

理屈はもういい、欲しいのは実践のお手本だ

書き込み式
演習型
ワークブック

専用ファイルを
ダウンロードして自習

本物の データ分析力が 身に付く本

経済産業省
「攻めのIT経営銘柄」選出
IT協会「IT総合賞」受賞

大阪ガス データサイエンティスト集団の ノウハウを公開!

著者
河村 真一
日置 孝一／野寺 綾／西脇 清行／山本 華世

1500人が受講した、
2日間のワークショップの
全てをこの1冊で
体験できる



	この本って何?.....4	
	用意するもの.....6	
プロローグ	データ分析の心構え7	
	1. よく起こる問題.....7	
	2. データ分析って何?.....9	
	3. データから考えない.....10	
	4. プロセスが全て.....11	
第1章	データ分析を設計する19	
	1-1. 分析ストーリーの見える化って何?.....20	
	1-2. なぜ「分析の概念図」を描くのか?.....22	
	1-3. 分析の概念図はこう描く.....23	
	1-4. 問題領域を決める.....24	
	(1) 問題領域を挙げる.....24	
	(2) 問題領域を選ぶ.....27	
	1-5. 評価軸を決める.....28	
	(1) 評価軸を挙げる.....28	
	(2) 評価軸を選ぶ.....31	
	1-6. 問題を文で表す.....32	
	1-7. 要因を挙げる.....34	
	【閑話休題】なぜ評価軸を挙げるのか?.....37	
	1-8. 要因を選ぶ.....38	
	(1) 重要度で仕分けする.....38	
	(2) 入手しやすさで仕分けする.....42	
	(3) 分析する要因を決める.....44	
	1-9. 部品をつなげる.....45	
	(1) 問題領域と評価軸を書く.....45	
	(2) 評価軸に要因をつなげる.....46	
	(3) 要因をグルーピングする.....46	
	(4) 要因同士の関係を推測してつなげる.....47	
	(5) 分析の流れを説明できるか確認する.....48	
	第1章の理解度チェック.....50	

第2章	データを事前にチェックする51
	2-1. なぜ事前チェックするのか?.....52
	2-2. データの出所をチェックする.....55
	(1) データの5W1H.....55
	(2) 一次情報かどうか.....55
	2-3. データの全体概要をチェックする.....57
	2-4. 個別の値をチェックする.....60
	(1) 欠損値をチェックする.....60
	(2) 外れ値をチェックする.....65
	(3) データの方向をチェックする.....70
	(4) データをクレンジングする.....71
	2-5. データの傾向をチェックする.....76
	第2章の理解度チェック.....78
第3章	分析方法を選ぶ79
	3-1. 代表値を使い分ける.....80
	(1) 3つの代表値を知る.....80
	(2) 代表値の得手・不得手.....82
	3-2. クロス集計する.....89
	(1) クロス集計とは?.....89
	(2) クロス集計の注意点.....94
	第3章の理解度チェック.....98
第4章	ケース実習「新商品の配置問題」99
	4-1. データ分析を設計する.....101
	(1) 問題領域を決める.....101
	(2) 評価軸を決める.....102
	(3) 問題を文で表す.....103
	(4) 要因を挙げる.....104
	(5) 要因を選ぶ.....106
	(6) 部品をつなげる.....107
	4-2. データを事前にチェックする.....109
	4-3. 分析方法を選ぶ.....110
	4-4. 分析を実行する.....114

第5章 標準偏差を使おう 117

- 5-1. 標準偏差って何?..... 118
- 5-2. 標準偏差はこう使う..... 120
 - (1) 多様性や格差を定量化する、比較する..... 120
 - (2) 不確実性を定量化する、比較する..... 121
 - (3) リスクを定量化する、比較する..... 122
 - (4) 平均値の信頼性を判断する、比較する..... 123
 - (5) 品質を管理する..... 124
- 5-3. 標準偏差を計算する..... 125
- 5-4. 標準偏差で意思決定する..... 128
 - (1) 標準偏差を使ってみる..... 128
 - (2) 外れ値を客観的な基準で特定する..... 130
 - (3) 2種類の標準偏差..... 139

第5章の理解度チェック..... 141

第6章 グループ間の差の確からしさを検証する 143

- 6-1. グループ間の大小関係は正しいとは限らない..... 144
- 6-2. 大小関係の確からしさをどう考える?..... 147
 - (1) 確率で考える..... 147
 - (2) 大小関係が確からしい確率は何で決まる?..... 148
- 6-3. 大小関係の確からしさを判断する..... 150
 - (1) 危険率を見る..... 150
 - (2) 何%以下なら確からしいか?..... 152
 - (3) 「対応なし」と「対応あり」..... 155

第6章の理解度チェック..... 159

第7章 分析結果の受け止め方と伝え方 161

- 7-1. 結果の解釈はここに注意..... 162
 - (1) 仮説検証バイアス..... 162
 - (2) アンカリング..... 165
 - (3) フレーミング..... 167
 - (4) プライミング..... 168
 - (5) 擬似相関..... 170
 - (6) まとめ..... 173
- 7-2. 結果の表現はここに注意..... 174
 - (1) データの集め方..... 175

- (2) グラフの見せ方..... 178
- (3) 言葉の表現の仕方..... 180
- (4) まとめ..... 182

- エビローク 1. 全体を振り返って..... 184
- 2. さらなる学習のために..... 185

謝辞..... 186

著者一覧..... 187

参考資料 本書のワークで使うエクセルの主な操作 (Windows版 Excel 2010の場合)

- 1. ワークシートの基本操作..... i
 - (1) 表示倍率の変更..... i
 - (2) ワークシートのコピー..... i
 - (3) ウィンドウ枠の固定..... i
 - (4) セル範囲の選択..... i
 - (5) コピー／貼り付けの速ワザ..... ii
 - (6) 行・列・セルの挿入／削除の速ワザ..... ii
 - (7) 行の並べ替え(ソート)..... ii
 - (8) 行の抽出(フィルター)..... iii
 - (9) データの表示書式の設定..... iii
 - (10) 直前の操作を繰り返す..... iv
 - (11) 直前の操作を取り消す..... iv
- 2. 基本の関数..... v
 - (1) 統計値(合計、平均、最大、最小、データの個数)..... v
 - (2) 条件による処理の分岐..... v
- 3. データの可視化..... vi
 - (1) 度数分布表とヒストグラムの作成..... vi
 - (2) 散布図の作成..... viii
- 4. クロス集計表の作成..... ix
 - (1) ピボットテーブルの作成..... ix
 - (2) 集計方法の変更..... x
 - (3) 複数の集計値の表示..... x
 - (4) ピボットテーブルの選択とコピー／貼り付け..... xi
 - (5) ピボットテーブルのレイアウト変更..... xii