

Fundamental Sampling Distributions
and Data Descriptions

8.1 Random Sampling

- Population consists of the totality of the observations with which we are concerned.
- A sample is a subset of a population.
- Any sampling procedure that produces inferences that consistently overestimate or consistently underestimate some characteristic of the population is said to be biased.
- Let X_1, X_2, \dots, X_n be n independent random variables, each having the same probability distribution $f(u)$. Define X_1, X_2, \dots, X_n to be a random sample of size n from the population $f(u)$ and write its joint probability distribution as
$$f(u_1, u_2, \dots, u_n) = f(u_1)f(u_2)\dots f(u_n).$$

8.2 Some Important Statistics

- Any function of the random variables constituting a random sample is called a statistic.

- Location Measures of a Sample: The Sample Mean, Median and Mode.

• Let X_1, X_2, \dots, X_n represent n random variables

a) Sample Mean

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

The term sample mean is applied to both the statistic \bar{X} and its computed value \bar{x} .

b) Sample Median

$$\tilde{x} = \begin{cases} u_{(n+1)/2}, & \text{if } n \text{ is odd.} \\ \frac{1}{2} (u_{n/2} + u_{n/2+1}), & \text{if } n \text{ is even.} \end{cases}$$

c). The sample mode is the value of the sample that occurs most often.

- Variability Measure of a Sample: The Sample Variance, Standard Deviation, and Range.

a) Sample variance:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Theorem 8.1: If S^2 is the variance of a random sample of size n , we may write

$$S^2 = \frac{1}{n(n-1)} \left[n \sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i \right)^2 \right]$$

b) Sample standard deviation

$$S = \sqrt{S^2}$$

where S^2 is the sample variance.

Let X_{\max} denote the largest of the X_i values and X_{\min} the smallest.

c) Sample range:

$$R = X_{\max} - X_{\min}.$$

-Skewness and Kurtosis

• A distribution is said to be symmetric if it can be folded along a vertical so that the two sides coincide. A distribution that lacks symmetry with respect to a vertical axis is said to be skewed.

$$S_k = \frac{\left(\sum_{i=1}^n (X_i - \bar{X})^3 \right) / n}{(S^2)^{3/2}}$$

• Kurtosis:

Tingkat ketajaman kurva distribusi

$$Kur = \frac{\left(\frac{1}{n} \sum (X_i - \bar{X})^4 \right)}{S^4}$$

8.3 Sampling Distributions

- Inference about the Population from Sample Information

• The probability distribution of a statistic is called a sampling distribution.

- The probability of \bar{X} is called the sampling distribution of the mean.

$$\bar{X} = \frac{1}{n} (X_1 + X_2 + \dots + X_n)$$

$$M_x = M_{\text{populasi}}$$

$$\sigma_x^2 = \frac{\sigma_{\text{populasi}}^2}{n}$$

- Distribusi sampel dari unknown distribution.
 - Jika sampling dilakukan dari populasi dengan distribusi yang tidak diketahui (finite/infinite), distribusi sampel akan diaproksimasi normal dengan mean $= M$ dan variansi $= \sigma^2/n$ dgn ukuran sampel yg besar.
 - Aproksimasi \bar{X} umumnya baik jika $n \geq 30$.
- Teorema Limit Pusat (Central Limit Theorem).

Bila \bar{X} rata-rata sampel acak berukuran n yang diambil dari populasi dengan rata-rata M dan variansi σ^2 yang berhingga, maka bentuk limit dan distribusinya =

$$Z = \frac{\bar{X} - M}{\sigma/\sqrt{n}}$$

bila $n \rightarrow \infty$, ialah distribusi normal baku $N(0,1)$.

- Probability Sampling.

- Simple random sampling: peluang terpilih setiap elemen sama.
- Stratified sampling: buat kelompok, lalu pilih secara random per strata.
- Systematic Sampling: buat sekuens elemen, pilih elemen pertama secara random, lalu pilih elemen berikutnya secara sistematis.
- Cluster Sampling: bentuk cluster, lalu pilih random cluster. Option: ambil semua elemen (single-stage) atau pilih random elemennya (two-stage).
- Multistage sampling: kombinasi beberapa teknik sampling.

- Non-Probability Sampling.

- Convenience Sampling: sampel dipilih berdasarkan ketersediaan data, biasanya untuk inisialisasi survei.
- Purposive Sampling: sampel dipilih berdasarkan tujuan study, atau segmen tertentu saja.
- Quota sampling: Sampel dipilih berdasarkan proporsi dari setiap karakteristik.
- Referral / Snowball Sampling: sampel dipilih berdasarkan rekomendasi.

- Sifat-sifat Kurva Normal
 - Modus = Mean (M)
 - simetri thd sb. tegak pada $x = M$
 - kurva mempunyai titik balik pada $x = M \pm \sigma$
 - Luas daerah di bawah kurva adalah 1.

- Distribusi Sampel Selisih 2 Rataan.

Bila sampel random independen ukuran n_1 dan n_2 diambil secara acak dari 2 populasi, diskrit/kontinu, msg² dan rataan M_1 dan M_2 dan variansi σ_1^2 dan σ_2^2 , maka distribusi sampel dari selisih rataan, $\bar{X}_1 - \bar{X}_2$, berdistribusi hampir normal, serta

$$M_{\bar{X}_1 - \bar{X}_2} = M_1 - M_2 ; \sigma_{\bar{X}_1 - \bar{X}_2}^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \text{ shg}$$

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (M_1 - M_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

secara hampiran adalah peubah normal baku. Aproksimasi \bar{X} umumnya baik jika $n \geq 30$

- Distribusi Sampel Variansi

Bila variansi sampel acak S^2 ukuran n diambil dari populasi normal dengan variansi σ^2 , maka statistik

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

berdistribusi chi-square (khi-kadrat) dengan derajat kebebasan $v = n-1$

- Tingkat kepercayaan $(1-\alpha)$, menerima klaim variansi populasi σ^2 , $\chi_{(1-\alpha)/2}^2 < \chi^2 < \chi_{\alpha/2}^2$

- Distribusi -t

- Penggunaan Central Limit Theorem mengasumsikan bahwa standar deviasi diketahui. Jika tidak diketahui, inferensi terhadap M menggunakan statistik =

$$T = \frac{\bar{X} - M}{S/\sqrt{n}}$$

- Jika ukuran sampel cukup besar, misalkan $n \geq 30$, distribusi T tidak berbeda jauh dari normal standar. Jika $n < 30$, sebaiknya menggunakan distribusi T . Misalkan Z variabel random normal baku dan V variabel random khi-kadrat dengan derajat kebebasan v . Bila Z dan V bebas, maka distribusi variabel random T :

$$T = \frac{Z}{\sqrt{V/v}} \text{ di mana } Z = \frac{\bar{X} - M}{\sigma/\sqrt{n}}, V = \frac{(n-1)S^2}{\sigma^2}, v = n-1$$

- Jika $v = \infty$, kurva sama dengan distribusi normal standar.

- Distribusi -F = Variance Ratio Distribution.

- Didefinisikan sebagai rasio dua variabel random khi-kadrat yang bebas, masing-masing dibagi dengan derajat

- Digunakan untuk comparative sampling

- Jadi dapat ditulis $F = \frac{U/U_1}{V/V_2}$

U dan V = variabel random bebar

U_1 dan V_2 = derajat kebebasan

Teorema :

- $f_{1-\alpha}(U_1, U_2) = \frac{1}{f_{\alpha}(U_2, U_1)}$

- Distribusi-F untuk Dua Variansi Populasi

- Bila S_1^2 dan S_2^2 variansi sampel random yang bebar ukuran n_1 dan n_2 yang diambil dari dua populasi,

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{S_1^2\sigma_2^2}{S_2^2\sigma_1^2}$$

berdistribusi-F dengan derajat kebebasan $U_1 = n_1 - 1$ dan $U_2 = n_2 - 1$