SSDSE を使ってみよう

― EXCEL を使った SSDSE データの使用例 ―

第５回　市区町村別の飲食店数（標本抽出の例）

◎　SSDSE-市区町村から算出した「人口千人当たりの飲食店数」のデータを使用し、全市区町村の中から一部の市区町村のデータを取り出すという形で、**標本調査（標本抽出）を行い標本の平均値を分析**します。

使用データ　　SSDSE-市区町村　2026年版（SSDSE-A-2026）

分析の内容　　母集団と標本、標本の抽出、標本平均、四分位数、箱ひげ図、ヒストグラム

【目次】

1　母集団と標本（標本調査） ---------- 2
2　分析用のデータセットを作成する ---------- 3
3　データの分布を確認する　（基本統計量、ヒストグラム、昇順の並べ替え） ---------- 5
4　標本（市区町村）を抽出し、標本平均を計算する　（標本の抽出、標本平均） ---------- 7
5　抽出を繰り返し、得られた標本平均を分析する　（箱ひげ図） ---------- 9
6　得られた標本平均のヒストグラムを作成する　（度数分布表、ヒストグラム） ---------- 14
7　標本調査（標本抽出）のまとめ ---------- 15

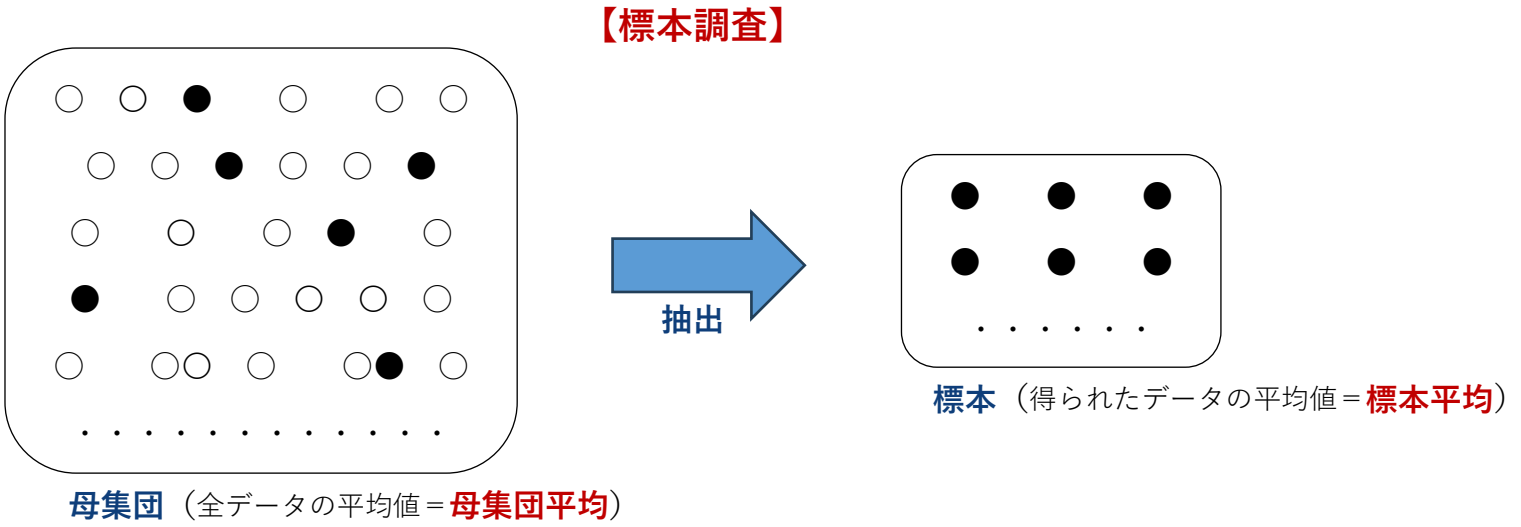
※　この資料では、Microsoft® Excel® for Microsoft 365 を使用しています。

1　母集団と標本（標本調査）

◎　本書では、全市区町村の「人口千人当たり飲食店数」のデータから、一部の市区町村のデータを取り出して分析します。

◎　これは、中学校の第３学年で学ぶ「標本調査」について、現実の統計データを材料として、Excelで実施する事例の紹介です。

◎　全市区町村のデータ（母集団）から一部の市区町村のデータ（標本）を無作為に取り出す（抽出する）方法を説明します。



▶本書で行う分析内容

①　標本の大きさ（データの数）を決めた上で標本調査（標本抽出）を行い、得られたデータの平均値（標本平均）を計算します。

②　①の標本抽出を繰り返して行い、各回の標本平均を記録し、得られた標本平均の分布について分析します。

③　①の標本の大きさを変えたり、②の標本抽出の繰り返しの回数を変えたりして、標本平均の分布の違いを分析します。

2　分析用のデータセットを作成する

◎　**SSDSE-市区町村**から、今回の分析で使用するセルを残し、他の行・列をすべて削除します。

◎　使用するデータ項目は「**A1101　総人口**」と「**H6131　飲食店数**」の２項目です。

SSDSE-A-202	Prefecture	Municipality	A1101	A110101		H6130	H6131	H6132		I6300	J250302
		年度	2020	2020		2021	2021	2021		2022	2023
地域コード	都道府県	市区町村	総人口	総人口（男）		小売店数	飲食店数	大型小売店数		薬剤師数	保育所等数（基本票）
R01100	北海道	札幌市	1973395	918682		10370	7354	348		5561	297
R01202	北海道	函館市	251084	113965		2163	1256	41		696	25
R01203	北海道	小樽市	111299	50136		1097	616	17		331	21
R01204	北海道	旭川市	329306	152108		2409	1503	67		853	55
R01205	北海道	室蘭市	82383	40390		621	447	13		201	10
R01206	北海道	釧路市	165077	77506		1224	1030	24		379	16
R01207	北海道	帯広市	166536	79623		1431	1123	35		431	29
R47359	沖縄県	伊平屋村	1126	607		19	10	0		0	1
R47360	沖縄県	伊是名村	1322	718		17	8	0		0	1
R47361	沖縄県	久米島町	7192	3823		87	65	0		10	3
R47362	沖縄県	八重瀬町	30941	15244		153	55	3		54	15
R47375	沖縄県	多良間村	1058	575		14	5	0		0	1
R47381	沖縄県	竹富町	3942	2033		58	71	0		0	0
R47382	沖縄県	与那国町	1676	923		24	23	0		0	1

◎　新しい列（データ項目）として「人口千人当たり飲食店数」を追加します。（計算式は、飲食店数÷総人口×1000）

地域コード	都道府県	市区町村	総人口	飲食店数	人口千人当たり飲食店数
R01100	北海道	札幌市	1973395	7354	3.73
R01202	北海道	函館市	251084	1256	5.00
R01203	北海道	小樽市	111299	616	5.53
R01204	北海道	旭川市	329306	1503	4.56
R01205	北海道	室蘭市	82383	447	5.43
R01206	北海道	釧路市	165077	1030	6.24
R01207	北海道	帯広市	166536	1123	6.74
R47359	沖縄県	伊平屋村	1126	10	8.88
R47360	沖縄県	伊是名村	1322	8	6.05
R47361	沖縄県	久米島町	7192	65	9.04
R47362	沖縄県	八重瀬町	30941	55	1.78
R47375	沖縄県	多良間村	1058	5	4.73
R47381	沖縄県	竹富町	3942	71	18.01
R47382	沖縄県	与那国町	1676	23	13.72

- 総人口が０の福島県双葉町（R07546）は計算結果がエラー（#DIV/0!）になるので、分析から除外するため双葉町の行を削除しておきます。
（以後、市区町村数は、1741－1＝1740 となります。）

- 人口千人当たり飲食店数の各セルには、数式が入っている状態なので、分析中に表示が変わらないよう、数値に置き換えておきます。

数式を数値に置き換え
・「人口千人当たり飲食店数」の列全体を選択
↓
・選択した状態で右クリックしてコピーをクリック
↓
・選択した状態のままで右クリックして、
形式を選択して貼り付け ＞ 値　をクリック
（または貼り付けのオプションで 123 をクリック）

（なお、ここでは、数値が小数点以下２桁で表示されるように書式設定しています。）

◎　結果として、縦１７４１行×横６列のデータになります。これを分析用データセットとして使用します。
（市区町村数は、1741－1＝1740 となります。）

3　データの分布を確認する

◎　**市区町村別の「人口千人当たり飲食店数」の分布の状況を確認します。**

◎　**平均値、四分位数等の主な統計量**は右の**表１**のとおりです。

◎　**市区町村別の「人口千人当たり飲食店数」のヒストグラム**を、Excelのグラフ作成機能により作成すると、**下図**のようになります。

表１　人口千人当たり飲食店数の主な統計量

	数値	使用した関数
平均値	**3.94**	AVERAGE
最小値	0.00	MIN
最大値	45.40	MAX
第１四分位数	2.52	QUARTILE.EXC
中央値	**3.46**	MEDIAN
第３四分位数	4.64	QUARTILE.EXC
四分位範囲	2.13	

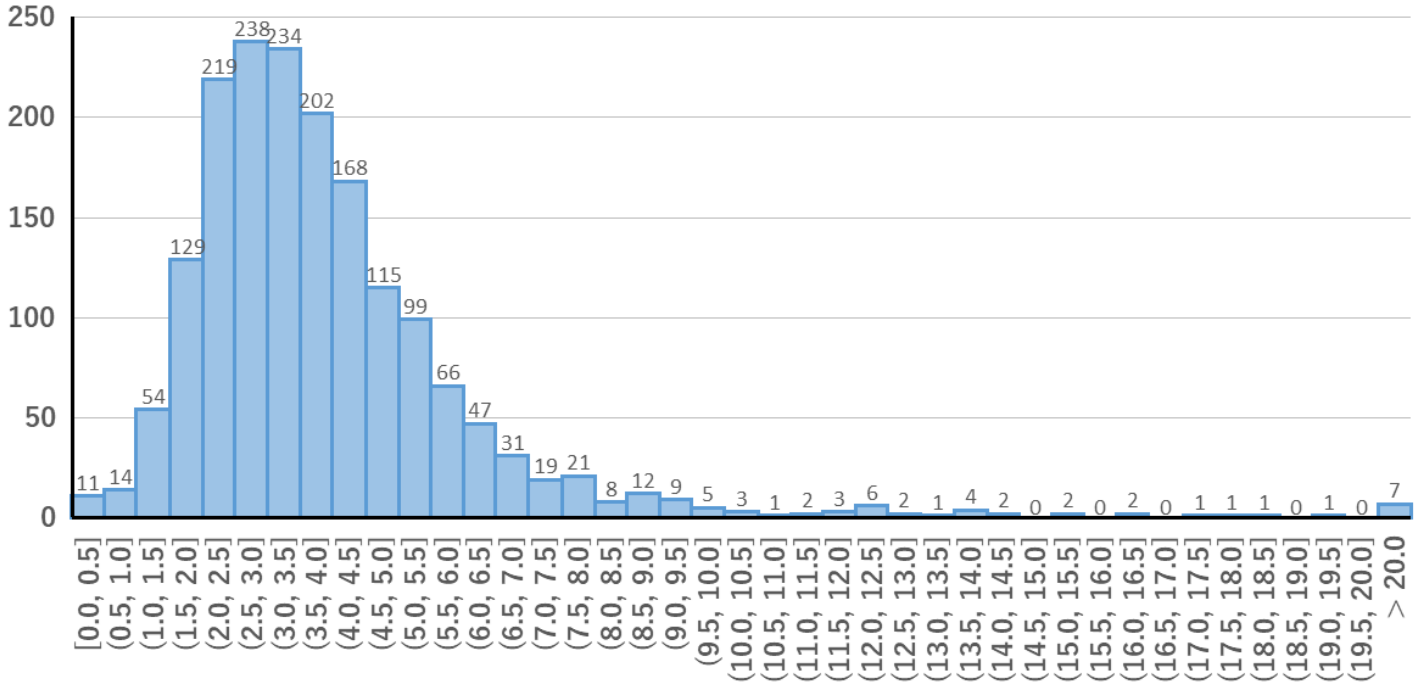


図１　人口千人当たり飲食店数のヒストグラム

◎　**データ（人口千人当たり飲食店数）を昇順に並べ替えた結果**は、**表２**のとおりです。

●　データの値が小さい市区町村は町村が占めていますが、データの値が大きい市区町村は、大都市と町村が混在していることがわかります。

▶　住んでいる市区町村の値などが、どのあたりの順位にあるか、確認してみましょう。

表２　人口千人当たり飲食店数を昇順に並べた結果

地域コード	都道府県	市区町村	総人口	飲食店数	人口千人当たり飲食店数
R07548	福島県	葛尾村	420	0	0.00
R07564	福島県	飯舘村	1318	0	0.00
R20307	長野県	北相木村	752	0	0.00
R43513	熊本県	球磨村	2433	0	0.00
R46303	鹿児島県	三島村	405	0	0.00
R46304	鹿児島県	十島村	740	0	0.00
R29345	奈良県	安堵町	7225	1	0.14
R05368	秋田県	大潟村	3011	1	0.33
R01425	北海道	上砂川町	2841	1	0.35
R40401	福岡県	小竹町	7151	3	0.42
R07364	福島県	檜枝岐村	504	9	17.86
R47381	沖縄県	竹富町	3942	71	18.01
R14382	神奈川県	箱根町	11293	220	19.48
R29446	奈良県	天川村	1176	24	20.41
R20409	長野県	平谷村	387	8	20.67
R15586	新潟県	粟島浦村	353	8	22.66
R13102	東京都	中央区	169179	4117	24.34
R47354	沖縄県	座間味村	892	23	25.78
R21604	岐阜県	白川村	1511	40	26.47
R13101	東京都	千代田区	66680	3027	45.40

4　標本（市区町村）を抽出し、標本平均を計算する

◎　**一部の市区町村を無作為に抽出**し、抽出した市区町村の「**人口千人当たり飲食店数**」の**平均値（標本平均）を計算**します。
ここでは、エクセルの**RAND関数**（乱数発生）と**RANK関数**（順位付け）を用いた抽出方法の例を紹介します。

◎　**分析用データセット**の右側に、次のように**㋑～㋓の３つの列を追加**して標本抽出を行い、標本平均を計算します。
（この例では**標本の大きさ（データの個数）を 25** として抽出します。）

表３　標本の大きさを 25 として抽出した結果（例）

	㋐	㋑	㋒	㋓
市区町村	人口千人当たり飲食店数	乱数	順位	抽出結果【25】
札幌市	3.73	0.02669807	1695	
函館市	5.00	0.084085849	1597	
小樽市	5.53	0.304318568	1201	
旭川市	4.56	0.381488743	1049	
室蘭市	5.43	0.942722534	107	
釧路市	6.24	0.998133491	6	6.24
帯広市	6.74	0.432876063	974	
北見市	4.61	0.573951817	733	
北大東村	6.78	0.281286063	1242	
伊平屋村	8.88	0.783371779	380	
伊是名村	6.05	0.202632583	1374	
久米島町	9.04	0.930434331	130	
八重瀬町	1.78	0.218536078	1345	
多良間村	4.73	0.955077711	92	
竹富町	18.01	0.845320306	270	
与那国町	13.72	0.337806993	1131	
データの数	1740			25
平均値	3.94			3.93

①　**列㋑にRAND関数により乱数**（０以上１未満の数）を生成します。
●㋑の各セルに　**=RAND ( )** を入力

②　**列㋒にRANK関数により列㋑の乱数の中の順位を表示**します。
●㋒の各セルに　**=RANK（同じ行の列㋑，列㋑の全範囲(赤枠)）** を入力

③　各行で**列㋒の順位**と**標本の大きさ**（この例では **25** ）とを比較し、
順位が標本の大きさ 25 以下の市区町村を抽出します。
列㋓には、**順位≦標本の大きさ**　なら、**人口千人当たり飲食店数**の値、
順位＞標本の大きさ　なら、**ブランク**を表示します。
●㋓の各セルに　**=IF（同じ行の列㋒ <=25，同じ行の列㋐，""）** を入力

列㋓に数値が表示された市区町村が、抽出された市区町村となります。

④　**列㋓緑枠にデータの数（標本の大きさ）と平均値（標本平均）**を表示します。
●**データの数**は　**=COUNT（列㋓の青太枠）**　（数値が入っているセルの数）
25になることを確認します。
●**平均値**は　**=AVERAGE（列㋓の青太枠）**　（ブランクのセルは計算から除外）

● **前ページは、標本の大きさを 25 とした場合の標本抽出の例ですが、**
これに準じて、標本の大きさを 25、50、100、200 の４通りにして標本抽出した例は、次の**表４**のようになります

表４　標本の大きさを４通りにして抽出した結果（例）

市区町村	人口千人当たり飲食店数	乱数	順位	抽出結果【25】	抽出結果【50】	抽出結果【100】	抽出結果【200】
札幌市	3.73	0.02669807	1695				
函館市	5.00	0.084085849	1597				
小樽市	5.53	0.304318568	1201				
旭川市	4.56	0.381488743	1049				
室蘭市	5.43	0.942722534	107				5.43
釧路市	6.24	0.998133491	6	6.24	6.24	6.24	6.24
帯広市	6.74	0.432876063	974				
北見市	4.61	0.573951817	733				
北大東村	6.78	0.281286063	1242				
伊平屋村	8.88	0.783371779	380				
伊是名村	6.05	0.202632583	1374				
久米島町	9.04	0.930434331	130				9.04
八重瀬町	1.78	0.218536078	1345				
多良間村	4.73	0.955077711	92			4.73	4.73
竹富町	18.01	0.845320306	270				
与那国町	13.72	0.337806993	1131				
データの数	1740			25	50	100	200
平均値	3.94			3.93	3.79	3.90	4.02

◎　この**表４**で**「再計算」を行うと、**
すべての市区町村の乱数が変わるので、
赤枠内の表示・数値が変動し、
抽出される市区町村が変わります。

「再計算」を行うたびに、
新たな標本抽出の結果が得られる
と考えることができます。

※ **再計算は 数式タブ ＞ 再計算**　（又は**Ｆ９キー**）

5　抽出を繰り返し、得られた標本平均を分析する

◎　標本抽出を繰り返して得られた結果は、**前ページの表４**のワークシートの、一番下にある**標本平均（緑太枠）を記録**することで得られます。

◎　右の**表５**は、**標本抽出を 30 回繰り返した結果**の一例です。
表４の**標本平均（緑太枠）**を、**表５**の各行に **コピー ＞ 貼り付け** たものです。

注）ここで、**コピー ＞ 貼り付け** は、**必ず数値に置き換えて貼り付け**ます。
形式を選択して貼り付け ＞ 値　をクリック
または、**貼り付けのオプション**　　をクリック

◎　このデータについて、平均値・四分位数を計算すると、下のようになります。
※　**表５のデータは抽出するたびに変動するので、参考値と考えてください。**

表６　標本平均の分布の平均値と四分位数の例（標本抽出 30 回）

	標本の大きさ【25】	標本の大きさ【50】	標本の大きさ【100】	標本の大きさ【200】
平均値	**4.01**	**3.97**	**4.01**	**4.01**
最小値	3.18	3.12	3.47	3.69
最大値	5.19	4.91	4.81	4.42
第１四分位数	3.69	3.68	3.71	3.88
中央値	**3.92**	**4.03**	**4.00**	**4.00**
第３四分位数	4.30	4.21	4.20	4.12
四分位範囲	**0.60**	**0.53**	**0.49**	**0.24**

※　母集団平均は 3.94

表５　標本平均の一例（標本抽出 30 回）

抽出回数	標本の大きさ【25】	標本の大きさ【50】	標本の大きさ【100】	標本の大きさ【200】
1	3.67	4.05	4.00	4.02
2	3.18	3.19	3.47	3.82
3	4.81	4.14	3.92	4.07
4	3.98	4.07	4.16	3.94
5	3.98	4.91	4.45	4.13
6	3.84	3.47	3.69	3.79
7	5.19	4.52	4.19	3.99
8	3.90	3.69	3.64	3.89
9	3.79	4.02	3.94	3.93
10	3.77	3.63	3.71	3.84
11	3.50	3.89	4.34	3.97
12	4.28	4.26	4.13	4.10
13	3.39	3.12	3.64	3.69
14	3.82	4.09	3.87	3.85
15	4.09	4.45	4.21	4.11
16	4.67	4.47	4.32	4.10
17	3.70	3.59	3.70	3.86
18	4.24	3.88	4.29	4.01
19	3.67	3.87	4.09	4.15
20	3.90	4.20	4.10	4.04
21	4.35	4.15	4.24	4.24
22	4.03	3.73	4.00	3.94
23	4.63	4.21	4.06	4.03
24	3.91	3.95	3.97	4.19
25	3.47	3.58	3.68	3.75
26	4.68	4.32	4.81	4.42
27	4.37	4.21	4.13	4.38
28	3.50	3.49	3.70	3.90
29	3.93	3.73	3.82	3.94
30	4.03	4.16	3.94	4.25

◎　箱ひげ図は右のようになります。

◎　これを見ると、標本の大きさが大きいほど、得られる
標本平均の範囲が小さくなる傾向がありそうです。

● **検証のため、さらに 30 回の標本抽出を 2 回実施**しました。

★　ここでは、**最初に実施**した 30 回の標本抽出を**試行 A**、
追加実施した 30 回の標本抽出を**試行 B**、**試行 C** と呼びます。

◎　**試行 B** と**試行 C** の箱ひげ図は下のとおりです。

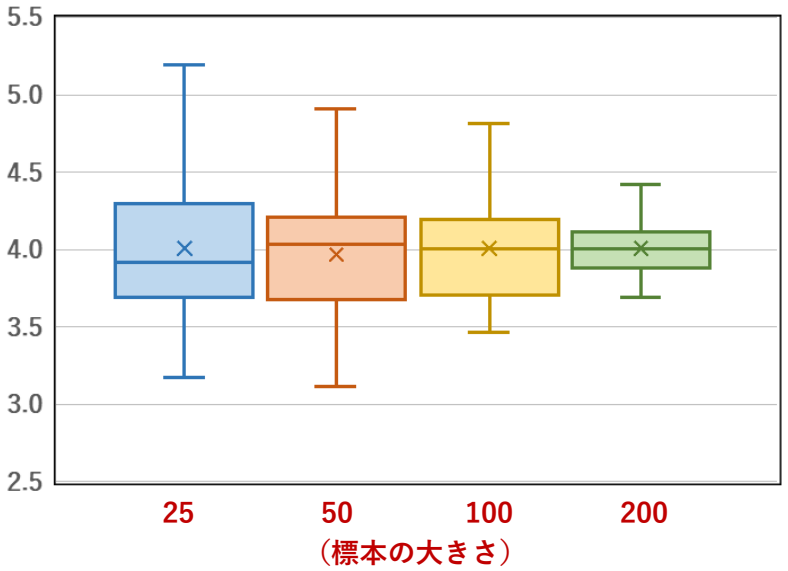


図 2 A　標本平均の分布の箱ひげ図の例（標本抽出 30 回）試行 A

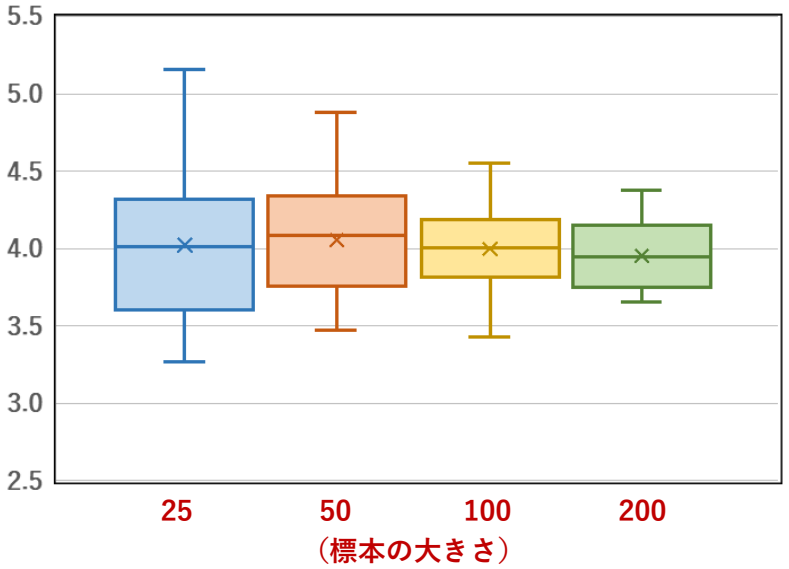


図 2 B 標本平均の分布の箱ひげ図の例（標本抽出 30 回）試行 B

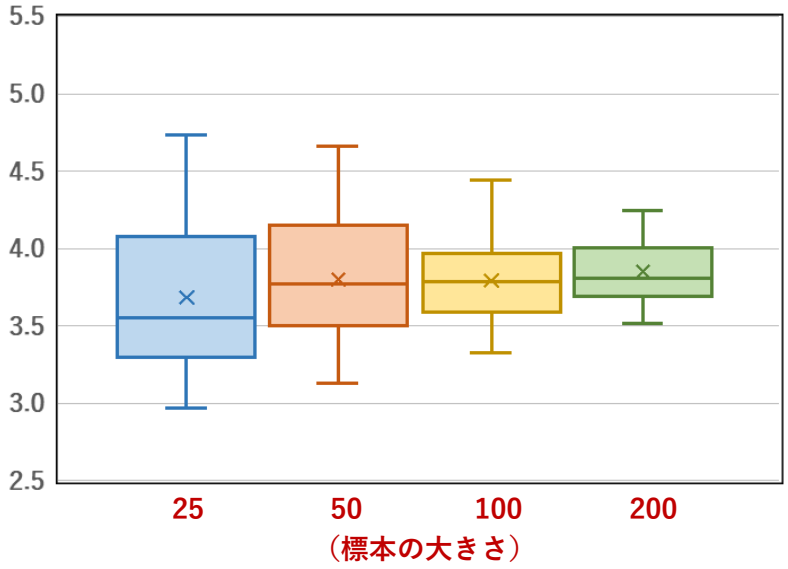


図 2 C 標本平均の分布の箱ひげ図の例（標本抽出 30 回）試行 C

◎　**図２**の箱ひげ図は、**試行Ａ**、**Ｂ**、**Ｃ**で、それぞれ形が異なりますが、

標本の大きさが大きいほど標本平均の分布範囲が狭くなる（＝四分位範囲が小さくなる）という傾向は（一部例外はありますが）、確認することができます。

表６Ａ　標本平均の分布の平均値と四分位数の例（表６の再掲）

（標本抽出 30 回）試行Ａ

	標本の大きさ【25】	標本の大きさ【50】	標本の大きさ【100】	標本の大きさ【200】
平均値	4.01	3.97	4.01	4.01
最小値	3.18	3.12	3.47	3.69
最大値	5.19	4.91	4.81	4.42
第１四分位数	3.69	3.68	3.71	3.88
中央値	3.92	4.03	4.00	4.00
第３四分位数	4.30	4.21	4.20	4.12
四分位範囲	0.60	0.53	0.49	0.24

表６Ｂ　標本平均の分布の平均値と四分位数の例

（標本抽出 30 回）試行Ｂ

	標本の大きさ【25】	標本の大きさ【50】	標本の大きさ【100】	標本の大きさ【200】
平均値	4.02	4.06	4.00	3.96
最小値	3.27	3.47	3.43	3.65
最大値	5.15	4.88	4.55	4.38
第１四分位数	3.61	3.76	3.82	3.75
中央値	4.01	4.08	4.01	3.95
第３四分位数	4.32	4.34	4.19	4.15
四分位範囲	0.71	0.58	0.37	0.40

表６Ｃ　標本平均の分布の平均値と四分位数の例

（標本抽出 30 回）試行Ｃ

	標本の大きさ【25】	標本の大きさ【50】	標本の大きさ【100】	標本の大きさ【200】
平均値	3.68	3.80	3.79	3.85
最小値	2.97	3.13	3.33	3.52
最大値	4.74	4.66	4.44	4.25
第１四分位数	3.30	3.50	3.59	3.69
中央値	3.55	3.77	3.78	3.81
第３四分位数	4.08	4.15	3.97	4.01
四分位範囲	0.78	0.65	0.38	0.31

【参考表】　表５Ｂ　標本平均の一例（標本抽出 30 回）試行Ｂ

抽出回数	標本の大きさ【25】	標本の大きさ【50】	標本の大きさ【100】	標本の大きさ【200】
1	5.01	4.33	4.00	4.14
2	3.58	3.58	3.43	3.69
3	4.03	3.79	3.86	4.12
4	4.07	3.94	3.63	3.68
5	3.66	4.11	4.20	4.38
6	3.62	3.71	3.94	4.04
7	3.99	3.76	3.57	3.82
8	3.27	3.56	3.46	3.71
9	3.70	4.49	4.36	4.17
10	4.36	3.98	3.82	3.65
11	3.69	3.74	4.21	4.25
12	4.92	4.35	4.05	3.79
13	5.12	4.23	4.42	4.08
14	3.41	3.84	3.91	3.74
15	3.27	4.05	3.91	3.75
16	3.31	3.47	3.80	3.67
17	5.15	4.88	4.15	4.17
18	3.99	4.59	4.46	4.23
19	4.29	4.40	4.15	3.88
20	4.32	4.24	4.06	3.89
21	4.47	4.20	4.15	3.81
22	3.57	3.52	3.78	3.85
23	4.25	4.47	4.19	4.15
24	4.03	4.35	4.28	4.01
25	3.54	4.29	4.55	4.24
26	3.92	3.78	3.89	3.76
27	4.06	3.93	4.01	4.00
28	4.09	4.29	3.97	4.16
29	3.74	3.62	3.54	3.75
30	4.33	4.18	4.12	4.07

表５Ｃ　標本平均の一例（標本抽出 30 回）試行Ｃ

抽出回数	標本の大きさ【25】	標本の大きさ【50】	標本の大きさ【100】	標本の大きさ【200】
1	3.85	3.51	3.33	3.57
2	3.56	3.87	3.95	3.91
3	3.54	3.93	4.18	4.15
4	3.32	3.32	3.59	3.93
5	4.18	4.42	4.44	4.25
6	3.86	4.15	3.89	3.92
7	4.13	4.27	3.94	3.97
8	3.11	3.49	3.84	3.66
9	3.51	3.54	3.81	3.78
10	4.04	4.66	4.04	4.10
11	3.51	3.13	3.64	3.78
12	3.44	3.47	3.68	3.77
13	2.97	3.41	3.74	4.00
14	3.43	3.59	3.62	4.13
15	3.87	3.71	3.76	3.80
16	3.12	3.88	3.89	4.04
17	3.39	3.44	3.43	3.81
18	4.74	4.22	4.11	4.02
19	4.26	4.17	3.90	3.79
20	3.19	3.53	3.45	3.65
21	4.04	3.90	4.02	3.97
22	3.29	3.52	3.59	3.74
23	4.12	4.08	3.89	3.76
24	4.10	3.58	3.72	3.70
25	3.27	3.83	3.66	3.68
26	3.30	3.89	3.57	3.68
27	3.82	3.58	3.45	3.52
28	4.07	4.33	4.03	3.88
29	3.11	3.35	3.51	3.54
30	4.43	4.21	4.18	4.10

【標本抽出 90 回の例】

● ここで、**試行Ａ**、**試行Ｂ**、**試行Ｃ**を合わせて**合計 90 回の標本抽出を実施**したと考えると、標本平均の分布は次のようになります。

表７　標本平均の分布の平均値と四分位数の例（標本抽出 90 回）

	標本の大きさ【25】	標本の大きさ【50】	標本の大きさ【100】	標本の大きさ【200】
平均値	3.91	3.94	3.93	3.94
最小値	2.97	3.12	3.33	3.52
最大値	5.19	4.91	4.81	4.42
第１四分位数	3.51	3.59	3.70	3.77
中央値	3.90	3.93	3.94	3.93
第３四分位数	4.20	4.22	4.15	4.10
四分位範囲	0.69	0.63	0.45	0.33

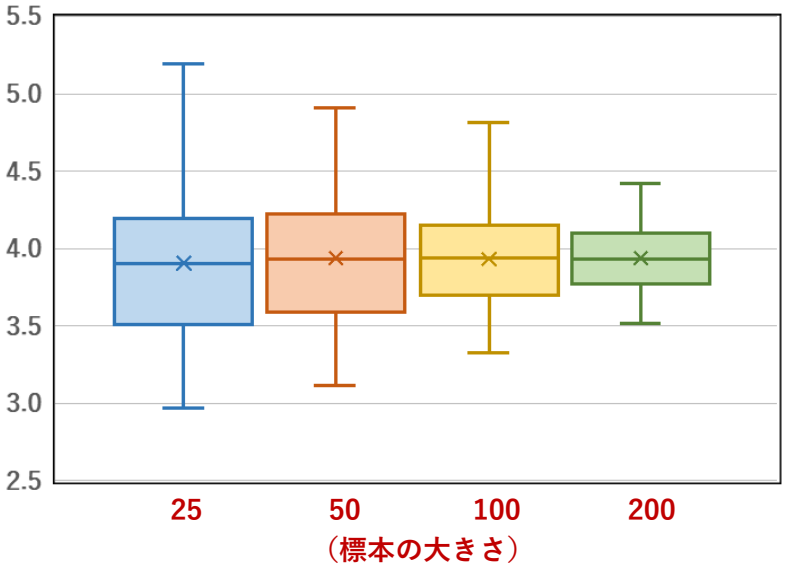


図３　標本平均の分布の箱ひげ図の例（標本抽出 90 回）

6　得られた標本平均のヒストグラムを作成する

◎　**標本抽出を 90 回実施した例**について、標本平均の度数分布表とヒストグラムは次のようになります。

- Excel の **FREQUENCY 関数**を用いて、それぞれの標本の大きさごとに度数分布表を作成。
- 得られた度数分布のデータから、棒グラフでヒストグラムを作成。

表 8　標本平均の度数分布表の例（標本抽出 90 回）

階級	標本の大きさ【25】	標本の大きさ【50】	標本の大きさ【100】	標本の大きさ【200】
≦3.0	1	0	0	0
(3.0, 3.2]	5	3	0	0
(3.2, 3.4]	9	2	1	0
(3.4, 3.6]	13	19	12	3
(3.6, 3.8]	10	11	17	25
(3.8, 4.0]	15	14	21	24
(4.0, 4.2]	15	14	25	29
(4.2, 4.4]	10	17	8	8
(4.4, 4.6]	2	7	5	1
(4.6, 4.8]	4	1	0	0
(4.8, 5.0]	2	2	1	0
>5.0	4	0	0	0

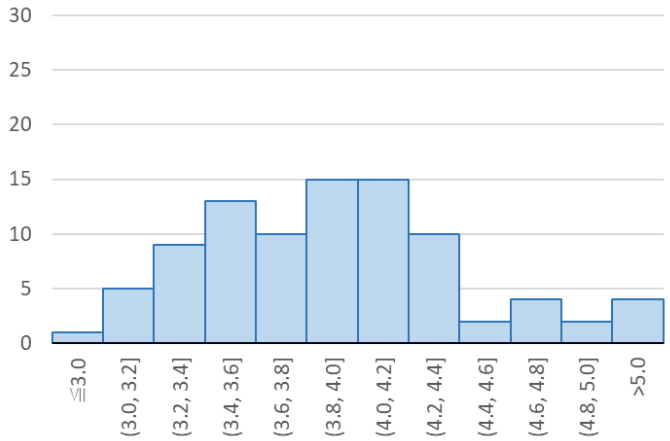


図 3 -1　標本平均の分布のヒストグラムの例
（標本の大きさ 25、標本抽出 90 回）

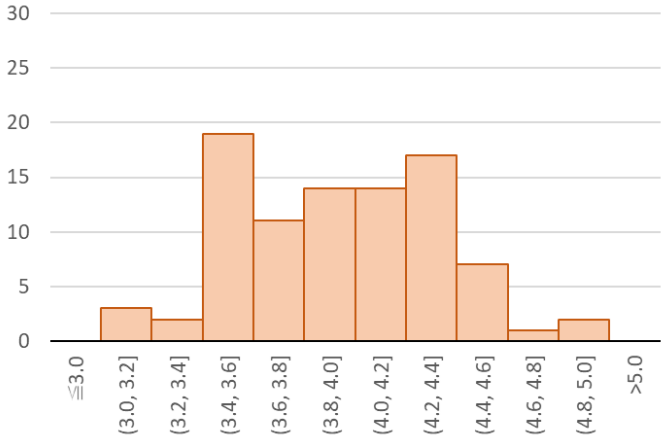


図 3 -2　標本平均の分布のヒストグラムの例
（標本の大きさ 50、標本抽出 90 回）

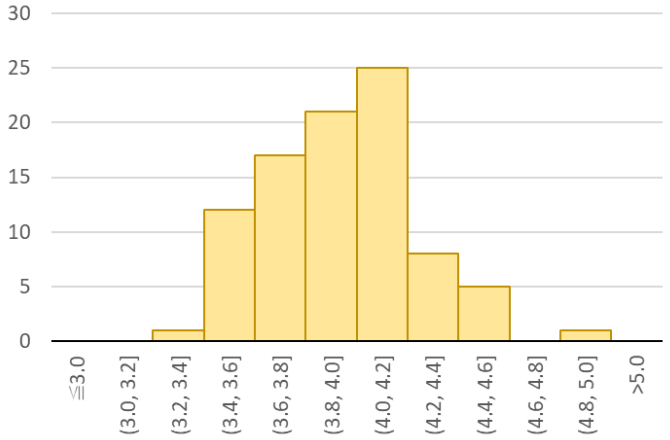


図 3 -3　標本平均の分布のヒストグラムの例
（標本の大きさ 100、標本抽出 90 回）

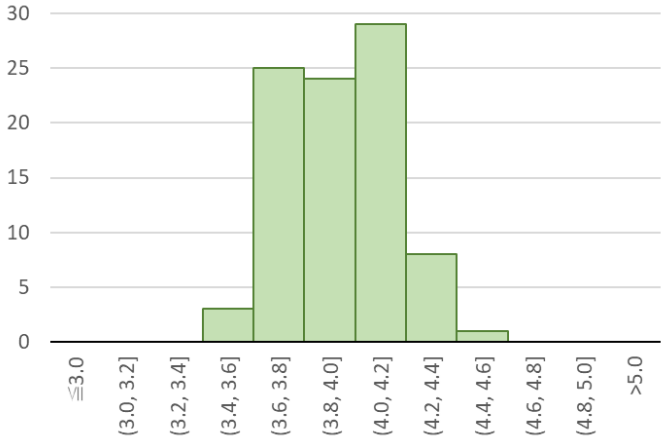


図 3 -4　標本平均の分布のヒストグラムの例
（標本の大きさ 200、標本抽出 90 回）

7　標本調査（標本抽出）のまとめ

◎　今回の分析の流れをまとめると、以下のとおりです。

表　題	内　容
1　母集団と標本（標本調査）	
2　分析用のデータセットを作成する	● 分析用のデータとして「人口千人当たり飲食店数」を計算
3　データの分布を確認する	● 分析用データの分布の状況を確認
4　標本（市区町村）を抽出し、標本平均を計算する	（標本抽出を１回行う） ● 標本の大きさを 25 として標本を抽出し、標本平均を計算 ● 標本の大きさが 50、100、200 の場合も同様に計算
5　抽出を繰り返し、得られた標本平均を分析する	（標本抽出を繰り返し行う） ● 標本の大きさ 25、50、100、200 について、それぞれ抽出を 30 回繰り返し、得られた標本平均の分布を確認 ● 抽出を 90 回繰り返した場合の、得られた標本平均の分布を確認
6　得られた標本平均のヒストグラムを作成する	

- 標本の抽出は乱数を用いるので、標本を抽出するたびに抽出される市区町村が変わり、標本平均の値も抽出ごとに変わります。
- 標本の抽出を繰り返すと、様々な値の標本平均が得られます。（例：30回繰り返すと、そのたびに異なる値が得られます。）
- 標本の大きさ 25、50、100、200 のそれぞれについて、抽出を同じ回数繰り返し、得られた標本平均の分布を確認すると、標本の大きさが大きいほど、標本平均の値のばらつきの幅（箱ひげ図の箱の幅＝四分位範囲）が小さくなる傾向がみられます。
- なお、抽出の繰返し回数を増やした場合、それぞれの標本の大きさの標本平均の値のばらつき幅には大きな差はみられませんが、標本平均のヒストグラムは滑らかな形になる傾向がみられます。（解説編を参照）

SSDSE（教育用標準データセット：Standardized Statistical Data Set for Education）は、
データ分析のための汎用素材として、独立行政法人統計センターが無償で提供しています。

https://www.nstac.go.jp/use/literacy/SSDSE/

・主要な統計を地域別に一覧できる表形式のデータセットです。
・６種類のSSDSEを公開しています。（令和８年11月現在）
・データサイエンス演習、統計教育などにご活用ください。

SSDSE-市区町村（SSDSE-A）	全国の全市区町村の、人口、経済、教育、労働、医療、福祉など、様々な分野の統計データを収録
SSDSE-県別推移（SSDSE-B）	人口、経済、教育、労働、医療、福祉など、様々な分野の統計データを、12年分の時系列で収録
SSDSE-家計消費（SSDSE-C）	１世帯当たりの食料の年間支出金額（消費額）を、魚介、肉、野菜、果物、菓子、飲料などに分類し、それぞれ詳細な品目別にデータを収録
SSDSE-社会生活（SSDSE-D）	男女別に、スポーツ・趣味・娯楽・ボランティアなどの詳細な活動データや、１日の睡眠・食事・学業・家事・仕事・趣味・娯楽などの時間配分データを収録
SSDSE-基本素材（SSDSE-E）	人口、経済、教育、文化、医療、福祉など、様々な分野の統計データを収録し、初学者にも扱いやすいデータセットです
SSDSE-気候値（SSDSE-F）	気温、気圧、風速、日照、降水量、雪など、様々な気象データについて、月・年別の平年値を収録

SSDSE を使ってみよう　― EXCEL を使った SSDSE データの使用例 ―

第５回　市区町村別の飲食店数（標本抽出の例）

公開：令和８年 11 月４日
作成：独立行政法人 統計センター　https://www.nstac.go.jp/

お問い合わせ先：　SSDSE 担当
〒162-8668　東京都新宿区若松町 19－1
統計センター 統計技術・提供部 技術研究開発課
ssdse_atmark_nstac.go.jp　（「_atmark_」を「@」に置き換えてください）