

07 確率分布②（正規分布）

1. 第6回：確率分布②（正規分布）

状況：ボテトは「基準 150g」で盛り付けているつもりでも、毎回少しづつ重さのズレが生じる（誤差が出る）

そこでやったこと：

- 30 人分（30 回分）のボ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

でも、疑問が生れた：

- 「平均との差〇(g)」は大きいの？ 普通なの？
- 「この重さは、よくある？ わい？」（判断の基準が欲しい）
- 「どこかが重すぎ」・「軽すぎ」と言えるのか？

2. ハンバーガーショップの話を思い出す

状況：ボテトは「基準 150g」で盛り付けているつもりでも、毎回少しづつ重さのズレが生じる（誤差が出る）

基礎統計量は出せたが「計算（判断）」ができない、という課題を話すと、この疑問を解決する道具として正規分布（山の形）がつながる。

3. 前回（離散）から今日（連続）へ

状況：同じ 30 人の調査データを使う。以降は Excel 実習に自然に移れるようにする。

基礎統計量（離散）と正規分布（連続）を比較する。

4. 今日使うデータ：ボテト 30 人分（重さの誤差）

状況：同じお店、同じ「基準 150g」のボテトでも、毎回ズレが生じる。

基礎統計量（離散）と正規分布（連続）を比較する。

5. 組い

6. 正規分布：一番よく出てくる「山の形」

状況：同じお店、同じ「基準 150g」のボテトでも、毎回ズレが生じる。

基礎統計量（離散）と正規分布（連続）を比較する。

7. でも、ヒストグラムだけで判断できる？

状況：同じ 30 人の調査データを使う。以降は Excel 実習に自然に移れるようにする。

ヒストグラムと正規分布曲線を比較する。

8. 次に必要な考え方は何か

状況：私たちが知りたいのは、単なる形ではなく、ヒストグラムの限界を理解するための道具として登場させる。

基礎統計量（離散）と正規分布（連続）を比較する。

9. この形は、実はよく現れる

状況：同じ 30 人の調査データを使う。以降は Excel 実習に自然に移れるようにする。

ヒストグラムと正規分布曲線を比較する。

10. 正規分布の事例

状況：自然現象（体長・体重・体力…）や人間社会の事象の多くは正規分布に従うと言われている

基礎統計量（離散）と正規分布（連続）を比較する。

11. なぜこの形が「判断」に使えるのか

状況：同じ 30 人の調査データを使う。以降は Excel 実習に自然に移れるようにする。

ヒストグラムと正規分布曲線を比較する。

12. 組い

13. 組い

14. 組い

15. 組い

16. 組い

17. 組い

18. 組い

19. 組い

20. 組い

21. 組い

22. 組い

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、どれくらいに起こりにくいか

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。

502020. アルゴリズム 2.00

12 / 25

24

25

26

27

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つの重さ
- 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

502020. アルゴリズム 2.00

13 / 25

28

29

30

31

16. 正規分布では「範囲」が割合を表す

正規分布の山は、全体で「全部（100%）」でした。

次の一步：

- 山の一部の面積 = 全体の中での 割合
- 割合を「確率」と呼ぶ

高さではなく、面積が意味を持つ



502020. アルゴリズム 2.00

14 / 25

32

33

34

35

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

- 普段の範囲：145g ~ 155g
- 知りたいのは——
- その範囲にどれくらいいるか（割合）
- 範囲の外はどれくらい珍しいか

502020. アルゴリズム 2.00

15 / 25

36

37

38

39

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いるか
- 全体の中でどれくらいの割合か

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

502020. アルゴリズム 2.00

20 / 25

40

41

42

43

22. なぜ「点」ではなく「範囲」で考えるのか

ここで重要な事があります。

- ボットの重さは「連続データ」
- 150.0g ちょうど、という重さは
- 無理に細かい数の中の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。

502020. アルゴリズム 2.00

21 / 25

44

45

46

47

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異な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」

狙い

ヒストグラム的な「神の高さ」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「山積=割合（確率）」を直感なしで理解できる土台を作る。

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——

- その範囲にどれくらいいるか
- 全体の中でどのくらいの割合か

正規分布では、「山の量」が意味を持つ

狙い

次に出てくる「山積=割合（確率）」を直感なしで理解できる土台を作る。

狙い

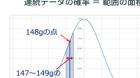
1点の判断が難しいことを解消させ、「範囲」という考え方を自然に受け入れさせる。

17. 連続データでは「点」より「範囲」で考える

ボトの重さは連続（小数までいくらでも細かい）です。

ここが重要：

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



502020. アルゴリズム 2.00

範囲

範囲は本来「全個体に対する割合」。
コインでも100個を数えて割合を出したが、ボトの重さは連続データなので同じ方法は使えない。
そこで正規分布では、山全体を100%と割り切る。その中で範囲が右側の割合を見る。

この「割合」を「確率」と呼ぶ。
意味が変わったのではなく、割合の振り方が変わっただけ。

狙い

「普通=範囲」をここでも細めで扱い切る。前スライドで「山の発見が入っているので、違和感なく接続できる」。

19. では、その面積（割合）をどうやって数値にする？

165g~155gの面積の割合を知りたい。でも、山の面積は定積では測れないのでは——

次の必要な道具：

- 山の位置を「範囲の物差し」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）



502020. アルゴリズム 2.00

狙い

ここで「標準化」や「2重」を理解しないと、必ず「面積を汎用化する必要がある」という必然を作る。次スライドで「平均」と「標準偏差」は「2重」を理解しないと、「面積を汎用化する必要がある」という必然を作る。

狙い

「普通=範囲」をここでも細めで扱い切る。前スライドで「山の発見が入っているので、違和感なく接続できる」。

21. なぜ、この割合を「確率」と呼ぶのか

理由はシンプルです。

- 同じ条件で何度も測る
- その範囲に入る割合は
- だいたい同じじ幅に落ちる

つまり——

長い目で見たときの割合

これが、確率の意味です。

502020. アルゴリズム 2.00

狙い

範囲=未だ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固定する。

23. 正規分布で、何をしているのか

ここまで話を整理します。

- 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、どれくらいいるか」

を考えているだけです。

502020. アルゴリズム 2.00

狙い

正規分布を「難しい式」ではなく、考え方の整理として理解させる。

狙い

「普通=範囲」をここでも細めで扱い切る。前スライドで「山の発見が入っているので、違和感なく接続できる」。

22. なぜ「点」ではなく「範囲」で考えるのか

ここで重要な事があります。

- ボットの重さは「連続データ」
- 150.0gちょうど、という重さは
- 無理に細かい数の中の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。

502020. アルゴリズム 2.00

22 / 25

44

45

46

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。

第7回・第8回（ σ 、標準化、中心極限定理）への自然な接続し。

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