商品コード: 110-2959 ISBN: 9784820729594 データアナリティクス基礎

◎本書の記述において下記のような誤りがありました。訂正してお詫び申し上げます。

【2024年5月16日現在】

更新日	刷	頁	訂正箇所	訂正前	訂正後		
↓本文	本文						
2023. 5.16	1	10	CONTENTS 1章 3	データマネジメント(層別・水準化・変換・欠測値処理)	データマネジメント(層別・水準化・変数変換・欠測値処理)		
2024. 5.16	1	21	本文上から4行目	(図表 1-3 の「レシート番号」)	削除		
2024. 5.16	1	28	本文上から7行目	疑似的な完全な	縮小した完全		
2024. 5.16	1	28	本文上から9行目	大きく縮小	縮小		
2024. 5.16	1	28	本文上から 10 行目	代入法は、欠測値を何らかの値で補い代入する方法です。リストワイズ除去のように、データが縮小する恐れがないことが利点です。この方法は、欠測部分に平均値を入力して補う方法が代表的です。また、欠測していないデータを用いて回帰分析を行い、その結果を用いて欠測値の予測値を代入する方法もあります。 このほか、欠測データの分布欠測値の予測値を複数抽出して欠測値を補う多重代入法や、欠測に対して最も得られやすかったであろう値を推定する期待値最大化法などが知られています。	代入法は、欠測値を適切な値で代入する方法です。実務上、欠 測部分に平均値を入力して補う方法が便宜的にとられることがあ ります。しかし、その変数の分散を過少に評価してしまうことや、 その他の変数との相関がある場合に相関関係が正しく評価でき なくなるので注意してください。このため、欠測値を複数の値で 補う多重代入法や期待値最大化法などのより精緻な方法が使わ れることもあります。		

2024. 5.16	1	40	図表 2-2	超計 3649 100.0% 全体の約7別の販売件数および販売会額は上位6種類で占められている 2	接計 3649 100.0%
				押 71,530 2.5% 2,748,440 95.8% 冷凍者島 40,540 1.4% 2,788,980 97.2% おかず 33,150 12% 2,822,130 98.4% 海鮮島 28,810 1.0% 2,850,940 99.4% 7ナ告 17,480 0.6% 2,858,420 100.0% 総計 2,868,420 100.0%	勝 71,530 2.5% 2.784,440 95.8% 冷凍食品 40,540 1.4% 2.789,980 97.2% おかず 33,150 1.2% 2.822,130 98.4% 無軽品 28,810 1.0% 2.850,940 99.4% ツナ缶 17,480 0.6% 2.868,420 100.0% 経計 2.868,420 100.0%
2024. 5.16	1	41	本文上から 4 行目、5 行目、7 行 目、12 行目	購入件数	購入回数
2024. 5.16	1	75	本文上から1行目	質的データのアナリティクスⅡには、連関係数とχ2検定の問題 も出題範囲に入っています。	出題範囲の質的データの分析には、連関係数とχ2検定の問題 も含まれています。
2024. 5.16	1	76	本文上から1行目	2:1 であると仮定すると、	2:1 であると仮定すると、「スポーツ中継」を見る人でかつ「商品への関心」がある人の期待度数は、
2024. 5.16	1	81	本文上から8行目	「商品への関心」の方が、「スポーツ中継」よりも「性別」	「性別」の方が、「スポーツ中継」よりも「商品への関心」
2024. 5.16	1	83	本文上から8行目	その値以上か起きる上側確率 p は	その値以上となる上側確率 p は
2024. 5.16	1	84	本文上から14行目	厳しくなりになる一方で	厳しくなる一方で
2024. 5.16	1	84	本文上から16行目	χ^2 検定検定	χ ² 検定
2024. 5.16	1	98	本文上から6行目	最も人数	最も事業所数
2024. 5.16	1	99	本文上から12行目	事業所数について変動係数を求めた。 最も変動係数が最も大 きくなったのは	5 つの地域をそれぞれに対し、事業所数の変動係数を求めた。 変動係数が最も大きくなった地域は

2024. 5.16	1	106	「 ※ Excel 2013 以前のヒストグラム の作成方法」の 4 行目	(「事業所数」)	(「個数/事業所数」)
2024. 5.16	1	106	「※Excel 2013 以前のヒストグラムの作成方法」の表	行ラベル 階級値 個数 / 事業所数	(行ラベル 階級値 個数/事業所数 0-999 500 5 1000-1999 1500 18 2000-2999 2500 9 3000-3999 3500 4 4000-4999 4500 2 5000-5999 5500 1 6000-6999 6500 2 7000-7999 7500 1 8000-8999 8500 1 9000-9999 9500 0 10000以上 10000以上 1
2024. 5.16	1	107	囲み1行目	分布の中心傾向を表す指標	分布の中心傾向や位置を表す指標
2024. 5.16	1	112	本文上から1行目	図表 4-2 の A	図表 4-2 の I
2024. 5.16	1	113	図表 4-5	図表 4-5 分散と標準偏差の計算手順	図表4-5 分散と標準偏差の計算手順 地域 図書館数 ② x/-x ③ (x/-x)² A 4 2.7 7.1 A 1 -0.3 0.1 A 1 -0.3 0.1 A 1 -0.3 0.1 A 1 -0.3 0.1 A 0 -1.3 1.8 ①平均 1.3 1.8 ①平均 1.3 */図表45では実際の値を小数第2位を四緒五人した形で示しています。 機準偏差 ⑥⑤の平方根 1.2
2024. 5.16	1	116	本文上から4行目	四分位数の注意と同様、	削除
2024. 5.16	1	118	注 5	デフォルトの設定として四分位範囲の	デフォルトの設定として第 1 四分位または第 3 四分位から四分位 範囲の

2024. 5.16	1	120	本文上から4行目	ギリシャ文字が使われます。	ギリシャ文字で表されることが多いです。
2024. 5.16	1	124	本文上から12行目	この場合、国語の変動係数は 12/86 = 0.1395(13.9%)、算数の変動係数は 20.4/78 = 0.2615(26.15%)となります。変動係数は単位が同じでも使用します。	この場合、算数の変動係数は 12/86 ≒ 0.1395 (13.95%)、国語の 変動係数は 20.4/78 ≒ 0.2615 (26.15%)となります。変動係数は 単位が同じでも使用します。
2024. 5.16	1	124	本文下から3行目	個々のデータの値からの距離を	個々のデータの値の平均値からの距離を
2023. 7.7	1	125	図表 4-14 ③の行 ID4	61.67	45.00
2023. 7.7	1	125	図表 4-14 ③の行 ID5	1.17	61.67
2024. 5.16	1	130	演習問題 4	小数第1位	小数第2位
2024. 5.16	1	130	演習問題 5	男子児童のうち、最も「ネット利用時間」の値の大きかった児童の、 「ネット利用時間」	最も「ネット利用時間」の値の大きかった男子児童の、男子児童に おける「ネット利用時間」
2024. 5.16	1	132	例題 5 2	① 人口密度(人/km²)② 平均年齢(歳)③ 核家族世帯(万世帯)④ 第3次就労者の割合(%)	① 平均年齢(歳) ② 核家族世帯(万世帯)③ 第3次就労者の割合(%)
2024. 5.16	1	133	例題 5 4	「平均年齢」が 50 歳である場合、	3 の回帰分析において「平均年齢」が 50 歳である場合、
2024. 5.16	1	133	本文下から2行目	まとめると、2 つの量的な	まとめると、相関係数は2つの量的な
2024. 5.16	1	136	本文下から2行目	標準偏差を積で	標準偏差の積で
2024. 5.16	1	137	図表 5-4	※値は小数第2位で四捨五入したもの	※値は小数第3位を四捨五入したもの
2024. 5.16	1	138	本文下から1行目	(r=0.52)とやや強い正の相関がみられ、「平均年齢(r=-0.43)」 とはやや強い負の相関があることがわかります。	とやや強い正の相関(r=0.52)がみられ、「平均年齢」とはやや強い負の相関(r=-0.43)があることがわかります。

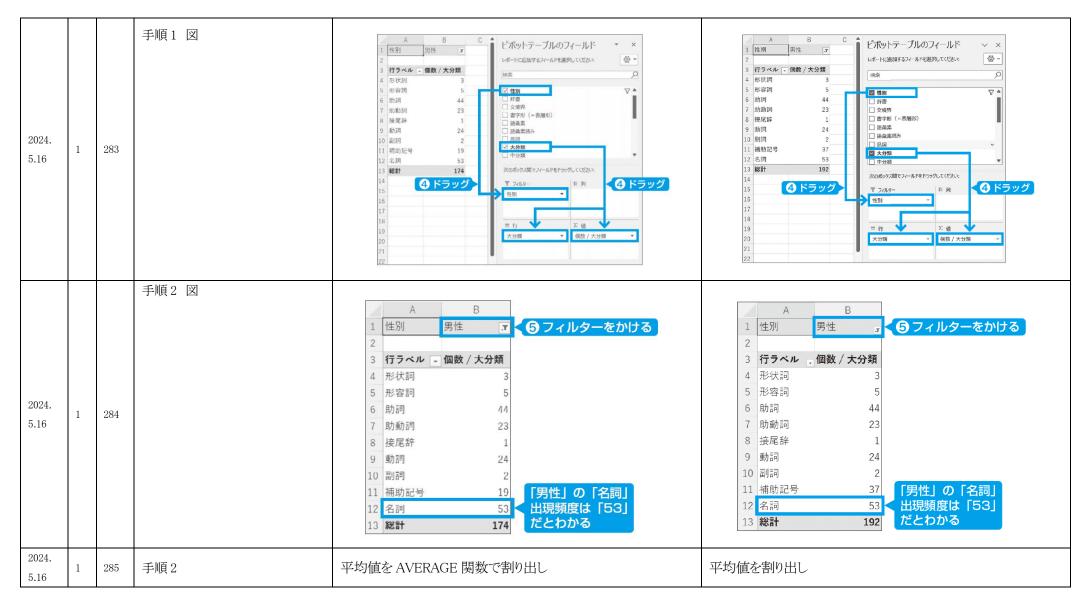
2024. 5.16	1	142	例題の解き方 3 行目	4	3
2024. 5.16	1	142	手順3下の注	リボンの[開発]タブをクリックし、[アドイン]グループにある[Excel アドイン]をクリックします。ダイアログボックスが表示されますので、[分析ツール]のチェックボックスをオンにして、[OK]ボタンをクリックします。	[ファイル]タブをクリックし、左側に表示されるファイルメニューの [オプション]をクリックします。メニューの中から[アドイン]を選択し、一覧の中から[分析ツール」を選択した状態で[設定]をクリックし、[分析ツール]にチェックを入れた状態で[OK]をクリックします。
2024. 5.16	1	146	本文上から6行目	b	а
2024. 5.16	1	146	本文下から6行目	二乗和を計算し、その合計が	二乗を計算し、その値の合計が
2024. 5.16	1	147	本文下から3行目	係数b	係数a
2024. 5.16	1	149	本文上から3行目	$R^2 = 0.28$	$R^2 = 0.29$
2024. 5.16	1	149	本文上から4行目	28%	29%
2024. 5.16	1	152	手順2 例題の解き方	例題 5 2 は、	例題 5 3 は、
2024. 5.16	1	154	例題の解き方6行目	本文のように	本文と同程度の
2024. 5.16	1	154	例題の解き方7行目	最も大きい市区町村	最も大きい市
2024. 5.16	1	154	例題の解き方 10 行目	最も大きい残差は	残差の絶対値が最も大きくなるのは
2024. 5.16	1	154	囲み	説明変数群	回帰分析による予測値
2024. 5.16	1	154	囲み	目的変数の標準誤差の推定値	残差の標準偏差
2024. 5.16	1	154	囲み	目的変数を	目的変数に対し
2024. 5.16	1	155	例題 5 4	「平均年齢」が 50 歳である場合、	3 の回帰分析において「平均年齢」が 50 歳である場合、

			演習問題 2		
2024. 5.16	1	156		3,400 3,200 3,000 2,800 2,600 2,400 2,200 2,000 1,800 2,200 2,400 2,600 2,800 3,000 3,200 3,400 3,600 3,800	3,400 3,200 3,000 2,800 2,600 2,400 2,200 2,000 1,800 2,200 2,400 2,600 2,800 3,000 3,200 3,400 3,600 3,800
2024. 5.16	1	163	本文上から8行目	その平均(期待値)と分散はいくつになるでしょうか。また、その中でテレワークを実施している企業の数が 5 社以下である確率	その中でテレワークを実施している企業の数の平均(期待値)と分散はいくつになるでしょうか。また、その企業の数が 5 社以下である確率
2024. 5.16	1	165	例題の解き方 7 行目	=BINOM.DIST (0,18, 0.43, false)	=BINOM.DIST(C5,18, 0.43, false)
2024. 5.16	1	172	Tips	p.168 の玉ねぎの問題	例題 6 1 の問題
2024. 5.16	1	174	Tips	x が 6 以上の	x が 6 以下の
2024. 5.16	1	176	Tips	tが2以上の確率P(T≥2)	tが2以下の確率P(T≦2)
2024. 5.16	1	180	本文上から4行目	両辺を $P(A)$ で割ると、	右の等号について両辺をP(A)で割ると、
2024. 5.16	1	192	本文下から10行目	誤差マージン	誤差のマージン
2024. 5.16	1	194	Tips	母標準偏差σ	母標準偏差
2024. 5.16	1	197	本文下から2行目	正規分布の性質より、信頼度 95%の信頼区間について、次の式が成り立ちます。	正規分布の性質より、次の式が成り立ちます。
2024. 5.16	1	198	図表 7-3 図版名	正規分布の信頼度 95%の信頼区間	正規分布の上側 2.5%点と下側 2.5%点

			1		-
2024. 5.16	1	199	本文上から11行目	t分布の性質より、信頼度 95%の信頼区間について、次の式が成り 立ちます	t分布の性質より、次の式が成り立ちます。
2024. 5.16	1	199	図表 7-4 図版名	t分布の信頼度 95%の信頼区間	t分布の上側 2.5%点と下側 2.5%点
2024. 5.16	1	201	本文上から1行目	正規分布の性質より、信頼度 95%の信頼区間について、次の 式が成り立ちます。	正規分布の性質より、次の式が成り立ちます。
2024. 5.16	1	203	例題の解き方 最後行	区間推定できます。	区間推定できます。したがって、例題7 5 について「②オンライン 授業の方が成績がよい」と判断ができます。
2023. 8.15	1	212	8 行目	過去のデータから $\sigma_0^2 = \sigma_0^2$ であること \sim	過去のデータから $\sigma^2 = \sigma_0^2$ であること \sim
2024. 5.16	1	212	本文上から9行目	検定統計量Zは正規分布に	検定統計量Zは標準分布に
2024. 5.16	1	222	本文上から4行目	$+((不偏分散S_2^2/標本数n_2)^2)/(標本数n_2-1)$	$+(不偏分散S_2^2/標本数n_2)^2/(標本数n_2-1))$
2024. 5.16	1	228	Tips	CHIQ.DIST 関数	CHISQ.DIST 関数
2024. 5.16	1	234	例題8 設問文	動物統計調査	動態統計調査
2024. 5.16	1	235	①指数 上から1行目	2000 年から 2018 年までの入場者数と売上高に関して、それぞれ 19 個の	2000 年から 2019 年までの入場者数と売上高に関して、それぞれ 20 個の時系列データがあります。
2024. 5.16	1	235	本文下から4行目	入場者数であれば 5,500 万人から 8,000 万人の間、売上高であれば 3,000 億円から 7,000 億円の間と、	入場者数であれば 5,500 万人から 8,500 万人の間、売上高であれば 2,500 億円から 7,500 億円の間と、
2024. 5.16	1	237	Tips 最後行		2019 142.1 240.6 を追加
2024. 5.16	1	237	本文下から4行目	2000 年の売上高 298,530(百万円)を基準にした 2008 年の売上 高 456,396(百万円)の増加率を求めてみると、(456,396 - 298,530) / 298,530≒0.529(52.9%)	2000 年の売上高 298,532(百万円)を基準にした 2008 年の売上 高 456,396(百万円)の増加率を求めてみると、(456,396 - 298,532)/298,532≒0.529(52.9%)

2024. 5.16	1	249	本文下から2行目	図表 8-5 からも 1 年を単位の周期性	図表 8-4 からも 1 年を単位とした周期性
2024. 5.16	1	254	図表 8-11	図表8-11 中心化移動平均の表とグラフ B C D E F G H I J K 広橋A販売額 中心化 (信万円) 移動平均 2 533 3 546 4 546 5 735 995 8 800 6 571 6039 750 7 5567 615.8 700 8 555 6272 9 786 637.5 800 10 611 044.6 800 11 610 049.1 810 12 578 0563.3 800 14 631 808 14 631 808	図表8-11 中心化移動平均の表とグラフ B
2024. 5.16	1	254	本文下から2行目	季節指数とは、以下のように月別の時系列データであれば一月当たりの平均を100とし(年合計は、1200)、各月の季節成分の大きさを表したものです。	季節指数とは、月別の時系列データがあれば、それぞれの月の 平均値を計算し、1 月あたりの平均が 100(年合計が 1200)になる ように調整して、各月の季節成分の大きさを表したものです。
2024. 5.16	1	257	本文下から4行目	2.8%	2.9%
2024. 5.16	1	257	図表 8-15	対前年同月比 A B C D E F G H J J	対前年司期比 A B C D E F G H J J 対称年同期比
2024. 5.16	1	259	図表 8-16	(前年同月比)	(前年同期比)
2024. 5.16	1	263	例題の解き方	「94.8」	「94.7」

2024. 5.16	1	264	EXCEL WORK 内 タイトル、手順1、本文(2 か所)、 例題の解き方	対前年同月比	対前年同期比
2024. 5.16	1	264	EXCEL WORK 内 手順1、本文	対前年同期比	対前年同月比
2024. 5.16	1	264	例題の解き方	計算の結果と問題が四捨五入して	計算の結果と問題が
2024. 5.16	1	265	例題の解き方	対前年同月比(対前年同期比)	対前年同期比(対前年同月比)
2024. 5.16	1	265	例題の解き方	計算の結果と問題が四捨五入して	計算の結果と問題が
2024. 5.16	1	269	演習問題2の解答 問題2 4	18434.8	18434.7
2024. 5.16	2	269	演習問題2の解答 問題2 4	18434.9	18434.7
2024. 5.16	1	272	例題 10 設問文	50人の	30人の
2024. 5.16	1	272	例題 10 2	語句について、平均出現回数の男女差は	語句の一人当たりの平均使用回数の男女差は
2024. 5.16	1	278	図表 9-5	図表9-5 女性(左:n=12) と男性(右:n=18) の品詞ごとの形態素の出現頻度	Yell
2024. 5.16	1	278	本文下から4行目	123/12=10.25 個、男性が 177/18≒9.83 個の品詞となる形態素 を意見として述べていたことがわかります。また、男性も女性も名 詞の出現頻度が最も高かったり、女性は動詞よりも補助記号に分 類されるような形態素を多く述べていたりと、その出現頻度の違い による比較が可能です。	135/12=11.25 個、男性が 192/18≒10.67 個の形態素があったことがわかります。また、男性も女性も名詞の出現頻度が最も高いといった、出現頻度の違いによる比較が可能です。



2024. 5.16	1	285	手順 2 図	D5	DS
2024. 5.16	1	285	例題の解き方	出現頻度の値を AVERAGE 関数で計算すると、「女性=2.83…」「男性=3.11…」となり、平均出現回数の男女差を計算すると、「0.28431…」となり、問題文の解答条件から小数第2位までを四捨五入すると、「0.28」となります。	出現頻度の平均値を計算すると、「女性=2.83…」「男性=2.944 …」となり、平均出現回数の男女差を計算すると、「0.111…」となり、問題文の解答条件から小数第3位を四捨五入すると、「0.11」となります。
2024. 5.16	1	290	例題 10 2	0.28	0.11
2024. 5.16	1	292	演習問題 4	出現回数が	出現回数の総計が
2024. 5.16	1	300	Tips	•VHISQ.INV	•CHISQ.INV
2024. 5.16	1	300	本文下から8行目	分布の期待値と分散はそれぞれ、 $E(X) = p \lor V(X) = p(1-p) \lor x$ ります。また、生起確率 p のベルヌーイ試行を n 回繰り返した際に、その結果が従う分布は	その事象が発生した場合は $X=1$ 、発生しなかった場合は $X=0$ とした確率変数 X の期待値と分散はそれぞれ、 $E(X)=p$ と $V(X)=p$ ($1-p$)となります。また、生起確率 p のベルヌーイ試行を n 回繰り返した際に、その事象が発生した回数 X が従う分布は
2024. 5.165	1	302	本文下から5行目	試行回数 15 の	試行回数 10 の

2024. 5.16	1	310	本文上から7行目	$_{n}C_{x}(1/2)^{x}(1/2)^{n-x} = _{\lambda}C_{x}/2^{n}$ $_{n}C_{x}(1/2)^{x}(1/2)^{n-x} = _{n}C_{x}/2^{n}$
2023. 5.16	1	313	例題11 5 解答	3 4
2023. 4.28	1	323	問題4のエクセル画像	A B C D E F G H 1 ID 職業 レベル 攻撃力 防御力 魔法力 回避力 体力 2 1 A 25 2528 1400 341 1162 873 3 2 A 35 3862 1678 473 2102 710 4 3 B 16 201 486 169 217 605 5 4 A 97 8219 5142 794 5203 3111 6 5 A 41 5549 1965 601 2290 387 7 6 B 37 504 856 420 570 808 8 7 A 78 7151 2438 872 3428 2355 9 8 A 66 7482 1119 1020 2437 690
2024. 5.16	1	329	問題8 4	Bの形式の授業を受講し、再びAの授業を希望すると答えた 受講した形式がBで、希望する形式をAとした