Chapter 1

Logan | Nick

`{r setup, include=FALSE} knitr::opts_chunk$set(fig.width=6, fig.height=4, fig.align="center", echo = FALSE)`

`{r message=FALSE} if (!require(mosaic)) install.packages("mosaic", repos = 'https://cloud.r-project.org') library(mosaic) # load the package mosaic to use its functions`

`library(float)`

#5.3, 5.4 Estimation

Def Parameter : a numerical value that describes the population. This value is fixed but often unknown in practice.

Def Statistic : a numerical value that describes a sample

- known from sample data
- varies from sample to sample
- use it to estimate the parameter
- use cap letters to denote R.V.
  - $\overline{X}$ is used for the sample mean R.V.
- use lower case letters to denote a calculated value of the statistic
  - $\overline{x}$ is used for a specific value of the sample mean

Def The R.V.'s $X_1, X_2, \ldots, X_n$ are said to form a simple random sample of size n. If

1. The X_i's are independent R.V's
2. Every X_i has the same probability distribution
    - independent and identically distributed (iid.)

The idea of a sampling distribution

1. a random sample is selected from a population
2. A statistic is calculated from the random data
    - a statistic is random as it arises from sample to sample according to random selection
    - the distribution of the statistic is called sampling distribution
    - the sampling distribution describes all possible values of the statistic and the probability (likelihood) of those values

Ex Let X be the # of packages being mailed by a random customer.

x	1	2	3	4
$p(x)$	.4	.3	.2	.1

a. Consider a random sample of size $n = 2$. Let $\overline{X}$ be the # of packages shipped.

Obtain the pdf of $\overline{X}$

x_1	x_2	$p(x_1, x_2)$	$\overline{x}$
1	1	$.4 * .4 = .16$	$\dfrac{1+1}{2} = 1$
1	2	$.4 * .3 = .12$	$\dfrac{1+2}{2} = 1.5$
1	3	$.4 * .2 = .08$	2
1	4	$.4 * .1 = .04$	2.5
2	1	.12	1.5
2	2	.09	2
2	3	.06	2.5
2	4	.03	3
$\vdots$	$\vdots$	$\vdots$	$\vdots$

$\overline{x}$	1	1.5	2	2.5	3	3.5	4
$p(\overline{x})$	.16	.24	.25	.2	.1	.04	.01

b. $P(\overline{x} \leq 2.5) = .16 + .24 + .25 + .2 = .85$

c. Find $E(\overline{X} = \sum \overline{x} p(\overline{x}) = 2$

Note $E(X) = 2$

Properties of the Sample Mean

Let $X_1, \ldots, X_n$ be simple random sample from a population with mean μ and standard deviation σ.

- $E(\overline{x}) = \mu_{\overline{x}} = \ldots = \mu$
- The variance of $\overline{X}$ is

$$\sigma^2/n = Var(\overline{X})$$

insert proof here

So the standard deviation of $\overline{X}$ is

$$\sigma_{\overline{X}} = \frac{\sigma_x}{\sqrt{n}} = \frac{\sigma}{\sqrt{n}}$$

Ex Let a SRS of size $n = 100$ be drawn from a population with mean 50 and standard deviation of 7

a. Find the mean and standard deviation of $\overline{X}$.
   - $\mu_{\overline{X}} = 50$
   - $\sigma_{\overline{X}} = \frac{7}{\sqrt{100}} = \frac{7}{10}$

The Central Limit Theorem

Let $X_1, \ldots, X_n$ be a random sample fro a distribution with mean μ and variance σ^2, then if n is sufficiently large $(n > 30)$.

$$\overline{X} \approx \text{ normal with mean } \mu_{\overline{X}} = \mu$$

$$Var(\overline{X}) = \sigma^2_{\overline{X}} = \frac{\sigma^2}{n}$$

$$Z = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma} = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

$T_0 = X_1 + \cdots + X_n \approx$ normal w/ mean

$$\mu_{T_0} = n\mu$$

$$\sigma^2_{T_0} = n\sigma^2$$

$$Z = \frac{T_0 - \mu_{T_0}}{\sigma_{T_0}} = \frac{T_0 - n\mu}{\sqrt{n}\sigma}$$

Note

1. This is true regardless of the shape of the distribution of the population for large samples $(n > 30)$
2. Always true for normal populations
3. true for discrete & normal populations

Ex Let X be the weights (lbs) of bags of feed. \$X \$ is normlally distributed w/ a mean of 100 lbs. & variance of 4.

- Find probability that the average of 20 random bags is less than 101 lbs.

$$\mu_X = 100 \qquad \sigma_X^2 = 4 \qquad n = 20 \qquad \sigma_X = 2 \qquad \sigma_{\overline{X}} = \frac{2}{\sqrt{20}}$$