

## 07 確率分布②（正規分布）

---



## 1. 第6回：確率分布②（正規分布）

- ・ 今日の主役：正規分布（ベルカーブ）
- ・ キーワード：平均（中心）とばらつき
- ・ 重要ルール：確率は「高さ」ではなく「面積」

題材：ハンバーガーショップのポテト

基準値：150g（盛り付けは毎回ズレる）

# 狙い

---

ポテトを“誤差が自然に出る連続データ”として導入し、  
平均だけでは判断できない問題を提示する。  
ばらつきをどう扱うかは、この後で新しい道具として導入する。

## 2. ハンバーガーショップの話を思い出そう

**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）

**そこでやったこと：**

- ・ 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

**でも、疑問が残った：**

- ・ 「平均との差が○ 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？

# 狙い

---

基礎統計量は出せたが「位置づけ（判断）」ができない、という課題を言語化する。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。

### 3. 前回（離散）から今日（連続）へ

- ・ 前回：コインを何回投げたか（0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）

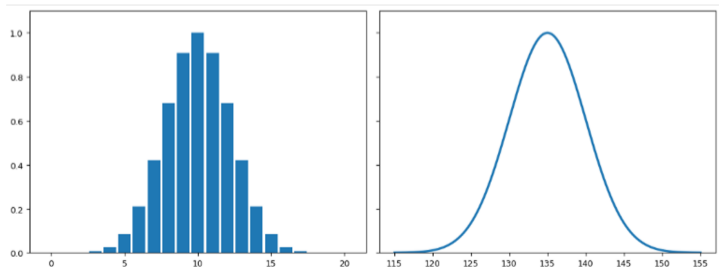
違いの一言：

離散＝「数える」／ 連続＝「測る」

離散（ギザギザ）

↔

連続（なめらか）



# 接続

---

“回数”は整数しか取らない。一方“重さ”は小数でいくらでも細かくなる。ここで「正規分布は連続データの代表」という位置づけに入る。

## 4. 今日使うデータ：ポテト 30 人分（重さの誤差）

状況：同じお店、同じ「基準 150g」のポテトでも、毎回ズレが出る。

- ・ データ数：30 個
- ・ 見たいこと：
  - ・ 150g 付近に集まる？（中心）
  - ・ どれくらい散らばる？（ばらつき）
  - ・ 「めったにない重さ」を判断できる？（確率）

# 狙い

---

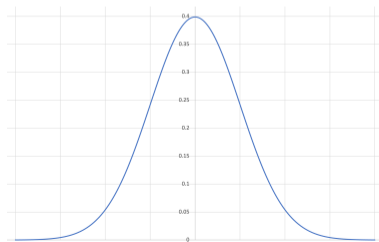
ここで“30 人の調査データを使う”宣言をして、以降は Excel 実習に自然に移れるようにする。

## 5. 正規分布：一番よく出てくる「山の形」

- ・ 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

今日の読み替え：

150g に近い重さが一番多い／極端に重い・軽いは珍しい



# ポイント

---

「正規分布＝普通に起きる誤差の集まり」であることを、ポテトに結びつけて言い切る。

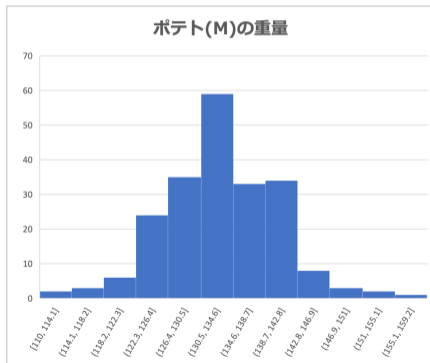
## 6. ヒストグラムで見たこと

### これまでの授業で行ったこと

- ・ 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

### そこから分かったこと

- ・ 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



# 狙い

---

学生に「知っている内容」「やったこと」を思い出させるスライド。ここでは評価や不足は言わず、事実の確認だけに徹する。

## 7. でも、ヒストグラムだけで判断できる？

### ここで疑問

- ・ 148g 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？

### ヒストグラムの役割

- ・ 分布の可視化：データの散らばりや偏りを「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

### 結論

観察はできるが、判断はできない

# 狙い

---

「ヒストグラムがダメ」ではなく、「役割が違う」という整理をする。ここで初めて「次の道具」が必要だと学生に感じさせる。

## 8. 次に必要な考え方は何か

私たちが知りたいのは、単なる形ではありません。

- ・ 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

そのために必要なのは：

データの形を、判断に使える「共通の考え方」として扱うこと

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。

# 狙い

---

正規分布を「突然の用語」ではなく、ヒストグラムの限界を埋める必然的な道具として登場させる。

## 9. この形は、実はよく現れる

ヒストグラムで見た「山の形」は、ポテトだけに特有のものではありません。

- ・ 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

**共通点：**

- ・ 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

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。

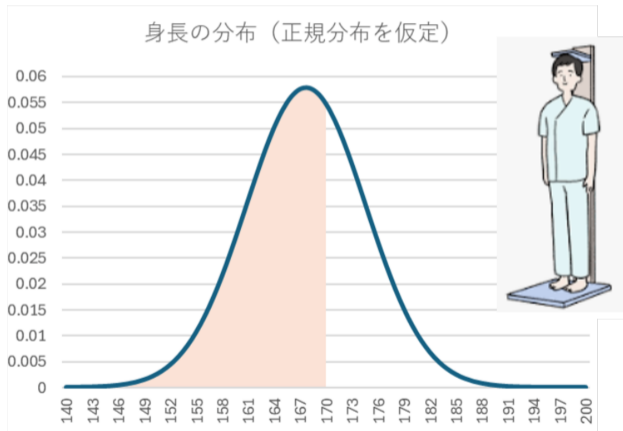
# 狙い

---

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前」として受け止めさせる。

## 10. 正規分布の事例

自然界（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



極端に背の高い人や低い人は滅多にいない



## 11. なぜこの形が「判断」に使えるのか

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。

しかし私たちが本当に知りたいのは——

- ・ この重さは、説明どおりと言えそうか
- ・ 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

ポイント：

- ・ 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

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。

# 狙い

---

「正規分布＝判断のための共通ルール」という位置づけを明確にする。ここではまだ計算には入らない。

## 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。



## 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に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」

# 狙い

---

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

## 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ポテトは多いのか？
- ・ それともあまり出ないのか？

正規分布では「範囲」で考える

# 狙い

---

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

## 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と——

- ・ その範囲に どれくらい含まれるか
- ・ 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「面積＝割合（確率）」を言葉なしで理解できる土台を作る。

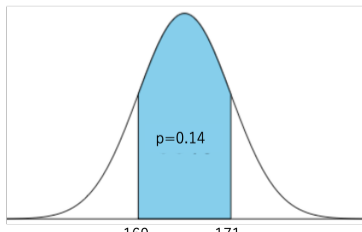
## 16. 正規分布では「面積」が割合を表す

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



# 狙い

---

「確率＝面積」をここで初めて言い切る。前スライド1で「量」の発想が入っているので、違和感なく接続できる。

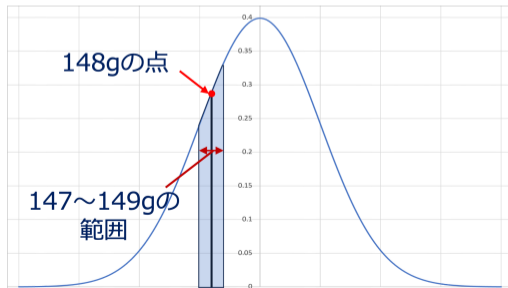
## 17. 連続データでは「点」より「範囲」で考える

ポテトの重さは **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



# 範囲

---

確率は本来「全体に対する割合」。

コイントスでは回数を数えて割合を出したが、ポテ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 % と約束し、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。

意味が変わったのではなく、割合の測り方が変わっただけ。

## 18.「普通の範囲」をどう表す？（例： $150 \pm 5\text{g}$ ）

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「 $150 \pm 5\text{g}$ 」と考えるなら

- ・ 普通の範囲： $145\text{g} \sim 155\text{g}$
- ・ 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 どれくらい珍しいか

# 狙い

---

「普通＝範囲」という言葉の形を確定させる。ここではまだ計算しない。次で“面積をどう求める？”に繋ぐ。

## 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、  
「山の上での位置」として表す

# 狙い

---

ここで「標準偏差」や「 $z$  値」を突然出さない。まず「位置を共通化する必要がある」という必然を作る。次スライドで「平均との差」「ばらつき（尺度）」の導入へ自然につなげる。

## 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** 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**

# 狙い

---

ここで初めて「確率」という言葉を正式に導入する。新しい考え方ではなく、これまで使ってきた割合に名前をつけただけだと強調する。

## 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。

# 狙い

---

確率＝未来予言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。

## 22. なぜ「点」ではなく「範囲」で考えるのか

ここで重要な事実があります。

- ・ ポテトの重さは **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。

# 狙い

---

「確率が 0」という表現で混乱させない。連続量では点に意味がない、という考え方だけを伝える。

## 23. 正規分布で、何をしているのか

ここまでの話を整理します。

- ・ 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、どれくらい含まれるか」

を考えているだけです。

# 狙い

---

正規分布を「難しい数式」ではなく、考え方の整理として理解させる。

## 24. なぜ、この形が判断の基準になるのか

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。

# 狙い

---

「仮定している」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。

## 25. 本日のまとめ

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

# 狙い

---

第 7 回・第 8 回（ $\sigma$ 、標準化、中心極限定理）への自然な橋渡し。