Lab 09

Fundamental Statistics

Applied Statistics and Experiments

October, 2024

Agenda

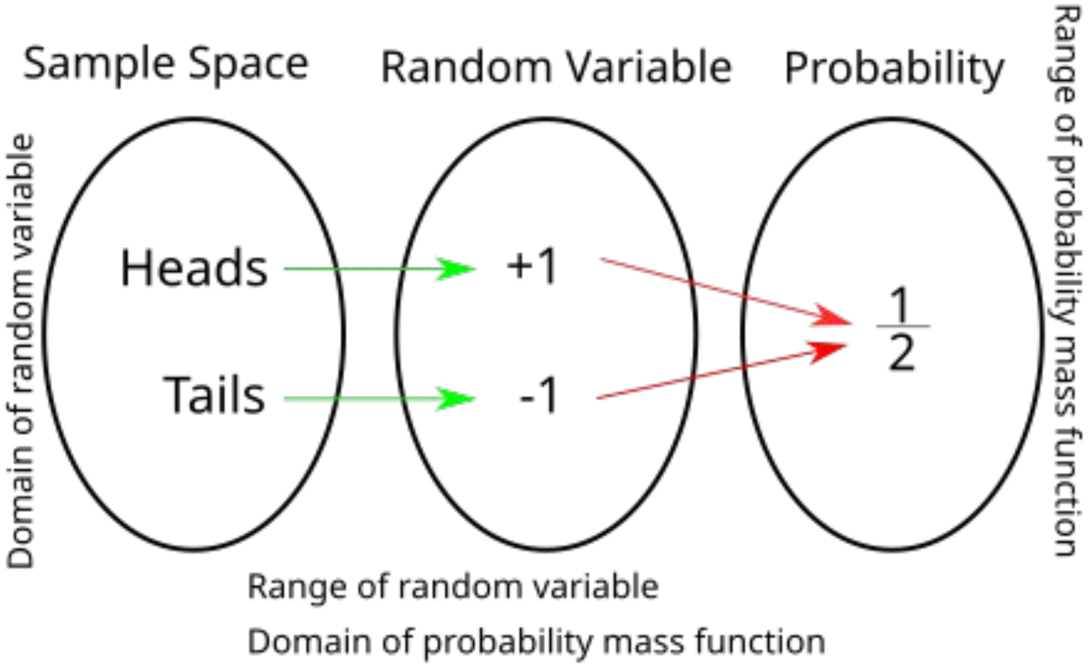
1. Discrete random variables
2. Continuous random variables
3. Descriptive statistics
4. Z-score

Lecture Recap

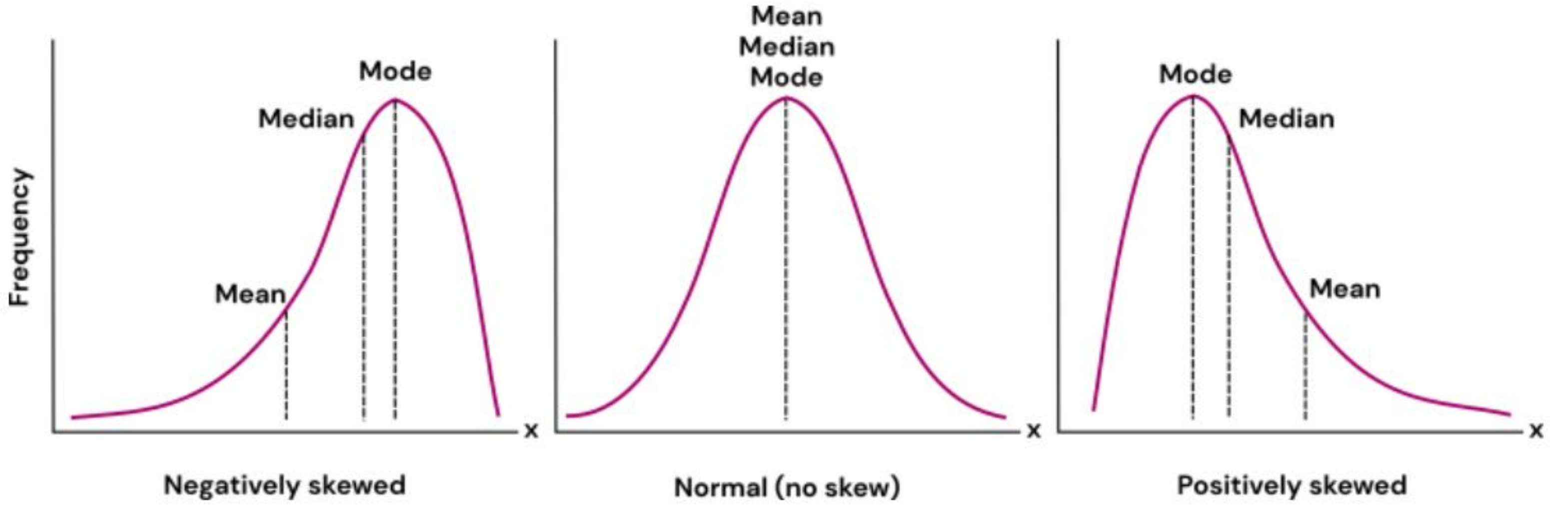
https://quizizz.com/join

Join and enter
game code

Random variable



Mean - Median - Mode



The formula for variance uses squares. Therefore, the variance has different units

Variance - Standard deviation

f(x) is the probability mass function

Discrete random variables

- the ***mean*** $E(x) = \mu = \sum x \cdot f(x)$
- the ***mode*** *the peak (or local maximum) of f(x)*
- the ***variance*** $\mathrm{Var(x)} = \sum (x-\mu)^2 \cdot f(x) = E(x^2) - \mu^2$
- the ***standard deviation*** $\sigma = \sqrt{Var(X)}$

Discrete random variables – Median

$$
\begin{aligned}
&\textit{Given a discrete random variable } X,\ \textit{the median is that value } m \textit{ for which}:\\
&P(X < m) \leqslant 0.5\\
&\textit{and}\\
&P(X \leqslant m)\ \geqslant 0.5\\
\\
&\textit{We have two cases}:\\
\\
&\textit{Case } 1\ (\textit{Unique median}):\\
&\textit{If } P(X < x) < 0.5 \textit{ then there is only one median which is } m = x.\\
\\
&\textit{Case } 2\ (\textit{multiple medians}):\\
&\textit{If } P(X < x_2) = 0.5 \textit{ then the previous value } x_1 < x_2 \textit{ holds } P(X \leqslant x_1) = 0.5\\
&\textit{and the median is any number between } x_1 \textit{ and } x_2.\ \textit{For convenience, we choose}\\
&\textit{the midpoint of the interval } [x_1,\ x_2] \textit{ which is } m = \frac{x_1 + x_2}{2}
\end{aligned}
$$

Source: https://www.stat.pitt.edu/stoffer/tsa4/intro_prob.pdf

Exercise 1

A discrete random variable X has the following probability distribution table:

x	0	2	4	6
$P(X = x)$	$\frac{1}{7}$	$\frac{3}{7}$	$\frac{1}{7}$	$\frac{2}{7}$

i. State the mode.

ii. Calculate this discrete random variable's mean value.

Exercise 1

A discrete random variable X has the following probability distribution table:

x	0	2	4	6
$P(X = x)$	$\frac{1}{7}$	$\frac{3}{7}$	$\frac{1}{7}$	$\frac{2}{7}$

i. State the mode.

ii. Calculate this discrete random variable's mean value.

$$i.\ mode = argmax_x f(x) = 2 \left(the\ value\ x\ at\ which\ f\ takes\ its\ maximum\ value\ max(f) = \frac{3}{7} \right)$$

$$ii.\ E(X) = \mu = \sum x * f(x) = 0 \cdot \frac{1}{7} + 2 \cdot \frac{3}{7} + 4 \cdot \frac{1}{7} + 6 \cdot \frac{2}{7} = 3.14$$

Exercise 2

A discrete random variable X has probability distribution table:

x	1	2	3	4	5
$P(X=x)$	$\frac{2}{10}$	$\frac{2}{10}$	k	$\frac{3}{10}$	$\frac{1}{10}$

i. Find the value of k.

ii. State the mode of X.

iii. Calculate the expected value $E(X)$.

Exercise 2

A discrete random variable X has probability distribution table:

x	1	2	3	4	5
$P(X=x)$	$\frac{2}{10}$	$\frac{2}{10}$	k	$\frac{3}{10}$	$\frac{1}{10}$

i. Find the value of k.

ii. State the mode of X.

iii. Calculate the expected value $E(X)$.

$$i.\ \frac{2}{10}+\frac{2}{10}+k+\frac{3}{10}+\frac{1}{10}=1 \Longrightarrow k=1-\frac{2+2+3+1}{10}=\frac{2}{10}=0.2$$

$$ii.\ mode = argmax_x f(x) = 4 \left(the\ value\ x\ at\ which\ f\ takes\ its\ maximum\ value\ max(f)=\frac{3}{10} \right)$$

$$iii.\ E(X)=1*\frac{2}{10}+2*\frac{2}{10}+3*\frac{2}{10}+4*\frac{3}{10}+5*\frac{1}{10}=\frac{29}{10}=2.9$$

Exercise 3

Find the variance and standard deviation for the following data:

57, 64, 43, 67, 49, 59, 61, 59, 44, 47

Exercise 3

Find the variance and standard deviation for the following data:

57, 64, 43, 67, 49, 59, 61, 59, 44, 47

$$\text{The mean}: \ \mu = \frac{57+64+43+67+49+59+61+59+44+47}{10} = \frac{550}{10} = 55$$

$$\text{The variance}: \ \sigma^2 = \frac{\sum(x-\mu)^2}{n} = \frac{2^2+9^2+(-12)^2+12^2+(-6)^2+4^2+6^2+4^2+(-11)^2+(-8)^2}{10} = \frac{662}{10} = 66.2$$

$$\text{The standard deviation}: \sigma = \sqrt{\sigma^2} = \sqrt{66.2} \simeq 8.13$$

Exercise 4

Given the discrete random variable X, with probability distribution function:

$$f(x) = x^2/90, \text{ where } x=\{2, 3, 4, 5, 6\}$$

1. Construct a probability distribution table for X.
2. State the mode of X.
3. Calculate the mean value of X.
4. Calculate standard deviation.
5. Construct a cumulative distribution table for X.
6. Find the median value of X.

Exercise 4

1. *Probility distribution function*

x	2	3	4	5	6
$f(x) = P(X = x)$	$\frac{4}{90}$	$\frac{9}{90}$	$\frac{16}{90}$	$\frac{25}{90}$	$\frac{36}{90}$

2. $mode = 6$

3. $E(X) = \mu = \frac{440}{90} = 4.89$

4. $\sigma^2 = (2-4.89)^2 * \frac{4}{90} + (3-4.89)^2 * \frac{9}{90} + (4-4.89)^2 * \frac{16}{90} + (5-4.89)^2 * \frac{25}{90} + (6-4.89)^2 * \frac{36}{90}$

$= 1.36543\bar{3}$

$\sigma = \sqrt{1.365433} = 1.17$

5. *Cumulative distribution table*

x	2	3	4	5	6
$F(x) = P(X \leqslant x)$	$\frac{4}{90}$	$\frac{13}{90}$	$\frac{29}{90}$	$\frac{54}{90}$	$\frac{90}{90}$

6. $Median = 5$

$P(X < 5) = \frac{29}{90} = 0.32 \leqslant 0.5$ *and* $P(X \leqslant 5) = \frac{54}{90} = 0.6 \geqslant 0.5$ (5 *is a unique median*)

$P(X < 4) = \frac{13}{90} = 0.14 \leqslant 0.5$ *but* $P(X \leqslant 4) = \frac{29}{90} = 0.32 < 0.5$ (4 *is not a median*)

Discrete variable	Continuous variable
Countable support	Uncountable support
Probability mass function	Probability density function
Probabilities assigned to single values	Probabilities assigned to intervals of values
Each possible value has strictly positive probability	Each possible value has zero probability

f(x) is the probability density function

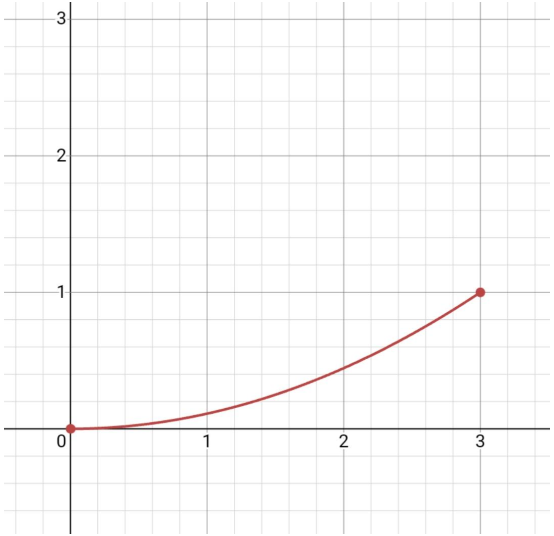
Continuous random variables

- the ***mean*** $\mu = \int_{-\infty}^{+\infty} x.\, f(x)dx$
- the ***median*** $\int_{-\infty}^{m} f(x)dx = \frac{1}{2}$
- the ***mode*** *the peak (or local maximum) of f(x)*
- the ***variance*** $Var(X) = E(X^2) - \mu^2$
- the ***standard deviation*** $\sigma = \sqrt{Var(X)}$

Exercise 5

A continuous random variable X has probability density function defined as:

$$f(x) = \begin{cases} x^2/9, & 0 \le x \le 3 \\ 0, & elsewhere \end{cases}$$



Calculate:

- mean
- median
- mode
- variance
- standard deviation

Exercise 5

$$mean = E(X) = \int_0^3 x f(x)\,dx = \int_0^3 x.\frac{x^2}{9}\,dx = \frac{x^4}{4*9}\Big|_0^3 = \frac{x^4}{36}\Big|_0^3 = \frac{3^4}{36} - 0 = \frac{81}{36} = \frac{9}{4} = 2.25$$

$$median = m \text{ such that } \int_0^m f(x)\,dx = 0.5$$

$$\int_0^m \frac{x^2}{9}\,dx = 0.5 \Longrightarrow \frac{x^3}{27}\Big|_0^m = 0.5 \Longrightarrow m^3 = \frac{27}{2} \Longrightarrow m = \sqrt[3]{\frac{27}{2}} \simeq 2.38$$

median is 2.38

$$mode = argmax_x f(x) = 3 \left(\text{the value } x \text{ at which } f \text{ takes its maximum value } max(f) = \frac{9}{9} = 1 \right)$$

$$variance = \sigma^2 = E\left(X^2\right) - E^2(X)$$

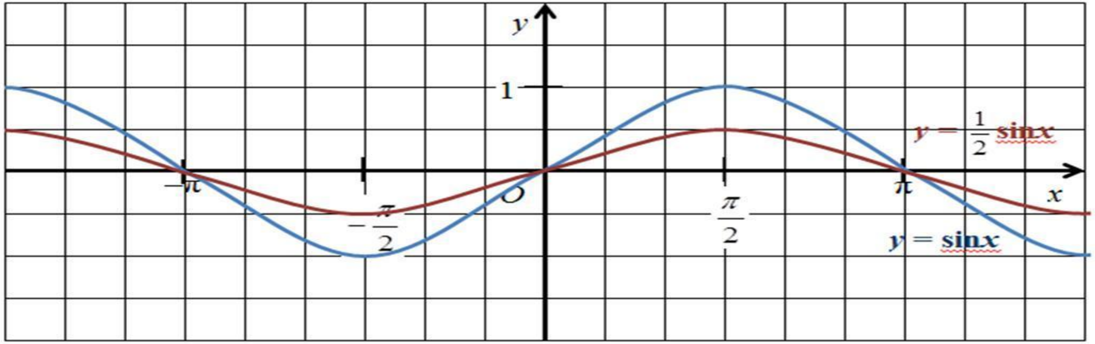
$$E(X)^2 = (2.25)^2 = 5.0625$$

$$E\left(X^2\right) = \int_0^3 x^2 f(x)\,dx = \int_0^3 x^2.\frac{x^2}{9}\,dx = \frac{x^5}{5*9}\Big|_0^3 = \frac{x^5}{45}\Big|_0^3 = \frac{3^5}{45} - 0 = 5.4$$

$$\sigma^2 = 5.4 - 5.0625 = 0.3375$$

$$\sigma = \sqrt{0.3375} \simeq 0.581$$

Exercise 6



A continuous random variable X has probability density function defined as:

$$f(x) = \begin{cases} \sin(x)/_2, & 0 \le x \le \pi \\ 0, & elsewhere \end{cases}$$

Calculate:

- mean
- median
- mode
- variance
- standard deviation

Exercise 6

Integration by Partial Fractions

$$mean = E(X) = \int_0^{\pi} x f(x)\, dx = \int_0^{\pi} x.\frac{sin(x)}{2}\, dx = \frac{1}{2}\int_0^{\pi} x.sin(x)\, dx$$

We can use integration by parts :

$$\int udv = u.v - \int vdu$$

$$u = x,\ dv = sin(x)dx$$
$$du = 1,\ v = -cos(x)$$

$$E(X) = \frac{1}{2}\int_0^{\pi} x.sin(x)\, dx = \frac{1}{2}\left(-xcos(x)\Big|_0^{\pi} - \int_0^{\pi} -cos(x)\, dx\right) = \frac{1}{2}\left(-xcos(x)\Big|_0^{\pi} + sin(x)\ \Big|_0^{\pi}\right)$$
$$= \frac{1}{2}(-\pi cos(\pi) + 0 + sin(\pi) - sin(0)) = \frac{1}{2}(\pi + 0 + 0 - 0) = \frac{\pi}{2}$$

$$median\ = m\ such\ that\ \int_0^{m} f(x)\, dx = 0.5$$

$$\int_0^{m}\frac{sin(x)}{2}\, dx = 0.5 \Longrightarrow \frac{-cos(x)}{2}\Big|_0^{m} = 0.5 \Longrightarrow cos(m) = -1 \Longrightarrow m = \pi + 2\pi k\ for\ integer\ k \geqslant 0$$

but the range of values for X is $[0,\ \pi]$, *so* $k = 0$ *and median is* $\frac{\pi}{2}$

$$mode = argmax_x f(x) = \frac{\pi}{2}\ \left(the\ value\ x\ at\ which\ f\ takes\ its\ maximum\ value\ max(f) = \frac{1}{2}\right)$$

Exercise 6

$$variance = \sigma^2 = E(X^2) - E^2(X)$$

$$E(X)^2 = \left(\frac{\pi}{2}\right)^2 = \frac{\pi^2}{4}$$

$$E(X^2) = \int_0^{\pi} x^2 f(x)\,dx = \int_0^{\pi} x^2 . \frac{sin(x)}{2}\,dx = \frac{1}{2}\int_0^{\pi} x^2 . sin(x)\,dx$$

To calculate this integral, you need to apply integration by parts twice.

$$E(X^2) = \frac{1}{2}\left(\left(2xsin(x) - (x^2 - 2)\,cos(x)\right)\Big|_0^{\pi}\right) = \frac{1}{2}\left\{\left(2\pi\ .\ sin(\pi) - (\pi^2 - 2)\,cos(\pi)\right) - (0 - (0-2)\ .\ cos(0))\right\}$$

$$= \frac{1}{2}(\pi^2 - 2 - 2) = \frac{\pi^2 - 4}{2}$$

$$\sigma^2 = \frac{\pi^2 - 4}{2} - \frac{\pi^2}{4} = \frac{\pi^2}{4} - 2 \simeq 0.467$$

$$\sigma = \sqrt{0.467} \simeq 0.683$$

Z-score – Exercise 7

Suppose that the weight of navel oranges is normally distributed with mean $\mu = 8$ ounces, and standard deviation $\sigma = 1.5$ ounces. We can write $X \sim N(8, 1.5)$.

Answer the following questions:

1. What proportion of oranges weigh more than 11.5 ounces?
2. What proportion of oranges weigh between 6.2 and 7 ounces?

Z-score – Exercise 7

1. What proportion of oranges weigh more than 11.5 ounces?
2. What proportion of oranges weigh between 6.2 and 7 ounces?

$$Z \sim N(0,1)$$

$$P(X > 11.5) = P(Z > \frac{11.5-8}{1.5}) = P(Z > 2.33) = 1 - 0.9901 = 0.0099$$

$$P(6.2 < X < 7) = P(\frac{6.2-8}{1.5} < Z < \frac{7-8}{1.5}) = P(-1.2 < Z < -0.67) = 0.2514 - 0.1151 = 0.1363$$

Exercise 8 – Median of discrete random variables

x	4	5	12	53	84	125
$f(x) = P(X = x)$	0.2	0.1	0.1	0.3	0.1	0.2
$F(x) = P(X \leqslant x)$	0.2	0.3	0.4	0.7	0.8	1

$P(X < 12) = 0.3 \leqslant 0.5$ *but* $P(X \leqslant 12) = 0.4 < 0.5$ (12 *is not a median*)
$P(X < 53) = 0.4 \leqslant 0.5$ *and* $P(X \leqslant 53) = 0.7 \geqslant 0.5$ (53 *is a unique median*)
median is $m = 53$

Exercise 9 – Median of discrete random variables

x	4	5	12	53	84	125
$f(x) = P(X = x)$	0.2	0.1	0.1	0.1	0.3	0.2
$F(x) = P(X \leqslant x)$	0.2	0.3	0.4	0.5	0.8	1

$P(X < 12) = 0.3 \leqslant 0.5$ *but* $P(X \leqslant 12) = 0.4 < 0.5$ (12 *is not a median*)
$P(X < 53) = 0.4 \leqslant 0.5$ *and* $P(X \leqslant 53) = 0.5 \geqslant 0.5$ (53 *is a candidate median*)
$P(X < 84) = 0.5 \leqslant 0.5$ *and* $P(X \leqslant 84) = 0.8 \geqslant 0.5$ (84 *is also a candidate median*)

The median is any number between 53 *and* 84. *For convenience, we choose the midpoint of the interval* $[53,\ 84]$ *which is* $\dfrac{53 + 84}{2} = 68.5$

Exercise 10 – Median of discrete random variables

x	4	5	12	53	84	125
$f(x) = P(X = x)$	0.2	0.1	0.1	0.05	0.3	0.25
$F(x) = P(X \leqslant x)$	0.2	0.3	0.4	0.45	0.75	1

$P(X < 53) = 0.4 \leqslant 0.5$ *but* $P(X \leqslant 53) = 0.45 < 0.5$ (53 *is not a median*)
$P(X < 84) = 0.45 \leqslant 0.5$ *and* $P(X \leqslant 84) = 0.75 \geqslant 0.5$ (84 *is a unique median*)
$m = 84$

Exercise 11 – Median of discrete random variables

x	0	1	4
$f(x) = P(X = x)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
$F(x) = P(X \leqslant x)$	$\frac{1}{3}$	$\frac{2}{3}$	1

$$P(X < 1) = \frac{1}{3} \leqslant 0.5 \textit{ and } P(X \leqslant 1) = \frac{2}{3} \geqslant 0.5 \text{ (1 } \textit{is a unique median})$$

$$P(X < 4) = \frac{2}{3} > 0.5 \text{ (4 } \textit{is not a median})$$

$$P(X < 0) = 0 \leqslant 0.5 \textit{ but } P(X \leqslant 0) = \frac{1}{3} < 0.5 \text{ (0 } \textit{is not a median})$$

$$m = 1$$

Exercise 12 – Median of discrete random variables

x	0	1	2	3
$f(x) = P(X = x)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
$F(x) = P(X \leqslant x)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{3}{4}$	1

$$P(X < 1) = \frac{1}{4} \leqslant 0.5 \text{ and } P(X \leqslant 1) = \frac{2}{4} = 0.5 \text{ (1 is a candidate median)}$$

$$P(X < 2) = \frac{2}{4} = 0.5 \text{ and } P(X \leqslant 2) = \frac{3}{4} \geqslant 0.5 \text{ (2 is a candidate median)}$$

$$P(X < 0) = 0 \leqslant 0.5 \text{ but } P(X \leqslant 0) = \frac{1}{4} < 0.5 \text{ (0 is not a median)}$$

$$m = \frac{1+2}{2} = 1.5$$

Note: X here is uniformly distributed and the median of a discrete uniform distribution is:

$$\frac{a+b}{2} = \frac{0+3}{2} = 1.5$$

Attendance

https://baam.duckdns.org

Questions?