



Copyright © SAS Institute Inc. All rights reserved.



◇相関関係 散布図 積率相関係数 偏相関係数

◇回帰分析 回帰式 偏回帰係数、t検定、決定係数







正の相関



無相関

相関関係には正の相関、負の相関、無相関。 点の集中度が関係の強さを測定する手がかり。



支店	広告費	売上高
北海道	92	44
東北	93	102
関東	332	288
北陸	78	54
中部	181	118
近畿	108	138
中国	113	138
四国	72	86
九州	243	152
沖縄	13	22

散布図の作成法 (SAS Studio)

1.SAS Studio にログインする。 2.SAS®Studio をクリックする。

	N SAS® OnDema Si	^{otices} nd for Academics gn In			
ප	SAS Profile email add	ress or user ID			
۳ ۲	Password				
	Accept the terms of use and condition	the license and the terms o	of		
		Applications	Enrollments	Courses	
		SAS® Studio Write and run SA Actions: <u>Clear r</u>	S code with a Web-k ny saved tabs.	based SAS development en	vironment.

3. 「SASプログラマ」をクリックし、「ビジュアルプログラマ」を選択する。



新規フォルダを作成する。 「ファイル(ホーム)」を右クリックー「新規作成」ー「フォルダ」をクリック、 「新規フォルダ名(データ分析の基礎3)」を入力し、「保存」をクリックする。

SAS® Studio						
・サーバーファイル ・サーバーファイル	ルとフォルダ 下 目 い spse1 ジショートカット	ti プロセスフロ ▶ 実行 │ 日 フロー + ▼ │ ケ	□-1 ×	⁵ の生成 X プロパティ チ 睯 🕯	■ 【 すべて選択	
■ sasu: ▷ ■ デー・ ▷ ■ デー・	 新規作成 デキスト形式でファ データのインボート マイタスクに追加 マイスニペットに追 作成 名前の変更 	・ イルを表示する 加	SAS プログラ データのイン クエリ プロセスフロ XML フォルダ フォルダショ	5ム (F4) ポート コー		
	削除 移動 コピー ファイルのアップロ ファイルのダウンロ プロパティ	- K - K		新規フォ. 場所: /hc 名前: デ	ルダ ome/u61364448 ニータ分析の基礎3	保存 キャンセ



×

5. ファイルをアップロードする。 「データ分析の基礎3」を右クリック、「ファイルのアップロード」をクリックし、 「ファイルの選択」から、ファイルを選択し「アップロード」をクリックする。

SAS [®] Studio		
 ・サーバーファイルとフォルダ ビ・ 命 点 字 目 い ・ ① odaws01-apse1 ● フォルダショートカット ▲ ファイル(ホーム) ● sasuser.v94 ▶ ● データ分析の基礎1 ▶ ● データ分析の基礎2 ▶ ● データ分析の基礎3 開く 	 ・実行 □ Q 図 □−Fの生成 ; フロー 結果 プロパティ + ▼ □ 色 ▼ 面 □ € № 	ファイルのアップロード ファイルのアップロード先: /home/u61364448/データ分析の基礎3 ファイルの選択
新規作成 テキスト形式で データのインボ マイタスクに追 マイスニペット 作成 名前の変更 削除 移動 コピー ファイルのアッ ファイルのダウ プロパティ	・ Eファイルを表示する Kート 自加 ・に追加 ・ プロード… フンロード	選択済みファイル: 1 XLSX EXCELデータ.xlsx 23.1 kb アップロード キャンセル

8

6.「データ分析の基礎3」フォルダを開きファイル(EXCELデータ)を 右側のプロセスフロー画面にドラッグし、右クリックして「開く」を選択する。



7.「ワークシート名(Data1)」を入力し、実行ボタンをクリックする。





8. 「出力データ」をクリックし、「テーブル(データセット名)」、データの内容を 確認する。

GAS® Studio					
AS [®] Studio ・サーバーファイルとフォルダ ● ・ 曲 ▲ 〒 目 い ・ 望 odaws01-apse1 ■ フォルダショートカット ・ ロ ファイル (ホーム) ■ sasuser.v94 ▶ ■ データ分析の基礎1 ▶ ■ データ分析の基礎2 ▲ 〒 データ分析の基礎3 ■ EXCELデータ.xlsx	このでは、「「「「」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」 この場所: /hom ファイル名: EXC ファイル名: IM うイブラリ: W 変更 ・オブション ファイルの種類: デフォルト(ファイ) ダ SAS 変数名の生成	XCELデータ.xlsx*のインボート ・ え Q しご SELデータ.xlsx me/u61364448/データ分析の基 ASApp IPORT /ORK 以拡張子に基づく) 炎	₩₩#3		
	 ✓ SAS 変数名の生成 コード ログ テーブル: WORK.IM 列 ダ すべて選択 	結果 PORT ▼ ビユー: 列石	 ● ■ \$1 ③ 合計行 	 ■ マフィルタ: 数: 10 合計列数: 2 広告費 ○○ 	なし) 売上高
	☑ 囫 広告費		1	92	44
	☑ 囫 売上高		2	93	102
マクトコーティリティ			3	332	288
X/CI)1))1		店	4	78	54
	プロバティ				
スニペット	フロバティ	<u>]</u> te	5	181	118

- ▶ タスク
- ▶ スニペッ

SAS[®] Stu

9. 「プロセスフロー」をクリックしてフロー画面に戻り、「タスクとユーティリティ」を 開く。

SAS [®] Studio	
▼ サーバーファイルとフォルダ	t *プロセスフロ−1 ×
は→ 竜 志 平 国 の	▶ 実行 🔒 😡 🕮 コードの生成 🔀
▲ 📳 odaws01-apse1	フロー 結果 プロパティ
🔁 フォルダショートカット	+ - 늘 色 - 盦 晶 ♂ № 喩 間 すべて選択
▲ □ ファイル (ホーム)	
 ■ データ分析の基礎1 	
▶ - データ分析の基礎2	
▲ ■ データ分析の基礎3	ス ^x EXCELデータ. ×lisx ^w のインボ ート
R EXCELデータ.xisx	Ø
▶ タスクとユーティリティ	
▶ スニペット	
> ライブラリ	
F 54559	

10. 「タスクとユーティリティ」→「タスク」→「グラフ」の「散布図」をフロー画面に ドラッグし、「EXCELデータ・・」と結合、右クリック-「開く」をクリックする。





11. 「散布図」を右クリック、開き、「データ」、「X軸(広告費)」、 「Y軸(売上高)」をセットする。



t *プロセスフロー1 ×	
<u>プロセスフロー 1</u> 〉 散布図	
設定 コード/結果 分割 📌 😡 🔀	
データ 表示 情報 ノード	
▼ データ	
WORK.IMPORT	▼
₹フィルタ:(なし)	
*X 軸: (1 項目)	竜 +
🐵 広告費	
*Y 軸: (1 項目)	命 +
❷ 売上高	
ガループ・クロー	⇒ ⊥
クルーノ: (7 項目)	ш т
∠ ♦ 2.3	
▶追加役割	

¹⁴ Sas

12. 「表示」をクリックし、「グラフサイズ」を幅「4.8」に変更し、出力する グラフを正方形にする。

t *プロセスフロー1 ×
<u>プロセスフロー1</u> 〉 散布図
設定 コード/結果 分割 📌 🐼 🕵
データ 表示 情報 ノード
▼近似曲線
Loess
□ 罰則付きB-スプライン
・マーカー
▶ × 華由
▶ 乂 車由
▶タイトルとフットノート
 ✓ グラフサイズ
単位: インチ (デフォルト) ▼
*幅: 4.8 -
*高さ: 4.8 -



13. 実行ボタンをクリックすると散布図が表示される。

* プロセスフロー1 ×	
<u>プロセスフロー1</u> 〉 散布図	
設定 コード/結果 分割	
データ 表示	
▼データ	
WORK.IMPORT	
₹フィルタ:(なし)	
▼役割	- 1
*X 軸: (1 項目)	益 +
🐵 広告費	
*Y 軸: (1 項目)	益 +
🐵 売上高	
グループ: (1項目)	曲 +
💊 列	
▶追加役割	

結果が表示される。





積率相関係数を求める (SAS Studio)

1. 「タスクとユーティリティ」→「タスク」→「統計量」の「相関分析」を フロー画面にドラッグし、「EXCELデータ・・」と結合、右クリック-「開く」をクリックする。





2. 「相関分析」を右クリック、開き、「データ」、「分析変数(広告費、 売上高)」をセットする。

SAS [®] Studio	
▶ サーバーファイルとフォルダ	t *プロセスフロー1 ×
▼ タスクとユーティリティ	<u>プロセスフロー1</u> > 相関分析
稽 ∞ 亩 民 目 55	設定 コード/結果 分割 🗶 😡 🎇
🗖 マイタスク	データ オプション 出力 情報 ノード
▲ 📫 タスク	 データ
▶ 🛄 データ	WORK.IMPORT
▶ 🛄 グラフ	▼フィルタ:(なし)
▶ 📫 マップ	▼ 沿割
▲ 🛄 統計量	
👪 データ探索	
醫 要約統計量	
₩ 分布分析	❷ 売上高
Ⅲ 一元度数表	
╱ 相関分析	
国 分割表分析	
H t 検定	
▶ 🛄 線形モデル	

19

Sas

3. 実行ボタンをクリックする。

ta *プロセスフロ−1 ×	
<u>プロセスフロー1</u> > 相関分析	
設定 コード/結果 分割 ★ ■ ★ ■	
データ オプション	•
▼データ	
WORK.IMPORT	
₹フィルタ: (なし)	
▼役割	- 1
分析変数:	+
🔞 広告費	
🛯 売上高	
相関変数: ▲ 💼 •	+



結果が表示される。

Pearsonの相関係数, N = 10			
	広告費 売上高		
広告費 広告費	1.00000	0.90236	
売上高 売上高	0.90236	1.00000	

積率相関係数(r)=0.902

積率相関係数(r) -1 ≤ r ≤ 1



積率相関係数(r)は相関関係の強さ

積率相関係数(r)の解釈

- r = 0 ⇒ 説明力は0%
- r = 1 ⇒ 説明力は100%
- r = 0.32 ⇒ 説明力は?%

0.32×0.32=0.1024 説明力は、10.24%

◇支店別広告費、売上高、人口

支店	広告費	売上高	人口
北海道	92	44	5506
東北	93	102	9335
関東	332	288	42604
北陸	78	54	5443
中部	181	118	18127
近畿	108	138	12912
中国	113	138	15554
四国	72	86	3976
九州	243	152	13204
沖縄	13	22	1393

S.Sas.

3 変数: 広告費 売上高 人口			
Pearson の相関係数. N=10			
	広告費	売上高	ᇇ
広告費 広告費	1.00000	0.90236	0.89471
売上高 売上高	0.90236	1.00000	0.95093
쎪	0.89471	0.95093	1.00000

広告費と売上高r = 0.902人口と広告費r = 0.895人口と売上高r = 0.951

広告宣伝費は売上高に貢献?



人口の影響を除いたときの広告費と売上高の 積率相関係数 → 偏相関係数

偏相関係数(SAS Studio)

1.「ワークシート名(Data3)」を入力し、実行ボタンをクリックする。



2. 「相関分析」を右クリック-「開く」をクリックする。





3. 「相関分析」を右クリック、開き、「データ」、「分析変数(広告費、売上高)」、 「部分変数(人口)」をセットする。





4. 実行ボタンをクリックする。



30

Sas

結果が表示される。

1 Partial 変数:	人口
2 変数:	広告費 売上高

Pearson の偏相関係数, N = 10		
	広告費	売上高
広告費 広告費	1.00000	0.37306
売上高 売上高	0.37306	1.00000

偏相関係数=0.373



都道府県別コンビニ件数、人口、甲子園の勝率

1		コンビニ数	人口	甲子園勝率
2	東京都	6847	13513.7	0.543
3	神奈川県	3431	9127.3	0.623
4	大阪府	3654	8838.9	0.616
5	愛知県	3576	7484.1	0.602
6	埼玉県	2597	7261.3	0.538
7	千葉県	2421	6224.0	0.554
8	兵庫県	1852	5537.0	0.563
9	北海道	1807	5383.6	0.335
10	福岡県	2040	5102.9	0.475
11	静岡県	1685	3701.2	0.493

Sas

Pearson の相関係数, N = 47			
	コンビニ数	人口	甲子園勝率
コンビニ数 コンビニ数	1.00000	0.98272	0.40987
人口 人口	0.98272	1.00000	0.43197
甲子園勝率 甲子園勝率	0.40987	0.43197	1.00000

・人口とコンビニ数:0.983

人口の多い都道府県はコンビニ件数が多い。

・人口と甲子園勝率: 0.432

人口が多いと高校生の数も多く、結果として野球のレベルも上がる。 ・コンビニ数と甲子園勝率:0.410?

人口の影響を除いた積率相関係数(偏相関係数)

偏相関係数を求める (SAS Studio)

1.「ワークシート名(Data5)」を入力し、実行ボタンをクリックする。



2. 「相関分析」を右クリック-「開く」をクリックする。



3. 「相関分析」を右クリック、開き、「データ」、「分析変数(コンビニ数、 甲子園勝率)」、「部分変数(人口)」をセットする。

SAS [®] Studio	
▶ サーバーファイルとフォルダ	t *Jロセスフロー1 ×
▼ タスクとユーティリティ	<u>プロセスフロー1</u> > 相関分析
粂→ 侖 民 目 \$5	設定 コード/結果 分割 大 😡 💱
	データ オプション 出力 情報 ノード
▲ ■ タスク	- 役割
▶ 🖪 データ	分析変数:
▶ 🛄 グラフ	🐵 コンビニ数
▶ 🖪 マップ	四甲子園勝率
▲ 🛄 統計量	
🙀 データ探索	
器 要約統計量	
12 分布分析	
Ⅲ 一元度数表	
∠ 相関分析	
3日分割表分析	
<u>H</u> t検定	
▶ ■ 線形モデル	
▶ ■ データマイニング	部分変数:
▶ ■8 計量経済	
 ▶ ■ 予測	
▷ 🗞 ユーティリティ	

Sas
結果が表示される。

1 Partial 変数:	人口
2 変数:	コンビニ数 甲子園勝率

Pearsonの偏相関係数, N = 47					
	コンビニ数	甲子園勝率			
コンビニ数 コンビニ数	1.00000	-0.08766			
甲子園勝率 甲子園勝率	-0.08766	1.00000			

偏相関係数 = -0.088



例)売上高と売上高に影響を与える要因との関係

売上高←広告宣伝費、人口、セールスマン数、・・・

従属変数←説明変数(独立変数) 説明変数が1つ:単回帰 説明変数が2つ以上:重回帰

目的1:回帰式を求め、予測する。

回帰式(y=a+bx+…)を求める。 a,b,・・・ 🔶 偏回帰係数

◇駅前コンビニの売上高と乗降客数

売上高←乗降客数

	売上高 (百万円/月)	乗降客数 (百人/日)
1	130	93
2	290	234
3	235	250
4	260	260
5	140	119
6	173	180
7	135	151
8	190	192
9	220	273
10	181	185

S.Sas.

2 変数: 売上高 乗降容数						
Pearson	の相関係数	枚, N = 10				
売上高 乗降客数						
売上高 売上高	1.00000	0.86747				
乗降客数 乗降客数	0.86747	1.00000				

積率相関係数=0.867



回帰式(売上高 = a + b×乗降客数)を求めて予測する。

回帰直線の求め方(最小二乗法)



各データと回帰直線との垂直距離の2乗和を最小にする



回帰分析 (SAS Studio)

1. 「ワークシート名(Data10)」を入力し、実行ボタンをクリックする。



2. 「タスクとユーティリティ」→「タスク」→「線形モデル」の「線形回帰分析」 をフロー画面にドラッグし、「EXCELデータ・・」と結合、右クリック-「開く」 をクリックする。



3. 「線形回帰分析」を右クリック、開き、「データ」、「従属変数(売上高)」、 「連続変数(乗降客数)」をセットする。



4. 「モデル」をクリックし、「モデル効果」の「編集」をクリックする。





5.「変数(乗降客数)」→「単一効果(追加)」をセットする。 「切片」にチェックがつき、「乗降客数」が表示されていることを確認し、 OKボタンをクリックする。





7. 実行ボタンをクリックする。



▼ モデル効果

▼ モデル効果

	民編集
切片	
乗降客数	



結果が表示される。

Root MSE	28.91642	R2 乗	0.7525
従属変数の平均	195.40000	調整済み R2 乗	0.7216
変動係数	14.79858		

バラメータの 推定						
変数 ラベル 自由度 パラメータ 推定値 標準誤差 t 値 Pr > t						
Intercept	Intercept	1	43.99116	32.03230	1.37	0.2069
乗降客数	乗降客数	1	0.78167	0.15849	4.93	0.0011

売上高=43.99+0.782×乗降客数

*乗降客数=180のときの売上高の予測 売上高=43.99+0.782×180=184.75

回帰モデルのチェック

◇自由度調整済み決定係数 R² = 0.7216 約72.16%説明できる。



乗降客数:P值(P值) = 0.0011



説明変数に「取扱品目数」を追加

	売上高	乗降客数	取扱品目数
	(百万円/月)	(百人/日)	(品)
1	130	93	150
2	290	234	311
3	235	250	182
4	260	260	245
5	140	119	149
6	173	180	160
7	135	151	98
8	190	192	180
9	220	273	113
10	181	185	105

1.「ワークシート名(Data11)」を入力し、実行ボタンをクリックする。



2. 「線形回帰分析」を右クリック-「開く」をクリックする。



3. 「線形回帰分析」を右クリック、開き、「データ」、「従属変数(売上高)」、 「連続変数(乗降客数、取扱品目数)」をセットする。



Sas

4. 「モデル」をクリックし、「モデル効果」の「編集」をクリックする。

€ *プロ	セスフロー	1 ×		
プロセ	<u> </u>	線形回帰	分析	
設定	コード/結果	分割	* 🛛	53
	データ	モデル	力	<u>۲</u>
 ₹5 	デル効果			
•	モデル効果			編集
	切片			



5.「変数(乗降客数、取扱品目数)」→「単一効果(追加)」をセットし、 OKボタンをクリックする。



6. 実行ボタンをクリックする。

は*プロ	コセスフロー1	1 ×	
<u>プロセ</u>	<u> zzju – 1</u> >	線形回帰分析	
設定	コード/結果	分割	
	データ	モデル オン・	•
∙ŧ	デル効果		
-	モデル効果		
		■「編集	
	切片		
	乗降客数		
	取扱品目数		



結果が表示される。

Root MSE	10.04926	R2 乗	0.9738
従属変数の平均	195.40000	調整済み R2 乗	0.9664
変動係数	5.14292		

バラメータの 推定							
変数	ラベル	自由度	バラメータ 推定値	標準誤差	t 値	Pr > [t]	
Intercept	Intercept	1	7.15385	12.11739	0.59	0.5735	
乗降客数	乗降客数	1	0.60153	0.05985	10.05	<.0001	
取扱品目数	取扱品目数	1	0.42368	0.05505	7.70	0.0001	

◇偏回帰係数のt検定

乗降客数 P値 = <0.0001 取扱品目数 P値 = 0.0001

Sas

売上高の予測

売上高=7.15+0.602×乗降客数+0.424×品目数

*乗降客数=200 取扱品目数が180のとき売上高の予測 売上高=7.15+0.602×200+0.424×180=203.87

◇自由度調整済決定係数

R² = 0.9664 約96.64%説明できる。

説明変数に世帯数をさらに追加

	売上高 (百万円/月)	乗降客数 (千人/日)	取扱品目数 (品)	世帯数
1	130	93.0	150	143
2	290	234.0	311	284
3	235	250.0	182	320
4	260	260.0	245	302
5	140	119.0	149	182
6	173	180.0	160	225
7	135	151.0	98	190
8	190	192.0	180	242
9	220	273.0	113	320
10	181	185.0	105	235

S.sas.

「ワークシート名(Data12)」を入力し、実行ボタンをクリックする。



t *プロセスフロ−1 ×
<u>プロセスフロー 1</u> > "EXCELデータ.xlsx" のインポート
設定 コード/結果 分割 🗶 😡 🖏
オプション ノード
▼ ファイル情報
ソース ファイル
ファイル名: EXCELデータ.xlsx
ソースの場所: /home/u61364448/テータ分析の基礎3
ワークシート名: Data12
出力データ
SAS Server: SASApp
データセット名: IMPORT
変更

重回帰分析 結果

Root MSE	10,65938	R2 乗	0.9748
従属変数の平均	195.40000	調整済み R2 乗	0.9622
変動係数	5.45516		

バラメータの推定								
変数	ラベル	自由度	バラメータ 推定値	標準誤差	t値	Pr > [t]		
Intercept	Intercept	1	-1.82664	23.00272	-0.08	0.9393		
乗降客数	乗降客数	1	0.42311	0.38429	1.10	0.3131		
取扱品目数	取扱品目数	1	0.421 44	0.05858	7.19	0.0004		
世带数	世带数	1	0.17978	0.38190	0.47	0.6544		

◇偏回帰係数のt検定

乗降客数P値 = 0.3131取扱品目数P値 = 0.0004世帯数P値 = 0.6544



売上高を乗降客数、取扱品目数、世帯数の3つの 説明変数で説明する重回帰モデルは、不成立!



①乗降客数&取扱品目数 ②乗降客数&世帯数 ③取扱品目数&世帯数

①説明変数:乗降客数&取扱品目数

Root MSE	10.04926	R2 乗	0.9738
従属変数の平均	195.40000	調整済み R2 乗	0.9664
変動係数	5.14292		

バラメータの推定								
変数	ラベル	自由度	バラメータ 推定値	標準誤差	t值	Pr > [t]		
Intercept	Intercept	1	7.15385	12.11739	0.59	0.5735		
乘降客数	乗降客数	1	0.601 53	0.05985	10.05	<.0001		
取扱品目数	取扱品目数	1	0.42368	0.05505	7.70	0.0001		

乗降客数P値 = <0.0001</th>取扱品目数P値 = 0.0001

Sas

②説明変数:乗降客数&世帯数

Root MSE	30.61 661	R2 乗	0.7572
従属変数の平均	195.40000	調整済み R2 乗	0.6879
変動係数	15,66869		

バラメータの推定								
変数	ラベル	自由度	バラメータ 推定値	標準誤差	t値	Pr > [t]		
Intercept	Intercept	1	23,40110	65.29769	0.36	0.7306		
乗降客数	乗降容数	1	0.37915	1.10365	0.34	0.7413		
世带数	世带数	1	0.40343	1.09329	0.37	0.7230		

乗降客数P値 = 0.7413世帯数P値 = 0.7230

③説明変数:取扱品目数&世帯数

Root MSE	10.81976	R2 乗	0.9697
従属変数の平均	195,40000	調整済み R2 乗	0.9610
変動係数	5,53723		

バラメータの推定								
変数	ラベル	自由度	バラメータ 推定値	標準誤差	t値	Pr > [t]		
Intercept	Intercept	1	-21.00871	15.24568	-1.38	0.21 06		
世带数	世带数	1	0.59449	0.06403	9.28	<.0001		
取扱品目数	取扱品目数	1	0.42041	0.05946	7.07	0.0002		

世帯数P値 = <0.0001</th>取扱品目数P値 = 0.0002

説明変数	t 検定
乗降客数、取扱品目数、世帯数	×
①乗降客数、取扱品目数	0
②乗降客数、世帯数	×
③取扱品目数、世帯数	0



Pearson の相関係数、N=10								
	売上高	乘降客数	取扱品目数	世帶数				
売上高 売上高	1.00000	0.86747	0.77224	0.86784				
乗降客数 乗降客数	0.86747	1.00000	0.391.08	0.98837				
取扱品目数 取扱品目数	0.77224	0.391.08	1.00000	0.39792				
世带数 世带数	0.86784	0.98837	0.39792	1.00000				





いずれも高い値⇒売上高を説明する説明変数として妥当



乗降客数と世帯数の値0.988は高い値

説明変数相互の積率相関係数は低い方が良い! 説明変数⇒独立変数



売上高	乗降客数	間口の広さ
130	93	150
290	234	148
235	250	182
260	260	245
140	119	149
173	180	160
135	151	135
190	192	180
220	273	113
181	185	105

データ10組のとき

「ワークシート名(Data15)」を入力し、実行ボタンをクリックする。

SAS [®] Studio	
 サーバーファイルとフォルダ ● 命 ▲ ● 目 ● ● のdaws01-apse1 ● フォルダショートカット ● ファイル(ホーム) ● sasuser.v94 ● データ分析の基礎1 ● データ分析の基礎2 ● データ分析の基礎3 ● データ分析の基礎3 	 ¹ 「フーセスフロー1 × ¹ 「ロセスフロー1 × 「EXCELデータ.xlsx" のインボート ¹ コード/結果 分割 ¹ ス ¹ ス



Root MSE	29,46589	R2 乗	0.7751
従属変数の平均	195,40000	調整済み R2 乗	0.71.09
変動係数	15.07978		

パラメータの推定								
変数	ラベル	自由度	バラメータ 推定値	標準誤差	t值	Pr > [t]		
Intercept	Intercept	1	17.72278	45.22169	0.39	0.7068		
乘降客数	乗降客数	1	0.74276	0.16802	4.42	0.0031		
間口の広さ	間口の広さ	1	0.21573	0.25704	0.84	0.4290		

間口の広さ: P値 = 0.4290

データ50組のとき

「ワークシート名(Data16)」を入力し、実行ボタンをクリックする。


Root MSE	25,42756	R2 乗	0.7751
従属変数の平均	195,40000	調整済み R2 乗	0.7656
変動係数	13.01308		

バラメータの推定									
変数	ラベル	自由度	バラメータ 推定値	標準誤差	t値	Pr > [t]			
Intercept	Intercept	1	17.72278	17.45207	1.02	0.3151			
乗降客数	乗降客数	1	0.74276	0.06484	11.45	<.0001			
間口の広さ	間口の広さ	1	0.21573	0.09920	2.17	0.0347			

間口の広さ: P値 = 0.0347

t 検定のP値はサンプルサイズの影響を受ける。

△洪口中Ⅲ本	満足度	機能	デザイン	男性 ダミー	満足度	機能	デザイン	男性 ダミー
◇ 洞 に 反 詞 且	5	5	4	1	5	2	5	0
	4	4	2	1	5	2	4	0
田州10夕	4	4	3	1	5	4	4	0
为住19石	4	3	5	1	5	3	3	0
女性21名	4	3	2	1	4	5	3	0
	3	4	3	1	4	2	5	0
	3	3	5	1	4	3	4	0
く従属変数>	3	3	2	1	4	5	3	0
、	3	3	4	1	4	3	5	0
测足区	3	3	3	1	3	2	4	0
	3	3	3	1	3	1	4	0
<説明変数>	3	3	5	1	3	5	2	0
・機能	3	3	3	1	3	5	3	0
・デザイン	2	2	4	1	3	3	2	0
• 牛 別	2	2	3	1	2	3	3	0
(里性1 女性0)	2	2	3	1	2	2	3	0
	2	2	3	1	2	3	2	0
	1	1	3	1	1	1	3	0
	1	1	4	1	1	3	2	0
	5	5	5	0	1	4	1	0

「ワークシート名(Data20)」を入力し、実行ボタンをクリックする。

SAS[®] Studio

- ▼ サーバーファイルとフォルダ
 - は~ 竜 出 平 国 い
 - ▲ 望 odaws01-apse1
 フォルダショートカット
 ▲ マアイル(ホーム)
 sasuser.v94
 ▶ データ分析の基礎1
 ▶ データ分析の基礎2
 ▲ データ分析の基礎3
 データ分析の基礎3



変更

重回帰分析の結果

Root MSE	0.91 407	R2 棄	0.4813
従属変数の平均	3.10000	調整済み R2 乗	0.4533
変動係数	29.48628		

バラメータの推定								
変数	ラベル	自由度	バラメータ 推定値	標準誤差	t値	Pr > [t]		
Intercept	Intercept	1	-0.74452	0.67665	-1.10	0.2783		
機能	機能	1	0.58538	0.1 2661	4.62	<.0001		
デザイン	デザイン	1	0.62340	0.1 41 72	4.40	<.0001		

- ・「機能」、「デザイン」の偏回帰係数のP値は小さい。
- ・自由度調整済み決定係数 = <u>0.4533</u>

データ (男性のみ) の場合 「ワークシート名 (Data21)」を入力し、実行ボタンをクリックする。





Root MSE	0.42484	R2 乗	0.8541
従属変数の平均	2,89474	調整済み R2 乗	0.8358
変動係数	14,67616		

バラメータの推定								
変数	544	自由度	バラメータ 推定値	標準誤差	t値	Pr > [t]		
Intercept	Intercept	1	0.05533	0.47289	0.12	0.9083		
機能	機能	1	0.95629	0.09883	9.68	<.0001		
デザイン	デザイン	1	0.03608	0.1 0498	0.34	0.7356		

・男性は、機能重視
 ・自由度調整済み決定係数 = 0.8358

データ(女性のみ)の場合

「ワークシート名(Data22)」を入力し、実行ボタンをクリックする。



Sas



Root MSE	0.89023	R2 乗	0.6274
従属変数の平均	3.28571	調整済み R2 乗	0.5860
変動係数	27.09389		

バラメータの推定								
ラベル	自由度	バラメータ 推定値	標準誤差	t値	Pr > [t]			
Intercept	1	-1.23401	0.88924	-1.39	0.1822			
機能	1	0.42718	0.15695	2.72	0.01 40			
デザイン	1	0.95315	0.17871	5.33	<.0001			
	<mark>ラベル</mark> Intercept 機能 デザイン	パラ うべル 自由度 Intercept 11 税能 11	パラメータの推定 ラベル パラメータ 自由度 パラメータ 推定値 1 物能 1 0.42718 デザイン 1 0.95315	パラメータの推定 うベル ሰ合由度 パラメータ 標準課差 Intercept 1 -1.23401 0.88924 機能 1 0.42718 0.15695 デザイン 1 0.95315 0.17871	バラメータの推定 ラベル 自由度 パラメータ 推定値 標準誤差 t値 Intercept 1 -1.23401 0.88924 -1.39 機能 1 0.42718 0.15695 2.72 デザイン 1 0.95315 0.17871 5.33			

・女性は、機能、デザイン共に重視機能<デザイン
 ・自由度調整済み決定係数 = 0.5860

男性は、機能重視 女性は、機能、デザイン共に重視

性別によって異なる。







まとめ

◇相関関係

- ·散布図、積率相関係数
- ·交絡要因(疑似相関)、偏相関係数

◇重回帰分析

- ·従属変数、説明変数(独立変数)
- ・回帰モデルのチェック
 - t 検定 P値(有意確率)
 - 自由度調整済み決定係数
- ・説明変数相互の積率相関係数
- ·交互作用

アンケートのお願い・ご質問 8月31日 データ分析の基礎-3

今後の参考にさせていただくため、ぜひともアンケートにご協力を お願いします。

・無記名
 ・所要時間目安: 1~3分

<u>アンケートURL</u>

https://sas.qualtrics.com/jfe/form/SV_b7TEny3AYhyH5ie

- ・お客様講演会のアーカイブは、2022年9月5日~2023年3月31日迄 視聴できます。
- 本日の内容に関するご質問は、以下宛にご連絡ください。
 que@datascience.co.jp
- ご視聴ありがとうございました。

