

Python 演習で初歩から学ぶ 実習統計学入門

序章	Python で統計学	15
	序-1 データサイエンスと統計学、そして Python	16
	～ データサイエンスで最も人気の言語	
	序-2 すぐに使える Python	18
	～ Google Colabo を開くだけ!	
	序-3 Python プログラミング、さあ始めよう (基本編)	22
	～ 基本は簡単な数学と英語	
	序-4 Python プログラミング、さあ始めよう (関数編)	28
	～ Python は関数の組み合わせ	
第1章	統計学のための資料整理	31
	1-1 度数分布表の作成法 ～ 資料を整理整頓	32
	● 資料の整理	32
	● 用語解説	32
	● 度数分布表と階級	33
	● 相対度数分布表	33
	● 累積度数分布表	33
	実習1 Python による統計分析	34
	◇ Pythonレッスン numpy.histogram 関数	35
	◇ Pythonレッスン astype 関数	36
	【参考1】 Python・Excel 二刀流	37
	1-2 ヒストグラム作成法 ～ 度数分布表はヒストグラムで可視化	38
	● ヒストグラムと度数折れ線	38
	実習2 Python による統計分析	38
	◇ Pythonレッスン Matplotlib の bar 関数、plot 関数	40
	【参考2】 Python・Excel 二刀流	41
	1-3 代表値の算出法 ～ 膨大な資料を一つの数に集約	42
	● 平均値	42
	● 中央値 (メジアン)	42
	● 最頻値 (モード)	43
	● 代表値は一長一短	43

実習 3 Python による統計分析	44
Python レッスン mean、median、mode 関数	45
【参考 3】 Python・Excel 二刀流	45
1-4 分散と標準偏差 ～ データの散らばりは貴重な情報	46
● 偏差	46
● 変動	46
● 分散	47
● 標準偏差	47
実習 4 Python による統計分析	48
Python レッスン numpy.mean、numpy.var、numpy.std 関数	49
【参考 4】 Python・Excel 二刀流	49
1-5 度数分布表から求める平均値と分散 ～ 実用的な平均値と分散の求め方	50
● 度数分布表から平均値と分散を求める	50
● 例で見てみよう	50
● 分散公式のアレンジ	51
実習 5 Python による統計分析	51
Python レッスン numpy.dot 関数	52
【参考 5】 Python・Excel 二刀流	53
<memo> 正しいサイコロを 100 回振ったら、実際に出る目の度数は？	53
1-6 変量の標準化 ～ データを規格化して見やすくする技法	54
● 変量の標準化	54
実習 6 Python による統計分析	55
Python レッスン 2 次元配列に[:, 0]、[:, 1] を付けると？	57
【参考 6】 Python・Excel 二刀流	58
1-7 相関図の意味と作成法 ～ 2 変量の間係を图示する方法	59
● 相関図	59
● 正の相関・負の相関	60
実習 7 Python による統計分析	60
Python レッスン matplotlib.pyplot.scatter 関数	61
【参考 7】 Python・Excel 二刀流	62
<memo> print 関数の引数 end	62
1-8 共分散と相関係数の意味と求め方 ～ 2 変量の間係の数値化	63
● 共分散	63
● 共分散の具体例	64
● 相関係数	64
実習 8 Python による統計分析	66
Python レッスン numpy.cov、numpy.corrcoef 関数	67
【参考 8】 Python・Excel 二刀流	68

第 2 章 確率論の基本

2-1 確率の定義 ～ 数学的な確率と統計的な確率	70
● 試行と事象	70
● 数学的な確率	70
● 統計的な確率	71
● 反復試行の確率の定理	71
実習 9 Python による統計分析	71
Python レッスン scipy.stats.randint.rvs 関数	73
Python レッスン 配列の型と要素数の取得	73
【参考 9】 Python・Excel 二刀流	74
<memo> 一様乱数からもサイコロの実験が可能	74
2-2 確率変数と確率分布 ～ 推測統計学の基本中の基本	75
● 確率変数	75
● 確率分布	75
● 連続的な確率変数と確率密度関数	76
● 累積分布関数	77
● パーセント点	77
● p 値	78
実習 10 Python による統計分析	78
Python レッスン numpy.random.choice 関数	80
Python レッスン np.count_nonzero 関数	80
Python レッスン 出力の表示形式	80
【参考 10】 Python・Excel 二刀流	81
2-3 確率変数の平均値・分散 ～ 変量の平均値・分散を抽象化	82
● 離散的な値をとる確率分布	82
● 分散の公式の変形	83
● 連続的な値をとる確率変数	83
● 標準偏差は確率分布の広がりの幅	84
実習 11 Python による統計分析	85
Python レッスン numpy.repeat 関数	87
【参考 11】 Python・Excel 二刀流	87
2-4 確率変数の標準化 ～ 不ぞろいな分布を統一化	88
● 確率変数の変換公式	88
● 確率変数の標準化	88
実習 12 Python による統計分析	89
Python レッスン エスケープシーケンス	91
【参考 12】 Python・Excel 二刀流	92

第3章 統計学のための有名な確率分布	93	3-6 正規分布の100 p %点の求め方 ~ 統計学で最も利用される数値	117
3-1 見てわかる一様分布 ~ 様々な確率分布の基本	94	• パーセント点の復習	117
• 一様分布	94	• Pythonで正規分布の100 p %点を求めるには	117
• 離散一様分布	94	実習18 Pythonによる統計分析	117
• 例で考えてみる	95	◇ Pythonレッスン norm.ppf 関数	118
実習13 Pythonによる統計分析	95	【参考18】 Python・Excel 二刀流	119
<memo> 2つの命令文を1行で表示するには	96	3-7 二項分布の正規分布近似 ~ 扱いにくい二項分布が優等生に	120
◇ Pythonレッスン scipy.stats.uniform.rvs 関数	97	• 二項分布の正規分布近似	120
◇ Pythonレッスン np.histogramの「density = True」指定	97	実習19 Pythonによる統計分析	121
<memo> 一様分布の見方	98	◇ Pythonレッスン binom.stats 関数	122
【参考13】 Python・Excel 二刀流	99	<memo> 二項分布の正規分布近似の大切さ	122
3-2 見てわかるベルヌーイ分布 ~ 比率を考えると基本となる分布	100	【参考19】 Python・Excel 二刀流	123
• ベルヌーイ分布	100	3-8 見てわかる t 分布 ~ 分散が不明な場合の推定・検定に活躍	124
• 例で見てみよう	100	• t 分布とは	124
実習14 Pythonによる統計分析	101	• 例で見てみよう	124
◇ Pythonレッスン reshape 関数	102	実習20 Pythonによる統計分析	125
【参考14】 Python・Excel 二刀流	103	◇ Pythonレッスン t.pdf 関数	126
3-3 見てわかる二項分布 ~ ベルヌーイ分布と融合して活躍	104	◇ Pythonレッスン t.stats 関数	127
• 二項分布とは	104	【参考20】 Python・Excel 二刀流	127
• 例で見てみよう	104	3-9 t 分布の100 p %点の求め方 ~ t .ppf 関数が活躍	128
実習15 Pythonによる統計分析	105	• Pythonで t 分布の100 p %点を求める	128
◇ Pythonレッスン binom.pmf 関数	107	実習21 Pythonによる統計分析	128
【参考15】 Python・Excel 二刀流	107	◇ Pythonレッスン t.ppf 関数	129
3-4 見てわかる正規分布 ~ 統計学の王道となる分布	108	【参考21】 Python・Excel 二刀流	130
• 正規分布	108	3-10 見てわかる χ^2 分布 ~ 分散の推定・検定に活躍	131
• 例で見てみよう	108	• χ^2 分布とは	131
実習16 Pythonによる統計分析	109	• 例で見てみよう	131
◇ Pythonレッスン scipy.stats.norm.pdf 関数	111	実習22 Pythonによる統計分析	132
◇ Pythonレッスン scipy.stats.norm.cdf 関数と累積分布関数	111	◇ Pythonレッスン chi2.pdf 関数	133
◇ Pythonレッスン numpy.linspace 関数	112	【参考22】 Python・Excel 二刀流	134
【参考16】 Python・Excel 二刀流	113	3-11 χ^2 分布の100 p %点の求め方 ~ chi2.ppf 関数が活躍	135
3-5 見てわかる標準正規分布 ~ 正規分布を標準化した分布	114	• Pythonで χ^2 分布の100 p %点を求める	135
• 標準正規分布	114	実習23 Pythonによる統計分析	135
• 標準化	114	◇ Pythonレッスン chi2.ppf 関数	136
実習17 Pythonによる統計分析	115	◇ Pythonレッスン roundとnp.round	136
◇ Pythonレッスン NumPyで正規分布のグラフ描画	116	【参考23】 Python・Excel 二刀流	137
【参考17】 Python・Excel 二刀流	116	<memo> χ^2 分布の命名者	137

3-12 見てわかる F 分布 ～ 分散を分析するときの基本となる分布	138
• F 分布	138
• 例で見てみよう	138
<memo> ガンマ関数	139
実習 24 Python による統計分析	140
Python レッスン f.pdf 関数	141
【参考 24】 Python・Excel 二刀流	142
3-13 F 分布の $100p\%$ 点の求め方 ～ f.ppf 数が活躍	143
• Python で F 分布の $100p$ パーセント点を求める	143
実習 25 Python による統計分析	143
Python レッスン f.ppf 関数	144
<memo> F 分布の由来	144
【参考 25】 Python・Excel 二刀流	145
<memo> legend 関数	145
(寄り道) 変数が 2 種ある場合の確率分布	146

第 4 章 Python でわかる母集団と標本の関係 147

4-1 母集団と標本抽出 ～ 標本変動の理解が推定・検定の基本	148
• 標本調査と推測統計学	148
• 母集団と標本	148
• 標本の抽出	148
• 母集団分布	149
• 正規母集団	150
• 統計量と標本分布	150
• 推定量と推定値	150
<memo> 復元抽出と非復元抽出	151
実習 26 Python による統計分析	152
Python レッスン scipy.stats.norm.rvs 関数で正規母集団からの抽出	153
【参考 26】 Python・Excel 二刀流	153
4-2 不偏性と不偏分散 ～ 実用的だが難しい自由度の考え方	154
• 分散の復習	154
• 不偏性と不偏分散	154
実習 27 Python による統計分析	155
Python レッスン 分散のパラメータ ddof	156
【参考 27】 Python・Excel 二刀流	157
<memo> 不偏推定量と自由度	158
4-3 中心極限定理 ～ すべてが正規分布に帰着するという実用的な定理	159
• 標本平均の分布と正規分布	159

実習 28 Python による統計分析	161
<memo> 複数の命令文を 1 行に書く	162
Python レッスン ax.bar、ax.plot の color、label 指定	163
【参考 28】 Python・Excel 二刀流	164

第 5 章 Python でわかる推定 165

5-1 見てわかる最尤推定法 ～ 「起こりやすいことが起こる」という考え方	166
• 画紙で分かる最尤推定法	166
• 最尤推定法をまとめてみよう	167
実習 29 Python による統計分析	167
Python レッスン ユーザ関数の定義法	168
Python レッスン scipy.optimize.fminbound 関数	169
<memo> 関数を利用するときのいろいろな表現	169
【参考 29】 Python・Excel 二刀流	170
5-2 見てわかる区間推定の考え方 ～ 推定に信頼度が付加される	171
• 区間推定の考え方	171
• 信頼度、信頼区間をグラフで見ると	172
• より確実に推定したければ	173
実習 30 Python による統計分析	174
Python レッスン 係数 1.96 の自動取得と norm.ppf 関数	175
【参考 30】 Python・Excel 二刀流	177
5-3 分散既知の正規母集団における母平均の推定 ～ 区間推定の基本問題	178
• 95% の信頼区間	178
• 99% の推定区間	179
• 母平均 μ を区間推定する手順	179
• 例で見てみよう	180
実習 31 Python による統計分析	180
Python レッスン floor、ceil 関数と丸め処理	182
【参考 31】 Python・Excel 二刀流	183
<memo> 標準誤差と SciPy	183
5-4 分散未知の正規母集団における母平均の推定 ～ t 分布が活躍する推定	184
• 母分散未知の母平均の推定の考え方	184
• 母平均 μ を区間推定する手順	185
• 例で見てみよう	185
実習 32 Python による統計分析	186
Python レッスン 実数型を整数型にする int 関数、astype(int)	187
【参考 32】 Python・Excel 二刀流	188
<memo> 直接信頼区間を算出する関数	188

5-5	大きな標本の場合の母平均の推定 ~ 中心極限定理が本領発揮する推定	189	● 右側検定 … 「母平均 $\mu > \mu_0$ 」を主張したい場合	213	
	● 大標本の場合の母平均の推定の考え方	189	● 例で見てみよう	214	
	● 母平均 μ を区間推定する手順	190	実習 36 Python による統計分析	215	
	● 例で見てみよう	190	◇ Pythonレッスン 棄却域の丸め方と ceil、floor 関数	216	
	実習 33 Python による統計分析	191	【参考 36】 Python・Excel 二刀流	217	
	◇ Pythonレッスン 分散、標準偏差の関数いろいろ	192			
	【参考 33】 Python・Excel 二刀流	192			
5-6	母比率の推定 ~ ベルヌーイ分布と中心極限定理が協力した区間推定	193	6-3	分散未知の正規母集団の母平均の検定 ~ t 分布が活躍する検定	218
	● 母比率と標本比率	193		● 母分散未知の母平均の検定の考え方	218
	● 標本と平均値を結ぶベルヌーイ分布	193		● 両側検定 … 「母平均 $\mu \neq \mu_0$ 」を主張したい場合	219
	● 母比率の推定の公式	194		● 右側検定 … 「母平均 $\mu > \mu_0$ 」を主張したい場合	219
	● 母平均 p を区間推定する手順	195		● 例で見てみよう	220
	● 例で見てみよう	196		実習 37 Python による統計分析	221
	実習 34 Python による統計分析	196		◇ Pythonレッスン ttest_1samp 関数と p 値による検定	222
	<memo> 実際の未婚率	197		【参考 37】 Python・Excel 二刀流	223
	◇ Pythonレッスン round 関数と型違い	198	6-4	大きな標本での母平均の検定 ~ 中心極限定理が本領発揮する検定	224
	【参考 34】 Python・Excel 二刀流	199		● 大きな標本の場合の母平均の検定の考え方	224
	<memo> print の書式設定	199		● 両側検定 … 「母平均 $\mu \neq \mu_0$ 」を主張したい場合	225
				● 右側検定 … 「母平均 $\mu > \mu_0$ 」を主張したい場合	225
				● 例で見てみよう	226
5-7	母分散の推定 ~ χ^2 分布が活躍する推定	200		実習 38 Python による統計分析	227
	● 母分散の推定の基本的な考え方	200		◇ Pythonレッスン 標準誤差と sem 関数	229
	● 母分散 σ^2 を区間推定する手順	201		【参考 38】 Python・Excel 二刀流	229
	● 例で見てみよう	201	6-5	母比率の検定 ~ ベルヌーイ分布と中心極限定理が協力した検定	230
	実習 35 Python による統計分析	202		● 母比率の検定の考え方	230
	◇ Pythonレッスン print の書式と f 文字列	203		● 両側検定 … 「母比率 $p \neq p_0$ 」を主張したい場合	231
	【参考 35】 Python・Excel 二刀流	204		● 右側検定 … 「母比率 $p > p_0$ 」を主張したい場合	232
	(発展) 関数一つで区間推定ができる!	205		● 例で見てみよう	232
	(寄り道) 本書の統計学とベイズ統計	206		実習 39 Python による統計分析	233
				◇ Pythonレッスン 有意水準の変更とパーセント点	235
				【参考 39】 Python・Excel 二刀流	235
第 6 章	Python でわかる検定	207	6-6	母分散の検定 ~ χ^2 分布が活躍する検定	236
6-1	統計的検定の仕組みと有意水準の意味 ~ 「そんなはずがない」が発想の原点	208		● 母分散の検定の考え方	236
	● 統計的検定の考え方	208		● 両側検定 … 「母分散 $\sigma^2 \neq \sigma_0^2$ 」を主張したい場合	237
	● 例で見てみよう	209		● 右側検定 … 「母分散 $\sigma^2 > \sigma_0^2$ 」を主張したい場合	237
	● 帰無仮説と対立仮説	209		● 例で見てみよう	238
	● 検定の手順	210		実習 40 Python による統計分析	239
	● 片側検定と両側検定	211		◇ Pythonレッスン p 値による検定と関数 stats.chi2.cdf	240
				<memo> p 値による検定	241
6-2	分散既知の正規母集団の母平均の検定 ~ 検定の基本となる問題	212		【参考 40】 Python・Excel 二刀流	242
	● 分散既知の正規母集団についての母平均の検定	212			
	● 両側検定 … 「母平均 $\mu \neq \mu_0$ 」を主張したい場合	213			

6-7 母比率の差の検定 ～標本比率の差が標準正規分布に従う……………	243
● 母比率の差の検定の考え方……………	243
● 両側検定 …「母比率 $\alpha \neq \beta$ 」を主張したい場合……………	244
● 右側検定 …「母比率 $\alpha > \beta$ 」を主張したい場合……………	244
● 例で見てみよう……………	245
実習 41 Python による統計分析……………	246
Pythonレッスン abs 関数……………	248
【参考 41】 Python・Excel 二刀流……………	248
6-8 等分散の検定 ～ F 分布が活躍する検定……………	249
● 等分散の検定の考え方……………	249
● 両側検定 …「母分散 $\sigma_A^2 \neq \sigma_B^2$ 」を主張したい場合……………	250
● 右側検定 …「母分散 $\sigma_A^2 > \sigma_B^2$ 」を主張したい場合……………	250
● 例で見てみよう……………	251
実習 42 Python による統計分析……………	252
Pythonレッスン p 値による検定と関数 stats.f.cdf……………	254
【参考 42】 Python・Excel 二刀流……………	255
<memo>f.stats 関数……………	255
6-9 第 2 種の誤りと検出力 ～有意水準と検定の正しさとの絶妙な関係……………	256
● 第 1 種の誤りと第 2 種の誤り……………	256
● 過誤のジレンマ……………	257
● 検出力……………	257
● 例で見てみよう……………	258
実習 43 Python による統計分析……………	259
【参考 43】 Python・Excel 二刀流……………	261

付録

A CSV ファイルを Google ドライブに格納する方法……………	262
B Excel データを CSV ファイルに変換する方法……………	264
C CSV ファイルを NumPy で読む方法 (Google ドライブ利用)……………	266
D Windows のコマンドプロンプトの利用法……………	268
E 個人のパソコンへ Python をインストールする方法……………	271
F Windows コマンドプロンプトで Python を実行……………	274
索引……………	277