324 Stat
Lecture Notes

(6) Fundamental Sampling Distribution and Data Discription

(Book*: Chapter 8 ,pg225)

Probability& Statistics for Engineers & Scientists
By Walpole, Myers, Myers, Ye

8.1 Random Sampling:

Population:

A population consists of the total observations with which we are concerned at a particular time.

Sample:

A sample is a subset of a population.

Random Sample

Let $X_1, X_2, ..X_n$ be n independent random variable; having the same probability distribution f(X), then $X_1, X_2, ..X_n$ is defined to be a random sample of size n from the population f(x)

Statistic :

Any function of the random sample is called a statistic.

8.2 Some important Statistics:

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

(a) Sample Mean:

If $X_1, X_2, \ldots X_n$ represent a random sample of size **n**, then the sample mean is defined by the statistic:

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

Properties of the Mean (Reading):

1. The mean is the most commonly used measure of certain location in statistics.
2. It employs all available information.
3. The mean is affected by extreme values.
4. It is easy to calculate and to understand.
5. It has a unique value given a set of data.

EX 1:

The length of time, in minutes, that **10** patients waited in a doctor's office before receiving treatment were recorded as follows: **5, 11, 9, 5, 10, 15, 6, 10, 5** and **10.** Find the mean.

Solution:

$$n = 10, \quad \sum x_i = 5 + 11 + \ldots + 10 = 86$$

$$\bar{X} = \frac{\sum x_i}{n} = \frac{86}{10} = 8.6$$

See Ex 8.4 pg 228

(b) Sample Median:

- If $X_1 \ldots X_n$ represent a random sample of size **n ,** arranged in increasing order of magnitude, then the sample median, which is denoted by Q_2 is defined by the statistic:

$$Q_2 = \left\{ \begin{array}{ll} X_{\frac{(n+1)}{2}} & if \quad n \quad is \quad odd \\ \\ \dfrac{X_{n/2+} X_{(n/2)+1}}{2} & if \quad n \quad is \quad even \end{array} \right\} \qquad (2)$$

Properties of the Median(Reading):

1. The median is easy to compute if the number of observations is relatively small.
2. It is not affected by extreme values.

EX 2:

The number of foreign ships arriving at an east cost port on **7** randomly selected days were **8, 3, 9, 5, 6, 8** and **5.**

Find the sample median.

Solution:

The arranged values are: **3 5 5 6 8 8 9**

$$n = 7 \quad , \quad \frac{n+1}{2} = \frac{7+1}{2} = 4$$

$$Q_2 = 6$$

EX 3:

The nicotine contents for a random sample of **6** cigarettes of a certain brand are found to be **2.3, 2.7, 2.5, 2.9, 3.1** and **1.9** milligrams. Find the median.

Solution:

The arranged values are: **1.9 ,2.3 ,2.5 ,2.7, 2.9, 3.1 .**

$$n = 6 \quad , \quad \frac{n}{2} = \frac{6}{2} = 3 \ , \qquad \frac{n}{2} + 1 = \frac{6}{2} + 1 = 3 + 1 = 4$$

$$Q_2 = \frac{2.5 + 2.7}{2} = 2.6$$

(c)Sample Mode:

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

Properties of the Mode(Reading):

1. The value of the mode for small sets of data is almost useless.
2. It requires no calculation.

EX(4):

The numbers of incorrect answers on a true – false test for a random sample of **14** students were recorded as follows: **2, 1, 3, 0, 1, 3, 6, 0, 3, 3, 2, 1, 4, and 2**, find the mode.

Solution:

mode=3

H.w.

Find the mean and median

Notes:

1. The sample means usually will not vary as much from sample to sample as will the median.
2. The median (when the data is ordered) and the mode can be used for qualitative as well as quantitative data.

2- Variability Measures of a Sample: The Sample Variance, Standard deviation and Range:

6.3.1 The Range:

(a)The range of a random sample $X_1 \ldots X_n$ is defined by the statistic $X_{(n)} - X_{(1)}$,

where $X_{(n)}$ *and* $X_{(1)}$ respectively the largest and the smallest observations which is denoted by **R**, then:

$$R = Max - Min = x_{(n)} - x_{(1)}$$

EX 5:

Let a random sample of five members of a sorority are **108,112,127,118** and **113**. Find the range.

Solution:

$$\mathbf{R}=127-108=19$$

(b) Sample Variance:

If $X_1 \ldots X_n$ represent a random sample of size **n**, then the sample variance, which is denoted by S^2 is defined by the statistic:

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

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right]$$

Ex 8.2 pg 229

A comparison of coffee prices at **4** randomly selected grocery stores in San Diego showed increases from the previous month of **12, 15, 17, 20,** cents for a **200** gram jar. Find the variance of this random sample of price increases.

Solution:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{12+15+17+20}{4} = 16$$

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1} = \frac{(12-16)^2 + (15-16)^2 + (17-16)^2 + (20-16)^2}{3}$$

$$= \frac{34}{3}$$

(b) Sample Standard Deviation:

The sample standard deviation s is given by:

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

Where S^2 is the sample variance.

EX 6 :

The grade – point average of **20** college seniors selected at random from a graduating class are as follows:

3.2, 1.9, 2.7, 2.4, 2.8, 2.9, 3.8, 3.0, 2.5, 3.3, 1.8, 2.5, 3.7, 2.8, 2.0, 3.2, 2.3, 2.1, 2.5, 1.9. Calculate the variance and the standard deviation.

Solution: $$\sum_{i=1}^{n} X_i = 53.3, \sum_{i=1}^{n} X_i^2 = 148.55$$

(answer: $\bar{X} = 2.665, S = 0.585, S^2 = 0.342$)

See Ex 8.3 pg 230

8.3 Sampling Distributions

Definition:

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

8.4 Sampling Distributions of Means and the Central Limit Theorem

Sampling Distributions of Means:

Suppose that a random sample of size n observation is taken from normal distribution then

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

Has a normal distribution with mean and variance

$$E(\bar{X}) = \mu_{\bar{X}} = \mu \quad , \quad V(\bar{X}) = \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} \qquad (3)$$

Central Limit Theorem:

If $\bar{X}$ is the mean of a random sample of size **n** taken from a population with mean μ and finite variance σ^2, then the limiting form of the distribution of:

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \quad as \quad n \to \infty \qquad (4)$$

is approximately the standard normal distribution■

Notes

- $f(Z) \sim N(0,1) \Rightarrow E(\bar{X}) = \mu$, $\underset{\bar{X}}{V}(\bar{X}) = \dfrac{\sigma^2}{n}$, $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}}$

$$\Rightarrow \bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \qquad (5)$$

- The approximation of will generally be good if n$\geq$ 30.

EX(8.4 pg 234):

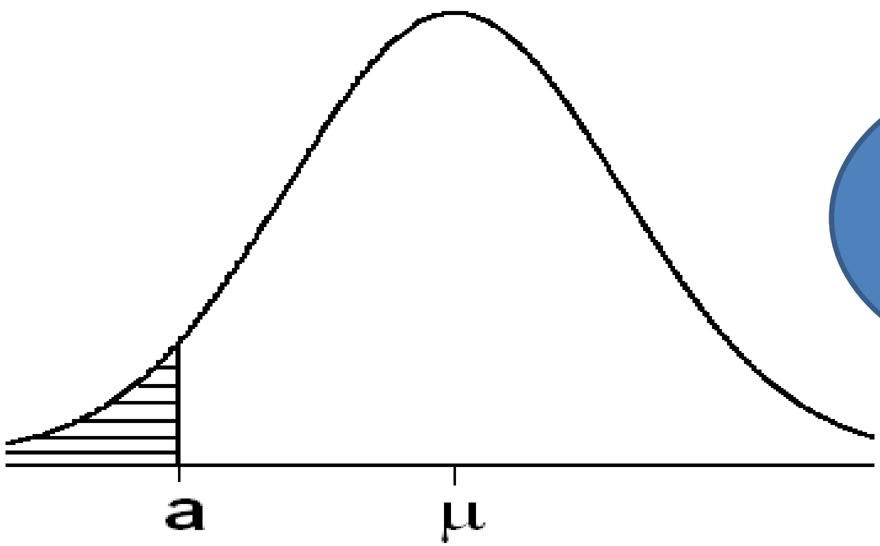
An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed with mean equal to **800** hours and a standard deviation of **40** hours. Find the probability that a random sample of **16** bulbs will have an average life of less than **775** hours.

Solution:

Let **X** be the length of life and $\bar{X}$ is the average life;

$$n = 16, \mu = 800, \ \sigma = 40$$

$$P(\bar{X} < 775) = P\left(Z < \frac{775 - 800}{40/\sqrt{16}}\right) = P(Z < -2.5) = 0.0062$$



See Ex 8.5 pg 237

Sampling distribution of the difference between two means

Theorem If independent samples of size $\mathbf{n_1}$ and $\mathbf{n_2}$ are drawn at random from populations, discrete or continuous with means $\mu_1 \ and \ \mu_2$ and variances $\sigma_1^2 \ and \ \sigma_2^2$ respectively, then the sampling distribution of the difference of means $\bar{X}_1 - \bar{X}_2$ is approximately normally distributed with mean and variance given by:

$$E(\bar{X}_1 - \bar{X}_2) = \mu_{(\bar{X}_1 - \bar{X}_2)} = \mu_1 - \mu_2 \quad , \quad \sigma^2_{(\bar{X}_1 - \bar{X}_2)} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \qquad (6)$$

Hence

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \qquad (7)$$

is approximately a standard normal variable.

EX(10):

A sample of size $\mathbf{n_1=5}$ is drawn at random from a population that is normally distributed with mean $\mu_1 = 50$ and variance $\sigma_1^2 = 9$ and the sample mean $\overline{X}_1$ is recorded. A second random sample of size $\mathbf{n_2=4}$ is selected independent of the first sample from a different population that is also normally distributed with mean $\mu_2 = 40$ and variance $\sigma_2^2 = 4$ and the sample mean $\overline{X}_2$ is recorded.

Find $P(\overline{X}_1 - \overline{X}_2 < 8.2)$

Solution:

$n_1 = 5 \qquad n_2 = 4$

$\mu_1 = 50 \qquad \mu_2 = 40$

$\sigma_1^2 = 9 \qquad \sigma_2^2 = 4$

See Case Study 8.2 pg 238

$$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2 = 50 - 40 = 10$$

$$\sigma^2_{\bar{X}_1 - \bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \frac{9}{5} + \frac{4}{4} = 2.8$$

$$P(\bar{X}_1 - \bar{X}_2 < 8.2) = P(Z < \frac{8.2 - 10}{1.673}) = P(Z < -1.08) = 0.1401$$

EX(11): (H.W)

The television picture tubes of manufacturer **A** have a mean lifetime of **6.5** years and a standard deviation of **0.9** year, while those of manufacturer **B** have a mean lifetime of **6** years and a standard deviation of **0.8** year. What is the probability that a random sample of **36** tubes from manufacture **A** will have a mean lifetime that at least **1** year more than the mean lifetime of a sample of **49** from tubes manufacturer **B**?

Solution:

Population 1	Population 2
$\mu_1 = 6.5$	$\mu_2 = 6.0$
$\sigma_1 = 0.9$	$\sigma_2 = 0.8$
$n_1 = 36$	$n_2 = 49$

$$\mu_{\bar{X}_1 - \bar{X}_2} = 6.5 - 6 = 0.5$$

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{0.81}{36} + \frac{0.64}{49}} = 0.189$$

$$P[(\bar{X}_1 - \bar{X}_2) > 1) = P(Z > \frac{1 - 0.5}{0.189}) = P(Z > 2.65)$$

$$= 1 - P(Z < 2.65) = 1 - 0.996 = 0.004$$

Sampling Distribution of the sample Proportion (Reading):

Let **X**= no. of elements of type **A** in the sample

P= population proportion = no. of elements of type A in the population / **N**

$\hat{p}$ **=** sample proportion **=** no. of elements of type A in the sample / n = x/n

$$\because x \sim binomial\,(n,p) \rightarrow E(x) = np, V(x) = npq$$

$$\therefore 1. E(\hat{p}) = E\left(\frac{x}{n}\right) = p$$

$$2. V(\hat{p}) = V\left(\frac{x}{n}\right) = \frac{pq}{n},\ q = 1 - p$$

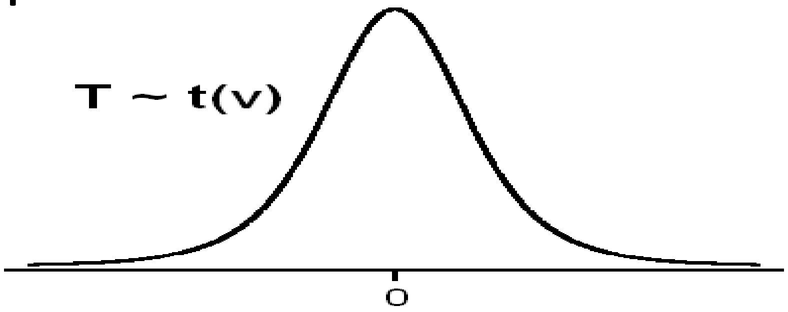
3. For large n, we have:

$$\hat{p} \sim N\left(p, \sqrt{\frac{pq}{n}}\right)$$

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} \sim N(0,1)$$

8.6 *t* – Distribution (pg 246):

* ***t*** distribution has the following properties:



1. It has mean of zero.
2. It is symmetric about the mean.
3. It ranges from $-\infty \ to \ \infty$.
4. Compared to the normal distribution, the ***t*** distribution is less peaked in the center and has higher tails.
5. It depends on the degrees of freedom **(n-1)**.
6. The t distribution approaches the normal distribution as **(n-1)**

   approaches ∞.

Notes

- Since the t-distribution is symmetric about zero we have

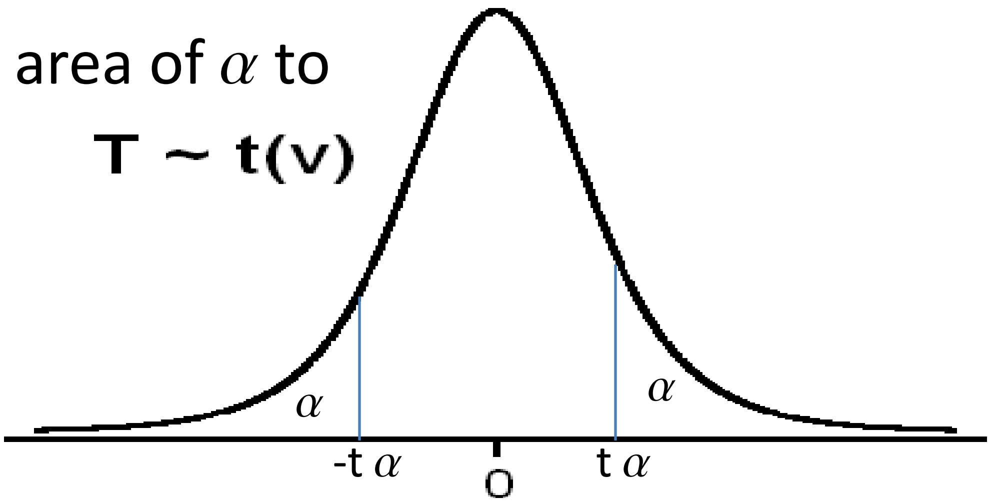
$t_{1-\alpha} = -t_{\alpha}$

- Table A4 pg 737-738 represent the

critical values of t-distribution

Where $t\alpha$ leaving an area of α to

The right.



Corollary 8.1

- Let X_1, X_2, ...X_n be independent random variables from normal with mean μ and standard deviation σ. Let
- $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})$

Then the random variable $\mathrm{T} = \dfrac{\bar{X}-\mu}{s/\sqrt{n}}$ has a t-distribution with υ =n-1 degrees of freedom.

EX(12):

Find:

$$(a)\, t_{0.025} \quad when \quad v = 14$$

$$(b)\, t_{0.01} \quad when \quad v = 10$$

$$(c)\, t_{0.995} \quad when \quad v = 7$$

(a) $t_{0.025}$ at $v = 14 \rightarrow t = 2.1448$

(b) $t_{0.01}$ at $v = 10 \rightarrow t = 2.764$

(c) $t_{0.995}$ at $v = 7 \rightarrow$

$$t = -t_{0.005} = -3.499$$

EX(13):

Find:

$(a)\ P(T < 2.365) \quad when \quad v = 7$

$(b)\ P(T > 1.318) \quad when \quad v = 24$

$(c)\ P(-t_{0.025} < T < t_{0.05})$

$(d)\ P(T > -2.567) \quad when \quad v = 17$

solution

(a) $P(T<2.356)=1-0.025=$ **0.975** at $\nu = 7$

(b) $P(T>1.318)= 0.1$ at $\nu =24$

(c) $P(-t_{0.025}<T<t_{0.05})$

$t_{0.05}$ leaves an area of 0.05 to the right and $-t_{0.025}$ leaves an area 0.025 to the left so the total area is

$1-0.05-0.025=0.925.$

(d) $P(T>-2.567)= 1-0.01=0.99$ at $\nu =17$