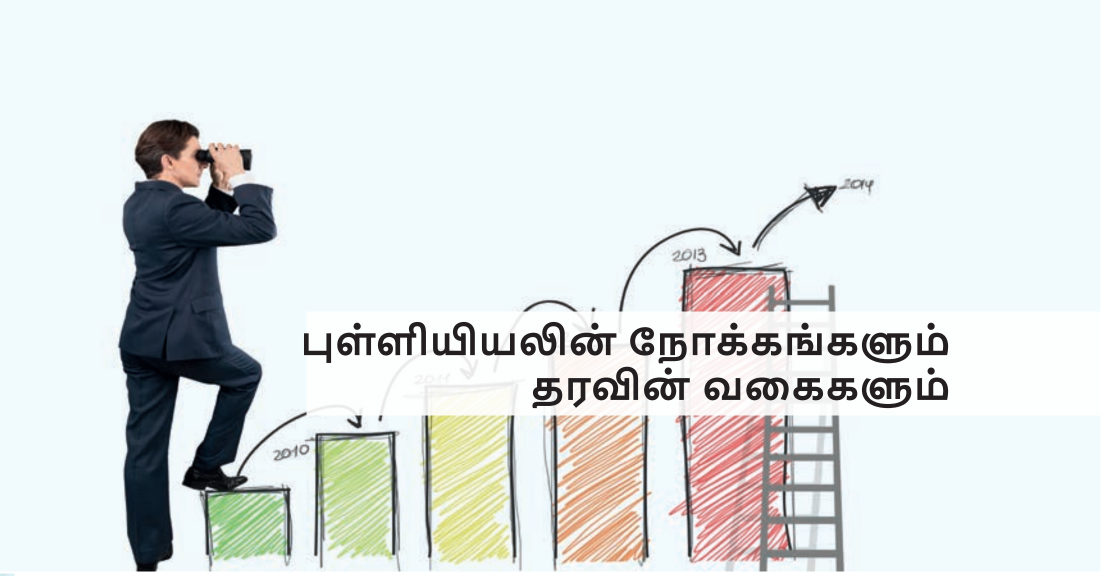
அத்தியாயம்

1



புள்ளியியலின் நோக்கங்களும் தரவின் வகைகளும்

பிரசந்த சந்திர மஹலனோபிஸ்
(29 ஜூன், 1893–28 ஜூன், 1972)

இந்தியப் புள்ளியியலின் தந்தை என P.C. மஹலனோபிஸ் அழைக்கப்படுகிறார். தொடக்கக் கல்வியைப் பிராமோ மாணவர் பள்ளியிலும், இயற்பியல் இளங்கலை பட்டத்தை கொல்கத்தாவில் உள்ள பிரசிடென்சி கல்லூரியிலும் பெற்றார். உயர்கல்வி தொடர்வதற்காக அவர் 1913 –ம் வருடம் இங்கிலாந்தில் உள்ள லண்டன் பல்கலைக் கழகத்திற்கு சென்றார். பயோமெட்ரிக்கா என்ற இதழ் வெளியீட்டாளர்களோடு ஏற்பட்ட அறிமுகம் அவருக்கு புள்ளியியலில் ஆர்வத்தை தூண்டியது.

ஒரு புள்ளியியல் ஆய்வகத்தை பிரசிடென்சி கல்லூரியில் தொடங்கினார். அதுவே பின்னர் கொல்கத்தாவிலுள்ள பிரசித்தி பெற்ற புள்ளியியல் நிறுவனத்தை உருவாக்குவதற்கு அடித்தளமிட்டது.

உங்களுக்குத் தெரியுமா?

P.C. மஹலனோபிஸ் பிறந்த தினமான ஜூன் 29 இந்தியப் புள்ளியியல் தினமாக கொண்டாடப்படுகிறது

பெருமளவிலான மாதிரி கணிப்பு ஆய்வுகளுக்கு அவர் பெரும் பங்களித்திருக்கிறார். துரித ஆய்வு பற்றி விவரங்களை அவர்தான் அறிமுகப்படுத்தினார். மேலும் மாதிரி கணிப்பு முறைகளின் பயன்பாட்டை உணர்த்தினார்.

'புள்ளியியலே அறிவியலின் இலக்கணம்' – கார்ல் பியர்ஸன்

நோக்கங்கள்

- புள்ளியியலின் பொருள் மற்றும் வரையறையை அறிமுகப்படுத்துதல்
- புள்ளியியலின் தோற்றமும் வளர்ச்சியும் பற்றி விளக்குதல்
- புள்ளியியலின் நோக்கம், பயன்பாடுகளை பட்டியலிடுதல்
- புள்ளியியலின் பல்துறை பயன்பாடுகளை விவரித்தல்
- தரவின் பொருளை அறிமுகப்படுத்துதல்
- பல்வேறு வகையான தரவுகளை வேறுபடுத்துதல்

VDAU9

Ch1_11th Stats TM.indd 1 07/12/2021 12:35:18

அறிமுகம்

இப்பாடத்தில் நாம் புள்ளியியலின் பொருள், பல்வேறு வரையறைகள், தோற்றம் மற்றும் வளர்ச்சி, செயல்பாடுகள், நோக்கம், விவசாயம், பொருளாதாரம் போன்ற பல்வேறு துறை பயன்பாடுகளின் விவரத்தை அறிந்து கொள்கிறோம். மேலும் தரவுகளின் வரையறை, வகைகள் மற்றும் அளவீட்டு அளவைகள் ஆகியவற்றை அறிந்து கொள்கிறோம்.

1.1 புள்ளியியலின் தோற்றமும் வளர்ச்சியும்

புள்ளியியலின் தோற்றமென்பது ஆதிகால மனிதனிடம் இருந்து ஆரம்பித்தது. அவன் தன் உடைமைகளை அறிந்துகொள்ள மரத்தில் குறியிட்டு அறிந்து கொண்டான். கி.மு 5000 ஆண்டு காலகட்டத்தில் மன்னர்கள் மக்கள் தொகை கணக்கெடுப்பு மற்றும் நாட்டின் வளத்தையும் இதன் மூலம் அறிந்துகொண்டனர். பண்டைய காலத்தில் போர்க் காலங்களில் ஏற்படும் இக்கட்டான சூழ்நிலையை எதிர்கொள்ள தங்களின் காலாட்படை, குதிரைப்படை, யானைப்படை மற்றும் எதிரிகளின் படைபலங்களின் புள்ளியியல் விவரங்களைக்கொண்டு முடிவுகள் எடுத்தனர். பின்னர் வரி மற்றும் அரசு நிர்வாகத்திற்கும் பயன்படுத்தப்பட்டது.

உங்களுக்குத் தெரியுமா?

"தரவு" (data) என்ற சொல் முதன் முதலில் 1640 களில் பயன்படுத்தப்பட்டது. 1946ல் "தரவு" என்ற சொல் "படிமாற்றத்திற்கு, கணிணியில் சேமித்து வைப்பதற்கு உகந்த" என்று பொருள்பட்டது. 1954ல் தகவல் செயலாக்கம் (data processing) என்ற சொல் அறிமுகப்படுத்தப்பட்டது. இது இலத்தீன் மொழியில் "கொடுத்த (அ) கொடுக்க" எனப் பொருள்படும்

'புள்ளியியல்' என்ற சொல் 'ஸ்டேட்டஸ்' (status) என்ற லத்தீன் சொல் அல்லது ஸ்டேட்டிஸ்டா (statista) என்ற இத்தாலிய சொல் அல்லது 'ஸ்டேடஸ்டிக்' (statistic) என்ற ஜெர்மன் சொல்லை மூலமாகக் கொண்டிருக்கிறது. இதன் பொருள் அரசியல் நிலைமை என்பதாகும். 'புள்ளியியல்' என்ற சொல் ஒரு நாட்டின் அரசாங்கத்தின் மக்கள், சமூகம் மற்றும் அரசியல் நிலைமை சார்ந்த புள்ளிவிவரங்கள் மற்றும் கணக்கெடுத்தல் போன்றவற்றைத் தருவதில் தொடர்புடையதாக இருந்தது. காலப்போக்கில் அதன் பரிணாம வளர்ச்சி அறிவியல் மற்றும் பெரும்பாலான துறைகளுக்கு அடிப்படையாக அமைந்தது. புள்ளியியலானது விவசாயத்துறை, மருத்துவத்துறை, தொழில்துறை, விளையாட்டு மற்றும் வணிகத்துறை ஆகியவற்றை உள்ளடக்கிய துறைகளின் கருத்துருவாக்கம் மற்றும் பயன்பாட்டுப் பகுதிகளின் வளர்ச்சியில் பயன்படுகிறது.

பண்டைய காலத்தில் புள்ளியியல் அரசு நிர்வாகத்திற்கும் போர்களுக்காகவும் பயன்படுத்தப்பட்டன. பின்னர் வரி வசூலிக்கும் நோக்கத்திற்காக பயன்படுத்தப்பட்டது என்பதற்குச் சான்றாக கௌடில்யரின் அர்த்தசாஸ்திரம் (கி.மு 324–300) என்ற நூலில் காணலாம். அக்பரின் நிதித்துறை அமைச்சரான இராஜா தோடர்மால் விவசாய நிலங்களைப் பற்றிய விவரங்களை சேகரித்தார். 17ஆம் நூற்றாண்டின் தொடக்க காலத்தில் வாழ்நிலைப்புள்ளியியல் என்ற அடிப்படைக் கருத்து தொடங்கி இப்போதைய நவீன காலத்தைய நிபுணத்துவ அளவீட்டு அறிவியல் (Actuarial science) வரை புள்ளியியலின் பயன்பாடுகள் வளர்ந்துள்ளன. 18ஆம் நூற்றாண்டின் இறுதியில் இயற்பியலில் பிழைகளின் கோட்பாடு என்ற கருத்தை காஸ் (Gauss) என்பவர் அறிமுகப்படுத்தினார்.

புள்ளியியல் என்பது அறிவியல் ரீதியாக எண் விவரங்களைச் சேகரிப்பது, அளிப்பது, பகுத்தாய்வது மற்றும் விளக்கமளிப்பது போன்றவற்றோடு தொடர்புடையது. எண் விவரங்களை முறையாகச் சேகரித்து தெளிவாக்குவதுடன் புள்ளியியல் தொடர்புடையதே புள்ளிவிவரம் என்ற சொல்லாகும்.

புள்ளியியல் என்ற சொல்லானது ஒருமை, பன்மை என இருவகைகளாக எடுத்துக்கொள்ளப்படுகிறது. பன்மையில் எடுத்துக்கொள்ளும் இது ஒரு குறிப்பிட்ட நோக்கத்துடன்,

Ch1_11th Stats TM.indd 2 07/12/2021 12:35:19

ஒரு திட்டமிட்ட முறையில் அளவிட்டோ, கணக்கிட்டோ பெறப்பட்ட எண்களின் விவரம் என்பதைக் குறிக்கிறது. எடுத்துக்காட்டாக நகரில் ஒரு நெருக்கடியான சாலையில் நடக்கும் விபத்துகளின் எண்ணிக்கை, ஒரு நாளில் ஒரு நாள்பட்ட நோய் காரணமாக இறந்தவர்களின் எண்ணிக்கை மற்றும் பலவற்றைக் கூறலாம். ஒருமையில் எடுத்துக்கொள்ளும்போது இது புள்ளியியல் கோட்பாடுகள் மற்றும் சேகரிப்பது, அளிப்பது, பகுத்தாய்வது மற்றும் எண்ணியியல் புள்ளி விவரங்களை விளக்குவது போன்றவற்றைக் குறிக்கிறது.

பின்னர் புள்ளியியலின் முக்கியத்துவம் உறுதியாக உணரப்பட்டு 20ஆம் நூற்றாண்டில் அபார வளர்ச்சி அடைந்தது. இந்த காலக்கட்டத்தில் நிறைய புதிய கோட்பாடுகளும் பல்வேறு துறைகளின் பயன்பாடுகளும் அறிமுகப்படுத்தப்பட்டன. புகழ்பெற்ற புள்ளியியலாளர்களின் பங்களிப்பின் காரணமாக பின்வரும் கருத்துகள் வெளிப்படுகின்றன. மாதிரியின் கோட்பாடு, நிகழ்தகவு கோட்பாடு, கருதுகோள் மதிப்பீடு மற்றும் புள்ளியியல் எடுகோள் சோதனை, போக்குபகுப்பாய்வு, ஒட்டுறவு பகுப்பாய்வு, பரிசோதனை, சோதனை வடிவமைத்தல், முன்கணிப்பு, காலத்தொடர், உயிரி புள்ளியியல், உளவியல் மற்றும் பல.

1900ஆம் ஆண்டுகளின் முற்பகுதியில், புள்ளியியலும் புள்ளியியலாளர்களும் அதிகம் முக்கியத்துவத்தை பெற்றிருக்கவில்லை. தொடர்ந்து வந்த ஆண்டுகளில் ஏற்பட்ட முன்னேற்றம் மற்றும் தொழில்நுட்ப வளர்ச்சியினால் எல்லா துறைகளிலும் பயன்படுத்தத் தக்கதாகவும், அனைவரின் கவனத்தை ஈர்க்கத்தக்கதாகவும் புள்ளியியல் துறை வளர்ச்சி அடைந்திருக்கிறது. இதன் விளைவாகப் பல புதிய இடைநிலைப் பிரிவுகள் உருவாகியது. தரவு சேகரித்தல் தரவின் ஆதாரமையம், புவியியல் தகவல் அமைப்பு, செயற்கை நுண்ணறிவு இன்னும் பல. இன்றைய காலகட்டத்தில் புள்ளியியலின் தொழில்நுட்பம் உயர் தகவலியல், குறிகையியல், தொலை தொடர்பு, பொறியியல், மருத்துவம், குற்றங்கள், சூழ்நிலையியல் இன்னும் பலவற்றில் பயன்படுகிறது.

இன்றைய வணிக மேலாளர்கள் நல்லமுடிவுகளை எடுத்து நல்ல விளைவுகளை பெறுவதற்கு பகுப்பாயும் திறனைக் கற்று வளர்த்துக்கொள்ள வேண்டிய தேவை இருக்கிறது. சிறந்த வணிக முடிவுகளை எடுப்பதற்குப் புள்ளியியல் உதவுகிறது. தங்களைச் சுற்றியுள்ள அபரிமிதமான விவரங்களை அறிந்துகொள்ள புள்ளியியல் சம்மந்தமான தரவுகள் தேவைப்படுகின்றன. மேலும் பரிவர்த்தனை நுட்பங்களான தீர்மான மரங்கள், ஒட்டுறவு பகுப்பாய்வு, தொகுப்பு போன்றவை வணிக செயல்முறைகள் மேம்பாட்டிற்குப் பயன்படுகின்றன.

1.2 வரையறைகள்

புள்ளியியலானது பல்வேறு ஆசிரியர்களால் வெவ்வேறு விதமாக வரையறுக்கப்பட்டிருக்கிறது..

- "புள்ளியியல் என்பது எண்ணுதலின் அல்லது கணக்கிடுதலின் அறிவியல்" – (ஏ.எல்.பௌலி)
- "புள்ளியியல் என்பது எண் விவரங்களைச் சேகரிப்பது, அளிப்பது, பகுத்தாய்வது மற்றும் விளக்கம் அளிப்பது போன்ற செயல்பாடுகளை கொண்ட அறிவியல் ஆகும்." –(கிராக்ஸ்டன் மற்றும் கௌடன்)

வாலிஸ்ட் மற்றும் இராபட்டின் புள்ளியியலின் வரையறை :

- "நிச்சயமற்ற எதிர்கொள்ளலில், முடிவுகளை எடுப்பதற்கான செயல்முறைகளின் வடிவமே புள்ளியியல் எனப்படும்."
- யா–லூன்–ஜோ என்ற அறிஞர் வாலிஸ்ட் மற்றும் இராபட்டின் புள்ளியியல் வரையறையை சிறிது மாற்றி கீழ்க்கண்டவாறு விவரிக்கிறார் : "புள்ளியியல் என்பது நிச்சயமற்ற நிகழ்வுகளின் விளைவால் வரும் எண்வடிவிலான தீர்மானிக்கப்பட்ட அளவிற்கு சாதகமற்ற விவரங்களைக் கொண்டு முடிவெடுப்பதற்கான ஒரு முறை"

Ch1_11th Stats TM.indd 3 07/12/2021 12:35:19

மேலே கொடுக்கப்பட்ட பெரும்பாலான வரையறைகள் ஓர் அரசுக்குத் தேவையான எண் விவரங்களை அளவிடுதல், கணக்கிடுதல் போன்றவற்றிற்கும் மட்டுமாக சுருக்கப்பட்டிருக்கிறது. ஆனால் சேக்ரிஸ்ட் போன்ற நவீன சிந்தனையாளர்கள் புள்ளியியலை கீழ்க்கண்டவாறு வரையறுக்கிறார்கள்.

'முன்னதாகவே தீர்மானிக்கப்பட்ட ஒரு நோக்கத்திற்காக ஒழுங்கான முறையில் சேகரிக்கப்பட்டதும் ஒன்றோடொன்று ஒப்பிடக்கூடியதாகவும், எண்ணிக்கையில் கூடியதாகவும் நியாயமான அளவுக்கு செம்மையாக மதிப்பிடத்தக்கதாகவும் பல்வகைக் காரணங்களால் குறிப்பிடத்தக்க அளவுக்கு பாதிக்கக்கூடியதுமான விவரங்களின் மொத்தமே புள்ளியியல் ஆகும்'.

மேற்கண்ட வரையறைகளில் கிராக்ஸ்டன் மற்றும் கௌடன் அவர்களின் வரையறை அதனுடைய முழுமைத் தன்மையினால் ஏற்றுக்கொள்ளக்கூடியதாக இருக்கிறது. இவ்வரையறையிலிருந்து புள்ளியியலானது கீழ்க்கண்ட குணங்களை வெளிப்படுத்துகிறது என்பது தெளிவாகிறது.

புள்ளியியலின் குணங்கள்:

(1) ஒரு குறிப்பிட்ட நோக்கத்திற்காக திட்டமிடப்பட்ட முறையில் சேகரிக்கப்பட்ட மொத்த விவரங்கள்.

சேகரிக்கப்பட்ட மொத்த விவரங்கள் மற்றும் எண்களைக் கொண்டு செயல்படுவது புள்ளியியல் ஆகும். ஓர் ஒற்றை எண் புள்ளியியல் விவரம் என்ற கூறமுடியாது. எடுத்துக்காட்டாக ஒரு மாணவனின் எடை 65 கிராம் எனில் இது புள்ளியியல் விவரம் ஆகாது. ஆனால் ஒரு வகுப்பில் உள்ள 60 மாணவர்களின் எடைகள் என்பது புள்ளியியல் விவரம் எனப்படும். ஏனெனில் அவற்றை சேர்த்துக் கணக்கிடலாம். மேலும் ஒன்றோடொன்று ஒப்பிடலாம். இவ்வுதாரணமானது "ஓர் இறப்பு என்பது துயரச் செய்தி பல இறப்புகள் என்பது புள்ளியியல் விவரம்" என்ற ஜோசப் ஸ்டாலினின் அறியப்பட்ட மேற்கோளை நினைவூட்டுகிறது. தரவு சேகரிக்கப்படும் நோக்கம் தெளிவாக இருத்தல் வேண்டும். சேகரிக்கப்பட்ட தரவு திட்டமிட்ட முறையில் இருத்தல் வேண்டும். முறைசாராததாக இருத்தல் கூடாது.

(2) அதிக எண்ணிக்கையிலான காரணிகளால் ஒரு குறிப்பிடக்கூடிய அளவிற்கு பாதிக்கக்கூடியவை.

பல்வகைக் காரணங்களால் குறிப்பிடத்தக்க அளவுக்கு பாதிக்கக்கூடியதுமான விவரங்களின் மொத்தமே புள்ளியியல் தரவுகள் ஆகும். இது புள்ளியியல் வல்லுநர்களுக்கு விளைவுகளை ஏற்படுத்தக்கூடிய காரணிகளை கண்டறிய உதவுகிறது. உதாரணமாக, சில பொருட்களின் விற்பனை விநியோகம், தேவை மற்றும் இறக்குமதி தரம் போன்ற காரணங்களால் பாதிக்கப்படுகிறது. இதேபோல் மில்லியன் இறப்புகள் ஏற்படும்போது நிர்வாகத்திற்கு அதற்குரிய காரணங்களை அறிந்து பின்னர் அதுபோல் மீண்டும் ஏற்படாதவாறு பார்த்துக்கொள்வர்.

(3) எண் வடிவில் விவரங்களை அளிப்பது

எல்லா புள்ளியியல் விவரங்களும் எண்வடிவில் சேகரிக்கப்படும். உதாரணமாக ஒரு தொலைபேசி நிறுவனத்தால் வழங்கப்படும் சேவையை குறை, சாராசரி, நிறை, மிக நிறை, சிறந்த போன்றவாறு வகைப்படுத்தலாம். இவைகள் இயற்கை பண்பு சார்ந்த விவரங்களாக அமைந்ததே தவிர புள்ளியியல் விவரங்களாக இல்லை. இவற்றையே எண் வடிவ தரவரிசையில்

Ch1_11th Stats TM.indd 4 07/12/2021 12:35:19

குறை என்பதற்கு 0, சராசரி 1, நிறை 2, மிகநிறை 3, சிறந்த 4 என்ற வகையில் குறியிடலாம். இந்த புள்ளி விவரங்கள் பகுப்பாய்வு செய்வதற்கு பொருத்தமானது. இவைகள் எண் வடிவில் அமையும்போது புள்ளியியல் விவரம் எனப்படும். மற்றொரு வகையான பண்பு சார்ந்த குணநலன்களான நேர்மை, அழகு, அறிவு போன்றவை எண்வடிவில் அளக்க முடியாது. எனவே அவை புள்ளியியல் விவரம் ஆகாது.

(4) விவரங்களைச் செம்மையான அளவிற்கு மதிப்பீடு செய்தல்

எண்ணியல் தரவுகளைச் சேகரிக்கும் போது உண்மையான கணக்கீடு அல்லது அளவீடு அல்லது மதிப்பீடு மூலம் சேகரிக்கப்படுகிறது. வரம்புக்குட்பட்ட தரவுகளைச் சேகரிப்பது விசாரணையின் நோக்கம் என்றால் உண்மையான எண்ணிக்கை அல்லது சிறந்த அளவீடு மூலம் சேகரிக்க வேண்டும். உதாரணமாக ஒரு மருத்துவமனையில் சேர்க்கப்பட்ட நோயாளிகளைக் கணக்கிட நோயாளிகளின் எண்ணிக்கையைக் கணக்கிட வேண்டும். நோயாளிகளின் உயரம் மற்றும் எடையைக் கணக்கிட வேண்டும். ஏனெனில் உண்மையான மதிப்பிடல் சேகரிப்பது சாத்தியமாகவோ, சாத்தியம் அல்லாமலும் இருக்கலாம். சாத்தியமாக இருப்பின் அது நேரமும் செலவும் அதிகரிக்கும். கணக்கிடப்பட்ட விவரங்கள் துல்லியமாகவும், தெளிவாகவும் இருக்காது. ஒர் அர்த்தமுள்ள பகுப்பாய்விற்கு ஒரு குறிப்பிட்ட தரத்திற்கு சரியான விவரங்கள் சேகரிக்கப்பட வேண்டும்.

(5) ஒன்றோடொன்று தொடர்புள்ள வகையில் விவரங்களை அமைத்தல்

புள்ளியியல் தரவுகளைச் சேகரிப்பதின் முக்கிய நோக்கங்களில் ஒன்று ஒப்பிடலாகும். அர்த்தமுள்ள மற்றும் ஏற்புடைய ஒப்பிடல் செய்ய தரவுகள் முடிந்த அளவிற்கு ஒருமைதன்மை கொண்டதாக இருக்க வேண்டும், பலவகைத் தன்மை பெற்றிருக்கக் கூடாது. உதாரணமாக ஒரு நிர்வாகத்தில் உள்ள ஆண் ஊழியர்கள் மற்றும் பெண் ஊழியர்கள் மாதாந்திர சேமிப்புகளை ஒப்பிடலாம். ஆனால் 20 வயதுடைய மரங்களின் உயரத்தை 20 வயதுடைய இளைஞர்களின் உயரத்தோடு ஒப்பிடுவது என்பது அர்த்தமற்றதாகும்.

பின்வருவன அனைத்தும் புள்ளியியலை ஒருமையில் எடுத்துக்கொள்ளும்போது, அதற்கான வரையறைகளாகும்.

புள்ளியியலை தரவுகள் என்ற பொருளில் எடுத்துக்கொள்ளும்போது, புள்ளியியல் என்பது பகுப்பாயத்தக்கதும், விளக்கமளிக்க ஏற்றவாறு எண்வடிவிலான விவரங்கள் ஆகும்.

புள்ளியியலை அறிவியலாக எடுத்துக்கொள்கையில், எந்தவொரு தள விசாரணையிலும் கிடைக்கும் எண்வடிவிலான தரவுகளைச் சேகரிப்பதற்கும், அளிப்பதற்கும், பகுப்பாய்வதற்கும், விளக்கமளிப்பதற்கும் பயன்படும் மற்றும் கோட்பாடுகளை ஆய்வதற்கான அறிவியல் ஆகும்.

1.3 புள்ளியியலின் செயல்பாடுகள்

புள்ளியியலின் முக்கிய ஏழு பண்புகளை இங்கு காண்போம்:

வ. எ	செயல்பாடுகள்	செயல்
1	விவரங்களைச் சேகரித்தல்	புள்ளியியலில் அடிப்படையானது தரவாகும். ஆகையால் மிகவும் துல்லியமாகவும் அறிவியல் ரீதியாகவும் சேகரிக்க வேண்டும்.
2	வகைப்படுத்துதல்	சேகரிக்கப்பட்ட தரவுகளை, பெரிய மற்றும் சிக்கலான தரவுகளை புரியும் வகையில் ஒத்த குணங்களுடைய குழுக்களாக வகைப்படுத்த வேண்டும்.

3	சுருங்கக் கூறுதல்	பகுப்பாய்விற்குத் தேவையான தகவல்களை இழந்துவிடாமல் தரவுகள் நேர்த்தியாக சுருக்கப்படவேண்டும். தரவுகளைச் சுருங்கக் கூறுதலின்போது தேவையான தகவலை இழந்துவிடாமல் துல்லியமாக புள்ளிவிவரத்தைப் பகுப்பாய்வு செய்யவேண்டும்.
4	ஒப்பிடல்	சிறந்த தரவுகளை அடையாளம் காண மற்றும் பிரிவுகளில் உள்ள ஒருங்கமைத்தன்மை அறிந்துகொள்வதற்குப் பயன்படுகிறது.
5	ஒட்டுறவு	மாறிகளுக்கு இடையே உள்ள உறவைக் கண்டறியப் பயன்படுகிறது.
6	காரணம் அறிதல்	சார்பற்ற மாறிகள் மீது சார்புடைய மாறிகளின் பாதிப்பை மதிப்பீடு செய்ய உதவுகிறது.
7	வாய்ப்பு	நிச்சயமற்ற தன்மையிலிருந்து சரியான தீர்மானத்தை எடுக்க புள்ளியியல் மிகவும் உதவுகிறது.

1.4 நோக்கம் மற்றும் பயன்பாடுகள்

பண்டைய காலங்களில், புள்ளியியலின் நோக்கம் மிகக் குறுகியது. புள்ளியியல் என்ற சொல்லைக் கேட்கும்போது, மக்கள் உடனடியாக விளையாட்டு தொடர்பான எண்கள் அல்லது கல்லூரியில் பயின்ற மிகக் குறைந்த அளவு மதிப்பெண் பெற்று தேர்ச்சி பெற்ற ஒரு பாடத்தைப் பற்றி சிந்திக்கின்றனர். புள்ளியியலை இந்நிலையில் கற்பிக்கும்போது, அதற்கு ஒரு விரிவான வாய்ப்பு இருப்பதை உணர முடிகிறது. இன்றைய காலங்களில், புள்ளியியலைப் பயன்படுத்தாத எந்த மனித நடவடிக்கைகளும் இல்லை எனலாம். புள்ளியியலில் விளக்கப் புள்ளியியல் (விவரிக்கும்) முறை, தீர்மானிக்கும் புள்ளியியல் (அனுமான) முறை என இரு பிரிவு முறைகள் உள்ளன. மேலும் இவ்விரு முறைகளும் முக்கியமானவை மற்றும் வெவ்வேறான நோக்கங்களைக் கொண்டுள்ளன.

விளக்க புள்ளியியல் முறையானது பெரிய அளவில் உள்ள தரவுகளைச் சுருக்கி தொகுத்தளிப்பதில் பயன்படுகிறது. உதாரணமாக கூட்டுச்சராசரி, இடைநிலை அளவு, முகடு, மையபோக்கு அளவைகள் விளக்க புள்ளியியல் ஆகும். விளக்க புள்ளியியலானது, சில குறிப்பிட்ட பிரச்சனைகளுக்குத் தீர்வு காண்பதற்காக, தரவுகளைச் சுருக்கிய வடிவில் தந்து விளக்குகிறது. இவை பெறப்பட்ட தரவுகளில் மட்டுமே செயல்படக்கூடியவைகளாகும். அதைத்தாண்டி செயல்பட கூடியவைகள் அல்ல.

மாறாக தீர்மானிக்கும் புள்ளியியலானது, பெருமளவிலான முழுமை தொகுதியிலிருந்து பெறப்பட்ட மாதிரிகளிலிருந்து அர்த்தமுள்ள சில முடிவுகளை எடுப்பதற்குப் பயன்படுத்தப்படுகிறது.

உதாரணமாக, ஒருவர் தாம் கண்டுபிடித்த ஒரு மருந்தானது, புராதன மருந்தைவிட, ஒரு குறிப்பிட்ட நோயைக் குணப்படுத்துவதில் சிறந்ததாக இருக்கிறதா என்பதனைக் கண்டறிய சோதனை நடத்தலாம். அதற்காக அக்குறிப்பிட்ட நோயால் பாதிக்கப்பட்ட ஒவ்வொரு நோயாளிக்கும் அம்மருந்தைக் கொடுத்து சோதிக்க முடியாது. ஆனால் அதை ஒரு தேர்வு செய்யப்பட்ட மாதிரி தரவுகளில் சோதனை செய்து நிரூபிக்கலாம். ஒரு தர கட்டுப்பாட்டு பொறியியலாளர் ஒரு தொழிற்சாலையில் தயாரிப்புகளின் தரத்தை அறிவதில் ஆர்வம் காட்டலாம். அதற்கு அவர் ஒர் ஆற்றல்மிக்க (சிறப்பான) முறையான ஏற்றுக்கொள்ளக்கூடிய மாதிரியைக் கொண்டு தயாரிப்பாளர் மற்றும் நுகர்வோர்களின் உணர்வுகளைப் பாதுகாக்கலாம். ஒரு வேளாண் விஞ்ஞானி உரங்களின் செயல்பாடுகளின் தன்மையை உணர்ந்துகொள்ள வடிவமைக்கப்பட்ட சோதனைகள் மூலம் அவற்றை சோதித்துத் தெரிந்து கொள்கிறார். இதன்மூலம் பண்ணையின் பரப்பு, பயன்படுத்தப்பட்ட

Ch1_11th Stats TM.indd 6 07/12/2021 12:35:19

நிலம், அறுவடை செய்யப்பட்ட பயிரின் அளவு போன்றவற்றையும் ஆர்வமிருப்பின், அறிந்துகொள்ளலாம்.

புள்ளியியல் முறையைப் பயன்படுத்துவதிலுள்ள மற்றொரு சிறப்பு என்னவெனில், ஒருவர் புள்ளியியலை அறிவியலின் எந்தவொரு பிரிவோடும், தொழில்நுட்பம் அல்லது பொருளாதாரம், வணிகம், பொறியியல், மருத்துவம், குற்றப்புலனாய்வு போன்ற பல சமூக அறிவியலோடும் சேர்த்துப் பயன்படுத்தலாம். புள்ளியியல் துறை அலுவல் என்பது பரபரப்பானது, சவாலானது மற்றும் வெகுமதியளிப்பதாகும். புள்ளியியல் வல்லுநர் என்பது மிக கௌரவமான பட்டமாகும். ஆனால் இடர்பாட்டுநிலை பகுப்பாய்வர், தரவு ஆய்வர், வணிக பகுப்பாய்வர் போன்றோர் புள்ளியியல் சார்ந்த துறைகளில் ஈடுபட்டுள்ளனர். புள்ளியியலிற்குப் பெருகி வரும் வாய்ப்பைக் கருத்தில்கொண்டு இந்திய பல்கலைக்கழகங்களும் மற்றும் பல்வேறு இடங்களில் உள்ள பல்கலைகழகங்களும் இளங்கலை மற்றும் முதுகலை பட்ட படிப்புகளில் புள்ளியியல் பாடங்களை பயிற்றுவிக்கின்றன. புள்ளியியலானது எல்லா வகையான துறைகளிலும் பயன்படுகிறது என்பதை ஏற்கனவே குறிப்பிட்டிருந்தோம். இப்பகுதியில் சில தேர்வு செய்யப்பட்ட துறை பிரிவுகளில் இதன் பயன்பாட்டைக் குறிப்பிட்டு காட்டுகிறோம்

1.4.1 புள்ளியியல் மற்றும் நிபுணத்துவ அளவீட்டு அறிவியல் (Statistics and actuarial science)

காப்பீட்டு அறிவியல் என்ற ஒரு அறிவியலானது புள்ளியியல் முறைகளை மிக அதிக அளவில் மற்ற எந்த பிரிவுகளைவிடவும் ஆயுட்காப்பீட்டு திட்டத்திலும், நிதி நிறுவனங்களிலும் பயன்படுத்தப்படுகின்றது. கல்வி மற்றும் அனுபவவாயிலாகவும் நிபுணத்துவ அளவீட்டு அறிவியலில் தகுதிபெற்றவர்கள் நிபுணத்துவ வல்லுநர்கள் என்றழைக்கப்படுகின்றனர். முன்பு ஒரு தீர்மானிக்கப்பட்ட மாதிரியையே காப்பீட்டு திட்டங்களுக்கான தவணைத் தொகையைத் தீர்மானிப்பதில் பயன்படுத்தினர். இன்றைய நாட்களில், நவீன கணினிகள் மற்றும் புள்ளியியல் முறைகள் இவ்வறிவியலை மிகப்பெரிய அளவில் வளர்த்திருக்கிறது. இந்தியாவில், 2006ஆம் ஆண்டிலிருந்து ஓர் அமைப்பு நிபுணத்துவ அளவீட்டு கல்வியை கொடுப்பதிலும், அதில் பயில்வோரின் திறமையை வளர்ப்பதிலும் கவனம் செலுத்துகிறது.

1.4.2 புள்ளியியல் மற்றும் வணிகம்

புள்ளியியல் முறைகள் வணிகத்தில் பெரும் அளவில் பயன்படுத்தப்படுகின்றன. மேலும் பண பரிவர்த்தனை பகுப்பாய்வு, சந்தை ஆய்வு மற்றும் மனித ஆற்றலைத் திட்டமிடல் போன்ற பல வர்த்தக தீர்வுகளுக்கும் புள்ளியியல் முறைகள் பயன்படுகிறது.

எந்த வகையான நிறுவனமாக இருப்பினும், பாகுபாடின்றி ஒவ்வொரு வர்த்தக நிறுவனமும் அதன் வளர்ச்சிக்காக புள்ளியியல் முறைகளைப் பயன்படுத்தியே ஆக வேண்டும். வர்த்தகர்கள் கடந்த கால தரவுகளிலிருந்தும், புள்ளியியல் முறைகளிலிருந்தும், பொருட்களின் விலை, வாங்குதல், விற்றல், இறக்குமதி, ஏற்றுமதி ஆகியவற்றின் போக்கைக் கணிக்கிறார்கள்.

யா–லூன்–சௌ என்ற புள்ளியியல் ஆசிரியர்; இன்றைய நாட்களில், "வர்த்தகத்தில் ஒவ்வொரு முடிவும் புள்ளியியல் தரவுகள் மற்றும் புள்ளியியல் முறைகள் ஆகியவற்றின் உதவியால்தான் எடுக்கப்படுகிறது" என்று கூறுகிறார்.

வெற்றிக் கதை

2004ஆம் ஆண்டு அமெரிக்கா ஐக்கிய நாடுகளின் கடற்கரை பிரதேசங்களை சாண்டி (Sandy) என்ற பெயரிடப்பட்ட புயல் கடுமையாகத் தாக்கியது. அதனால் பல்லாயிரக்கணக்கான மக்கள் அங்கு நிலவிய மோசமான பருவ நிலையாலும், மின்சாரம் இல்லாததாலும் பெரிதும் துன்பமுற்றனர்.

அச்சமயத்தில்தான் அப்பகுதியில் சில்லறை வணிகத்தில் ஈடுபட்டுள்ள புகழ் பெற்ற ஒரு பன்னாட்டு நிறுவனம் தமது தரவுத்தளத்தை ஆய்வு செய்த போது வியக்கத்தக்க முடிவுகளைக் கண்டறிந்தது. அதில் அப்பகுதி மக்களுக்கு அவசரகால கருவிகளும், உறைநிலை உணவுப் பொருட்களும் அதிகம் தேவைப்படுகிறது எனத் தெரிய வந்தது.

இம்முறையில் பகுப்பாய்வுகளை நடத்திய அப்பன்னாட்டு நிறுவனம் 2012 ஆம் ஆண்டு, பிரான்ஸ் (France) என்று பெயரிடப்பட்ட மற்றொரு புயல் வருவதை முன்கூட்டியே கணித்து, அதற்கேற்ப தனது வணிக யுத்தியை வகுத்துக்கொண்டது. முன்னேற்பாடாக, அந்நிறுவனம் அதிக அளவில் அவசரகால கருவிகளையும், உறைநிலை உணவுப் பொருட்களையும், பேரிடர் நிகழும் காலங்களில் தேவைப்படும் மற்ற பொருட்களையும் ஹரிக்கேன் புயல் பாதிப்பிற்கு உட்பட்ட பகுதியிலுள்ள எல்லா விற்பனை நிலையங்களுக்கும் அனுப்பி வைத்தது. அதனால் அந்நிறுவனத்தின் விற்பனை, மற்ற எல்லா நிறுவனங்களை விட மிகமிக அதிக அளவில் அமைந்ததால் விற்பனையில் மிகப்பெரும் வெற்றியைப் பெற்றது.

அதற்கு மிக முக்கிய காரணமாக இருந்தது, அப்பன்னாட்டு நிறுவனம் பெரும் தரவு பகுப்பாய்வைத் (Bigdata Analysis) தக்க முறையில் பயன்படுத்திக் கொண்டதேயாகும். இதன் மூலம் கணிப்புகளை மேற்கொள்வதால் அப்பன்னாட்டு நிறுவனங்கள், நேரம், செலவு, உழைப்பு போன்றவற்றைக் குறைத்து தங்கள் விற்பனையைப் பெரிதும் பெருக்கிக் கொள்ள உதவுகிறது.

1.4.3 புள்ளியியல் மற்றும் பொருளாதாரம்

கடைபிடிக்க வேண்டிய கொள்கைகள் மற்றும் பொது நிதிக்கொள்கைகள் போன்ற பொருளாதார விவரங்களைப் புரிந்துகொள்ள புள்ளியியல் முறைகள் மிகவும் பயனுள்ளவைகளாக உள்ளன. இந்த நவீன உலகில், பொருளாதார பாடமானது, நாட்டின் ஒரு சரியான சேவைக்கான பாடமாகவே நடத்தப்படுகிறது. இம்முயற்சி புள்ளியியலின் மிக அதிக அளவிலான பயன்பாட்டினால் உருவாக்கப்படுகிறது.

பொருளாதார பகுப்பாய்வில் பயன்படுத்தப்படும் சில முக்கியமான புள்ளியியல் முறைகள் (நுட்பங்கள்) காலத்தொடர்கள், குறியீட்டு எண்கள், மதிப்பீட்டு கருத்துருவாக்கல், வேறுபாட்டின் முக்கியத்துவத்தை மிகைக்காண் சோதனைகளை அறியும் சோதனைகள், வாய்ப்பு சார்ந்த மாதிரி வடிவம் போன்றவைகளாகும். என்பெர்க் (Enberg) அவர்களின் கூற்றுப்படி "எந்தவொரு பொருளாதார வல்லுநரும், புள்ளியியல் தரவுகளை விரிவாக ஆராயாமல் நாட்டின் வளத்தைப் பெருக்குவது அல்லது விநியோகிப்பது பற்றிய முடிவுகளை எடுப்பதற்கு முயற்சிப்பதில்லை."

Ch1_11th Stats TM.indd 8 07/12/2021 12:35:19

நமது நாட்டில், பல மாநிலங்களில், அம்மாநிலங்களில் பெறப்பட்ட பொருளாதார தரவுகளை ஆராய்ந்து முடிவெடுப்பதற்காக வளர்ச்சியடைந்த துறையாக, பொருளாதார மற்றும் புள்ளியியல் துறையை நிறுவியிருக்கின்றன

1.4.4 புள்ளியியல் மற்றும் மருத்துவம்

மருத்துவ துறையில், புள்ளியியல் முறைகள் பெரிதும் பயன்படுத்தப்படுகின்றன. மருத்துவ வெளியீடுகளைக் கவனித்துப் பார்த்தால், புள்ளியியல் முறைகள் மருத்துவத்தில் எந்த அளவிற்கு ஒரு முக்கியமான பங்கு அளிக்கிறது என்பதை உணர முடிகிறது. மருத்துவ புள்ளியியல், மிகைக்காண் சோதனைகள், நம்பிக்கை இடைவெளிகள் மற்றும் பெருவாரியாக விரைவாக பரவுகின்ற நோய் மற்றும் பொதுநலம் உள்ளடக்கிய உடல்நல அறிவியலில் பயன்படுகிறது. நவீன புள்ளியியல் முறைகள், நாள்பட்ட வியாதியால் பாதிக்கப்பட்ட ஒருவர் இன்னும் எவ்வளவு காலங்கள் உயிரோடிருப்பார் என்றும், எந்தெந்த காரணிகள் அவரின் இறப்பிற்குக் காரணங்களாக இருக்கின்றன என்பது போன்ற ஒரு மருத்துவரின் கேள்விகளுக்கு விடையளிக்கக் கூடியதாக உள்ளன.

உடல்நலம் காப்பதில் புரட்சி

– The Economist, Feb 03,2018

இணையம் மூலமாக தமக்குப் பொருத்தமான மருத்துவப் பரிந்துரைகளைப் பெறுவது பற்றி நாம் அறிந்திருக்கிறோம். தற்போதுள்ள தொழில்நுட்ப மேம்பாடுகளினால் நமது கைபேசி வழியாகவே நமது நடல்நலத்தைப் பேணும் வழிமுறைகளை அறிந்து கொள்ளலாம்.

நமது முந்தையை மருத்துவக் குறிப்புகளுடன் விடுபட்ட தேவையான மருத்துவ குறிப்புகளை கைபேசியின் மூலமாக பதிவு செய்து வழங்குவதன் மூலம் நோயின் தன்மையை கண்டறியும் திறன் பன்மடங்கு பெருகுகிறது. அதனால் திறமையின்மை குறைக்கப்பட்டு, சிக்கலான மருத்துவ நிரல்களுக்கு தீர்வு காணப்படுகிறது. இது போன்ற மருத்துவ செயலிகள் மூலம் நம் உடல் நலம் பேணிக் காப்பது பற்றி நாமும் சிந்திக்கலாமே...

1.4.5 புள்ளியியல் மற்றும் வேளாண்மை

பரிசோதித்தல் அதன் அடிப்படையில் உணர்தல் என்பது பொதுவான அறிவியல் முறையின் முக்கியமான பண்புகளாகும். வேளாண்மை அறிவியல் முறையிலும் சோதனைகள் நடத்தி ஒரு குறிப்பிட்ட வகை பயிர் மற்ற எந்த வகைப்பயிரை விடவும் அதிக விளைச்சலைக் கொடுக்கிறதா என்பதை உணர்ந்து முடிவுகள் எடுக்கப்படுகிறது. ஒரு குறிப்பிட்ட வகை உரம் மற்ற எந்த வகையான உரங்களைவிடவும் அதிக பலன் அளிக்கக்கூடியதா? என்பதை உணர்ந்து முடிவுகள் எடுக்கப் பயன்படுகிறது.

விவசாய ஆராய்ச்சியில் ஈடுபடும் பல நிறுவனங்கள் உள்ளன. அவை மாறுபாட்டளவை பகுப்பாய்வு போன்ற சோதனை வடிவமைப்புகளில் புள்ளியியல் முறைகளைப் பயன்படுத்துகின்றன. விவசாய புள்ளியியலில் ஆய்வுகள் செய்வதற்கென்றே (IASRI) என்ற நிறுவனம் புது தில்லியில் உள்ளது.

1.4.6 புள்ளியியல் மற்றும் தொழில்துறை

அறிவியல் மற்றும் தொழில் நுட்பத்தின் நவீனமயமான பயன்பாட்டிற்குப் புள்ளியியல் முறைகள் முக்கிய பங்கு வகிக்கிறது. பல புள்ளியியல் முறைகள் கண்டுபிடிக்கப்பட்டுள்ளன. மேலும் அவை தொழிற்துறையில் ஏற்படும் பல்வேறுபட்ட பிரச்சனைக்கு தீர்வுகாண பயன்படுகிறது. உதாரணமாக, தயாரிக்கப்படும் பொருளின் தரத்தை பராமரிக்க (நீட்டிக்க) புள்ளியியல் தரக்கட்டுப்பாடு முறைகளைப் பயன்படுத்துகின்றன.

Ch1_11th Stats TM.indd 9 07/12/2021 12:35:19

காலச்சூழலில் தரம், எந்திரங்கள், மின்பொருட்கள், மின்னணு பொருட்கள் ஆகியவற்றில் நம்பகத்தன்மை அறியும் முறை விரிவடைந்திருக்கிறது. இவைகளில் புள்ளியியல் முறைகளே பயன்படுத்தப்படுகிறது. மொத்த தர மேலாண்மை மற்றும் சிக்ஸ் சிக்மா கருத்துரு (six sigma theory) போன்றவற்றில் புள்ளியியல் விதிகளே பயன்படுத்தப்படுகிறது.

1.4.7 புள்ளியியல் மற்றும் தகவல் தொழில் நுட்பம்

தகவல் தொழில்நுட்பத்தின் மூலம் கணினி மற்றும் தொலைத்தொடர்பு சாதனங்களைப் பயன்படுத்தி தரவுகளை சேகரிக்கவும், திரும்பப் பெறவும், பரிமாற்றம் மற்றும் செயல்படுத்தவும் முடிகிறது. இந்நாட்களில், பெரும்பாலான தொழிற்சாலைகள் தகவல் தொழில் நுட்பத்துடன் இணைந்திருக்கின்றன. மேலும் பெருமளவிலான தரவுகள் ஒவ்வொரு நாளும் சேகரிக்கப்படுகின்றன. இவ்வகையான தரவுகள் சரியான முறையில் பகுப்பாயப்படுவதால், இத்தரவுகளிலுள்ள தகவல்களை பயன்பாட்டாளர்கள் பயன்படுத்த ஏதுவாகிறது. ஆய்வு களத்தைப் பற்றி அறிந்துகொள்ள, தரவு எடுத்தல் (Data mining) என்பது வலைதளத்தில் உருவாக்கப்பட்டுள்ளது. தரவு எடுத்தல் (Data mining) என்பது கணினி அறிவியலின் உள்ளடங்கிய ஒரு பகுதியாகும். இது பெரும்அளவிலான தரவு தொகுப்பிலிருந்து, குறிப்பிட்ட முறைகளை தேர்ந்தெடுப்பதற்கான கணக்கிடும் முறையாகும். இம்முறையில் செயற்கை நுண்ணறிவு மற்றும் புள்ளியியல் விபரங்களும் பயன்படுத்தப்படுகின்றன. புள்ளியியல் முறைகளில் பயிற்சியும், கணினி அறிவும் பெற்றவர்கள் தரவு பகுப்பாய்வாளர்களாக பெருமளவிலான தரவுகளை ஆராய்வதற்காகப் பணிபுரிகின்றனர்.

1.4.8 புள்ளியியல் மற்றும் அரசு நிர்வாகம்

புள்ளியியலானது, புள்ளியியல் தரவுகளை அரசுக்கு அளித்து, அரசின் திட்ட வளர்ச்சியைக் காணவும், சட்டம் ஒழுங்கைப் பேணவும், நலத்திட்டங்களை உருவாக்கவும், அரசின் மற்ற பிற திட்டங்களை மதிப்பிடவும் பயன்படுகிறது. குறிப்பாக, புள்ளியியல் தகவல்கள், மாநிலத்தின் மொத்த ஆளுமைக்கும் இன்றியமையாததாகும். உதாரணமாக, புள்ளியியலானது மக்கட்தொகை, விவசாய உற்பத்தி, தொழில் உற்பத்தி, வளம், இறக்குமதி, ஏற்றுமதி, குற்றங்கள், பிறப்பு விகிதம், வேலையின்மையின் அளவு, கல்வி, கனிமவளம் போன்ற இன்னும் பலவற்றைப் பற்றிய தகவல்களை அரசிற்குக் கொடுக்கிறது.

1.5 பெரும் தரவுகள் (Big Data)

மிகவும் அதிக எண்ணிக்கையால் ஆன தரவு உறுப்புகளைக் கொண்டதும் சிக்கலானதுமான தரவு அமைப்புகள் பெருந்தரவுகள் எனப்படும். அப்பெரும் எண்ணிக்கையிலான தரவு உறுப்புகளைச் சேமித்து வைப்பதும் பிறகு அவற்றை வழக்கமான தரவுத்தள மென்பொருள்களைக் கொண்டு தகவல்களைப் பெறுவதும் எளிதன்று. தினமும் மில்லியன் அளவிலான தகவல் தரவுகள் பதிவேற்றம் செய்யப்படுகின்றன. உலகத்தின் தகவல் தொகுப்புகளில் சுமார் 90% அளவிற்கு சமீப ஆண்டுகளில் தான் உருவாக்கப்பட்டுள்ளன.

இன்றைய காலம், மகத்தான தகவல்கள் மூலம் இயக்கப்படுகிறது. எனவே தரவுகளைக் கையாளும் முறையும் அதனை விளக்குவதற்குமான வல்லுனர்களும் அதிகபடியான அளவில் தேவைப்படுவர். ஆதலால் வருங்காலத்தில் இது எல்லோரையும் ஈர்க்கும் ஒரு முக்கிய துறையாக மாறும்.

பெரும்தரவுகளின் பயன்பாடுகள்

மக்களைக் கருதாமல் பெருந்தரவுகளைப் பற்றி பேசுவது சரியல்ல. ஏனெனில் பெருந்தரவுகளால் மக்கள் பெறும் பயன்கள் அதிகமாகும். பொதுவாக இன்றைய நிலவரங்களில்

Ch1_11th Stats TM.indd 10 07/12/2021 12:35:19

எல்லா தொழிலகங்களும் பெரும் தரவுகளைப் பயன்படுத்த ஆரம்பித்துவிட்டன. இனி பெரும் தரவுகள் பயன்படுத்தப்படும் சில துறைகளைக் காண்போம்.

உடல் நலம் பேணும் துறை: பல்வேறு நோயாளிகளின் உடல்நிலை சம்பந்தமான தரவுகள் கணினிகளில் பீட்டாபைட் (Petabyte) கணக்கில் சேமிக்கப்பட்டிருக்கும். அத்தரவுகளிலிருந்து மருத்துவம் சார்ந்த நிறுவனங்கள், தமக்கு வேண்டிய பயனுள்ள தகவல்களைப் பெற்று நோயாளிகளின் நலம் சார்ந்த செயல்களுக்கு முன்பாகவே உருவாக்கம் செய்ய முடியும்..

சில்லறை விற்பனை செய்யும் பெரும் நிறுவனங்கள்: பெரும் தரவுகளால் மிகப்பெரிய பயனை அடைவோர் சில்லறை விற்பனையில் ஈடுபடும் பெரும் நிறுவனங்களே. இப்பெரும் தரவு விவரங்களிலிருந்து நுகர்வோரின் மனப்பாங்கை அறிந்து, அவர்களின் மனநிலைக்கேற்ப பொருட்களை வழங்கி, விற்பனையைப் பெருக்குகின்றனர்.

உற்பத்தித்துறை: பெரும் தரவுகளைப் பகுப்பாய்வு செய்து, அதன்மூலம் எந்திரப் பொருட்களின் உதிரிபாகங்களிலுள்ள குறைகளைக் குறைக்கவும், பொருட்களின் தரத்தை உயர்த்தவும் திறன் மேம்பாடு அடையவும், நேரம், பணம் ஆகியவற்றை மிச்சப்படுத்தவும் உதவுகிறது.

போக்குவரத்து கண்காணிப்பு: உலக அளவில் பெரிய நகரங்களில் காணப்படும் போக்குவரத்து நெரிசலைச் சமாளிப்பது சவால் நிறைந்ததாகும். அதிக அடர்த்தியுள்ள மக்கள் தொகையுள்ள நகரங்களில் சென்சார்களைக் (sensor) கொண்டு, பெரும் தரவுகளைப் பெற்று போக்குவரத்து நெரிசல் தீர்க்கப்பட்டு மேம்படுத்தப்படுகிறது.

தேடல் தரம்: கூகுல் (google) தேடுபொறி மூலம் தேடும் ஒவ்வொரு தகவலுக்கும் நாம் அச்சொல்லை கூகுளுக்குள் சேர்த்து வைக்கிறோம். பெரும் தரவுகளுள் ஒன்றாக இதையும் சேமித்து அடுத்த முறை தேடும்போது தேடும் தரத்தை உயர்த்துகிறது.

விற்பனை பெருக்கத்திற்கான விளம்பரம்: மிகச் சிறந்த விளையாட்டு வீரர்கள், நடிகர்கள், புகழ்வாய்ந்தவர்கள் பற்றிய செய்திகள், செய்தித்தாள்கள், சமூக ஊடகங்கள், வலைதளங்களில் இடம் பெறுவதால் அவர்களை பெரும் நிறுவனங்கள் தத்தம் விளம்பரத் தூதுவர்களாக நியமித்து தங்கள் வணிகத்தைப் பெருக்குகின்றன.

பெரும் தரவுகளுக்குள்ள சவால்கள்

பெரும் தரவுகளால் பெரும்பயன் பெற்றாலும் சில சவால்களைச் சந்திக்க வேண்டியுள்ளது. அவை தரவுகளைச் சேமிப்பதற்கான மிகப்பெரும் சேமிக்கும் இடம், தரவுகளின் தரம், பகுப்பாய்வுத் திறன், திறமையின்மை போன்றவையாகும்.

இச்சவாலை சமாளிக்கும் விதமாக ஹடூப் (Hadoop) என்ற மென்பொருள் பெருமளவில் உதவிபுரிகிறது. இது ஜாவா மொழியை அடிப்படையாகக் கொண்ட திறந்த வகையிலான நிரல்களைக் கொண்ட அமைப்பாகும். இம்மென்பொருள் அமைப்பு பரவலான பங்கீட்டு முறை சூழ்நிலையில் கணக்கிடுவதற்கும் ஏராளமான தரவு கணங்களைக் கையாள்வதற்கும் ஏற்றதாகும்.

1.6 தரவுகளின் பொருளும் வகைகளும்

முடிவு எடுப்பதற்காக, சேகரிக்கப்பட்ட தகவல்கள் குறிப்பாக விவரங்கள் (அல்லது) தரவுகள் எனப்படும். தரவுகள் எண் சார்ந்தவையாகவோ தீர்மானமானவையாகவோ (வகைப்படுத்தப்படக்கூடிய) இருக்கலாம். தரவுகள் ஒரு மாறி சார்ந்தும் உருவாக்கப்படலாம்.

மாறி: இடத்திற்கு இடம், நபருக்கு நபர், முயற்சிக்கு முயற்சி மேலும் பல சூழல்களில், மாறக்கூடிய ஒன்று மாறி எனப்படும். உதாரணமாக, உயரம் ஒரு மாறி, வசிப்பிடம் ஒரு மாறி ஏனெனில் இவை நபருக்கு நபர் மாறுபடுகிறது.

ஒரு மாறியானது அளவிடக்கூடியதாகவும் எண் உருவில் குறிப்பிடக்கூடியதாகவும் இருப்பின் அது எண் சார்ந்த மாறி எனப்படும்.

ஒரு மாறியானது அளவிடமுடியாததாகவும், எண்ணுருவில் குறிப்பிட முடியாததாகவும் இருப்பின் அது பண்பு சார்ந்த மாறி எனப்படும். இம்மாறிகள் தீர்மானமான மாறிகள் (வகைப்படுத்தப்படக் கூடிய) என்றும் அழைக்கப்படும்.

உயரம் என்பது எண் சார்ந்த மாறி, ஏனெனில் இதை அளவிட முடியும். மேலும் எண்ணால் குறிப்பிட முடியும். அதே வேளையில், வசிப்பிடம் ஒரு பண்பு சார்ந்த மாறி ஏனெனில் இதை அளவிட முடியாது. மேலும் கிராமம் அல்லது நகரம் என்றே குறிக்க முடிகிறது. இவைகள் எண்ணால் குறிப்பிடப்படுவதில்லை என்பதை கவனத்தில் கொள்ளவும்.

எண் சார்ந்த தரவுகள்:

அளவுகள் சேகரிக்கப்பட்டு, எண்களாக பதிவிடப்பட்டிருப்பின் அம்மாதிரியான தரவுகள் எண்சார்ந்த தரவுகள் எனப்படும். உதாரணமாக உயரம், எடை போன்ற பல எண் சார்ந்த தரவுகளாகும்.

பண்புசார்ந்த தரவுகள்:

இயற்கையாக எண் முறைப்படி, அளவிட முடியாததாக உள்ள தரவுகள் பண்பு சார்ந்த தரவுகள் எனப்படும். உதாரணமாக, இரத்தத்தின் பிரிவானது Rh காரணியோடு சேர்த்து O,A,B என வகைப்படுத்தப்பட்டுள்ளது. இவை ஏற்கனவே சொல்லப்பட்டிருக்கிற அல்லது வடிவமைக்கப்பட்டிருக்கிற ஏதாவது ஒன்றாக வகைப்படுத்தப்படுகிறது.

1.7 அளவீட்டு அளவைகள்

தரவுகள் நான்கு வகை அளவீட்டு அளவைகளாக பிரிக்கப்படுகின்றன. அவை பெயரளவு (பண்பு சார்ந்த) வரிசைப்படுத்தக்கூடிய இடைவெளிகளில் கொடுக்கக்கூடிய மற்றும் விகிதாசாராத்தில் கொடுக்கக்கூடிய தரவுகள் எனப்படும். அளவீட்டு அளவைகள் ஸ்டான்லி ஸ்டீவன்ஸ் என்பவரால் உருவாக்கப்பட்டது.

Ch1_11th Stats TM.indd 12 07/12/2021 12:35:20

1.7.1 பெயரளவு அளவுகள் (Nominal Scale):

பெயரளவு அளவுகள் (பண்பு சார்ந்த அளவுகள்) ஒரு மாறியை எந்தவொரு எண் அளவு இல்லாமல் குறிப்பதற்காக கொடுக்கப்படும் அளவுகள். உதாரணமாக உங்கள் பாலினம் என்ன? என்று ஒரு வினாபட்டியலில் நாம் ஒரு கேள்வியை எழுப்பலாம். அதற்கு ஆண் இனம் அல்லது பெண் இனம் என விடையாக இருக்கும். இங்கு பாலினம் என்பது பெயரளவு மாறியாகும். மேலும் ஆண் இனம் எனில் 1 எனவும் பெண் இனம் எனில் 2 எனவும் இம்மாறிகளை எண் மதிப்போடு தொடர்புபடுத்தலாம். இவைகள் பெயரளவிலே எண்களாக குறிப்பிடப்படுகிறது. மேலும் ஒருவர் தனியாகவோ, திருமணமானவராகவோ, விதவையாக, விவாகரத்து பெற்றவராகவோ இருப்பின் முறையே 1, 2, 3, 4 என்ற எண்களால் குறிப்பிடலாம். இந்த எண்கள், நாம் அன்றாட வாழ்வில் பயன்படுத்தும் எண்களின் பண்புகள் எதையும் கொண்டிருப்பதில்லை. இங்கு $4 > 1$ or $2 < 3$ or $1+3 = 4$ என நாம் கூற முடியாது. இவற்றை வரிசைப்படுத்துவது என்பது முரணானதாகும். இவ்வகையான தரவுகளை வரிசைப்படுத்தியோ, கணித விதிகளைப் பயன்படுத்தியோ செய்யப்படும் எந்தவொரு புள்ளியியல் ஆய்வும் அர்த்தமற்றதாகும்.

1.7.2 வரிசைப்படுத்தக்கூடிய அளவுகள் :

இந்த மாதிரியான அளவுகள் எண்ணியலின் சில பண்புகளைப் பெற்றிருக்கின்றன. ஆனால் அனைத்து பண்புகளையும் பெற்றிருப்பதில்லை. உதாரணமாக மகிழுந்துகளை நாம் சிறிய, நடுத்தர மற்றும் பெரிய அளவிலான மகிழுந்து என அவற்றின் அளவுகளைப் பொறுத்து சொல்கிறோம். வரிசைப்படுத்துதலில் அளவுகளின் வரிசையே மிக முக்கியம். ஆனால் அவ்வளவுகளில் வேறுபாடு தெரிவதில்லை. கீழ்க்கண்ட எடுத்துக்காட்டைக் காண்போம்.

நேற்று வேடந்தாங்கலுக்கு நாம் சென்ற உல்லாசப் பயணம் எப்படி இருந்தது என்ற கேள்விக்கு விடையாக : (1) மிக்க மகிழ்ச்சி இல்லை (2) மகிழ்ச்சியில்லை (3) பரவாயில்லை (4) மகிழ்ச்சி (5) மிக்க மகிழ்ச்சி எனக் கிடைக்கலாம். இவ்வரிசையில் எண் 5, 4 ஐ விட சிறந்தது அல்லது 3 ஐ விட சிறந்தது என அறிவோம். ஆனால் எவ்வளவு சிறந்தது என அறிய முடிவதில்லை. உதாரணமாக பரவாயில்லை மற்றும் மகிழ்ச்சி இல்லை என்பதற்கும் உள்ள வித்தியாசமும் மிக மகிழ்ச்சி மற்றும் மகிழ்ச்சி என்பதற்கு இடையேயுள்ள வித்தியாசமும் சமம் என நம்மால் கூற இயலாது.

இதைப்போலவே மருத்துவர், ஒரு நோயாளியின் நிலையை நன்று, பரவாயில்லை, மோசம், மிகமிக மோசம் எனக் கூறலாம். மேலும் அவர் நன்றுக்கு 1 எனவும், பரவாயில்லை என்பதற்கு 2 எனவும், மோசம் என்பதற்கு 3 எனவும், மிக மோசம் என்பதற்கு 4 எனவும் எண் குறியிடலாம். நோயாளியின் நிலைமையின் அளவானது 1 இல் இருந்து 4 வரை கீழ்க்கண்டவாறு குறிக்கலாம் $1 < 2$ அல்லது $2 < 3$ அல்லது $3 < 4$ எனினும் மேல்கண்டவை நோயாளியின் நிலையை மட்டுமே குறிக்கிறது. வரிசைப்படுத்துதல் அளவில் நாம் மாறிகளின் வரிசையை மட்டுமே உணருகிறோம். இம்மாதிரியான தரவுகளில் சராசரி அளவிட முடியாது. ஆனால் இடைநிலை அளவு மற்றும் முகடு அளவிட முடியும்.

1.7.3 இடைவெளி அளவுகள்:

இடைவெளி அளவுகளில் மாறிகளுக்கு இடையே உள்ள வித்தியாசத்தை கணக்கிட முடியும். ஆனால் அவற்றின் பெருக்கல் மற்றும் வகுத்தல்களைக் காண முடியாது. வேறுவிதமாக இடைவெளி மாறிகளுக்கு இடையில் உள்ள தூரங்களை மட்டுமே அறிய முடியும்.

Ch1_11th Stats TM.indd 13 07/12/2021 12:35:20

உதாரணமாக ஃபாரன்ஹீட் அளவுகளில்: $60^\circ, 65^\circ, 88^\circ, 105^\circ, 115^\circ$, and 120° என்று வெப்பநிலை அளவுகள் கொடுக்கப்பட்டிருப்பது எனில் நாம் $105^\circ > 88^\circ$ அல்லது $60^\circ < 65^\circ$ என எழுதலாம். இவைகள் 105° என்பது 88° விட வெப்பமானது என்றும் 60° என்பது 65° ஐ விட குளிரானது என்பதைக் குறிக்கின்றன. மேலும் $65^\circ - 60^\circ = 120^\circ - 115^\circ$ ஐ விட குளிரானது என்பதைக் குறிக்கின்றன. மேலும் 650 – 600 = 1200 – 1150 என எழுதலாம். ஏனெனில் சம வெப்பநிலை வித்தியாசங்கள் சமம். அதாவது வெப்பநிலையானது 60° –லிருந்து 65° ஆவதற்கு தேவையான வெப்பம் 115° –லிருந்து 120° அளவிற்கு ஆவதற்கு தேவைப்படும் வெப்பத்திற்கு சமம் ஆதலில் 120° என்பது 60° –ல் வகுபடும்போது 2 என இருப்பினும் வெப்பநிலை இருமடங்கு அல்ல. அதற்கான காரணம் பாரன்ஹீட் அளவை செல்சியஸ் அளவில் மாற்றும்போது காணலாம்.

$$60^\circ F = \frac{5}{9}[60^\circ - 32^\circ]C = 15.570^\circ C \text{ மற்றும்}$$

$$120^\circ F = \frac{5}{9}[120^\circ - 32^\circ]C = 48.870^\circ C$$

குறிப்பு

'0' என்பது ஆதியாக இல்லாதபோது விகிதங்களை காண்பது இயலாது..

மேலே கண்ட சமன்பாடுகளில் இடதுபக்க அளவுகளில் $120^\circ F$ என்பது $60^\circ F$ இருமடங்கு என்பதும் வலது பக்க அளவுகளில் $48.87^\circ C$ என்பது $15.57^\circ C$ மூன்று மடங்கைவிட பெரியது என தெளிவாக தெரிகிறது. இந்த பிரச்சனைக்கான காரணம் என்னவெனில் பாரன்ஹீட், செல்சியஸ் அளவைகள் செயற்கை ஆதிகளைப் பெற்றிருக்கின்றன. வெப்பமின்மை என்ற ஒன்றே இல்லை.

1.7.4 விகித அளவுகள்

அளவிடக்கூடிய அளவுகள் என்று வரும்போது விகித அளவுகள் முக்கியமானவை. ஏனெனில் அவை வரிசைகளைக் கூறுகளுக்கான சரியான அளவுகளைக் குறிக்கின்றன. மேலும் பூஜ்ஜியம் என்ற மதிப்பைக் கொண்டுள்ளது. இது விளக்கப் புள்ளியியல் மற்றும் உய்த்துணர்வு புள்ளியியல் ஆகியவற்றில் பயன்படுத்தக்கூடியதாக இருக்கிறது. விகித மாறிகளுக்கு சரியான உதாரணங்களில் உயரம் மற்றும் எடை ஆகியவை உள்ளன. விகித மாறிகள் புள்ளியியல் பகுப்பாய்வுகள் விகித அளவைகள் ஏராளமான வழிகளைத் தருகிறது. இந்த மாறிகளைக் கூட்டல், கழித்தல், பெருக்கல், வகுத்தல் போன்ற கணித செயல்பாடுகளுக்குப் பயன்படுத்த முடியும். விகித அளவைகளைப் பயன்படுத்தி மைய அளவையான முகடு, இடைநிலை அளவு, சராசரி ஆகியவற்றையும் சிதறல் அளவைகளான திட்டவிலக்கம் மற்றும் மாறுபாட்டுக்கெழு போன்றவற்றைக் கணக்கிடலாம்.

தொகுப்பாக பண்பு சார்ந்த அளவைகள் குறியிடுவதற்கும் பயன்படுகிறது. வரிசை அளவைகள் வரிசையிடுதலில் நல்ல தகவல்களைத் தருகின்றன. வாடிக்கையாளர் திருப்தி ஆய்வில் இருப்பதைவிட இடைவெளி அளவுகள் வரிசைப்படுத்தலையும் மாறிகளுக்கிடையேயுள்ள வித்தியாசத்தில் அறியப்படுகின்றன. இறுதியாக விகித அளவுகள் வரிசைப்படுத்துதல், இடைவெளி மதிப்பிடுதல் விகிதங்களை கணக்கிடுதல் ஆகியவற்றிற்கு பயன்படுத்தப்படுகிறது. ஏனெனில் இங்கு பூஜ்ஜியம் வரையறுக்கப்பட்டுள்ளது. பண்பு, வரிசை, இடைவெளி மற்றும் விகித அளவைகளுக்கிடையே உள்ள வித்தியாசங்களை நாம் விரிவாகக் கண்டோம். ஏனெனில் இவற்றின் பயன்பாடுகள் கணிணியிலும் புள்ளியியல் மென்பொருள்களான SPSS, SAS, R, STATA போன்ற பலவற்றில் காணப்படுகின்றன.

Ch1_11th Stats TM.indd 14 07/12/2021 12:35:22

நினைவில் கொள்க...	
● புள்ளியியலின் குணங்கள்	ஒரு குறிப்பிட்ட நோக்கத்திற்காக திட்டமிடப்பட்ட முறையில் சேகரிக்கப்பட்ட மொத்த விவரங்கள் அதிக எண்ணிக்கையிலான காரணிகளால் ஒரு குறிப்பிடக்கூடிய அளவிற்கு பாதிக்கக்கூடியவைகள் எண் வடிவில் விவரங்களை அளிப்பது விவரங்களைச் செம்மையான அளவிற்கு மதிப்பீடு செய்தல் ஒன்றோடொன்று தொடர்புள்ள வகையில் விவரங்களை அமைத்தல்
● பெரும்தரவுகளின் பயன்பாடுகள்	சுகாதாரத் துறை, சில்லறை விற்பனை செய்யும் பெரும் நிறுவனங்கள், போக்குவரத்து கண்காணிப்பு, உற்பத்தித்துறை, தர நிர்ணயம், விற்பனை பெருக்கத்திற்கான விளம்பரம்
● நோக்கம் மற்றும் பயன்பாடுகள்	விவரங்களைச் சேகரித்தல், வகைப்படுத்துதல், சுருங்கக் கூறுதல், ஒப்பிடல், ஒட்டுறவு, காரணம் அறிதல், வாய்ப்பு அறிதல்
● நோக்கம் மற்றும் பயன்பாடுகள்	புள்ளியியல் மற்றும் வணிகம், புள்ளியியல் மற்றும் பொருளாதாரம், புள்ளியியல் மற்றும் மருத்துவம், புள்ளியியல் மற்றும் விவசாயம், புள்ளியியல் மற்றும் தொழில்துறை, புள்ளியியல் மற்றும் தகவல் தொழில் நுட்பம், புள்ளியியல் மற்றும் அரசு நிர்வாகம்
● தரவுகளின் வகைகள்	எண் சார்ந்த தரவுகள், பண்பு சார்ந்த தரவுகள்
● அளவீட்டு அளவைகள்	பெயரளவு அளவுகள், வரிசைப்படுத்தக்கூடிய அளவுகள், இடைவெளி அளவுகள், விகித அளவுகள்

பயிற்சி

QLG9C

I. மிகச் சரியான விடையைத் தெரிவு செய்க:

1. ஒரு தொழிலாளி ஒரு வருடத்தில் எடுத்த விடுமுறை நாட்களின் எண்ணிக்கை

   (a) பண்பு அளவு (b) வரிசை அளவு (c) இடைவெளி அளவு (d) விகித அளவு

2. நிறத்தைப் பொறுத்து வகையிடக்கூடிய தரவு என்பது

   (a) பண்பு அளவு (b) பெயரளவு (c) இடைவெளி அளவு (d) விகித அளவு

3. படங்களைத் தரமிடுதல் என்பது

   (a) பண்பு அளவு (b) வரிசை அளவு (c) இடைவெளி அளவு (d) விகித அளவு

4. மருத்துவமனையில் இருந்த நேரத்தில் ஒரு நோயாளியின் வெப்பநிலை 100^{0}F என்பது

   (a) பண்பு அளவு (b) வரிசை அளவு (c) இடைவெளி அளவு (d) விகித அளவு

Ch1_11th Stats TM.indd 15 07/12/2021 12:35:22

II. கோடிட்ட இடங்களை நிரப்புக:

5. புள்ளியியல் ____________________ ஐ சார்ந்தது

6. புள்ளியியலை ____________________ மற்றும் ____________________ என்று வகைப்படுத்தலாம்

7. பண்டைய காலப் புள்ளியியலில் ____________________ அளவை பயன்படுத்தப்பட்டது.

8. மாணவர்களின் வயது ____________________ மாறியாகும்.

9. இந்திய புள்ளியியல் நிறுவனத்தை (ISI) தோற்றுவித்தவர் ____________________

10. உன் நகரத்தின் தட்பவெப்ப நிலை ____________________ அளவீட்டு அளவையாகும்.

11. ஒரு மாணவரின் அறிவுத்திறன் விகிதம் ____________________ மாறியாகும்.

12. ஒரு காரின் நிறம் ____________________ அளவீட்டு அளவையாகும்.

13. ஒரு நாடாளுமன்ற பிரதிநிதியின் பெயரைக் குறிப்பிடுவது ____________________ அளவீடு.

14. ஒரு விற்பனையாளரின் செயல்பாடு சிறந்தது. அச்சிறந்த என்பது ____________________ அளவீடு.

III. குறுவினா (ஓரிரு சொற்களில் விடையளி):

15. புள்ளியியல் என்பதை வரையறு.

16. தரவுகள் என்றால் என்ன?

17. நிபுணத்துவ மதிப்பீட்டு அறிவியலில் புள்ளியியல் பயன்பாடுகளைக் கூறுக.

IV. சிறுவினா (ஓரிரு சொற்றொடர்களில் விடையளி):

18. கிராக்ஸ்டன் மற்றும் கௌடன் அவர்களின் புள்ளியியல் வரையறையை எழுதுக.

19. புள்ளியியலின் குணங்களைப் பட்டியலிடுக?

20. பண்புசார்ந்த மற்றும் எண்சார்ந்த மாறிகளைப் பற்றி நீவீர் அறிவன யாவை?

V. விரிவாக விடையளி:

21. P.C. மஹலனோபிஸ் அவர்களின் பங்களிப்பைப் பற்றி ஒரு குறிப்பு வரைக.

22. புள்ளியியலின் தோற்றமும் வளர்ச்சியும் பற்றி ஒரு குறிப்பு வரைக

23. புள்ளியியலின் பணிகளைப் பற்றி விளக்குக.

24. விவசாயம் மற்றும் தொழில்துறையில் புள்ளியியலின் பயன்பாடுகளைப் பற்றி விளக்குக.

விடைகள்

I. 1. d 2. b 3. b 4. c

II. 5. எண் விவரங்கள் 6. ஒருமை மற்றும் பன்மை 7. எண்ணுதல் 8. எண் சார்ந்த

9. P.C.மஹாலனோபிஸ் 10. இடைவெளி 11. எண் சார்ந்த

12. பெயரளவு 13. பெயரளவு 14. வரிசைப்படுத்தக்கூடிய

Ch1_11th Stats TM.indd 16 07/12/2021 12:35:22