සංඛ්‍යාන විද්‍යාව - 5

දත්ත සාරාංශනය

පර්යේෂණ හා පුහුණු කිරීමේ නිලධාරි
මහින්ද පෙරේරා විසිනි

පසුගිය කලාපවල මෙම මාතෘකාව යටතේ පළ වූ ලිපි මගින් පිළිවෙලින් පහත සඳහන් මාතෘකා හා සම්බන්ධ කරුණු පැහැදිලි කළෙමු.

1. **නියැදීම යනු කුමක්ද?**
2. **පොදුවේ භාවිතා වෙන නියැදි මෝස්තර මොනවාද?**
3. **නියැදියකින් දත්ත ලබා ගන්නේ කෙසේද?**
4. **දත්ත සකස් කර ඉදිරිපත් කරන්නේ කෙසේද?**

මෙම ලිපි මාලාවේ අවසාන ලිපිය වන මෙයින් බලාපොරොත්තු වන්නේ දත්ත සම්පිණ්ඩනයෙහි හෙවත් සාරාංශනයෙහි ලා පොදුවේ භාවිතය කරන මිනුම් කිහිපයක් හඳුන්වා දීමටය.

ජාල රේඛයක් මගින් දත්ත ඉදිරිපත් කරන අන්දම පසුගිය ලිපියෙන් අපි ඉදිරිපත් කළෙමු. දත්ත වර්ගීකරණයකට යටත් කර, සංඛ්‍යාත ව්‍යාප්තියක් තනා, එය ජාල රේඛයක් මගින් ඉදිරිපත් කිරීමෙහි ප්‍රධාන පරමාර්ථය වූයේ දත්තවල ව්‍යාප්තියේ ස්වභාවය අවබෝධ කර ගැනීමය. නිරීක්ෂිත දත්ත සතුව ඇති සාමාන්‍ය තොරතුරු සෑහෙන ප්‍රමාණයක් ජාලරේඛයක් මගින් ප්‍රකාශ වෙන නමුත්, වඩාත් පැහැදිලි, ප්‍රයෝජනවත් සහ සම්පිණ්ඩනය කළ තොරතුරු ප්‍රකාශ කළ හැක්කේ සංඛ්‍යාත ව්‍යාප්තියේ ගණිතමය විස්තරයකින්ය. එවැනි ගණිතමය විස්තර කිරීමක් සඳහා සංඛ්‍යාන මිනුම් කිහිපයක් භාවිතය කෙරේ.

සංඛ්‍යාත ව්‍යාප්තියක් ගණිතමය ලෙස විස්තර කිරීමට භාවිතය කරන මිනුම් ප්‍රධාන වශයෙන් වර්ග හතරකි. ඒවායින් වඩාත් වැදගත් මිනුම් වර්ග දෙක සහ ඒවා යටතේ මෙහිදී විස්තර කරන මිනුම් පහත සඳහන් වේ.

1. පිහිටීමේ හෙවත් කේන්ද්‍රික ප්‍රවනතා වේ මිනුම් –
   (අ) මධ්‍යනාය (ආ) මධ්‍යස්ථය
   (ඇ) මාතය
2. විසිරීමේ හෙවත් අපකිරණයේ මිනුම්
   (අ) පරාසය (ආ) විචලතාව සහ සම්මත අපගමනය

මධ්‍යනාය පන්තියක සිටින ළමයින්ගේ සාමාන්‍ය වයස, ගමක පදිංචි පවුල්වල සිටින සාමාන්‍ය සාමාජික සංඛ්‍යාව යනාදි වශයෙන් අපි කතා කිරීමට පුරුදු වී ඇති බව මෙහිදී සිහිපත් කරන්න. මධ්‍යනාය යනුවෙන් හඳුන්වන්නේ අපට හුරුපුරුදු මෙම 'සාමාන්‍ය' අගයය. නිරීක්ෂණ සමූහයක මධ්‍යනාය ගණනය කරන්නේ එම නිරීක්ෂණ සියල්ල එකතු කර, එම එකතුව නිරීක්ෂණ සංඛ්‍යාවෙන් බෙදීමෙන්ය. උදාහරණයක් වශයෙන්, වත්තකින් කපා ගත් තැඹිලි වලු 10 ක, එක් එක් තැඹිලිවල්ලේ තිබූ ගෙඩි ගණන, 8, 7, 10, 14, 13, 10, 15, 12, 9 සහ 11 වී නම්, තැඹිලිවල්ලක ඇති ගෙඩි ගණනෙහි මධ්‍යනාය හෙවත් තැඹිලි වල්ලක ඇති සාමාන්‍ය ගෙඩි ගණන,

$$= (8+7+10+14+13+10+15+12+9+11) \div 10 = 109 \div 10 = 10.9$$ වේ.

පොදුවේ භාවිතා කළහැකි නිරීක්ෂණය සමූහයක මධ්‍යනාය ලබා ගැනීම සඳහා සූත්‍රයක් මෙසේ සඳහන් කළ හැකිය. අප හමුවේ නිරීක්ෂණය ඇතැයිද, ඒවා $x_1, x_2, x_3, \ldots, x_n$ යනුවෙන් නම් කර ඇතැයිද

සිතමු. එවිට, මාධ්‍යනාය $\bar{x}$ මගින් සඳහන් කරන අතර,

$$\text{මධ්‍යනාය}, \quad \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

යන සූත්‍රයෙන් එම ගණනය කළහැකි වේ. මෙහි Σ සංකේතය මගින් 'එකතු කරන්න' යන අදහස ප්‍රකාශ වේ. මේ අනුව $i = 1$ සිට n දක්වා x_i අගයන් එකතු කළ යුතුය. ඉහත සළකා බැලූ උදහරණයේ $n = 10$ වන අතර, $x_1 = 8$, $x_2 = 7$ .... යනාදී වශයෙන් වේ. තවද, එහි $\Sigma x_i = 109$ වේ.

නිරීක්ෂණ සමූහයක මධ්‍යනාය යනු එම නිරීක්ෂණ සියල්ල වෙනුවෙන් පෙනී සිටිය හැකි හෙවත් එම නිරීක්ෂණය සියල්ල නිරූපනය කරන කිසියම් අගයක් වේ. මධ්‍යනාය සඳහා ප්‍රතිඵල වන අගය නිරීක්ෂණ සමූහයේ දක්නට ඇති අගයන්ම විය යුතු නොවේ. අපගේ උදහරණයේ, තැඹිලිවල්ලක ඇති ගෙඩි ගණනෙහි මධ්‍යනාය 10.9 විය පැහැදිලිවම එම අගය අපගේ නිරීක්ෂණ අතර දක්නට නොමැත. ඇත්ත වශයෙන්ම, කිසිම තැඹිලි වල්ලක ගෙඩි 10.9 ක් අපට කිසි දිනෙක මුණ නොගැසෙනු ඇත. තැඹිලි වල්ලක ඇති ගෙඩි ගණනෙහි මධ්‍යනාය 10.9 යයි සඳහන් කිරීමෙන් පොදුවේ අපි අදහස් කරන්නේ, මෙම වත්තෙන් තැඹිලි වලු 100 ක් ගතහොත්, එහි මුළු මනින්ම තැඹිලි ගෙඩි 10.9×100 = 1,090ක් (පමණ) ඇති බවය.

මධ්‍යස්ථය– නිරීක්ෂණ සමූහයක් ආරෝහණ පිළිවෙලට (එනම්, අඩු අගයේ සිට වැඩි අගය දක්වා) සකස් කළවිට, හරි මැද පිහිටි නිරීක්ෂණයේ අගය මධ්‍යස්ථය නමින් හඳුන්වනු ලැබේ.

නිරීක්ෂණයේ ඔත්තේ සංඛ්‍යාවක් ඇති විට හරි මැද නිරීක්ෂණය තෝරා ගැනීම සහ ඒ නයින් මධ්‍යස්ථය ලබා ගැනීමේ අපහසුවක් පැන නොනගී. එහෙත් නිරීක්ෂණ ඉරට්ටේ සංඛ්‍යාවක් ඇති විට හරි මැදට යෙදෙන නිරීක්ෂණයක් නොමැත. එවැනි අවස්ථාවක මැදට යෙදී ඇති නිරීක්ෂණ දෙකෙහි සාමාන්‍යය හෙවත් මධ්‍යනාය ගෙන එය අදාළ මධ්‍යස්ථය හැටියට නම් කරනවා ඇත.

ඉහත සළකා බැලූ උදහරණයේ ඇති නිරීක්ෂණය ආරෝහණ පිළිවෙලට සකස් කළවිට, 7 8, 9, 10, 10, 11, 12, 13, 14, 15 ආකාරය ගනී. මෙහි නිරීක්ෂණය ඉරට්ටේ සංඛ්‍යාවක්, එනම් 10, ඇති නිසා මධ්‍යස්ථයේ අගය මෙහිදී ලබා ගන්නේ 5 වැනි සහ 6 වැනි නිරීක්ෂණ දෙක එකතු කර 2 න් බෙදීමෙන්ය. එවිට, මෙම නිරීක්ෂණ වල මධ්‍යස්ථය $(10+11) \div 2 = 10.5$ වේ.

මධ්‍යස්ථය යනුවෙන් සඳහන් කරන අගයෙන් අපි අදහස් කරන්නේ නිරීක්ෂණ වලින් අඩක් හෙවත් 50% ක් එම අගයට අඩු අගයන් ගන්නා බව සහ නිරීක්ෂණ වලින් අඩක් හෙවත් 50% ඒ එම අගයට වැඩි අගයක් ගන්නා බවය. ඒ අනුව අපගේ උදහරණයේ, තැඹිලි වලු වලින් 50% ක ගෙඩි 10.5 ට අඩුවෙන්ද, 50% ක ගෙඩි 10.5 ට වැඩියෙන් ඇති බවද ප්‍රකාශ වේ.

මාතය– නිරීක්ෂණ සමූහයක වැඩියෙන්ම යෙදී ඇති නිරීක්ෂණය හෙවත් වඩා පොදු නිරීක්ෂණය මාතය නමින් හඳුන්වනු ලැබේ. ඒ අනුව, අප සළකා බැලූ උදහරණයේ වැඩියෙන්ම යෙදී ඇති නිරීක්ෂණය 10 නිසා (10 අගය දෙවරක්ද, අනිකුත් අගයන් එක් වර බැගින්ද යෙදී ඇත.) එහි මාතය 10 වේ. මාතයෙන් ප්‍රකාශ වන්නේ වඩා පොදු අගයය. මාත අගය පිළිබඳ හැඟීම අපගේ එදිනෙද කටයුතුවලදී බොහෝවිට පාවිච්චි කෙරේ. වෙළෙන්දෙක් ලෙලි ගැසූ පොල් ගොඩකින් සාම්පලයක වශයෙන් ගෙඩියක් ගෙන ඒ අනුව යම් මිළක් තීරණය කෙරේ. එහිදී පොල් ගොඩෙන් ඔහු තෝරා ගන්නේ වඩාත් පොදු ප්‍රමාණයේ ගෙඩියකි. එනම් මාත අගය ගැන ඔහු සැළකිල්ල දක්වා ඇත. මසා නිම කළ ඇඳුම්, පාවහන් යනාදිය වෙළඳ පොලට සැපයීමේදී නිෂ්පාදකයා මාත අගයන් කෙරෙහි උනන්දු වේ.

පොදු අගයන් වැඩි ගණනක් ඇති විට මාත අගයන් එකකට වැඩියෙන් වේ. උදහරණයක් වශයෙන්, 20, 25, 28, 28, 30,

33, 33, 40, 41, 45, 50 යන නිරීක්ෂණ සමූහයේ 28 සහ 33 යන අගයන් දෙවාරය බැගින් යෙදී ඇත. එබැවින් එහි මාත දෙකකි. ඒවා 28 සහ 33 වේ.

පරාසය- නිරීක්ෂණ සමූහයක විසිරීම හෙවත් අපකිරණය ප්‍රකාශ කිරීම සඳහා යොදා ගන්නා මිනුම් අතරින් සරලතම මිනුම පරාසය වේ. නිරීක්ෂණ සමූහයක විශාලම අගය සහ කුඩාම අගය අතර, අන්තරය පරාසය නමින් හඳුන්වනු ලැබේ සමහරවිට කුඩාම අගය සහ විශාලම අගය සඳහන් කිරීමෙන් අදාළ දත්තවල පරාසය සඳහන් කරනු ඇත. උදාහරණයක් වශයෙන්, මධ්‍යනාය විස්තර කිරීමේදී ඉහත සළකා බැලූ නිරීක්ෂණවල පරාසය (7:15) හෝ 15–7 = 8 හෝ යයි සඳහන් කළ හැකිය. ඒ මගින් අධ්‍යයනය යටතේ ඇති නිරීක්ෂණ පිහිටා ඇති ප්‍රාන්තරය ප්‍රකාශ වේ. ඉතා පහසුවෙන් පරාසය ගණනය කළහැකි වුවද, සංඛ්‍යානමය අතින් එය අපකිරණයේ ප්‍රබල මිනුමක් නොවේ.

විචලතාව සහ සම්මත් අපගමනය

දත්තයන්හි විසිරීම ප්‍රකාශ කිරීම සඳහා ප්‍රචලිතව භාවිතා කරන වැදගත් මිනුමක් හැටියට විචලතාව හඳුන්වා දිය හැකිය. සම්මත අපගමනය යනු, වෙන කිසිවක් නොව, විචලතාවේ ධනවර්ග මූලයය. විචලතාව නැතහොත් සම්මත අපගමනය මගින් ඇත් තෙන්ම මිනුම් කරනු ලබන්නේ මධ්‍යනාය වටා නිරීක්ෂණයන්ගේ විසිරීමය.

එක් එක් නිරීක්ෂණයෙන් අදාළ මධ්‍යනාය අඩු කළ විට, ප්‍රතිඵල වන අගයන්ට (ඒවා ධන අගයන් මෙන්ම, රින අගයන්ද විය හැකිය.) අපගමනය යයි කියනු ලැබේ. මෙම එක් එක් අපගමනයෙහි වර්ගය ගෙන එකතු කර නිරීක්ෂණය ගණනින් බෙදූ විට විචලතාව ප්‍රතිඵල වේ. අප හමුවේ නිරීක්ෂණ n ඇතැයිද, ඒවා $x_1, x_2, \ldots\ldots, x_n$ යනුවෙන් නම් කර ඇතැයිද, නැවතත් සිතමු. තවද එම නිරීක්ෂණවල මධ්‍යනාය $\bar{x}$ යයිද සිතමු. එවිට විචලතාව S^2 මගින් සඳහන් කරන අතර,

$$\text{විචලතාව,}\quad S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

යන සූත්‍රයෙන් එය ගණනය කළහැකි වේ. විචලතාවේ ධන වර්ග මූලය සම්මත අප ගමනය යනුවෙන් නම් කරන නිසා,

$$\text{සම්මත අපගමනය,}\quad S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

වේ. මෙහිදී අදාළ ගණනය කිරීම් සැහැල්ලු කිරීමට මහක් පාද ගැනීමට හැකි වේ. කුමක් නිසාද, ඉහත වූ පළමු සූත්‍රය සුළු කළවිට පහත සඳහන් ආකාරය ගන්නා නිසාය. එනම්,

$$\text{විචලතාව,}\quad S^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\bar{x})^2$$

ආරම්භයේ සළකා බැලූ තැඹිලි වලුවල ඇති ගෙඩි ගණන හා සම්බන්ධ උදාහරණ ඇති නිරීක්ෂණවල විචලතාව දැන් මෙසේ ගණන කළහැකි වේ. එහි $n=10$, $\bar{x}=10.9$ සහ

$$\sum_{i=1}^{10} x_i^2 = 8^2 + 7^2 + \ldots\ldots + 11^2 = 1249$$

$$\text{නිසා විචලතාව,}\quad S^2 = \frac{1}{10}(1249) - (10.9)^2$$

$$= 124.9 - 118.8 = 6.1$$

ඒනයින්, සම්මත අපගමනය, $S = \sqrt{6.1} = 2.47$

විචලතාවේ ඒකක නිරීක්ෂණ මිනුම් කළ ඒකකවල වර්ග ලෙසද, සම්මත අපගමනයේ ඒකක මිනුම්වල ඒකක වලින්ද ප්‍රකාශ වේ. ඒ අනුව අපගේ උදාහරණයේ, විචලතාව (ගෙඩි)^2 වලින් හා සම්මත අපගමනය ගෙඩි වලින්ද ප්‍රකාශ වේ. එනම් සම්මත අප ගමනය ගෙඩි 2.47 වේ.

කලින් සඳහන් කළ අපගමනය $(= x_i - \bar{x})$ මත විචලතාව සහ සම්මත අපගමනය පදනම් වී ඇති බව සහ කුඩා අපගමන

මගින් මධ්‍යන්‍යය වඩා නිරීක්ෂණවල ඒක රාශි වීමද, විශාල අපගමන වලින් මධ්‍යන්‍යය වටා නිරීක්ෂණවල විසිරීමද ප්‍රකාශ වෙන බව සනිටුහන් කරන්න. ඒ නයින් බලන විට, විචලතාවේ හෝ සම්මත අපගමනයේ කුඩා අගයන් මගින් නිරීක්ෂණවල ඒකරාශි වීමද, විශාල අගයන් මගින් නිරීක්ෂණවල විසිරීමද ප්‍රකාශ වේ.

මිණුම් අතර අන්‍යොන්‍ය වැදගත්කම

අධ්‍යයනයේ පරමාර්ථ අනුව සහ නිරීක්ෂණවල ස්වභාවය ගැන සළකා ඒ ඒ මිණුම් භාවිතා කළයුතු බව අමුතුවෙන් සඳහන් කළයුතු නොවේ. නිරීක්ෂණවල පිහිටීම පිළිබඳ අධ්‍යයනය කරන විට මධ්‍යන්‍යය, මධ්‍යස්ථය සහ මාතය යන මිනුම් වලින් එකක් හෝ කිහිපයක් භාවිතය කළහැකිවේ. විසිරීම පිළිබඳ අධ්‍යයනය කරන විට බොහෝවිට විචලතාව හෝ සම්මත අපගමනය භාවිතා කිරීමට සිදු වේ.

නිරීක්ෂණ සමූහයක් මගින් අපට ලබා දෙන ජාලරේඛාවේ හැඩය අනුව සළකා බලා පිහිටීම මිණුම් කිරීම සඳහා ඇති මිණුම් තුනෙන් කුමන ඒවා අදාළ අධ්‍යයනයට යෝග්‍යදැයි අපට තීරණය කිරීමට සිදු වේ. ජාලරේඛය සමමිතික හැඩය (එනම්, චිත්‍රයේ මැද උස්ම සහ දෙපැත්ත පහත්ව තරමක් සමාන ආකාරය) ගනී නම් පිහිටීම මිණුම් කිරීම මධ්‍යන්‍ය යෝග්‍ය මිනුමක් වේ. එහෙත් ජාලරේඛීය කුටික හැඩයක් ගනී නම් (එනම්) සමමිතිකයෙන් බැහැර ගොස් තිබේ නම්, මධ්‍යස්ථය සහ මාතය යෝග්‍ය මිණුම් වේ. උදහරණයක් වශයෙන් පුද්ගලයින්ගේ වයස, උස, බර, ළමයින්ගේ විභාග ලකුණු, බෝග වල අස්වැන්න වැනි නිරීක්ෂණ සඳහා මධ්‍යන්‍යය වඩා යෝග්‍යවේ. මාසික ආදායම, ගෙවල්වල වර්ගඵලය, ඉඩම් කැබලි වල ප්‍රමාණය යනාදි නිරීක්ෂණ සඳහා මධ්‍යන්‍යය හා මාතය වඩා යෝග්‍ය වේ.

ආපසු හැරී බැලීමක්

ආරම්භක ලිපියේ සඳහන් කළ පරිදි මෙම ලිපි පෙළේ පරමාර්ථය වූයේ 'සංඛ්‍යානය පිළිබඳ මූලික අවබෝධයක් නැති අයෙකුට වුව පහසුවෙන් අවබෝධ කරගත හැකි වන ලෙසින් සංඛ්‍යානයේ මූලික මූල ධර්ම හා භාවිතා පිළිබඳ අවබෝධයක් ලබා දීමය. සෑහෙන තරම් දුරට මෙම පරමාර්ථය සපුරාලන ලෙස මෙම ලිපි පෙළ සකස් වී යයි සිතමු.

සංඛ්‍යානය ඉතා ගැඹුරු, වඩාත් සංකීර්ණ එමෙන්ම පුළුල් විෂයයකි. ගණිතය පිළිබඳ හෝ නිපුනත්වයක් නොමැතිව සංඛ්‍යානය සෑහෙන දුරට හෝ හැදෑරීම දුෂ්කර කාර්යයකි. කරුණු මෙසේ හෙයින්, ඇත්ත වශයෙන්ම මෙම ලිපි පෙළෙන් උත්සාහ කර ඇත්තේ සංඛ්‍යානයේ 'අ' යන්න හඳුන්වා දීමට තැත් කිරීමය.