

平成27年度数学科リレー講座

「統計入門」

～主観と客観の狭間に屹立する
尖塔から見える風景～

2日目「平均と確率」

縫部拓也

今日の内容

- ・ 平均とは
 - ・ 中央値と最頻値
 - ・ 分散と標準偏差
 - ・ 確率とは
 - ・ 確率変数と期待値
 - ・ 連続型確率変数
- 
- Several white lines of varying lengths and slopes are positioned in the bottom right corner of the slide, creating a modern, abstract design element.

1. 平均とは

Q.あるA社とB社の社員の平均月収をそれぞれ6人に聞いて調べました。

A社は28万でB社は30万でした。仕事内容は全く同じです。

みなさんならB社の方を志望しますか？

定義

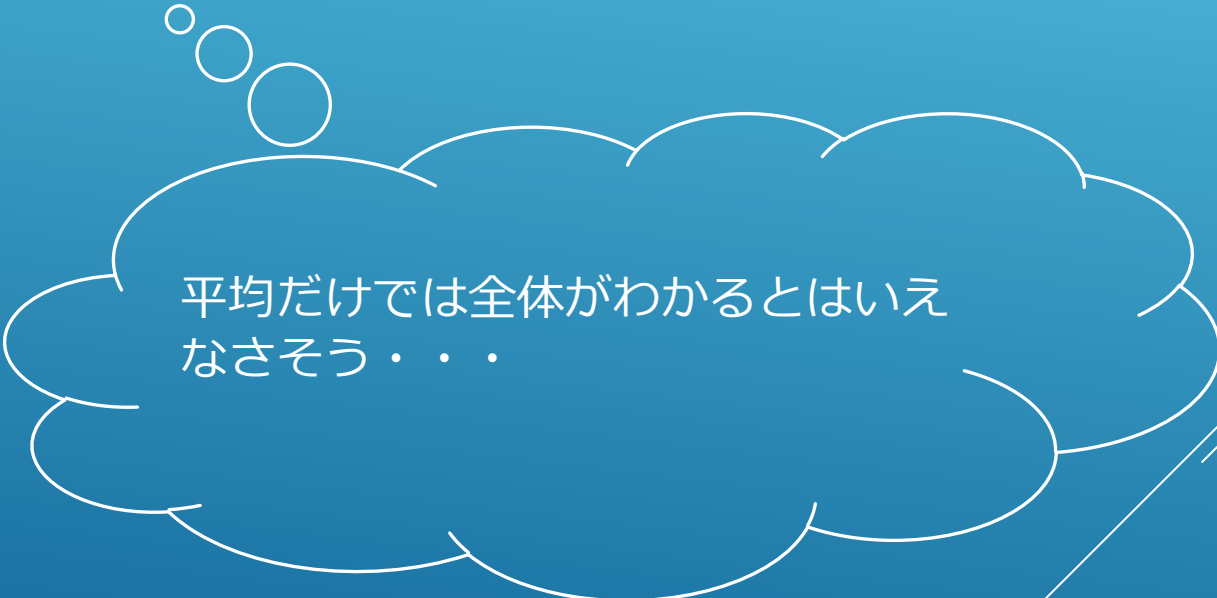
n 個のデータの値を $X_1, X_2, \dots, X_n$ とすると、平均 $\bar{X}$ は以下で与えられる。

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

Q.
あるA社とB社の平均月収をそれぞれ6人に聞いて調べるとA社は28万でB社は30万でした。
仕事内容は全く同じです。

A社 33 25 35 25 24 26

B社 10 10 10 10 130 10



平均だけでは全体がわかるとはいえない
なさそう・・・

2. 中央値と最頻値

▶ 定義

資料の値を大きさ順に並べたとき、その中央の値を **中央値** という。

ただし、資料の個数が偶数の場合は中央に並ぶ2つの値の平均をとるものとする。

▶ 定義

資料の値の中で最も頻繁に現れる値を **最頻値** という。

例 3) 15 人の小テストの結果

4 10 2 3 4 4 2 10 6 5 1 8 7 6 9

点数が低い方から順に並べると

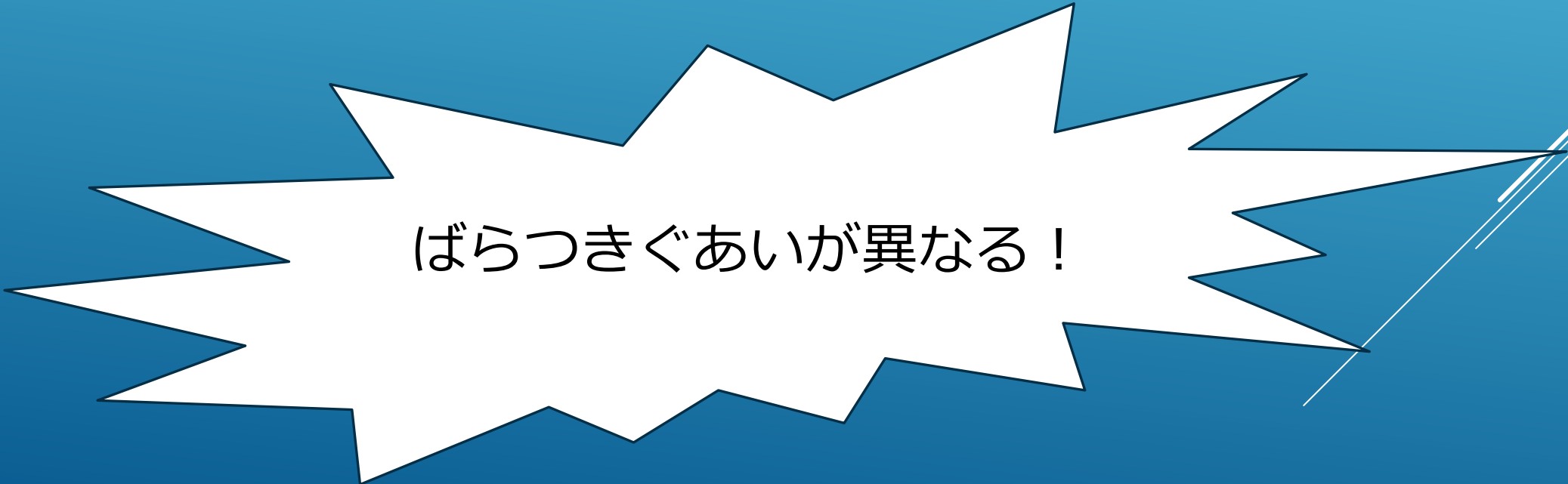
1 2 2 3 4 4 4 5 6 6 7 8 9 10 10

真ん中は8番目で、中央値は5点
最頻値は4点

3. 分散と標準偏差

C社	10	15	30	30	45	50
D社	25	28	30	30	32	35

2社とも平均月収は30万、中央値30万、最頻値30万



ばらつきぐあい異なる！

▶ 定義

n 個のデータの値を $X_1, X_2, \dots, X_n$ とし、平均を $\bar{X}$ とする。分散 $V(X)$ は次で与えられる。

$$V(X) = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{n}$$



▶ 定義

分散 $V(X)$ が与えられているとき、標準偏差 s は次で定義される。

$$s = \sqrt{V(X)}$$

補足

$\sqrt{\quad}$ （ルート、根号）に関して

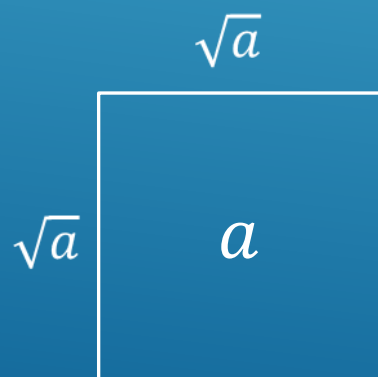
$x^2 = a$ をみたす x を a の平方根とよぶ。

例えば9の平方根は 3 と -3 である。

では10の平方根は・・・これが $\sqrt{10}$ と $-\sqrt{10}$ なのである。

実際、 $\sqrt{10}=3.162277\dots$ となっている。

別の見方をすれば、 $\sqrt{a}$ は面積 a の正方形の一辺の長さともみることができる。



に与えられているとき、次の問いに答えよ。

X : 10, 12, 7, 6, 13, 10, 7, 15

- (1) X の平均値を求めよ。
- ▶ (2) X の分散を求めよ。
- (3) X の標準偏差を求めよ。

解

$$(1) \frac{10+12+7+6+13+10+7+15}{8} = 10$$

$$(2) \frac{(10-10)^2+(12-10)^2+(7-10)^2+(6-10)^2+(13-10)^2+(10-10)^2+(7-10)^2+(15-10)^2}{8} = 9$$

$$(3) \sqrt{9} = 3$$

(参考)

おなじみの「(学力) 偏差値」はここで出てくる標準偏差とは異なるものです。

▶ 次のような式で求めることができます。

$$\text{(学力) 偏差値} = \frac{\text{(自分の得点)} - \text{(全体の平均)}}{\text{全体の標準偏差}} \times 10 + 50$$

このように計算すると実は偏差値の分布は平均値 50、標準偏差 10 になる。

4. 確率とは

- ▶ 試行・・・コインやサイコロを投げる場合や、カードやくじを引く場合のように、同じ条件で何度も繰り返すことができる実験や観測。
- ▶ 事象・・・試行の結果起こる事柄。
- ▶ 数学的確率：100本のくじのうち2本が当たりの時、当たる確率は $1/50$
- ▶ 統計的確率：何度も試行を行ったときの事象が起こった割合

例 4) 統計的確率と数学的確率

- ・サイコロを振るとき 1 の目が出る確率は、目は1～6が出得るから $1/6$ である. (数学的確率)
- ・サイコロを10回振ったら1の目は1回しか出なかったので $1/10$ (統計的確率)

2つの試行が互いに他方に影響を及ぼさないとき、これら2つの試行は**独立**という。

○大数の（弱）法則

↑回の試行でAが起こる確率を p とする。

n 個の独立試行でAが起こる回数を r とするとき

n が大きくなるにしたがって、 $\frac{r}{n}$ が確率 p に近づく。

～実験～

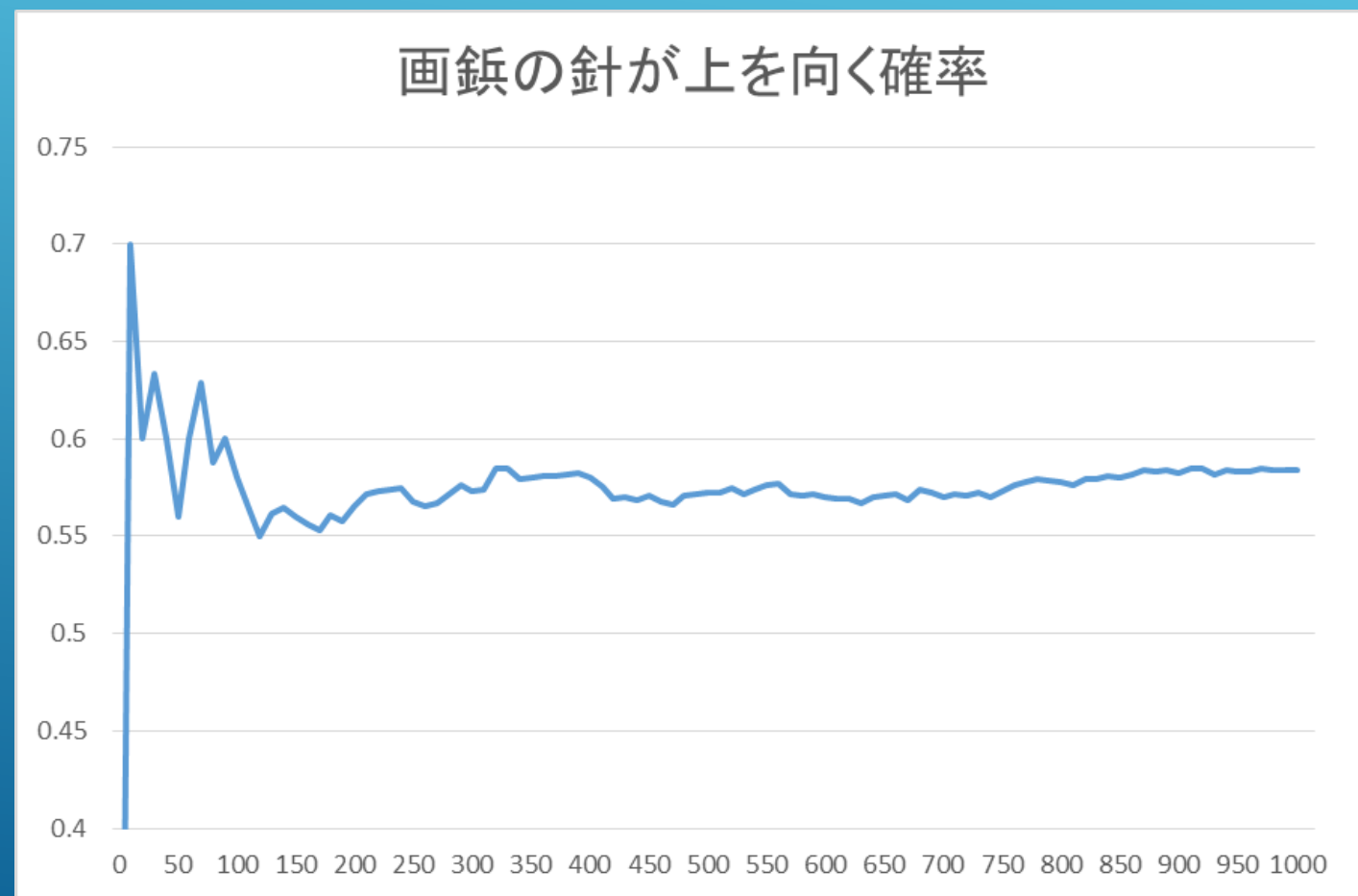
画鋲の針が上を向く確率はいくらか。

実際に1000回投げて統計的確率を調べてみると...



合計回数	針が上を向いた数	確率
10	7	0.7
100	58	0.58
200	113	0.565
⋮	⋮	⋮
800	462	0.5775
900	542	0.5822
1000	584	0.584

(ちなみに指に刺さった回数は6回...)



- ▶ 『全く区別のつかないサイコロ2個を同時に投げるとき目の和が偶数となる確率を求めよ。』という問題にAさんは
 - ▶ $\frac{18}{36} = \frac{1}{2}$, Bさんは $\frac{12}{21} = \frac{4}{7}$, Cさんは $\frac{6}{11}$ と答えました。
- どの人が正しい？

(Aさんの考え)

(1,1)(1,2)(1,3)(1,4)(1,5)(1,6)(2,1)(2,2)(2,3)(2,4)(2,5)(2,6)
(3,1)(3,2)(3,3)(3,4)(3,5)(3,6)(4,1)(4,2)(4,3)(4,4)(4,5)(4,6)
(5,1)(5,2)(5,3)(5,4)(5,5)(5,6)(6,1)(6,2)(6,3)(6,4)(6,5)(6,6)

(Bさんの考え)

(1,1)(1,2)(1,3)(1,4)(1,5)(1,6)
(2,2)(2,3)(2,4)(2,5)(2,6)
(3,3)(3,4)(3,5)(3,6)
(4,4)(4,5)(4,6)
(5,5)(5,6)
(6,6)

(Cさんの考え)

2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

正しいのは・・・Aさん

Bさんの考え方では(1,3)は(1,1)より2倍起こりやすい。

Cさんの考え方では和が6となるときは2となるときの5倍起こりやすい。

確率は起こりやすさを表す数なので1つ1つの要素の起こりやすさが異なっていてはダメ。

「どの根元事象が起きても同程度期待できる」といえるか、

▶ どの根元事象が起こることも同程度期待できるとき、これらの根元事象は**同様に確からしい**という。

▶ 定義

▶ 全事象 U に属する根元事象の個数を $n(U)$

事象 A に属する根元事象の個数を $n(A)$ とする。

どの根元事象も同様に確からしい試行において、事象 A の確率は次で与えられる。

$$P(A) = \frac{n(A)}{n(U)}$$

5. 確率変数と期待値

サイコロを1個投げるとき出る目を X とすると

X	1	2	3	4	5	6	計
確率 P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

上の例では、 X は1,2,3,4,5,6のいずれかの値をとる変数で、 X の値によって確率が定まる。

このような X を **確率変数** とよぶ。

○表記として

$$P(X = 1) = \frac{1}{6} \left(X = 1 \text{ となる確率は } \frac{1}{6} \right)$$

▶
$$P(X = 2) = \frac{1}{6} \left(X = 2 \text{ となる確率は } \frac{1}{6} \right)$$

$$P(2 \leq X \leq 6) = \frac{5}{6} \left(X \text{ が } 2 \text{ 以上 } 6 \text{ 以下 となる確率は } \frac{5}{6} \right)$$

他の例としては、サイコロ2個投げるとき出る目の和を X とすると

X	2	3	4	5	6	7	8	9	10	11	12	計
P	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	1

このような X と P の対応関係を、 X の **確率分布** とよぶ。

Q.参加すると、くじを一本引ける。10本のくじがあり3本が当たりだという。

当たりが出ると500円がもらえて、はずれると100円もらえる。

参加費100円だったらもちろん参加しますね。でも参加費500円だったら参加しないはず。では、参加費250円だったら・・・みなさんはどうしますか？？

▶ 定義

確率変数 X が次の表に表された分布に従うとする。

X	x_1	x_2	$\dots$	x_n	計
P	p_1	p_2	$\dots$	p_n	1

このとき、

$$x_1p_1 + x_2p_2 + \dots + x_np_n$$

を確率変数 X の**期待値**または**平均**といい $E(X)$ または m で表す。

6.連続型確率変数

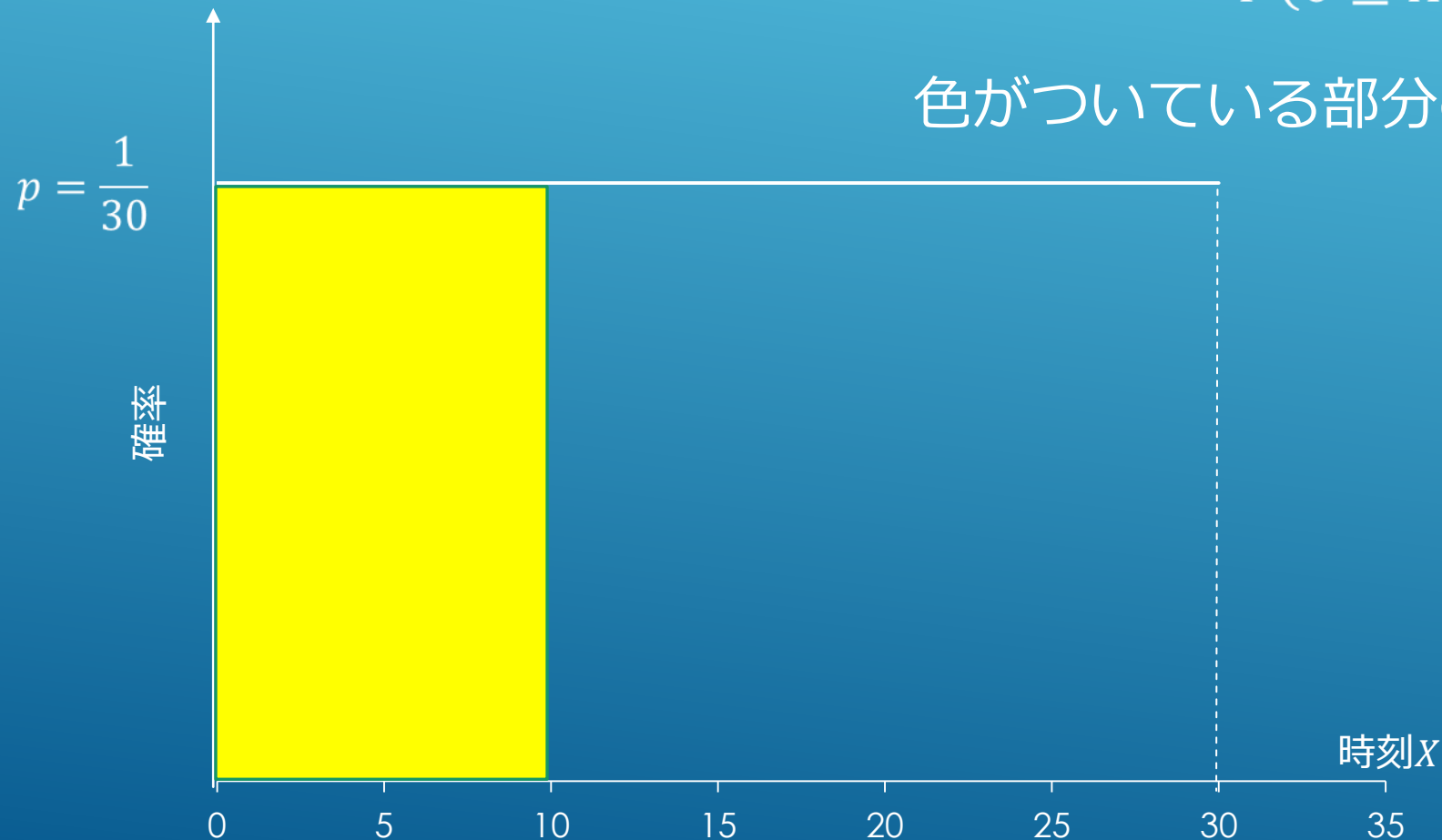
- ▶ Q.3 0分ごとに発車するバスに乗る。ある時刻にバス停に着いたとして10分以内にバスが発車する確率はいくらか。



確率変数 X は時刻である。
これはとびとびの値ではなく連続的に変化する。
このような確率変数を**連続型確率変数**という。

$$P(0 \leq X \leq 10) = 10 \times \frac{1}{30} = \frac{1}{3}$$

色がついている部分の面積として確率が求まる！



ご清聴ありがとうございました！
明日は相関についてです。お楽しみに。

