

<p>07 確率分布②（正規分布）</p>	<p>1</p>	<p>2</p>	<p>3</p>
<p>2.ハンバーガーショップの話を出そう</p> <p>状況：ポテトは「基準 150g」で盛り付けているつもりでも、毎回少しずつ重さのズレが生じる（誤差が出る）</p> <p>そこでやったこと：</p> <ul style="list-style-type: none"> ・ 30 人分（30 回分）のポテトの重さを調査してデータ化した ・ 平均・中央値・最小/最大など、基礎統計量を Excel で計算した <p>でも、疑問が残った：</p> <ul style="list-style-type: none"> ・ 「平均との差が〇g」は 大きいの？ 普通なの？ ・ 「この重さ」は よくある？ 珍しい？（判断の基準が欲しい） ・ どこからが 重すぎ/軽すぎ と言えるのか？ 	<p>狙い</p> <p>基礎統計量は出せたが「何割-何割（何割）」がでない、という課題を克服する。この疑問を解決する道具として正規分布（山の形）へつなぐ。</p>	<p>3.前回（離散）から今日（連続）へ</p> <ul style="list-style-type: none"> ・ 前回：コインを何回投げたか（10 回~10 回）離散（とびとび） ・ 今日：ポテトの重さ（149.8g, 150.2g ...）連続（つながる） <p>違いの一言：</p> <p>離散 = 「数える」／連続 = 「測る」</p> 	<p>接続</p> <p>「割数」は割数しか取らない。一月「場合」は割数でいくらかでも割かくなる。ここで「正規分布は連続データの代用」という位置づけに入る。</p>
<p>4.今日使うデータ：ポテト 30 人分（重さの調査）</p> <p>状況：同じお店、同じ「基準 150g」のポテトでも、毎回ズレが出る。</p> <ul style="list-style-type: none"> ・ データ数：30 個 ・ 見たいこと： <ul style="list-style-type: none"> ・ 150g 付近に集まる？（中心） ・ どれくらい集まる？（ばらつき） ・ 「めったにない重さ」を判断できる？（確率） 	<p>狙い</p> <p>ここで「30 人の調査データを使う」意味をして、以降は Excel 実習に自然に移れるようにする。</p>	<p>5.正規分布：一番よく出てくる「山の形」</p> <ul style="list-style-type: none"> ・ 中央が高く、左右対称（ベルカーブ） ・ 平均に近い値が出やすい ・ 平均から離れるほど出にくい <p>今日の読み替え：</p> <p>150g に近い重さが一番多い／極端に重い・軽いはずしい</p> 	<p>ポイント</p> <p>「正規分布＝一番端に集まる調査の集まり」であることを、ポテトに結びつけて言い換える。</p>
<p>8</p>	<p>9</p>	<p>10</p>	<p>11</p>
<p>6.ヒストグラムで見たこと</p> <p>これまでの授業で行ったこと</p> <ul style="list-style-type: none"> ・ 30 人分のポテトの重さデータを使って ・ ヒストグラムを作成した <p>そこから分かったこと</p> <ul style="list-style-type: none"> ・ 真ん中あたり（150g 付近）にデータが多い ・ 左右に少しずつ広がっている ・ 全体として「山のような形」になっている 	<p>狙い</p> <p>学生に「知っている内容」「やったこと」を思い起こせるスライド。ここでは評価や不足は言わず、事実の羅列だけに留める。</p>	<p>7.でも、ヒストグラムだけで判断できる？</p> <p>ここで疑問</p> <ul style="list-style-type: none"> ・ 149g のポテトは、どれくらい珍しい？ ・ 「重すぎ」「軽すぎ」と言える 基準は？ ・ そのような重さは 何%くらい起きる？ <p>ヒストグラムの役割</p> <ul style="list-style-type: none"> ・ 分布の可視化：データの散らばりや集まりを「山の形」で捉える ・ 異常の発見：全体から外れた「極端な値」をひと目で特定する ・ 背景の推測：山の数や位置から「データの裏にある事実」を探る <p>結論</p> <p>観察はできるが、判断はできない</p>	<p>狙い</p> <p>「ヒストグラムがダメ」ではなく、「疑問が湧く」という状態をする。ここで初めて「山の道」が必要だと学生に感じさせる。</p>
<p>12</p>	<p>13</p>	<p>14</p>	<p>15</p>
<p>8.次に必要な考え方は何か</p> <p>私たちが知りたいのは、単なる形ではありません。</p> <ul style="list-style-type: none"> ・ ハンバーガーショップの説明と 合っているか ・ その重さが どれくらい起こりやすいか ・ 「普通」と「外れている」を 同じ基準で 判断したい <p>そのために必要なのは：</p> <p>データの形を、判断に使える「共通の考え方」として使うこと</p> <p>このあと、その代表的な考え方として正規分布を導入します。</p>	<p>狙い</p> <p>正規分布を「突然の用語」ではなく、ヒストグラムの観察を導ける必然的な道具として登場させる。</p>	<p>9.この形は、実はよく現れる</p> <p>ヒストグラムで見た「山の形」、ポテトだけに特有のものではありません。</p> <ul style="list-style-type: none"> ・ この形は自然界に多く見られる ・ 人の身長・体重 ・ テストの点数、作業時間や誤差 ・ 機械で作られる製品の重さや精度 <p>共通点：</p> <ul style="list-style-type: none"> ・ 真ん中あたりが一番多い ・ 左右に少しずつ広がる ・ 極端な値はあまり起きない <p>このような「よくある山の形」には、正規分布 という名前がついています。</p>	<p>狙い</p> <p>正規分布を「突然の数学用語」ではなく、「よく見かける形の名称」として受け取らせる。</p>
<p>16</p>	<p>17</p>	<p>18</p>	<p>19</p>
<p>10.正規分布の事例</p> <p>自然界（身長・体重・体力→）や人間社会の現象の多くは正規分布に従うと言われている</p> <p>身長分布（正規分布を仮定）</p>  <p>極端に背の高い人や低い人は滅多にない</p>	<p>狙い</p> <p>「正規分布＝判断のための代通（＝代）」という役割づけを明確にする。ここではまだ計算には入らない。</p>	<p>11.なぜこの形が「判断」に使えるのか</p> <p>ヒストグラムは「今回集めたデータの様子」を見せてくれました。</p> <p>しかし私たちが本当に知りたいのは――</p> <ul style="list-style-type: none"> ・ この重さは、説明どおりと見えそうか ・ どの範囲までを 普通 と考えるか ・ どれくらい外れたら 外れている と言えるか <p>ポイント：</p> <ul style="list-style-type: none"> ・ ヒストグラム：今回のデータを見る道具 ・ 正規分布：多くの場合に成り立つ共通の考え方 <p>だから正規分布は「この店の説明は妥当か？」を考えるための基準として使える。</p>	<p>狙い</p> <p>「正規分布＝判断のための代通（＝代）」という役割づけを明確にする。ここではまだ計算には入らない。</p>
<p>20</p>	<p>21</p>	<p>22</p>	<p>23</p>

12.正規分布を使うと、何ができるのか

ここまでで分かったことを整理します。

- ヒストグラムは「今回のデータ」を見る道具
- 正規分布は「判断の基準」を与えてくれる考え方

正規分布を使うと：

- どの範囲を「普通」と考えるか
- どれくらい外れたら「外れている」と言えるか
- そのような重さが、どれくらい起こりにくいのか

今日はまず、「正規分布で判断するとはどういうことか」を考えます。

10/20/2019, アルゴリズム 2 / 107

12 / 25

24

14.重さを「範囲」で考えてみる

次の2つを比べてみましょう。

- 148.7g という1つの値
- 145g〜155g という 重さの範囲

私たちが本当に知りたいのは――

- このくらいの重さのボテトは多いのか？
- それともあまり出ないのか？

正規分布では「範囲」で考える

10/20/2019, アルゴリズム 2 / 107

14 / 25

28

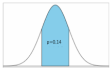
16.正規分布では「面積」が割合を表す

正規分布の山は、全体で「全部（100%）」でした。

次の一歩：

- 山の 一部の面積 = 全体の中の 割合
- 割合を「確率」と呼ぶ

高さではなく、面積が意味を持つ



10/20/2019, アルゴリズム 2 / 107

16 / 25

32

18.「普通の範囲」をどう表す？（例：150±5g）

ハンバーガーショップの説明を、範囲で書き直します。

例：普通を「150±5g」と考えるなら

- 普通の範囲：145g ～ 155g
- 知りたいのは：
- その範囲にどれくらい入るか（割合）
 - 範囲の外はどれくらい多いのか

10/20/2019, アルゴリズム 2 / 107

18 / 25

36

20.「割合」を、何と呼ぶか

ここまで私たちは、ずっと **割合** という言葉を使ってきました。

- ある範囲にどれくらい含まれるか
- 全体の中でどれくらいいる割合か

この「割合」に、統計では名前があります。

この割合を「確率」と呼ぶ

10/20/2019, アルゴリズム 2 / 107

20 / 25

40

22.なぜ「点」ではなく「範囲」で考えるのか

ここまで重要な事実があります。

- ボテトの重さは **連続データ**
- 150.0g ちょうど、という値は
- 無限に細かい数の中の1点

その結果――

1点だけの確率は 0

だから私たちは必ず「**範囲**」に含まれる**確率**を考えます。

10/20/2019, アルゴリズム 2 / 107

22 / 25

44

13.判断するとき、山のどこを見るか

ヒストグラムでは、これまで「**どこが高いか**」を見てきました。

しかし、正規分布で判断するときは**見る場所が変わります**。

- 1つの重さ（1点）を見るのではない
- 山の **一部の広がり**に注目する

判断のポイントは「高さ」ではなく「**範囲**」

10/20/2019, アルゴリズム 2 / 107

13 / 25

26

15.山の「量」を考えるという発想

正規分布の山は、全体で「**全部**」を表しています。

その山の一部を切り取ると――

- その範囲にどれくらい含まれるか
- 全体の中でどれくらいの割合か

正規分布では、「山の量」が意味を持つ

10/20/2019, アルゴリズム 2 / 107

15 / 25

30

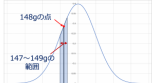
17.連続データでは「点」より「範囲」で考える

ボテトの重さは **連続**（小数までいくらでも細かい）です。

ここが重要：

- 148.0g ちょうどが出る確率は「ほぼ0」
- だから 148g 付近のように「**範囲**」で考える

連続データの確率 = 範囲の面積



10/20/2019, アルゴリズム 2 / 107

17 / 25

34

19.では、その面積（割合）をどうやって数値にする？

145g〜155g の **面積の割合**を知りたい。でも、山の面積は **定規では測れない**ので――

次に必要になる道具：

- 山の位置を「共通の物差し」に変換する（標準化）
- その結果から **面積（割合）**を読み取る方法（表・関数）

次のステップ：

145g と 155g を、「山の上下の位置」として表す

10/20/2019, アルゴリズム 2 / 107

19 / 25

38

21.なぜ、この割合を「確率」と呼べるのか

理由はシンプルです。

- 同じ条件で何度も測ると
- その範囲に入る割合は
- だいたい同じ値に落ち着く

つまり――

長い目で見たときの割合

これが、**確率**の意味です。

10/20/2019, アルゴリズム 2 / 107

21 / 25

42

23.正規分布で、何をしているのか

ここまでを語を整理します。

- 山全体：すべて（100%）
- 山の一部：その範囲に入る割合
- その割合を：確率と呼ぶ

つまり正規分布では――

「どの範囲に、どれくらい含まれるか」

を考えているだけです。

10/20/2019, アルゴリズム 2 / 107

23 / 25

46

狙い

ヒストグラム的な「山の高さ」思考から、正規分布的な「範囲」思考へ視点を切り替える。

狙い

次に出てくる「面積＝割合（確率）」を言葉なしで理解できるようにする。

範囲

確率は本来「全体に対する割合」。

ポイントとしては同数を数えて割合を出したが、ボテトの重さは連続データなので同じ方法では考えられない。

そこで正規分布では、山全体を 100 % と標準し、その中で範囲が占める割合を見る。

この「割合」を、確率と呼ぶ。

面積が広がったので数値も、割合の値が広がったのだ。

狙い

ここで「標準化」や「Z 値」を登場させない。まず「位置を読み取る必要がある」という視点を作る。次ステップで「平均と偏差」「Z 値」の導入へ自然につなげる。

狙い

確率＝未来予測ではないことを明確にする。「繰り返したときの割合」という統計的な意味づけに限定する。

狙い

正規分布を「新しい視点」ではなく、考え方の整理として理解させる。

47

この「山の形」は、特別なものではありません。

- ・誤差が少しずつ重なる
- ・小さなズレが積み重なる
- ・プラスとマイナスが混ざる

その結果、自然にこの形に近づくことが多い。
だから正規分布は**判断の基準**として使われる。

「既定している」のではなく「よく現れるから使う」という現実的な理由を与える。

今日学んだことを整理します。

- ・ヒストグラムは「観察」の道具
- ・正規分布は「判断」の基準
- ・連続データでは「範囲」で考える
- ・山の面積は「割合」＝「確率」

次回は――

この山を、どうやって数値で扱うか

第7回・第8回（ σ 、標準化、中心極限定理）への自然な接続し。