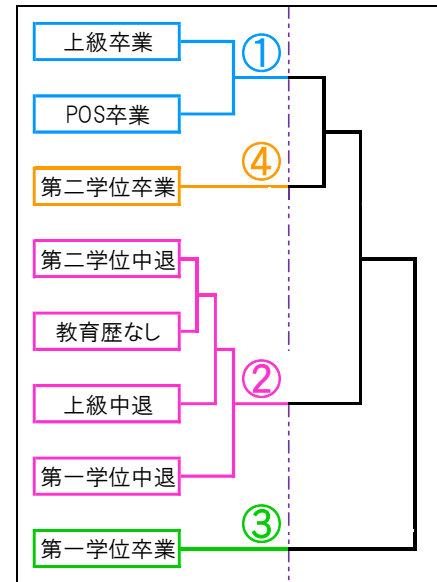


図7.デンドログラム(男性)

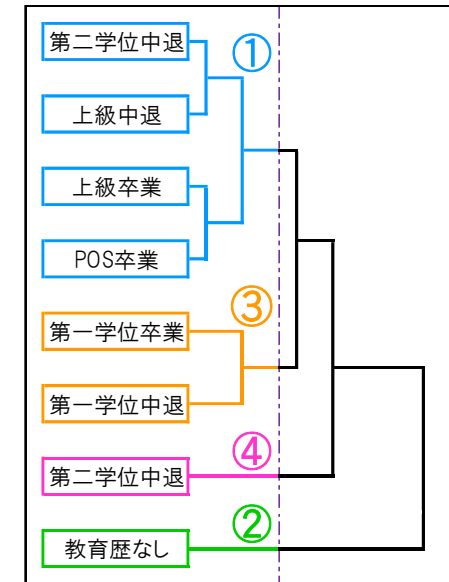


図8.デンドログラム(女性)

性別で学歴毎に分類するために「Visual Mining Studio」を用いて階層クラスタ分析を行った。距離計算法は最も一般的なユークリッド距離を採用した。分析方法は分離感度が良いとされるWard法を採用した。デンドログラムから男女共に4クラスターに分けて分析を進めていくのが有効だと判断した。

男性は第一学位卒業、第二学位卒業、女性は第二学位中退、教育歴なしが独立していることが判明した。

9. クラスター分析(男性)

表8. クラスター毎の各平均(男性)

クラスター	データ数	平均								
		年齢	主収入	不動産価値	商品価格	頭金	借入金	分割数	利子	未払い者率
第1クラスター	588	37.6歳	R\$ 2,212.72	R\$ 135,891.27	R\$ 7,455.10	R\$ 751.75	R\$ 7,209.84	44.2回	8.1%	17.9%
第2クラスター	1117	35.9歳	R\$ 1,530.71	R\$ 97,683.85	R\$ 6,793.80	R\$ 695.42	R\$ 6,606.20	45.3回	8.4%	25.8%
第3クラスター	337	38.8歳	R\$ 1,218.96	R\$ 510,691.78	R\$ 6,402.11	R\$ 701.77	R\$ 6,187.88	46.3回	9.4%	23.4%
第4クラスター	5662	35.6歳	R\$ 1,549.24	R\$ 194,819.35	R\$ 6,802.29	R\$ 682.17	R\$ 6,627.42	45.8回	9.0%	26.8%
全体	7704	35.9歳	R\$ 1,582.74	R\$ 190,055.46	R\$ 6,833.38	R\$ 690.26	R\$ 6,649.57	45.6回	8.9%	25.9%

表9. クラスター毎の在住地域率(男性)

クラスター	データ数	地域				
		北部	北東部	中西部	南東部	南部
第1クラスター	588	13.6%	23.5%	11.1%	42.7%	9.2%
第2クラスター	6263	25.2%	22.8%	12.5%	29.0%	10.5%
第3クラスター	516	19.6%	34.7%	15.7%	21.1%	8.9%
第4クラスター	337	18.6%	30.4%	14.1%	25.6%	11.3%
全体	7704	19.2%	29.0%	13.7%	27.2%	10.9%

平均年齢、平均頭金、平均分割回数は全クラスターで類似している。第1クラスター以外は全体の平均主収入を下回っていることから貧困層が集まっていると言える。しかし、第3クラスターは平均主収入は低いですが平均不動産価値においては最も高い。このことから未払い者率が高い第2クラスター、第4クラスターが真の貧困層だと言える。在住地域については平均主収入が最も高い第1クラスターに南東部が集中しているおり地域ごとの割合は偏っている。一方で他クラスターは2割以上の大きな偏りがないことが読み取れる。

男性の未払い者率においては主収入、不動産価値が共に低い場合に起こる確率が高いと推測される。

9. クラスタ分析(女性)

表10. クラスタ毎の各平均(女性)

クラスター	データ数	平均								
		年齢	主収入	不動産価値	商品価格	頭金	借入金	分割数	利子	未払い者率
第1クラスター	3149	36.0歳	R\$ 1,251.80	R\$ 246,344.91	R\$ 6,126.39	R\$ 522.68	R\$ 6,088.62	45.8回	9.8%	19.9%
第2クラスター	81	35.9歳	R\$ 1,162.01	R\$ 103,762.10	R\$ 6,466.01	R\$ 707.57	R\$ 6,268.78	43.8回	9.4%	8.6%
第3クラスター	129	42.3歳	R\$ 980.26	R\$ 78,938.76	R\$ 5,939.38	R\$ 572.73	R\$ 5,853.08	45.3回	10.8%	23.3%
第4クラスター	140	39.3歳	R\$ 1,131.41	R\$ 94,894.50	R\$ 5,971.04	R\$ 580.21	R\$ 5,887.33	44.3回	11.1%	17.9%
全体	3499	36.3歳	R\$ 1,234.90	R\$ 230,812.57	R\$ 6,121.14	R\$ 531.11	R\$ 6,076.05	45.7回	9.9%	19.7%

表11. クラスタ毎の在住地域率(女性)

クラスター	データ数	地域				
		北部	北東部	中西部	南東部	南部
第1クラスター	3149	19.8%	30.9%	15.1%	23.0%	11.3%
第2クラスター	81	23.5%	14.8%	17.3%	34.6%	9.9%
第3クラスター	129	28.7%	34.9%	12.4%	22.5%	1.6%
第4クラスター	140	19.3%	26.4%	13.6%	22.9%	17.9%
全体	3499	20.1%	30.5%	15.0%	23.2%	11.2%

男性と同様に平均年齢、平均頭金、平均分割回数は全クラスターで類似しているが男性よりは全体的に値が低い。第1クラスターは、平均主収入、平均不動産価値が共に高いにもかかわらず、未払い者率が19.9%と高い事から、**女性は男性とは異なり平均主収入、平均不動産価値が共に高くとも未払いになりやすいと思われる。**また、女性の場合は南東部の在住率が高い第2クラスターよりも北東部に在住率が集中している第1クラスターの方が平均主収入が高い。しかし、未払い率においては、第2クラスターの方が低い事から**女性は経済状況よりも在住地域の環境によって未払い率が変化すると推測される。**

10. まとめ・今後の課題

まとめ

今回の分析で同じ国であっても性別によって債務者のデータには、**異なった傾向・特徴がある**ことが判明した。厚木の研究^[5]で判明した「収入額が低いからローン未払いになるとは言い切れず、学歴に関しては平均主収入額が多い第二学位卒業の方が平均主収入額の低い上級学位中退者よりも未払い者になりやすい」という点に関しては、クラスター分析より、女性の第二学位卒業者と上級学位中退者は同じクラスターだったが、男性においては、別々のクラスターに所属しており、第二学位卒業の所属クラスターが未払い率が高かったことより、全体を対象にして行った分析で、データ数の多い男性のデータが大きく影響したと考えられる。

今後の課題

今回の研究では男性と女性ではデータによる特徴・傾向が異なっていたことは判明したが、実際に未払いにどれほどの影響を与えているかまでは確認することが出来なかった。今後は性別を考慮し、ジェンダー差別化された予測モデルの作成を行い、未払いにおける性別の影響を鑑みたいと思う。

参考文献

- [1] World Bank Group – International Development, Poverty, & Sustainability
<https://www.worldbank.org/>
(最終閲覧日:2020/11/6)
- [2] Investing.com - ブラジル政策金利発表
<https://jp.investing.com/economic-calendar/brazilian-interest-rate-decision-415>
(最終閲覧日:2020/11/6)
- [3] G1 – Frota de motos cresce 170% em 10 anos no Brasil – notícias em Motos
<http://g1.globo.com/carros/motos/noticia/2015/05/frota-de-motos-cresce-170-em-10-anos-no-brasil.html>
(最終閲覧日:2020/11/25)
- [4] ブラジルのバイク売上ランキング ブラジル／サンパウロ特派員ブログ
https://tokuhain.arukikata.co.jp/sao_paulo/2009/10/ranking_modelos_no_brasil.html
(最終閲覧日:2020/11/25)
- [5] 厚木麻里,朝日弓未:ローン未払い者の経歴の特徴についての分析,東海大学紀要情報通信学部
Vol.11, No.2, 2018, pp.27-33
- [6] Visual Mining Studio 製品パンフレット
<https://www.msi.co.jp/vmstudio/materials/VMSpamphlet.pdf>
(最終閲覧日:2020/11/25)
- [7] データマイニングツール Visual Mining Studio | 商品情報 | ユサコオンラインショップ
https://www2.usaco.co.jp/shop/pages/product_visualminingstudio.aspx
(最終閲覧日:2020/11/25)

The image features a central white hexagonal area with a thin yellow border. This hexagon is set against a background of four overlapping triangles that meet at the center. The top-left and bottom-right triangles are dark blue, while the top-right and bottom-left triangles are a lighter teal color. The word "Appendix" is centered within the white hexagon in a black, serif font.

Appendix

App. 欠損値・異常値除去

- 欠損値について

VMSのクリーニング機能を用いて欠損値の補完を行った。連続変数については- INF に加えて生の無限大(INF)、非数(NaN)を欠損値として検出し、欠損値に正常データから判明している結果への影響が小さいことから、列の平均値を代入することで補完した。また、借入金等のローンを組むうえで必要不可欠なデータが欠損している連続変数に関してはデータに問題があるとみなし該当行を削除した。質的変数については欠損値がある行を除外した。

- 異常値について

厚木の研究^[5]をもとに正常値の基準に該当しないデータを削除した。

App. VMSについて

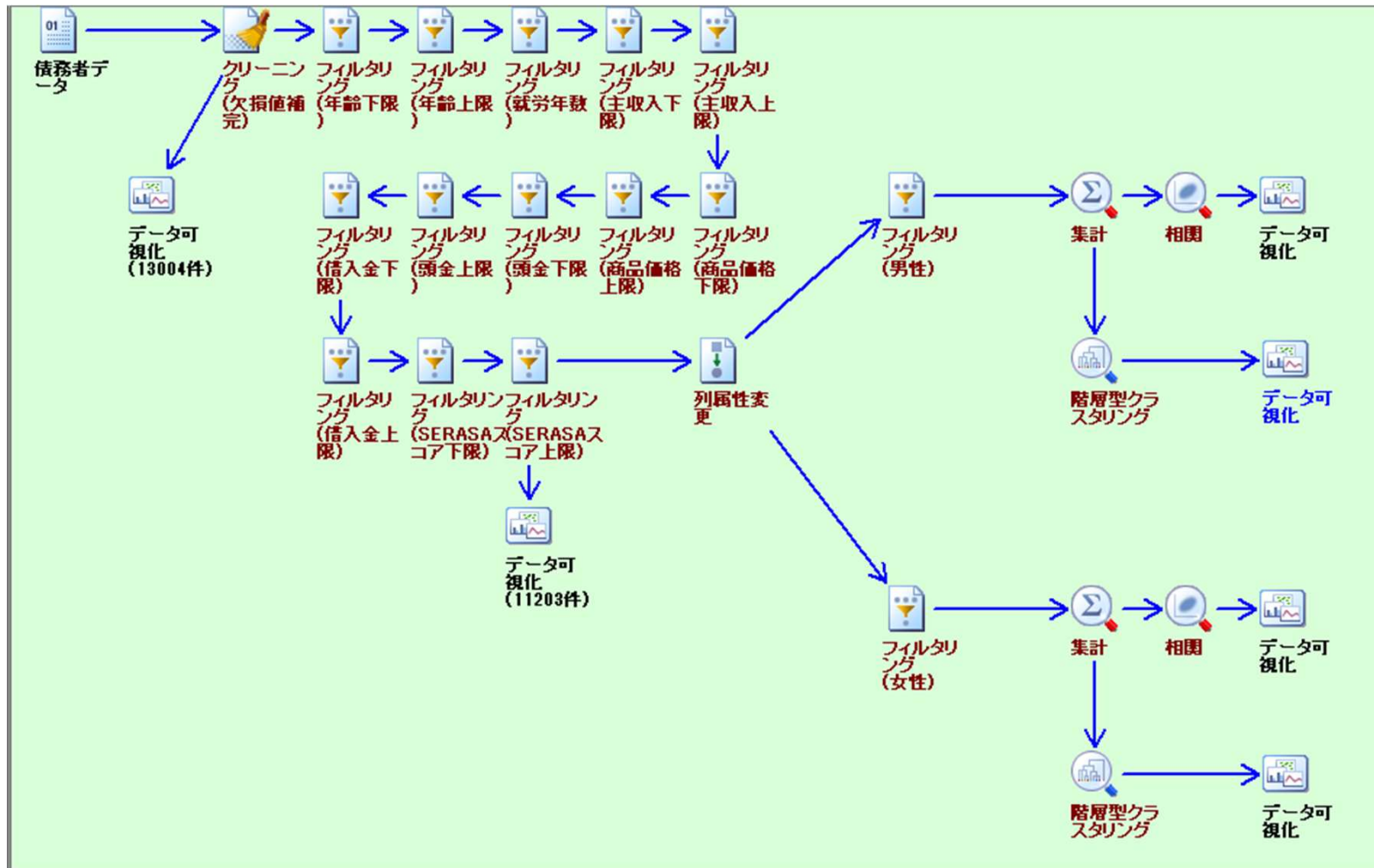


図9.VMS用いた作業全体図

App. VMS分析設定



図10.VMS用いた相関分析の設定

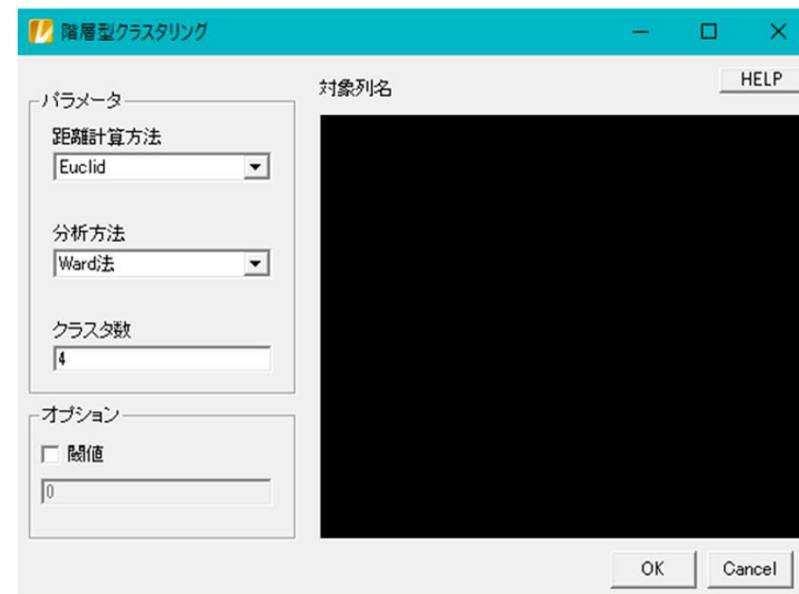


図11.VMS用いた階層クラスタ分析の設定

相関分析、階層クラスタ分析共に男女で同一設定で実行。