(i) ডেটা : প্রায় সমস্ত DBMS-এ যে ধরনের ডেটা রাখা যায় সেগুলি হল পূর্ণসংখ্যা, ভগ্নাংশ, অক্ষর বা টেক্সট, তারিখ এবং যুক্তিমূলক ডেটা (True বা False type) । কয়েকটি DBMS-এ ছবি এবং মাল্টিমিডিয়া ডেটাকেও রাখা যায়।

(ii) সাধারণ তথ্য বিশ্লেষণ : প্রায় সমস্ত DBMS-এ ডেটা বিশ্লেষণের সাধারণ পদ্ধতিগুলি থাকে, যেমন—Sort করা অর্থাৎ ছোটো থেকে বড় বা বড় থেকে ছোটো হিসাবে সাজানো, search করা বা নির্দিষ্ট ডেটা খুঁজে বের করা, সেগুলি filter করা অর্থাৎ খুঁজে একজায়গায় জড়ো করা ইত্যাদি। এগুলি সাধরণত কম্যান্ড হিসাবেই থাকে। এই কাজগুলি মনিটরে দেখতে পাওয়া যায় বলে ব্যবহারকারীর খুব সুবিধা হয়।

(iii) ডেটা কম্যানিগুলেশন লেঙ্গুয়েজ : ডেটাবেস থেকে প্রয়োজনীয় তথ্য বের করা, বিশ্লেষণ করা এবং উপস্থাপনা করার জন্য যে ল্যাঙ্গুয়েজ ব্যবহার করা হয়, সেটি হল DML। এই DMLএক ধরনের query ল্যাঙ্গুয়েজ। আজকাল প্রায় সমস্ত DBMS-এ SQL (Structuered Query Language) ব্যবহার করা হয়।

(iv) প্রোগ্রামিং ল্যাঙ্গুয়েজ : কয়েকটি DBMS-এ Query ল্যাঙ্গুয়েজ ছাড়াও, রেকর্ড স্তরে নিয়ন্ত্রণের জন্য এবং অ্যাপ্লিকেশন প্রোগ্রাম তৈরী করার জন্য অন্য আর একটি অবজেক্ট ওরিয়েন্ট ল্যাঙ্গুয়েজ থাকে।

(v) ফাইল স্ট্রকচার : ডেটাবেস থেকে যাতে প্রয়োজনীয় ডেটা দ্রুত বের করা যায় তার জন্য সমস্ত ডেটাকে ফাইলের মধ্যে বিশেষ পদ্ধতিতে রাখতে হয়। প্রত্যেক DBMS-এর এরকম নিজস্ব পদ্ধতি আছে। তবে বেশীরভাগ BDMS সফটওয়্যার ওপেন ডেটাবেস কালেক্টিভিটি (ODBC) নামের একটি বিশেষ সংক্ষেপ ব্যবস্থার সাহায্যে অন্য ধরনের DBMS-এর সঙ্গে ডেটা আদান-প্রদান করতে পারে।

সম্পর্কযুক্ত ডেটাবেস বা রিলেশনাল ডেটাবেস : রিলেশনাল ডেটাবেস হল অনেকগুলি টেবিলের একটি সট। এই একাধিক টেবিলকে একসঙ্গে জুড়ে বা relate করে প্রয়োজনীয় তথ্য বের করার এটি একটি পদ্ধতি। সাধারণত, ডেটাবেস তৈরী করার সময় ডেটাগুলিকে এমনভাবে বিভিন্ন টেবিলে ভাগ করে রাখা হয় যাতে একই ডেটা একাধিকবার এণ্ট্রি করতে না হয়। এই সমস্ত টেবিলে ছড়িয়ে থাকা ডেটা থেকে সম্পূর্ণ ডেটা পেতে হলে প্রয়োজনমত একাধিক টেবিলকে জুড়ে নিতে হয়। দুটি টেবিল জোড়ার জন্য দুটি টেবিলেই একটি common field থাকা দরকার, যাকে বলা হয় Key field। একটি টেবিল আর একটি টেবিলের সাথে জুড়ে দেওয়ার প্রক্রিয়াই হল relationship বা সম্পর্ক স্থাপন। ডেটাবেসের বিভিন্ন টেবিলের মধ্যে নির্দিষ্ট সম্পর্ক বা relatonship থাকে বলেই এই ধরনের ডেটাবেসকে বলা হয় রিলেশনাল ডেটাবেস।

রিলেশনার ডেটাবেস যে সফটওয়্যারের সাহায্যে তৈরী ও ব্যবহার করা হয় তাকে বলা হয় রিলেশনাল ডেটাবেস ম্যানেজমেন্ট সিস্টেম (RDBMS)। এখানকার প্রায় সমস্ত উল্লেখযোগ্য BDMS সফ্টওয়্যার আসলে RDBMS সফ্টওয়্যার, যেমন—Oracle, মাইক্রোসফট SQL Server, Informise, Microsoft Access ইত্যাদি।

(ii) ফিল্ড ও রেকর্ড : এক একটি টেবিলে অনেকগুলি রো (row) বা সারি ও কলাম (column) থাকে। এই প্রতিটি রো-কে বলা হয় **রেকর্ড** ও প্রতিটি কলামকে বলা হয় **ফিল্ড**। যে বিষয়ের জন্য টেবিলটি তৈরী, সেই সম্পর্কীত ডেটা এক একটি রেকর্ডে থাকে। অন্যদিকে ডেটা আইটেমগুলি টেবিলের কলামে রাখা হয়।

বস্তুত : ডেটবেস -এ সমস্ত ডেটাকে কেন্দ্রীয়ভাবে রাখা হয়। ফলে একটি ডেটাকে কেবলমাত্র একবার একজায়গায় এন্ট্রি করতে হয়। ডেটাগুলি পরস্পর সম্পর্কীত বলে ডেটাবেস থেকে সমস্ত জায়গায় ডেটা পাওয়া যায়। এছাড়াও একজায়গায় ডেটা পরিবর্তন করা হলে বাকি সমস্ত জায়গাতেই তার প্রতিফলন ঘটে। তবে কেন্দ্রীয়ভাবে ডেটা রাখা হলে কতকগুলি বিষয়ের উপর গুরত্ব দিতে হয়। এগুলি হল ডেটার সুরক্ষা এবং গোপনীয়তা, ডেটার গুণমাণ ও ডেটার বিশুদ্ধতা।

● **ডেটাবেস ম্যানেজমেন্ট সিস্টেম** (Database Management System)

ডেটাবেস তৈরী ও ব্যবহার করার জন্য এক বিশেষ ধরনের প্রোগ্রামের প্রয়োজন হয়। এই প্রোগ্রামকে বলা হয় ডেটাবেস ম্যানেজমেন্ট সিস্টেম বা সংক্ষেপে DBMS। DBMS-এর প্রধান কাজগুলি হল—

(i) প্রয়োজনীয় টেবিল তৈরী করা;

(ii) টেবিলে ডেটা এন্ট্রির ব্যবস্থা করা;

(iii) টেবিলে ডেটা এডিট করার ব্যবস্থা করা;

(iv) টেবিলগুলিকে সম্পর্কযুক্ত করা;

(v) দ্রুত প্রয়োজনীয় তথ্য বের করা;

(vi) একসাথে একাধিক ব্যক্তি যাতে ডেটাবেস ব্যবহার করতে পারে তার ব্যবস্থা করা,

(vii) ডেটা সুরক্ষার ব্যবস্থা করা।

DBMS-এর প্রধান সুবিধা হল ব্যবহারকারী খুব সহজেই ডেটা ভরতে পারে, এডিট করতে পারে ও প্রয়োজনীয় তথ্য সংগ্রহ করতে পারে। এজন্য তাকে কিভাবে ডেটাবেসটি তৈরী করা হয়েছে বা কিভাবে ডেটাগুলিকে পরস্পরের সাথে সম্পর্কযুক্ত করা হয়েছে তা জনতে হয় না। বস্তুত DBMS নিজে কোনো ডেটাবেস নয় একটি সফটওয়্যার। একে ব্যবহার করে নির্দিষ্ট কাজের জন্য প্রয়োজনীয় ডেটাবেস তৈরী করতে হয়। যেমন, Excel-কে ব্যবহার করে কোনো স্প্রেডশিট তৈরী করা যায় বাMs-Word-কে ব্যবহার করে কোনো বিষয় লেখা যায়। সাধারণত দু'ধরনের DBMS সফ্টওয়্যার পাওয়া যায়—ডেস্কটপ DBMS এবং ক্লায়েন্ট/সার্ভার DBMS. মাইক্রোসফট্ Access হল একটি ডেস্কটপ DBMS এবং Oracle বা মাইক্রোসফট SQL Server হল ক্লায়েন্ট/সার্ভার DBMS । যেসব ক্ষেত্রে ডেটার পরিমাণ খুবই বেশী এবং ডেটা সুরক্ষার বিষয়টি অত্যন্ত গুরুত্বপূর্ণ সেখানে ক্লায়েন্ট/সার্ভার DBMS ব্যবহার কার হয়। তবে সোশ্যাল ওয়ার্ক গবেষণার ক্ষেত্রে ডেস্কটপ DBMS যথেষ্ট উপযোগী।

● **DBMS-এর বৈশিষ্ট :** DBMS-এর গুণগত বৈশিষ্ট্য যেসব বিষয়ের উপর নির্ভর করে সেগুলি হল নিম্নরূপ :

তথ্যচিত্র তৈরি করা প্রয়োজন হয়। Excel -এ বিভিন্ন ধরনের চার্ট যেমন বার চার্ট, কলম চার্ট, লাইন চার্ট, পাই চার্ট, XY চার্ট, সারফেস চার্ট ইত্যাদি এছাড়া ত্রিমাত্রিক টার্টও Excel ব্যবহার করে তৈরি করা যায়। চার্ট তৈরি করার জন্য সাধারণত চার্ট উইজার্ড (Chart Wizard) এর সাহায্য নেওয়া হয়। মেনুবারে Insert–Chart-এ ক্লিক করে বা টুলবারের চার্ট উইর্জাড বার্টনে ক্লিক করে অগ্রসর হতে হয়। এর ফলে একটি চার্ট উইর্জাড উইন্ডো আসে যেখানে চারটি ধাপের সাহায্যে চার্ট তৈরী করা যায়।

উপরের আলোচনা থেকে বলা যায় যে সোশ্যাল ওয়ার্ক গবেষণায় বা সমীক্ষায় সংগৃহীত তথ্য সহজে, নির্ভুলভাবে ও অত্যন্ত দ্রুততার সাথে বিশ্লেষণ করার জন্য গবেষকের MS-Excel সংক্রান্ত জ্ঞান ও দক্ষতা থাকলে তার পক্ষে, গবেষণার কাজ চালানো অনেক সহজ হয়। এছাড়াও বিভিন্ন ধরনের গ্রাফ, চার্ট ইত্যাদি তৈরী করতেও বিশেষ সুবিধা হয় ও সময়ের সাশ্রয় ঘটে।

৩.৪ ডেটাবেস ম্যানেজমেন্ত (Database Management)

গবেষণার ক্ষেত্রে বিভিন্ন উৎস থেকে বিভিন্ন ধরনের উপাত্ত বা ডেটা সংগ্রহ করা হয়। এই ডেটাগুলি হল গবেষণার মূল ভিত্তি। কারণ সংগৃহীত ডেটাগুলি বিশ্লেষণ করে যেসব তথ্য পাওয়া যায় সেগুলিই গবেষণার ফলাফলকে চূড়ান্ত রূপ দেয়। আগে ডেটা সংগ্রহ থেকে বিশ্লেষণ প্রতিটি কাজই গবেষক বা সমীক্ষককে নিজের হাতে করতে হত। এতে শ্রম, সময় ও ব্যয় বেশী হত এবং ভুল থাকার সম্ভাবনাও অনেক বেশী ছিল। কম্পিউটার আসার পর এসব কাজ বর্তমানে কম্পিউটারই করে দেয় এবং অনেক নির্ভুলতা ও দ্রুততার সাথে এই কাজ করে দেয় বলে গবেষকের অনেক বেশী সুবিধা হয়। অবশ্য এজন্য গবেষকের কম্পিউটার অপারেটিং সংক্রান্ত জ্ঞান থাকা প্রয়োজন।

ডেটাবেস হল যে কোনো ধরনের ডেটা কম্পিউটারে সঞ্চয় করে রাখার একটি সংগঠিত ও বৈজ্ঞানিক ব্যবস্থা। ডেটা সঞ্চয় করে রাখার অন্য পদ্ধতির সাথে এর পার্থক্য হল যে ডেটাবেস থেকে অত্যন্ত দ্রুততার সাথে প্রয়োজনীয় তথ্য নির্ভুলভাবে বের করা যায়, গবেষণায় যেসব ডেটা সংগৃহীত হয়, বেশীরভাগ ক্ষেত্রেই সেগুলি পরস্পর সম্পর্কযুক্ত। কাজেই সেগুলি একসঙ্গে বিশ্লেষণ করে সিদ্ধান্ত নিতে হয়। এজন্য কেন্দ্রীয়ভাবে ডেটা অ্যাকসেসের ব্যবস্থা থাকা দরকার। ডেটাগুলি কেন্দ্রীয়ভাবে না রাখা হলে সাধারণত দুটি প্রধান অসুবিধা হয়। এগুলি হল—কাজের বিভিন্ন স্তরে একই ডেটাকে অনেকবার এন্ট্রি করতে হয় এবং সমস্ত ডেটাকে একই সময়ে, একই জায়গায় একই সঙ্গে পাওয়া যায় না। ডেটাগুলি আলাদা আলাদা জায়গায় মজুত করে রাখলে অযথা সময় নষ্ট হয়, ডেটা মজুতের পরিমাণ বাড়ে ও ভুল হবার সম্ভবনা থাকে। ডেটাবেস তৈরি করার ক্ষেত্রে দুটি বিষয় সম্বন্ধে ধারণা থাকা আবশ্যক। এগুলি হল :

(i) টেবিল : ডেটাবেসের মধ্যে এক বা একাধিক টেবিলে ডেটা রাখা থাকে। গবেষণার তথ্য সংগ্রহের এক একটি বিষয় বা ঘটনা সম্পর্কীত ডেটার জন্য এক একটি টেবিল থাকে। সমস্ত টেবিল মিলিয়ে ডেটাবেসটি তৈরী হয়।

আগে হিসাব করার জন্য এবং তথ্য লিপিবদ্ধ করার জন্য স্প্রেডশিট ব্যবহার করা হত। কিন্তু কম্পিউটার আসার পর ইলেকট্রনিক স্প্রেডশিট বা স্প্রেডশিট সফ্‌টওয়্যারের প্রচলন শুরু হয়। কারণ সাধারণ স্প্রেডশিটের মূল অসুবিধা ছিল হাতে করে হিসাব করা। এতে সময় বেশী লাগত এবং হিসাবে অর্থাৎ যোগ-বিয়োগ ভুল হবার সম্ভাবনা থাকত। তাই হিসাব ঠিক হয়েছে কিনা পরীক্ষা করে দেখতে হত। তাছাড়া স্প্রেডশিটে যেসব তথ্য থাকত সেগুলি থেকে রিপোর্ট তৈরী করা অনেকসময়ই সম্ভব হত না।

গবেষণার ক্ষেত্রে তথ্য প্রক্রিয়াকরণ ও বিশ্লেষণের প্রাথমিক কাজ স্প্রেডশিট দ্বারা করা যায়। গবেষণায় সংগৃহীত বিভিন্ন ডেটা এবং স্প্রেডশিটে এন্ট্রি করে পরবর্তী পর্যায়ে বিভিন্ন তথ্য সংগ্রহ করা যায়। আর ডেটা থেকে প্রয়োজনীয় তথ্য সংগ্রহের কাজটি কম্পিউটারই করে দেয়। কারণ বিভিন্ন পরিসংখ্যানমূলক পদ্ধতি এই সফ্‌টওয়্যারের মধ্যেই দেওয়া থাকে। ইলেকট্রনিক স্প্রেডশিটে সাধারণ স্প্রেডশিটের অসুবিধাগুলি থাকে না। একটি জটিল গণনার ক্ষেত্রে কোনো একটি ডেটাকে পাল্টালে বাকী সমস্ত ডেটা নিমেষের মধ্যে আবার গণনা করে পাল্টে যায় এবং সেগুলি দেখতে পাওয়া যায়। এই ব্যাপারটা একদিকে যেমন গুরুত্বপূর্ণ, তেমনি আকর্ষণীয়ও বটে। এই সুবিধা সাধারণ স্প্রেডশিটে চিন্তা করাও যায় না। ইলেকট্রনিক স্প্রেডশিটের মধ্যে (built in) হিসাব করার জন্য অজস্র ফর্মুলা থাকে, যেগুলি খুব সহজেই ব্যবহার করা যায়, প্রথম দিকে পার্সোনাল কম্পিউটারের (P.C.) জনপ্রিয়তার প্রধান কারণ হল ওয়ার্ড প্রসেসিং সফ্‌টওয়্যার ও স্প্রেডশিট সফ্‌টওয়্যারের ব্যবহার। বর্তমানে স্প্রেডশিট সফ্‌টওয়্যার অনেক উন্নত হয়েছে। এর অন্তর্ভুক্ত ফর্মুলাগুলি আর কেবল যোগ-বিয়োগ বা গুণ-ভাগের মধ্যে সীমাবদ্ধ নেই। স্ট্যাটিসটিক্‌স, হায়ার ম্যাথমেটিক্স, ত্রিকোণমিতি, ইঞ্জিনিয়ারিং ইত্যাদি বিভিন্ন বিষয়ের ফর্মুলা এতে যুক্ত হয়েছে। এছাড়াও, চার্ট, রিপোর্ট, ডেটাবেস টেবিল এমনকি প্রোগ্রামিং সুবিধাও এতে যুক্ত করে বর্তমানে সংখ্যা সংক্রান্ত তথ্য বিশ্লেষণে এটি অত্যন্ত শক্তিশালী হয়ে উঠেছে। ফিনান্স এবং অ্যাকাউন্টস সংক্রান্ত সমস্ত কাজের সাথে সাথে ইঞ্জিনিয়ারিং ডিজাইন এবং গবেষণার তথ্য বিশ্লেষণের কাজেও স্প্রেডশিট সফ্‌টওয়্যারের ব্যাপক ব্যবহার ঘটেছে।

Windows-এর Microsoft-Excel এরূপ একটি বহু-ব্যবহৃত সফ্‌টওয়্যার। এছাড়াও Lotus 1-2-3, Quattro Pro ইত্যাদি স্প্রেডশিট সফ্‌টওয়্যার যেগুলি তথ্য বিশ্লেষণে ব্যবহৃত হয়। তবে এদেরও মধ্যে Excel সর্বাধিক জনপ্রিয়। MS-Office-এর যে চারটি ভাগ রয়েছে তার একটি হল Excel. অন্যগুলি হল Word, Access ও Power Point। Excel-এর আইকনটি সাদারণত প্রোগ্রাম মেনুর মধ্যে থাকে। এই আইকনে ক্লিক করলে মনিটরে স্প্রেডশিটটি ভেসে আসে। এই স্প্রেডশিটের প্রতিটি রো এবং কলাম নিয়ে যে এক-একটি ঘরের সৃষ্টি হয় তাদের বলে সেল (Cell)। এই প্রতিটি সেল-এর একটি ঠিকানা থাকে এবং বিভিন্ন ধরনের গণনার ক্ষেত্রে যে ফর্মুলা ব্যবহার করা হয় তাতে এই Cell-এর ঠিকানা ব্যবহৃত হয়।

গবেষণার প্রতিবেদন তৈরীর ক্ষেত্রে অনেকসময় চার্ট বা গ্রাফের প্রয়োজন হয়। কারণ ওয়ার্কশিটে যে সমস্ত ডেটা থাকে তাদের বৈশিষ্ট্যগুলি রেখা বা ছবির মাধ্যমে দেখতে পারলে অনেক বেশি সহজবোধ্য হয়। কারণ কেবলমাত্র সংখ্যা দেখে অনেকসময়ই তাদের ধরন বোঝা যায় না। সেজন্য পরিসংখ্যানমূলক

(vi) প্রয়োজনমাফিক যতখুশি লেখার প্রিন্ট পাওয়ার সুবিধা, এবং

(vii) লেখাকে মজুত (store) করে রাখার সুবিধা।

এর মধ্যে শেষের বিষয়টি সবচেয়ে গুরুত্বপূর্ণ কারণ লেখা হল একধরনের তথ্য এবং ভবিষ্যতে প্রয়োজনের সময় সেটি দ্রুত ও নির্ভুলভাবে পাওয়া অত্যন্ত গুরুত্বপূর্ণ। স্বাভাবিকভাবেই কম্পিউটার হল এই কাজের আদর্শ মাধ্যম। কম্পিউটারে লেখা সংক্রান্ত এই কাজগুলি করাকে বলা হয় **ওয়ার্ড প্রসেসিং।** গবেষণার ক্ষেত্রে যেহেতু বিভিন্ন বিষয় লিখে রাখতে হয় তাই গবেষকের এ-ব্যাপারে জ্ঞান থাকা বাঞ্ছণীয়। প্রকৃতপক্ষে ওয়ার্ড প্রসেসিং সিস্টেম এতটাই শক্তিশালী যে বর্তমানে এর সাহায্য ছাড়াই গবেষণা চালানো অসম্ভব হয়ে পড়ে। সংগৃহীত ডেটা কম্পিউটারে লিখে রাখা ও প্রয়োজনমতো ব্যবহার করার জন্য এই সিস্টেম খুবই কার্যকরী হয়।

ওয়ার্ড প্রসেসিং প্রোগ্রামের সাহায্যে লেখা কোনো বিষয় কম্পিউটারের ডিস্কে ফাইল হিসাবে রাখা থাকে। প্রয়োজন অনুযায়ী এক ফালি খোলা যায় ও পরিবর্তন করা যায়। এই প্রোগ্রামের সাহায্যে লেখার সময় পুরো বিষয়টি মনিটর-এ দেখা যায়। লেখার জন্য যে আকারের ও যে ধরনের অক্ষর ব্যবহার হয়, কি-বোর্ডে টাইপ করলেই ঠিক সেই অক্ষরগুলিই পাওয়া যায়। এছাড়া লাইন, অনুচ্ছেদ ও পাতার বিন্যাস যেভাবে করা হয়, মনিটরে ঠিক সেভাবেই দেখায়। ফলে পুরো বিষয়টি লিখতে সুবিধা হয়। ফাইলে লেখা ছাড়াও টেবিল, ছবি ইত্যাদিও রাখা যায়, এমনকি অন্য কোনো প্রোগ্রাম থেকে পাওয়া তথ্যও সরাসরি রাখা যায়। অক্ষর ও পাতার বিন্যাস ব্যবহারকারীর পছন্দের উপর নির্ভর করে অর্থাৎ লেখায় যে কোনো বিন্যাস ব্যবহার করা যেতে পারে বা একই বিন্যাস লেখায় ব্যবহার করা যেতে পারে।

লেখার মধ্যে বানান বা ব্যাকরণ সংক্রান্ত কোনো ভুল থাকলে টাইপ করার সময়েই ওয়ার্ড প্রসেসিং প্রোগ্রাম সেটি দেখিয়ে দেয় এবং সেই অনুযায়ী কারেকশন করা যায়। পুরো লেখাটি বিষয় ও পছন্দ অনুযায়ী সাজানো হয়ে গেলে প্রিন্টারের সাহায্যে পুরো লেখাটির যত খুশি প্রিন্ট নেওয়া যায়। আবার প্রিন্ট করার আগে Print Preview থেকে সম্ভাব্য কপির প্রিন্টও দেখা যায় এবং প্রয়োজনমত পরিবর্তন করা যায়। ওয়ার্ড প্রসেসিং-এর জন্য সবচেয়ে বেশী ব্যবহৃত ওয়ার্ড প্রসেসিং সফ্টওয়্যার হল MS-Word। তবে অন্যান্য ওয়ার্ড প্রসেসিং সফ্টওয়্যার যেমন Word-Perfect, Word Pro ইত্যাদির সাহায্যেও এই কাজ করা যায়।

৩.৩ ইলেকট্রনিক স্প্রেডশিট (Electronic Spreadsheet)

সাধারণভাবে স্প্রেডশিট হল একধরনের হিসাবের খাতা যাতে অনেকগুলি রো (Row) এবং কলাম (Column) থাকে। এই রো এবং কলাম মিলিয়ে যে ঘরগুলি হয় তাতে বিভিন্ন হিসাব বা তথ্য লেখা হয়। স্প্রেডশিটে ইলম্ব লাইনগুলি কলাম ও সমান্তরাল লাইনগুলি রো-র সৃষ্টি করে। কম্পিউটার ব্যবহারের

সফ্‌টওয়্যারে। কম্পিউটার অত্যন্ত দ্রুততার সাথে সেই নির্দেশ পালন করে। সফ্‌টওয়্যার দু'ধরনের হয়। এগুলি হল—

(i) সিস্টেম সফ্‌টওয়্যার : একটি অবশ্য প্রয়োজনীয় সিস্টেম সফ্‌টওয়্যার হল অপারেটিং সিস্টেম (Operating system)। এটি না থাকলে কম্পিউটার চালানো যাবে না। এরূপ সফ্‌টওয়্যার-এর উদাহরণ হল Windows, Linux ইত্যাদি। এছাড়াও আর একটি গুরুত্বপূর্ণ সিস্টেম সফ্‌টওয়্যারের হল প্রোগ্রামিং ল্যাঙ্গুয়েজের (Programming language) কমপাইলার, যেমন—C^{++}, Fortran ইত্যাদি যার সাহায্যে সফ্‌টওয়্যার তৈরী করা যায়।

(ii) অ্যাপ্লিকেশন সফ্‌টওয়্যার : দৈনন্দিন কাজের জন্য যে সমস্ত সফ্‌টওয়্যার ব্যবহার করা হয় তাদের বলা হয় অ্যাপ্লিকেশন সফ্‌টওয়্যার। যেমন লেখার জন্য ওয়ার্ড প্রসেসিং সফ্‌টওয়্যার MS-Word, Word Perfect ইত্যাদি, হিসাব করার জন্য স্প্রেডশিট সফ্‌টওয়্যার MS Excel, Lotus 1-2-3 ইত্যাদি, ছবি প্রসেসিং সফ্‌টওয়্যার Adobe Photoshop, Corel Photo Paint ইত্যাদি।

৩.২ ওয়ার্ড প্রসেসিং (Word Processing)

গবেষণার ক্ষেত্রে লেখার বিষয়টি ওতপ্রোতভাবে জড়িয়ে রয়েছে। কারণ উপাত্ত ও তথ্য সংক্রান্ত যে কোনো বিষয়ই লেখার বিষয় হতে পারে। আগে কেবলমাত্র হাতে লেখাই প্রচলিত ছিল। পরবর্তীকালে টাইপরাইটারের উদ্ভব হলেও হাতে লেখার সমস্যাগুলি কিন্তু থেকেই গিয়েছিল। যেমন একই লেখার অনেক বেশি কপি পাওয়া সম্ভব হত না। লেখায় কোনো ভুল থাকলে পুরো লেখা আবার লিখতে হত। ইলেকট্রিক টাইপরাইটারে যদিও এই অসুবিধাগুলি অনেকাংশে দূর করা সম্ভব হয়েছিল তবুও এর মেমরীর পরিমাণ খুবই কম হওয়ায় বেশ কিছু সীমাবদ্ধতা ছিল। তাছাড়া লেখার কোনো অংশ রঙিন করা, বড় করা, অক্ষরের ধরন সংক্রান্ত, টেবিল ও ছবি সংক্রান্ত সমস্যাগুলি থেকেই গিয়েছিল। এই সব সমস্যা কম্পিউটার উদ্ভাবনের সাথে সাথে নিরসন ঘটেছে। বর্তমানে কম্পিউটারের মাধ্যমে লেখার ক্ষেত্রে যেসব সুবিধা পাওয়া যায় সেগুলি হল—

(i) ভালভাবে, নির্ভুলভাবে এবং সুবিধাজনকভাবে লিখতে পারা।

(ii) বিষয় অনুযায়ী অক্ষর, শব্দ, লাইন এবং পাতার সুষ্ঠু ও সুন্দর বিন্যাস করতে পারা,

(iii) প্রয়োজনবোধে অন্যান্য তথ্য, যেমন, ছবি, টেবিল, সমীকরণ ইত্যাদি যোগ করতে পারা ;

(iv) লেখার বানান ও ব্যাকরণ পরীক্ষা করার সুবিধা ;

(v) লেখা পরিবর্তন করার সুবিধা ;

কম্পিউটার সম্বন্ধে জানতে হলে দুটি বিষয় সম্বন্ধে ধারণা গড়ে তোলা দরকার। এগুলি হল— (1) কম্পিউটারের বিভিন্ন যন্ত্রাংশ যাদের এককথায় বলা হয় **হার্ডওয়্যার** (Hardware) এবং (2) যা দিয়ে কম্পিউটার চালানো হয় অর্থাৎ কম্পিউটারকে দিয়ে কাজ করানো হয়, যাকে এককথায় বলে **সফ্টওয়্যার** (Software)।

(1) কম্পিউটার হার্ডওয়্যার : কম্পিউটারের ভিতরে ও কম্পিউটারের সাথে যে যন্ত্রাংশগুলি লাগানো থাকে সেগুলিকে বলে হার্ডওয়্যার। অন্যভাবে বলা যায় কম্পিউটারের যেসব যন্ত্রাংশ আমরা দেখতে পাই, স্পর্শ করতে পারি এবং যাদের আকার ও আয়তন রয়েছে সেগুলিই হার্ডওয়্যারের অন্তর্ভুক্ত। কম্পিউটারের মূল কাঠামোটিই হল হার্ডওয়্যার। এর অন্তর্ভুক্ত তিনটি গুরুত্বপূর্ণ যন্ত্রাংশ হল :

(i) সেন্ট্রাল প্রসেসিং ইউনিট বা CPU : এই যন্ত্রটি কম্পিউটারের সমস্ত যন্ত্রাংশ নিয়ন্ত্রণ করে এবং সমস্ত ধরনের তথ্য বিশ্লেষণ করে। সেজন্য একে প্রসেসর-ও (Processor) বলা হয়। এটিই হল কম্পিউটারের মূল অংশ। এই প্রসেসর যত শক্তিশালী হবে কম্পিউটারের তত বেশী ডেটা অল্প সময়ের মধ্যে প্রসেস করতে পারবে। বর্তমানে লার্জ স্কেল ইন্ট্রিগ্রেশন প্রযুক্তির সাহায্যে এর ক্ষমতা বহুগুণ বাড়ানো সম্ভব হয়েছে।

(ii) মূল মেমরী : প্রসেসরের শক্তি বাড়ালে অনেক বেশী ডেটা প্রসেস করা সম্ভব কিন্তু এই প্রসেসিং মূল মেমরীর পরিমাণের উপরও নির্ভরশীল। কারণ অনেক ডেটা প্রসেস করতে গেলে প্রসেসিং-এর সময় সেই ডেটা রাখার জন্য সমতুল মেমরীরও প্রয়োজন হয়। কারণ প্রসেসিং-এর সময় মেমরী থেকে প্রসেসরে ডেটা আদান-প্রদান হয়। সেকারণে বর্তমানে ডেটা মজুত করার জন্য ডিস্ক মেমরীর ব্যবহার বৃদ্ধি পেয়েছে।

(iii) ইনপুট-আউটপুট ইউনিট : কম্পিউটারে ডেটা এন্ট্রির জন্য যেসব যন্ত্রাংশ ব্যবহার করা হয় তাদের এককথায় ইনপুট ডিভাইস (Input devise) বলা হয়। কম্পিউটারের সাথেই লাগানো কি বোর্ড, মাউস, স্ক্যানার ইত্যাদি ইনপুট ডিভাইস-এর উদাহরণ। অন্যদিকে যেসব যন্ত্রের সাহায্যে কম্পিউটার থেকে তথ্য বা ফলাফল পাওয়া যায় তাদের আউটপুট ডিভাইস (output devise) বলে। এরূপ যন্ত্রাংশের উদাহরণ হল মনিটর, প্রিন্টার, সাউন্ড বক্স ইত্যাদি।

(2) সফ্টওয়্যার : অসাধারণ দ্রুততায় গণনা করতে পারলেও কম্পিউটার আসলে একটি যন্ত্র মাত্র। প্রয়োজন অনুযায়ী বিভিন্ন ধরনের কাজ করার জন্য এই যন্ত্রটিই যথেষ্ট নয় কারণ কম্পিউটার নিজে নিজে কাজ করতে পারে না। কম্পিউটারকে নির্দেশ (Command) দিলে তবেই সে কাজ করে। আর এই নির্দেশ দেওয়ার কাজটি সফ্টওয়্যার-এর মাধ্যমে পালিত হয়। প্রকৃতপক্ষে সফ্টওয়্যার এক ধরনের কম্পিউটার প্রোগ্রাম। যে কাজটি করতে হবে সেটি কিভাবে করতে হবে তার নির্দেশ থাকে এই

গ বিভাগ ❐ সামাজিক কর্ম গবেষণায় কম্পিউটারের ব্যবহার (Use of Computer in Social Work Research)

একক : ৩

৩.১ কম্পিউটার কি? (What is computer)

বর্তমান যুগকে কম্পিউটার যুগ বলা হয়। এটি হল একটি বহু উদ্দেশ্য সঠিক যন্ত্র। সেজন্য একে যন্ত্রগণক বা পরিগণকও বলা হয়। কম্পিউটার কথার অর্থ হল 'Common purpose users terminal' অর্থাৎ সাধারণ উদ্দেশ্যসাধক ব্যবহারকারীদের গন্তব্যস্থল। এই যন্ত্র মানুষের দৈনন্দিন প্রয়োজনীয় বিভিন্ন ধরনের কাজ করতে পারে এবং মানবিক কার্যকলাপের প্রতিটি ক্ষেত্রেই এর উপস্থিতি ও ব্যবহার ক্রমাগত বৃদ্ধি পাচ্ছে। এই যন্ত্রের সাহায্যে মূলত তিনটি কাজ অত্যন্ত দ্রুত ও নির্ভুলভাবে করা যায়—(i) তথ্য সংগ্রহ (ii) তথ্য বিশ্লেষণ বা গণনা ও (iii) তথ্য মজুত। যেকোনো গবেষণার ক্ষেত্রে কম্পিউটারের ব্যবহার আজ সর্বজনস্বীকৃত এবং সোশ্যাল ওয়ার্ক রিসার্চের ক্ষেত্রেও এর ব্যতিক্রম ঘটে না।

বিগত চারদশকে কম্পিউটারের ক্ষেত্রে অভূতপূর্ব অগ্রগতি ঘটেছে। তাই বর্তমানে analogue Computer-এর Digital Computer-এ উত্তরণ ঘটেছে। এছাড়াও ব্যবহারকারীদের প্রয়োজন অনুযায়ী Desk Top Computer, Laptop Computer ও Super Computer-এর সৃষ্টি হয়েছে। বর্তমান যন্ত্রসভ্যতা কম্পিউটার-নির্ভর সভ্যতায় পরিণত হয়েছে। প্রযুক্তির ক্ষেত্রে এক নতুন বিষয়ের আবির্ভাব ঘটেছে। যার নাম দেওয়া হয়েছে তথ্য প্রযুক্তি (Information Technology)। আর এই প্রযুক্তির প্রধান স্তম্ভটি হল কম্পিউটার। কম্পিউটারের সাহায্যে সংগৃহীত **উপাত্তকে** (Data) **তথ্যে** (Information) পরিণত করা সম্ভব হয় এবং ভবিষ্যতের জন্য মজুত করে রাখাও যায়।

Reference Books :

1. Spiegel M. R.	:	Theory and Problems of Probability and Statistics (Schaum's outline series) : Mc Graw Hill Book Co.
2. Goon A. M., Gupta M. K. & Das Gupta B.	:	An Outline of Statistical Theory, Vol. I & II (The World Press Pvt. Ltd.)
3. P. K. Giri & J. Banerjee	:	Statistical Tools and Techniques
4. Das N. G.	:	Statistical Methods Vol. I & II
5. Ghosh and Saha	:	Business Mathematics and Statistics
6. Nag N. K.	:	Business Mathematics

এক্ষেত্রে 6টি শ্রেণী আছে। $\therefore$ $d.f.$ = 6 − 1 = 5

এবার যেহেতু χ-এর পর্যবেক্ষ লব্ধ মান 3·4 যা χ^2-এর প্রদত্ত মান অপেক্ষা কম ($\chi^2_{\cdot 01}$ = 15·09 $d.f.$ 5-এর জন্য) যেহেতু H_0-কে 1% গুরুত্বের স্তরে বাতিল করা যায় না। মন্তব্য হল ছক্কা যে পক্ষপাত শূন্য এই অনুমানের সঙ্গে উপাত্তের পূর্ণ সামঞ্জস্য বিদ্যমান। $\therefore$ ছক্কা পক্ষপাতশূন্য (unbiased)

অনুশীলনী

1. পরিসংখ্যানগত অনুমান (Hypothesis) বলতে কি বোঝ? অনুমানগুলি কয় প্রকার ও কি কি উদাহরণ সহযোগে ব্যাখ্যা কর।

2. কয়েকটি ধারণা সম্পর্কে ব্যাখ্যা দাও ঃ (i) স্ট্যাটিস্‌টিক, (ii) প্যারামিটার, (iii) গুরুত্বের স্তর, (iv) প্রথম ও দ্বিতীয় ধরনের ত্রুটি।

3. অনুমানের পরীক্ষার বিভিন্ন পর্যায়গুলি উল্লেখ কর।

4. টেস্ট স্ট্যাটিস্‌টিক ও χ^2 বিভাজন সম্পর্কে সংক্ষেপে আলোচনা কর।

5. কোন সমগ্রক থেকে 10-এর একটি নমুনা নেওয়া হল। যৌগিক গড় থেকে প্রদত্ত নমুনার অন্তরগুলির বর্গের সমষ্টি হল 50। অনুমানটি যে সমগ্রকের ভেদমান গুরুত্বের 5 শতাংশ স্তরে হল 5।

[χ^2-এর টেবিল মান 5% স্তরে $d.f.$ 9-এর জন্য হল 16,92] [Ans. (H_0 গৃহীত)]

6. মুদ্রাটি যথার্থ (perfect) কিনা পরীক্ষা করার জন্য তা 5 বার ছোঁড়া (toss) হল। মুদ্রার যথার্থতার (perfectness) নাল অনুমান বাতিল করা হয় যদি 4-এর অধিক বার হেড আসে। এক্ষেত্রে প্রথম ধরনের ত্রুটির সম্ভাবনা কি? যদি হেড আসার সংশ্লিষ্ট সম্ভাবনা = 0·2 তাহলে দ্বিতীয় ধরনের ত্রুটির সম্ভাবনা নির্ণয় কর। $\left[\text{Ans.}\ \frac{1}{32}, \frac{3124}{3125}\right]$

7. 5 জন শিশু রয়েছে এমন পরিবারের 320টির ক্ষেত্রে অনুসন্ধান লব্ধ রাশিতথ্যের বিভাজন নিম্নে দেওয়া হল :

বালকের সংখ্যা	:	5	4	3	2	1	0
বালিকার সংখ্যা	:	0	1	2	3	4	5
পরিবারের সংখ্যা	:	14	56	110	88	40	12

উপরোক্ত উপাত্তসমূহ কি এরূপ একটি অনুমানের সঙ্গে সংগতিপূর্ণ যে পুরুষ ও স্ত্রী জন্মের সম্ভাবনা সমান যেখানে ছকে উল্লিখিত χ^2-এর মান $d.f.$ 5-এর জন্য গুরুত্বের 5% স্তরে হল 11·07 ?

[Ans. সম্ভাবনা সমান গৃহীত সিদ্ধান্ত]

যেহেতু χ^2 শূন্য নয় সেইহেতু উপাত্তগুলি নিরপেক্ষ নয় (data are not independent)। তবে χ^2-এর মান থেকে যোগসূত্রের অভিমুখ (direction of association) জানা যায় না। সিদ্ধান্তমূলক রাশিবিজ্ঞানে χ^2 ছক অথবা কমপিউটার প্রোগ্রামের প্রয়োজন হয় যোগসূত্রের মূল্যায়ন করার জন্য (for evalute the association)। x^2 ছকের পুঙ্খানুপুঙ্খ বিশ্লেষণের মধ্যে না গিয়ে এই ধরনের যোগসূত্র হঠাৎভাবে ঘটতে পারে এমনকি 1000 বারের মধ্যে 1-এর কমবার হতে পারে। 9টি কোষ (cell) সম্পন্ন ছকে $\chi^2 = 45{\cdot}7$ গুরুত্বপূর্ণ বিবেচিত হয় ·0001 স্তরে।

χ^2 বিভাজনে প্যারামিটার হল (degree of freedom) বা *d.f.*। *d.f.* কার্যত বর্গগুলির সমষ্টিতে নিরপেক্ষ পর্যবেক্ষণগুলির সংখ্যা নির্দেশ করে। উপরোক্ত ক্ষেত্রে d.f. হল 1 যেহেতু একটি মাত্র সমক স্বাভাবিক চলকের বর্গ উল্লেখিত হয়েছে।

অনুরূপভাবে, ধরা যাক $z_1, z_2, \ldots z_k$ হল K সংখ্যক নিরপেক্ষ স্বাভাবিক চলকের একক—অর্থাৎ প্রতিটি z স্বাভাবিক সমসম্ভাবনামূলক চলক (Normal random variable) যার যৌগিক গড় হল শূন্য এবং ভেদমান = 1। এবার যদি প্রতিটি z-এর বর্গ নেওয়া হয় তাহলে দেখান যেতে পারে যে এই বর্গগুলির সমষ্টি একটি χ^2 বিভাজন প্রকাশ করবে যার *d.f.* হবে K

$$\sum z^2 = z_1^2 + z_2^2 + \cdots + z_k^2 \sim \chi^2(K)$$

উদাহরণ 1 ঃ একটি ছক্কা (die) 60 বার ছোঁড়ার ফলগুলি নিম্নে প্রদত্ত হল। die বা ছক্কা পক্ষপাত-শূন্য (unbiased) এরূপ অনুমানের সঙ্গে প্রদত্ত উপাত্তসমূহ কি সংগতিপূর্ণ যেখানে $\chi^2_{\cdot 01} = 15{\cdot}09$ *d.f.* 5-এর জন্য—

তল (Face)	ঃ	1	2	3	4	5	6	Total
পরিসংখ্যা (f_0)	ঃ	6	10	8	13	11	12	60

H_0 হল যে ছক্কা পক্ষপাত শূন্য (unbiased) ঃ

$\therefore$ প্রতিটি তলের সম্ভাবনা $\frac{1}{6}$; যেহেতু পাশার তল হল 6টি।

এবং প্রত্যাশিত পরিসংখ্যা (= f^c) $= 60 \times \frac{1}{6} = 10$ প্রত্যেক ক্ষেত্রে

$\therefore$ f_0	:	6	10	8	13	11	12
f_1	:	10	10	10	10	10	10
$(f_0 - f_1)^2$	:	16	0	4	9	1	4

$$\therefore \quad \chi^2 = \frac{16}{10} + \frac{0}{10} + \frac{4}{10} + \frac{9}{10} + \frac{1}{10} + \frac{4}{10} = 3{\cdot}4$$

উপরোক্ত ছকের সংশ্লিষ্ট কোষে পর্যবেক্ষণলব্ধ মান = f_0 দেওয়া হয়েছে।
প্রত্যাশিত মান = [(স্তম্ভ সমষ্টি × সারি সমষ্টি) ÷ সামগ্রিক সমষ্টি] = f_e

ছাত্রদের উচ্চতা	গবেষণা প্রক্রিয়ার গ্রেড C	B	A	(Total) মোট
সুউচ্চ (Tall)	17·5	15	17·5	50
মধ্যম উচ্চতা (Medium)	17·5	15	17·5	50
নিম্ন উচ্চতা (Short)	30	20	50	100
(Total) মোট	70	60	70	200

(উদাহরণ—সুউচ্চ পর্যায়ের ছাত্রদের *C* গ্রেডের প্রত্যাশিত মান = (70 × 50) ÷ 200 = 17·5)
উপরোক্ত ছকের সংশ্লিষ্ট কোষে (cell) প্রত্যাশিত মানগুলি উল্লিখিত হয়েছে (= f_1)
পর্যবেক্ষণলব্ধ মান ও প্রত্যাশিত মানের অন্তরের ছক (Difference Table)
অন্তর (Difference) = পর্যবেক্ষণলব্ধ মান (f_0) – প্রত্যাশিত মান (f_1)
(উদাহরণ—সুউচ্চ ছাত্রদের *C* গ্রেডের অন্তর সম্পর্কিত তথ্যরাশি = 30 – 17·5 = 12·5)

গবেষণা প্রক্রিয়ায় গ্রেড

ছাত্রদের উচ্চতা	গবেষণা প্রক্রিয়ার গ্রেড C	B	A	(Total) মোট
সুউচ্চ (Tall)	12·5	–5	–7·5	0
মধ্যম উচ্চতা (Medium)	–7·5	15	–7·5	0
নিম্ন উচ্চতা (Short)	–5	–10	15	0
(Total) মোট	0	0	0	0

∴ χ^2 = (প্রতিটি অন্তরের বর্গের সমষ্টি) ÷ (সংশ্লিষ্ট কোষের প্রত্যাশিত মান)

$$= \sum \frac{f_0 - f_1}{f_1}$$

[উদাহরণ—সুউচ্চ ছাত্রদের C গ্রেডের ক্ষেত্রে $(12{\cdot}5)^2$ ÷ 17·5 = 156·25 ÷ 17·5 – 8·93]

∴ χ^2 = প্রথম সারি (8·93 + 1·67 + 3·21) +
দ্বিতীয় সারি (3·21 + 15 + 3·21) +
তৃতীয় সারি (0·71 + 3·33 + 6·43) = 45·7

যেখানে K ধ্রুবক, n হল $d.f.$। যে চলকটি কাই-স্কয়ার বিভাজন অনুসরণ করে সেই চলকটিকে কাই-স্কয়ার চলক বলা হয়।

χ^2 বিভাজনের বৈশিষ্ট্য :

(1) যৌগিক গড় = n, সমক বিচ্যুতি $= \sqrt{2n}$ যেখানে n হল χ^2 বিভাজনের $d.f.$

(2) χ^2 রেখা ধনাত্মক প্রতি বৈষম্যযুক্ত হয় এবং মূল বিন্দু থেকে উত্থিত হয়ে ডান দিকে অসীম পর্যন্ত বিস্তৃত হয়।

(3) যদি x এবং y দুটি নিরপেক্ষ কাই-স্কয়ার চলকদ্বয় হয় যাদের d.f. হয় যথাক্রমে n_1 এবং n_2 তাহলে চলকদুটির সমষ্টি $(x + y)$ কার্যত χ^2 বিভাজন অনুসরণ করে যার $d.f.$ হল $(n_1 + n_2)$

(4) যখন $d.f. = n$ বৃহৎ হয় তখন $\sqrt{2\chi^2} - \sqrt{2n-1}$ আসন্নভাবে সমক স্বাভাবিক বিভাজন অনুসরণ করে।

χ^2 বিভাজন কার্যত বৃহৎ ও ক্ষুদ্র উভয় ধরনের নমুনা পরীক্ষায় ব্যবহৃত হয়। পর্যবেক্ষণলব্ধ মানগুলি কাল্পনিক বিভাজনের সঙ্গে সুসামঞ্জস্যপূর্ণ আছে কিনা যাকে 'Goodness of fit' বলা হয় তা পরীক্ষা করার জন্য χ^2 বিভাজন ব্যবহৃত হয়। আবার গুণগুলির বা attributes-এর নিরপেক্ষতা পরীক্ষায় χ^2 বিভাজন ব্যবহার করা হয়। ক্ষুদ্র নমুনার ক্ষেত্রে বিশেষভাবে উল্লেখিত সমক বিচ্যুতি জন্যও পরীক্ষায় χ^2 বিভাজন ব্যবহার করা হয়। χ^2 বিভাজন প্রধানত ব্যবহৃত হয়।

(1) পর্যবেক্ষণভিত্তিক মান ও কাল্পনিক বিভাজনের মধ্যে সুসামঞ্জস্য আছে কিনা তার জন্য পরীক্ষা বা Goodness of fit Test

(2) গুণাবলীর মধ্যে নিরপেক্ষতা পরীক্ষা (Test for independence of attributes)

(3) বিশেষভাবে উল্লেখিত সমক বিচ্যুতির জন্য পরীক্ষা (Test for a specified standard deviation)

উদাহরণের সাহায্যে χ^2 পরীক্ষা (χ^2 – test) পদ্ধতি ব্যাখ্যা করা যেতে পারে। নিম্নের উদাহরণে উচ্চতা এবং গ্রেডের মধ্যে যোগসূত্র ছকের মাধ্যমে উল্লেখ করা হল :

পর্যবেক্ষণলব্ধ উপাত্তের ছক

ছাত্রদের উচ্চতা	গবেষণা প্রক্রিয়ার গ্রেড			
	C	B	A	(Total) মোট
সুউচ্চ (Tall)	30	10	10	50
মধ্যম উচ্চতা (Medium)	10	30	10	50
নিম্ন উচ্চতা (Short)	30	20	50	100
(Total) মোট	70	60	70	200

(Significant)। এর দরুন নাল অনুমান H_0-কে গুরুত্বের 5% স্তরে বাতিল করতে হয় এবং সিদ্ধান্তে আসা যায় যে শহরের গরিষ্ঠ সংখ্যক মানুষ ধূমপায়ী যা উপাত্ত দ্বারা সমর্থিত।

(ঘ) কাই-স্কয়ার (χ^2) বিবাজন (Chi-Square (χ^2) Distribution) : কাই-স্কয়ার পরীক্ষা (χ^2 test) হল পরিসংখ্যানগত ক্রিয়া-প্রক্রিয়া বহুল ব্যবহৃত এবং সরলতম পরীক্ষাগুলির মধ্যে অন্যতম। 1900 সালে কার্ল পিয়ারসন (Karl Pearson) সর্বপ্রথম এরূপ পরীক্ষা করেন। χ কার্যত তত্ত্বও পর্যবেক্ষণের মধ্যে তারতম্যের মাত্রা ব্যাখ্যা করে। নিম্নলিখিতভাবে একে সংজ্ঞায়িত করা যায় :

$$\text{কাই-স্কয়ার} = \chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e}$$

যেখানে f_0 ও f_e হল যথাক্রমে পর্যবেক্ষণপ্রসূত পরিসংখ্যা এবং প্রত্যাশিত পরিসংখ্যা (Observed frequency and expected frequency)।

χ^2 দুভাবে ব্যবহৃত হয়। এর ফলে একটা বিভ্রান্তি দেখা দিতে পারে। বর্ণনামূলক রাশিবিজ্ঞানের শাখায় χ^2-এর দ্বারা দুটি চলকের মধ্যে যোগসূত্রের মাত্রা বোঝান হয়। পক্ষান্তরে রাশিবিজ্ঞানের সিদ্ধান্তমূলক শাখায় χ^2-এর মাধ্যমে কোন যোগসূত্র যা সম্ভাবনা নির্ভর তার সম্ভাবনা (Probability) সম্পর্কে ধারণা পাওয়া যায়। মাত্রাগত সংখ্যার দ্বারা প্রকাশযোগ্য চলকগুলির (attribute) ক্ষেত্রে এর ব্যাপক ব্যবহার দেখা যায়।

যখন যৌগিক গড় (μ_i) এবং ভেদমান σ_i^n সহ একটি স্বাভাবিকভাবে বিভাজিত সমসম্ভাবনা চলক X ধরা হয় (অর্থাৎ χ-এর বিভাজনটি একটি Normal distribution) অর্থাৎ $\chi \sim N(\mu, \sigma^2)$ তখন $z = \frac{(\chi - \mu)}{\sigma}$ হয় একটি সমক স্বাভাবিক চলক (Standard normal variable)—অর্থাৎ $z \sim N(0, 1)$। পরিসংখ্যানগত তত্ত্ব দেখায় যে একটি সমক সাধারণ চলক χ^2 বিভাজনের মত বিভাজিত হয় যার 'degree of freedom' (= d.f.) হল 1।

সাংকেতিকভাবে

$\chi^2(1) = z^2$ যেখানে (1) হল χ^2-এর d.f.

স্বাবাবিক বিভাজেনর যৌগিক গড় বা ভেদমান যেমন প্যারামিটার ঠিক তেমন χ^2 বিভাজনের প্যারামিটার হল d.f.

একটি সমসম্ভাবনা চলক χ কার্যত χ^2 বিভাজন অনুরূপ ধরা হবে যদি চলকটির সম্ভাবনা বিভাজন অপেক্ষকটি (*P.d.f.*) হয় :

$$f(x) = K.e^{-\frac{x}{2}}.\, x^{(n/2)-1}; \ (0 < x < \infty)$$

করা মান ও সংকটপূর্ণ মান (Critical value) তুলনা করেই H_0 বাতিল করা বা অন্যভাবে সিদ্ধান্ত নেওয়া হয়।

(7) এবার উপসংহার সরলভাবে প্রকাশ করা হয়। যদি H_0 বাতিল করা হয় তাহলে ব্যাখ্যা হবে নিম্নরূপ ঃ অনুমান H_0 যে সত্য এর সঙ্গে উপাত্তগুলি (data) সামঞ্জস্যপূর্ণ নয় এবং এই জন্য H_0 গ্রহণযোগ্য নয়। আবার H_0 যদি বাতিল করা না হয় তাহলে উল্লেখ করা হয় যে উপাত্তগুলি H_0-এর বিপক্ষে কোনরূপ প্রমাণ দিতে পারে না এবং একারণে H_0 গ্রহণ করা যেতে পারে। উপসংহার কিন্তু সমস্যার বক্তব্য অনুসারে কথায় প্রকাশ করাই বিধেয়।

উদাহরণ ঃ একটি বড় শহরে 600 জন ব্যক্তির মধ্যে 325 জন ধূমপায়ী। এই তথ্য থেকে ধরা যেতে পারে কি শহরের গরিষ্ঠ সংখ্যক ব্যক্তি (Majority of the persons) ধূমপায়ী ?

সমাধান ঃ নাল অনুমান হয় যে সমগ্র শহরে ধূমপায়ীদের অনুপাত হল 50% অর্থাৎ $\frac{50}{100} = 0{\cdot}5$

$\therefore\ H_0\ (P = 0{\cdot}5)$

শহরে ধূমপায়ীদের অনুপাত 50%-এর অধিক কিনা বিচার কার আবশ্যক। অতএব বিকল্প অনুমানটি হল ঃ $H_1\ (P > 0{\cdot}5)$

নমুনায় ধূমপায়ীর অনুপাত হয় 600 = n-এর মধ্যে 325

নমুনা অনুপাত ব্যবহার করে,

পর্যবেক্ষণ লব্ধমান $(P) = \frac{325}{600} = 0{\cdot}542$

যদি H_0 সত্য হয় তাহলে,

প্রত্যাশিত মান $(P_0) = 0{\cdot}5$

এবং সমক ত্রুটি বা $S.E.(P) = \sqrt{\frac{0{\cdot}5(1-0{\cdot}5)}{600}} = 0{\cdot}0204$

টেস্ট স্ট্যাটিসটিক $z = \frac{\text{[illegible]} - \text{[illegible]}}{S.E.} = \frac{0{\cdot}542 - 0{\cdot}5}{0{\cdot}0204} = 2{\cdot}1$

যেহেতু বিকল্প অনুমান, $H_1\ (P > 0{\cdot}5)$ হয় একমুখী, সেইহেতু পরীক্ষার সংকটপূর্ণ ক্ষেত্র (Critical region) হল একপ্রান্ত পর্যায়ী।

গুরুত্বের 5% স্তরে (at 5% level of significance) সংকটপূর্ণ ক্ষেত্র হয় $z \geq 1{\cdot}645$

উল্লেখ করা যেতে পারে যে z-এর জন্য সমক স্বাভাবিক রেখার প্রান্ত পর্যায়ের ক্ষেত্র হল 5% এবং তা $\geq 1{\cdot}645$।

টেস্ট স্ট্যাটিসটিক z-এর মান 2·1 যেহেতু Critical region-এর মধ্যে রয়েছে এবং তাই গুরুত্বপূর্ণ

(2) পরবর্তী পদক্ষেপ হল যথাক্রমে টেস্ট স্ট্যাটিস্টিক, T উল্লেখ কর। এবং নাল অনুমানটি সত্য এর পরিপ্রেক্ষিতে T-এর নমুনা বিভাজনটিও উল্লেখ কর। বৃহৎ নমুনা পরীক্ষায় (in large sample test) $z = \dfrac{(T - \theta_0)}{S.E(T)}$ যা আসন্নভাবে সমক স্বাভাবিক বিভাজন (Standard normal distribution) অনুসরণ করে প্রায়শই ব্যবহার করা হয়। কিন্তু স্বল্প নমুনা পরীক্ষার ক্ষেত্রে, সমগ্রকটিকে স্বাভাবিক ধরা হয় এবং বিভিন্ন টেস্ট স্ট্যাটিসটিকগুলি ব্যবহার করা হয় যেগুলি প্রকৃতভাবে সম্যক স্বাভাবিক বিভাজন, কাই-স্কয়ার (chi-square), t বিভাজন অথবা F বিভাজন অনুসরণ করে।

(3) পরবর্তী পদক্ষেপ হল পরীক্ষার ক্ষেত্রে গুরুত্বের স্তর α নির্বাচন করা বা স্থির করা। যদি না তা সমস্যায় উল্লিখিত থাকে। এ হল প্রথম ধরনের ত্রুটি করার সর্বোচ্চ সম্ভাবনা—অর্থাৎ একটি পরীক্ষা প্রক্রিয়া দ্বারা ভ্রান্ত সিদ্ধান্ত গ্রহণ করা যখন ঘটনাক্রমে নাল অনুমানটি প্রকৃতই সত্য। সাধারণত 5% অথবা 1% এর স্তর গুরুত্বের স্তর হিসাবে ধরা হয়। যখন এ সম্পর্কে কোন উল্লেখ থাকে না তখন 5% এর স্তর ধরাই বিধেয়।

(4) পরবর্তী পদক্ষেপ হল গুরুত্বের পছন্দ করা স্তরে পরীক্ষার Critical region নির্ণয় করা এর দ্বারা টেস্ট স্ট্যাটিস্টিকের একগুচ্ছ মানগুলি প্রকাশ করা হয় যেগুলি নাল অনুমান বাতিলের পক্ষে রায় দেয়। Critical region সর্বদা বিভাজনের একপ্রান্তের বা উভয় প্রান্তের পর্যায়ে দেখা দেয় যা বিকল্প অনুমান একমুখী না দুই-মুখী তার উপর নির্ভর করে। বিভাজনের দুই প্রান্ত পর্যায়ের ক্ষেত্র (area in the two tails) যা ‘Critical region’-এর আয়তন বলে অভিহিত হয় তা অতি অবশ্যই গুরুত্বের স্তর α-এর সমান হবে। এক প্রান্তিক পর্যায়ের পরীক্ষার জন্য α কার্যত বিভাজনের একপ্রান্ত পর্যায়ে প্রকাশ পায় এবং দু-প্রান্ত পর্যায়ের পরীক্ষায় $\alpha/2$ বিভাজনের প্রতি প্রান্ত পর্যায়ে প্রকাশ পায়। সংকটপূর্ণ এলাকা বা Critical region হল :

$T \geq T_{\alpha/2}$ অথবা যখন $H_1(\theta \neq \theta_0)$

$T \geq T_{\alpha}$ যখন

$T \leq T_{1-\alpha}$ যখন $H_1(\theta < \theta_0)$

যেখানে T_α হল T-এর এমন মান যে এর ডান দিকের ক্ষেত্রে হল α।

(5) পরবর্তী পদক্ষেপ হল নমুনা উপাত্তের ভিত্তিতে (on the basis of sample data) টেস্ট স্ট্যাটিস্টিক T-এর মান নির্ণয় করা এবং H_0 গণনা করা। বৃহৎ নমুনা পরীক্ষায় যদি কিছু প্যারামিটার অজ্ঞাত হয় সেগুলি কার্যত নমুনা থেকে গণনা করা হয়।

(6) পরবর্তী পদক্ষেপ হল টেস্ট স্ট্যাটিস্টিক, T-এর গণনা করা মান যদি সংকটপূর্ণ ক্ষেত্রে বা Critical region-এ থাকে তাহলে H_0-কে বাতিল করা ; নতুবা H_0-কে বাতিল না করা। T-এর গণনা

(8) গুরুত্বের স্তর (Level of Significance) : সর্বোচ্চ সম্ভাবনা যার ভিত্তিতে একটি সত্য নাল অনুমান (H_0) বাতিল করা হয় তাকে পরীক্ষার ক্ষেত্রে গুরুত্বের স্তর বলে অভিহিত করা হয় এবং তা α দ্বারা নির্দেশিত হয়। সিদ্ধান্ত নিয়মগুলি রচনার ক্ষেত্রে পরিসংখ্যাগত সিদ্ধান্তের গুরুত্ব অনুযায়ী গুরুত্বের স্তর (Level of significance) ইচ্ছামত (arbitrarily) স্থির করা হয়। প্রথাগতভাবে গুরুত্বের স্তর 5% অথবা 1% ধরা হয়। যদিও অন্যান্য স্তর যেমন 2% অথবা $\frac{1}{2}\%$ ও ব্যবহার করা হয়। প্রথম ধরনের ত্রুটি (Type 1 error) করার সম্ভাবনার উচ্চতর সীমা অর্থাৎ 'critical region'-এর আয়তন নির্দেশ করার জন্য গুরত্বের স্তর, α ব্যবহার করা হয়।

(9) প্রথম ধরনের ত্রুটি (Error of first type) : নাল অনুমানটি প্রকৃতই সত্য হওয়া সত্ত্বেও পরীক্ষার দ্বারা তা বাতিল করা হলে যে ত্রুটি হল তাকে প্রথম ধরনের ত্রুটি বলে। 'Critical region'টি এমনভাবে নির্ধারিত যে প্রথম ধরনের ত্রুটির সম্ভাবনা কিন্তু পরীক্ষার গুরুত্বের স্তরকে ছাড়িয়ে যায় না।

(10) দ্বিতীয় ধরনের ত্রুটি (Error of second type) : অকার্যকর অনুমানটি প্রকৃতই মিথ্যা হওয়া সত্ত্বেও পরীক্ষার ভিত্তিতে তা গৃহীত হলে যে ত্রুটি হয় তাকে দ্বিতীয় ধরনের ত্রুটি বলে। দ্বিতীয় ধরনের ত্রুটির সম্ভাবনা কার্যত বিকল্প অনুমান (H_1)-এর বিশেষভাবে উল্লেখিত মানের উপর নির্ভর করে এবং পরীক্ষার দক্ষতা মূল্যায়নে ব্যবহৃত হয়।

(11) পরীক্ষার ক্ষমতা (Power of the test) : একটি মিথ্যা নাল অনুমান বাতিল করার সম্ভাবনাকে পরীক্ষাটির ক্ষমতা বলা হয়। সুতরাং 'ক্ষমতা' (power) হল পরীক্ষার দ্বারা সঠিক উপসংহার টানা যখন নাল অনুমানটি মিথ্যা। বিকল্প অনুমান (H_1)-এর সঙ্গে সঙ্গতিপূর্ণ প্যারামিটারে বিশেষভাবে স্থিরিকৃত একটি মানের জন্য ক্ষমতা = 1—দ্বিতীয় ধরনের ত্রুটির সম্ভাবনা। বিশেষভাবে স্থিরিকৃত সকল বিকল্পগুলির পরিপ্রেক্ষিতে 'ক্ষমতা' (power) যদি লেখচিত্রের সাহায্যে প্রকাশ করা হয় তাহলে যে রেখাটি পাওয়া যায় তাকে ক্ষমতা রেখা (power curve) বলে অভিহিত করা হয়।

(গ) অনুমানের বা গুরুত্বের পরীক্ষায় বিভিন্ন পদক্ষেপ (Steps in test of Hypothesis or Significance) :

(1) প্রথম পদক্ষেপ হল নাল অনুমান H_0 এবং বিকল্প অনুমান H_1 স্থির করা বা উপস্থাপিত করা। প্রদত্ত সমস্যার পরিপ্রেক্ষিতে H_0 এবং H_1 স্থির করতে হয়। নাল অনুমান সাধারণত সমগ্রক সংশ্লিষ্ট কোন প্যারামিটারের মান সম্পর্কিত অনুমান হয়ে থাকে। H_0 ($\theta = \theta_0$), যেখানে θ হল একটি প্যারামিটার এবং θ_0 প্যারামিটারটির অনুমিত মান। এবার বিকল্প অনুমান, H_1 নিম্নের যে কোন একভাবে হতে পারে $H_1(\theta \neq \theta_0), H_1(\theta > \theta_0), H_1(\theta < \theta_0)$ বিকল্প অনুমানের ধরনটি কার্যত এক-প্রান্ত পরীক্ষা (One-tail Test) হবে না দুই-প্রান্ত পরীক্ষা (Two-tail test) হবে তা স্থির করে দেয়।

(3) অনুমান পরীক্ষা বা গুরুত্ব পরীক্ষা (Test of Hypothesis or Test of significance) : বিচার্য অনুমানটি গ্রহণ করা হবে কিম্বা বাতিল করা হবে এ বিষয়ে সিদ্ধান্ত নেওয়ার জন্য একগুচ্ছ নিয়মাধীন এক প্রক্রিয়া হল অনুমান পরীক্ষা (Test of Hypothesis) বা গুরুত্ব পরীক্ষা (Test of significance)।

(4) নাল বা অকার্যকর অনুমান (Null hypothesis) : এ হল পরিসংখ্যানগত এমন একটি অনুমান যে অনুমানের বলবৎযোগ্যতা কার্যত নমুনা পর্যবেক্ষণসমূহের ভিত্তিতে (on the basis of sample observations) সম্ভাব্য বাতিল করার পক্ষে পরীক্ষিত হয়। নাল অনুমান সাধারণত H_0 দ্বারা প্রকাশ করা হয় এবং বিকল্প অনুমানগুলির পরিপ্রেক্ষিতে পরীক্ষা করা হয় অনুমান পরীক্ষা (Test of hypothesis) কার্যত শুধু নাল অনুমান গ্রহণ অথবা বাতিলকরণ সম্পর্কিত হয়।

(5) বিকল্প অনুমান (Alternative hypothesis) : একটি পরিসংখ্যানগত অনুমান বা নাল অনুমান থেকে পৃথক বা ভিন্নতাকে একটি বিকল্প অনুমান বলে এবং তা সাধারণত H_1 দ্বারা নির্দেশিত হয়। বিকল্প অনুমান পরীক্ষা করা হয় না কিন্তু এই অনুমান গ্রহণ করার (বা বাতিল করার) সিদ্ধান্তটি কার্যত নির্ভর করে নাল অনুমানটি বাতিল করার (বা গ্রহণ করার) উপর। নাল অনুমানের বিপরীত হয় বিকল্প অনুমান। যথাযথ সংকটপূর্ণ এলাকা (Critical region) নির্বাচন কার্যত বিকল্প অনুমানের ধরনের উপর নির্ভর করে।

(6) টেস্ট স্ট্যাটিস্‌টিক (Test statistic) : নমুনা পর্যবেক্ষণসমূহের একটি পরিমাপক বা একটি স্ট্যাটিস্‌টিক যার গণনাকৃত মানের ভিত্তিতে H_0 গ্রহণ করা অথবা বাতিল করার বিষয়ে চূড়ান্ত সিদ্ধান্ত নির্ধারিত হয় তাকে টেস্ট স্ট্যাটিস্‌টিক বলা হয়। অতি সতর্কতার সহিত যথোপযুক্ত টেস্ট স্ট্যাটিস্‌টিকের নমুনা বিভাজন সম্পর্কে ধারণা থাকা আবশ্যক। টেস্ট স্ট্যাটিস্‌টিকের মান যদি সঙ্কটপূর্ণ এলাকার (critical region) মধ্যে পড়ে তাহলে অকার্যকর অনুমান বাতিল করা হয়।

(7) সংকটপূর্ণ এলাকা (Critical region) : সংকটপূর্ণ এলাকা বা 'Critical region' বলতে স্টেট স্ট্যাটিস্‌টিকের একটি সেট বা দলকে বোঝায় যেগুলি নাল অনুমান বাতিল করার পথ প্রদর্শন করে। একটি সত্য নাল অনুমান পরীক্ষা সাপেক্ষে বাতিল হওয়ার সম্ভাবনা প্রায়শই critical region-এর আকার নির্দিষ্ট করে। জ্যামিতিকভাবে n আকার বিশিষ্ট $(x_1, x_2 \dots x_n)$, একটি নমুনা একটি বিন্দুর (x) দ্বারা প্রকাশ করা হয় এবং এই বিন্দুকে নমুনা বিন্দু (Sample point) বলে এবং যে জায়গায় বা সমতলে সম্ভাব্য সকল নমুনা বিন্দুগুলি থাকে তাকে 'স্যাম্পল স্পেস' (Sample space) বলে (w)। অতএব 'Critical region'-কে (Sample space) w-এর সেই সকল নমুনা বিন্দুর একটি উপদল (Subset) হিসাবে সংজ্ঞায়িত করা যায় যে সকল নমুনা বিন্দুগুলির জন্য নাল অনুমান বাতিল করতে হয়।

২.৬ পরিসংখ্যাগত অনুমানসমূহের পরীক্ষা (Test of Hypothesis)

বহু ব্যবহারিক সমস্যার ক্ষেত্রে পরিসংখ্যায়কদের নমুনা পর্যবেক্ষণসমূহের ভিত্তিতে পরিসংখ্যাগত সমগ্রক সম্পর্কে সিদ্ধান্ত গ্রহণ করার প্রয়োজন হয়। **উদাহরণস্বরূপ**, একটি সমসম্ভাবনা নমুনা দেওয়া থাকলে, যে সমগ্রক থেকে নমুনাচয়ন করা হয়েছে তা যৌগিক গড় = 40 এবং সম্যকবিচ্যুতি = 3 সহ একটি স্বাভাবিক বিভাজন (Normal distribution) কিনা সে বিষয়টি ঠিক করার প্রয়োজন হতে পারে, এরূপ বিষয়ে সিদ্ধান্তে পৌঁছানোর প্রচেষ্টায় সমগ্রকের বৈশিষ্ট্যসমূহ সম্পর্কে কিছু অনুমান ও আন্দাজ করা আবশ্যক হয় বিশেষত সম্ভাবনা বিভাজন অথবা প্যারামিটারের মান সম্পর্কে কিছু অনুমান করার প্রয়োজন পড়ে। এরূপ অনুমান অথবা সমগ্রক সম্পর্কে বক্তব্যকে পরিসংখ্যাগত অনুমান (Statistical hypothesis) বলা হয়। নমুনা বিশ্লেষণ করে অনুমানের বলবৎযোগ্যতা পরীক্ষা করা হয়। যে প্রক্রিয়ার সাহায্যে কোন বিশেষ অনুমান সত্য কিম্বা সত্য নয় তা নির্ধারণ করা হয় তাকে অনুমানের পরীক্ষা (Test of hypothesis) অথবা গুরুত্বের পরীক্ষা (Test of significance) বলা হয়।

(ক) পরিসংখ্যাগত অনুমিতি বা সিদ্ধান্ত (Statistical Inference) ঃ নমুনা বিশ্লেষণের মুখ্য উদ্দেশ্য হল নমুনা নামে সমগ্রকের একটি অংশ পরীক্ষা করে সমগ্রক সম্পর্কে সিদ্ধান্তে উপনীত হওয়া। এভাবে গৃহীত সিদ্ধান্তকে পরিসংখ্যানগত সিদ্ধান্ত (Statistical inference) বলা হয়। নমুনাগত তথ্যাবলীর ভিত্তিতে সমগ্রকের পরিসংখ্যাগত পরিমাপক (সমগ্রক যৌগিক গড়, সমগ্রক ভেদমান ইত্যাদি) বা প্যারামিটার গণনা করা এবং প্যারামিটারগুলি সম্পর্কে অনুমান পরীক্ষা করা—এই দুটি হল পরিসংখ্যানগত সিদ্ধান্তসমূহের দুটি প্রধান শাখা। সুতরাং সমগ্রকের প্রকৃতি বা গঠন সম্পর্কে অথবা প্যারামিটার সম্পর্কে অনুমানকে পরিসংখ্যাগত অনুমান বলা হয়। এরূপ অনুমান সরল অথবা যৌগিক (Simple বা Composite) হতে পারে।

(খ) কতকগুলি উপযোগী ধারণা (Some useful concepts) :

(1) সরল অনুমান (Simple hypothesis) : পরিসংখ্যাগত যে অনুমান সমগ্রককে সম্পূর্ণরূপে বিশেষভাবে উল্লেখ করে (অর্থাৎ সম্ভাবনা বিভাজন, অপেক্ষকগত ধরন এবং সব প্যারামিটারগুলি জ্ঞাত) তাকে সরল অনুমান বলে।

(2) যৌগিক অনুমান (Composite hypothesis) : পরিসংখ্যাগত যে অনুমান সমগ্রককে সম্পূর্ণরূপে বিশেষভাবে উল্লেখ করে না। (অর্থাৎ সম্ভাবনা বিভাজনের ধরন অথবা কিছু প্যারামিটার অজ্ঞাত) তাকে যৌগিক অনুমান বলে।

universe) বলে। পক্ষান্তরে সমগ্রকের অন্তর্ভুক্ত এককের সংখ্যা অসীম হলে তাকে **অসীম সমগ্রক** (Infinite population or universe) বলে। **উদাহরণ স্বরূপ,** কোন এক অঞ্চলে অবস্থিত গৃহসমূহের সমগ্রক হচ্ছে সসীম সমগ্রক, পক্ষান্তরে গৃহের বিভিন্ন স্থানে আবহাওয়াগত চাপের সমগ্রক (Population of atmospheric pressures) হল অসীম।

সমগ্রক আবার অস্তিত্ব সম্পন্ন (esistent) অথবা কাল্পনিক (hypothetical) হতে পারে। সমগ্রকের একগুলির বাস্তব অস্তিত্ব থাকলে সেই সমগ্রককে অস্তিত্ব সম্পন্ন সমগ্রক বলে। পক্ষান্তরে সমগ্রক যদি কাল্পনিক একক নিয়ে গঠিত হয় তাকে কাল্পনিক সমগ্রক বলা হয়।

নমুনাচয়ন (Sampling) সাধারণভাবে দুটি শ্রেণীতে ভাগ করা যায়। মানসিক (Subjective) এবং উদ্দেশ্য সম্পর্কিত বা বিষয়গত (Objective)। মানসিক নমুনাচয়নের ক্ষেত্রে এককগুলির নির্বাচন নমুনাচয়নকারীর বিচার বিবেচনা অথবা স্বেচ্ছানুসারে হয়ে থাকে। বিষয়গত নমুনাচয়নের ক্ষেত্রে কিন্তু এক বিশেষ নিয়ম অনুসারে এককগুলি নির্বাচন করা হয়। সুতরাং বিষয়গত নমুনাচয়ন কার্যতঃ নমুনাচয়নকারীর বিচার বিবেচনা নিরপেক্ষ হয়।

নমুনাচয়ন আবার সম্ভাবনা নিরপেক্ষ, সম্ভাবনা ভিত্তিক ও মিশ্র এই তিন শ্রেণীতে ভাগ করা যায়। এককগুলি চয়নের ক্ষেত্রে যখন কোনরূপ সম্ভাবনার বিষয়টি জড়িত থাকে না। কেবল বিশেষ নিয়ম অনুসারে নমুনাচয়ন করা হয় তা হল **সম্ভাবনা-নিরপেক্ষ নমুনাচয়ন** (Non-Prabalistic Sampling)। নমুনাচয়নের ক্ষেত্রে যখন সমগ্রকের প্রতিটি একক নির্বাচিত হওয়ার সমসম্ভাবনা থাকে তাকে **সমসম্ভাবনা নমুনাচয়ন** (Random sampling) বা **সম্ভাবনাভিত্তিক নমুনাচয়ন** (Probalistic sampling) বলা হয়। অংশত সম্ভাবনাভিত্তিক এবং অংশত সম্ভাবনা নিরপেক্ষ নমুনাচয়ন হল **মিশ্র নমুনাচয়ন** (Mixed sampling)।

২.৫ নমুনাচয়ন পদ্ধতি (Sampling Methods)

বিভিন্ন পদ্ধতিতে বা উপায়ে নমুনা চয়ন করা হয় নমুনাচয়নের কয়েকটি গুরুত্বপূর্ণ পদ্ধতি সম্পর্কে ধারণা দেওয়া যেতে পারে। নমুনাচয়নের এই পদ্ধতিগুলি হল ঃ (1) সরল সমসম্ভাবনাযুক্ত নমুনাচয়ন পদ্ধতি (Simple Random Sampling Method), (2) উদ্দেশ্যমূলক নমুনাচয়ন পদ্ধতি (Purposive Sampling Method), (3) স্তরবিন্যস্ত নমুনাচয়ন পদ্ধতি (Stratified Sampling Method), (4) পদ্ধতি মাফিক নমুনাচয়ন পদ্ধতি (Systematic Sampling Method), (5) বহুপর্যায় নমুনাচয়ন পদ্ধতি (Multi-stage Sampling Method)।

এই সমস্ত পদ্ধতি বিস্তারিতভাবে আলোচিত হয়েছে প্রথম এককে। সেজন্য তার পুনরাবৃত্তি করা হল না।

অনুশীলনী

1. নমুনাচয়নের বিভিন্ন পদ্ধতিগুলি কি কি? সমগ্রক থেকে নমুনাচয়ন কিভাবে করা হয়?

২.৪ সমগ্রক থেকে নমুনাচয়ন (Determination of Sample from Population)

নমুনাচয়ন (Sampling) বলতে পরিসংখ্যানগত সমগ্র বিষয় থেকে একটি অংশ নির্বাচনকে বোঝায় যার সাহায্যে সমগ্র বিষয় সম্পর্কে সঠিক তথ্য অনুধাবন করা সম্ভব হয়। সকল একক বা ব্যষ্টির বিশেষ এক চরিত্র বা প্রকৃতি সম্পর্কিত পরিসংখ্যানগত তথ্যাবলীর সমষ্টি বা সমগ্র পরিসংখ্যানগত তথ্যমালাকে **সমগ্রক** (Population or universe) বলে। এবং যে অংশ সমগ্রকের প্রকৃতি নিশ্চিত করার জন্য চয়ন করা হয় তাকে **নমুনা** (Sample) বলা হয়।

পরিসংখ্যানগত অনুসন্ধানের জন্য ব্যষ্টিসমূহের একটি গোষ্ঠীর কিছু প্রকৃতি বিশ্লেষণ করার প্রয়োজন হয়, এরূপ অনুসন্ধানাধীন সমগ্র গোষ্ঠী হল সমগ্রক। সমগ্রকের সদস্যরা বা এককগুলি হতে পারে যেমন কোন এক শিল্পের কর্মচারীগণ বা চুপড়ির আপেলগুলি, বা একটি এলাকায় আবাদযোগ্য জোতগুলি ইত্যাদি।

প্রায়শই অর্থ, সময় ও শ্রমশক্তির সীমাবদ্ধতার দরুণ সমূহ সমগ্রক বিচার-বিশ্লেষণ করা বাস্তবসম্মত হয় না অথবা সমগ্রকের এককসমূহ অসীম হলে সমূহ সমগ্রক (whole population) বিশ্লেষণ করা যায় না। সুতরাং সমগ্রকে একটি অংশ চয়ন করে বিচার-বিশ্লেষণের দ্বারা সমগ্রকের প্রকৃতি বা বৈশিষ্ট্য নির্ধারণ করা যায়। সমগ্রকের একটি অংশ যার দ্বারা সমূহ সমগ্রক প্রদর্শিত হয় তা হল নমুনা (Sample)।

যথাযথ নমুনা নিয়ে তদন্ত করাকে নমুনা তদন্ত (Sample survey) বলা হয়। আবার পূর্ণ তদন্তের ক্ষেত্রে (in case of complete enumeration or complete census) সমূহ সমগ্রকের প্রতিটি একক নিয়ে তদন্ত করা হয়। সাধারণত নমুনা তদন্তই অধিকতর পছন্দ করা হয়। এর কয়েকটি কারণ উল্লেখ করা যায়।

(1) ব্যয় সংক্ষেপ হয় (Reduction of Cost)

(2) দ্রুত তদন্ত সম্পন্ন করা যায় (Greater speed)

(3) অধিকতর সঠিকতার মান বজায় করা যায় (Greater accuracy)

(4) এই পদ্ধতি হল সঠিকতার পরিমাপক (Measure of accuracy)

(5) এর অধিকতর প্রয়োগ যোগ্যতা (Greater applicability) রয়েছে।

সুতরাং পূর্ণতদন্ত অপেক্ষা আংশিক তদন্ত অধিকতর গ্রহণযোগ্য এবং বহুল প্রচলিত। যথাযথ পদ্ধতিতে নমুনা সংগ্রহ করে তদন্ত করা হয়। এবং তদন্তলব্ধ সিদ্ধান্তকে সমূহ সমগ্রকের ক্ষেত্রে সমানভাবে প্রযোজ্য বলে গণ্য করা হয়। সমূহ সমগ্রক সম্পর্কে এরূপ সিদ্ধান্ত গ্রহণের পদ্ধতিকে **আংশিক চয়ন পদ্ধতি** বা **নমুনা চয়ন পদ্ধতি** (Sampling method) বলা হয়।

সমগ্রকের অন্তর্ভুক্ত এককের সংখ্যা সীমিত হলে তাকে **সীমিত সমগ্রক** (finite population or

অনুশীলনী

1. নিম্নলিখিত তথ্য থেকে সহগতি সহগাঙ্ক নির্ণয় কর :

X	:	6	2	10	4	8
Y	:	9	11	5	8	7

[Ans. +0·98]

2. X ও Y চলক দুটির মান নিম্নে দেওয়া হল :

X	:	–3	–1	+1	+3
Y	:	9	1	1	9

দেখাও যে সহগতি সহগাঙ্ক হচ্ছে শূন্য। চলক দুটি কি নিরপেক্ষক (independent) যদি তা না হয় তাহলে সহগতি সহগাঙ্ক শূন্য হওয়ার কারণ কি? [Hints : সম্পর্কটি সরল নয় (non-linear) : $y = x^2$ সহগতি সহগাঙ্ক কিন্তু সরল যোগসূত্রের পরিমাপক]

3. নিম্নে প্রদত্ত তথ্যগুলি থেকে (i) সহগতি সহগাঙ্ক, (ii) y-এর পরিপ্রেক্ষিতে x-এর নির্ভরণ সমীকরণ নির্ণয় কর।

$$\sum x = 56, \sum y = 40, \sum x^2 = 524, \sum y^2 = 256, \sum xy = 364, n = 8$$

[Ans. +0·98, $x = 1{\cdot}5y - 0{\cdot}5$]

4. সহগতি সহগাঙ্ক যে নির্ভরণ গুণাঙ্কদ্বয়ের গুণোত্তর গড় তা দেখাও।

5. একটি প্রতিযোগিতায় দুজন বিচারক 7 জন ছাত্রকে নিম্নলিখিতভাবে অনুক্রমিক মাত্রা প্রদান করল [Ans. $\bar{x} = 1, \bar{y} = 2, \sigma_y = 2, r = -\frac{\sqrt{3}}{2}$:]

ছাত্র	:	A	B	C	D	E	F	G
বিচারক I-এর মাত্রাগত তথ্য	:	2	1	4	5	3	7	6
বিচারক π-এর মাত্রাগত তথ্য	:	3	4	2	5	1	6	7

অনুক্রমিক সহগাঙ্ক নির্ণয় কর : [Ans. $r_R = 0{\cdot}64$]

6. x ও y চলক দুটির মধ্যে সহগতি সহগাঙ্ক, r = 0·60, এবার যদি $\sigma_x = 1{\cdot}50$, $\sigma_y = 2{\cdot}00$, $\bar{x} = 10, \bar{y} = 20$ হয় তাহলে (1) x-এর সাপেক্ষে y-এর (ii) y-এর সাপেক্ষে x-এর নির্ভরণ সমীকরণ দুটি নির্ণয় কর। [Ans. $y = 0{\cdot}8x + 12$; $x = 0{\cdot}45y + 1$]

7. x ও y চলকদ্বয়ের নির্ভরণ সমীকরণ দুটি হল : $y = 5{\cdot}6 + 1{\cdot}2x$ এবং $x = 12{\cdot}5 + 0{\cdot}6y.x$ ও y-এর যৌগিক গড় ও উহাদের মধ্যে সহগতি সহগাঙ্ক নির্ণয় কর।

[Ans. 56·64, 73·57 + 0·85]

8. দুটি নির্ভরণ সমীকরণ হল $x + 2y = 5$ এবং $2x + 3y = 8$ এবং $\sigma_x^2 = 12\, \bar{x}, \bar{y}, \sigma,$ এবং r-এর মান নির্ণয় করো।

$x - \bar{x}\, b_{xy}(y - \bar{y})$ যেখানে

$$\bar{x} = \frac{\sum x}{n} = \frac{24}{4} = 6,\ \bar{y} = \frac{\sum y}{n} = \frac{44}{4} = 11$$

সহভেদমান $(x, y) = \frac{306}{4} - \frac{24}{4} . \frac{44}{4} = 10{\cdot}5$

ভেদমান $(y) = \sigma_y^2 = \frac{574}{4} - \left(\frac{44}{4}\right)^2 = 22{\cdot}5$

$$b_{xy} = \frac{10{\cdot}5}{22{\cdot}5} = 0{\cdot}467$$

$\therefore$ নির্ণেয় নির্ভরণ সমীকরণটি হল : $x - 6 = 0{\cdot}467\ (4 - 11)$

অথবা $x = 0{\cdot}467y + 0{\cdot}86$

$y = 6$ হলে $x = 0{\cdot}467 \times 6 + 0{\cdot}86 = 3{\cdot}7$

উদাহরণ 2 : যদি $4u = 2x + 7$ এবং $6v = 2y - 15$ এবং x-এর উপর y-এর নির্ভরণ গুণাঙ্ক 3 হয় তাহলে u-এর উপর v-এর নির্ভরণ গুণাঙ্ক নির্ণয় কর।

সমাধান : $u = \frac{1}{2}x + \frac{7}{4}$ এবং $v = \frac{1}{3}y - \frac{5}{2}$

$\therefore\ u - \bar{u} = \frac{1}{2}(x - \bar{x})$ এবং $v - \bar{v} = \frac{1}{3}(y - \bar{y})$

ভেদমান $(u) = \left(\frac{1}{2}\right)^2$ ভেদমান $(x) = \frac{1}{4}$ ভেদমান (x)

এবং সহভেদমান $(u, v) = \frac{1}{2} \cdot \frac{1}{3}$ সহভেদমান (x, y)

$$\therefore\ b_{uv} = \frac{(u, v)}{(u)} = \frac{\frac{1}{2} \cdot \frac{1}{3}\ (x, y)}{\frac{1}{4}\ (x)} = \frac{2}{3} b_{yx} = \frac{2}{3} \cdot 3 = 2$$

যখন $r = \pm 1$ তখন $\tan \boldsymbol{\theta} = 0$ অর্থাৎ $\boldsymbol{\theta} = 0$ এবং দুটি নির্ভরণ রেখা তখন পরস্পরের উপর সমপাতিত হয় (Coincide), আবার যখন r = 0, cot $\boldsymbol{\theta} = 0$ অর্থাৎ তখন নির্ভরণ রেখা দুটি পরস্পরকে লম্ব আকারে বিন্দুতে ছেদ করে।

(2) নির্ভরণ গুণাঙ্ক দুটি হল : নিম্নরূপ

$$b_{yx} = r\frac{\sigma_y}{\sigma_x} \text{ এবং } b_{xy} = r\frac{\sigma_x}{\sigma_y}$$

$$\therefore \quad b_{yx} . b_{xy} = r\frac{\sigma_y}{\sigma_x} . r\frac{\sigma_x}{\sigma_y} = r^2$$

$$\therefore \quad |r| = \sqrt{b_{yx} . b_{xy}}$$

সুতরাং সংখ্যাগতভাবে সহগতি সহগাঙ্ক হল নির্ভরণ গুণাঙ্ক দুটির গুণোত্তর গড় এবং r-এর চিহ্ন হবে গুণাঙ্ক দুটির সাধারণ চিহ্ন (sign)

(3) ধনাত্মক সংখ্যাগুলির যৌগিক গড় যেহেতু সংখ্যাগুলির গুণোত্তর গড় অপেক্ষা অধিক হয় অথবা উহার সমান হয় সেইহেতু,

$$\frac{|b_{yx}| + |b_{xy}|}{2} \geq \sqrt{|b_{yx}||b_{xy}|}$$

(যেহেতু b_{yx}, b_{xy} একই চিহ্ন যুক্ত হয়)

$$= \sqrt{r^2} = |r|$$

সহগাঙ্ক, r নির্ভরণ গুণাঙ্কের সংখ্যাগত মানগুলির যৌগিক গড় অপেক্ষা অধিক হতে পারে না।

উদাহরণ 1. নিম্নলিখিত তথ্যগুলি থেকে y-এর উপর x-এর নির্ভরণ সমীকরণ নির্ণয় কর এবং $y = 6$ হলে x-এর মান নির্ধারণ কর।

$$\sum x = 24 \qquad \sum y = 44 \qquad \sum xy = 306$$

$$\sum x^2 = 164 \qquad \sum y^2 = 574 \qquad n = 4$$

সমাধান ঃ y-এর উপর x-এর নির্ভরণ সমীকরণের সাধারণ রূপ হল :

যেহেতু ভেদমান $(e) \geq 0$ সেইহেতু অথবা,

অথবা,

অথবা, এই ফল পূর্বেই প্রতিষ্ঠা করা হয়েছে।

(3) সহভেদমান $(x,\ \mathrm{e})$ যেহেতু $\bar{e} = 0$

$$=$$

$$= \frac{1}{n}(0 - \bar{x}.0) = 0$$

সুতরাং $r_{xe} = 0$

সুতরাং, e-কে y এবং সেই অংশ হিসাবে দেখান যেতে পারে যে অংশটির সঙ্গে x-এর কোন সম্পর্ক থাকে না।

(ঠ) নির্ভরণ রেখা ও গুণাঙ্ক সম্পর্কিত কয়েকটি ফল (Some results relating to Regression Line & Coefficient) :

(1) দুটি নির্ভরণ রেখা হল :

$$Y = y + r\frac{\sigma_y}{\sigma_x}(x - \bar{x}) \quad \ldots \quad \text{(i)}$$

এবং $X = \bar{x} + r\frac{\sigma_x}{\sigma_y}(y - \bar{y}) \quad \ldots \quad \text{(ii)}$

(i)-এর ঢাল (Gradiant) হল $r\frac{\sigma_y}{\sigma_x} = m_1$ (ধরি) এবং (ii)-এর ঢাল হল $r\frac{\sigma_x}{\sigma_y} = m_2$ ধরি, এবার যদি দুটি নির্ভরণ রেখার মধ্যবর্তী সূক্ষ্ম কোণ (acute angle) হয় **θ,** তাহলে

$$\theta = \tan^{-1}\left|\frac{m_2 - m_1}{1 + m_1 m_2}\right| = \tan^{-1}\left|\frac{\frac{\sigma_y}{r\sigma_x} - r\frac{\sigma_y}{\sigma_x}}{1 + \frac{\sigma_y}{r\sigma_x}\cdot\frac{\sigma_y}{r\sigma_x}}\right|$$

$$= \tan^{-1}\left|\frac{1 - r^2}{r}\cdot\frac{\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2}\right|$$

এবং অপর কোণটি হবে **π − θ**

$$\therefore r^2 = \frac{(Y)}{(y)} = \frac{\sigma_y^2}{\sigma_y^2}$$

অথবা, $|r| = \dfrac{\sigma_Y}{\sigma_y}$

সহগতি সহগাঙ্কের সংখ্যাগত মানকে x-এর উপর y-এর নির্ভরণ রেখার পরিপ্রেক্ষিতে y-এর সমক বিচ্যুতির অনুপাত (Proportion of the total variability of y) হিসাবে দেখান যায়। অনুরূপভাবে দেখান যায় যে x-এর নির্ভরণ রেখার জন্য $|r|\dfrac{\sigma_X}{\sigma_x}$, যেখানে $\boldsymbol{\sigma}_X$ হল x -এর নির্ধারিত মানগুলির সমক বিচ্যুতি (Standard deviation) এবং $\boldsymbol{\sigma}_x$ হল x-এর প্রদত্ত মানগুলির সমক বিচ্যুতি।

r^2-কে নির্ধারণ সহগাঙ্ক (Coefficient of determination) বলা হয় এবং নির্ধারণ সূত্র হিসাবে সরল নির্ভরণ সমীকরণগুলির ব্যবহারিক উপযোগের পরিমাপকও বলে গণ্য করা হয়।

(2) যখন $e = 0$ তখন X-এর উপর y-এর সরল নির্ভরণ থেকে ভেদমান (e) হয় $\sigma_y^2(1 - r^2)$। এক্ষেত্রে e-এর সমক বিচ্যুতি (Standard deviation) হল $\sigma_y\sqrt{1-r^2}$ এবং একে X-এর উপর Y-এর সরল নির্ভরণ থেকে y-এর গণনার সমক বিচ্যুতি (Standard error of estimation of y) বলা হয়।

ভেদমান (e) $= \dfrac{1}{n}\sum_i e_i^2$, যেহেতু $e = 0$

$$= \frac{1}{n}\sum_i (y_i - Y)^2 = \frac{1}{n}\sum_i \left\{(y_i - \overline{Y}) - r\frac{\sigma_y}{\sigma_x}(x_i - \overline{x})\right\}^2$$

$$= \frac{1}{n}\sum_i (y_i - \overline{y})^2 - 2r\frac{\sigma_y}{\sigma_x}\frac{1}{n}\sum_i (y_i - \overline{x})(x_i - \overline{x}) + r^2\frac{\sigma_y^2}{\sigma_x^2}\frac{1}{n}\sum_i (x_i - \overline{x})^2$$

$$= \sigma_y^2 - 2r\frac{\sigma_y}{\sigma_x}r.\sigma_x\sigma_y + r^2\frac{\sigma_y^2}{\sigma_x^2}.\sigma_y^2 = \sigma_x^2 - 2r^2\sigma_y^2 + r^2\sigma_y^2$$

$$= \sigma_y^2(1 - r^2)$$

$\therefore\ x - \bar{x} = c(u - \bar{u})$ এবং

অতএব ভেদমান (x) = c^2 ভেদমান (u) এবং ভেদমান (y) = d^2 ভেদমান (v)

এবং সহভেদমান (x, y) = cd সহভেদমান (u, v)

অনুরূপভাবে, $b_{xy} = \frac{c}{d} b_{uv}$

(2) x-এর উপর y-এর নির্ভরণ রেখা থেকে দেখা যায় যে y-এর নির্ধারিত মান (predicted value) হল Y_i যখন $x = x_i$, 3.1(a) নং বিক্ষেপন চিত্রে x-এর উপর y-এর নির্ভরণ রেখার উপর নির্ধারিত বিন্দু (x_i y_i) এবং প্রদত্ত বিন্দু (x_i y_i) দেখান হয়েছে।

ধরা যাক, $Y_i = \bar{y} + r\frac{\sigma_y}{\sigma_x}(x_i - \bar{x})$

$$\therefore \sum_i Y_i = n\bar{y} + r\frac{\sigma_y}{\sigma_x}\sum_i (x_i - \bar{x}) \ \left[\because \sum(x_i - \bar{x}) = 0\right]$$

y-এর নির্ধারিত মানগুলির যৌগিক গড় উহাদের প্রদত্ত মানগুলির যৌগিক গড়ের সমান হয়েছে। এরূপ অনুমান সাপেক্ষে বিচ্যুতিসমূহের গড় হবে।

x চলকের ক্ষেত্রেও অনুরূপ ফল পাওয়া যায়।
গড় বিচ্যুতি (Average error) শূন্য হলে ($\bar{e} = 0$) আরও কয়েকটি ফল পাওয়া যায় :

(1) ভেদমান (Y) (যেহেতু $\bar{Y} = \bar{y}$)

হল y-এর উপর x-এর নির্ভরণ রেখা যা y-এর মান থেকে x-এর মান নির্ধারণের জন্য ব্যবহার করা হয়। এই রেখা পাওয়ার জন্য নিম্নলিখিত সম্পর্কটি ধরা যেতে পারে।

$$x = c + dy \qquad ..... \quad (v)$$

যেখানে c + dy_i হল x-এর নির্ধারিত মান (Predicted value) যখন y = y_i. অতএব c ও d নির্ধারণের জন্য x-এর হিসাবের বিচ্যুতি বা ভ্রান্তিসমূহের বর্গের সমষ্টিকে নূন্যতম করতে হয়।

অর্থাৎ $\sum_i (x_i - c - dy_i)^2 = \sum_i e_i^2$-কে c ও d-এর পরিপ্রেক্ষিতে নূন্যতম করতে হয়।

পূর্বের অনুরূপ প্রক্রিয়ায় অগ্রসর হয়ে y-এর উপর x-এর নির্ভরণ সমীকরণ নিম্নবর্ণিতরূপে পাওয়া যায়।

.... (vi)

যেখানে $b_{xy} = r\dfrac{\sigma_x}{\sigma_y}$ হল y-এর উপর x-এর নির্ভরণ গুণাঙ্ক এবং তা হল y বৃদ্ধির জন্য x-এর বৃদ্ধি 3.1(b) নং চিত্রে (vi) নং সমীকরণের জ্যামিতিক প্রতিরূপ হিসাবে y-এর উপর x-এর নির্ভরণ রেখাটি দেখান হয়েছে।

মন্তব্য : (1) নির্ভরণ রেখা বিন্দুতে পরস্পরকে ছেদ করে যেহেতু দেখা যায় যে উভয় সমীকরণ (iv এবং vi) $x = \bar{x}$ এবং দ্বারা সিদ্ধ হয়।

(2) দুটি নির্ভরণ রেখা হল পৃথক দুটি রেখা যেহেতু রেখা দুটি ভিন্ন শর্তাধীনে উদ্ভূত হয়েছে। অবশ্য দুটি রেখা পরস্পরের উপর সমপাতিত হয় (coincide) যখন অর্থাৎ চলক দুটির মধ্যে সম্পর্ক যখন প্রকৃতই সরল (exactly linear) হয়।

(ট) কিছু গুরুত্বপূর্ণ ফল (Some Important Results) :

(1) গণনার সুবিধার জন্য মূল বিন্দু বা কেন্দ্র এবং পরিমাপ মাত্রা উভয়ই পরিবর্তন করা যায়। x ও y চলক দুটির কেন্দ্র (a, b) বিন্দুতে স্থানান্তরিত হয় এবং পরিমাপ যথাক্রমে c, d মাত্রায় পরিবর্তন করা হয় যখন এবং ,

তখন এবং $b_{xy} = \dfrac{c}{d} b_{uv}$ হবে

প্রমাণ : x = a + cu এবং y = b + dv

$$= \frac{\frac{1}{n}\sum_i x_i y_i - \overline{xy}}{\frac{1}{n}\sum_i x_i^2 - \bar{x}^2} = \frac{(x, y)}{(x)} = r\frac{\sigma_y}{\sigma_x}$$

সুতরাং, $b = r\frac{\sigma_y}{\sigma_x}$

সমীকরণ (ii) হতে পাই,

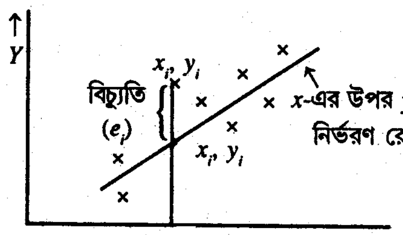
$$\bar{y} = a + b\bar{x}$$

অথবা,

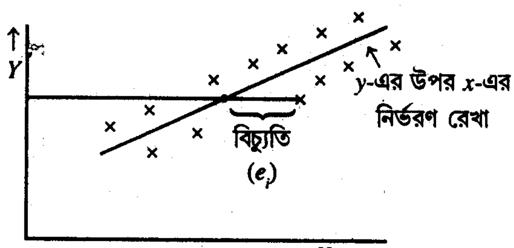
a এবং b-এর নিরূপিত মানগুলিতে সমীকরণ (i)-এ বসাইয়া x-এর উপর y-এর (y on x) সরল নির্ভরণ সমীকরণটি পাওয়া যায় এবং এই সমীকরণটি হল নিম্নরূপ :

$$y = \bar{y} + r\frac{\sigma_y}{\sigma_x}(x - \bar{x}) \quad \ldots \quad \text{(iv)}$$

সুতরাং, $b_{yx} = r\frac{\sigma_y}{\sigma_x}$ যেখানে b_{yx}-কে x-এর উপর y-এর নির্ভরণ গুণাঙ্ক (regression coefficient of y on x) বলা হয়। b_{yx} হল x-এর এক একক বৃদ্ধির জন্য y-এর বৃদ্ধি



চিত্র নং 3.1(a)



চিত্র নং 3.1(b)

3.1(a) নং বিক্ষেপণ চিত্রে x-এর উপর y-এর নির্ভরণ রেখাটি ও ভ্রান্তি (error) দেখান হয়েছে। অনুরূপভাবে লঘিষ্ঠ বর্গ পদ্ধতি অনুসারে আর একটি নির্ভরণ রেখা পাওয়া যায়—অপর এই রেখাটি

সুতরাং y_i-এর জন্য $a + bx_i$ নেওয়ার ক্ষেত্রে একটি ভ্রান্তি (error) দেখা দেয় এই ভ্রান্তি, e হল :

$e_i = y_i - (a + bx_i)$ একে গণনার বিচ্যুতি বা ভ্রান্তি (error of estimation) বলা হয়। লঘিষ্ঠ বর্গ পদ্ধতিতে a এবং b এমনভাবে নির্ধারিত হয়, যাতে a এবং b-এর পরিপ্রেক্ষিতে $\sum_i e_i^2 = \sum (y_i - a - bx_i)^2$ হয় ন্যূনতম। a এবং b নির্ধারণের জন্য সমীকরণগুলি হল নিম্নরূপ :

$$\frac{\partial}{\partial a}\left(\sum_i e_i^2\right) = 0 \text{ এবং } \frac{\partial}{\partial b}\left(\sum_i e_i^2\right) = 0$$

সমীকরণ দুটি থেকে পাওয়া যায় $\sum_i (y_i - a - bx_i) = 0$

অথবা, $\sum_i y_i = na + b\sum x_i \quad \ldots \quad$ (ii)

এবং

অথবা, $\sum_i x_i y_i = a\sum_i x_i + b\sum_i x_i^2 \quad \ldots \quad$ (iii)

$$\left[\sum_i x_i(y_i - a - bx_i) = 0\right]$$

$$\sum_i x_i \times (ii) - n \times (iii)$$

এই সমীকরণগুলিকে সুসম্বন্ধ সমীকরণ (Normal equations) বলে। সমীকরণ (iii)-এর সঙ্গে n গুণ করে এবং তা থেকে সমীকরণ (ii)-এর সঙ্গে -এর গুণফল বাদ দিলে

পাওয়া যায় :

$$n\sum_i x_i y_i - \left(\sum_i x_i\right)\left(\sum_i y_i\right) = b\left[n\sum_i x_i^2 - \left(\sum_i x_i\right)^2\right]$$

$$\therefore b = \frac{n\sum_i x_i y_i - \left(\sum_i x_i\right)\left(\sum_i y_i\right)}{n\sum_i x_i^2 - \left(\sum_i x_i\right)^2}$$

$$r_R = \frac{\frac{n^2-1}{12} - \frac{Tu+Tv}{2} - \frac{1}{2n}\sum_i d_i^2}{\sqrt{\frac{n^2-1}{12} - Tu}\sqrt{\frac{n^2-1}{12} - Tv}}$$

$$= \frac{2 \cdot 88895}{\sqrt{40 \cdot 46346}} = 0 \cdot 45$$

(ঝ) সরল নির্ভরণ বিশ্লেষণ (Linear Regression Analysis) : একটি চলক (ধরি) y-এর অপর একটি চলক (ধরি) x উপর নির্ভরণ (Regression) বলতে x-এর উপর y-নির্ভরতা বোঝায়। দ্বিচলক সম্পর্কিত বিশ্লেষণে একটি সমস্যা হল স্বাধীন চলক x-এর মান জানা থাকলে ঐ চলকটির উপর নির্ভরশীল চলক y-এর মান নির্ধারণ করা। এই সমস্যা খুবই সহজ সমাধানযোগ্য হয় যদি y-কে x-এর একটি গাণিতিক অপেক্ষক রূপে প্রকাশ করা যায় ; ধরি $y = f(x)$; তখন এই সমীকরণটিকে x উপর y-এর নির্ভরণ সমীকরণ (Regression equation) বলা হয়। সহজতম ক্ষেত্রে y-এর সঙ্গে x প্রকৃত অথবা প্রায় সরলভাবে সম্পর্কযুক্ত (linearly related) হয় তখন তা নিম্নরূপে প্রকাশ করা হয় :

$$y = a + bx$$

এই সম্পর্ক অনুসারে $a + bx_0$ হল y-এর নির্ধারিত মান যখন $x = x_0$ এই অংশে আলোচনা কেবল সরল নির্ভরণ এর ক্ষেত্রে সীমাবদ্ধ রাখা হল।

(ঞ) সরল নির্ভরণ সমীকরণ নির্ণয় (Derivation of Linear Regression Equation) :

ধরা যাক সরল নির্ভরণ সমীকরণ হল : $y = a + bx$ .... (i)

যেহেতু এই সমীকরণের ভিত্তিতে x-এর মানের জন্য y-এর মান নির্ধারণের উদ্দেশ্যে ব্যবহার করা হবে সেইহেতু x ও y-এর প্রদত্ত মানসমূহের ভিত্তিতে a এবং b ধ্রুবক দুটি নির্ধারণ বা গণনা করতে হয়। ধরা যাক n সংখ্যক মান যুগল (x_i, y_i) দেওয়া আছে যেখানে $i = 1, 2, ..., n$. এবং a, b নির্ধারণের জন্য সম্ভাব্য পদ্ধতিগুলির মধ্যে **লঘিষ্ঠ বর্গ পদ্ধতি** (Least Square Method) ব্যবহার করা বিধেয় যেহেতু এই পদ্ধতির বহু কাম্য বৈশিষ্ট্য রয়েছে।

যখন $x = x_i$ তখন y-এর প্রদত্ত মান (observered value) হল y_i কিন্তু নির্ধারিত মান (Predicated value) হল $a + bx_i$

উদাহরণ 2. এক পরীক্ষায় 9 জন ছাত্র ইংরেজী ও গণিতে নিম্নবর্ণিত নম্বর পেয়েছে। এক্ষেত্রে স্পেয়ারম্যানের অনুক্রমিক সহগাঙ্ক নির্ণয় করো :

ছাত্র (রোল নং)	:	1	2	3	4	5	6	7	8	9
ইংরেজীতে প্রাপ্ত নম্বর	:	45	60	32	45	32	32	58	56	47
গণিতে প্রাপ্ত নম্বর	:	51	51	38	54	54	38	62	58	38

সমাধান ঃ প্রাপ্ত নম্বর অনুযায়ী ছাত্রদের অনুক্রমিক মানগুলি সংযোজন করা হল।

রোল নং	:	1	2	3	4	5	6	7	8	9
ইংরেজী (u_i)	:	$5\frac{1}{2}$	1	8		8	8	2	3	4
গণিত (v_i)	:			8			8	1	2	8

প্রথম শ্রেণীভুক্ত (ইংরেজী) অনুক্রমিক মানগুলিতে 2 এবং 3 দৈর্ঘ্যের দুটি বন্ধন (ties) রয়েছে।

[illegible]

দ্বিতীয় শ্রেণীর (গণিত) অন্তর্গত অনুক্রমিক মানগুলিতে 2, 3 এবং 2 দৈর্ঘ্যের তিনটি বন্ধন (ties) রয়েছে।

$$\therefore\ TV = \frac{1}{12n}\left[\left(2^3-2\right)+\left(3^3-3\right)+\left(2^3-2\right)\right] = \frac{36}{12\times 9} = \frac{1}{3} = 0\cdot 3333$$

আবার,

এবং $\frac{1}{2n}\sum d_i^2 = \frac{1}{2\times 9}\left[0+\frac{81}{4}+0+4+\frac{81}{4}+0+1+1+16\right]$

$$= \frac{1}{18}\times\frac{250}{4} = \frac{125}{36} = 3\cdot 4722$$

অতএব স্পেয়ারম্যানের অনুক্রমিক সহগাঙ্ক, r_R

অনুক্রমিক মানসমূহের দুটি শ্রেণীর মধ্যে যদি সম্পূর্ণরূপে অসমাঞ্জস্য থাকে তাহলে $v_i = n - u_i + 1$ এবং u এবং v চলক দুটির মধ্যে সম্পর্কটি প্রকৃতই এমন একটি সরলরেখার দ্বারা প্রকাশিত হবে যে সরলরেখার ঢাল হবে ঋণাত্মক অর্থাৎ $r_R = -1$

প্রসঙ্গাক্রমে কেন্‌ডালস্ অনুক্রমিক সহগাঙ্গা ও (Kendall's Correlation coefficient) উল্লেখ করা যেতে পারে।

উদাহরণ 1. একটি প্রতিযোগিতায় 10 জন প্রতিযোগিকে দুটি বিচার অনুযায়ী যে ক্রমে সাজান হয় তা নিম্নে প্রদত্ত হল—দুটি বিচারের মধ্যে অনুক্রমিক সহগাঙ্ক নির্ণয় করো।

প্রতিযোগী :		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
বিচারের ক্রম :	প্রথম ঃ	6	11	9	3	7	10	5	2	8	4
	দ্বিতীয় ঃ	4	9	11	5	10	8	2	3	7	6

সমাধান ঃ প্রথম এবং দ্বিতীয় বিচারের ক্রমসমূহ যথাক্রমে u_i, v_i ধরা হল যেখানে $i = 1, 2 \ldots 10$.

বিচারের অনুক্রমিক সহগাঙ্ক নির্ণয়ের জন্য ছক

	u_i	v_i	$d_i = u_i - v_i$	d^2
	6	4	2	4
	11	9	2	4
	9	11	–2	4
	3	5	–2	4
	7	10	–3	9
	10	8	2	4
	5	2	3	9
	2	3	–1	1
	8	7	–1	1
	4	6	–2	4
মোট				44

প্রতিযোগী সংখ্যা $= n = 10$

$\therefore$ নির্ণেয় অনুক্রমিক সহগাঙ্ক

$$r_R = 1 - \frac{6\sum_{i=1}^{10} d_i^2}{n(n^2-1)} = 1 - \frac{6 \times 44}{10(10^2-1)} = \frac{726}{990} = \cdot 73$$

(আসন্ন)

যেহেতু অনুক্রমিক মানসমূহের বর্গ $\frac{1}{12}(m^3 - m)$ পরিমাণ হ্রাসপ্রাপ্ত হয় সেইহেতু অনুক্রমিক মান শ্রেণীর যৌগিক গড় অপরিবর্তিত থাকবে, কিন্তু ভেদমান $\frac{1}{12m}(m^3 - m)$ পরিমাণ হ্রাসপ্রাপ্ত হবে।

ধরা যাক, u চলকের অনুক্রমিক মানগুলির ক্ষেত্র $m_1, m_2, \ldots m_k$ দৈর্ঘ্যের K সংখ্যক বন্ধন রয়েছে এবং V চালকের অনুক্রমিক মানগুলির ক্ষেত্রে $m_1', m_2' \ldots m_L'$ দৈর্ঘ্যের L সংখ্যক বন্ধন রয়েছে। এই ক্ষেত্রে প্রতিটি বন্ধনের জন্য সংশ্লিষ্ট ভেদমান হ্রাসপ্রাপ্ত হবে এবং যেহেতু বন্ধন প্রক্রিয়া যোগ ধর্মাধীন (additive) সেইহেতু উভয় চলকের ক্ষেত্রে নতুন ভেদমান হবে ঃ

$$\text{এবং } \sigma_v^2 = \frac{n^2 - 1}{12} - Tv$$

$$\text{যেখানে, } Tu = \frac{1}{12n}\sum_{i=1}^{K}(m_i^3 - m_i) \text{ এবং } Tv = \frac{1}{12n}\sum_{i=1}^{L}(m_i'^3 - m_i')$$

$$\text{আবার যেহেতু } \frac{1}{n}\sum_{i=1}^{n} d_i^2 = \sigma_u^2 + \sigma_v^2 - 2 \text{ ভেদমান } (u, v)$$

বন্ধনযুক্ত অনুক্রমিক মানগুলির ক্ষেত্রে

$$\text{ভেদমান } (u,v) = \frac{n^2 - 1}{12} - \frac{Tu + Tv}{2} - \frac{1}{2n}\sum_{i=1}^{n} d_i^2$$

সুতরাং বন্ধন যুক্ত অনুক্রমিক মানগুলির ক্ষেত্রে স্পেয়ারম্যানের অনুক্রমিক সহগাঙ্ক হল ঃ

$$r_R = \frac{\frac{n^2 - 1}{12} - \frac{Tu + Tv}{2} - \frac{1}{2n}\sum_{i=1}^{n} d_i^2}{\sqrt{\frac{n^2 - 1}{12} - Tu}\ \sqrt{\frac{n^2 - 1}{12} - Tv}}$$

অনুক্রমিক মানসমূহের দুটি শ্রেণীর মধ্যে যদি সম্পূর্ণ সামঞ্জস্য (Perfect agreement) থাকে তাহলে প্রতি i-এর জন্য $u_i = v_i$ এবং $d_i = 0$, $T_u = T_v$ এবং $\boldsymbol{\sigma}_u = \boldsymbol{\sigma}_v$ হবে।

$\therefore\ r_R = 1$

এক্ষেত্রে u এবং v চলক দুটির মধ্যে সম্পর্কটি প্রকৃতই এমন একটি সরলরেখার সাহায্যে প্রকাশিত হবে যে রেখার ঢাল, $r_R = 1$ হবে—অর্থাৎ ধনাত্মক হবে।

অতএব, $r_{uv} = \dfrac{\text{[illegible]}\,(u, v)}{\sigma u\, \sigma v} = \dfrac{\dfrac{n^2-1}{12} - \dfrac{1}{2n}\sum\limits_i d_i^2}{\dfrac{n^2-1}{12}}$

$$= \frac{6\sum\limits_i d_i^2}{n(n^2-1)} = r_R$$

(b) সম অনুক্রমিক মানসম্পন্ন যখন কিছু ব্যষ্টি—এরূপ ক্ষেত্রে (When there are ties)

এখন ধরা যাক m সংখ্যক ব্যষ্টি একই অনুক্রমিক মান সম্পন্ন। তাহলে বলা যায় যে m দৈর্ঘ্যের একটি বন্ধন (a tie) রয়েছে। এরূপ একই অনুক্রমিক স্থাপনে একাধিক বন্ধন থাকতে পারে। বিশেষ কোন এক ক্ষেত্রে যদি m বন্ধন বিশিষ্ট ব্যষ্টিগুলির অব্যবহিত পূর্বেস্থিত ব্যষ্টির অনুক্রমিক মান r হয় তাহলে বন্ধনভুক্ত m সংখ্যক ব্যষ্টির প্রতিটির অনুক্রমিক মান হবে

$$\frac{(r+1)+(r+2)+\ldots+(r+m)}{m} = r + \frac{m+1}{2}$$

এবার এরূপ বন্ধন না থাকলে অনুক্রমিক মানগুলির বর্গের সমষ্টি, S_1 হত

$$= mr^2 + 2r(1+2+\ldots+m) + (1^2 + 2^2 + \ldots m^2)$$

$$= mr^2 + m(m+1)r + \frac{1}{6}m(m+1)(2m+1)$$

এবার বন্ধন বিশিষ্ট অনুক্রমিক মানগুলির বর্গের সমষ্টি, S_2 হত

$$S_2 = m\left(r + \frac{m+1}{2}\right)^2$$

এখন, $S_1 - S_2 = m(m+1)\left(\dfrac{2m+1}{6} - \dfrac{m+1}{4}\right) = \dfrac{m(m+1)(m-1)}{12} = \dfrac{1}{12}(m^3 - m)$

যেখানে হল যথাক্রমে u_i ও v_i-এর মানগুলির যৌগিক গড় u-এর বিভিন্ন মানগুলির ভেদমান, এবং v-এর বিভিন্ন মানগুলির ভেদমান হল নিম্নরূপ ঃ

$$= \frac{1}{n}\sum u^2 - \overline{u}^2$$

$$= \frac{1}{n}\frac{n(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2$$

$$= \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{n^2-1}{12}$$

অনুরূপভাবে,

$$\sigma_v^2 = \frac{n^2-1}{12}$$

আবার,

$$= \frac{1}{n}\sum_i \left\{(u_i - \overline{u}) + (v_i - \overline{v})\right\}^2,$$ (যেহেতু)

$$= \sigma_u^2 + \sigma_v^2 - 2 \text{ সহভেদমান } (u,\ v)$$

$\therefore$ সহভেদমান $(u,\ v)$

$$r_R = 1 - \frac{6\sum_i d_i^2}{n(n^2 - 1)}$$

অনুক্রমিক মানগুলির দুটি শ্রেণীর মধ্যে যখন পূর্ণ সামঞ্জস্য (Perfect agreement) থাকে—অর্থাৎ প্রত্যেক ব্যষ্টি সংশ্লিষ্ট দুটি ক্রমিক মান যখন সমান হয় তখন চলক দুটির মধ্যে যোগসূত্রটি (association) ধনাত্মকভাবে সম্পূর্ণ (Positively perfect) বলে গণ্য করা হয়। সেই ক্ষেত্রে প্রতিটি i-এর জন্য $u_i = v_i$ এবং $\sum_i d_i^2 = 0$ সুতরাং $r_R = 1$ হয়।

আবার অনুক্রমিক মানগুলির দুটি শ্রেণীর মধ্যে যখন পূর্ণ সামঞ্জস্য থাকে না (অর্থাৎ) তখন চলক দুটির মধ্যে যোগসূত্রটি ঋণাত্মকভাবে সম্পূর্ণ (negatively perfect) বলে গণ্য করা হয় এবং সেই ক্ষেত্রে প্রতিটি i-এর জন্য $v_i = n - u_i + 1$

এবং

$$= 4\sum_i u_i^2 - 4(n+1)\sum_i u_i + n(n+1)^2$$

$$= 4.\frac{n(n+1)(2n+1)}{6} - 4(n+1)\frac{n(n+1)}{2} + (n+1)^2$$

$$= \frac{2}{3}n(n+1)(2n+1) - n(n+1)^2$$

$$= \frac{1}{3}n(n^2+1)$$

$$\therefore r_R = 1 - \frac{6n(n^2-1)}{3} \cdot \frac{1}{n(n^2-1)} = 1 - 2 = -1$$

u_i এবং v_i-কে চলক দুটির বিভিন্ন মান ধরে সরল গুণন পরিঘাত সহগতি সহগাঙ্ক (Simple product moment correlation coefficient) থেকে স্পেয়ারম্যানের অনুক্রমিক সহগাঙ্ক পাওয়া যায়।

যেহেতু অনুক্রমিক মানগুলি 1, 2, ... n ব্যতীত অন্য কোন মান গ্রহণ করে না

সেইহেতু, $\sum_i u_i = \sum v_i = 1 + 2 + \ldots + n = \frac{n(n+1)}{2}$

$$= \frac{50 \times 30 - (-13) \times (-37)}{\sqrt{50 \times 71 - (-37)^2}\sqrt{50 \times 55 - (-13)^2}} = \frac{1019}{\sqrt{2181}\sqrt{2581}}$$

$$= \frac{1019}{46 \cdot 70 \times 50 \cdot 80} = 0 \cdot 43$$

(ছ) অনুক্রমিক সহগাঙ্গ (Rank Correlation Co-efficient) : যখন ব্যষ্টিসমূহের একটি গোষ্ঠী (a group of individuals) একটি বিশেষ চরিত্র বা গুণের মাত্রা অনুসারে সাজান হয় তখন তাদের অনুক্রমিক বিন্যাস (ranking) হয়েছে বলে ধরা হয়। এরূপ অনুক্রমিক বিন্যাসের অন্তর্গত কোন ব্যষ্টি যে নির্দিষ্ট স্থানে অবস্থিত সেই স্থান সংখ্যা হল মাত্রাগত সংখ্যা (ordinal number) এবং সেই স্থানসংখ্যা ঐ ব্যষ্টির অনুক্রমিক মান (rank) হিসাবে গণ্য হয়।

একই গোষ্ঠীভুক্ত ব্যষ্টিসমূহের জন্য যখন দুটি ভিন্ন চরিত্র বা গুণ সংশ্লিষ্ট দুটি অনুক্রমিক মান শ্রেণী পাওয়া যায় অথবা যখন দুজন বিচারক একই চরিত্র বা গুণের জন্য দুটি ক্রমিক মান শ্রেণী প্রদান করে তখন এই অনুক্রমিক মান শ্রেণী দুটির মধ্যে কোনরূপ যোগসূত্র আছে কিনা তা বিচার করার আগ্রহ দেখা দেয়। একই গোষ্ঠীভুক্ত ব্যষ্টিসমূহের জন্য এরূপ দুটি অনুক্রমিক মান শ্রেণীর মধ্যে যোগসূত্র বা সম্পর্ককে **অনুক্রমিক সহগতি** (Rank correlation) বলা হয় এবং এরূপ সহগতি সংশ্লিষ্ট সহগাঙ্ক হল অনুক্রমিক সহগাঙ্ক।

(জ) স্পেয়ারম্যানের অনুক্রমিক সহগাঙ্ক (Spearman's rank Correlation Co-efficient) :

(a) প্রত্যেক ব্যষ্টি পৃথক অনুক্রমিক মান বিশিষ্ট এরূপ ক্ষেত্রে (Case when there are no ties) ধরা যাক n সংখ্যক ব্যষ্টির অনুক্রমিক মানগুলির দুটি শ্রেণী হল $u_1, u_2, \ldots u_n$ এবং $v_1, v_2, \ldots v_n$.

যেহেতু অনুক্রমিক মানসমূহ (ranks) 1, 2, ... n ব্যতীত অন্য কোন মান গ্রহণ করে না এবং u এবং v চলক দুটি পৃথক অনুক্রমিক মান সম্পন্ন, সেইহেতু u_i এবং v_i হল প্রথম n সংখ্যক স্বাভাবিক সংখ্যার বিন্যাস (Permutation of n natural members)। আরও ধরা যাক, $d_i = u_i - v_i$ যেখানে $i = 1, 2, \ldots n$ তাহলে r_R দ্বারা চিহ্নিত স্পেয়ারম্যানের অনুক্রমিক সহগাঙ্ক নিম্নলিখিতভাবে প্রকাশ করা হয় :