Chapter 8
Sampling Distributions

8.1 Random Sampling

Objectives:

1. **To review some of the important definitions.**
2. **To define the random sample.**

Some definitions:

The population and the sample were defined earlier. A random sample should be always representative and unbiased.

Definition: If $X_1, X_2, ..., X_n$ are **independent** r.v.'s each having the **same** probability distribution with a pdf $f(x)$, then $X_1, X_2, ..., X_n$ is defined to be a **random sample** of size n from the r.v. X.

Note: The numerical values of the random sample $X_1, X_2, ..., X_n$ are denoted by $x_1, x_2, ..., x_n$.

8.2 Some Important Statistics

Objectives:

1. **To define the parameter and the statistic.**
2. **To review some of the important statistics.**

Definition: A **parameter** is a characteristic value of the population and usually, practically, is unknown and need to be estimated.

Definition: A **statistic** is a function defined on the sample units and is used as an estimator for the population parameter. It is considered as a r.v. since its value varies from one r.s. to another.

Note: The sample mean, sample standard deviation, median, range and mode are all examples of sample statistics used to estimate their population counterparts.

Note: For a given r.s. $X_1, X_2, ..., X_n$, the following two r.v.'s are defined;

the sample **mean** $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ and the sample **variance** $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$.

8.4 Sampling Distributions

Objective: To define the sampling distribution.

Definition: The sampling distribution is the **probability distribution of a statistic**. It depends on the population size, the sample size, and the method of sampling.

Note: The sampling distribution of $\bar{X}$, for instance, from a sample of size n is the distribution that results when an experiment is conducted over and over (with the same sample size) and the many values of

$\bar{x}$ results which gives information about the variability of $\bar{x}$ values around μ in repeated experiments.

8.5 Sampling Distributions of Means

Objectives:

1. To introduce the central limit theorem (CLT).
2. To identify the sampling distribution for the sample mean.
3. To identify the sampling distribution for the difference of two (independent or dependent) sample means.

Theorem: (CLT) If $\bar{X}$ is the mean of a r.s. of size n from a population with mean μ and finite variance σ^2, then $\bar{X}$ is approximately normally distributed with $\mu_{\bar{X}} = \mu$ and $\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$, i.e. $\bar{X}$ ~N(μ , $\frac{\sigma^2}{n}$), and hence $Z = \frac{\bar{X}-\mu}{\sigma_{\bar{X}}} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ ~N(0 , 1) ⇔ $n \to \infty$.

Notes:

1. If the sample is drawn from N(μ , σ^2) then $\bar{X}$ ~N(μ , $\frac{\sigma^2}{n}$).
2. The normal approximation for the sample mean is good ⇔ $n \geq 30$, but if $n < 30$ then the approximation will be good ⇔ the sample is drawn from a normal population ⇒ $\bar{X} : N(\mu , \frac{\sigma^2}{n})$.

Ex.1 (8.6/210): If X : N(800 , 1600) and n = 16, then find $P(\bar{X} < 775)$.
Ex.2 (8.7/211): Let X : N(5 , 0.01) and n = 100. Find $P(|\bar{X} - 5| \geq 0.027)$.

Theorem: If two independent samples of sizes n_1 and n_2, are drawn from two, discrete or continuous, populations with means μ_1 & μ_2 and finite variances σ_1^2 & σ_2^2, then the sampling distribution of the **means difference**, $\bar{X}_1 - \bar{X}_2$, is approximately normal with a mean $\mu_{\bar{X}_1-\bar{X}_2} = \mu_1 - \mu_2$ and $\sigma^2_{\bar{X}_1-\bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$. Hence,

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{X}_1-\bar{X}_2}} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0 , 1).$$

Ex.3 (8.8/213): If $n_A = n_B = 18$, $\sigma_A^2 = \sigma_B^2 = 1$, and $\mu_1 = \mu_2$, then find $P(\bar{X}_1 - \bar{X}_2 > 1.0)$.
Ex.4 (8.9/214): $n_1 = 36$, $\mu_1 = 6.5$, $\sigma_1^2 = 0.81$, $n_2 = 49$, $\mu_2 = 6.0$, and $\sigma_2^2 = 0.64$, then find $P(\bar{X}_1 - \bar{X}_2 > 1.0)$.
Ex.5: If X is a r.v. with the following pdf;

X	1	2	3	4
f(x)	0.15	0.25	0.4	0.2

a. Find μ_X and σ_X^2.
b. Find $\mu_{\bar{X}}$ and $\sigma_{\bar{X}}^2$ for a r.s. of size 64 from X.

c. What is the probability that the average of a sample of size 64 will exceed 2.85?

Ex.6: Problem (10/216), $\mu = 3.2$ min., $\sigma = 1.6$ min., and $n = 64$.

Ex.7: Problem (12/216), $n_1 = 25$, $\mu_1 = 80$, $\sigma_1^2 = 25$, $n_2 = 36$, $\mu_2 = 75$, and $\sigma_2^2 = 9$. Find $P(3.4 \leq \overline{X}_1 - \overline{X}_2 \leq 5.9)$.

8.6 The Sampling Distribution of S^2

Objective: To find the sampling distribution of the sample variance.

Theorem: If S^2 is the variance of a r.s. of size n from a normal population with a finite variance σ^2, then the statistic;

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\sigma^2} \sim \chi^2_{n-1} \text{ where } Z^2 = \chi_1^2 .$$

Ex.1 (8.10/218): $\mu = 3$, $\sigma = 1$, and $n = 5$. The sample 1.9, 2.4, 3.0, 3.5, 4.2 has $S^2 = 0.815 \Rightarrow \chi_5^2 = 3.26$. The sample will be accepted if its variance is within 95% of the data.

Ex.2: Problem (1/227), a. $\chi^2_{0.025;15} = ?$, b. $\chi^2_{0.01;7} = ?$

Ex.3: Problem (4/227), a. $P(\chi^2 > \chi^2_\alpha) = 0.01$, $\nu = 21$, b. $P(\chi^2 < \chi^2_\alpha) = 0.95$, $\nu = 6$.

Ex.4: Problem (5/227), if $n = 25$ and $\sigma^2 = 6$, then find $P(9.1 < S^2)$ and $P(3.462 < S^2 < 10.745)$.

8.7 The t-Distribution

Objectives:

1. **To define the t-distribution.**
2. **To find the probabilities of a t r.v.**

Definition: If X : N(μ , σ^2), n is small ($n < 30$), and σ is unknown, then the r.v. $T = \dfrac{\overline{X} - \mu}{s_{\overline{X}}} = \dfrac{\overline{X} - \mu}{s/\sqrt{n}}$: t_{n-1} where n–1 are called the degrees of freedom denoted by d.f.

Theorem: Let the r.v. Z:N(0,1), and the r.v. V: χ_r^2. If Z and V are independent then $T = \dfrac{Z}{\sqrt{V/r}}$: t_r.

Corollary: If $X_1, X_2, ..., X_n$ are members of a r.s. each having N(μ , σ^2), then the r.v. $T = \dfrac{\overline{X} - \mu}{\sqrt{S^2/n}} = \dfrac{\overline{X} - \mu}{S/\sqrt{n}}$.

Properties of the t-curve:

1. It is bell shaped but shorter and more spread than the normal curve.
2. For longer runs, as $n \to \infty$, $t \to Z$:N(0,1).
3. By symmetry $t_\alpha = -t_{1-\alpha}$.

Ex.1: Problem (9/228). Find: a. $P(T_7 < 2.365)$, b. $P(1.318 < T_{24})$, c. $P(-1.356 < T_{12} < 2.179)$, d. $P(T_{17} > -2.567)$.

Ex.2 (8.13/222): Find k such that $P(k < T_{14} < -1.761) = 0.045$.

Ex.3 (8.14/223): $\mu = 500$, $n = 25$, $\overline{X} = 518$, and $S = 40$.