



# **BIOSTATISTIKA UNTUK PERGURUAN TINGGI**

**PENULIS :**

**Neny Rochyani, Iha Nursolihah, Nur Al-Faida, Bambang Suhartawan,  
Khairul Alim, Samsul Bahri Loklomin, Indah Resti Ayuni Suri,  
Iya Purnama Sari**

ISBN 978-623-125-294-4



9 786231 252944

# **BIOSTATISTIKA UNTUK PERGURUAN TINGGI**

**Neny Rochyani  
Iha Nursolihah  
Nur Al-Faida  
Bambang Suhartawan  
Khairul Alim  
Samsul Bahri Loklomin  
Indah Resti Ayuni Suri  
Iya Purnama Sari**



**GET PRESS INDONESIA**

# **BIOSTATISTIKA UNTUK PERGURUAN TINGGI**

## **Penulis :**

Neny Rochyani  
Iha Nursolihah  
Nur Al-Faida  
Bambang Suhartawan  
Khairul Alim  
Samsul Bahri Loklomin  
Indah Resti Ayuni Suri  
Iya Purnama Sari

**ISBN : 978-623-125-294-4**

**Editor :** Yuliatr Novita., M.Hum

**Penyunting :** Tri Putri Wahyuni., S.Pd

**Desain Sampul dan Tata Letak :** Atyka Trianisa, S.Pd

**Penerbit :** GET PRESS INDONESIA

Anggota IKAPI No. 033/SBA/2022

## **Redaksi :**

Jln. Palarik Air Pacah No 26 Kel. Air Pacah  
Kec. Koto Tangah Kota Padang Sumatera Barat

Website : [www.getpress.co.id](http://www.getpress.co.id)

Email : [adm.getpress@gmail.com](mailto:adm.getpress@gmail.com)

Cetakan pertama, Juni 2024

Hak cipta dilindungi undang-undang  
Dilarang memperbanyak karya tulis ini dalam bentuk dan  
dengan cara apapun tanpa izin tertulis dari penerbit.

## **KATA PENGANTAR**

Segala Puji dan syukur atas kehadiran Allah SWT dalam **segala** kesempatan. Sholawat beriring salam dan doa kita sampaikan kepada Nabi Muhammad SAW. Alhamdulillah atas Rahmat dan Karunia-Nya penulis telah menyelesaikan Buku Biostatistika Untuk Perguruan Tinggi ini.

Buku Ini Membahas Pengambilan Sampel, Pengolahan Data, Penataan Data, Analisis Data, Distribusi Probabilitas, Distribusi Sampel, Pengujian Hipotesis.,

Proses penulisan buku ini berhasil diselesaikan atas kerjasama tim penulis. Demi kualitas yang lebih baik dan kepuasan para pembaca, saran dan masukan yang membangun dari pembaca sangat kami harapkan.

Penulis ucapkan terima kasih kepada semua pihak yang telah mendukung dalam penyelesaian buku ini. Terutama pihak yang telah membantu terbitnya buku ini dan telah mempercayakan mendorong, dan menginisiasi terbitnya buku ini. Semoga buku ini dapat bermanfaat bagi masyarakat Indonesia.

Padang, Juni 2024

Penulis

## DAFTAR ISI

|                                                                |           |
|----------------------------------------------------------------|-----------|
| <b>KATA PENGANTAR .....</b>                                    | <b>i</b>  |
| <b>DAFTAR ISI .....</b>                                        | <b>ii</b> |
| <b>DAFTAR GAMBAR .....</b>                                     | <b>iv</b> |
| <b>DAFTAR TABEL .....</b>                                      | <b>v</b>  |
| <b>BAB 1 PENGAMBILAN SAMPEL .....</b>                          | <b>1</b>  |
| 1.1 Pendahuluan .....                                          | 1         |
| 1.2 Konsep Dasar Pengambilan Sampel.....                       | 2         |
| 1.3 Strategi Pengambilan Sampel.....                           | 3         |
| 1.4 Faktor-Faktor yang Mempengaruhi Pengambilan<br>Sampel..... | 6         |
| 1.5 Etika Pengambilan Sampel .....                             | 8         |
| 1.6 Kesimpulan.....                                            | 9         |
| DAFTAR PUSTAKA.....                                            | 10        |
| <b>BAB 2 PENGOLAHAN DATA.....</b>                              | <b>11</b> |
| 2.1 Pendahuluan .....                                          | 11        |
| 2.2 Aplikasi Statistik dalam Pengolahan Data .....             | 12        |
| 2.3 Tahapan Pengolahan Data .....                              | 14        |
| DAFTAR PUSTAKA.....                                            | 24        |
| <b>BAB 3 PENATAAN DATA .....</b>                               | <b>25</b> |
| 3.1 Pendahuluan .....                                          | 25        |
| 3.2 Pengumpulan Data .....                                     | 25        |
| 3.2.1 Jenis-jenis pengumpulan Data.....                        | 26        |
| 3.3 Pembersihan Data.....                                      | 27        |
| 3.4 Transformasi Data .....                                    | 29        |
| 3.5 Penyusunan Data.....                                       | 30        |
| 3.6 Dokumentasi Data.....                                      | 32        |
| 3.7 Penyimpanan Data.....                                      | 33        |
| 3.8 Pemeliharaan Data.....                                     | 34        |
| DAFTAR PUSTAKA.....                                            | 37        |
| <b>BAB 4 ANALISIS DATA .....</b>                               | <b>39</b> |
| 4.1 Pendahuluan .....                                          | 39        |
| 4.2 Tahapan Analisis Data dalam Bidang Biostatistika....       | 39        |
| DAFTAR PUSTAKA.....                                            | 53        |
| <b>BAB 5 TEORI PROBABILITAS .....</b>                          | <b>55</b> |
| 5.1 Pendahuluan .....                                          | 55        |

|                                            |           |
|--------------------------------------------|-----------|
| 5.2 Ruang Sampel dan Kejadian.....         | 56        |
| 5.2.1 Ruang Sampel.....                    | 56        |
| 5.2.2 Kejadian .....                       | 57        |
| 5.3 Peluang.....                           | 58        |
| 5.3.1 Syarat Peluang.....                  | 58        |
| 5.3.2 Peluang Gabungan dan Irisan.....     | 59        |
| 5.3.3 Peluang Bersyarat.....               | 60        |
| 5.3.4 Kejadian Bebas atau Independen ..... | 62        |
| 5.3.5 Rumus Bayes.....                     | 63        |
| DAFTAR PUSTAKA.....                        | 65        |
| <b>BAB 6 DISTRIBUSI PROBABILITAS.....</b>  | <b>67</b> |
| 6.1 Pendahuluan .....                      | 67        |
| 6.2 Distribusi Bernoulli.....              | 68        |
| 6.3 Distribusi Binomial.....               | 69        |
| 6.4 Distribusi Poisson .....               | 71        |
| 6.5 Distribusi Geometri.....               | 73        |
| 6.6 Distribusi Normal.....                 | 75        |
| DAFTAR PUSTAKA.....                        | 77        |
| <b>BAB 7 DISTRIBUSI SAMPEL .....</b>       | <b>79</b> |
| 7.1 Pendahuluan .....                      | 79        |
| 7.2 Pengertian Distribusi Sampling.....    | 79        |
| 7.3 Sifat-Sifat Ditribusi Sampling .....   | 80        |
| 7.4 Kesimpulan.....                        | 89        |
| DAFTAR PUSTAKA.....                        | 90        |
| <b>BAB 8 PENGUJIAN HIPOTESIS.....</b>      | <b>91</b> |
| 8.1 Pengertian Pengujian Hipotesis.....    | 91        |
| 8.2 Langkah Pengujian Hipotesis .....      | 93        |
| 8.3 Rumus dalam Pengujian Hipotesis.....   | 95        |
| 8.4 Contoh Soal Pengujian Hipotesis.....   | 98        |
| DAFTAR PUSTAKA.....                        | 103       |
| <b>BIODATA PENULIS</b>                     |           |

## DAFTAR GAMBAR

|                                                                                                              |    |
|--------------------------------------------------------------------------------------------------------------|----|
| <b>Gambar 2.1.</b> Halaman Lembar Kerja Excel .....                                                          | 13 |
| <b>Gambar 2.2.</b> Halaman Lembar Kerja SPSS ( <i>Data View</i> ). 14                                        |    |
| <b>Gambar 2.3.</b> Halaman Lembar Kerja SPSS ( <i>Variable View</i> ).....                                   | 14 |
| <b>Gambar 2.4.</b> Contoh Coding dalam Lembar Kerja Excel                                                    | 17 |
| <b>Gambar 2.5.</b> Contoh Input Data Pada Program SPSS<br>( <i>Data View</i> ) .....                         | 20 |
| <b>Gambar 2.6.</b> Contoh Input Data Pada Program SPSS<br>( <i>Variable View</i> ).....                      | 20 |
| <b>Gambar 2.7.</b> Contoh <i>Missing Data</i> Menggunakan Tabel<br>Distribusi Frekuensi SPSS .....           | 21 |
| <b>Gambar 2.8.</b> Contoh Kesalahan Variasi Data<br>Menggunakan Tabel Distribusi Frekuensi<br>SPSS .....     | 22 |
| <b>Gambar 2.9.</b> Contoh Kesalahan Konsistensi Data<br>Menggunakan Tabel Distribusi Frekuensi<br>SPSS ..... | 22 |
| <b>Gambar 8.1.</b> Langkah dalam pemilihan uji hipotesis<br>yang akan digunakan.....                         | 94 |
| <b>Gambar 8.2.</b> Batasan dalam penerimaan atau<br>penolakan hipotesis .....                                | 94 |
| <b>Gambar 8.3.</b> Daerah penerimaan dan penolakan $H_0$ .....                                               | 97 |

## DAFTAR TABEL

|                                                                                                           |     |
|-----------------------------------------------------------------------------------------------------------|-----|
| <b>Tabel 8.1.</b> Cara membuat keputusan dari contoh<br>situasi yang ada .....                            | 92  |
| <b>Tabel 8.2.</b> Hasil penilaian atau skoring terhadap 8<br>karyawan sebelum dan setelah pelatihan ..... | 101 |



# BAB 1

## PENGAMBILAN SAMPEL

*Oleh Neny Rochyani*

### 1.1 Pendahuluan

Pengambilan sampel merupakan tahapan awal yang kritis dalam setiap penelitian biostatistik. Langkah ini menjadi pondasi bagi keseluruhan penelitian, memastikan bahwa data yang dianalisis dapat mewakili populasi yang lebih besar. Dalam konteks biostatistik, di mana penelitian seringkali melibatkan fenomena biologis atau kesehatan manusia, keberhasilan penelitian sangat bergantung pada kecermatan dan kevalidan sampel yang diambil. Dalam bagian ini, kita akan menjelajahi konsep dasar pengambilan sampel dan pentingnya memahami populasi target. Populasi dalam penelitian biostatistik mengacu pada kelompok lengkap dari unit atau individu yang memiliki karakteristik tertentu yang menjadi fokus penelitian. Misalnya, populasi dapat berupa semua penderita penyakit tertentu, kelompok usia tertentu, atau bahkan seluruh spesies tertentu. Seringkali, keterbatasan sumber daya memaksa peneliti untuk tidak mengumpulkan data dari seluruh populasi. Oleh karena itu, dipilihlah sampel, yaitu sebagian kecil dari populasi yang dianggap dapat memberikan informasi yang mewakili keseluruhan. Kunci dari pengambilan sampel yang baik adalah representativitas. Sampel harus mencerminkan diversitas dan karakteristik utama dari populasi secara keseluruhan. Jika sampel tidak representatif, analisis statistik dan generalisasi hasil penelitian menjadi sulit dilakukan.

Pada tahap pengambilan sampel, peneliti dapat menghadapi dua jenis kesalahan: kesalahan acak (*random error*) dan kesalahan non-acak (*systematic error*). Kesalahan acak adalah variasi alamiah yang muncul secara kebetulan, sedangkan kesalahan non-acak disebabkan oleh kekurangan dalam desain penelitian atau metode pengambilan sampel.

Kesalahan dalam pengambilan sampel dapat menyebabkan hasil penelitian tidak akurat atau bahkan menyesatkan. Oleh karena itu, pemahaman terhadap tingkat kesalahan yang mungkin terjadi menjadi krusial dalam menafsirkan hasil penelitian.

## 1.2 Konsep Dasar Pengambilan Sampel

Pengambilan sampel adalah suatu proses penting dalam penelitian biostatistik yang memerlukan pemahaman mendalam tentang konsep-konsep dasar agar hasil penelitian dapat dipercaya dan bermakna. Dalam konteks ini, ada beberapa konsep dasar yang perlu diperhatikan:

1. Populasi dan Sampel:

Populasi: Merupakan keseluruhan unit atau individu yang memiliki karakteristik tertentu dan menjadi fokus penelitian. Contoh, populasi penderita diabetes tipe 2.

Sampel: Sebagian kecil dari populasi yang diambil untuk mewakili keseluruhan. Penting bahwa sampel adalah representatif dan mencerminkan variasi dalam populasi.

2. Representativitas Sampel:

Sampel harus mencakup berbagai karakteristik yang ada dalam populasi sehingga hasil penelitian dapat diterapkan secara lebih luas. Representativitas memastikan bahwa temuan penelitian memiliki validitas eksternal.

3. Pengambilan Sampel Acak dan Non-Acak:

Pengambilan Sampel Acak: Setiap anggota populasi memiliki kesempatan yang sama untuk dipilih. Ini termasuk metode seperti pengambilan sampel acak sederhana.

Pengambilan Sampel Non-Acak: Metode pengambilan yang tidak mengikuti prinsip acak. Contohnya termasuk pengambilan sampel convenience atau purposive.

4. Kesalahan Pengambilan Sampel:

Kesalahan Acak (*Random Error*): Kesalahan yang terjadi secara kebetulan dan dapat dikelola dengan meningkatkan ukuran sampel.

Kesalahan Non-Acak (*Systematic Error*): Kesalahan yang bersifat konsisten dan dapat merugikan hasil penelitian.

- Dapat muncul dari desain penelitian atau proses pengambilan sampel.
5. **Ukuran Sampel:**  
Menentukan jumlah sampel yang tepat penting untuk mendapatkan hasil yang akurat. Ukuran sampel yang kecil dapat menghasilkan estimasi yang tidak stabil, sementara ukuran sampel yang terlalu besar bisa menghamburkan sumber daya.
  6. **Margin of Error:**  
Sejauh mana hasil dari sampel dapat bervariasi dari hasil populasi sebenarnya. Tingkat kepercayaan yang tinggi akan meningkatkan margin of error.
  7. **Deviasi Standar dan Keragaman Populasi:**  
Deviasi Standar: Menunjukkan sejauh mana data tersebar dari rata-rata. Pengambilan sampel dari populasi yang lebih bervariasi memerlukan ukuran sampel yang lebih besar.  
Keragaman Populasi: Semakin besar variasi dalam populasi, semakin besar ukuran sampel yang dibutuhkan.  
Pemahaman konsep dasar ini menjadi dasar yang kuat untuk merancang dan melaksanakan pengambilan sampel yang tepat dalam penelitian biostatistik. Dengan memperhatikan prinsip-prinsip ini, peneliti dapat meminimalkan kesalahan dan meningkatkan kepercayaan pada hasil penelitian mereka.

### **1.3 Strategi Pengambilan Sampel**

Pengambilan sampel adalah suatu proses yang kompleks dan strategis, yang melibatkan keputusan-keputusan penting untuk memastikan representativitas dan kevalidan hasil penelitian. Berikut adalah penjelasan detail mengenai beberapa strategi pengambilan sampel yang umum digunakan dalam penelitian biostatistik:

#### **1. Pengambilan Sampel Acak Sederhana:**

##### **a. Konsep:**

Pengambilan sampel acak sederhana melibatkan pemilihan unit atau individu secara acak dari populasi target.

Setiap anggota populasi memiliki probabilitas yang sama untuk dipilih.

b. Langkah-Langkah:

Penentuan ukuran sampel yang diinginkan.

Memberikan label atau nomor unik pada setiap elemen populasi.

Menggunakan metode randomization untuk memilih sampel.

c. Kelebihan dan Kelemahan:

Kelebihan: Sederhana, mudah dilaksanakan, dan meminimalkan bias.

Kelemahan: Tidak mempertimbangkan struktur internal populasi, mungkin tidak efisien untuk populasi besar.

## **2. Pengambilan Sampel Stratifikasi:**

a. Konsep:

Populasi dibagi menjadi subkelompok (strata) yang homogen berdasarkan karakteristik tertentu.

Sampel kemudian diambil dari setiap stratum secara independen.

b. Langkah-Langkah:

Identifikasi faktor-faktor kritis yang membedakan populasi.

Pembagian populasi menjadi strata.

Pengambilan sampel acak sederhana dari setiap stratum.

c. Kelebihan dan Kelemahan:

Kelebihan: Memastikan representasi dari setiap subkelompok, dapat meningkatkan akurasi hasil.

Kelemahan: Memerlukan informasi lebih lanjut untuk pembagian strata, membutuhkan perencanaan lebih cermat.

## **3. Pengambilan Sampel Kluster:**

a. Konsep:

Populasi dibagi menjadi kelompok (kluster) yang heterogen, mirip dengan strata.

Sampel diambil dengan memilih beberapa kluster secara acak dan mengambil semua elemen dalam kluster tersebut.

b. Langkah-Langkah:

Identifikasi kluster dalam populasi.

Pengambilan sampel acak beberapa kluster.

Pengambilan semua elemen dalam kluster yang dipilih.

c. Kelebihan dan Kelemahan:

Kelebihan: Efisien untuk populasi yang tersebar luas, biaya pengambilan sampel lebih rendah.

Kelemahan: Potensial untuk menghasilkan hasil yang kurang representatif jika kluster tidak homogen.

**4. Pengambilan Sampel Berlapis (Multistage Sampling):**

a. Konsep:

Kombinasi dari beberapa strategi pengambilan sampel.

Pengambilan sampel dilakukan dalam beberapa tahap.

b. Langkah-Langkah:

Identifikasi strata dan kluster.

Pengambilan sampel acak dari setiap kluster.

Pemilihan unit sampel dalam kluster.

c. Kelebihan dan Kelemahan:

Kelebihan: Fleksibel, dapat digunakan untuk populasi yang kompleks.

Kelemahan: Memerlukan perencanaan yang lebih kompleks, potensial untuk meningkatkan kesalahan.

Pemahaman mendalam tentang strategi pengambilan sampel membantu peneliti memilih metode yang paling sesuai dengan tujuan penelitian dan karakteristik populasi. Kombinasi strategi ini juga dapat digunakan untuk meningkatkan validitas dan representativitas hasil penelitian biostatistik.

## 1.4 Faktor-Faktor yang Mempengaruhi Pengambilan Sampel

Pemilihan sampel yang tepat dalam penelitian biostatistik sangat dipengaruhi oleh berbagai faktor yang harus diperhitungkan. Berikut adalah penjelasan detail mengenai beberapa faktor utama yang mempengaruhi pengambilan sampel:

### 1. Ukuran Sampel:

Ukuran sampel merujuk pada jumlah elemen atau individu yang akan diambil dari populasi. Ukuran sampel yang tidak memadai dapat mengakibatkan hasil yang tidak representatif.

#### **Faktor-faktor yang Mempengaruhi:**

Kepercayaan Tingkat Tinggi: Semakin tinggi tingkat kepercayaan yang diinginkan, semakin besar ukuran sampel yang diperlukan.

Ketepatan (*Precision*): Semakin kecil *margin of error* yang diinginkan, semakin besar ukuran sampel yang dibutuhkan.

### 2. Tingkat Kepercayaan dan Tingkat Signifikansi:

Tingkat kepercayaan menunjukkan sejauh mana peneliti yakin bahwa hasil penelitian mencerminkan populasi.

Tingkat signifikansi menentukan batas kesalahan yang dapat diterima saat menarik kesimpulan.

#### **Faktor-faktor yang Mempengaruhi:**

Kepercayaan Tingkat Tinggi: Memerlukan ukuran sampel yang lebih besar.

Tingkat Signifikansi Rendah: Memerlukan ukuran sampel yang lebih besar untuk meningkatkan keakuratan hasil.

### 3. Keragaman Populasi:

Keragaman atau variasi dalam populasi merujuk pada sejauh mana elemen atau individu dalam populasi berbeda satu sama lain.

#### **Faktor-faktor yang Mempengaruhi:**

Tingkat Keragaman yang Tinggi: Memerlukan ukuran sampel yang lebih besar untuk mencakup variasi yang lebih besar.

Analisis Subgrup: Jika ada subgrup dalam populasi dengan keragaman yang tinggi, perlu dipertimbangkan untuk mengambil sampel dari setiap subgrup.

**4. Desain Penelitian:**

Desain penelitian mencakup cara penelitian dirancang, termasuk metode pengambilan sampel.

**Faktor-faktor yang Mempengaruhi:**

Metode Pengambilan Sampel: Desain penelitian harus mempertimbangkan apakah akan menggunakan pengambilan sampel acak sederhana, stratifikasi, atau kluster.

Karakteristik Penelitian: Apakah penelitian bersifat eksperimental atau observasional akan memengaruhi strategi pengambilan sampel.

**5. Tujuan Penelitian:**

Tujuan penelitian menentukan informasi apa yang akan dikumpulkan dari pengambilan sampel.

**Faktor-faktor yang Mempengaruhi:**

Generalisasi Hasil: Jika tujuan penelitian adalah untuk membuat generalisasi yang lebih luas, maka pengambilan sampel harus mewakili populasi secara keseluruhan.

Spesifik Tujuan Penelitian: Jika tujuan penelitian bersifat spesifik, pengambilan sampel dapat difokuskan pada karakteristik tertentu dalam populasi.

**6. Sumber Daya Tersedia:**

Sumber daya mencakup waktu, dana, dan tenaga manusia yang tersedia untuk penelitian.

**Faktor-faktor yang Mempengaruhi:**

Keterbatasan Dana: Jika sumber daya terbatas, peneliti mungkin harus memilih ukuran sampel yang memadai tetapi tetap efisien.

Waktu: Pengambilan sampel yang memakan waktu lama dapat memerlukan strategi yang berbeda dari pengambilan sampel yang cepat.

**7. Karakteristik Populasi:**

Karakteristik populasi mencakup aspek-aspek seperti homogenitas atau heterogenitas populasi.

### **Faktor-faktor yang Mempengaruhi:**

Homogenitas Populasi: Jika populasi homogen, ukuran sampel yang lebih kecil mungkin cukup representatif.

Heterogenitas Populasi: Jika populasi heterogen, perlu mempertimbangkan strategi pengambilan sampel yang lebih cermat.

## **1.5 Etika Pengambilan Sampel**

Pengambilan sampel dalam penelitian biostatistik tidak hanya mencakup aspek teknis, tetapi juga melibatkan pertimbangan etika yang mendalam. Kehati-hatian dan keadilan dalam setiap tahapan pengambilan sampel merupakan landasan integritas penelitian. Berikut adalah penjelasan detail mengenai etika pengambilan sampel dalam konteks penelitian biostatistik:

### **1. Hak dan Kewajiban Partisipan:**

#### **a. Hak Partisipan:**

Partisipan penelitian memiliki hak untuk diinformasikan secara jelas tentang tujuan penelitian, prosedur pengambilan sampel, dan potensi risiko atau manfaat yang mungkin terjadi.

#### **b. Kewajiban Peneliti:**

Peneliti memiliki tanggung jawab untuk memastikan bahwa partisipan memberikan persetujuan informir sebelum terlibat dalam penelitian. Pemeliharaan privasi dan kerahasiaan partisipan harus diutamakan.

### **2. Transparansi dan Reproductibilitas:**

#### **a. Transparansi:**

Peneliti harus transparan dalam menjelaskan metode pengambilan sampel yang digunakan, termasuk kriteria pemilihan sampel dan prosedur yang diikuti.

#### **b. Reproductibilitas:**

Penelitian yang dapat direproduksi adalah prinsip dasar etika dalam penelitian biostatistik. Detail metodologi dan pengambilan sampel harus cukup jelas sehingga peneliti lain dapat mengulang penelitian dengan akurasi.

### **3. Perlindungan Hak dan Keamanan Partisipan:**

- a. Perlindungan Risiko:  
Peneliti memiliki tanggung jawab untuk mengidentifikasi dan mengurangi potensi risiko bagi partisipan selama pengambilan sampel. Risiko harus sebanding dengan manfaat yang diharapkan.
- b. Keamanan Data:  
Informasi pribadi partisipan harus dijaga keamanannya. Data harus disimpan dengan metode yang aman dan hanya diakses oleh pihak yang berwenang.
4. Konsultasi dengan Pihak Terkait:
  - a. Komunikasi dengan Pihak Terkait:  
Jika penelitian melibatkan populasi tertentu, konsultasi dengan masyarakat setempat atau kelompok terkait sangat penting. Keterlibatan komunitas dapat membantu merancang pengambilan sampel yang lebih etis dan kontekstual.
5. Pertimbangan Khusus untuk Populasi Rentan:
  - a. Populasi Rentan:  
Populasi seperti anak-anak, lansia, atau kelompok dengan kapasitas kognitif terbatas dianggap rentan. Peneliti harus memastikan perlindungan ekstra dan persetujuan yang benar ketika melibatkan populasi ini.
  - b. Keadilan dalam Pengambilan Sampel:  
Pemastian keadilan dan pencegahan eksploitasi harus menjadi perhatian utama. Penelitian harus memberikan manfaat kepada populasi tersebut tanpa menyebabkan kerugian yang tidak perlu.

## **1.6 Kesimpulan**

Bab ini memberikan pemahaman mendalam tentang pengambilan sampel dalam penelitian biostatistik. Dengan memahami konsep dasar, strategi, faktor-faktor yang mempengaruhi, dan aspek-etika dalam pengambilan sampel, peneliti dapat meningkatkan validitas dan reliabilitas hasil penelitian mereka. Dalam bab berikutnya, kita akan membahas analisis statistik yang dapat diterapkan pada data sampel yang telah dikumpulkan.

## **DAFTAR PUSTAKA**

- Cochran, W. G. (1977). *Sampling Techniques*. Wiley.
- Kothari, C. R. (2004). *Research Methodology: Methods and Techniques*. New Age International.
- Krejcie, R. V., & Morgan, D. W. (1970). Determining Sample Size for Research Activities. *Educational and Psychological Measurement*, 30(3), 607-610.
- Lohr, S. (2019). *Sampling: Design and Analysis*. Cengage Learning.
- Rosner, B. (2015). *Fundamentals of Biostatistics*. Cengage Learning.
- Salkind, N. J. (2010). *Encyclopedia of Research Design*. SAGE Publications.
- Sarndal, C. E., Swensson, B., & Wretman, J. (2003). *Model Assisted Survey Sampling*. Springer.

# BAB 2

## PENGOLAHAN DATA

*Oleh Iha Nursolihah*

### 2.1 Pendahuluan

Pengolahan data merupakan salah satu kegiatan statistika yang harus dilakukan. Seperti pengertiannya bahwa statistika adalah bidang ilmu yang mempelajari tentang pengumpulan data, pengolahan data, analisis data, penyajian data, hingga pengambilan kesimpulan untuk mengungkap pola atau masalah yang diteliti. Statistika sangat bermanfaat untuk membantu dalam pembuatan keputusan di berbagai aspek kehidupan. Untuk itu, mempelajari statistika menjadi penting karena statistika memberikan pengetahuan dan kemampuan untuk melakukan evaluasi terhadap data. Selain itu, dengan pengetahuan statistika, seseorang dapat menerima, meragukan, bahkan menolak suatu data untuk kebenarannya dan keberlakuannya. Maka menjadi penting untuk melakukan kegiatan-kegiatan statistika dengan baik, salah satunya dalam pengolahan data.

Pengolahan data adalah rangkaian kegiatan statistika setelah pengumpulan data, data mentah (*raw data*) diolah sehingga menjadi informasi yang menjawab tujuan suatu riset. Proses pengolahan data meliputi *editing*, *coding*, *processing*, dan *cleaning*. Sebelum mempelajari lanjut mengenai tiap prosesnya, pemahaman mengenai jenis data penting dipelajari agar proses pengolahan data sesuai menurut jenis datanya.

Berdasarkan sumbernya, data dibagi menjadi data primer dan data sekunder. Data Primer adalah data yang dikumpulkan atau didapat secara langsung dari responden/informan. Sedangkan data sekunder adalah data yang diperoleh dari pihak lain atau sumber yang menjadi rujukan, seperti data dari Badan Pusat Statistik (BPS), data survei yang telah dilakukan

diantaranya Riset Kesehatan Dasar (Riskedas), Survei Sosial Ekonomi Nasional (Susenas), dan lain-lain.

Berdasarkan sifatnya, data dibedakan menjadi data kategorik (kualitatif) dan data numerik (kuantitatif).

1. Data kategorik adalah data yang berbentuk kata-kata dan merupakan pengkategorian, sehingga bentuknya bukan dalam angka. Skala pengukuran data kategorik terdiri dari nominal dan ordinal. Nominal merupakan skala pengukuran data yang disusun berdasarkan pembeda atau kategori, sehingga antar data sederajat. Contohnya: data agama (Islam, Katolik, Protestan, Budha, Hindu, dan Khonghucu), data jenis kelamin (Laki-laki, Perempuan) dan lain-lain. Sedangkan ordinal merupakan skala pengukuran data kategorik dengan ciri antar data berjenjang atau terdapat tingkatan. Contohnya: data pendidikan (SD, SMP, SMA, dan Perguruan Tinggi), data penghasilan/ekonomi (Rendah, Menengah, dan Tinggi) dan lain-lain.
2. Data numerik adalah data dari hasil pengukuran atau perhitungan, sehingga bentuknya adalah angka-angka. Skala pengukuran data numerik terdiri dari interval dan rasio. Interval merupakan skala pengukuran dengan ciri antar datanya terdapat tingkatan, selain itu terdapat jarak antara satu data dengan data lainnya, namun kelipatan antar data tidak diketahui. Contohnya: data tekanan darah sistol (90 mmHg, 100 mmHg, 110 mmHg, 120 mmHg, dst), data IQ (50, 100, 150, dst). Sedangkan rasio merupakan skala yang antar datanya memiliki tingkatan, jarak, dan kelipatan. Contohnya: data berat badan (40 Kg, 50 Kg, 60 Kg, dst), data tinggi badan (150 cm, 155 cm, 160 cm, dst), dan lain-lain.

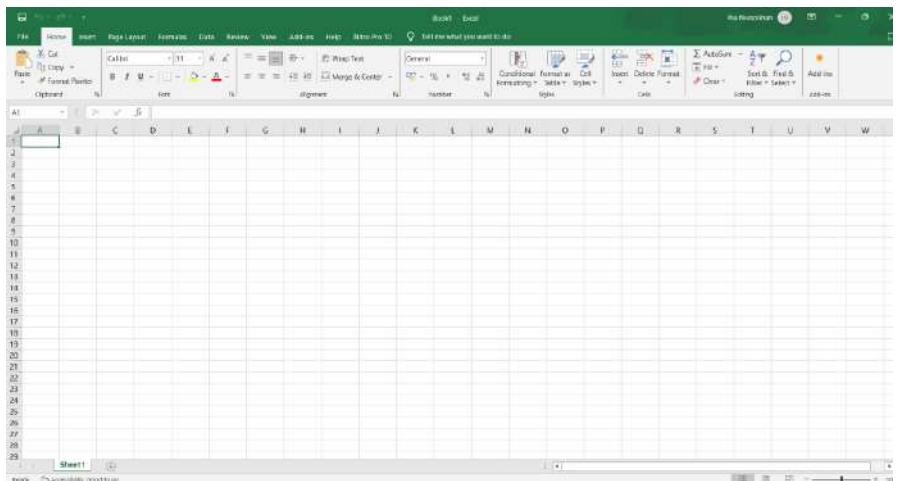
## **2.2 Aplikasi Statistik dalam Pengolahan Data**

Pengolahan data adalah tahapan kegiatan statistik yang penting. Data yang terkumpul dalam tahapan pengumpulan data akan menjadi sia-sia jika tahapan pengolahan data tidak dilakukan. Selain itu, jika kegiatan pengolahan data tidak dilakukan maka hasil akhir atau informasi yang disimpulkan dari data yang dikumpulkan dapat diragukan ketepatannya.

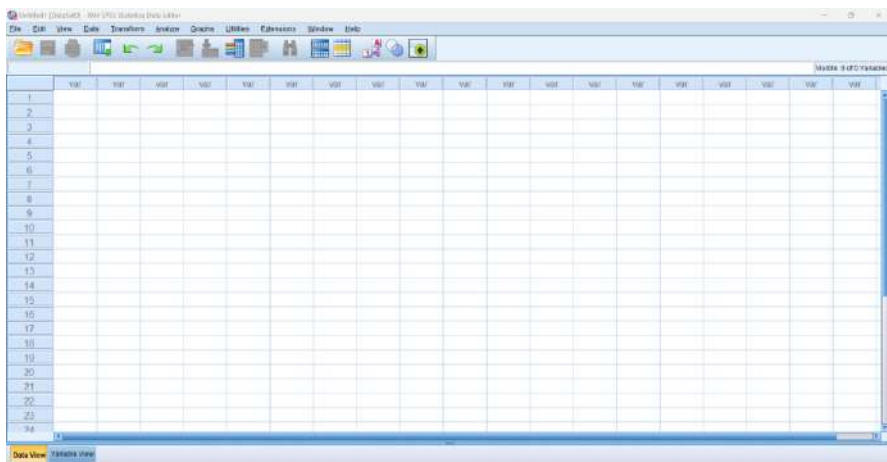
Pengolahan data semakin mudah dengan adanya berbagai program atau aplikasi yang membantu dalam pengolahan data. Aplikasi yang sering dan populer digunakan dalam pengolahan data di perguruan tinggi adalah Excel dan SPSS.

Microsoft excel adalah program perangkat lunak (*software*) lembar bentang yang paling populer di industri dan merupakan alat professional untuk menganalisis dan menampilkan data. Sedangkan SPSS merupakan singkatan dari *Statistical Program for Social Science* adalah paket program aplikasi komputer untuk mengolah dan menganalisis data statistik, hasilnya dapat dijadikan laporan dalam berbagai bentuk seperti tabulasi, grafik, diagram, dan berbagai distribusi, statistik deskriptif, ataupun statistik yang lebih kompleks lainnya.

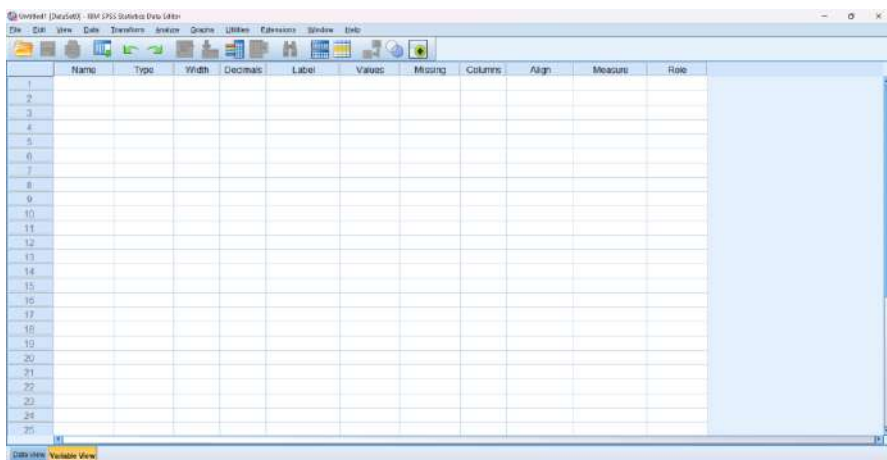
Terdapat persamaan dalam format penyajian data pada Excel dan SPSS, keduanya menyajikan data dalam format baris dan kolom. Baris menyajikan data nomor urut kuesioner dan data tiap responden, sedangkan kolom menyajikan data variabel yang diteliti. Berikut gambaran lembar kerja Excel dan SPSS:



**Gambar 2.1.** Halaman Lembar Kerja Excel



**Gambar 2.2.** Halaman Lembar Kerja SPSS (*Data View*)



**Gambar 2.3.** Halaman Lembar Kerja SPSS (*Variable View*)

## 2.3 Tahapan Pengolahan Data

Pengolahan data terbagi menjadi beberapa tahapan, diantaranya: *editing*, *coding*, *processing*, dan *cleaning*. Seluruh tahapan tersebut harus dilakukan secara sistematis dan teliti, berikut penjelasan dari masing-masing tahapan:

### 1. *Editing*

*Editing* adalah kegiatan statistik dalam memeriksa atau mengecek kembali isian dari formulir atau kuesioner yang diisi oleh para responden. *Editing* merupakan kegiatan

pertama dalam pengolahan data, prosesnya dilakukan langsung setelah pengumpulan data. Peneliti langsung memeriksa isian dari kuesioner yang disebarnya, sehingga jika terjadi kekeliruan atau kesalahan maka peneliti mengoreksinya langsung kepada responden yang bersangkutan. Berikut hal-hal yang perlu diperiksa pada isian kuesioner:

- a. Kelengkapan jawaban pada kuesioner  
Pastikan seluruh jawaban dalam kuesioner terisi lengkap. Periksa kembali setiap pertanyaan diisi oleh responden, baik pada pertanyaan tertutup ataupun terbuka.
- b. Jawaban pada kuesioner tertulis dan terbaca jelas  
Selain memeriksa bahwa seluruh pertanyaan diisi oleh responden, peneliti juga perlu mengecek kejelasan jawaban. Pastikan hanya satu jawaban yang diberikan pada kuesioner tertutup, sehingga jawaban tidak ambigu. Pastikan juga tulisan terbaca jelas dan dapat dimengerti oleh peneliti pada setiap pertanyaan terbuka.
- c. Jawaban yang diberikan responden relevan  
Relevansi jawaban dapat diperiksa dari pertanyaan-pertanyaan pada setiap kuesioner. Misalnya, suatu riset dengan responden ibu hamil, responden memberikan jawaban bahwa ini adalah kehamilannya yang ke-4, sedangkan pertanyaan usia dijawab bahwa responden berusia 20 tahun. Maka peneliti harus menanyakan kembali atas jawaban yang tidak relevan tersebut yaitu hamil ke-4 di usia 20 tahun, apakah data usia atau data jumlah kehamilan yang mungkin keliru.
- d. Jawaban yang diberikan responden konsisten  
Konsistensi jawaban dapat terlihat pada setiap pertanyaan yang berkaitan. Misalnya, suatu riset dengan responden Wanita Usia Subur (WUS), responden memberikan jawaban “Tidak” pada pertanyaan apakah menggunakan KB. Namun, pertanyaan selanjutnya mengenai Jenis KB yang

digunakan saat ini, responden tersebut menjawab KB suntik 3 bulan. Maka jawaban responden tersebut tidak konsisten. Perlu adanya pengecekan kembali atas jawaban-jawaban yang tidak konsisten tersebut.

Tahapan *editing* dalam pengolahan data ini penting dilakukan, agar data yang nantinya akan diolah terisi lengkap dan data yang ada merupakan jawaban riil dari responden.

## **2. Coding**

*Coding* adalah kegiatan statistik dalam merubah data yang berbentuk huruf menjadi berbentuk angka atau bilangan. Angka tersebut merupakan kode (nomor) khususnya pada variabel kategorik. Namun, pengkodean juga dapat dilakukan pada data numerik, yaitu membuat koding baru pada data numerik yang akan dijadikan variabel kategorik.

*Coding* merupakan proses pengolahan data yang dapat memudahkan dalam melakukan analisis data selanjutnya, karena beberapa program komputer membutuhkan data dalam bentuk angka untuk dapat mengolahnya. Jika dilihat pada Excel tampilan data sebelum dan setelah dilakukan coding akan seperti berikut ini:

| No | Pendidikan | Pendidikan | Status Bekerja | Status Bekerja |
|----|------------|------------|----------------|----------------|
| 1  | SMP        | 2          | Bekerja        | 1              |
| 2  | SMA        | 3          | Bekerja        | 1              |
| 3  | PT         | 4          | Tidak Bekerja  | 0              |
| 4  | SMA        | 3          | Bekerja        | 1              |
| 5  | SMA        | 3          | Tidak Bekerja  | 0              |
| 6  | SD         | 1          | Tidak Bekerja  | 0              |
| 7  | SMP        | 2          | Tidak Bekerja  | 0              |
| 8  | SD         | 1          | Bekerja        | 1              |
| 9  | SMA        | 3          | Bekerja        | 1              |
| 10 | SMA        | 3          | Tidak Bekerja  | 0              |
| 11 | PT         | 4          | Tidak Bekerja  | 0              |
| 12 | PT         | 4          | Bekerja        | 1              |
| 13 | SMP        | 2          | Tidak Bekerja  | 0              |
| 14 | SMA        | 3          | Bekerja        | 1              |
| 15 | SD         | 1          | Tidak Bekerja  | 0              |
| 16 | SMP        | 2          | Bekerja        | 1              |
| 17 | PT         | 4          | Tidak Bekerja  | 0              |
| 18 | SMP        | 2          | Tidak Bekerja  | 0              |
| 19 | PT         | 4          | Bekerja        | 1              |
| 20 | SD         | 1          | Bekerja        | 1              |

**Gambar 2.4.** Contoh Coding dalam Lembar Kerja Excel

Gambar diatas menunjukkan variabel pendidikan dengan kategori data SD, SMP, SMA, dan Perguruan Tinggi, dilakukan pengkodean dengan “1=SD” “2=SMP” “3=SMA” “4=Perguruan Tinggi” dan variabel status bekerja, dilakukan koding “1=tidak bekerja” “2=bekerja”. Namun, dalam memberikan angka untuk pengkodean perlu memperhatikan beberapa faktor, seperti jenis data, konsistensi, dan tujuan penelitian.

#### a. Jenis Data

Seperti diketahui bahwa data kategorik terdiri dari skala nominal dan skala ordinal. Keduanya memiliki perbedaan ketika dilakukan koding.

1) Data kategorik nominal, seperti pengertian dari skala nominal bahwa tidak ada tingkatan didalam pengkategorianya, maka tidak ada asumsi urutan

atau jarak antar datanya, sehingga angka pada koding data kategorik tidak ada perbedaan atau tingkatan yang membedakan.

Contohnya: variabel Jenis kelamin dikategorikan Laki-laki dan Perempuan, dapat dilakukan koding "1=Laki-laki" dan "2=Perempuan" atau sebaliknya "1=Perempuan" dan "2=Laki-laki" hal tersebut tidak menjadi masalah jika datanya merupakan data kategorik nominal.

- 2) Data kategorik ordinal, ciri dari skala ordinal adalah adanya tingkatan dalam tiap pengkategorian. Maka angka pada koding data ordinal diurutkan berdasarkan tingkatan atau peringkatnya.

Contohnya: variabel pendidikan yang dikategorikan menjadi SD, SMP, SMA, dan Perguruan Tinggi, maka pemberian angka harus sesuai urutan, menjadi "1=SD" "2=SMP" "3=SMA" "4=Perguruan Tinggi" atau "1=Perguruan Tinggi" "2=SMA" "3=SMP" "4=SD". Penggunaan urutan paling rendah ke yang paling tinggi atau paling tinggi ke yang paling rendah bergantung pada konsistensi dengan variabel lainnya.

#### **b. Konsistensi**

Melakukan koding juga harus memperhatikan konsistensi antar variabelnya, khususnya pada variabel dengan skala ordinal. Contohnya suatu riset ingin mengetahui pengaruh pendidikan terhadap pengetahuan seseorang. Maka koding antara variabel pendidikan dan pengetahuan harus konsisten. Jika pada variabel pengetahuan kategori yang paling rendah diletakkan pada kode yang rendah, maka variabel pendidikan dengan kategori rendah diberikan angka yang rendah juga, begitupun sebaliknya.

*Coding* yang dapat diberikan:

| <b>Pengetahuan</b> | <b>Pendidikan</b> |
|--------------------|-------------------|
| 1= Kurang          | 1=SD              |
| 2= Cukup           | 2=SMP             |

|                    |                     |
|--------------------|---------------------|
| 3= Baik            | 3=SMA               |
|                    | 4=Perguruan Tinggi  |
| <b>atau</b>        |                     |
| <b>Pengetahuan</b> | <b>Pendidikan</b>   |
| 1=Baik             | 1= Perguruan Tinggi |
| 2=Cukup            | 2=SMA               |
| 3=Kurang           | 3=SMP               |
|                    | 4=SD                |

### c. Tujuan Penelitian

Melakukan koding juga harus memperhatikan tujuan penelitian atau tujuan dari analisis yang akan digunakan. Contohnya: pada kuesioner, pertanyaan pendidikan terdiri dari enam kategori antara lain: SD, SMP, SMA, Strata 1, Strata 2, Strata 3. Namun, tujuan penelitian ingin melihat pendidikan yang dibedakan atas pendidikan dasar, pendidikan menengah, dan pendidikan tinggi. Maka perlu dilakukan koding atas variabel tersebut.

| <b>Sebelum koding</b> | <b>Setelah koding</b>       |
|-----------------------|-----------------------------|
| <b>Pendidikan</b>     | <b>Pendidikan</b>           |
| 1=SD                  | 1=Pendidikan Dasar (SD-SMP) |
| 2=SMP                 | 2=Pendidikan Menengah (SMA) |
| 3=SMA                 | 3=Pendidikan Tinggi (S1-S3) |
| 4=S1<br>5=S2<br>6=S3  |                             |

## 3. *Processing*

*Processing* adalah kegiatan statistik dalam memproses data agar data yang sudah terkumpul dan dilakukan koding dapat dianalisis. Memproses data dilakukan dengan cara memasukkan data ke paket program komputer. Seluruh variabel yang akan diteliti dimasukkan atau diinput ke dalam program komputer, dapat input langsung menggunakan SPSS atau lanjut menggunakan Excel. Berikut

contoh data yang sudah diinput dalam SPSS:

data driving EM classroom [data5] - IBM SPSS Statistics Data Editor

File Edit View Data Transformations Analyze Graphs Utilities Extensions Window Help

NO.ID

ses

age

age.kat

udu

kosa

P1

P2

P3

P4

P5

P6

P7

P8

P9

P10

P11

P12

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

2

**Gambar 2.5.** Contoh Input Data Pada Program SPSS (*Data View*)

|    | Name    | Type    | Width | Decimals | Label               | Values             | Missing | Columns | Align | Measure | Role  |
|----|---------|---------|-------|----------|---------------------|--------------------|---------|---------|-------|---------|-------|
| 1  | NO.ID   | Numeric | 8     | 0        |                     | None               | None    | 6       | Right | Scale   | Input |
| 2  | ses     | Numeric | 8     | 0        | Jenis Kelamin       | (1, Perempuan)     | None    | 6       | Right | Nominal | Input |
| 3  | age     | Numeric | 8     | 0        | Usia Responden      | None               | None    | 6       | Right | Scale   | Input |
| 4  | age.kat | Numeric | 8     | 0        |                     | (1, <40 tahun)     | None    | 10      | Right | Nominal | Input |
| 5  | udu     | Numeric | 8     | 0        | Pendidikan Terakhir | (1, SD)            | None    | 6       | Right | Nominal | Input |
| 6  | kosa    | Numeric | 8     | 0        | Pekerjaan Responden | (1, Tidak bekerja) | None    | 6       | Right | Nominal | Input |
| 7  | P1      | Numeric | 8     | 0        |                     | None               | None    | 6       | Right | Nominal | Input |
| 8  | P2      | Numeric | 8     | 0        |                     | None               | None    | 6       | Right | Nominal | Input |
| 9  | P3      | Numeric | 8     | 0        |                     | None               | None    | 6       | Right | Nominal | Input |
| 10 | P4      | Numeric | 8     | 0        |                     | None               | None    | 6       | Right | Nominal | Input |
| 11 | P5      | Numeric | 8     | 0        |                     | None               | None    | 6       | Right | Nominal | Input |
| 12 | P6      | Numeric | 8     | 0        |                     | None               | None    | 6       | Right | Nominal | Input |
| 13 | P7      | Numeric | 8     | 0        |                     | None               | None    | 6       | Right | Nominal | Input |
| 14 | P8      | Numeric | 8     | 0        |                     | None               | None    | 6       | Right | Nominal | Input |
| 15 | P9      | Numeric | 8     | 0        |                     | None               | None    | 6       | Right | Nominal | Input |
| 16 | P10     | Numeric | 8     | 0        |                     | None               | None    | 6       | Right | Nominal | Input |
| 17 | P11     | Numeric | 8     | 0        |                     | None               | None    | 6       | Right | Nominal | Input |
| 18 | P12     | Numeric | 8     | 0        |                     | None               | None    | 6       | Right | Nominal | Input |
| 19 | P13     | Numeric | 8     | 0        |                     | None               | None    | 6       | Right | Nominal | Input |
| 20 | P14     | Numeric | 8     | 0        |                     | None               | None    | 6       | Right | Nominal | Input |
| 21 | P15     | Numeric | 8     | 0        |                     | None               | None    | 6       | Right | Nominal | Input |
| 22 | P16     | Numeric | 8     | 0        |                     | None               | None    | 6       | Right | Scale   | Input |
| 23 | P17     | Numeric | 8     | 0        |                     | None               | None    | 6       | Right | Scale   | Input |
| 24 | P18     | Numeric | 8     | 0        |                     | None               | None    | 6       | Right | Scale   | Input |
| 25 | P19     | Numeric | 8     | 0        |                     | None               | None    | 6       | Right | Scale   | Input |

**Gambar 2.6.** Contoh Input Data Pada Program SPSS (*Variable View*)

#### 4. Cleaning

*Cleaning* atau pembersihan data merupakan kegiatan pengecekan kembali data yang sudah dimasukkan dalam program komputer, apakah terdapat kesalahan atau tidak, seperti: missing data, variasi data, atau konsistensi data.

Data yang telah diinput pada langkah *processing* disebut data mentah (*raw data*) yang belum dapat dianalisis lanjut sesuai dengan tujuan atau hipotesis, maka data tersebut perlu dibersihkan dari kesalahan umum yang sering terjadi, seperti salah ketik saat input, salah membaca kuesioner, atau masih ada data yang belum terinput (*missing data*).

*Cleaning* dapat dilakukan dengan membuat tabel distribusi frekuensi. Tabel distribusi frekuensi dapat digunakan untuk mengecek kembali adanya kesalahan data. Data dari sebuah variabel dapat dilihat kesalahannya berupa:

- Kesalahan ketik. Seperti data berat badan 55 Kg, namun yang tertulis 550 Kg. Kesalahan ini mungkin terjadi saat input data.
- Jawaban yang sama dengan kode *missing value*. Seperti usia 99 tahun. Angka 99 biasanya digunakan untuk kode *missing value* atau untuk jawaban “tidak tahu”. Maka perlu dicek kembali ketepatan datanya.

Proses *cleaning* dengan tabel distribusi frekuensi juga dilakukan untuk melihat *missing data*, seperti pada tabel distribusi frekuensi berikut ini:

**S1**

|         |        | Frequency | Percent | Valid Percent | Cumulative Percent |
|---------|--------|-----------|---------|---------------|--------------------|
| Valid   | 0      | 2         | 2.7     | 3.2           | 3.2                |
|         | 1      | 49        | 66.2    | 77.8          | 81.0               |
|         | 4      | 12        | 16.2    | 19.0          | 100.0              |
|         | Total  | 63        | 85.1    | 100.0         |                    |
| Missing | System | 11        | 14.9    |               |                    |
| Total   |        | 74        | 100.0   |               |                    |

**Gambar 2.7.** Contoh *Missing Data* Menggunakan Tabel Distribusi Frekuensi SPSS

Tabel diatas menunjukkan bahwa soal nomor 1 dari 74 responden, 11 orang tidak menjawab. Maka perlu dipastikan

kembali, apakah 11 responden tersebut memang tidak mengisi atau belum terinput datanya.

Selanjutnya tabel distribusi frekuensi dapat melihat variasi dan konsistensi data yang salah seperti contoh dibawah ini:

| Jenis Kelamin Responden |           |           |         |               |                    |
|-------------------------|-----------|-----------|---------|---------------|--------------------|
|                         |           | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid                   | Perempuan | 28        | 37.8    | 37.8          | 37.8               |
|                         | Laki-laki | 44        | 59.5    | 59.5          | 97.3               |
|                         | 10        | 1         | 1.4     | 1.4           | 98.6               |
|                         | 22        | 1         | 1.4     | 1.4           | 100.0              |
|                         | Total     | 74        | 100.0   | 100.0         |                    |

**Gambar 2.8.** Contoh Kesalahan Variasi Data Menggunakan Tabel Distibusi Frekuensi SPSS

| Status Bekerja Responden |               |           |         |               |                    |
|--------------------------|---------------|-----------|---------|---------------|--------------------|
|                          |               | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid                    | Tidak Bekerja | 29        | 39.2    | 39.2          | 39.2               |
|                          | Bekerja       | 45        | 60.8    | 60.8          | 100.0              |
|                          | Total         | 74        | 100.0   | 100.0         |                    |

| Pekerjaan Responden |               |           |         |               |                    |
|---------------------|---------------|-----------|---------|---------------|--------------------|
|                     |               | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid               | Tidak bekerja | 25        | 33.8    | 33.8          | 33.8               |
|                     | Nelayan       | 26        | 35.1    | 35.1          | 68.9               |
|                     | Wiraswasta    | 10        | 13.5    | 13.5          | 82.4               |
|                     | Karyawan      | 1         | 1.4     | 1.4           | 83.8               |
|                     | Pedagang      | 9         | 12.2    | 12.2          | 95.9               |
|                     | Buruh         | 3         | 4.1     | 4.1           | 100.0              |
|                     | Total         | 74        | 100.0   | 100.0         |                    |

**Gambar 2.9.** Contoh Kesalahan Konsistensi Data Menggunakan Tabel Distibusi Frekuensi SPSS

Pada tabel jenis kelamin responden terjadi kesalahan variasi data, kategori jenis kelamin dikoding dengan 1=Perempuan dan 2=Laki-laki, tetapi hasil tabel mendapatkan kesalahan variasi yang mungkin terjadi akibat kesalahan ketik, terdapat variasi 10 dan 22, pada data tersebut perlu dilakukan *cleaning* kembali dengan membetulkan data yang keliru. Selanjutnya kesalahan konsistensi data terdapat pada gambar 3.9 yang menunjukkan tabel status bekerja dan pekerjaan responden, pada tabel status bekerja mendapatkan hasil bahwa responden yang tidak bekerja jumlahnya 29 orang dan yang bekerja sebanyak 45 orang, sedangkan pada pertanyaan selanjutnya yang menanyakan jenis pekerjaan responden mendapatkan hasil bahwa responden yang tidak bekerja berkurang menjadi 25 orang yang seharusnya 29 orang. Maka diperlukan *cleaning* kembali pada dua pertanyaan mengenai pekerjaan responden, sehingga datanya konsisten.

## DAFTAR PUSTAKA

- Ariyani, H., Hairuddin, H., Palilingan, R.A., Nugroho, H., Sarumi, R. and Aji, R., 2023. Metodologi Penelitian Kesehatan dan Statistika.
- Roflin, E., Zulvia, F.E. and Liberty, I.A., 2021. *Pengolahan dan Penyajian Data Penelitian Bidang Kedokteran*. Penerbit NEM.

# **BAB 3**

## **PENATAAN DATA**

*Oleh Nur Al-Faida*

### **3.1 Pendahuluan**

Penataan data adalah proses pengaturan, organisasi, dan pengolahan data agar mudah diakses, dikelola, dan dimanfaatkan secara efektif (William H. Inmon, 2015).

Proses penataan data ini melibatkan beberapa tahapan, termasuk pengumpulan, pembersihan, penyusunan, dan dokumentasi data. Berikut adalah yang terkait dengan penataan data statistik:

1. Pengumpulan Data
2. Pembersihan Data
3. Transformasi Data
4. Penyusunan Data
5. Dokumentasi Data
6. Penyimpanan Data
7. Pemeliharaan Data

### **3.2 Pengumpulan Data**

Pengumpulan data adalah mencari, mencatat, dan mengumpulkan semua secara objektif dan apa adanya sesuai dengan hasil observasi dan wawancara di lapangan yaitu pencatatan data dan berbagai bentuk data yang ada di lapangan (Sugiyono, 2010).

Pengumpulan data juga merupakan serangkaian metode atau cara yang digunakan dalam proses penelitian untuk mengumpulkan data yang diperlukan untuk menjawab pertanyaan penelitian.

### 3.2.1 Jenis-jenis pengumpulan Data

Dalam melakukan pengumpulan data, terdapat beberapa jenis pengumpulan data yang paling sering digunakan dalam penelitian yaitu sebagai berikut :

#### 1. Observasi

Observasi diartikan sebagai cara mengadakan pencatatan secara sistematis mengenai tingkah laku dengan melihat atau mengamati tingkah laku individu atau kelompok yang diteliti secara langsung (Basrowi dan Suwandi, 2008).

Pengumpulan data observasi dilakukan melalui pengamatan langsung dilakukan oleh peneliti kepada objek yang diteliti secara langsung di tempat terjadinya peristiwa. Bila berupa perilaku manusia, fenomena atau proses perubahan.

#### 2. Wawancara

Wawancara adalah proses memperoleh keterangan untuk tujuan penelitian dengan cara tanya jawab sambil bertatap muka antara si penanya atau wawancara dengan menggunakan alat yang dinamakan *interview guide* atau panduan wawancara (Nazir, 1988).

Wawancara merupakan pengumpulan data dengan mengambil data secara langsung dengan melakukan komunikasi secara langsung dengan responden atau informan. Wawancara bisa memberikan pemahaman mendalam mengenai pengalaman, pandangan, atau pengetahuan individu mengenai topik penelitian.

#### 3. Kuesioner

Angket atau kuesioner merupakan teknik pengumpulan data yang dilakukan dengan cara memberi seperangkat pertanyaan atau pernyataan tertulis kepada responden untuk dijawab (Sugiyono, 2017).

Kuesioner atau angket digunakan untuk mendapatkan jawaban secara langsung dari responden. Metode ini bisa membantu mendapatkan jawaban responden dalam jumlah yang besar. Tipe pertanyaan kuesioner dibagi menjadi dua yaitu terbuka dan tertutup. Namun, diperlukan penyusunan pertanyaan angket yang akurat untuk mendapatkan jawaban yang sesuai.

4. Dokumentasi

Dokumentasi adalah suatu cara yang digunakan untuk memperoleh data dan informasi dalam bentuk buku, arsip, dokumen, tulisan angka, dan gambar yang berupa laporan serta keterangan yang dapat mendukung penelitian (Sugiyono, 2017). Dengan dokumentasi, peneliti dapat menggunakan sumber-sumber yang telah ada untuk mendukung hipotesis mereka, dan juga untuk mendapatkan pemahaman yang lebih mendalam tentang subjek penelitian tanpa perlu melakukan pengumpulan data secara langsung melalui metode observasi atau wawancara.

5. Studi Literatur

Studi literatur adalah merupakan penelitian yang dilakukan oleh peneliti dengan mengumpulkan sejumlah buku, majalah, atau literatur lainnya yang berkaitan dengan masalah dan tujuan penelitian (Danial dan Warsiah, 2009).

Studi literatur atau studi pustaka bertujuan untuk mencari berbagai teori-teori yang relevan dengan permasalahan yang sedang diteliti sebagai bahan rujukan dalam pembahasan hasil penelitian.

### **3.3 Pembersihan Data**

Pembersihan data adalah adalah proses memperbaiki atau menghapus kesalahan, ketidak konsistenan, dan ketidak akuratan dalam Kumpulan data.

Pembersihan data merupakan tahapan yang perlu dilakukan sebelum analisis data karena biasanya data mentah mengandung informasi yang tidak akurat, tidak tersusun rapi, atau tidak lengkap.

Pembersihan data atau cleaning adalah proses penting dalam penyusunan data yang bertujuan untuk mengidentifikasi, memperbaiki dan menghapus anomali, kesalan, atau ketidaksesuaian dalam kumpulan data. Berikut adalah langkah-langkah umum dalam proses pembersihan data:

1. Identifikasi Kesalahan: Langkah pertama adalah mengidentifikasikan potensi kesalahan dalam data. Ini bisa

termaksud nilai yang hilang, entri yang tidak valid, atau outlier yang tidak biasa.

2. **Penanganan Nilai yang Hilang:** Data sering kali memiliki nilai yang hilang atau kosong. Langkah ini melibatkan mengidentifikasi nilai yang hilang dan memutuskan cara penanganannya, seperti mengisi nilai yang hilang dengan nilai rata-rata, median, atau menggunakan teknik imputasi data.
3. **Identifikasi Kesalahan:** Langkah pertama adalah mengidentifikasikan potensi kesalahan dalam data. Ini bisa termaksud nilai yang hilang, entri yang tidak valid, atau outlier yang tidak biasa.
4. **Penanganan Nilai yang Hilang:** Data sering kali memiliki nilai yang hilang atau kosong. Langkah ini melibatkan mengidentifikasi nilai yang hilang dan memutuskan cara penanganannya, seperti mengisi nilai yang hilang dengan nilai rata-rata, median, atau menggunakan teknik imputasi data.
5. **Pendeteksian Duplikasi:** Kadang-kadang, data bisa mengandung entri duplikat yang perlu diidentifikasi dan dihapus untuk mencegah distorsi dalam analisis. Ini berdasarkan kriteria tertentu, seperti variabel tertentu atau kombinasi variabel.
6. **Penanganan Anomali:** Anomali atau outlier adalah nilai yang jauh dari pola umum dalam data. Langkah ini melibatkan identifikasi outlier dan memutuskan apakah harus dihapus, dinormalisasi, atau ditinjau lebih lanjut.
7. **Validasi Format:** Memastikan bahwa data memenuhi format yang diharapkan, seperti tanggal yang sesuai, format teks yang benar atau satuan yang konsisten.
8. **Konsistensi:** Mengkonfirmasi konsistensi data diseluruh kumpulan data. Misalnya, memastikan bahwa entri yang sama memiliki nilai yang seragam diseluruh dataset.
9. **Penyaringan Noise:** Noise adalah gangguan acak atau tidak relevan dalam data yang dapat mengaburkan sinyal yang sebenarnya. Langkah ini melibatkan identifikasi dan penghapusan noise dari data.

10. Verifikasi Data: Sebelum menyelesaikan proses pembersihan, penting untuk memverifikasi bahwa data dibersihkan secara tepat dan tidak mengandung kesalahan yang signifikan.

Pembersihan data yang efektif memastikan bahwa data yang digunakan untuk analisis atau pengambilan Keputusan adalah akurat, konsisten, dan dapat dipercaya. Hal ini mendukung hasil analisis yang lebih baik dan Keputusan yang lebih tepat.

### **3.4 Transformasi Data**

Transformasi data merupakan proses mengubah data dari skala pengukuran data asli menjadi skala bentuk lain dengan tujuan untuk memperlakukan karakteristik data dan memudahkan analisis.

Transformasi data dapat dilakukan dengan berbagai cara, seperti mengubah satuan ukuran data, mengubah distribusi data, atau mengubah bentuk data (Ghozalli, 2018).

Transformasi data adalah proses mengubah atau mentransformasi data dari satu bentuk atau representasi ke bentuk atau representasi lainnya. Tujuan utama dari transformasi data adalah untuk meningkatkan pemahaman, analisis, atau kinerja model dalam konteks tertentu. Berikut adalah beberapa teknik transformasi data umum:

1. Normalisasi: Normalisasi adalah proses mengubah nilai-nilai dalam suatu dataset ke dalam rentang tertentu, seperti 0 hingga 1 atau -1 hingga 1. Ini membantu dalam membandingkan variabel yang memiliki skala atau unit yang berbeda.
2. Standardisasi: Standardisasi adalah proses mengubah nilai-nilai dalam suatu dataset sehingga memiliki mean nol dan deviasi standar satu. Ini membantu dalam menghilangkan efek skala dan memudahkan perbandingan antar variabel.
3. Transformasi Logaritmik: Transformasi logaritmik melibatkan pengambilan logaritma dari nilai-nilai dalam suatu variabel. Ini sering digunakan untuk mengatasi

distribusi data yang miring ke kanan atau memiliki variasi yang besar.

4. Transformasi Box-Cox: Transformasi Box-Cox adalah teknik yang lebih umum digunakan untuk mengubah distribusi data menjadi lebih normal. Ini menghasilkan parameter transformasi yang optimal berdasarkan data yang diberikan.
5. Pengurangan Dimensi: Pengurangan dimensi adalah proses mengurangi jumlah variabel dalam dataset. Ini termaksud teknik seperti analisis komponen utama dan analisis faktor untuk mengurangi kompleksitas dan mempercepat analisis.
6. *Discretisasi*: *Discretisasi* adalah proses mengubah variabel kontinu menjadi variabel kategoris dengan membagi rentang nilai menjadi interval atau bin.
7. *Smoothing*: *Smoothing* adalah proses mengurangi fluktuasi atau noise dalam data dengan menggunakan teknik seperti *moving averages* atau *kernel smoothing*.
8. *Detrending*: *Detrending* adalah proses menghapus tren atau pola jangka panjang dari data untuk fokus pada fluktuasi jangka pendek atau variabilitas.
9. Imputasi Data: Imputasi data melibatkan pengisian nilai yang hilang dalam dataset dengan nilai yang diprediksi berdasarkan pola atau hubungan dalam data yang tersedia.

Transformasi data dapat meningkatkan interpretasi, analisis, atau kinerja model, tergantung pada kebutuhan dan karakteristik data yang diberikan.

### 3.5 Penyusunan Data

Penyusunan data adalah proses mengumpulkan, mengelompokkan, menyusun dan menyediakan data dalam bentuk yang bermanfaat untuk pengambilan keputusan (William G. Zikmund, 1997).

Penyusunan data merujuk pada proses mengumpulkan, mengorganisir, dan mengelola data agar dapat diakses dan digunakan dengan efisien. Berikut langkah-langkah umum dalam penyusunan data:

1. **Pengumpulan Data:** Langkah pertama adalah mengumpulkan data dari sumber-sumber yang relevan, seperti survei, database, atau sumber online. Pastikan untuk memperhatikan kualitas dan keakuratan data yang dikumpulkan.
2. **Pembersihan Data:** Setelah data terkumpul, langkah selanjutnya adalah pembersihan data dari noise, kesalahan, atau duplikasi. Ini melibatkan identifikasi dan koreksi masalah dalam data seperti nilai yang hilang, outlier, atau kesalahan pengetikan.
3. **Pengorganisasian Data:** Data kemudian diorganisasikan ke dalam format yang sesuai untuk analisis lebih lanjut. Ini bisa melibatkan pengelompokan data ke dalam tabel database, atau format lainnya, serta menetapkan struktur dan tata letak yang tepat.
4. **Integrasi Data:** Jika data berasal dari berbagai sumber, mereka mungkin perlu diintegrasikan agar dapat dianalisis bersama. Proses ini dapat melibatkan penggabungan data, penyesuaian skema, atau penghapusan inkonsistensi.
5. **Analisis Data:** Setelah data disusun dengan baik, Anda dapat menerapkan teknik analisis data untuk mendapatkan wawasan atau informasi yang berguna. Ini bisa melibatkan berbagai metode seperti statistik deskriptif, analisis regresi, atau *machine learning*.
6. **Interpretasi dan Visualisasi Data:** Hasil dari analisis data kemudian diinterpretasikan dan disajikan secara visual menggunakan grafik, diagram, atau laporan. Ini membantu untuk memahami pola tren, atau anomali dalam data.
7. **Penyimpanan dan Manajemen Data:** Terakhir, data yang sudah disusun dan dianalisis perlu disimpan dengan aman dan diatur dengan baik untuk penggunaan dan referensi di masa mendatang.

Penting untuk diingat bahwa penyusunan data adalah proses berkelanjutan yang dibutuhkan perhatian terus-menerus terhadap kualitas dan integritas data.

### 3.6 Dokumentasi Data

Dokumentasi Data adalah suatu cara yang digunakan untuk memperoleh data dan informasi dalam bentuk buku, arsip, dokumen, tulisan angka dan gambar yang berupa laporan serta keterangan yang dapat mendukung penelitian (Sugiyono, 2018).

Dokumentasi data adalah proses mendokumentasikan informasi tentang data yang dikumpulkan, termasuk sumbernya, struktur, format, konten, dan metode pengolahan. Dokumentasi ini penting untuk memudahkan pemahaman dan penggunaan data oleh orang lain, termasuk rekan kerja, peneliti, atau pemangku kepentingan lainnya. Berikut adalah beberapa langkah penting dalam membuat dokumentasi data:

1. **Deskripsi Umum:** Mulailah dengan memberikan deskripsi umum tentang data, termasuk tujuan pengumpulan data, sumbernya, konteksnya, dan pemiliknya. Ini membantu orang lain untuk memahami konteks dan kegunaan data.
2. **Metadata:** Metadata adalah informasi tentang data itu sendiri, seperti nama variabel, tipe data, unit ukuran, dan deskripsi singkat. Metadata membantu dalam pemahaman dan interpretasi data, serta memastikan konsistensi dan akurasi dalam penggunaannya.
3. **Struktur Data:** Dokumentasikan struktur data, termasuk format file, kolom, dan baris data. Jelaskan bagaimana data disusun, apakah dalam format tabel, database, atau file lainnya. Ini membantu pengguna untuk memahami cara mengakses dan memanipulasi data.
4. **Panduan Penggunaan:** Sertakan panduan penggunaan yang jelas untuk data, termasuk langkah-langkah untuk mengakses, memproses, dan menganalisis data. Berikan instruksi tentang cara menginterpretasikan hasil dan memperoleh wawasan dari data.
5. **Pengolahan Data:** Jika data telah diproses atau dimodifikasi secara signifikan, dokumentasikan langkah-langkah pengolahan data yang telah dilakukan. Ini membantu untuk memahami transformasi yang telah diterapkan pada data.

6. **Kontrol Versi:** Penting untuk melacak versi data dan dokumentasinya, terutama jika data diperbarui atau dimodifikasi secara berkala. Pastikan untuk menandai setiap perubahan dengan jelas dan menjaga catatan revisi yang akurat.
7. **Izin dan Keamanan:** Jika data memiliki batasan akses atau izin tertentu, dokumentasikan kebijakan keamanan dan hak akses yang berlaku. Pastikan untuk menjaga keamanan data dan mengikuti regulasi privasi yang berlaku.
8. **Referensi dan Sumber:** Cantumkan referensi dan sumber data dengan jelas. Ini membantu untuk melacak asal-usul data dan memberikan penghargaan kepada pemilik asli atau institusi yang terlibat dalam pengumpulan data.

Agar dokumentasi data baik, pengguna lain dapat dengan mudah memahami, mengakses, dan menggunakan data dengan efektif dan efisien.

### **3.7 Penyimpanan Data**

Penyimpanan data adalah proses dan teknologi untuk menyimpan informasi secara terstruktur dan terorganisir agar dapat diakses, dikelola, dan dipulihkan ketika diperlukan (Sugiarto dan Wahyono, 2014). Beberapa aspek penting dari penyimpanan data meliputi:

1. **Pengolahan Kapasitas:** Memastikan bahwa kapasitas penyimpanan cukup untuk menampung semua data yang dihasilkan oleh organisasi, dan dapat diperluas sesuai kebutuhan di masa mendatang.
2. **Keamanan Data:** Melindungi data dari akses yang tidak sah, kerusakan, atau kehilangan. Ini melibatkan penerapan kontrol akses, enkripsi, dan langkah-langkah keamanan lainnya.
3. **Integritas Data:** Memastikan bahwa data tetap utuh dan tidak diubah secara tidak sah selama penyimpanan dan pengambilan. Ini dapat mencakup penggunaan checksum, redundansi data, dan mekanisme verifikasi lainnya.

4. **Ketersediaan Data:** Memastikan bahwa data tersedia saat diperlukan oleh pengguna atau aplikasi. Ini melibatkan penggunaan teknologi penyimpanan yang tahan bencana dan strategi pemulihan bencana yang efektif.
5. **Efisiensi Penyimpanan:** Memanfaatkan ruang penyimpanan secara efisien dengan mengompresi data, deduplikasi, dan teknik lainnya untuk mengurangi *overhead* penyimpanan.
6. **Fleksibilitas:** Menyediakan infrastruktur penyimpanan yang dapat menangani berbagai jenis data dan skenario penggunaan, termasuk penyimpanan di lokasi lokal, cloud, atau hibrida.
7. **Manajemen Data:** Melakukan tugas-tugas seperti backup, pemulihan, penghapusan data yang tidak lagi diperlukan, dan migrasi data antara sistem penyimpanan.

Penyimpanan data telah berkembang pesat seiring dengan kemajuan teknologi, dengan munculnya solusi penyimpanan yang lebih canggih seperti penyimpanan berbasis cloud, penyimpanan flash, dan penyimpanan definisi perangkat lunak penyimpanan flash, dan penyimpanan definisi perangkat lunak (*software-defined storage*). Ini memberikan organisasi berbagai pilihan untuk memenuhi kebutuhan penyimpanan data mereka dengan cara yang paling efisien dan efektif.

### **3.8 Pemeliharaan Data**

Pemeliharaan data mencakup upaya untuk memastikan bahwa data mematuhi kebijakan, regulasi, dan standar yang berlaku, serta memastikan bahwa infrastruktur teknologi yang mendukung data tetap berjalan dengan baik (Sudrajat, 2011).

Dengan kata lain, pemeliharaan data bukan hanya sekadar menjaga integritas teknis data, tetapi juga melibatkan aspek kebijakan, keamanan, dan kepatuhan yang krusial untuk memastikan bahwa data dapat digunakan secara efektif dan aman oleh organisasi.

Pemeliharaan data adalah serangkaian kegiatan yang dilakukan secara teratur untuk memastikan bahwa data yang dimiliki oleh sebuah organisasi tetap dalam kondisi yang baik,

akurat, relevan, dan tersedia untuk digunakan sesuai kebutuhan. Proses pemeliharaan data ini mencakup beberapa aspek penting:

1. **Validasi Data:** Proses memastikan bahwa data yang masuk ke dalam sistem atau basis data telah diverifikasi kebenarannya. Ini melibatkan pemeriksaan data untuk memastikan bahwa setiap entri memiliki format yang benar, nilai yang valid, dan sesuai dengan aturan validasi yang ditetapkan.
2. **Pembersihan Data:** Langkah untuk mengidentifikasi dan memperbaiki kesalahan atau anomali dalam data. Ini meliputi penghapusan atau koreksi entri yang tidak lengkap, tidak akurat, atau tidak relevan, serta penanganan nilai yang hilang.
3. **Pengamanan Data:** Upaya untuk melindungi data dari akses yang tidak sah, perubahan yang tidak diizinkan, atau kerusakan yang disengaja atau tidak disengaja. Ini melibatkan penerapan kontrol akses, enkripsi data, dan langkah-langkah keamanan lainnya.
4. **Pemantauan Kinerja Data:** Proses memantau kinerja sistem dan proses yang berhubungan dengan data untuk mengidentifikasi masalah atau penyimpangan yang memengaruhi integritas atau ketersediaan data.
5. **Backup dan Pemulihan:** Tindakan untuk membuat salinan cadangan data secara teratur dan menyimpannya di lokasi yang aman. Jika terjadi kehilangan data atau kerusakan pada sistem, salinan cadangan ini dapat digunakan untuk memulihkan data yang hilang.
6. **Penghapusan Data yang tidak diperlukan:** Praktik untuk menghapus data yang sudah tidak relevan, usang, atau tidak diperlukan lagi. Ini membantu mengurangi kompleksitas sistem, mengurangi risiko keamanan, dan mematuhi regulasi privasi data.
7. **Dokumentasi dan Metadata:** Penting untuk mendokumentasikan struktur, asal, dan konteks data, serta menyertakan metadata yang memudahkan pencarian, pemahaman, dan penggunaan data oleh orang lain.

8. Kepatuhan Regulasi: Memastikan bahwa pemeliharaan data mematuhi peraturan, standar, dan kebijakan yang berlaku, seperti GDPR, HIPAA, atau PCI DSS, tergantung pada jenis data yang dikelola dan industri tempat organisasi beroperasi.

Pemeliharaan data yang efektif membantu untuk memastikan bahwa data yang mereka miliki dapat dipercaya, berguna, dan aman digunakan untuk pengambilan keputusan, analisis, dan operasi bisnis sehari-hari.

## DAFTAR PUSTAKA

- Basrowi dan Suwandi, 2008. Memahami Penelitian Kualitatif. Jakarta : Rieka Cipta.
- Danial dan Warsiah, 2009. Metode Penulisan Karya Ilmiah. Bandung: Laboratorium Pendidikan Kewarganegaraan UPI.
- Ghozali, 2018. Aplikasi Analisis Multivariate dengan Program IBM SPSS 25. Badan Penerbit Universitas Diponegoro: Semarang.
- Nazir, 1988. Metode Penelitian. Jakarta : Ghalia Indonesia.
- Sudrajat. 2011. Pedoman Praktis Manajemen Perawatan Mesin Industri, Bandung : PT Refika Aditama.
- Sugiarto dan Wahyono. 2014. Manajemen Kearsipan Elektronik. Yogyakarta: Penerbit Gava Media.
- Sugiyono, 2010. Metode Penelitian Pendidikan Pendekatan Kuantitatif, kualitatif dan R&D. Bandung: Alfabeta.
- Sugiyono, 2017. Metode Penelitian Pendidikan Pendekatan Kuantitatif, kualitatif dan R&D. Bandung: Alfabeta.
- William G. Zikmund, 1997. Business Research Methods (Eight edition). South Western, USA: Cengage Learning.
- William H. Inmon, 2015. Building The Data Warehouse. John Wiley & Sons.



# **BAB 4**

## **ANALISIS DATA**

*Oleh Bambang Suhartawan*

### **4.1 Pendahuluan**

Analisis data dalam bidang biostatistika adalah proses interpretasi, pengolahan, dan pemodelan data yang berkaitan dengan ilmu biologi, kesehatan, dan kedokteran menggunakan metode statistika. Tujuannya adalah untuk memahami hubungan antara variabel-variabel yang diamati, mengidentifikasi pola atau tren, menguji hipotesis, dan membuat prediksi yang relevan dalam konteks ilmu biologi atau kedokteran.

Dalam biostatistika, analisis data dapat melibatkan berbagai teknik statistika, seperti statistika deskriptif (misalnya, mean, median, dan distribusi frekuensi), statistika inferensial (misalnya, uji hipotesis, interval kepercayaan, dan analisis regresi), dan metode statistika lanjutan (misalnya, analisis multivariat, analisis survival, dan analisis longitudinal).

Analisis data dalam biostatistika biasanya dilakukan menggunakan perangkat lunak statistika seperti R, Python, SPSS, atau SAS, yang memungkinkan para peneliti untuk menerapkan berbagai metode statistika yang sesuai dengan data yang mereka miliki, serta untuk membuat visualisasi yang dapat membantu dalam pemahaman dan interpretasi hasil analisis.

### **4.2 Tahapan Analisis Data dalam Bidang Biostatistika**

Analisis data dalam biostatistika merupakan proses penting dalam mengeksplorasi, menganalisis, dan menginterpretasi data yang berkaitan dengan studi-studi di bidang biologi, kesehatan, dan kedokteran. Tujuan utama dari analisis data dalam biostatistika adalah untuk menyelidiki hubungan antara variabel yang diamati, mengidentifikasi pola

atau tren dalam data, menguji hipotesis yang diajukan, serta membuat generalisasi atau prediksi yang relevan terhadap fenomena yang diamati. Berikut adalah tahapan umum dalam analisis data dalam biostatistika:

### **1. Penyusunan Data**

Tahap awal dalam analisis data adalah penyusunan data mentah yang diperoleh dari studi atau percobaan. Ini melibatkan pengumpulan data yang relevan sesuai dengan desain penelitian yang telah ditetapkan.

Penyusunan data merupakan tahapan awal yang krusial dalam analisis data dalam biostatistika. Proses ini melibatkan pengumpulan dan penyusunan data mentah yang diperoleh dari studi atau percobaan sesuai dengan desain penelitian yang telah ditetapkan sebelumnya. Berikut adalah penjelasan lebih detail tentang penyusunan data beserta contoh penerapannya:

Pengumpulan Data:

- a. Perencanaan Pengumpulan Data: Sebelum memulai pengumpulan data, peneliti perlu merencanakan proses pengumpulan dengan cermat. Hal ini melibatkan pemilihan instrumen pengukuran yang sesuai, pemilihan sampel yang representatif, serta perencanaan jadwal dan lokasi pengumpulan data.
- b. Pengumpulan Data: Setelah perencanaan selesai, data dikumpulkan sesuai dengan desain penelitian yang telah ditetapkan. Ini bisa dilakukan melalui survei, observasi langsung, wawancara, atau pengamatan laboratorium, tergantung pada jenis studi yang dilakukan.

Pengelompokan dan Pengkodean Data:

- a. Pengelompokan Data: Setelah data terkumpul, langkah selanjutnya adalah mengelompokkan data ke dalam kategori-kategori yang sesuai. Misalnya, dalam studi tentang jenis penyakit, data dapat dikelompokkan menjadi kategori penyakit tertentu seperti flu, diabetes, atau kanker.

- b. Pengkodean Data: Data kemudian dikodekan agar dapat diproses dan dianalisis lebih lanjut. Pengkodean ini bisa berupa penggunaan angka atau kode tertentu untuk merepresentasikan kategori atau variabel yang diamati. Misalnya, jika dalam studi ditemukan tiga jenis penyakit: flu, diabetes, dan kanker, maka flu dapat dikode sebagai 1, diabetes sebagai 2, dan kanker sebagai 3.

**Verifikasi dan Pembersihan Data:**

- a. Verifikasi Data: Data yang telah terkumpul perlu diverifikasi untuk memastikan keakuratannya. Ini melibatkan pengecekan kembali setiap entri data untuk memastikan tidak ada kesalahan pencatatan atau kekurangan data.
- b. Pembersihan Data: Setelah verifikasi, data kemudian dibersihkan dari data yang tidak lengkap, data yang tidak sesuai, atau outlier yang mungkin mempengaruhi hasil analisis. Misalnya, jika terdapat nilai yang tidak mungkin atau ekstrem yang tidak masuk akal, maka nilai tersebut mungkin perlu dihapus atau diperiksa lebih lanjut.

**Contoh Penerapan:**

Misalnya, dalam sebuah studi observasional tentang pola makan dan risiko penyakit jantung, peneliti akan mengumpulkan data tentang pola makan (misalnya, konsumsi buah, sayuran, lemak, dll.) dan riwayat penyakit jantung dari sejumlah partisipan yang dipilih secara acak. Data kemudian akan dikelompokkan berdasarkan variabel-variabel yang diamati, seperti jenis makanan yang dikonsumsi dan riwayat penyakit jantung. Selanjutnya, data akan dikodekan, diverifikasi, dan dibersihkan untuk memastikan keakuratannya sebelum dilanjutkan ke tahapan analisis lebih lanjut.

## **2. Deskripsi Data**

Langkah selanjutnya adalah menganalisis statistika deskriptif dari data. Ini termasuk menghitung ukuran-

ukuran pusat (seperti mean, median, modus) dan ukuran penyebaran (seperti rentang, simpangan baku) untuk setiap variabel yang diamati. Visualisasi seperti histogram, diagram batang, atau diagram pencar juga sering digunakan untuk memberikan gambaran visual tentang distribusi data. Deskripsi data merupakan langkah penting dalam analisis data yang melibatkan penentuan karakteristik-karakteristik dasar dari kumpulan data yang diamati. Ini membantu peneliti untuk memahami distribusi data, tendensi sentral, serta variabilitasnya. Berikut adalah uraian detail tentang deskripsi data beserta contoh penerapannya dalam bidang biostatistika:

a. Menghitung Ukuran-Ukuran Pusat:

- 1) Mean: Merupakan nilai rata-rata dari semua pengamatan dalam dataset. Ini dihitung dengan menjumlahkan semua nilai dalam dataset dan kemudian membaginya dengan jumlah pengamatan.
- 2) Median: Merupakan nilai tengah dalam dataset ketika data diurutkan dari yang terkecil hingga yang terbesar. Jika jumlah pengamatan ganjil, median adalah nilai tengah; jika genap, median adalah rata-rata dari dua nilai tengah.
- 3) Modus: Merupakan nilai yang paling sering muncul dalam dataset.

b. Menghitung Ukuran Penyebaran:

- 1) Rentang: Merupakan selisih antara nilai maksimum dan minimum dalam dataset.
- 2) Simpangan Baku: Merupakan ukuran sebaran data dari mean. Ini mengukur seberapa jauh setiap nilai dalam dataset dari mean.

Contoh Penerapan dalam Bidang Biostatistika:

Misalnya, dalam studi klinis tentang efek obat terhadap tekanan darah, berikut adalah contoh penerapan analisis deskriptif:

- a. Menghitung Mean: Peneliti menghitung rata-rata tekanan darah (sistolik dan diastolik) dari semua pasien yang berpartisipasi dalam studi.
- b. Menghitung Median: Jika terdapat 100 pasien dalam studi, median tekanan darah dihitung dengan mengurutkan semua nilai tekanan darah dari yang terkecil hingga yang terbesar, lalu menemukan nilai tengahnya.
- c. Menghitung Modus: Jika dalam data tekanan darah ada nilai yang paling sering muncul (misalnya, tekanan darah sistolik 120 mmHg muncul paling sering), maka nilai tersebut merupakan modus.
- d. Menghitung Rentang: Rentang tekanan darah dihitung dengan mengurangi nilai tekanan darah maksimum dengan nilai tekanan darah minimum.
- e. Menghitung Simpangan Baku: Simpangan baku tekanan darah dihitung untuk mengetahui sebaran data dari nilai rata-rata tekanan darah. Semakin besar simpangan baku, semakin bervariasi data tekanan darah dari rata-rata.

#### Visualisasi Data:

- a. Histogram: Digunakan untuk menampilkan distribusi frekuensi dari data numerik. Misalnya, histogram tekanan darah akan menunjukkan sebaran frekuensi tekanan darah dalam dataset.
- b. Diagram Batang: Dapat digunakan untuk menunjukkan perbandingan antara kategori-kategori dalam dataset. Misalnya, diagram batang dapat digunakan untuk membandingkan jumlah pasien dengan penyakit tertentu di berbagai kelompok usia.
- c. Diagram Pencar: Digunakan untuk menunjukkan hubungan antara dua variabel numerik. Misalnya, diagram pencar dapat digunakan untuk melihat hubungan antara tekanan darah sistolik dan diastolik pada pasien-pasien tertentu.

Penerapan analisis deskriptif ini membantu peneliti untuk memahami dan menggambarkan karakteristik dasar dari data yang diamati dalam studi kesehatan atau biologi, yang merupakan langkah awal penting sebelum melanjutkan ke analisis statistik inferensial atau pemodelan lebih lanjut.

### **3. Pemilihan Metode Analisis**

Berdasarkan pertanyaan penelitian dan sifat data yang ada, peneliti memilih metode analisis yang paling sesuai. Ini bisa meliputi analisis univariat, bivariat, atau multivariat, serta teknik-teknik statistik tertentu seperti uji hipotesis, analisis regresi, atau analisis survival, tergantung pada kebutuhan penelitian.

Pemilihan metode analisis merupakan langkah penting dalam analisis data dalam bidang biostatistika. Pemilihan metode yang tepat bergantung pada pertanyaan penelitian yang diajukan, jenis data yang dikumpulkan, dan tujuan dari analisis tersebut. Berikut adalah uraian detail tentang pemilihan metode analisis beserta contoh penerapannya dalam biostatistika:

#### **a. Analisis Univariat:**

Analisis univariat fokus pada satu variabel tunggal dalam dataset dan bertujuan untuk memahami karakteristik dasar dari variabel tersebut.

Contoh Penerapan:

Dalam studi epidemiologi, peneliti mungkin melakukan analisis univariat untuk melihat distribusi frekuensi penyakit tertentu dalam populasi yang diamati. Misalnya, menghitung jumlah kasus penyakit malaria pada setiap kelompok usia dalam populasi.

#### **b. Analisis Bivariat:**

Analisis bivariat memeriksa hubungan antara dua variabel dalam dataset dan bertujuan untuk mengetahui apakah ada hubungan atau korelasi di antara keduanya.

Contoh Penerapan:

Dalam studi kesehatan masyarakat, peneliti mungkin melakukan analisis bivariat untuk mengevaluasi hubungan antara konsumsi alkohol dan risiko kanker hati. Ini melibatkan memeriksa apakah ada hubungan antara jumlah alkohol yang dikonsumsi (variabel independen) dan risiko kanker hati (variabel dependen).

c. Analisis Multivariat:

Analisis multivariat memeriksa hubungan antara tiga atau lebih variabel dalam dataset dan dapat mengidentifikasi pola yang kompleks atau interaksi antara variabel.

Contoh Penerapan:

Dalam studi genetika, peneliti mungkin melakukan analisis multivariat untuk memeriksa hubungan antara beberapa gen tertentu dan risiko terkena penyakit jantung koroner. Ini melibatkan mempertimbangkan efek beberapa gen (variabel independen) terhadap risiko penyakit jantung koroner (variabel dependen) secara bersamaan.

d. Teknik-teknik Statistik Tertentu:

Selain itu, terdapat berbagai teknik statistik khusus yang dapat digunakan tergantung pada jenis data dan pertanyaan penelitian yang diajukan. Contoh teknik-teknik ini meliputi:

- 1) Uji Hipotesis: Digunakan untuk menentukan apakah perbedaan atau hubungan antara variabel dalam dataset adalah hasil kebetulan ataukah signifikan secara statistik.
- 2) Analisis Regresi: Digunakan untuk memodelkan hubungan antara satu atau lebih variabel independen (prediktor) dengan satu variabel dependen (variabel yang ingin diprediksi).
- 3) Analisis Survival: Digunakan untuk memeriksa waktu yang diperlukan untuk terjadinya suatu peristiwa tertentu (seperti kematian atau kejadian klinis) dalam studi longitudinal.

#### Contoh Penerapan:

Dalam studi klinis, peneliti mungkin menggunakan analisis regresi untuk memeriksa hubungan antara faktor risiko tertentu (misalnya, merokok, tekanan darah) dan perkembangan penyakit jantung koroner. Analisis survival juga dapat digunakan untuk memeriksa waktu yang diperlukan bagi pasien untuk mengalami kekambuhan penyakit setelah pengobatan.

Pemilihan metode analisis yang tepat sangat penting dalam memastikan validitas dan interpretasi yang tepat dari hasil analisis data dalam bidang biostatistika. Dengan memahami sifat data dan tujuan penelitian, peneliti dapat memilih metode analisis yang paling sesuai untuk menjawab pertanyaan penelitian yang diajukan.

#### **4. Analisis Inferensial**

Pada tahap ini, peneliti melakukan analisis inferensial untuk menggeneralisasi temuan dari sampel ke populasi yang lebih luas. Ini sering melibatkan penggunaan uji hipotesis statistik untuk menentukan apakah perbedaan atau hubungan antara variabel adalah hasil kebetulan ataukah nyata.

Analisis inferensial merupakan tahapan penting dalam proses analisis data dalam bidang biostatistika. Pada tahap ini, peneliti melakukan generalisasi dari temuan yang diperoleh dari sampel ke populasi yang lebih luas. Tujuan utamanya adalah untuk membuat kesimpulan atau inferensi tentang populasi berdasarkan data yang diperoleh dari sampel. Analisis inferensial sering melibatkan penggunaan uji hipotesis statistik untuk menentukan apakah perbedaan atau hubungan antara variabel yang diamati adalah hasil kebetulan atau nyata. Berikut adalah uraian detail tentang analisis inferensial beserta contoh penerapannya dalam bidang biostatistika:

a. Uji Hipotesis:

Uji hipotesis adalah teknik statistik yang digunakan untuk menguji pernyataan tentang parameter populasi berdasarkan informasi yang diperoleh dari sampel. Prosedur umum dalam uji hipotesis melibatkan pembentukan hipotesis nol ( $H_0$ ) yang menyatakan tidak ada perbedaan atau hubungan yang signifikan, serta hipotesis alternatif ( $H_1$ ) yang menyatakan adanya perbedaan atau hubungan yang signifikan.

Contoh Penerapan:

Dalam studi klinis tentang efektivitas suatu obat terhadap penyakit tertentu, peneliti mungkin ingin menguji hipotesis nol bahwa tidak ada perbedaan dalam rerata pengurangan gejala antara kelompok pasien yang diberi obat dan kelompok kontrol yang tidak diberi obat. Hipotesis alternatifnya adalah bahwa ada perbedaan dalam rerata pengurangan gejala antara kedua kelompok tersebut.

b. Interval Kepercayaan:

Interval kepercayaan adalah kisaran nilai yang mungkin berisi parameter populasi dengan tingkat keyakinan tertentu. Interval kepercayaan sering digunakan untuk mengevaluasi ketidakpastian dalam estimasi parameter populasi.

Contoh Penerapan:

Dalam studi epidemiologi tentang prevalensi penyakit tertentu dalam populasi, peneliti mungkin ingin mengestimasi persentase populasi yang menderita penyakit tersebut. Mereka dapat menggunakan interval kepercayaan untuk memberikan perkiraan kisaran persentase populasi dengan tingkat keyakinan tertentu, misalnya, 95% interval kepercayaan untuk prevalensi penyakit.

c. Analisis Regresi:

Analisis regresi digunakan untuk memeriksa hubungan antara satu atau lebih variabel independen dengan variabel dependen. Ini memungkinkan peneliti untuk

mengevaluasi pengaruh variabel independen terhadap variabel dependen dan memprediksi nilai variabel dependen berdasarkan nilai variabel independen.

Contoh Penerapan:

Dalam studi nutrisi tentang hubungan antara asupan kalori dan berat badan, peneliti dapat menggunakan analisis regresi untuk menentukan seberapa kuat hubungan antara asupan kalori (variabel independen) dengan berat badan (variabel dependen). Analisis ini dapat membantu dalam memahami pengaruh asupan kalori terhadap perubahan berat badan.

d. Uji Beda (T-Test, ANOVA, dll.):

Uji beda digunakan untuk membandingkan rata-rata atau proporsi antara dua atau lebih kelompok yang berbeda. Ini memungkinkan peneliti untuk menentukan apakah ada perbedaan yang signifikan antara kelompok-kelompok tersebut.

Contoh Penerapan:

Dalam studi klinis tentang efek diet tertentu terhadap kadar kolesterol, peneliti mungkin ingin membandingkan rata-rata kadar kolesterol antara kelompok yang mengikuti diet tersebut dengan kelompok kontrol yang tidak mengikuti diet. Uji beda seperti uji t-tidak berpasangan dapat digunakan untuk menentukan apakah ada perbedaan yang signifikan antara kedua kelompok tersebut.

Penerapan analisis inferensial ini penting dalam menyimpulkan temuan dari data sampel ke populasi yang lebih luas dalam bidang biostatistika. Dengan memahami teknik-teknik analisis inferensial yang tepat, peneliti dapat membuat kesimpulan yang akurat dan relevan tentang hubungan dan perbedaan antar variabel yang diamati dalam studi biologi, kesehatan, dan kedokteran.

## 5. Interpretasi Hasil

Setelah analisis selesai, peneliti menginterpretasikan hasil-hasil yang diperoleh dalam konteks pertanyaan penelitian yang diajukan. Ini melibatkan penafsiran makna statistik dari temuan yang dihasilkan dan mengaitkannya kembali dengan teori atau literatur yang ada dalam bidang tersebut. Interpretasi hasil merupakan tahap penting dalam analisis data dalam bidang biostatistika. Setelah analisis statistik selesai dilakukan, peneliti perlu menginterpretasikan temuan yang diperoleh dan menghubungkannya kembali dengan pertanyaan penelitian yang diajukan serta teori atau literatur yang relevan. Berikut adalah uraian detail tentang interpretasi hasil beserta contoh penerapannya dalam bidang biostatistika:

### a. Penafsiran Makna Statistik:

- 1) Menginterpretasikan Hasil Uji Hipotesis: Peneliti harus memahami apakah hasil uji hipotesis menunjukkan adanya perbedaan atau hubungan yang signifikan antara variabel yang diamati. Ini melibatkan penafsiran nilai p-nilai dan keputusan apakah menolak atau gagal menolak hipotesis nol.
- 2) Menginterpretasikan Hasil Analisis Regresi: Dalam analisis regresi, peneliti harus memahami seberapa kuat hubungan antara variabel independen dan dependen serta signifikansinya dalam konteks penelitian.

### b. Kaitkan Kembali dengan Pertanyaan Penelitian:

Relevansi dengan Tujuan Penelitian: Hasil analisis harus dikaitkan kembali dengan pertanyaan penelitian yang diajukan. Peneliti harus menjelaskan apakah temuan mendukung atau menolak hipotesis penelitian serta implikasinya dalam konteks penelitian.

### c. Mengaitkan dengan Teori atau Literatur yang Ada:

Hubungan dengan Pengetahuan Terdahulu: Peneliti harus menghubungkan temuan mereka dengan teori atau pengetahuan yang ada dalam bidang tersebut. Ini memungkinkan mereka untuk melihat apakah hasil

penelitian mereka konsisten atau bertentangan dengan penelitian sebelumnya.

Contoh Penerapan dalam Bidang Biostatistika:

Misalkan sebuah studi observasional dilakukan untuk mengevaluasi hubungan antara konsumsi buah-buahan dan risiko penyakit jantung koroner. Setelah analisis selesai, berikut adalah contoh interpretasi hasil:

- a. Makna Statistik: Hasil analisis menunjukkan bahwa terdapat hubungan negatif yang signifikan antara konsumsi buah-buahan dan risiko penyakit jantung koroner (nilai  $p$ -nilai  $< 0.05$ ). Ini menunjukkan bahwa semakin tinggi konsumsi buah-buahan, semakin rendah risiko penyakit jantung koroner.
- b. Kaitkan Kembali dengan Pertanyaan Penelitian: Hasil ini mendukung hipotesis penelitian bahwa konsumsi buah-buahan dapat mengurangi risiko penyakit jantung koroner.
- c. Mengaitkan dengan Teori atau Literatur yang Ada: Temuan ini konsisten dengan penelitian sebelumnya yang menunjukkan bahwa diet tinggi buah-buahan berkaitan dengan penurunan risiko penyakit kardiovaskular.

Dengan cara ini, interpretasi hasil dalam bidang biostatistika memungkinkan peneliti untuk menyajikan temuan mereka dengan jelas dan memberikan kontribusi pada pemahaman ilmiah yang lebih luas tentang topik yang diteliti.

## **6. Pelaporan dan Presentasi**

Hasil analisis data disajikan dalam laporan penelitian atau publikasi ilmiah. Biasanya, hasil ini juga dipresentasikan secara lisan di konferensi atau seminar ilmiah.

Pelaporan dan presentasi merupakan tahap akhir dalam proses analisis data dalam bidang biostatistika. Pada tahap ini, hasil analisis data disajikan dalam bentuk laporan penelitian atau publikasi ilmiah, serta dapat dipresentasikan secara lisan di konferensi atau seminar ilmiah. Berikut

adalah uraian detail tentang pelaporan dan presentasi hasil analisis data, beserta contoh penerapannya dalam bidang biostatistika:

a. Laporan Penelitian atau Publikasi Ilmiah:

- 1) Struktur Laporan Penelitian: Laporan penelitian biasanya terdiri dari bagian-bagian seperti pendahuluan, metode, hasil, dan kesimpulan. Setiap bagian memberikan informasi yang relevan tentang studi yang dilakukan, termasuk desain penelitian, metode pengumpulan data, hasil analisis, dan interpretasi temuan.
- 2) Presentasi Hasil Analisis: Hasil analisis data disajikan dengan jelas dan disertai dengan tabel, grafik, atau diagram yang mendukung. Penjelasan statistik dan interpretasi hasil diberikan secara rinci, dan kesimpulan yang didasarkan pada temuan yang diperoleh disampaikan dengan jelas.

Contoh Penerapan:

Dalam studi klinis tentang efek terapi baru terhadap penyakit tertentu, laporan penelitian akan mencakup deskripsi tentang desain studi, karakteristik sampel, metode analisis data, hasil uji hipotesis, dan interpretasi temuan. Hasil analisis data akan disajikan dalam bentuk tabel atau grafik, dan kesimpulan akan menyajikan implikasi klinis dari temuan tersebut.

b. Presentasi Lisan di Konferensi atau Seminar Ilmiah:

- 1) Struktur Presentasi: Presentasi lisan biasanya terdiri dari bagian-bagian seperti latar belakang, tujuan penelitian, metode, hasil, dan kesimpulan. Presentasi disusun secara sistematis untuk menjelaskan studi yang dilakukan dan menyoroti temuan utama.
- 2) Visualisasi: Presentasi lisan sering didukung oleh visualisasi seperti slide PowerPoint, grafik, atau diagram untuk membantu penonton memahami temuan dengan lebih baik.

### Contoh Penerapan:

Seorang peneliti dalam bidang epidemiologi dapat menyampaikan hasil studi mereka dalam konferensi ilmiah tentang kesehatan masyarakat. Mereka dapat menggunakan slide PowerPoint untuk memperlihatkan grafik yang menggambarkan prevalensi penyakit dalam populasi yang diteliti, serta memberikan penjelasan tentang interpretasi hasil dan implikasinya dalam konteks kesehatan masyarakat.

Dengan cara ini, pelaporan dan presentasi hasil analisis data dalam bidang biostatistika memungkinkan peneliti untuk berbagi temuan mereka dengan masyarakat ilmiah secara efektif, serta memperluas pemahaman tentang topik yang diteliti dan mendorong diskusi dan kolaborasi lebih lanjut.

Selama seluruh proses analisis data, penting untuk memperhatikan prinsip-prinsip statistika yang baik, seperti memperhatikan asumsi model, menghindari bias, dan mempertimbangkan kekuatan dan keterbatasan dari teknik analisis yang digunakan. Keseluruhan, analisis data dalam biostatistika tidak hanya memberikan wawasan mendalam tentang fenomena yang diamati, tetapi juga memungkinkan peneliti untuk membuat keputusan informasional yang berbasis bukti dalam ilmu biologi, kesehatan, dan kedokteran.

## DAFTAR PUSTAKA

- Boyd, D.M., & Ellison, N.B. (2007). "Social Network Sites: Definition, History, and Scholarship." *Journal of Computer-Mediated Communication*, 13 (1), article 11, <http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html>.
- Gall, A.G., Gall, J.P., & Borg, W.R. (2003). *Educational Reserach: An Introduction*. New York: Pearson Education.
- Howell, D. C. (2011). *Fundamental Statistics for the Behavioral Sciences* (7rd ed). Belmont, CA: Wadsworth.
- Johnson, R. A. (2010). *Statistics: Principles and Methods* (6 rd ed).Hoboken, NJ: John Wiley & Sons
- Nurdyansyah. N., Andiek Widodo, (2010), *Manajemen Sekolah Berbasis ICT*. Sidoarjo: Nizamia Learning Center.
- Ramli, K., Sarwoto, dan Rusadi, (2010), *Komunikasi dan Informatika Indonesia Whitepaer 2010*, Jakarta: Pusat Data Kementerian Komunikasi dan Informatika.
- Ratnawati Susanto, (2019), *Modul 2, Kebijakan Pemerintah dalam ICT dan Pendidikan*, Universitas Esa Unggul.
- Rogantina Meri Andri, (2017), *Peran dan Fungsi Teknologi Dalam Peningkatan Kualitas Pembelajaran*, *Jurnal Ilmiah Research Sains*, Vol. 3, No. 1.
- Undang-Undang RI Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional
- Wiwik Angranti, *Educational Management Based on Community Needs: A New Paradigm of Education Management in Indonesia*, (2018), *Journal of Education and Practice*, Vol. 9, No. 14.



# BAB 5

## TEORI PROBABILITAS

*Oleh Khairul Alim*

### 5.1 Pendahuluan

Dalam percakapan sehari-hari, istilah "probabilitas" digunakan untuk mengukur tingkat keyakinan seseorang terhadap kemungkinan bahwa suatu peristiwa akan terjadi di masa depan. Namun, kami ingin memahami konteks probabilitas, bagaimana itu diukur, dan bagaimana itu membantu membuat inferensi.

Bekerja dengan mekanisme fisika, biologis, atau sosial yang menghasilkan pengamatan yang tidak dapat diprediksi dengan pasti membutuhkan konsep probabilitas. Sebagai contoh, tekanan darah seseorang pada titik tertentu tidak dapat diprediksi, dan jembatan tidak pernah tahu berapa banyak beban yang akan ditanggungnya sebelum roboh ke sungai. Meskipun peristiwa acak seperti ini tidak dapat diprediksi dengan pasti, frekuensi relatif dengan jumlah yang terjadi selama berbagai uji coba seringkali sangat stabil. Peristiwa acak atau stokastik adalah peristiwa yang memiliki ciri-ciri ini. Jika observasi di masa depan dilakukan, konsekuensi relatif yang stabil dalam jangka panjang ini memberikan ukuran yang bermakna secara intuitif terhadap keyakinan kita bahwa suatu peristiwa acak akan terjadi.

Misalnya, meskipun tidak mungkin untuk memprediksi dengan pasti berapa banyak kepala yang muncul pada satu kali lemparan koin yang seimbang, kita dapat mengatakan dengan keyakinan yang wajar bahwa persentase kepala akan sangat dekat dengan 0,5 dalam beberapa uji coba yang panjang. Ini karena, dari sudut pandang penjudi, frekuensi relatif ini umumnya digunakan sebagai ukuran keyakinan dalam hasil satu kali lemparan. Dia mempertaruhkan uang pada satu kali lemparan koin, bukan serangkaian lemparan. Frekuensi relatif

dari kepala dalam serangkaian lemparan yang panjang, yang disebut penjudi sebagai probabilitas dari kepala, memberinya ukuran kemungkinan untuk menang pada satu kali lemparan. Jika koin tidak seimbang dan memberikan 90% kepala dalam serangkaian lemparan yang panjang, penjudi akan mengatakan probabilitas dari kepala adalah 0,9, dan dia akan cukup yakin dengan kemunculan.

Contoh sebelumnya memiliki analogi yang masuk akal dan dapat digunakan. Semua orang adalah penjudi dalam banyak hal. Dokter peneliti telah menginvestasikan waktu dan sumber daya finansial untuk proyek penelitian ini, dan pada satu kesempatan, dia berharap keberhasilannya. Dengan cara yang sama, investasi modal dalam pabrik manufaktur baru adalah seperti bermain game dengan harapan besar untuk sukses. Pengusaha hanya tertarik pada fraksi investasi serupa yang berhasil dalam berbagai uji coba karena mereka memberikan bukti keyakinan dalam hasil yang sukses dari satu investasi.

Meskipun mudah dipahami, konsep frekuensi relatif probabilitas tidak memberikan definisi yang ketat tentang probabilitas. Banyak konsep lain telah diajukan, seperti probabilitas subjektif, yang memungkinkan tingkat kemungkinan suatu peristiwa bervariasi tergantung pada orang yang melakukan evaluasinya. Namun demikian, untuk tujuan kami, kami menerima interpretasi frekuensi relatif sebagai ukuran yang signifikan dari keyakinan kami bahwa suatu peristiwa akan terjadi. Selanjutnya, kami akan memeriksa hubungan yang disediakan oleh probabilitas antara pengamatan dan inferensi.

## **5.2 Ruang Sampel dan Kejadian**

### **5.2.1 Ruang Sampel**

Dalam teori probabilitas, ruang sampel adalah himpunan semua hasil yang mungkin dari suatu eksperimen atau situasi yang diamati. Secara sederhana, ruang sampel adalah kumpulan dari semua kemungkinan hasil yang dapat terjadi dalam suatu percobaan dinotasikan dengan  $S$ .

Misalnya, jika kita melempar sebuah koin, ruang sampelnya akan berisi dua hasil yang mungkin: "kepala" atau "ekor". Oleh karena itu, ruang sampel untuk percobaan melempar koin adalah  $S = \{K, E\}$ , di mana "K" menunjukkan kepala dan "E" menunjukkan ekor.

Dalam konteks biostatistik, ruang sampel seringkali melibatkan hasil-hasil yang berkaitan dengan peristiwa biologis atau kesehatan. Misalnya, jika kita mengukur tingkat kolesterol dalam sampel darah, ruang sampelnya mungkin adalah semua kemungkinan nilai-nilai kolesterol yang mungkin kita peroleh dalam pengujian tersebut.

Penting untuk dicatat bahwa ruang sampel harus mencakup semua kemungkinan hasil yang mungkin, dan setiap hasil individual dalam ruang sampel harus eksklusif dan tidak saling tumpang tindih.

Dengan memahami ruang sampel, kita dapat mengidentifikasi semua hasil yang mungkin dari suatu percobaan atau pengamatan, yang merupakan langkah penting dalam menganalisis data dan membuat inferensi dalam teori probabilitas dan statistik.

### 5.2.2 Kejadian

Dalam teori probabilitas, kejadian (event) merujuk kepada satu atau lebih hasil yang mungkin terjadi dari suatu eksperimen atau pengamatan yang dilakukan. Dengan kata lain, kejadian adalah subset dari ruang sampel, yang terdiri dari satu atau lebih hasil dalam ruang sampel.

Misalnya, dalam percobaan melempar sebuah dadu, ruang sampelnya adalah himpunan nilai-nilai yang mungkin muncul saat dadu dilempar, yaitu  $S = \{1, 2, 3, 4, 5, 6\}$ . Sebagai contoh kejadian, kita dapat mengambil "munculnya angka genap" sebagai kejadian, yang terdiri dari himpunan  $G = \{2, 4, 6\}$ . Kejadian lainnya mungkin adalah "munculnya angka lebih besar dari 4", yang terdiri dari himpunan  $H = \{5, 6\}$ .

Dalam konteks biostatistik, kejadian dapat berhubungan dengan hasil-hasil pengamatan atau peristiwa yang terjadi dalam penelitian di bidang biologi dan kesehatan. Contoh

kejadian dalam bidang ini mungkin termasuk "peningkatan risiko penyakit jantung koroner pada perokok" atau "keberhasilan pengobatan pada pasien dengan kondisi tertentu".

Kejadian dapat memiliki sifat-sifat tertentu, seperti kejadian yang saling lepas (*mutually exclusive events*) yang tidak bisa terjadi bersamaan, atau kejadian independen yang tidak dipengaruhi oleh hasil kejadian lainnya.

Dengan memahami konsep kejadian, kita dapat menganalisis probabilitas dari berbagai hasil yang mungkin terjadi dalam berbagai situasi, dan ini merupakan landasan penting untuk pengambilan keputusan yang berbasis data dalam berbagai disiplin ilmu, termasuk biostatistik.

Untuk sebarang dua kejadian  $E$  dan  $F$  dari ruang sampel  $S$  kita dapat mendefinisikan kejadian baru  $E \cup F$  sebagai kejadian yang memuat titik-titik yang berada dalam  $E$  dan  $F$  atau dalam kedua  $E$  dan  $F$ . Ini berarti kejadian  $E \cup F$  akan terjadi bila salah satu  $E$  dan  $F$  terjadi. Sebagai contoh,  $G = \{2,4,6\}$  dan  $H = \{5,6\}$  maka  $G \cup H = \{2,4,5,6\}$ .

Selain itu, untuk sebarang dua kejadian  $E$  dan  $F$ , kita juga dapat mendefinisikan kejadian baru  $E \cap F$  yang biasa disebut irisan  $E$  dan  $F$ . Himpunan  $E \cap F$  terdiri dari semua titik yang berada dalam kedua  $E$  dan  $F$  sekaligus. Ini berarti kejadian  $E \cap F$  terjadi bila dan hanya bila  $E$  dan  $F$  kedua-duanya terjadi. Sebagai contoh,  $G = \{2,4,6\}$  dan  $H = \{5,6\}$  maka  $G \cap H = \{6\}$ .

## 5.3 Peluang

### 5.3.1 Syarat Peluang

Perhatikan suatu percobaan dengan ruang sampel  $S$ . Untuk setiap kejadian  $E$  dari ruang sampel  $S$ , diandaikan bilangan  $P(E)$  terdefinisi dan memenuhi tiga syarat berikut

1.  $0 \leq P(E) \leq 1$
2.  $P(S) = 1$
3. Untuk sebarang barisan kejadian  $E_1, E_2, \dots$  yang saling asing, yaitu kejadian  $E_i \cap E_j = \emptyset$  untuk  $i \neq j$  maka  $P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$ .

Kita menyebut  $P(E)$  sebagai peluang atau probabilitas dari kejadian  $E$ .

**Contoh:**

Pada pelemparan mata uang, bila kita andaikan muka (M) mempunyai kemungkinan yang sama dengan belakang (B) untuk muncul, maka kita mempunyai probabilitas  $P(\{B\}) = P(\{M\}) = \frac{1}{2}$ . Di lain pihak, bila kita mempunyai mata uang tidak seimbang dan merasa bahwa muka dua kali kemungkinannya untuk muncul dibandingkan belakang, maka kita mempunyai  $P(\{M\}) = \frac{2}{3}$  dan  $P(\{B\}) = \frac{1}{3}$ .

**Contoh:**

Pada pelemparan sebuah dadu, bila kita andaikan keenam sisi mempunyai kemungkinan yang sama untuk muncul, maka kita mempunyai  $P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = \frac{1}{6}$ . Dari syarat ketiga peluang di atas, maka probabilitas mendapatkan mata dadu genap adalah  $P(\{2,4,6\}) = P(\{2\}) + P(\{4\}) + P(\{6\}) = \frac{1}{2}$ .

Karena kejadian  $E$  dan  $E^c$  selalu saling asing dan karena  $E \cup E^c = S$  maka dari syarat peluang kedua dan ketiga di dapat  $P(S) = P(E) + P(E^c) = 1$  atau  $P(E^c) = 1 - P(E)$ . Hal ini memiliki arti bahwa peluang suatu kejadian tidak terjadi sama dengan satu dikurangi peluangnya terjadi.

Perhatikan  $S \cup \emptyset = S$  dan  $S \cap \emptyset = \emptyset$ . Sehingga  $P(S) + P(\emptyset) = P(S)$  atau  $1 + P(\emptyset) = 1$  atau  $P(\emptyset) = 0$ .

### 5.3.2 Peluang Gabungan dan Irisan

Diberikan dua kejadian  $E$  dan  $F$  dari suatu ruang sampel. Peluang  $P(E \cup F)$  adalah peluang dari kejadian-kejadian yang berada dalam  $E$  atau  $F$ . Lebih lanjut,  $P(E) + P(F)$  adalah peluang semua kejadian dalam  $E$  ditambah peluang semua kejadian dalam  $F$ . Jika kedua kejadian ini beririsan, artinya ada kejadian yang terjadi dalam  $E$  sekaligus dalam  $F$  dan kejadian ini dihitung dua kali pada  $P(E) + P(F)$ , tetapi hanya satu kali dalam  $P(E \cup F)$ , sehingga akan berlaku

$$P(E) + P(F) = P(E \cup F) + P(E \cap F)$$

atau secara ekuivalen

$$P(E \cup F) = P(E) + P(F) - P(E \cap F).$$

Perhatikan lagi jika dalam kasus  $E$  dan  $F$  saling asing, maka  $E \cap F = \emptyset$ , sehingga

$$P(E \cup F) = P(E) + P(F).$$

### Contoh:

Pada eksperimen melempar dua mata uang seimbang, yang berarti memiliki ruang sampel  $S = \{(M, M), (B, M), (M, B), (B, B)\}$  berkemungkinan semua, atau setiap kejadian mempunyai peluang  $\frac{1}{4}$ . Andaikan  $E = \{(M, M), (B, M)\}$  dan  $F = \{(M, M), (M, B)\}$ .

Dengan menggunakan rumus sebelumnya, maka

$$\begin{aligned} P(E \cup F) &= P(E) + P(F) - P(E \cap F) \\ &= \frac{1}{2} + \frac{1}{2} - P(\{(M, M)\}) \\ &= 1 - \frac{1}{4} = \frac{3}{4}. \end{aligned}$$

Peluang ini dapat juga dihitung secara langsung, yaitu  $P(E \cup F) = P(\{(M, M), (B, M), (M, B)\}) = \frac{3}{4}$ .

### 5.3.3 Peluang Bersyarat

Misalkan kita melemparkan dua dadu dan misalkan dari 36 hasil yang mungkin masing-masing mempunyai kemungkinan yang sama untuk muncul, yang berarti masing-masing mempunyai peluang  $\frac{1}{36}$ . Misalkan kita mengetahui dadu pertama menghasilkan mata empat. Maka dengan informasi ini, kita akan menentukan peluang jumlah mata dadu yang tampak sama dengan enam.

Karena dadu pertama menghasilkan mata empat, maka paling banyak terdapat 6 hasil dari eksperimen di atas, yaitu  $(4,1)$ ,  $(4,2)$ ,  $(4,3)$ ,  $(4,4)$ ,  $(4,5)$ , dan  $(4,6)$ . Karena masing-masing hasil awalnya mempunyai peluang yang sama untuk muncul, maka masing-masing hasil akhirnya pun memiliki peluang yang sama. Ini berarti, dengan diketahuinya dadu pertama menghasilkan mata empat, maka peluang (bersyarat) setiap

hasil (4,1), (4,2), (4,3), (4,4), (4,5), dan (4,6) adalah  $\frac{1}{6}$  sedangkan peluang (bersyarat) 30 titik yang lain dalam ruang sampel adalah 0. Karena itu, peluang yang ditanyakan adalah  $\frac{1}{6}$ . Bila kita misalkan  $E$  menyatakan kejadian jumlah mata dadu enam dan  $F$  menyatakan kejadian dadu pertama empat, maka peluang yang baru dihitung disebut peluang bersyarat  $E$  terjadi jika diketahui  $F$  telah terjadi dan dinyatakan dengan  $P(E|F)$ .

Rumus umum untuk  $P(E|F)$  yang berlaku untuk semua kejadian  $E$  dan  $F$  dirumuskan dengan cara yang sama seperti di atas. Katakanlah, bila kejadian  $F$  terjadi, maka agar  $E$  terjadi, kejadian sebenarnya haruslah berupa titik-titik yang berada dalam  $E$  dan  $F$ , yaitu harus berada dalam  $E \cap F$ . Sekarang, karena kita mengetahui  $F$  telah terjadi, maka  $F$  menjadi ruang sampel kita yang baru dan akibatnya peluang kejadian  $E \cap F$  terjadi sama dengan peluang  $E \cap F$  relatif dengan peluang  $F$ . Ini berarti

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

Perhatikan bahwa persamaan di atas hanya terdefinisi jika  $P(F) > 0$  dan akibatnya  $P(E|F)$  hanya terdefinisi jika  $P(F) > 0$ .

### **Contoh:**

Misalkan kartu-kartu yang diberi nomor satu sampai sepuluh ditempatkan pada sebuah topi, dikocok, dan kemudian diambil satu kartu. Bila kita diberi tahu bahwa nomor kartu yang diambil paling sedikit lima, tentukan peluang bersyarat kartu yang terambil bernomor 10.

### **Penyelesaian:**

Misalkan  $E$  menyatakan kejadian nomor kartu yang terambil sepuluh, dan  $F$  kejadian nomor yang terambil paling sedikit lima. Peluang yang dicari adalah  $P(E|F)$ .

Untuk menghitung peluang tersebut, gunakan rumus

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

Tetapi  $E \cap F = E$ , karena nomor kartu yang mempunyai sifat nomor 10 dan paling sedikit 5 jika dan hanya jika nomor itu 10. Maka

$$P(E|F) = \frac{\frac{1}{10}}{\frac{1}{6}} = \frac{1}{6}.$$

### 5.3.4 Kejadian Bebas atau Independen

Dua kejadian  $E$  dan  $F$  disebut independen, jika  $P(E \cap F) = P(E)P(F)$ . Dengan menggunakan rumus peluang bersyarat, berarti  $E$  dan  $F$  independen bila  $P(E|F) = P(E)$  (yang berarti juga  $P(F|E) = P(F)$ ). Ini berarti terjadinya  $E$  bebas atau independen dengan  $E$  terjadi atau tidak. Dua kejadian  $E$  dan  $F$  yang tidak independen disebut *dependen*.

#### Contoh:

Misalkan kita melempar dua dadu seimbang. Misalkan  $E_1$  menyatakan kejadian jumlah mata dadu yang tampak enam dan  $F$  menyatakan kejadian dadu pertama menghasilkan mata empat.

Maka  $P(E_1 \cap F) = P(\{4,2\}) = \frac{1}{36}$  sedangkan  $P(E_1)P(F) = \frac{5}{36} \cdot \frac{1}{6} = \frac{5}{6}$  dan akibatnya  $E_1$  dan  $F$  tidak independen.

Sekarang misalkan  $E_2$  kejadian yang menyatakan jumlah mata dadu tujuh. Apakah  $E_2$  independen dengan  $F$ ? Jawabannya adalah ya, karena  $P(E_1 \cap F) = P(\{4,3\}) = \frac{1}{36}$  sedangkan  $P(E_2)P(F) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$  dan akibatnya  $E_2$  dan  $F$  independen.

Definisi independensi di atas dapat diperluas untuk lebih dari dua kejadian. Kejadian-kejadian  $E_1, E_2, \dots, E_n$  disebut *independen* jika untuk setiap himpunan bagian  $E_1, E_2, \dots, E_r, r \leq n$  dan kejadian-kejadian tersebut berlaku

$$P(E_1 \cap E_2 \cap \dots \cap E_r) = P(E_1)P(E_2) \dots P(E_r).$$

Secara intuitif, kejadian-kejadian  $E_1, E_2, \dots, E_n$  disebut independen jika pengetahuan terjadinya salah satu kejadian tersebut tidak mempunyai pengaruh pada peluang terjadinya kejadian lainnya.

### 5.3.5 Rumus Bayes

Misalkan  $E$  dan  $F$  kejadian. Kita bisa menyatakan  $E$  sebagai

$$E = (E \cap F) \cup (E \cap F^c).$$

Karena  $E \cap F$  dan  $E \cap F^c$  saling asing maka kita mempunyai

$$\begin{aligned} P(E) &= P(E \cap F) + P(E \cap F^c) \\ &= P(E|F)P(F) + P(E|F^c)P(F^c) \\ &= P(E|F)P(F) + P(E|F^c)(1 - P(F)) \end{aligned}$$

Persamaan di atas menyatakan bahwa peluang kejadian  $E$  adalah rata-rata tertimbang peluang bersyarat  $E$  diketahui  $F$  telah terjadi dan peluang bersyarat  $E$  dengan  $F$  tidak terjadi.

#### Contoh:

Misalkan kita mempunyai dua kotak. Kotak pertama memuat dua bola kuning dan tujuh bola hitam, sedang kotak kedua memuat lima bola kuning dan enam bola hitam. Kita melemparkan sebuah mata uang seimbang dan mengambil satu bola dari kotak pertama atau kotak kedua tergantung apakah hasilnya muka atau belakang. Tentukan peluang bersyarat hasil lemparan muka bila diketahui bola kuning yang terambil.

#### Penyelesaian:

Misalkan  $K$  menyatakan kejadian bola kuning yang terambil dan  $M$  kejadian mata uang menghasilkan muka. Peluang yang dicari  $P(M|K)$  dapat dihitung sebagai berikut

$$\begin{aligned} P(M|K) &= \frac{P(M \cap K)}{P(K)} \\ &= \frac{P(K|M)P(M)}{P(K)} \\ &= \frac{P(K|M)P(M)}{P(K|M)P(M) + P(K|M^c)P(M^c)} \\ &= \frac{\frac{2}{9} \times \frac{1}{2}}{\frac{2}{9} \times \frac{1}{2} + \frac{5}{11} \times \frac{1}{2}} = \frac{22}{67}. \end{aligned}$$

Persamaan di atas dapat diperluas pada kejadian-kejadian  $F_1, F_2, \dots, F_n$  kejadian-kejadian yang saling asing sedemikian hingga  $\bigcup_{i=1}^n F_i = S$ . Dengan perkataan lain, tepat satu dari kejadian-kejadian  $F_1, F_2, \dots, F_n$  akan terjadi.

Dengan menulis

$$\begin{aligned}
 E &= (E \cap F_1) \cup (E \cap F_2) \cup \dots \cup (E \cap F_n) \\
 &= \bigcup_{i=1}^n (E \cap F_i)
 \end{aligned}$$

dan menggunakan kejadian-kejadian  $E \cap F_i, i = 1, 2, \dots, n$  saling asing, didapat

$$\begin{aligned}
 P(E) &= \sum_{i=1}^n P(E \cap F_i) \\
 &= \sum_{i=1}^n P(E|F_i)P(F_i).
 \end{aligned}$$

Jadi, persamaan di atas menunjukkan bagaimana untuk kejadian-kejadian  $F_1, F_2, \dots, F_n$  di mana satu dan hanya satu harus terjadi, kita dapat menghitung  $P(E)$  dengan pertama-tama "mensyaratkan" pada salah satu  $F_i$  yang terjadi. Ini menyatakan bahwa  $P(E)$  sama dengan rata-rata tertimbang  $P(E|F_i)$ , dengan masing-masing suku berboboti peluang kejadian yang disyaratkan. Sekarang misalkan  $E$  telah terjadi dan kita tertarik untuk menentukan mana salah satu dari  $F_i$  yang juga terjadi. Sehingga, diperoleh

$$\begin{aligned}
 P(F_j|E) &= \frac{P(E \cap F_j)}{P(E)} \\
 &= \frac{P(E \cap F_j)P(F_j)}{\sum_{i=1}^n P(E|F_i)P(F_i)}
 \end{aligned}$$

Persamaan ini disebut dengan *rumus bayes*.

## DAFTAR PUSTAKA

- Bain, L.J. and Engelhardt, M. 2014, *Introduction to Probability and Mathematical Statistics*. Andover: Cengage Learning.
- Hoel, P.G. 2005. *Introduction to Mathematical Statistics*. New York: John Wiley & Sons.
- Hogg, R.V., McKean, J.W. and Craig, A.T. 2020. *Introduction to Mathematical Statistics*. Harlow, Essex: Pearson Education Limited.
- Kerns, G.J. 2010. *Introduction to Probability and Statistics Using R*, s.n., S.l.
- Rasch, D. and Schott, D. 2018. *Mathematical Statistics*. Hoboken, NJ: John Wiley & Sons, Inc.



# BAB 6

## DISTRIBUSI PROBABILITAS

*Oleh Samsul Bahri Loklomin*

### 6.1 Pendahuluan

Dalam statistika terdapat dua konsep yang yang agak mirip yaitu distribusi frekuensi dan distribusi probabilitas. Perlu diketahui perbedaan distribusi frekuensi dengan distribusi probabilitas. Distribusi frekuensi hanya mendata semua hasil akhir yang memungkinkan dari suatu eksperimen dan menghitung jumlah frekuensi semua hasil akhir yang muncul. Hitungan tersebut kemudian disusun menjadi grafik yang membuatnya mudah divisualisasikan dan dipahami. Distribusi probabilitas menggambarkan semua hasil akhir yang memungkinkan dari suatu variabel, bukan menggambarkan frekuensinya. Kedua pemahaman ini kedengarannya serupa, namun kedua pengertian tersebut adalah dua konsep yang berbeda. Tabel yang berisikan probabilitas dari setiap kemungkinan hasil suatu percobaan, disebut dengan tabel distribusi probabilitas atau tabel distribusi peluang.

Dalam distribusi probabilitas terdapat 2 jenis distribusi berdasarkan variabel randomnya yaitu distribusi probabilitas diskrit dan distribusi probabilitas kontinu. Jika kemungkinan hasil dari suatu percobaan berupa variabel random diskrit maka jenis distribusinya adalah distribusi probabilitas diskrit. Sedangkan jika kemungkinan hasil dari suatu percobaan merupakan variabel random kontinu maka jenis distribusinya adalah distribusi probabilitas kontinu. Diberikan percobaan random, kita dapat menemukan himpunan dari semua kemungkinan hasil yang dikenal sebagai ruang sampel. Objek di ruang sampel tidak hanya angka. Jadi, kita menggunakan gagasan variabel random untuk mengukur elemen kualitatif ruang sampel. Variabel random dicirikan baik dengan fungsi kepadatan probabilitas atau fungsi distribusi kumulatifnya.

Karakteristik lain dari variabel random adalah rata-rata, varians dan fungsi pembangkit momen. Pada bagian ini akan dibahas lebih lanjut beberapa distribusi yang umumnya digunakan di bidang biostatistika yaitu distribusi Bernoulli, Binomial, Poisson, Geometri dan distribusi Normal.

## 6.2 Distribusi Bernoulli

### Prinsip Dasar

Percobaan Bernoulli adalah eksperimen random yang di dalamnya terdapat dua kemungkinan hasil yaitu sukses (S) atau gagal (F). Ruang sampel dari 2 kemungkinan tersebut dapat didefinisikan  $\{S, F\}$  ke dalam himpunan bilangan real sebagai  $X(S) = 1$  dan  $X(F) = 0$  dimana 1 adalah sukses dan 0 adalah gagal. Sifat-Sifat Percobaan Bernoulli :

1. Percobaan itu terdiri dari  $n$  pengulangan
2. Setiap hasil pengulangan dapat diidentifikasi sebagai sukses atau gagal
3. Probabilitas sukses dapat dinyatakan dengan  $p$  sedangkan probabilitas gagal adalah  $q = 1 - p$
4. Setiap pengulangan dengan pengulangan lainnya saling bebas.

### Definisi

Variabel random  $X$  berdistribusi Bernoulli jika fungsi kepadatan probabilitasnya berbentuk

$$f(x) = p^x(1 - p)^{1-x}, \quad x = 0, 1$$

dimana  $p$  adalah kemungkinan sukses. Variabel random Bernoulli dapat ditulis  $X \sim BER(p)$

### Contoh Soal

Kemungkinan kesalahan 20% dalam sebuah pemeriksaan kesehatan. Jika satu pasien dipilih secara acak dari 50 pasien, hitunglah peluang keberhasilan dari pemeriksaan tersebut.

### Jawab :

$q = 20\%$  maka  $20\% \times 50 = 10$  pasien

$50 - 10 = 40$  pasien

$$P(X = 1) = \left(\frac{40}{50}\right)^1 \left(1 - \frac{40}{50}\right)^{1-1}$$

$$\begin{aligned}
 &= \frac{4}{5} \\
 &= 0,8
 \end{aligned}$$

Jadi peluang keberhasilan sebesar 0,8 atau 80%.

### **Teorema**

Distribusi Bernoulli  $f(x; p)$  mempunyai nilai rata-rata dan varians

$$E(x) = \mu = p \text{ dan } Var(x) = \sigma^2 = pq = p(1 - p)$$

dimana  $p$  adalah kemungkinan sukses,  $q$  adalah gagal.

## **6.3 Distribusi Binomial**

### **Prinsip Dasar**

Distribusi Binomial adalah distribusi probabilitas dari variabel random diskrit dengan jumlah keberhasilan dalam  $n$  percobaan sukses atau gagal yang independen, dimana setiap hasil eksperimen memiliki probabilitas. Prinsip dasar dalam distribusi Binomial merupakan percobaan Bernoulli yang dilakukan secara berulang. Pada distribusi ini hanya diketahui 2 hasil misalnya sukses/gagal, sehat/sakit, setuju/tidak setuju.

### **Definisi**

Variabel random  $X$  disebut variabel random Binomial dengan parameter  $p$  dan  $n$  jika fungsi kepadatan probabilitasnya adalah

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad x = 0, 1, \dots, n$$

dimana  $0 < p < 1$  adalah kemungkinan sukses. Variabel random Binomial dengan parameter  $p$  dan  $n$  ditulis sebagai  $X \sim \text{BIN}(n; p)$ .

### **Contoh Soal**

Probabilitas seorang bayi tidak diimunisasi polio adalah 0,2. Pada suatu hari di Puskesmas 'DAMAI' ada 5 orang bayi. Hitunglah probabilitas dari bayi tersebut 2 orang belum imunisasi polio.

**Jawab :**

Berdasarkan soal dapat diketahui  $x = 2$ ,  $n = 5$ ,  $p = 0,2$  sehingga dapat ditulis  $b(2; 5; 0,2)$

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

$$\begin{aligned} P(X = 2) &= \binom{5}{2} (0,2)^2 (0,8)^3 \\ &= \frac{5!}{2! (5 - 2)!} (0,2)^2 (0,8)^3 \\ &= \frac{5 \times 4}{2 \times 1} (0,2)^2 (0,8)^3 \\ &= 10(0,04)(0,512) \\ &= 0,205 \end{aligned}$$

Jadi probabilitas 2 bayi yang belum imunisasi polio sebesar 0,205.

**Teorema**

Jika  $X$  adalah variabel random binomial dengan parameter  $p$  dan  $n$ , maka rata-rata, varians masing-masing diberikan oleh

$$\text{Rata-rata : } \mu = np$$

$$\text{Variansi : } \sigma^2 = np(1 - p)$$

**Contoh Soal**

Probabilitas seorang bayi tidak diimunisasi polio adalah 0,2. Pada suatu hari di Puskesmas 'DAMAI' ada 5 orang bayi. Hitunglah rata-rata dan variansnya.

**Jawab**

$$\text{Diketahui: } p = 0,2$$

$$\begin{aligned} \text{Rata-rata : } \mu &= np \\ &= 5(0,2) \\ &= 0,10 \end{aligned}$$

$$\begin{aligned} \text{Varians : } \sigma^2 &= np(1 - p) \\ &= 5(0,2)(1 - 0,2) \\ &= 0,008 \end{aligned}$$

Jadi rata-rata dari 5 bayi yang tidak diimunisasi sebesar 0,10 . Ini dapat diartikan bahwa sebanyak 5 bayi dengan probabilitas 0,2 setiap bayi yang tidak diimunisasi didapatkan nilai rata-rata sebesar 0,10 dari keseluruhan 5 bayi. Sedangkan variansi dari 5 bayi yang diimunisasi sebesar 0,008 . Ini dapat diartikan bahwa sebanyak 5 bayi dengan probabilitas 0,2 setiap bayi yang tidak diimunisasi didapatkan variansi sebesar 0,008 kemencengan terhadap nilai rata-rata.

## 6.4 Distribusi Poisson

### Prinsip Dasar

Ketika percobaan menghasilkan variabel random yang menyatakan banyaknya sukses dalam derah tertentu atau selama interval waktu tertentu, maka percobaan tersebut disebut percobaan Poisson. Distribusi Poisson merupakan distribusi dengan variabel random diskrit. Dalam distribusi ini nilai rata-rata dinyatakan sebagai  $\lambda$  yang pada dasarnya juga sama dengan  $\mu$ .

### Definisi

Suatu variabel random  $X$  mempunyai distribusi Poisson, jika dan hanya jika distribusinya diberikan dengan

$$f(x; \lambda) = \frac{e^{-\lambda} \lambda^x}{x!} \quad ; x = 1, 2, \dots$$

dimana  $\lambda$  adalah rata-rata kejadian sukses dari banyaknya percobaan dan  $e = 2,71828$ .

### Teorema

Distribusi Poisson  $f(x; \lambda)$  mempunyai nilai rata-rata dan variansi

$$\begin{aligned} E(x) &= \mu = \lambda \\ Var(x) &= \sigma^2 = \lambda \end{aligned}$$

### Contoh Soal

Probabilitas terjadinya shock pada saat imunisasi dengan sebuah vaksin adalah 0,07. Jika di suatu daerah jumlah orang yang akan dilakukan vaksinasi sebanyak 50. Hitunglah

probabilitas 4 orang yang akan terjadi shock dan probabilitas paling sedikit 6 orang yang mengalami shock.

### Jawab

Diketahui :

$$p = 0,07, n = 50, x = 4 \text{ dan } P(X \geq 6)$$

1. Probabilitas 4 orang yang akan terjadi shock

$$\lambda = np = 50(0,07) = 3,5$$

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

$$\begin{aligned} P(X = 4) &= \frac{2,71828^{-3,5} \times 3,5^4}{4!} \\ &= \frac{0,03 \times 150,063}{4 \times 3 \times 2 \times 1} \\ &= \frac{4,532}{24} \\ &= 0,188 \end{aligned}$$

Jadi peluang bahwa tepat 3 orang yang mengalami shock adalah 0,188 atau 18,88%.

2. Probabilitas paling sedikit 6 orang yang akan terjadi shock

$$P(X \geq 6) = 1 - P(0) - P(2) - P(3) - P(4) - P(5)$$

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0,1,2,3,4,5$$

$$P(X = 0) = \frac{2,71828^{-3,5} \times 3,5^0}{0!} = \frac{0,03 \times 1}{1} = 0,03$$

$$P(X = 1) = \frac{2,71828^{-3,5} \times 3,5^1}{1!} = \frac{0,03 \times 3,5}{1} = 0,106$$

$$P(X = 2) = \frac{2,71828^{-3,5} \times 3,5^2}{2!} = \frac{0,03 \times 12,25}{2} = 0,185$$

$$P(X = 3) = \frac{2,71828^{-3,5} \times 3,5^3}{3!} = \frac{0,03 \times 42,87}{6} = 0,216$$

$$\begin{aligned} P(X = 4) &= \frac{2,71828^{-3,5} \times 3,5^4}{4!} = \frac{0,03 \times 150,06}{24} \\ &= 0,189 \end{aligned}$$

$$P(X = 5) = \frac{2,71828^{-3,5} \times 3,5^5}{5!} = \frac{0,03 \times 525,22}{120} = 0,132$$

Maka probabilitas paling sedikit 6 orang mengalami shock

$$\begin{aligned} P(X \geq 6) &= 1 - P(0) - P(2) - P(3) - P(4) - P(5) \\ &= 1 - 0,03 - 0,106 - 0,185 - 0,216 - 0,189 \\ &\quad - 0,132 \\ &= 0,1424 \end{aligned}$$

Jadi probabilitas bahwa paling sedikit 6 orang yang shock adalah 0,1424 atau 14,24%

## 6.5 Distribusi Geometri

### Prinsip Dasar

Pada distribusi geometri merupakan distribusi dari variabel random diskrit dan mencakup kejadian khusus distribusi binomial negatif. Distribusi ini berfokus pada mencari nilai untuk  $k=1$ . Distribusi geometri dicirikan oleh probabilitas jumlah percobaan yang diperlukan untuk mencapai keberhasilan. Dengan kata lain, distribusi ini merepresentasikan eksperimen random hingga sukses yang pertama kali terjadi. Sifat-sifat distribusi geometri diantaranya

1. Percobaan dilakukan berulang
2. Diantara ulangan dalam percobaan saling bebas
3. Probabilitas sukses pada tiap ulangan sama
4. Percobaan ini dilakukan/diulang sampai diperoleh sukses pertama.

### Definisi

Bila percobaan yang saling bebas dan dilakukan secara berulang kali menghasilkan sukses dengan peluang  $p$ , gagal dengan peluang  $q = 1 - p$  maka distribusi peluang variabel random  $X$  yaitu banyaknya percobaan sampai saat terjadi sukses pertama diberikan oleh

$$f(x) = g(x; p) = pq^{x-1}$$

dimana  $p$  menunjukkan probabilitas keberhasilan dalam satu percobaan Bernoulli. Jika  $X$  memiliki distribusi Geometri, dapat ditulis sebagai  $X \sim \text{GEO}(p)$ .

### **Teorema**

Jika  $X$  adalah variabel random geometri dengan parameter  $p$  maka rata-rata, varians adalah

$$E(x) = \mu = \frac{1}{p}$$
$$\sigma^2 = \frac{(1-p)}{p^2}$$

$p$  dan  $q$  adalah parameter (probabilitas sukses dan gagal)

### **Contoh Soal**

Probabilitas seseorang sembuh dari sebuah penyakit setelah diberikan obat tertentu adalah 70%, hitunglah :

1. Berapa probabilitas pertama kali sembuh dari 6 orang yang diberi obat ?
2. Berapakah rata-rata dan variansnya

### **Jawab**

Diketahui :

$$x = 6; \quad p = 70\% = 0,7; \quad q = 1 - p = 1 - 0,7 = 0,3$$

1. Probabilitas 6 orang yang diberi obat, orang pertama yang sembuh adalah :

$$g(x; p) = pq^{x-1}$$
$$= 0,7(0,3)^{6-1}$$
$$= 0,7(0,3)^5$$
$$= 0,0017$$

Jadi probabilitas 6 orang yang diberi obat, orang pertama yang sembuh adalah 0,0017 atau 0,17%.

2. Rata-rata dan varians

$$E(x) = \frac{1}{p}$$
$$= \frac{1}{0,7}$$
$$= 1,4286, \quad \text{dan}$$
$$\sigma^2 = \frac{(1-p)}{p^2}$$
$$= \frac{0,3}{0,7^2}$$
$$= 0,6147$$

Maka rata-rata orang yang akan sembuh sebesar 1,4286 dan variansnya sebesar 0,147.

## 6.6 Distribusi Normal

### Prinsip Dasar

Distribusi normal adalah jenis distribusi variabel random kontinu. Salah satu ciri distribusi normal adalah mengikuti kurva atau grafik berbentuk lonceng. Persamaan distribusi normal mencakup fungsi densitas. Fungsi probabilitas dalam distribusi normal ini menunjukkan variabel atau penyebaran distribusi. Distribusi ini memiliki dua jenis parameter yaitu nilai rata-rata dan standar deviasi atau simpangan baku.

### Definisi

Variabel random  $X$  disebut distribusi normal jika fungsi kepadatan probabilitasnya adalah

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

dimana  $-\infty < \mu < \infty$  dan  $0 < \sigma^2 < \infty$  adalah parameter yang berubah-ubah. Jika  $X$  berdistribusi normal dengan parameter  $\mu$  dan  $\sigma^2$  maka dapat ditulis  $X \sim N(\mu, \sigma^2)$ .

### Teorema

Jika  $X$  adalah variabel random kontinu dan berdistribusi normal dengan parameter  $\mu$  dan  $\sigma^2$  maka rata-rata dan varians adalah

$$\begin{aligned} E(x) &= \mu \\ Var(x) &= \sigma^2 \end{aligned}$$

### Contoh Soal

Di sebuah Rumah Sakit dilakukan pemeriksaan kadar glukosa terhadap 100 pasien. Hasil laboratorium dari pengukuran kadar glukosa darah sewaktu-waktu diperoleh rata-rata 152 mg% dan standar deviasi 55 mg%, hitunglah probabilitas kadar glukosa antara 80 mg% dan 120 mg%.

## Jawab

Diketahui :

$$\mu = \bar{x} = 152; \quad s = \sigma = 55$$

Langkah pertama adalah mencari nilai  $P(80\% < x < 152\%)$

$$\begin{aligned} z &= \frac{x - \mu}{\sigma} \\ &= \frac{80 - 152}{55} \\ &= -1,31 \end{aligned}$$

dengan nilai  $z = -1,31$  lihat tabel  $Z = 0,0951$

Langkah kedua mencari nilai  $P(120\% < x < 152\%)$

$$\begin{aligned} z &= \frac{x - \mu}{\sigma} \\ &= \frac{120 - 152}{55} \\ &= -0,58 \end{aligned}$$

dengan nilai  $z = -0,58$  lihat tabel  $Z = 0,2810$ . Selanjutnya menghitung probabilitas  $P(80\% < x < 120\%)$

$$\begin{aligned} &= P(120\% < x < 152\%) - P(80\% < x < 152\%) \\ &= 0,2810 - 0,0951 \\ &= 0,1859 \end{aligned}$$

Jadi probabilitas 100 orang dipilih secara random memiliki glukosa antara 80 mg% dan 120 mg% adalah 0,1859 atau 18,59%.

## DAFTAR PUSTAKA

- Sugiyarto, A. (2021) *Pengantar Statistika Matematika 1*. Yogyakarta: Magnum Pustaka Utama.
- Wibowo, A. (2022) *Statistik Probabilitas*. Semarang: Yayasan Prima Agus Teknik.
- Machfoed, I. (2016) *Bio Statistika*. Yogyakarta: Fitramaya
- Sugiyono (2015) *Statistik Untuk Penelitian*. Yogyakarta: Andi Offset
- Santjaka, A. (2011) *Statistik Untuk Penelitian Kesehatan dan Kedokteran*. Jakarta: Nuha Medika.
- Riwidikdo, H. (2012) *Statistika Kesehatan : Belajar Mudah Teknik Analisa Data Dalam Penelitian Kesehatan*. Yogyakarta: Mitra Cendekia Press.



# BAB 7

## DISTRIBUSI SAMPEL

*Oleh Indah Resti Ayuni Suri*

### 7.1 Pendahuluan

Statistika atau dikenal juga dengan metode statistik mempunyai peranan yang sangat penting dalam kehidupan manusia. Bidang statistika yang semula hanya berkaitan dengan urusan pemerintahan atau kenegaraan, kini telah meluas ke bidang lain seperti pertanian, biologi, bisnis, kimia, komunikasi, ekonomi, pendidikan, elektronika, kedokteran, fisika, ilmu politik, psikologi, sosiologi, dan beberapa bidang ilmiah dan teknis lainnya (Sri Wahyuning, 2021).

Bidang ilmu statistika erat kaitannya dengan kesimpulan dan prediksi yang diperoleh dari kemungkinan hasil suatu percobaan. Jika kita ingin menarik kesimpulan dari suatu eksperimen, biasanya kita akan mengambil beberapa sampel observasi dari suatu populasi. Oleh karena itu, lebih masuk akal jika menggunakan sampel dari suatu populasi untuk membuat perkiraan/prediksi terhadap suatu populasi. Dengan menggunakan pola ini kita bisa menduga distribusi dari populasi, mean, varians, dan sifat-sifat distribusi populasi lainnya. Untuk membuat estimasi ini, pertama-tama kita perlu mendistribusikan fungsi variabel acak yang nilainya membentuk sampel, yang disebut statistik. Salah satu contoh statistik adalah mean sampel (*sample mean*) (Wawan Hafid Syaifudi and Achmad Choiruddin, 2021).

### 7.2 Pengertian Distribusi Sampling

Distribusi sampling adalah distribusi nilai rata-rata yang diambil berulang kali dari suatu populasi. Dalam populasi tak terbatas, pengambilan sampel secara acak dilakukan berulang kali hingga semua sampel yang mungkin dapat diperoleh dari populasi tersebut. Pengambilan sampel secara acak atau

random sampling adalah suatu proses pengambilan sampel yang memungkinkan setiap anggota populasi mempunyai kesempatan yang sama untuk dipilih menjadi sampel.

Sampling adalah proses pemilihan sejumlah individu atau serangkaian objek pengukuran dari suatu kelompok atau populasi dengan banyak pilihan sampel yang berbeda (Azwar Habibi, 2021). Parameter adalah nilai yang diukur dalam suatu populasi. Simbol yang digunakan:

- $\mu$  = mean populasi
- $\sigma$  = simpangan baku populasi
- $P$  = Proporsi populasi =  $X/N$
- $\mu_1 - \mu_2$  = perbedaan 2 mean populasi
- $P_1 - P_2$  = perbedaan 2 proporsi populasi

Nilai statistik adalah angka atau data yang diamati atau nilai observasi dari sampel yang menjadi dasar untuk memperkirakan nilai parameter. Simbol yang digunakan:

- $\bar{x}$  = rata-rata dari sampel
- $s$  = standar deviasi dari sampel
- $\hat{p}$  = Proporsi dari sampel, yang dapat dihitung sebagai  $\frac{x}{n}$
- $\bar{x}_1 - \bar{x}_2$  = perbedaan antara dua rata-rata sampel
- $\hat{p}_1 - \hat{p}_2$  = perbedaan 2 antara dua proporsi sampel

### 7.3 Sifat-Sifat Ditribusi Sampling

Distribusi sampel atau distribusi sampling rata-rata harga, memiliki karakteristik berikut:

Ketika sampel diambil dengan pengembalian (*with replacment*), atau ketika jumlah sampel ( $N$ ) mendekati tak terhingga, atau sangat besar dibandingkan dengan ( $n$ ) (Alifah Ratnawati, dkk, 2021).

$$\begin{aligned}\mu_{\bar{x}} &= \mu \\ \sigma_{\bar{x}} &= \sigma/\sqrt{n}\end{aligned}$$

Jika sampel diambil tanpa pengembalian (*without replacment*):

$$\mu_{\bar{x}} = \mu$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

Apabila ukuran sampel besar ( $n \geq 30$ ), distribusi sampel rata-rata harga dianggap mendekati distribusi normal. Nilai standar unitnya dijelaskan sebagai:

$$Z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Ketika ukuran sampel kecil ( $n < 30$ ), digunakan distribusi sampel harga t (*distribusi student t*) dengan derajat kebebasan (*degree of freedom*) =  $n - 1$ .

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

#### Distribusi Sampling Harga Proporsi

Jika sampel diambil dengan pengembalian (*with replacement*) atau jika ukuran populasi sangat besar dan rasio proporsi (P) tidak terlalu mendekati nol atau satu, distribusi sampel harga proporsi dianggap mendekati distribusi normal dengan karakteristik:

$$\begin{aligned}\mu_{\hat{p}} &= P \\ \sigma_{\hat{p}} &= \sqrt{\frac{P(1-P)}{n}}\end{aligned}$$

Jika pengambilan sampelnya tanpa pengembalian (*without replacement*) maka:

$$\begin{aligned}\mu_{\hat{p}} &= P \\ \sigma_{\hat{p}} &= \sqrt{\frac{P(1-P)}{n}} \sqrt{\frac{N-n}{N-1}}\end{aligned}$$

Standar unit nilai dijelaskan sebagai:

$$Z = \frac{\hat{p} - P}{\sigma_{\hat{p}}} = \frac{\hat{p} - P}{\sqrt{\frac{P(1-P)}{n}}}$$

#### Distribusi Sampling Harga Perbedaan Dua Mean

$$\mu_{(\bar{x}_1 - \bar{x}_2)} = \mu_1 - \mu_2$$

$$\sigma_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Nilai standar unit dinyatakan dengan:

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{(\bar{x}_1 - \bar{x}_2)}}$$

Jika ( $n < 30$ ):

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\left[ \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \right] \cdot \left[ \frac{1}{n_1} + \frac{1}{n_2} \right]}}$$

Distribusi Sampling Harga Perbedaan Dua Proporsi

$$\mu_{(\hat{p}_1 - \hat{p}_2)} = P_1 - P_2$$

$$\sigma_{(\hat{p}_1 - \hat{p}_2)} = \sqrt{\frac{P_1(1 - P_1)}{n_1} + \frac{P_2(1 - P_2)}{n_2}}$$

Nilai standar unit dinyatakan dengan:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (P_1 - P_2)}{\sqrt{\frac{P_1(1 - P_1)}{n_1} + \frac{P_2(1 - P_2)}{n_2}}}$$

Distribusi Sampling Harga Simpangan baku

Jika sampel diambil dengan pengembalian (*with replacement*)

$$\mu_s = \sigma$$

$$\sigma_s = \frac{\sigma}{\sqrt{2n}}$$

Jika pengembalian sampelnya tanpa pengembalian (*without replacement*)

$$\mu_s = \sigma$$

$$\sigma_s = \frac{\sigma}{\sqrt{2n}} \sqrt{\frac{N - n}{N - 1}}$$

Nilai standar unit dinyatakan dengan:

$$Z = \frac{S - \sigma}{\sigma_s}$$

### Contoh Soal

#### Distribusi Sampling rata-rata Nilai (Harga Mean)

Pada ujian mata kuliah Fisika di laboratorium, nilai rata-rata mahasiswa adalah 70 dengan standar deviasi adalah 25. Sebuah sampel 40 mahasiswa diambil. Berapa probabilitas bahwa mahasiswa tersebut:

Memiliki nilai paling tinggi 80

Memiliki nilai paling sedikit 60

Memiliki nilai paling sedikit 75

Jawab:

$$\mu = 70$$

$$\sigma = 25$$

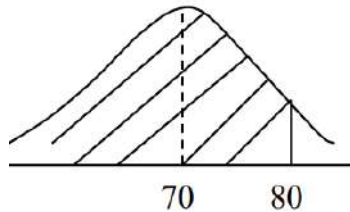
$$n = 40$$

Memiliki nilai paling tinggi 80

$$Z = \frac{x - \mu}{\sigma/\sqrt{n}} = \frac{80 - 70}{25/\sqrt{40}} = 2,5298$$

$$Z_{2,5} = 0,4943$$

Memiliki Probabilitasnya:  $0,4943 + 0,5 = 0,9943 = 99,43\%$

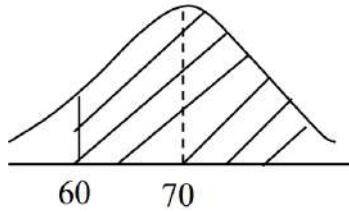


Memiliki nilai paling sedikit 60

$$Z = \frac{x - \mu}{\sigma/\sqrt{n}} = \frac{60 - 70}{25/\sqrt{40}} = -2,529$$

$$Z_{-2,529} = 0,4945$$

Probabilitasnya:  $0,4945 + 0,5 = 0,9945$

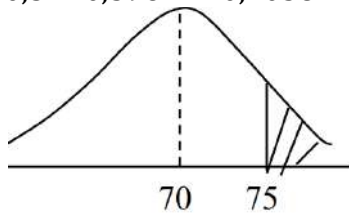


Memiliki nilai paling sedikit 75

$$Z = \frac{x - \mu}{\sigma/\sqrt{n}} = \frac{75 - 70}{25/\sqrt{40}} = 1,26$$

$$Z_{1,26} = 0,3962$$

Probabilitasnya:  $0,5 - 0,3962 = 0,1038$



Sampel acak yang terdiri dari 50 anggota diambil dari suatu populasi dengan rata-rata 32 dan simpangan baku 10. Hitunglah probabilitasnya bahwa rata-rata dari sampel ini akan berada

Antara 28 dan 35

Kurang dari 30

Lebih dari 35

Jawab:

Jawaban pertanyaan di atas didasarkan pada distribusi sampel rata-rata harga dengan karakteristik sebagai berikut:

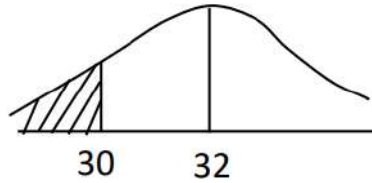
$$\mu_x = \mu \quad \sigma_x = \frac{\sigma}{\sqrt{n}}$$

$$\mu_x = 32 \quad \sigma = 10 \quad n = 50$$

$$Z_1 = \frac{x - \mu}{\sigma/\sqrt{n}} = \frac{28 - 32}{10/\sqrt{50}} = -2,828 \quad \text{luasnya} \quad 0,4977$$

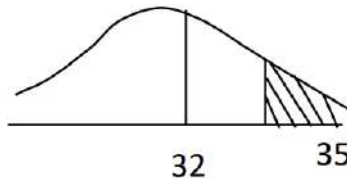
$$Z_2 = \frac{x - \mu}{\sigma/\sqrt{n}} = \frac{35 - 32}{10/\sqrt{50}} = 2,121 \quad \text{luasnya} \quad 0,4830$$

Jadi, jumlah luasnya =  $0,9807 = 98,07\%$



$$Z = \frac{x - \mu}{\sigma/\sqrt{n}} = \frac{30 - 32}{10/\sqrt{50}} = -1,41 \rightarrow \text{luas} = 0,4207$$

Maka, luas daerah atau area yang diarsir  
 $= 0,5 - 0,4207$   
 $= 0,0793$   
 $= 7,93\%$



$$Z = \frac{x - \mu}{\sigma/\sqrt{n}} = \frac{35 - 32}{10/\sqrt{50}} = 2,121 \rightarrow \text{luas} = 0,4830$$

Jadi, luas daerah yang diarsir =  $0,5 - 0,4830 = 0,017 = 1,7\%$

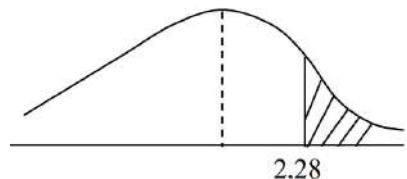
#### Distribusi Sampling Harga Proporsi

60% anak suka belajar menggambar. Berapa probabilitasnya bahwa dalam sampel acak yang terdiri dari 500 anak, lebih dari 325 anak akan berminat belajar menggambar?

Jawab:

$P(\text{menggambar}) = 0,6$

$$\begin{aligned} Q &= 1 - P \\ &= 1 - 0,6 \\ &= 0,40 \end{aligned}$$



$$\hat{P} = \frac{x}{n} = \frac{325}{500} = 0,65$$

$$Z = \frac{\frac{x}{n} - P}{\sqrt{\frac{P \cdot Q}{n}}} = \frac{0,65 - 0,6}{\sqrt{\frac{0,6 \cdot 0,4}{500}}} = 2,28$$

$$Z_{(2,28)} = 0,4887$$

Jadi probabilitasnya adalah

$$= 0,5 - 0,4887$$

$$= 0,0113$$

$$= 1,13\%$$

5% dari produk tembakau yang diproduksi oleh perusahaan mengalami kerusakan. sebanyak 100 bungkus rokok akan diambil. Berapa probabilitasnya bahwa:

Yang baik paling banyak 90 bungkus

Yang baik paling sedikit 98 bungkus

Yang baik antara 92 dan 97 bungkus

Jawab:

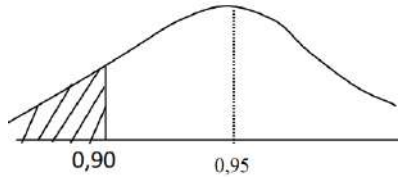
$$Q \text{ (rusak)} = 5\% = 0,05$$

$$P \text{ (baik)} = 1 - 5\% = 95\% = 0,95$$

$$N = 100$$

Yang baik paling banyak 90 bungkus

$$\frac{x}{n} = \frac{90}{100} = 0,90$$



$$Z = \frac{\frac{x}{n} - P}{\sqrt{\frac{P \cdot Q}{n}}} = \frac{\frac{90}{100} - 0,95}{\sqrt{\frac{0,95 \cdot 0,05}{100}}} = -2,50$$

$$Z_{(2,50)} = 0,4938$$

Jadi probabilitasnya:

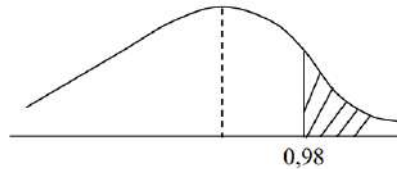
$$= 0,5 - 0,4938$$

$$= 0,0062$$

$$= 0,62\%$$

Yang baik paling sedikit 98 bungkus

$$\frac{x}{n} = \frac{98}{100} = 0,98$$



$$Z = \frac{\frac{x}{n} - P}{\sqrt{\frac{P \cdot Q}{n}}} = \frac{0,98 - 0,95}{\sqrt{\frac{0,95 \cdot 0,05}{100}}} = 1,50$$

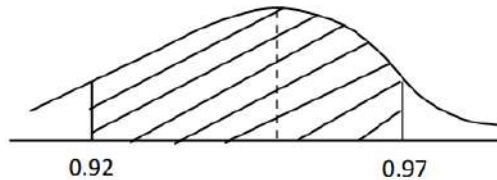
$$Z_{(1,50)} = 0,4332$$

Jadi probabilitasnya:

$$= 0,5 - 0,4332$$

$$= 0,0688$$

Yang baik antara 92 dan 97 bungkus



$$Z = \frac{\frac{x}{n} - P}{\sqrt{\frac{P \cdot Q}{n}}} = \frac{0,92 - 0,95}{\sqrt{\frac{0,95 \cdot 0,05}{100}}} = -1,50$$

$$Z_{(1,50)} = 0,4332$$

$$Z = \frac{\frac{x}{n} - P}{\sqrt{\frac{P \cdot Q}{n}}} = \frac{0,97 - 0,95}{\sqrt{\frac{0,95 \cdot 0,05}{100}}} = 1,00$$

$$Z_{(1,00)} = 0,3413$$

Jadi probabilitasnya:

$$= 0,4332 + 0,3413$$

$$= 0,7745$$

Distribusi Sampling Harga Perbedaan Dua Proporsi

Penelitian telah dilakukan mengenai pilihan ibu rumah tangga dalam menonton sebuah film telenovela. Berdasarkan hasil penelitian, 80% dari ibu rumah tangga di Kota Metro menyukai telenovela, sementara di Kotabumi sebanyak 75% menyatakan kesukaannya terhadap jenis film tersebut. Jika disetiap wilayah diambil sampel secara acak sebanyak 500 ibu rumah tangga. Berapa probabilitasnya bahwa:

Di kota Metro, tingkat kesukaan terhadap telenovela lebih tinggi maksimal 4% dibandingkan dengan Kotabumi.

Di kota Metro, tingkat kesukaan terhadap telenovela lebih rendah 6% dibandingkan dengan Kotabumi.

Jawab:

$P(\text{Metro}) = 80\%$  atau 0,8

$P(\text{Kotabumi}) = 75\%$  atau 0,75

$n = 500$

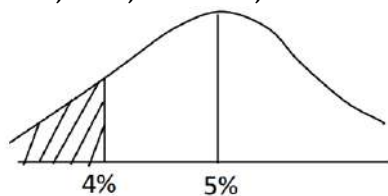
Di kota Metro yang menyukai film telenovela paling banyak 4% lebih besar dari pada Kotabumi.

$$Z = \frac{4\% - (80\% - 75\%)}{\sqrt{\frac{80\% \cdot 20\%}{500} + \frac{75\% \cdot 25\%}{500}}} = \frac{4\% - 5\%}{\sqrt{0,00032 + 0,000375}}$$

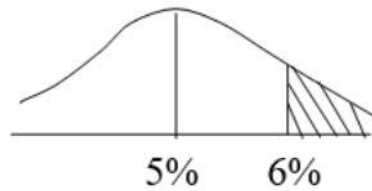
$$= \frac{-0,01}{0,0264} = -0,387 \approx -0,38$$

$$Z_{(-0,38)} = 0,1480$$

$$\text{Probabilitasnya} = 0,5 - 0,1480 = 0,352$$



Di kota Metro, tingkat kesukaan terhadap telenovela paling rendah 6% lebih tinggi dibandingkan dengan Kotabumi.



$$Z = \frac{6\% - 5\%}{\sqrt{0,00032 + 0,000375}} = \frac{0,01}{0,0264} = 0,38$$

$$Z_{(0,38)} = 0,1480$$

$$\text{Probabilitasnya} = 0,5 - 0,1480 = 0,352$$

## 7.4 Kesimpulan

Sampel adalah sebagian dari anggota populasi yang dipilih dengan cara tertentu yang akan diteliti sifat-sifatnya dalam penelitian. Nilai-nilai yang berasal dari data sampel dinamakan dengan Statistik. Distribusi sampling adalah distribusi probabilitas dan suatu statistik tergantung dari ukuran populasi, ukuran sampel, dan metode dalam memilih sampel.

Distribusi sampling adalah distribusi dari mean yang diambil secara berulang kali dari suatu populasi. Bila pada suatu populasi tak terhingga dilakukan pengambilan sampel secara acak berulang-ulang hingga semua sampel yang mungkin dapat ditarik dari populasi tersebut. Random Sampling atau sampling secara acak adalah suatu proses pengambilan sampai dimana setiap anggota populasi memiliki kesempatan yang sama untuk terpilih sebagai sampel. Kemudian, ada beberapa sifat pada distribusi sampling antara lain:

1. Distribusi sampling harga mean
2. Distribusi Sampling Harga Proporsi
3. Distribusi Sampling Harga Perbedaan Dua Mean
4. Distribusi Sampling Harga Perbedaan Dua Proporsi
5. Distribusi Sampling Harga Simpangan baku

## DAFTAR PUSTAKA

- Habibi, Azwar. *Buku Ajar Statistika Bisnis Teori Dan Aplikasi Program R*. Pamekasan: Duta Media Publishing, 2021.
- Ratnawati, Alifah, Chrisna Suhendi, and Mulyana. *Statistik Bisnis*. Purwokerto: CV. IRDH, 2021.
- Syaifudi, Wawan Hafid, and Achmad Choiruddin. *Pengantar Teori Probabilitas Dan Statistika*. Bengkulu: Elmarkazi, 2021.
- Wahyuning, Sri. *Dasar-Dasar Statistik*. Semarang: Yayasan Prima Agus Teknik, 2021.

# **BAB 8**

## **PENGUJIAN HIPOTESIS**

*Oleh Iya Purnama Sari*

### **8.1 Pengertian Pengujian Hipotesis**

Pengujian hipotesis merupakan tahap akhir prosedur pengambilan keputusan secara statistika, yaitu kesimpulan atau inferensia terhadap populasi berdasarkan informasi contoh. Secara alami, keputusan yang diambil dapat benar atau dapat keliru. Pengujian hipotesis adalah untuk membantu peneliti menjawab pertanyaan terkait dengan permasalahan untuk menolak atau menerima suatu hipotesis.

Secara pasti kebenaran dari suatu hipotesis tidak dapat ditentukan, kecuali dilakukan pengamatan terhadap seluruh anggota populasi. Melakukan hal ini tentunya sangat tidak efisien, mengingat besarnya ukuran populasi. Hal ini membutuhkan penarikan sejumlah sampel secara acak dari populasi, kemudian diamati karakteristiknya untuk dibandingkan dengan hipotesis yang dibuat atau diajukan (Anuraga et al., 2021). Apabila sampel acak ini memberikan indikasi yang dapat mendukung hipotesis yang dibuat atau diajukan tersebut maka hipotesis tersebut dapat diterima, sedangkan jika sampel tersebut menghasilkan kesimpulan yang berbeda terhadap hipotesis yang diajukan, maka hipotesis tersebut ditolak (Taeger and Kuhnt, 2014).

Pengujian hipotesis menjadi salah satu bagian prosedur penting dalam biostatistik. Beberapa pengujian yang dilakukan oleh peneliti yaitu untuk membuktikan apakah asumsi awal atau pendapat yang diajukan atau dipercaya peneliti tersebut dapat diterima (benar) atau ditolak. Pengujian hipotesis dapat membantu peneliti untuk membuktikan permasalahan tersebut yang apakah benar-benar terbukti menjadi fakta atau tidak dan hanya sekedar teori. Berikut beberapa contoh keputusan yang

dapat diambil dari berbagai situasi atau keadaan (Satyawana and Sari, 2023).

**Tabel 8.3.** Cara membuat keputusan dari contoh situasi yang ada

| Keputusan         | Situasi                       |                                    |
|-------------------|-------------------------------|------------------------------------|
|                   | Tidak Aman                    | Aman                               |
| Tunggu            | Benar                         | Tidak Benar                        |
| Menyeberang       | Tidak Benar                   | Benar                              |
| Keputusan         | Situasi                       |                                    |
|                   | Proses terkendali             | ProsesTidak Terkendali             |
| Teruskan Produksi | Benar                         | Error (Tipe II)                    |
| Hentikan Produksi | Error (Tipe I)                | Benar                              |
| Keputusan         | Situasi                       |                                    |
|                   | $H_0$ Benar atau $H_1$ Keliru | $H_0$ Tidak Benar atau $H_1$ Benar |
| Gagal Tolak $H_0$ | Benar                         | $\beta$                            |
| Tolak $H_1$       | $\alpha$                      | Benar                              |

Keterangan:

- $\alpha$ : peluang menolak  $H_0$  padahal  $H_0$  benar =  $P \{Tolak H_0|H_0 \text{ Benar}\}$ ;
- $1-\alpha$ : tingkat kepercayaan penolakan  $H_0$ ;
- $\beta$ : peluang menolak  $H_1$  padahal  $H_1$  benar =  $P \{Tolak H_1|H_1 \text{ Benar}\}$ ;
- $1-\beta$ : kuasa Uji (Kriteria sahih tidaknya prosedur pengujian).

## 8.2 Langkah Pengujian Hipotesis

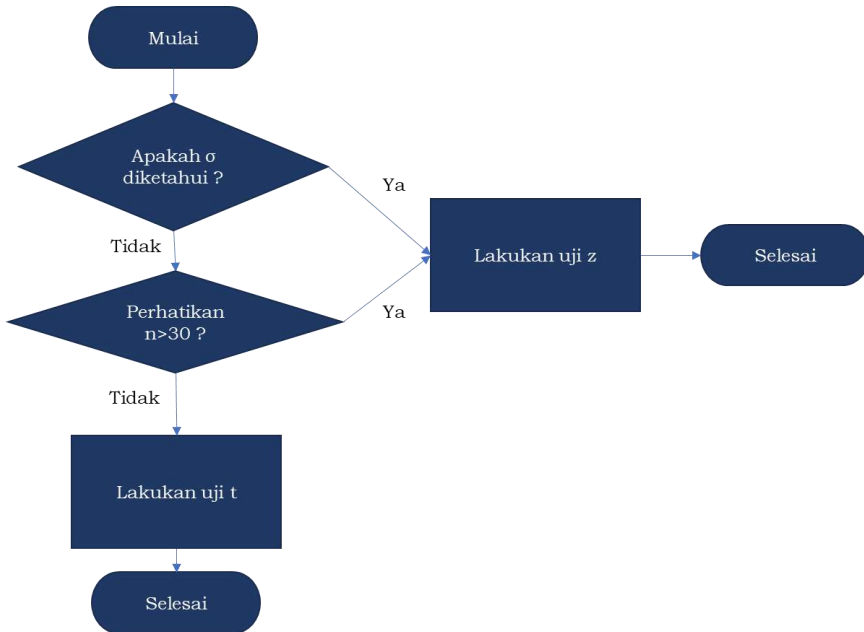
Tahapan yang perlu diperhatikan dalam pengujian hipotesis (Gambar 8.1), yaitu:

1. Pertama tentukan  $H_0$  dan  $H_1$ . Apabila tujuan pengujian hipotesis untuk menghitung rata-rata populasi sama dengan suatu nilai tertentu (misal  $\mu_0$ ), maka hipotesis "*sama dengan* ( $=$ )" selalu dinyatakan sebagai  $H_0$  dan hipotesis "*tidak sama dengan* ( $\neq$ )" selalu dinyatakan sebagai  $H_1$ .
2. Berikan batasan terhadap statistik uji disesuaikan dengan data contoh. Pengujian rata-rata populasi untuk ukuran contoh  $n$  yang besar ( $n \geq 30$ ), maka dapat memakai uji  $z$ , sedangkan untuk ukuran contoh  $n$  yang kecil ( $n < 30$ ) dapat menggunakan uji  $t$ , masing-masing:

$$z_{hit} = \frac{\bar{x} - \mu_0}{s_{\bar{x}}} \quad \text{dan} \quad t_{hit} = \frac{\bar{x} - \mu_0}{s_{\bar{x}}} \quad \text{dengan} \quad s_{\bar{x}} = \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \quad \text{untuk} \quad \frac{n}{N} > 0,05$$

$$\frac{N-n}{N-1} = 1 \quad \text{untuk } N \text{ tidak diketahui.}$$

3. Buat batasan untuk rentang daerah penolakan hipotesis  $H_0$  berdasarkan taraf nyata  $\alpha$  yang ditentukan (umumnya yaitu 5% atau 0,05). Apabila nilai hasil uji statistik berada pada daerah ini maka tolak  $H_0$  (Gambar 2).
4. Lakukan perhitungan uji statistik uji, mengikuti langkah 2 lalu bandingkan dengan titik kritis pada langkah 3 kemudian ambil keputusan apakah tolak  $H_0$  atau gagal tolak  $H_0$ .
5. Buatlah kesimpulan atau interpretasi berdasarkan hasil uji sesuaikan dengan permasalahan yang diteliti. Pemberian pernyataan di bagian kesimpulan yang baik harus dapat memberikan ringkasan dari hasil analisis data dan terbebas dari bahasa statistika.



**Gambar 8.4.** Langkah dalam pemilihan uji hipotesis yang akan digunakan



**Gambar 8.5.** Batasan dalam penerimaan atau penolakan hipotesis

### 8.3 Rumus dalam Pengujian Hipotesis

Beberapa rumus dalam pengujian hipotesis terhadap rata-rata populasi (n besar), yaitu:

1. Uji Dua Arah :  $H_0 : \mu = \mu_0$  Vs  $H_1 : \mu \neq \mu_0$

Tolak  $H_0$  jika:  $|z_{hit}| > z_{\alpha/2}$  sedangkan  $z_{hit} = \frac{\bar{x} - \mu_0}{s_{\bar{x}}}$   
( $z_{\alpha/2} = 1,96$  untuk  $\alpha = 0,05$ )

2. Uji Satu Arah:  $H_0 : \mu = \mu_0$  Vs  $H_1 : \mu > \mu_0$  atau  $H_0 : \mu = \mu_0$  Vs  $H_1 : \mu < \mu_0$

Tolak  $H_0$  jika:  $z_{hit} > z_{\alpha}$  atau  $z_{hit} < -z_{\alpha}$  ( $z_{\alpha} = 1,64$  untuk  $\alpha = 0,05$ )

3. Populasi terhingga dengan  $\frac{n}{N} > 0,05$ ,  $s_{\bar{x}} = \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$  dan  $s_{\bar{x}} = \frac{s}{\sqrt{n}}$  untuk kasus ukuran populasi lainnya.

Beberapa rumus dalam pengujian hipotesis terhadap rata-rata populasi (n kecil), yaitu:

1. Uji Dua Arah :  $H_0 : \mu = \mu_0$  Vs  $H_1 : \mu \neq \mu_0$

Tolak  $H_0$  jika:  $|t_{hit}| > t_{\alpha/2}$  sedangkan  $t_{hit} = \frac{\bar{x} - \mu_0}{s_{\bar{x}}}$  (Untuk nilai  $t_{\alpha/2, n-1}$  dapat dilihat pada tabel t dengan  $df = n - 1$ )

2. Uji Satu Arah:  $H_0 : \mu = \mu_0$  Vs  $H_1 : \mu > \mu_0$  atau  $H_0 : \mu = \mu_0$  Vs  $H_1 : \mu < \mu_0$

Tolak  $H_0$  jika:  $t_{hit} > t_{\alpha, n-1}$  atau  $t_{hit} < -(t_{\alpha, n-1})$   
(Untuk nilai  $t_{\alpha, n-1}$  dapat dilihat pada tabel t dengan  $df = n - 1$ )

3. Populasi terhingga dengan  $\frac{n}{N} > 0,05$ ,  $s_{\bar{x}} = \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$  dan  $s_{\bar{x}} = \frac{s}{\sqrt{n}}$  untuk kasus ukuran populasi lainnya.

#### Perbedaan pengujian hipotesis dua arah dan satu arah

Perusahaan yang bergerak dalam bidang pertanian membeli bibit tanaman dalam jumlah yang besar. Bibit disimpan didalam kotak khusus yang masing-masing berisikan 20 bibit. Berdasarkan pengalaman pembelian sebelumnya diketahui selalu ditemukan kondisi bibit yang kurang bagus

dalam setiap pengiriman ke perusahaan tersebut. Hal ini menyebabkan penjual bibit tanaman tersebut akan memberikan diskon 12% setiap kotaknya (1 dari 20 bibit dalam kotak tidak harus dibayar). Agar pemeriksaan tidak harus dilakukan secara berulang terhadap kotak bibit tersebut oleh perusahaan, maka perusahaan melakukan pengambilan contoh acak sebanyak 200 kotak. Dari rata-rata bibit yang kurang bagus pada setiap kotak dan simpangan bakunya, perusahaan memerlukan kriteria untuk memutuskan, yaitu:

1. Apakah rata-rata bibit yang kurang bagus mengalami perubahan;
2. Apakah kualitas bibit yang dikirimkan semakin memburuk; atau
3. Apakah kualitas bibit yang dikirimkan semakin membaik.

Hipotesis :

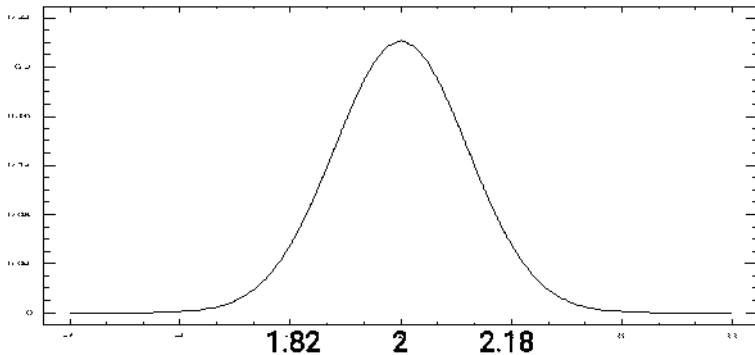
- a)  $H_0: \mu = 2$  vs  $H_1: \mu \neq 2$  (hipotesis dua arah)
- b)  $H_0: \mu = 2$  vs  $H_1: \mu > 2$  (hipotesis satu arah)
- c)  $H_0: \mu = 2$  vs  $H_1: \mu < 2$  (hipotesis satu arah)

Jawab :

Dari contoh berukuran  $n = 200$  diperoleh  $\bar{x} = 1,35$  dan  $s = 1,50$ .  
Jika  $H_0$  benar, maka 95% rata-rata contoh haruslah terletak antara  $\mu + 1,96 \frac{s}{\sqrt{n}}$  dan  $\mu - 1,96 \frac{s}{\sqrt{n}}$  atau antara 1,82 dan 2,18.

Dengan arti kata, semua  $\bar{x} < 1,82$  atau  $\bar{x} > 2,18$  akan menyebabkan  $H_0$  ditolak, sedangkan semua  $1,82 < \bar{x} < 2,18$  akan menyebabkan  $H_0$  gagal ditolak (harus diterima).

Dari contoh berukuran  $n = 200$  diperoleh  $\bar{x} = 1,35$  dan  $s = 1,50$



**Gambar 8.6.** Daerah penerimaan dan penolakan  $H_0$

Oleh karena  $\bar{x} = 2$  atau  $\bar{x} > 2,18$  maka  $H_0$  harus ditolak.

### Hipotesis Dua Arah

$H_0 : \mu=1$  vs  $H_1 : \mu \neq 1$  dapat diuji untuk menentukan rata-rata kertas yang rusak mengalami perubahan.

Hitung statistik uji:  $z_{hit} = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{1,35 - 1,00}{1,50/\sqrt{200}} = \frac{0,35}{0,11} = 3,30$

Tentukan taraf nyata  $\alpha=0,05(5\%)$  dan  $z_{\alpha/2} = z_{0,025} = 1,96$  (hipotesis dua arah)

Oleh karena  $z_{hit} = 3,30 > z_{0,025} = 1,96$  , maka Tolak  $H_0$

Kesimpulan atau Interpretasi:

Informasi yang diperoleh dari contoh dapat digunakan untuk menggugurkan hipotesis yang menyatakan bahwa hanya 1 dari 20 bibit tanama yang kurang bagus dalam kotak tersebut.

Oleh karena hipotesis yang diuji adalah hipotesis dua arah, hipotesis ini tidak dapat digunakan untuk menyatakan bahwa kualitas pengiriman telah membaik atau memburuk.

### Hipotesis Satu Arah

Dari contoh berukuran  $n = 200$  diperoleh  $\bar{x} = 1,35$  dan  $s=1,50$

$H_0 : \mu=1$  vs  $H_1 : \mu > 1$  dapat diuji untuk menentukan apakah bibit yang dikirimkan semakin membaik atau justru memburuk.

Hitung statistik uji:  $z_{hit} = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{1,35 - 1,00}{1,50/\sqrt{200}} = \frac{0,35}{0,11} = 3,30$

Tentukan taraf nyata  $\alpha = 0,05$  (5%) dan  $z_{\alpha} = z_{0,05} = 1,64$  (hipotesis satu arah)

Oleh karena  $z_{hit} = 3,30 > z_{0,05} = 1,64$ , maka Tolak  $H_0$

Kesimpulan atau Interpretasi:

Informasi yang diperoleh yaitu berdasarkan SK 95% dipercaya bahwa kualitas pengiriman memburuk berdasarkan uji satu arah.

## 8.4 Contoh Soal Pengujian Hipotesis

Beberapa contoh kasus dalam pengujian hipotesis, adalah sebagai berikut:

### Contoh 1: Uji z

Salah satu perusahaan yang bergerak dibidang pengolahan ikan mempekerjakan tenaga kerja menerima keluhan dari pekerjanya. Keluhan terkait bahwa berdasarkan 50 pekerja yang dipilih secara acak, diperoleh kesimpulan bahwa pekerja di perusahaan tersebut ternyata telah dibayar secara rata-rata kurang dari Rp 50.000.- per harinya. Tentukan proses pengambilan keputusan yang dilakukan para pekerja tersebut.

Hipotesis yang akan diuji adalah:

$H_0 : \mu = 50.000$  vs  $H_1 : \mu < 50.000$  (hipotesis satu arah)

Dari contoh berukuran  $n=50$  diperoleh  $\bar{x} = 45.000$  dan  $s=7.000$

Diperoleh Hitung statistik uji:  $z_{hit} = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{45.000 - 50.000}{7.000/\sqrt{50}} =$

$$\frac{-5.000}{989,95} = -5,05$$

Oleh karena  $z_{hit} = -5,05 < z_{0,05} = -1,64$ , maka Tolak  $H_0$

Kesimpulan atau Interpretasi:

Keluhan dari para pekerja terbukti memiliki fakta yang sangat kuat untuk menyatakan bahwa para pekerja di perusahaan pengolahan ikan tersebut telah dibayar kurang dari Rp 50.000.- perharinya.

### Contoh 2: Uji z

Survei pendahuluan dari seorang peneliti menunjukkan bahwa penduduk miskin perkotaan rata-rata memiliki 5 orang anak. Membuktikan hal tersebut maka perlu dilakukan pengambilan contoh terhadap 100 orang penduduk miskin di perkotaan dan diperoleh rata-rata penduduk tersebut memiliki anak 6 orang. Tentukanlah apakah pernyataan di atas dapat diterima secara statistika (Berdasarkan penelitian terdahulu diketahui  $\sigma = 0,9$ ).

Hipotesis yang akan diuji adalah:  $H_0 : \mu=5$  vs  $H_1 : \mu \neq 5$  (hipotesis dua arah)

Dari contoh berukuran  $n=100$  diperoleh  $\bar{x} = 6$  dan  $\sigma = 0,9$

Diperoleh Hitung statistik uji: 
$$z_{hit} = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{6-5}{0,9 / \sqrt{100}} = \frac{1}{0,09} =$$

11,11

Oleh karena  $z_{hit} = 11,11 > z_{0,025} = 1,96$ , maka Tolak  $H_0$

Kesimpulan atau interpretasi:

Hasil survei pendahuluan yang menyatakan bahwa penduduk miskin perkotaan rata-rata memiliki 5 orang anak tidak dapat dibuktikan.

### Contoh 3: Uji z

Data dari dinas pengelolaan pasar menunjukkan bahwa terjadi peningkatan jumlah pedagang kaki lima dari tahun ke tahun rata-rata 43% setiap tahunnya. Seorang peneliti ingin menguji pernyataan tersebut dan diambillah data terkait jumlah pedagang kaki lima dari beberapa pasar, sehingga terkumpul sampel berukuran  $n=73$  (pasar) dan diperoleh rata-rata peningkatan jumlah pedagang kaki lima mencapai 44,56% setiap tahunnya (Berdasarkan penelitian terdahulu diketahui  $\sigma = 15$ ).

Hipotesis yang akan diuji adalah:  $H_0 : \mu \leq 43$  vs  $H_1 : \mu > 43$  (hipotesis satu arah)

Dari contoh berukuran  $n=73$  diperoleh  $\bar{x} = 44,56$  dan  $\sigma = 15$

Diperoleh Hitung statistik uji:  $z_{hit} = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{44,56 - 43}{15 / \sqrt{73}} = \frac{1,56}{1,76} =$

0,89

Oleh karena  $z_{hit} = 0,89 < z_{0,1} = 1,28$ , maka Gagal Tolak  $H_0$  ( $\alpha = 10\%$ )

Kesimpulan atau interpretasi:

Tidak cukup fakta untuk membuktikan bahwa rata-rata jumlah pedagang kaki lima meningkat.

#### Contoh 4: Uji t

Seorang peneliti menyatakan bahwa rata-rata kepadatan ikan nila yang terbaik dengan sistem resirkulasi dengan filter zeolit serta hewan uji berupa benih ikan nila untuk ukuran 4-6 cm adalah dengan padat tebar rata - rata 10 ekor ikan. Suatu contoh acak berukuran  $n=5$  hasil penelitian lainnya yang terkait menghasilkan rata-rata jumlah padat tebar ikan nila dengan kriteria yang sama yaitu 12 ekor ikan dengan simpangan baku 5. Ujilah apakah pernyataan tersebut beralasan disertai dengan interpretasi.

Hipotesis yang akan diuji adalah:

$H_0 : \mu=10$  vs  $H_1 : \mu \neq 10$  (hipotesis dua arah)

Dari contoh berukuran  $n=5$  diperoleh  $\bar{x} = 12$  dan  $s = 5$

Diperoleh Hitung statistik uji:  $t_{hit} = \frac{\bar{x} - \mu}{s / \sqrt{n}} = \frac{12 - 10}{5 / \sqrt{5}} = \frac{2}{2,24} =$

0,90

Oleh karena  $t_{hit} = 0,90 < t_{\frac{0,05}{2}, 25-1} = 2,064$ , Gagal Tolak  $H_0$

Kesimpulan/Interpretasi:

Terdapat cukup bukti bahwa kepadatan ikan nila yang terbaik dengan sistem resirkulasi dengan filter zeolit serta hewan uji berupa benih ikan nila untuk ukuran 4-6 cm adalah dengan padat tebar rata - rata 10 ekor ikan.

#### Contoh 5: Pengujian hipotesis untuk data berpasangan

Direktur dari suatu perusahaan menginginkan adanya perubahan kinerja dari karyawannya. Target ini dapat tercapai

dengan memberikan pelatihan untuk menambah skill karyawannya sehingga berdampak terhadap perubahan kinerjanya. Perusahaan tersebut mengirimkan karyawannya untuk melaksanakan pelatihan di suatu lembaga pelatihan terkait. Berdasarkan hal tersebut maka ujilah apakah pelatihan tersebut efektif sehingga berdampak terhadap kinerja karyawan perusahaan tersebut?

Untuk itu peneliti mengambil sampel acak terhadap 8 orang karyawan terkait dengan kriteria tingkat keberhasilan dalam menangani proyek kerjanya diatas 50% diberikan skor 1 sisanya 0. Penilaian atau skoring diberikan terhadap proyek yang dikerjakan oleh karyawan tersebut sebelum dan setelah mereka melaksanakan pelatihan. Hasil penilaian dan skoring terhadap 8 orang karyawan tersebut sebagai berikut:

**Table 8.4.** Hasil penilaian atau skoring terhadap 8 karyawan sebelum dan setelah pelatihan

| Karyawan          | 1 | 2  | 3 | 4 | 5 | 6 | 7 | 8 |
|-------------------|---|----|---|---|---|---|---|---|
| Sebelum pelatihan | 0 | 1  | 1 | 1 | 0 | 0 | 0 | 0 |
| Setelah pelatihan | 1 | 0  | 1 | 0 | 1 | 1 | 1 | 1 |
| Selisih ( $D_i$ ) | 1 | -1 | 0 | 1 | 1 | 1 | 1 | 1 |

Pelatihan tersebut apakah telah merubah kinerja karyawan setelah melaksanakan pelatihan di lembaga terkait, ujilah pada taraf nyata 1%?

Hipotesis yang akan diuji adalah:

$H_0 : \mu_D = 0$  vs  $H_1 : \mu_D \neq 0$  (hipotesis dua arah)

Diketahui:

$$\bar{D} = \frac{\sum D_i}{n} = \frac{5}{8} = 0,63$$

$$\begin{aligned}
 S_D &= \sqrt{\frac{\sum (D_i - \bar{D})^2}{n-1}} \\
 &= \sqrt{\frac{((1-0,63)^2 + (-1-0,63)^2 + (0-0,63)^2 + (1-0,63)^2 + (1-0,63)^2 + (1-0,63)^2 + (0-0,63)^2 + (1-0,63)^2)}{8-1}} \\
 &= 0,74
 \end{aligned}$$

Dari contoh berukuran  $n=8$  diperoleh  $\bar{D} = 0,63$  dan  $S_D = 0,55$

Diperoleh Hitung statistik uji:  $t_{hit} = \frac{\bar{D} - \mu_D}{S_D / \sqrt{n}} = \frac{0,63 - 0}{0,74 / \sqrt{8}} = 2,41$

Karena  $t_{hit} = 2,41 < t_{0,01/2} = 4,604$ , Gagal Tolak  $H_0$

Kesimpulan/Interpretasi:

Setelah melaksanakan pelatihan di lembaga terkait, tidak terdapat perubahan terhadap kinerja karyawan.

## DAFTAR PUSTAKA

- Anuraga, G., Indrasetianingsih, A., Athoillah, M., 2021. Pelatihan pengujian hipotesis statistika dasar dengan software r. BUDIMAS: Jurnal Pengabdian Masyarakat 3, 327–334.
- Satyawan, N.M., Sari, I.P., 2023. Buku Ajar berbasis Keunggulan Lokal STATISTIKA TERAPAN Teori dan Aplikasi pada Bidang Perikanan Tangkap. Deepublish Publisher, Yogyakarta.
- Taeger, D., Kuhnt, S., 2014. Statistical hypothesis testing with SAS and R. John Wiley & Sons.



## BIODATA PENULIS



### **Neny Rochyani**

Dosen Program Studi Teknik Kimia  
Fakultas Teknik Universitas PGRI Palembang

Penulis lahir di Palembang tanggal 09 Juni, menikah dengan Gonan Sumadi, dikaruniai 2 orang putra Ahmad Fadlan Abdullah latif Al-Ghozali (alm.) dan Muhammad Aufar Nurhasyim Ar-rambani. Menyelesaikan pendidikan S1, S2 pada Jurusan Teknik kimia dan melanjutkan S3 pada program studi ilmu-ilmu lingkungan dengan BKU : Agro Industri Energi di UNSRI-Palembang, pendidikan profesi keinsinyuran (Ir.) pada Unika Atmajaya Jakarta dan memegang registrasi sebagai Insinyur Profesional Madya (IPM) serta ASEAN Eng. Pada bidang penulisan karya ilmiah, sampai dengan saat ini telah menghasilkan karya ilmiah utamanya penelitian yang berhubungan dengan materi ilmu kimia dan lingkungan dengan kajian utama berfokus terhadap restorasi dan pendalaman karakteristik lingkungan sebagai objek penelitian yang menjadi kekuatan pengelolaan lingkungan. Dalam ranah penulisan buku, sampai dengan saat ini telah menghasilkan buku materi ajar dan modul paktikum yang dipakai dan diaplikasikan pada lingkungan perguruan tinggi.

## BIODATA PENULIS



### **Iha Nursolihah, S.S.T., M.K.M.**

Dosen Program Studi Administrasi Rumah Sakit  
Fakultas Ilmu Kesehatan Universitas Singaperbangsa Karawang

Penulis lahir di Karawang tanggal 23 Oktober 1991. Tahun 2019, penulis menyelesaikan Strata2 (S2) pada peminatan Biostatistika dan Kependudukan, Fakultas Kesehatan Masyarakat Universitas Indonesia. Setelah menamatkan S2, penulis bekerja di Rumah Sakit Anak dan Bunda Harapan Kita, sebagai Analis Data dan Informasi Rawat Inap pada Substansi Pelayanan Medik. Sejak Maret 2022, Penulis menjadi salah satu dosen tetap PNS (Pegawai Negeri Sipil) di Fakultas Ilmu Kesehatan, Universitas Singaperbangsa Karawang. Penulis merupakan dosen pada program studi Administrasi Rumah Sakit, beberapa mata kuliah yang diampu diantaranya Sistem Informasi Manajemen Rumah Sakit, Biostatistik dan Statistika Kesehatan, Manajemen dan Analisis Data Rumah Sakit, dan mata kuliah lainnya.

## **BIODATA PENULIS**



**Nur Al-faida, SKM., M. Kes**  
Dosen Program Studi Sarjana Gizi  
STIKes Persada Nabire Papua Tengah

Penulis lahir di Sumulluk tanggal 25 September 1993. Penulis adalah dosen tetap pada Program Studi Sarjana Gizi STIKes Persada Nabire sejak tahun 2020. Tahun 2020-2025 penulis mendapat amanah sebagai Ketua Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM) di Sekolah Tinggi Ilmu Kesehatan Persada Nabire (STIKPEN).

Penulis menyelesaikan pendidikan S1 pada Fakultas Kesehatan dengan Jurusan Kesehatan Masyarakat Konsentrasi Epidemiologi di Universitas Muhammadiyah Parepare pada tahun 2016 dan melanjutkan S2 pada Jurusan Kesehatan Masyarakat Konsentrasi Epidemiologi lulus tahun 2018 di Universitas Muslim Indonesia.

Penulis menjadi pengampu pada mata kuliah: Matematika, Dasar Sosiologi, Fisiologi, Ilmu Bahan Makanan, Biostatistik, Gizi Ibu dan Anak 1, Antropologi Sosial Gizi, Epidemiologi Gizi, Teknologi Pangan Gizi, Metodologi Penelitian Gizi, dan Inovasi Ilmu Pangan Lokal.

## BIODATA PENULIS



### **Drs. Bambang Suhartawan, M.MT**

Dosen Jurusan Teknik Lingkungan  
Fakultas Teknik Sipil dan Perencanaan  
Universitas Sains dan Teknologi Jayapura

Penulis lahir di Blitar Tahun 1964. Penulis adalah dosen tetap dengan status PNS-DPK pada Jurusan Teknik Lingkungan Fakultas Teknik Sipil dan Perencanaan (FTSP) Universitas Sains dan Teknologi Jayapura. Menyelesaikan pendidikan S1 pada Program Studi Kimia Jurusan Matematika dan Ilmu Pengetahuan Alam (MIPA) Fakultas Keguruan dan Ilmu Pendidikan Universitas Cenderawasih Jayapura pada tahun 1989. Pendidikan Pasca Sarjana diperoleh dari Institut Teknologi Sepuluh Nopember (ITS) Surabaya jurusan Manajemen Teknik Lingkungan pada tahun 2002. Beberapa buku yang telah ditulis terkait bidang lingkungan antara lain : Pengelolaan Sumber Daya Air; Penyehatan Udara; Adaptasi dan Mitigasi Perubahan Iklim di Indonesia; Ekotoksikologi, Pengelolaan Limbah Industri; Sistem Lingkungan Industri; Analisis Kualitas Lingkungan; Kesehatan Lingkungan Bancana; Manajemen Penyakit Tripos Berbasis Lingkungan; Pengelolaan Limbah Padat, Limbah Industri dan B3; Lingkungan : Pemahaman Masalah dan Solusi; Kesehatan Lingkungan : Memahami Dampak Lingkungan Terhadap Kesehatan; Kimia Ramah Lingkungan; Pencemaran Udara dan Perubahan Iklim; Entomologi dalam Kesehatan

Lingkungan; Pengelolaan Limbah Cair; Kimia Lingkungan; Kesehatan Lingkungan Global; Pengenalan Pembangunan Berkelanjutan dan Memahami Konsep dan Teori Ekologi Lingkungan;

## **BIODATA PENULIS**



**Khairul Alim, S.Si., M.Sc.**

Dosen Program Studi Matematika  
Fakultas Sains dan Teknologi Universitas Jambi

Penulis bernama lengkap Khairul Alim, S.Si., M.Sc. dan lahir di Kuala Tungkal tanggal 22 Februari 1994. Penulis adalah dosen tetap pada Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Jambi. Menyelesaikan pendidikan S1 Matematika dan S2 Magister Matematika, Universitas Gadjah Mada. Penulis menekuni keilmuan Matematika Aktuaria.

Ini merupakan pengalaman pertama penulis dalam menulis buku. Semoga buku ini dapat bermanfaat bagi pembaca.

## **BIODATA PENULIS**



**Samsul Bahri Loklomin**  
Program Studi Statistika  
Fakultas MIPA Universitas Pattimura

Penulis lahir di Seram Bagian Timur, Provinsi Maluku pada 07 Desember 1992. Penulis adalah dosen tetap pada Program Studi Statistika Fakultas MIPA Universitas Pattimura. Menyelesaikan pendidikan S1 pada Jurusan Matematika Fakultas MIPA Universitas Pattimura dan melanjutkan S2 pada Jurusan Statistika Fakultas MIPA Institut Teknologi Sepuluh November – Surabaya. Beberapa mata kuliah yang pernah diampu penulis diantaranya; Statistika Dasar, Statistika Pendidikan, Statistika Sosial, Teori Probabilitas, Statistika Matematika, Analisis Data Survival, Algoritma dan Pemrograman, Biostatistika, Statistika Bayesian, Pengendalian Kualitas Statistika, Pengantar Teori Keputusan dan Statistika Komputasi.

## **BIODATA PENULIS**



**Indah Resti Ayuni Suri, S.Si., M.Si.**

Dosen Program Studi Pendidikan Matematika  
Fakultas Tarbiyah dan Keguruan UIN Raden Intan Lampung

Penulis lahir di Kotagajah Lampung Tengah tanggal 30 Maret 1988. Penulis adalah dosen program studi pendidikan matematika fakultas tarbiyah dan keguruan UIN Raden Intan Lampung. Menyelesaikan pendidikan S1 di UIN Malang Jurusan Matematika dan melanjutkan S2 di UB Malang Jurusan Matematika. Penulis menekuni bidang Menulis.

## BIODATA PENULIS



### **Iya Purnama Sari, S.Pi., M.Si.**

Dosen Program Studi Perikanan Tangkap  
Politeknik Kelautan dan Perikanan Jembrana

Penulis lahir di Kampung Dalam, Solok, Sumatera barat pada 25 November 1994. Penulis diterima di Institut Pertanian Bogor (IPB) melalui jalur SNMPTN tahun 2013. Penulis menyelesaikan studi di Departemen Manajemen Sumberdaya Perairan, Fakultas Perikanan dan Ilmu Kelautan, IPB University pada tahun 2017. Selain itu, pada tahun yang sama 2017, penulis berkesempatan melanjutkan Studi Magister melalui program Sinergi (*Fast track*) S1-S2 pada program studi Pengelolaan Sumberdaya Pesisir dan Lautan, di IPB University dan lulus pada tahun 2019.

Selama perkuliahan penulis menjadi asisten mata kuliah Metode Statistika (2016-2018) dan Pengelolaan Sumberdaya Perikanan (2018/2019) di Departemen MSP FPIK-IPB, serta Analisis Statistika (2018/2019) di Sekolah Pascasarjana FPIK IPB. Penulis pernah mengikuti *Delegate at ASIA Young Sociopreneurship Leader Exchange* di Singapore tahun 2016. Tahun 2016, penulis meraih penghargaan sebagai “*The Most Inspiring Student MSP 50*” dalam acara Malam Karya FPIK 2016 Jembatan Karya Untuk Kejayaan FPIK, Fakultas Perikanan dan Ilmu Kelautan (FPIK), Institut Pertanian Bogor (IPB). Serta di tahun 2016 penulis juga memperoleh pendanaan Program

Kreativitas Mahasiswa kategori Penelitian oleh DIKTI dengan judul *“Sedimen Danau Sunter Selatan Sebagai Pupuk Tanaman Hydrophytik”*. Sebelumnya di tahun 2014 juga memperoleh pendanaan Program Kreativitas Mahasiswa kategori Gagasan Tertulis oleh DIKTI dengan judul *“JFI’ Jakarta Flood Insurance”*.

Penulis menjadi Dosen tetap Program Studi Perikanan Tangkap, Politeknik Kelautan dan Perikanan Jember tahun 2021 hingga saat ini. Penulis mengampu mata kuliah Statistika, Tingkah Laku Ikan dan Daerah Penangkapan Ikan, Teknik dan Metode Penangkapan Ikan, Penanganan dan Penyimpanan Ikan Hasil Tangkapan dan Hukum Maritim. Rumpun keilmuan yang didalami penulis adalah Pendugaan Stok Ikan, Perikanan Berkelanjutan, dan Manajemen Perikanan.

Penulis aktif pada seminar nasional maupun internasional, salah satunya yaitu seminar internasional *“The 10th International Conference on Fisheries and Aquaculture 2023 (ICFA 2023)”*. Beberapa Buku karya penulis pada tahun 2023 adalah Buku Ajar dengan judul *“Biologi Perikanan”* dan *“Buku Ajar Berbasis Keunggulan Lokal Statistika Terapan Teori dan Aplikasi pada Bidang Perikanan Tangkap”*, serta Buku Referensi dengan judul *“Pengembangan Sumber Daya Kelautan dan Perikanan Indonesia”*. Selain Dosen, penulis juga menjabat sebagai Sekretaris Satuan Pengawas Internal (SPI) Politeknik Kelautan dan Perikanan Jember pada tahun 2021-2023 dan Kepala SPI pada tahun 2024. Serta menjadi anggota Tim Teknis Tata Kelola Kinerja di Lingkungan Politeknik Kelautan dan Perikanan Jember Tahun 2021-2024 bagian Sub-Tim Perencanaan dan Verifikasi Kinerja.