بسم الله الرحمن الرحيم

STAT 105

Prepared by

Abdulrahman Alfaifi

King Saud University

Department of Statistics and Operation Research

Email: alfaifi@ksu.edu.sa

Twitter: @AlfaifiStat

ملاحظة : ليس بالضرورة ان تكون المذكرة شاملة للمقرر

Oct 2019

CHAPTER 1

- *Discrete Random Variables.*
- *Binomial Distribution.*
- *Hyper geometric Distribution.*
- *Poisson Distribution.*

Discrete Random Variables:

- $0 \leq f(x) \leq 1$
- $\sum f(x) = 1$
- $f(x) = P(X = x)$
- $\mu_x = E(X) = \sum x\, f(x)$
- $\sigma_x^2 = Var(X) = E(X^2) - E(X)^2$
- $E(X^2) = \sum x^2 f(x)$
- $E(aX \pm b) = aE(X) \pm b$
- $Var(aX \pm b) = a^2 Var(X)$
- $F(x) = P(X \leq x)$

1. *Let the random variable X having the probability distribution (p.m.f) as:*

x	-3	6	9
$f(x)$	$1/6$	$1/2$	$1/3$

(A) Find the probability that:

i. *The random variable X assumes a non-negative value.*

$$P(X = 6) + P(X = 9) = \frac{1}{2} + \frac{1}{3} = \frac{5}{6}$$

ii. *The random variable X assumes a value less than 7.*

$$P(X < 7) = P(X = -3) + P(X = 6) = \frac{1}{6} + \frac{1}{2} = \frac{4}{6}$$

(B) Find μ_x and σ_x^2 and then deduce each of μ_y and σ_y^2; where $Y = 3X - 6$.

$$\mu_x = E(X) = \sum x\, f(x) = \left(-3 \times \frac{1}{6}\right) + \left(6 \times \frac{1}{2}\right) + \left(9 \times \frac{1}{3}\right) = \frac{11}{2}$$

$$\boxed{\sigma_x^2 = E(X^2) - E(X)^2}$$

$$E(X^2) = \sum x^2 f(x) = \left(-3^2 \times \frac{1}{6}\right) + \left(6^2 \times \frac{1}{2}\right) + \left(9^2 \times \frac{1}{3}\right) = \frac{93}{2}$$

$$\sigma_x^2 = E(X^2) - E(X)^2 = \frac{93}{2} - \left(\frac{11}{2}\right)^2 = \frac{65}{4}$$

$$\mu_y = E(Y) = E(3X - 6) = 3 \times E(X) - 6 = 3 \times \frac{11}{2} - 6 = \frac{21}{2}$$

$$\sigma_Y^2 = Var(Y) = Var(3X - 6) = 3^2 Var(X) = 9 \times \frac{65}{4}$$

2. ***A large industrial firm purchases several word processors at the end of each year, the exact number depending of the frequency of repairs in the previous year. Suppose that the number of word processors, X that are purchased each year has the following probability distribution:***

x	*0*	*1*	*2*	*3*
$f(x)$	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{4}{10}$	$\frac{2}{10}$

Find μ_x and σ_x^2.

- $\mu_x = \left(0 \times \frac{1}{10}\right) + \left(1 \times \frac{3}{10}\right) + \left(2 \times \frac{4}{10}\right) + \left(3 \times \frac{2}{10}\right) = \frac{17}{10}$

$$\boxed{\sigma_x^2 = E(X^2) - E(X)^2}$$

- $E(X^2) = \sum x^2 f(x)$

$$= \left(0^2 \times \frac{1}{10}\right) + \left(1^2 \times \frac{3}{10}\right) + \left(2^2 \times \frac{4}{10}\right) + \left(3^2 \times \frac{2}{10}\right) = \frac{37}{10}$$

$$\sigma_x^2 = E(X^2) - E(X)^2 = \frac{37}{10} - \left(\frac{17}{10}\right)^2 = \frac{81}{100}$$

3. ***Let X be a discrete random variable with probability mass function:***

$$f(x) = cx \qquad ; \qquad x = 1,2,3,4$$

What is the value of c ?

x	*1*	*2*	*3*	*4*
$f(x)$	c	$2c$	$3c$	$4c$

$$c + 2c + 3c + 4c = 1 \quad \Longrightarrow c = \frac{1}{10}$$

Then probability mass function:

x	*1*	*2*	*3*	*4*
$f(x)$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{3}{10}$	$\frac{4}{10}$

4. *Suppose that the number of cars X pass through a car wash between 4:00pm and 5:00pm on any sunny Friday has the following probability distribution:*

x	4	5	6	7	8	9
$f(x)$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$

Let $g(x) = 2X - 1$ *represent the amount of dollars, paid to the attendant by the manger. Find the attendant's expected earnings (and it variance) for this particular time period.*

$$\mu_x = \left(4 \times \frac{1}{12}\right) + \left(5 \times \frac{1}{12}\right) + \left(6 \times \frac{1}{4}\right) + \left(7 \times \frac{1}{4}\right) + \left(8 \times \frac{1}{6}\right) + \left(9 \times \frac{1}{6}\right) = \frac{41}{6}$$

$$\boxed{\sigma_x^2 = E(X^2) - E(X)^2}$$

$$E(X^2) = \left(4^2 \times \frac{1}{12}\right) + \left(5^2 \times \frac{1}{12}\right) + \cdots + \left(9^2 \times \frac{1}{6}\right) = \frac{293}{6}$$

$$\sigma_x^2 = E(X^2) - E(X)^2 = \frac{293}{6} - \left(\frac{41}{6}\right)^2 = \frac{77}{36}$$

$$\mu_y = E(Y) = E(2X - 1) = 2 \times E(X) - 1 = 2 \times \frac{41}{6} - 1 = \frac{38}{3}$$

$$\sigma_Y^2 = Var(Y) = Var(2X - 1) = 2^2 Var(X) = 4 \times \frac{77}{36}$$

Binomial Distribution:

$$f(x) = \binom{n}{x} p^x q^{n-x} \;\; ; \quad x = 0,1 \ldots.,n$$

$$* \, E(X) = np \qquad * \, Var(X) = npq$$

$$q = 1 - p$$

Q1. Suppose that 33% of the buildings in a certain city violate the building code. A building engineer randomly inspects a sample of 3 new buildings in the city.

(a) Find the (p.m.f) of the random variable X representing the number of buildings that violate the building code in the sample.

$$p = \frac{1}{3} \;\; , \;\; n = 3$$

$$f(x) = \binom{3}{x}\left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{3-x} \quad ; \quad x = 0,1,2,3$$

(b) Find the probability that:

(i) <u>None of the buildings</u> in the sample violating the building code.

$$P(X = 0) = f(0) = \binom{3}{0}\left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^3 = 0.296$$

(ii) <u>One building</u> in the sample violating the building code.

$$P(X = 1) = f(1) = \binom{3}{1}\left(\frac{1}{3}\right)^1 \left(\frac{2}{3}\right)^2 = 0.44$$

(iii) <u>At least one building</u> in the sample violating the building code.

$$P(X \geq 1) = 1 - P(X < 1) = 1 - f(0) = 1 - 0.296 = 0.704$$

(c) Find the expected number of buildings that violate the building code $E(X)$.

$$E(X) = np = 3 \times \frac{1}{3} = 1$$

(d) Find Var(X).

$$Var(X) = npq = 3 \times \frac{1}{3} \times \frac{2}{3} = \frac{2}{3}$$

Q3. Suppose that the probability that a person dies when he or she contracts a certain disease is 0.4. A sample of 10 persons who contracted this disease is randomly chosen.

(1) What is the expected number of persons who will die in this sample?

$$p = 0.4 \quad , \quad n = 10$$

$$E(X) = np = 10 \times 0.4 = 4$$

(2) What is the variance of the number of persons who will die in this sample?

$$Var(X) = npq = 10 \times 0.4 \times 0.6 = 2.4$$

(3) What is the probability that <u>exactly 4 persons</u> will die among this sample?

$$f(x) = \binom{10}{x}(0.4)^x(0.6)^{10-x} \quad ; \quad x = 0,1,\ldots.,10$$

$$P(X = 4) = f(4) = \binom{10}{4}(0.4)^4(0.6)^6 = 0.251$$

(4) What is the probability that <u>less than 3 persons</u> will die among this sample?

$$P(X < 3) = f(0) + f(1) + f(2)$$

$$= \binom{10}{0}(0.4)^0(0.6)^{10} + \binom{10}{1}(0.4)^1(0.6)^9 + \binom{10}{2}(0.4)^2(0.6)^8 = 0.167$$

(5) What is the probability that <u>more than 8 persons</u> will die among this sample?

$$P(X > 8) = f(9) + f(10)$$

$$= \binom{10}{9}(0.4)^9(0.6)^1 + \binom{10}{10}(0.4)^{10}(0.6)^0 = 0.0017$$

Q9. If X~ Binomial (n,p), E(X)=1, and Var(X)=0.75, find P(X=1).

$$X \sim Binomial(n,p) \quad \& \quad E(X) = 1 \quad \& \quad Var(X) = 0.75$$

$$\frac{Var(X)}{E(X)} = \frac{0.75}{1} \Rightarrow \frac{npq}{np} = \frac{0.75}{1} \Rightarrow q = 0.75 \Rightarrow p = 0.25.$$

$$E(X) = 1 \Rightarrow np = 1 \Rightarrow n \times 0.25 = 1 \Rightarrow n = 4.$$

$$f(x) = \binom{4}{x}(0.25)^x(0.75)^{4-x} \quad ; \quad x = 0,1,2,3,4$$

$$P(X = 1) = f(1) = \binom{4}{1}(0.25)^1(0.75)^3 = 0.422$$

Q11. A traffic control engineer reports that 75% of the cars passing through a checkpoint are from Riyadh city. If at this checkpoint, five cars are selected at random.

(1) The probability that <u>none of them is from Riyadh</u> city equals to:

$$p = 0.75 \quad , \quad n = 5$$

$$f(x) = \binom{5}{x}(0.75)^x(0.25)^{5-x} \quad ; \quad x = 0,1,2,3,4,5$$

$$P(X = 0) = f(0) = \binom{5}{0}(0.75)^0(0.25)^5 = 0.00098$$

(2) The probability that <u>four of them are from Riyadh</u> city equals to:

$$P(X = 4) = f(4) = \binom{5}{4}(0.75)^4(0.25)^1 = 0.3955$$

(3) The probability that <u>at least four of them are from Riyadh</u> city equals to:

$$P(X \geq 4) = f(4) + f(5)$$
$$\binom{5}{4}(0.75)^4(0.25)^1 + \binom{5}{5}(0.75)^5(0.25)^0 = 0.6328$$

(4) The expected number of cars that are from Riyadh city equals to:

$$E(X) = np = 5 \times 0.75 = 3.75$$

Hyper geometric Distribution:

$$f(x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \quad ; \quad x = 0,1, \ldots . min(n,k)$$

$$* \, E(X) = n \times \frac{k}{N} \qquad * \, Var(X) = n \times \frac{k}{N}\left(1 - \frac{k}{N}\right)\left(\frac{N-n}{N-1}\right)$$

Q1. A shipment of 7 television sets contains 2 defective sets. A hotel makes a random purchase of 3 of the sets.

(i) *Find the probability distribution function of the random variable X representing the number of defective sets purchased by the hotel.*

$$N = 7 \quad , \quad n = 3 \quad , \quad k = 2$$

$$f(x) = \frac{\binom{2}{x}\binom{5}{3-x}}{\binom{7}{3}} \quad ; \quad x = 0,1,2$$

(ii) *Find the probability that the hotel purchased <u>no defective</u> television sets.*

$$P(X = 0) = f(0) = \frac{\binom{2}{0}\binom{5}{3}}{\binom{7}{3}} = 0.29$$

(i) *What is the expected number of defective television sets purchased by the hotel?*

$$E(X) = n \times \frac{k}{N} = 3 \times \frac{2}{7} = \frac{6}{7}$$

(ii) *Find the variance of X.*

$$Var(X) = n \times \frac{k}{N}\left(1 - \frac{k}{N}\right)\left(\frac{N-n}{N-1}\right) = 3 \times \frac{2}{7}\left(\frac{5}{7}\right)\left(\frac{7-3}{7-1}\right) = 0.41$$

Q2. Suppose that a family has 5 children, 3 of them are girls and the rest are boys. A sample of 2 children is selected randomly and without replacement.

$$N = 5 \quad , \quad n = 2 \quad , \quad k = 3$$

$$f(x) = \frac{\binom{3}{x}\binom{2}{2-x}}{\binom{5}{2}} \quad ; \quad x = 0,1,2$$

a. *The probability that no girls are selected is*

$$P(X = 0) = f(0) = \frac{\binom{3}{0}\binom{2}{2}}{\binom{5}{2}} = 0.1$$

b. *The probability that <u>at most one</u> girls are selected is*

$$P(X \leq 1) = f(0) + f(1) = \frac{\binom{3}{0}\binom{2}{2}}{\binom{5}{2}} + \frac{\binom{3}{1}\binom{2}{1}}{\binom{5}{2}} = 0.7$$

c. *The expected number of girls in the sample is*

$$E(X) = n \times \frac{k}{N} = 2 \times \frac{3}{5} = \frac{6}{5}$$

d. *The variance of the number of girls in the sample is*

$$Var(X) = n \times \frac{k}{N}\left(1 - \frac{k}{N}\right)\left(\frac{N-n}{N-1}\right) = 2 \times \frac{3}{5}\left(1 - \frac{3}{5}\right)\left(\frac{5-2}{5-1}\right) = 0.36$$

Q3. A random committee of size 4 is selected from 2 chemical engineers and 8 industrial engineers.

(1) Write a formula for the probability distribution function of the random variable X representing the number of chemical engineers in the committee.

$$N = 10 \quad , \quad n = 4 \quad , \quad k = 2$$

$$f(x) = \frac{\binom{2}{x}\binom{8}{4-x}}{\binom{10}{4}} \quad ; \quad x = 0,1,2$$

(2) Find the probability that there will be <u>no chemical engineers</u> in the committee.

$$P(X = 0) = f(0) = \frac{\binom{2}{0}\binom{8}{4}}{\binom{10}{4}} = 0.33$$

(3) Find the probability that there will be <u>at least one chemical</u> engineer in the committee.

$$P(X \geq 1) = 1 - P(X < 1) = 1 - f(0) = 0.67$$

(4) What is the expected number of chemical engineers in the committee?

$$E(X) = n \times \frac{k}{N} = 4 \times \frac{2}{10} = 0.8$$

(5) What is the variance of the number of chemical engineers in the committee?

$$Var(X) = n \times \frac{k}{N}\left(1 - \frac{k}{N}\right)\left(\frac{N-n}{N-1}\right) = 4 \times \frac{2}{10}\left(1 - \frac{2}{10}\right)\left(\frac{10-4}{10-1}\right) = 0.43$$

Q9. A shipment of 20 digital voice recorders contains 5 that are defective. If 10 of them are randomly chosen (without replacement) for inspection, then:

(1) The probability that 2 will be defective is:

$$N = 20 \quad , \quad n = 10 \quad , \quad k = 5$$

$$f(x) = \frac{\binom{5}{x}\binom{15}{10-x}}{\binom{20}{10}} \quad ; \quad x = 0,1,2,3,4,5$$

$$P(X = 2) = f(2) = \frac{\binom{5}{2}\binom{15}{8}}{\binom{20}{10}} = 0.35$$

(2) The probability that at most 1 will be defective is:

$$P(X \leq 1) = f(0) + f(1) = \frac{\binom{5}{0}\binom{15}{10}}{\binom{20}{10}} + \frac{\binom{5}{1}\binom{15}{9}}{\binom{20}{10}} = 0.15$$

(3) The expected number of defective recorders in the sample is:

$$E(X) = n \times \frac{k}{N} = 10 \times \frac{5}{20} = \frac{5}{2}$$

(4) The variance of the number of defective recorders in the sample is:

$$Var(X) = n \times \frac{k}{N}\left(1 - \frac{k}{N}\right)\left(\frac{N-n}{N-1}\right) = 10 \times \frac{5}{20}\left(\frac{15}{20}\right)\left(\frac{20-10}{20-1}\right) = 0.99$$

<u>***Poisson distribution:***</u>

$$f(x) = \frac{e^{-\lambda}\ \lambda^x}{x!} \quad ; \quad x = 0,1,2,\ldots$$

$$E(X) = Var(X) = \lambda$$

Q1. On average, a certain intersection results in 3 traffic accidents per day. Assuming Poisson distribution,

(iii) What is the probability that at this intersection:

(1) <u>No accidents</u> will occur in a given <u>day</u>?

$$\lambda_{one\ day} = 3$$

$$f(x) = \frac{e^{-3}(3)^x}{x!} \ ; \quad x = 0,1,2,\ldots$$

$$P(X = 0) = f(0) = \frac{e^{-3}(3)^0}{0!} = 0.05$$

(2) <u>More than 3</u> accidents will occur in a given <u>day</u>?

$$P(X > 3) = 1 - P(X \le 3) = 1 - [f(0) + f(1) + f(2) + f(3)]$$

$$= 1 - [\frac{e^{-3}(3)^0}{0!} + \frac{e^{-3}(3)^1}{1!} + \frac{e^{-3}(3)^2}{2!} + \frac{e^{-3}(3)^3}{3!}] = 0.35$$

(3) Exactly 5 accidents will occur in a period of two days?

$$\lambda_{two\ days} = 6$$

$$f(x) = \frac{e^{-6}(6)^x}{x!} \quad ; \quad x = 0, \ldots\ldots\infty$$

$$P(X = 5) = f(5) = \frac{e^{-6}(6)^5}{5!} = 0.16$$

(iv) What is the average number of traffic accidents in a period of 4 days?

$$E(X) = \lambda_{4\ days} = 4 \times 3 = 12$$

Q3. Suppose that the number of telephone calls received per day has a Poisson distribution with mean of 4 calls per day.

(a). The probability that <u>2 calls</u> will be received in a given <u>day</u> is

$$\lambda_{\,one\ day} = 4$$

$$f(x) = \frac{e^{-4}(4)^x}{x!} \quad ; \quad x = 0,1,2,\dots$$

$$P(X = 2) = f(2) = \frac{e^{-4}(4)^2}{2!} = 0.15$$

(b). The expected number of telephone calls received in a given week is

$$E(X) = \lambda_{week} = 4 \times 7 = 28$$

(c). The probability that <u>at least 2</u> calls will be received in a period of <u>12 hours</u> is

$$\lambda_{\,12hours} = 2$$

$$f(x) = \frac{e^{-2}(2)^x}{x!} \quad ; \quad x = 0,1,2,\dots$$

$$P(X \geq 2) = 1 - P(X < 2) = 1 - [f(0) + f(1)]$$

$$= 1 - \left[\frac{e^{-2}(2)^0}{0!} + \frac{e^{-2}(2)^1}{1!}\right] = 0.59$$

Q6. Suppose that X~ Binomial (1000,0.002). By using Poisson approximation, P(X=3) is approximately equal to

When we have a Binomial distribution with a small p and a large n, we can make an approximation to Poisson distribution:

$$p = 0.002 \;\;\&\;\; n = 1000 \;\;\&\;\; \lambda = np = 2$$

$$P(X = 3) = f(3) = \frac{e^{-2}2^3}{3!} = 0.18045$$

CHAPTER 2

- *Continuous Random Variables.*
- *The Uniform Distribution.*
- *The Exponential Distribution.*
- *Normal Distribution.*

Continuous Random Variables:

- $0 \le f(x) \le 1$
- $\int_{-\infty}^{\infty} f(x)dx = 1$
- $P(a < X < b) = \int_a^b f(x)dx$
- $E(X) = \int_{-\infty}^{\infty} x\, f(x)dx$
- $Var(X) = E(X^2) - E(X)^2$
- $E(X^2) = \int_{-\infty}^{\infty} x^2\, f(x)dx$
- $E(aX \pm b) = aE(X) \pm b$
- $Var(aX \pm b) = a^2 Var(X)$

Q1. Suppose $\boldsymbol{X}$ ***is a continuous random variable what is*** $\boldsymbol{P(X = 16)}$?

$$P(X = 16) = 0$$

$Note,\ for\ any\ continuous\ random\ variable, P(x = a) = 0$

Q2: Suppose we have the (p.d.f): $\boldsymbol{f(x) = k\sqrt{x}\ \ ,\ \ 0 < x < 1}$

(1): Find the value of k

$$\int_{-\infty}^{\infty} f(x)dx = 1 \Rightarrow \int_0^1 k\sqrt{x}\, dx = 1$$

$$\int_0^1 k\sqrt{x}\, dx = 1 \Rightarrow \int_0^1 kx^{0.5}\, dx = 1 \Rightarrow k\left[\frac{x^{1.5}}{1.5}\right]_0^1 = 1 \Rightarrow \frac{k}{1.5} = 1 \Rightarrow \boldsymbol{k = 1.5}$$

$$then, \qquad \boxed{f(x) = 1.5\sqrt{x}\ \ ,\ \ 0 < x < 1}$$

(2): Find the probability $P(0.3 < X < 0.6)$*:*

$$P(0.3 < X < 0.6) = \int_{0.3}^{0.6} f(x)dx$$

$$\int_{0.3}^{0.6} 1.5x^{0.5}dx = 1.5\left[\frac{x^{1.5}}{1.5}\right]_{0.3}^{0.6} = [x^{1.5}]_{0.3}^{0.6} = 0.6^{1.5} - 0.3^{1.5} = \boldsymbol{0.3004}$$

(3):Find $E(X)$*:*

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)dx$$

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)dx = \int_0^1 1.5x^{1.5}dx = 1.5\left[\frac{x^{2.5}}{2.5}\right]_0^1 = \frac{1.5}{2.5}[x^{2.5}]_0^1 = \boldsymbol{0.6}$$

Q3: Suppose we have the (p.d.f):

$$\boldsymbol{f(x) = 3x^2 \quad , \quad 0 < x < 1}$$

1. ***Find the mean μ:***

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)dx = \int_0^1 3x^3 dx = 3\int_0^1 x^3 dx = \frac{3}{4}\,[x^4]_0^1 = \frac{3}{4}$$

2. $\boldsymbol{P(X > 0.5) =}$

$$= \int_{0.5}^{1} f(x)dx = \int_{0.5}^{1} 3x^2\, dx = 3\int_{0.5}^{1} x^2\, dx = \frac{3}{3}\,[x^3]_{0.5}^{1} = 0.875$$

3. $\boldsymbol{P(0.4 < X < 0.6) =}$

$$= \int_{0.4}^{0.6} f(x)dx = \int_{0.4}^{0.6} 3x^2\, dx = 3\int_{0.4}^{0.6} x^2\, dx = \frac{3}{3}\,[x^3]_{0.4}^{0.6} = 0.154$$

Q4: Suppose we have the (p.d.f):

$$\boldsymbol{f(x) = c(4 - x)} \quad , \qquad \boldsymbol{-2 < x < 2}$$

1. ***Find the value of c:***

$$f(X) = \int_{-\infty}^{\infty} f(x)dx = 1$$

$$\Longrightarrow \int_{-2}^{2} c(4 - x)\, dx = 1$$

$$\Longrightarrow c \int_{-2}^{2} 4 - x\, dx = 1$$

$$\Longrightarrow c \quad \left[4x - \frac{x^2}{2}\right]_{-2}^{2} = 1$$

$$\Longrightarrow c \quad \left[\left(8 - \frac{4}{2}\right) - \left(-8 - \frac{4}{2}\right)\right] = 1$$

$$\Longrightarrow c \quad [16] = 1$$

$$\Longrightarrow c \quad = \frac{1}{16}$$

2. $\boldsymbol{P(X > 0) =}$

$$= \int_{0}^{2} f(x)dx = \frac{1}{16}\int_{0}^{2} 4 - x\, dx = \frac{1}{16}\left[4x - \frac{x^2}{2}\right]_{0}^{2} = \frac{1}{16}\left[\left(8 - \frac{4}{2}\right)\right] = \frac{6}{16}$$

3. $\boldsymbol{P(X < 3) = 1}$
4. $\boldsymbol{P(X < -3) = 0}$

5. ***Find*** $\boldsymbol{E(X)}$

$$= \int_{-2}^{2} x\, f(x)dx = \frac{1}{16}\int_{-2}^{2} 4x - x^2 \; dx = \frac{1}{16}\left[2\,x^2 - \frac{x^3}{3}\right]_{-2}^{2} = -\frac{1}{3}$$

The Uniform Distribution:

$$f(x) = \frac{1}{b-a} \quad ; \quad a < x < b$$

$$* \, E(X) = \frac{a+b}{2} \qquad * \, Var(X) = \frac{(b-a)^2}{12}$$

Q1. If the random variable X has a uniform distribution (0,10), then

$$f(x) = \frac{1}{10} \quad ; \; 0 < x < 10$$

(1):Find $P(X < 6)$

$$P(X < 6) = \int_0^6 \frac{1}{10} \, dx = \frac{1}{10} [x]_0^6 = \frac{1}{10} [6-0] = \frac{6}{10}$$

(2): The mean of X is

$$E(X) = \frac{a+b}{2} = \frac{0+10}{2} = \frac{10}{2} = 5$$

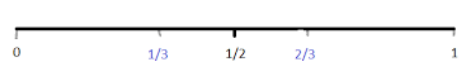
(3): The variance X is

$$Var(X) = \frac{(b-a)^2}{12} = \frac{(10-0)^2}{12} = \frac{100}{12} = 8.33$$

Q2. Suppose that the random variable X has the following uniform distribution:

$$f(x) = 3 \;\; ; \;\; \frac{2}{3} < x < 1$$

1. $P(0.33 < X < 0.5) = 0$
2. $P(X > 1.25) = 0$



$$Var(X) = \frac{(b-a)^2}{12} = \frac{\left(1 - {}^2/_3\right)^2}{12} = 0.00926$$

Q3. Suppose that the continuous random variable X has the following Probability density function (pdf)

$$f(x) = 0.2 \quad ; \quad 0 < x < 5$$

(1): ***Find*** $\boldsymbol{P(X > 1)}$

$$P(X > 1) = \int_1^5 0.2\, dx = 0.2[x]_1^5 = 0.2[5 - 1] = 0.8$$

(2): ***Find***$\boldsymbol{P(X \geq 1)}$

$P(X \geq 1) = 0.8$, *the same as part (1) because we had a continuous random variable, and hence,* $P(X = 1) = 0$

*(3):****Find*** $\boldsymbol{E(X)}$

$$E(X) = \frac{a + b}{2} = \frac{0 + 5}{2} = 2.5$$

(4): ***Find*** $\boldsymbol{Var(X)}$:

$$Var(X) = \frac{(b - a)^2}{12} = \frac{(5 - 0)^2}{12} = \frac{25}{12} = 2.0833$$

The Exponential Distribution:

$$f(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}} \quad ; \quad x > 0$$

$$* \; E(X) = \beta \qquad\qquad * \; Var(X) = \beta^2$$

Q1. If the random variable X has an exponential distribution with the mean 4, then:

$$X \sim expo(4) \quad , \quad f(x) = \frac{1}{4} e^{-\frac{x}{4}} \quad ; \quad x > 0$$

*(1): **Find $P(X < 8)$***

$$P(X < 8) = \int_0^8 \frac{1}{4} e^{-\frac{x}{4}} dx = \frac{1}{4} \int_0^8 e^{-\frac{x}{4}} dx = \frac{1}{4} \left[-4e^{-\frac{x}{4}} \right]_0^8 = 1 - e^{-2} = 0.8647$$

*(2): **The variance of X is***

$$Var(X) = \beta^2 = 4^2 = 16$$

Q2. The lifetime of a specific battery is a random variable X with probability density function given by:

$$f(x) = \frac{1}{200} e^{-\frac{x}{200}} \quad ; \quad x > 0$$

*(1): **The mean life time of the battery equals to***

$$E(X) = \beta = 200$$

*(2): **$P(X > 100)$ is***

$$P(X > 100) = \int_{100}^{\infty} \frac{1}{200} e^{-\frac{x}{200}} dx = \frac{1}{200} \int_{100}^{\infty} e^{-\frac{x}{200}} dx$$

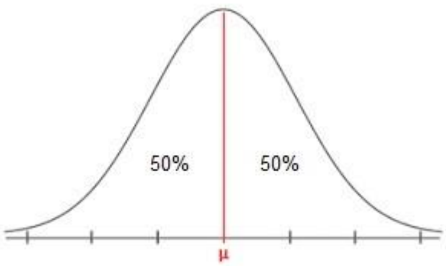
$$= \frac{1}{200} \left[-200e^{-\frac{x}{200}} \right]_{100}^{\infty} = e^{-0.5} = 0.6065$$

*(3): **$P(X = 200)$ is***

$P(X = 200) = 0$, *because we had a continuous random variable, and hence,*

$$P(X = a) = 0$$

The Normal Distribution:



Normal distribution $X \sim N(\mu, \sigma)$

Standard Normal $Z \sim N(0\ , 1)$

Q1. Suppose that Z is distributed according to the standard normal distribution.

1) *The area under the curve to the left of is* $Z = 1.43$:

$$P(Z < 1.43) = 0.9236$$

2) *The area under the curve to the left of is* $Z = 1.39$:

$$P(Z < 1.39) = 0.9177$$

3) *The area under the curve to the right of is* $Z = -0.89$:

$$P(Z > -0.89) = 1 - P(Z < -0.89) = 1 - 0.1867 = 0.8133$$

4) *The area under the curve between* $Z = -2.16$ *and* $Z = -0.65$ *is:*

$$P(-2.16 < Z < -0.65)$$
$$= P(Z < -0.65) - P(Z < -2.16)$$
$$= \quad 0.2578 \quad - \quad 0.0154 \quad = 0.2424$$

5) *The value of* k *such that is* $P(0.93 < Z < k)$:

$$P(0.93 < Z < k) \quad = \ 0.0427$$
$$P(Z < k) - P(Z < 0.93) \ = 0.0427$$
$$P(Z < k) - 0.8238 \ = 0.0427$$
$$P(Z < k) = 0.8665$$

P (Z < اطراف الجدول) = داخل الجدول $\qquad \Rightarrow k = 1.11$

Q2. The finished inside diameter of a piston ring is normally distributed with a mean of 12 centimeters (c.m) and a standard deviation of 0.03 centimeter. Then,

1) *The proportion of rings that will have inside diameter less than 12.05 is:*

$$X \sim N(\mu, \sigma)$$
$$X \sim N(12, 0.03)$$
$$P(X < 12.05) = P\left(Z < \frac{12.05-\mu}{\sigma}\right)$$
$$= P\left(Z < \frac{12.05-12}{0.03}\right)$$
$$= P(Z < 1.67) = 0.9525$$

2) *The proportion of rings that will have inside diameter exceeding 11.97 is:*

$$P(X > 11.97) = P\left(Z > \frac{11.97-\mu}{\sigma}\right)$$
$$= P\left(Z > \frac{11.97-12}{0.03}\right)$$
$$= P(Z > -1)$$
$$= 1 - P(Z < -1)$$
$$= 1 - 0.1587 = 0.8413$$

3) *The probability that a piston ring will have an inside diameter between 11.95 and 12.05 is:*

$$P(11.95 < X < 12.05)$$
$$= P\left(\frac{11.95-12}{0.03} < Z < \frac{12.05-12}{0.03}\right)$$
$$= P(-1.67 < Z < 1.67)$$
$$= P(Z < 1.67) - P(Z < -1.67)$$
$$= 0.9525 - 0.0475 = 0.905$$

Q6. The weight of a large number of fat persons is nicely modeled with a normal distribution with mean of 128 kg and a standard deviation of 9 kg.

$$X \sim N(\mu, \sigma)$$
$$X \sim N(128, 9)$$

(1): The percentage of fat persons with weights at most 110 kg is

$$P(X \leq 110) = P\left(Z < \frac{110-128}{9}\right)$$
$$= P(Z < -2) = 0.0228$$

(2): The percentage of fat persons with weights more than 149 kg is

$$P(X > 149) = P\left(Z > \frac{149-128}{9}\right)$$
$$= 1 - P(Z < 2.33)$$
$$= 1 - 0.9901 = 0.0099$$

(3): The weight x above which 86% of those persons will be

$$P(X > x) = 0.86$$
$$1 - P(X < x) = 0.86$$
$$P(X < x) = 0.14$$
$$P\left(Z < \frac{x-128}{9}\right) = 0.14$$

By searching inside the table for 0.14, and transforming X to Z, we got:

$$\frac{x-128}{9} = -1.08 \Rightarrow x = 118.28$$

(4): The weight x below which 50% of those persons will be

$P(X < x) = 0.5$, *by searching inside the table for 0.5, and transforming X to Z*

$$\frac{x-128}{9} = 0 \Rightarrow x = 128$$

Q8. If the random variable X has a normal distribution with the mean μ and the variance σ^2, then $P(X < \mu + 2\sigma)$ equals to

$$P(X < \mu + 2\sigma) = P\left(Z < \frac{(\mu + 2\sigma) - \mu}{\sigma}\right) = P(Z < 2) = 0.9772$$

Q9. If the random variable X has a normal distribution with the mean μ and the variance 1, and if P(X<3)=0.877, then μ equals to

$$Given\ that\ \ \sigma = 1$$

$$P(X < 3) = 0.877 \Rightarrow P\left(Z < \frac{3 - \mu}{1}\right) = 0.877$$

$$3 - \mu = 1.16 \Rightarrow \mu = 1.84$$

Q10. Suppose that the marks of the students in a certain course are distributed according to a normal distribution with the mean 70 and the variance 25. If it is known that 33% of the student failed the exam, then the passing mark x is

$$X \sim N(70{,}5)$$

$$P(X < x) = 0.33 \Rightarrow P\left(Z < \frac{x - 70}{5}\right) = 0.33$$

By searching inside the table for 0.33, and transforming X to Z, we got:

$$\frac{x - 70}{5} = -0.44 \Rightarrow x = 67.8$$

	$f(x)$ or PDF		$E(X)$	$Var(X)$
Binomial	$\binom{n}{x} p^x q^{n-x}$	$x = 0, 1, 2, \ldots., n$	np	npq
Hyper geometric	$\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$	$x = 0,1, \ldots. min(n, k)$	$n \times \frac{k}{N}$	$n \times \frac{k}{N}\left(1 - \frac{k}{N}\right)\left(\frac{N-n}{N-1}\right)$
Poisson	$\frac{\lambda^x e^{-\lambda}}{x!}$	$x = 0, 1, 2, \ldots.$	λ	λ
Continuous Uniform	$\frac{1}{b-a}$	$a < X < b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential	$\frac{1}{\beta} e^{-\frac{x}{\beta}}$	$X > 0$	β	β^2

CHAPTER 3

- *Sampling Distribution :*
  - *Single mean.*
  - *Two means.*
  - *Single proportion.*
  - *Two proportions.*

- *The Student (T) Distribution.*
- *Chi-square* (χ^2) *Distribution.*
- *Fisher (F) Distribution.*

Sampling Distribution

Sampling Distribution: Single Mean

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

$$* \ E(\bar{X}) = \bar{X} = \mu \qquad * \ Var(\bar{X}) = \frac{\sigma^2}{n}$$

Q1. The average life of a certain battery is 5 years, with a standard deviation of 1 year. Assume that the live of the battery approximately follows a normal distribution.

1) *The sample mean X of a random sample of 5 batteries selected from this product has a mean $E(\bar{X}) = \mu$ equal to:*

$$\mu = 5 \ \ ; \ \ \sigma = 1 \ \ ; \ \ n = 5$$

$$E(\bar{X}) = \mu = 5$$

2) *The variance $Var(\bar{X})$ of the sample mean X of a random sample of 5 batteries selected from this product is equal to:*

$$Var(\bar{X}) = \frac{\sigma^2}{n} = \frac{1}{5} = 0.2$$

4) *The probability that the average life of a random sample of size 16 of such batteries will be between 4.5 and 5.4 years is:*

$$n = 16 \rightarrow \frac{\sigma}{\sqrt{n}} = \frac{1}{4}$$

$$P(4.5 < \bar{X} < 5.4) = P\left(\frac{4.5-\mu}{\frac{\sigma}{\sqrt{n}}} < Z < \frac{5.4-\mu}{\frac{\sigma}{\sqrt{n}}}\right) = P\left(\frac{4.5-5}{\frac{1}{4}} < Z < \frac{5.4-5}{\frac{1}{4}}\right)$$

$$= P(-2 < Z < 1.6)$$

$$= P(Z < 1.6) - P(Z < -2)$$

$$= 0.9452 - 0.0228 = 0.9224$$

5) *The probability that the average life of a random sample of size 16 of such batteries will be less than 5.5 years is:*

$$P(\bar{X} < 5.5) = P\left(Z < \frac{5.5-\mu}{\frac{\sigma}{\sqrt{n}}}\right) = P\left(Z < \frac{5.5-5}{1/4}\right) = P(Z < 2) = 0.9772$$

6) *If* $P(\bar{X} > a) = 0.1492$ *where X represents the sample mean for a random sample of size 9 of such batteries, then the numerical value of a is:*

$P(\bar{X} > a) = 0.1492 \quad ; \quad n = 9$

$$P\left(Z > \frac{a-\mu}{\frac{\sigma}{\sqrt{n}}}\right) = 0.1492$$

$$\Rightarrow 1 - P\left(Z < \frac{a-5}{\frac{1}{3}}\right) = 0.1492$$

$$\Rightarrow P\left(Z < \frac{a-5}{\frac{1}{3}}\right) = 0.8508$$

$$\frac{a-5}{\frac{1}{3}} = 1.04$$

$$a = 5 + \frac{1.04}{3} = 5.347$$

Q. Suppose that you take a random sample of size $n = 64$ from a distribution with mean $\mu = 55$ and standard deviation $\sigma = 10$. Let X be the sample mean.

(a) What is the approximated sampling distribution of X ?

$$\mu = 55 \quad ; \quad \sigma = 10 \quad ; \quad n = 64$$

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) = \bar{X} \sim N\left(55, \frac{10}{8}\right)$$

(b) What is the mean of X ?

$$E(\bar{X}) = \mu = 55$$

(c) What is the standard error (standard deviation) of X ?

$$S.D(\bar{X}) = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{64}} = \frac{10}{8}$$

(d) Find the probability that the sample mean X exceeds 52.

$$(a) P(\bar{X} > 52) = P\left(Z > \frac{52-55}{\frac{10}{8}}\right)$$

$$= P(Z > -2.4)$$

$$= 1 - P(Z < -2.4)$$

$$= 1 - 0.0082 = 0.9918$$

<u>Sampling Distribution: Two Means</u>

$$* \bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2 \ , \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

$$* E(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2 \qquad * Var(\bar{X}_1 - \bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Q. A random sample of size $n_1 = 36$ is taken from a normal population with a mean $\mu_1 = 70$ and a standard deviation $\sigma_1 = 4$. A second independent random sample of size $n_2 = 49$ is taken from a normal population with a mean $\mu_2 = 85$ and a standard deviation $\sigma_2 = 5$. Let $\bar{X}_1$ and $\bar{X}_2$ be the averages of the first and second samples, respectively.

$$n_1 = 36 , \mu_1 = 70 , \sigma_1 = 4$$

$$n_2 = 49 , \mu_2 = 85 , \sigma_2 = 5$$

a. *Find $E(\bar{X}_1 - \bar{X}_2)$:*

$$E(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2 = 70 - 85 = -15$$

a. *Find $Var(\bar{X}_1 - \bar{X}_2)$:*

$$Var(\bar{X}_1 - \bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \frac{16}{36} + \frac{25}{49} = 0.955$$

$$S.D(\bar{X}_1 - \bar{X}_2) = \sqrt{0.955}$$

b. *Find $P(\bar{X}_1 - \bar{X}_2 > -16)$:*

$$= P\left(Z > \frac{-16-(-15)}{\sqrt{0.955}}\right) = 1 - P\left(Z < \frac{-16-(-15)}{\sqrt{0.955}}\right)$$

$$= 1 - P(Z < -1.02) = 0.8461$$

Q. The distribution of heights of a certain breed of terrier has a mean of 72 centimeters and a standard deviation of 10 centimeters, whereas the distribution of heights of a certain breed of poodle has a mean of 28 centimeters with a standard deviation of 5 centimeters. Assuming that the sample means can be measured to any degree of accuracy, find the probability that the sample mean for a random sample of heights of 64 terriers exceeds the sample mean for a random sample of heights of 100 poodles by <u>less than 44.2</u> centimeter

$$n_1 = 64\,,\ \ \mu_1 = 72\,,\ \sigma_1 = 10$$

$$n_2 = 100\,,\mu_2 = 28\,,\sigma_2 = 5$$

$$E(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2 = 72 - 28 = 44$$

$$Var(\bar{X}_1 - \bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \frac{100}{64} + \frac{25}{100} = 1.8125$$

$$P(\bar{X}_1 - \bar{X}_2 < 44.2) =$$

$$= P\left(Z < \frac{44.2 - (44)}{\sqrt{1.8125}}\right)$$

$$= \quad P(Z < 0.15)$$

$$= \quad 0.5596$$

<u>Sampling Distribution: Single Proportion</u>

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

$$* E(\hat{p}) = p \qquad * Var(\hat{p}) = \frac{pq}{n}$$

Q. Suppose that you take a random sample of size n=100 from a population with proportion of diabetic equal to p =0.25. Let p^ be the sample proportion of of diabetic.

(a) What is the mean and the standard error of $\hat{p}$?

$$p = 0.25 \quad ; \quad n = 100$$

$E(\hat{p}) = p = 0.25$

$Var(\hat{p}) = \frac{pq}{n} = \frac{0.25 \times 0.75}{100} = 0.003$

(b) What is the approximated sampling distribution of $\hat{p}$?

$$\hat{p} \sim N(0.25, \sqrt{0.003})$$

(c) Find the probability that the sample proportion $\hat{p}$ is less than 0.2.

$$P(\hat{p} < 0.2) = P\left(Z < \frac{0.2 - 0.25}{\sqrt{0.003}}\right) = P(Z < -0.91) = 0.1814$$

<u>Sampling Distribution: Two Proportions</u>

$$* \ \hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2, \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}\right)$$

$$* \ E(\hat{p}_1 - \hat{p}_2) = p_1 - p_2 \qquad * \ Var(\hat{p}_1 - \hat{p}_2) = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$$

Q. Suppose that 25% of the male students and 20% of the female students in a certain university smoke cigarettes. A random sample of 5 male students is taken. Another random sample of 10 female students is independently taken from this university. Let and be the proportions of smokers in the two samples, respectively.

$$p_1 = 0.25 \ ; \quad n_1 = 5$$

$$p_2 = 0.20 \ ; \quad n_2 = 10$$

(1): $E(\hat{p}_1 - \hat{p}_2) = p_1 - p_2 = 0.25 - 0.2 = 0.05$

(2): $Var(\hat{p}_1 - \hat{p}_2) = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2} = \frac{0.25 \times 0.75}{5} + \frac{0.2 \times 0.8}{10} = 0.054$

(3): $\hat{p}_1 - \hat{p}_2 \sim N\left(0.05, \sqrt{0.054}\right)$

(4): $P(0.1 < \hat{p}_1 - \hat{p}_2 < 0.2) = \left(\frac{0.1 - 0.05}{\sqrt{0.054}} < Z < \frac{0.2 - 0.05}{\sqrt{0.054}}\right)$

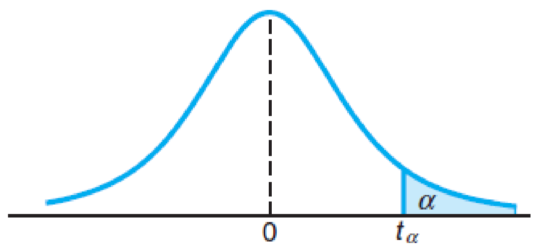
$$= (0.22 < Z < 0.65)$$

$$P(Z < 0.65) - P(Z < 0.22)$$

$$= 0.7422 - 0.5871 = 0.1551$$

Sampling Distribution		
single Population	Mean	$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$
	Proportion	$\hat{p} \sim N\left(p, \sqrt{\frac{pq}{n}}\right)$
Two Populations	Means	$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2\ , \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$
	Proportions	$\hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2, \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}\right)$

The student (t) Distribution:



$$t_{v,\alpha} \Longrightarrow P\left(T > t_{v,\alpha}\right) = \alpha$$

$$\Longrightarrow P\left(T < t_{v,\alpha}\right) = 1 - \alpha \qquad ; v = n - 1$$

- $If\ P(T < t_{22}) = 0.99 \Longrightarrow P(T > t_{22}) = 0.01$
  $\Longrightarrow t_{22,0.01} = 2.508$

- $If\ P\left(T > t_{18,0.975}\right) = 0.975$
  $\Longrightarrow t_{18,0.975} = -t_{18,0.025} = -2.101$

- $If\ P\left(T > t_{v,\alpha}\right) = \alpha\ where\ v = 24, \alpha = 0.995$
  $\Longrightarrow t_{24,0.995} = -t_{24,0.005} = -2.797$

- $If\ P\left(T > t_{v,\alpha}\right) = \alpha\ where\ v = 7, \alpha = 0.975$
  $\Longrightarrow t_{7,0.975} = -t_{7,0.025} = -2.365$

Q2. A random sample of size 15 selected from a normal distribution with <u>unknown variance</u>, and let $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ ***and k such that*** $P(1.761 < T < k) = 0.045$ ***,then the value of k is:***

$$t_{14,0.05} = 1.761 \ \ (from\ "t"\ table\ )$$

$$P\left(t_{14,0.05} < T < t_{14,\alpha}\right) = 0.045$$

$$\Longrightarrow P\left(T < t_{14,\alpha}\right) - P\left(T < t_{14,0.05}\right) = 0.045$$

$$\Longrightarrow P\left(T < t_{14,\alpha}\right) - \quad 0.95 \quad = 0.045$$

$$\Longrightarrow P\left(T < t_{14,\alpha}\right) = 0.995$$

$$\Longrightarrow P\left(T > t_{14,\alpha}\right) = 0.005$$

$$\Longrightarrow k = t_{14,\alpha} = 2.977$$

Q3. A random variable T with degree 17 and $P(-1.333 < T < k) = 0.75$ ***,then the value of k is:***

$$t_{17,0.90} = -1.333\ (from\ "t"\ table\ )$$

$$P\left(t_{17,0.90} < T < t_{17,\alpha}\right) = 0.75$$

$$\Longrightarrow P\left(T < t_{17,\alpha}\right) - P\left(T < t_{17,0.90}\right) = 0.75$$

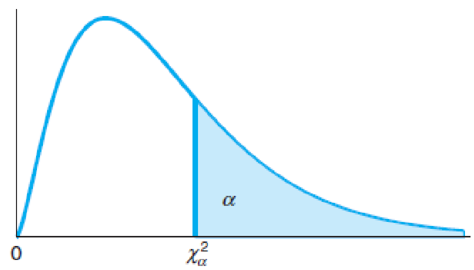
$$\Longrightarrow P\left(T < t_{17,\alpha}\right) - \quad 0.10 \quad = 0.75$$

$$\Longrightarrow P\left(T < t_{17,\alpha}\right) = 0.85$$

$$\Longrightarrow P\left(T > t_{17,\alpha}\right) = 0.15$$

$$\Longrightarrow k = t_{17,\alpha} = 1.069$$

Chi-square Distribution:



$$* \ E(\chi^2) = v \qquad\qquad * \ Var(\chi^2) = 2v$$

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{v=(n-1)} \qquad\qquad \chi^2_{v,\alpha} \Rightarrow P\left(\chi^2 > \chi^2_{v,\alpha}\right) = \alpha$$

$$\Rightarrow P\left(\chi^2 < \chi^2_{v,\alpha}\right) = 1 - \alpha$$

Q. For a chi-squared distribution:

(1) ***the value*** $\boldsymbol{\chi^2_\alpha}$ ***such that*** $\boldsymbol{P(\chi^2 > \chi^2_\alpha) = 0.01}$ ***when*** $\boldsymbol{v = 21}$ ***is:***

$$\Rightarrow \chi^2_{v,\alpha} = \chi^2_{21,0.01} = 38.93$$

(2) ***the value*** $\boldsymbol{\chi^2_\alpha}$ ***such that*** $\boldsymbol{P(\chi^2 > \chi^2_\alpha) = 0.95}$ ***when*** $\boldsymbol{v = 14}$ ***is:***

$$\Rightarrow \chi^2_{v,\alpha} = \chi^2_{14,0.95} = 6.571$$

(3) ***the value*** $\boldsymbol{\chi^2_\alpha}$ ***such that*** $\boldsymbol{P(\chi^2_\alpha < X < 23.209) = 0.015}$ ***when*** $\boldsymbol{v = 10}$ ***is:***

$$\boxed{\chi^2_{0.01,10} = 23.209 \ (from \ "\chi^2" \ table)}$$

$$\Longrightarrow P\left(\chi^2_{\alpha,10} < X < \chi^2_{0.01,10}\right) = 0.015$$

$$\Longrightarrow P\left(X < \chi^2_{0.01,10}\right) - P\left(X < \chi^2_{\alpha,10}\right) = 0.015$$

$$\Longrightarrow \quad 0.99 \quad - P\left(X < \chi^2_{\alpha,10}\right) = 0.015$$

$$\Longrightarrow P\left(X < \chi^2_{\alpha,10}\right) = 0.975$$

$$\Longrightarrow P\left(X > \chi^2_{\alpha,10}\right) = 0.025$$

$$\chi^2_{0.025,10} = 20.483$$

(4) ***the value*** $\boldsymbol{\chi^2_\alpha}$ ***such that*** $\boldsymbol{P(6.57 < X < \chi^2_\alpha) = 0.85}$ ***when*** $\boldsymbol{v = 14}$ ***is:***

$$\boxed{\chi^2_{0.95,14} = 6.571 \ (from \ "\chi^2" \ table)}$$

$$\Longrightarrow P\left(\chi^2_{0.95,14} < X < \chi^2_{\alpha,14}\right) = 0.85$$

$$\Longrightarrow P\left(X < \chi^2_{\alpha,14}\right) - P\left(X < \chi^2_{0.95,14}\right) = 0.85$$

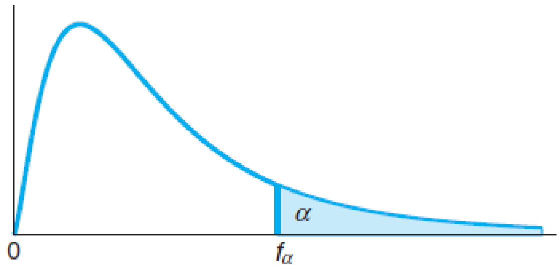
$$\Longrightarrow P\left(X < \chi^2_{\alpha,14}\right) - \quad 0.05 \quad = 0.85$$

$$\Longrightarrow P\left(X < \chi^2_{\alpha,14}\right) = 0.90$$

$$\Longrightarrow P\left(X > \chi^2_{\alpha,14}\right) = 0.10$$

$$\Longrightarrow \chi^2_{0.1,14} = 21.064$$

Fisher (F) Distribution



$* \; E(F) = \frac{v_1}{v_2-2}$ $\quad * \; Var(F) = \frac{2v_2^2(v_1+v_2-2)}{v_1(v_2-2)^2(v_2-4)} \; ; \; v_2 > 4$

$\frac{\chi^2_{v_1}}{\chi^2_{v_2}} \sim F_{v_1,v_2}$ $\quad F_{v_1,v_2,\alpha} = \frac{1}{F_{v_2,v_1,1-\alpha}}$ $\quad F_{v_1,v_2} \Rightarrow P\left(F > F_{v_1,v_2,\alpha}\right) = \alpha$

$\Rightarrow P\left(F < F_{v_1,v_2,\alpha}\right) = 1-\alpha$

Q.1. For the F distribution, then the value F such that $\boldsymbol{P\left(F > F_{8,6}\right) = 0.05}$ ***is:***

$$F_{8,6,0.05} = 4.15$$

Q.2. For the F distribution, then the value F such that $\boldsymbol{P\left(F > F_{8,6}\right) = 0.99}$ ***is:***

$$F_{8,6,0.99} = \frac{1}{F_{6,8,0.01}} = \frac{1}{6.37} = 0.157$$

Q.3.For the F distribution with degrees 8 and 6, then:

(a). the value a such that $P(F > a) = 0.95$

$$a = F_{8,6,0.95} = \frac{1}{F_{6,8,0.05}} = \frac{1}{3.58} = 0.2793$$

(b). the value b such that $P(F > b) = 0.01$

$$b = F_{8,6,0.01} = 8.10$$

CHAPTER 4

- *Estimation (point estimation and confidence interval):*
  - *Single mean*
  - *Two means*
  - *Single proportion*
  - *Two proportion*
  - *Single variance.*
  - *The ratio of two variances.*

Estimation and Confidence Interval

Estimation and Confidence Interval: Single Mean:

- *To find the confidence intervals for a single mean:*

  1- $\bar{X} \pm \left(Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$

  2- $\bar{X} \pm \left(t_{\frac{\alpha}{2},n-1} \frac{S}{\sqrt{n}}\right)$

- $Z_{\frac{\alpha}{2}} = a \Longrightarrow P(Z > a) = \frac{\alpha}{2}$

- *To estimate an error:*

$$e = Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

- *To estimate the sample size with particular error:*

$$n = \left(\frac{Z_{\frac{\alpha}{2}} \sigma}{e}\right)^2$$

Q1. Suppose that we are interested in making some statistical inferences about the mean, μ, of a normal population with standard deviation $\sigma = 2$. Suppose that a random sample of size $n = 49$ from this population gave a sample mean $\bar{X} = 4.5$.

$$\sigma = 2 \;\;\&\;\; \bar{X} = 4.5 \;\;\&\;\; n = 49$$

(1) The distribution of $\bar{X}$:

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \Longrightarrow \bar{X} \sim N\left(\mu, \frac{2}{\sqrt{49}}\right) \Longrightarrow \bar{X} \sim N\left(\mu, \frac{2}{7}\right)$$

(2) A good point estimate of μ is:

$$\hat{\mu} = \bar{X} = 4.5$$

(3) The standard error of $\bar{X}$ is:

$$S.E\,(\bar{X}) = \frac{2}{7} = 0.2857$$

(4) A 95% confidence interval for μ is:

$$\bar{X} \pm \left(Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$$

$$95\% \rightarrow \alpha = 0.05$$

$$Z_{\frac{\alpha}{2}} = Z_{0.025} = 1.96$$

$$4.5 \pm \left(1.96 \times \frac{2}{7}\right)$$

The 95% confidence interval is: $(3.94,\; 5.06)$

(5) If the upper confidence limit of a confidence interval is 5.2, then the lower confidence limit is

$\text{The upper limit} = 4.5 + e = 5.2$

$\text{and the lower limit} = 4.5 - e = ?$

$$4.5 + e = 5.2 \;\Rightarrow\; e = 5.2 - 4.5 \Rightarrow (e = 0.7)$$

$\text{Then, the lower limit} = 4.5 - e = 4.5 - 0.7 = 3.8$

(6) The confidence level of the confidence interval (3.88, 5.12) is

We have the interval: $(3.88, 5.12)$*:*

$$\text{The upper limit} = 5.12$$

$$\Rightarrow \bar{X} + \left(Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 5.12$$

$$\Rightarrow 4.5 + \left(Z_{\frac{\alpha}{2}} \times \frac{2}{7}\right) = 5.12$$

$$\Rightarrow Z_{\frac{\alpha}{2}} = 2.17 \Rightarrow \frac{\alpha}{2} = 1 - 0.985$$

$$\Rightarrow \frac{\alpha}{2} = 0.015 \Rightarrow \alpha = 0.03$$

Hence, the confidence level is 97%.

<u>Or, we can do the same thing with the lower limit:</u>

$$\text{The lower limit} = 3.88$$

$$\Rightarrow \bar{X} - \left(Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 3.88$$

$$\Rightarrow 4.5 - \left(Z_{\frac{\alpha}{2}} \times \frac{2}{7}\right) = 3.88$$

$$\Rightarrow Z_{\frac{\alpha}{2}} = 2.17 \Rightarrow \frac{\alpha}{2} = 1 - 0.985$$

$$\Rightarrow \frac{\alpha}{2} = 0.015 \Rightarrow \alpha = 0.03$$

Hence, the confidence level is 97%

(7) If we use $\overline{X}$ to estimate μ, then we are 95% confident that our estimation error will not exceed.

$95\% \rightarrow \alpha = 0.05 \quad \rightarrow \quad Z_{\frac{\alpha}{2}} = Z_{0.025} = 1.96$

$$e = \frac{Z_{\frac{\alpha}{2}} \times \sigma}{\sqrt{n}} = \frac{1.96 \times 2}{7} = 0.56$$

(8) If we want to be 95% confident that the estimation error will not exceed e=0.1 when we use $\overline{X}$ to estimate μ, then the sample size n must be equal to

$95\% \rightarrow \alpha = 0.05 \quad \rightarrow \quad Z_{\frac{\alpha}{2}} = Z_{0.025} = 1.96 \quad \& \quad e = 0.1$

$$n = \left(\frac{Z_{\frac{\alpha}{2}} \times \sigma}{e}\right)^2 = \left(\frac{1.96 \times 2}{0.1}\right)^2 = 1536.64 \simeq 1537$$

Q3. The following measurements were recorded for lifetime, in years, of certain type of machine: 3.4, 4.8, 3.6, 3.3, 5.6, 3.7, 4.4, 5.2, and 4.8. Assuming that the measurements represent a random sample from a normal population, then a 99% confidence interval for the mean life time of the machine is

$$99\% \rightarrow \alpha = 0.01 \rightarrow t_{\frac{\alpha}{2}, n-1} = t_{0.005, 8} = 3.355.$$

$$\bar{X} \pm \left(t_{\frac{\alpha}{2}, n-1} \times \frac{S}{\sqrt{n}}\right)$$

$$= 4.31 \pm \left(3.355 \times \frac{0.84}{3}\right)$$

$$\bar{X} = \frac{\sum x_i}{n} = 4.31$$

$$S^2 = \frac{\sum (x_i - \bar{X})^2}{n-1} = 0.71$$

$$S^2 = 0.71 \Rightarrow S = 0.84$$

The 99% confidence interval is: $(3.37, \ 5.25)$

Q4. A researcher wants to estimate the mean lifespan of a certain light bulbs. Suppose that the distribution is normal with standard deviation of 5 hours.

1. ***Determine the sample size needed on order that the researcher will be 90% confident that the error will not exceed 2 hours when he uses the sample mean as a point estimate for the true mean.***

$$\sigma = 5 \quad \& \quad e = 2 \quad \& \quad \alpha = 0.10$$

$$\alpha = 0.10 \quad \rightarrow \quad Z_{\frac{\alpha}{2}} = Z_{0.05} = 1.645$$

$$n = \left(\frac{Z_{\frac{\alpha}{2}} \times \sigma}{e}\right)^2 = \left(\frac{1.645 \times 5}{2}\right)^2 = 16.9 \simeq 17$$

2. ***Suppose that the researcher selected a random sample of 49 bulbs and found that the sample mean is 390 hours.***

(i) Find a good point estimate for the true mean μ.

$$\hat{\mu} = \bar{X} = 390$$

(ii) Find a 95% confidence interval for the true mean μ.

$$\sigma = 5 \quad \& \quad n = 49 \quad \& \quad \bar{X} = 390$$

$$\bar{X} \pm \left(Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$$

$$390 \pm \left(1.96 \frac{5}{\sqrt{49}}\right)$$

$$95\% \rightarrow \alpha = 0.05$$

$$Z_{\frac{\alpha}{2}} = Z_{0.025} = 1.96$$

The 95% confidence interval is: $(388.6, \ 391.3)$

Q5. The amount of time that customers using ATM (Automatic Teller Machine) is a random variable with a standard deviation of 1.4 minutes. If we wish to estimate the population mean μ by the sample mean , and if we want to be 96% confident that the sample mean will be within 0.3 minutes of the population mean, then the sample size needed is X

$$\sigma = 1.4 \ \ \& \ \ e = 0.3 \ \ \& \ \ \alpha = 0.04$$

$$\alpha = 0.04 \ \ \rightarrow \ \ Z_{\frac{\alpha}{2}} = Z_{0.02} = 2.055$$

$$n = \left(\frac{Z_{\frac{\alpha}{2}} \times \sigma}{e}\right)^2 = \left(\frac{2.055 \times 1.4}{0.3}\right)^2 = 91.9 \simeq 92$$

	population normal or not normal n large $(n \geq 30)$		population normal n small $(n < 30)$	
	σ known	σ unknown	σ known	σ unknown
Estimation	$\bar{X} \pm \left(Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$	$\bar{X} \pm \left(Z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right)$	$\bar{X} \pm \left(Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$	$\bar{X} \pm \left(t_{1-\frac{\alpha}{2},n-1} \times \frac{S}{\sqrt{n}}\right)$
testing	$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$	$Z = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$	$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$	$T = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$

<u>Estimation and Confidence Interval: Two Means</u>

To find the confidence intervals for two means:

1- $(\bar{X}_1 - \bar{X}_2) \pm \left(Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right)$

2- $(\bar{X}_1 - \bar{X}_2) \pm \left(t_{\frac{\alpha}{2}, n_1+n_2-2}\, Sp \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right)$

$$Sp^2 = \frac{S_1^2(n_1 - 1) + S_2^2(n_2 - 1)}{n_1 + n_2 - 2}$$

Q1.The tensile strength of type I thread is approximately normally distributed with standard deviation of 6.8 kilograms. A sample of 20 pieces of the thread has an average tensile strength of 72.8 kilograms. Then,

The tensile strength of type II thread is approximately normally distributed with standard deviation of 6.8 kilograms. A sample of 25 pieces of the thread has an average tensile strength of 64.4 kilograms. Then for the 98% confidence interval of the difference in tensile strength means between type I and type II, we have:

$$Theard\ 1: \quad n_1 = 20, \bar{X}_1 = 72.8, \sigma_1 = 6.8$$

$$Thread\ 2: \quad n_2 = 25, \bar{X}_2 = 64.4, \sigma_2 = 6.8$$

$$98\% \rightarrow \alpha = 0.02 \rightarrow \quad Z_{\frac{\alpha}{2}} = Z_{0.01} = 2.33$$

$$(\bar{X}_1 - \bar{X}_2) \pm \left(Z_{\frac{\alpha}{2}} \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right)$$

$$(72.8 - 64.4) \pm \left(2.33 \times \sqrt{\frac{6.8^2}{20} + \frac{6.8^2}{25}} \right)$$

$$8.4 \pm (2.33)(2.04) = (3.65, 13.15)$$

Q2.Two random samples were independently selected from two normal populations with equal variances. The results are summarized as follows.

	First sample	*Second sample*
Sample size (n)	*12*	*14*
Sample mean $(\bar{X})$	*10.5*	*10*
Sample variance (S^2)	*4*	*5*

1. ***Find a point estimate for*** $\boldsymbol{\mu_1 - \mu_2}$.

$$E(\bar{X}_1 - \bar{X}_2) = 10.5 - 10 = 0.5$$

2. ***Find 95% confidence interval for*** $\boldsymbol{\mu_1 - \mu_2}$.

$$\alpha = 0.05 \rightarrow \quad t_{\frac{\alpha}{2}, n_1+n_2-2} = t_{0.025\,,24} = 2.064$$

$$(\bar{X}_1 - \bar{X}_2) \pm \left(t_{\frac{\alpha}{2}, n_1+n_2-2}\, Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

$$(0.5) \quad \pm \left(2.064 \times 2.13 \times \sqrt{\frac{1}{12} + \frac{1}{14}}\right)$$

$$(-1.23\,,2.23)$$

$$Sp^2 = \frac{S_1^2(n_1 - 1) + S_2^2(n_2 - 1)}{n_1 + n_2 - 2}$$

$$= \frac{4(11)+5(13)}{24} = 4.54$$

$$Sp^2 = 4.54 \Rightarrow Sp = 2.13$$

Estimation and Confidence Interval: Single Proportion

$$* \text{ Point estimate for } P \text{ is: } \frac{x}{n}$$

$$* \text{ Interval estimate for } P \text{ is: } \hat{p} \pm \left(Z_{\frac{\alpha}{2}} \times \sqrt{\frac{\hat{p}\hat{q}}{n}} \right)$$

Q. A random sample of 200 students from a certain school showed that 15 students smoke. Let p be the proportion of smokers in the school.

1. ***Find a point Estimate for p.***

$$n = 200 \quad \& \quad x = 15$$

$$\hat{p} = \frac{x}{n} = \frac{15}{200} = 0.075 \rightarrow \hat{q} = 0.925$$

2. ***Find 95% confidence interval for p.***

$$95\% \rightarrow \alpha = 0.05 \quad \rightarrow \quad Z_{\frac{\alpha}{2}} = Z_{0.025} = 1.96$$

$$\hat{p} \pm \left(Z_{\frac{\alpha}{2}} \times \sqrt{\frac{\hat{p}\hat{q}}{n}} \right)$$

$$= 0.075 \pm \left(1.96 \times \sqrt{\frac{0.075 \times 0.925}{200}} \right)$$

The 95% confidence interval is: $(0.038, \ 0.112)$

Estimation and Confidence Interval: Two Proportions

$$* \text{ Point estimate for } P_1 - P_2 = \hat{p}_1 - \hat{p}_2 = \frac{x_1}{n_1} - \frac{x_2}{n_2}$$

$$* \text{ Interval estimate for } P_1 - P_2 \text{ is: } (\hat{p}_1 - \hat{p}_2) \pm \left(Z_{\frac{\alpha}{2}} \times \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} \right)$$

Q3. A random sample of 100 students from school "A" showed that 15 students smoke. Another independent random sample of 200 students from school "B" showed that 20 students smoke. Let p1 be the proportion of smokers in school "A" and p2 is the proportion of smokers in school "B".

(1) Find a point Estimate for $p_1 - p_2$.

$$n_1 = 100 \quad x_1 = 15 \quad \rightarrow \hat{p}_1 = \frac{15}{100} = 0.15$$

$$n_2 = 200 \quad x_2 = 20 \rightarrow \hat{p}_2 = \frac{20}{200} = 0.10$$

$$\hat{p}_1 - \hat{p}_2 = 0.15 - 0.1 = 0.05$$

(2) Find 95% confidence interval for $p_1 - p_2$.

$$95\% \rightarrow \alpha = 0.05 \quad \rightarrow \quad Z_{\frac{\alpha}{2}} = Z_{0.025} = 1.96$$

$$(\hat{p}_1 - \hat{p}_2) \pm \left(Z_{\frac{\alpha}{2}} \times \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} \right)$$

$$= (0.05) \pm \left(1.96 \times \sqrt{\frac{(0.15)(0.85)}{100} + \frac{(0.1)(0.9)}{200}} \right)$$

$$= 0.05 \pm \left(1.96 \times \sqrt{0.001725} \right)$$

The 95% confidence interval is: $(-0.031, \ 0.131)$

Confidence Interval for the Population Variance and Standard Deviation:

$$P\left(\frac{(n-1)S^2}{\chi^2_{(n-1),\frac{\alpha}{2}}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{(n-1),1-\frac{\alpha}{2}}}\right) = 1-\alpha$$

$$P\left(\frac{S_1^2}{S_2^2} \times \frac{1}{F_{v_1,v_2,\frac{\alpha}{2}}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} \times F_{v_2,v_1,\frac{\alpha}{2}}\right) = 1-\alpha$$

Q1. We are interested in the content of a soft-drink dispensing machine. A random sample of 25 drinks gave a variance of 2.03 deciliters^2. Assume that the contents are approximately normally distributed.

$$\boldsymbol{n = 25 \quad , \quad S^2 = 2.03}$$

(1) The point estimate of the population variance of the contents is:

$$S^2 = 2.03$$

(2) The lower bound of the of the 90 % confidence interval of the population variance σ^2 is:

$$\frac{(n-1)S^2}{\chi^2_{(n-1),\frac{\alpha}{2}}} = \frac{(25-1)\times 2.03}{\chi^2_{(25-1),\frac{0.1}{2}}} = \frac{(24)\times 2.03}{\chi^2_{(24),0.05}} = \frac{48.72}{36.415} = 1.34$$

(3) The upper bound of the of the 90 % confidence interval of the population variance σ^2 is:

$$\frac{(n-1)S^2}{\chi^2_{(n-1),1-\frac{\alpha}{2}}} = \frac{(25-1)\times 2.03}{\chi^2_{(25-1),1-\frac{0.1}{2}}} = \frac{(24)\times 2.03}{\chi^2_{(24),095}} = \frac{48.72}{13.848} = 3.52$$

Q2. In a series of experiments to determine the absorption rate of certain pesticides into skin, measured amounts of two pesticides were applied to several skin specimens. For pesticide A, the variance of the amounts absorbed in $\boldsymbol{n_1 = 6}$ specimens was $\boldsymbol{S_1^2 = 2.3}$, while for pesticide B, the variance of the amounts absorbed in $\boldsymbol{n_2 = 10}$ specimens was $\boldsymbol{S_2^2 = 0.6}$. Assume that for each pesticide, the amounts absorbed are a simple random sample from a normal population.

$$\boldsymbol{n_1 = 6 \quad , \ S_1^2 = 2.3}$$

$$\boldsymbol{n_2 = 10 \ , \ S_2^2 = 0.6}$$

1. *Find A point estimate of the ratio σ_1^2/σ_2^2 of the two population variances is:*

$$\frac{S_1^2}{S_2^2} = \frac{2.3}{0.6} = 3.833$$

2. *The <u>lower bound</u> of the of the 90 % confidence interval of the ratio σ_1^2/σ_2^2*

$$\frac{S_1^2}{S_2^2} \times \frac{1}{F_{v_1,v_2,\frac{\alpha}{2}}} = \frac{2.3}{0.6} \times \frac{1}{F_{5,9,\frac{0.1}{2}}} = 3.833 \times \frac{1}{F_{5,9,0.05}} = 3.833 \times \frac{1}{3.48} = 1.1$$

3. *The <u>upper bound</u> of the of the 90 % confidence interval of the ratio σ_1^2/σ_2^2*

$$\frac{S_1^2}{S_2^2} \times F_{v_2,v_1,\frac{\alpha}{2}} = \frac{2.3}{0.6} \times F_{9,5,\frac{0.1}{2}} = 3.833 \times F_{9,5,0.05} = 3.833 \times 4.77 = 18.28$$

Estimation and Confident intervals		
single Population	Mean	$\bar{X} \pm \left(Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$ σ *known*
		$\bar{X} \pm \left(t_{\frac{\alpha}{2},n-1} \frac{S}{\sqrt{n}}\right)$ σ *unknown*
	Proportion	$\hat{p} \pm \left(Z_{\frac{\alpha}{2}} \times \sqrt{\frac{\hat{p}\hat{q}}{n}}\right)$
	Variance	$\frac{(n-1)S^2}{\chi^2_{(n-1),\frac{\alpha}{2}}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{(n-1),1-\frac{\alpha}{2}}}$
Two Populations	Means	$(\bar{X}_1 - \bar{X}_2) \pm \left(Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$ σ_1 *and* σ_2 *known*
		$(\bar{X}_1 - \bar{X}_2) \pm \left(t_{\frac{\alpha}{2},n_1+n_2-2}\, Sp \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$ σ_1 *and* σ_2 *unknown* $S_P^2 = \frac{S_1^2(n_1-1) + S_2^2(n_2-1)}{n_1+n_2-2}$
	Proportions	$(\hat{p}_1 - \hat{p}_2) \pm \left(Z_{\frac{\alpha}{2}} \times \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}\right)$
	Ratio of two Variance	$\frac{S_1^2}{S_2^2} \times \frac{1}{F_{v_1,v_2,\frac{\alpha}{2}}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} \times F_{v_2,v_1,\frac{\alpha}{2}}$

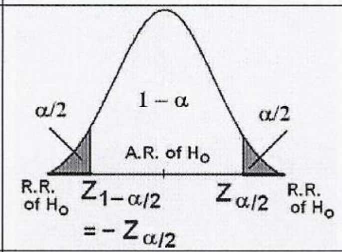
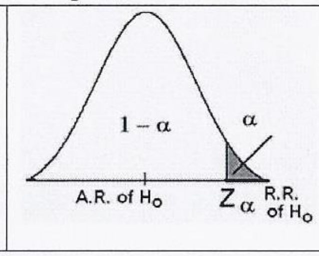
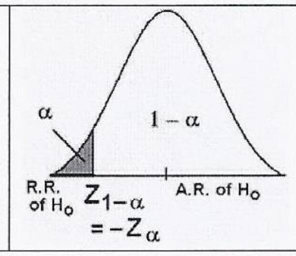
CHAPTER 5

- *Testing of The Hypothesis for:*
  - *Single mean*
  - *Two means*
  - *Single proportion*
  - *Two proportion*
  - *Single variance.*
  - *Two variances.*
  - *Two Samples Paired Observation*

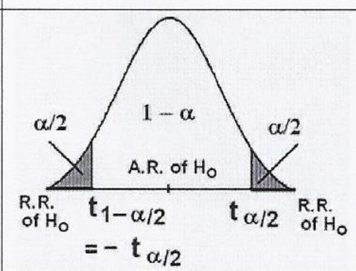
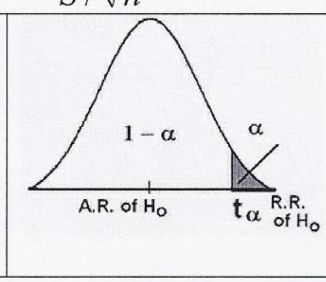
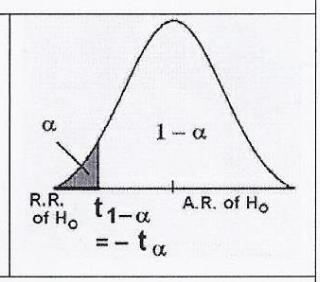
<u>***Hypotheses Testing***</u>

<u>*1-Single Mean*</u>

$(\,if\ \sigma\ known\,)$:

Hypotheses	$H_o: \mu = \mu_o$ $H_1: \mu \neq \mu_o$	$H_o: \mu = \mu_o$ $H_1: \mu > \mu_o$	$H_o: \mu = \mu_o$ $H_1: \mu < \mu_o$
Test Statistic (T.S.)	$Z = \dfrac{\overline{X} - \mu_o}{\sigma / \sqrt{n}} \sim N(0,1)$		
R.R. and A.R. of H_o			
Decision:	Reject H_o (and accept H_1) at the significance level α if:		
	$Z > Z_{\alpha/2}$ or $Z < -Z_{\alpha/2}$ Two-Sided Test	$Z > Z_{\alpha}$ One-Sided Test	$Z < -Z_{\alpha}$ One-Sided Test

$(\,if\ \sigma\ unknown\,)$:

Hypotheses	$H_o: \mu = \mu_o$ $H_1: \mu \neq \mu_o$	$H_o: \mu = \mu_o$ $H_1: \mu > \mu_o$	$H_o: \mu = \mu_o$ $H_1: \mu < \mu_o$
Test Statistic (T.S.)	$T = \dfrac{\overline{X} - \mu_o}{S / \sqrt{n}} \sim t(n-1)$		
R.R. and A.R. of H_o			
Decision:	Reject H_o (and accept H_1) at the significance level α if:		
	$T > t_{\alpha/2}$ or $T < -t_{\alpha/2}$ Two-Sided Test	$T > t_{\alpha}$ One-Sided Test	$T < -t_{\alpha}$ One-Sided Test

Question 1:

Suppose that we are interested in making some statistical inferences about the mean μ , of a normal population with standard deviation σ =2. Suppose that a random sample of size n=49 from this population gave a sample mean $\bar{X} = 4.5$.

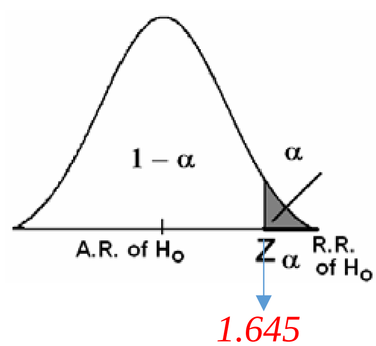
$$\sigma = 2 \quad , \quad n = 49 \quad , \quad \bar{X} = 4.5$$

(1)If we want to test $H_0: \mu = 5 \quad vs \quad H_1: \mu > 5$, then <u>the test statistic equals</u> to

$$Z = \frac{\bar{X} - \mu_o}{\sigma/\sqrt{n}} = \frac{4.5-5}{2/7} = -1.75$$

(2): If we want to test, $H_0: \mu = 5 \quad vs \quad H_1: \mu > 5$, <u>the Rejection Region</u> of H_0

$$\alpha = 0.05 \rightarrow Z_\alpha = Z_{0.05} = 1.645$$



The Rejection Region (R.R) is $(1.645, \infty)$

(3): If we want to test $H_0: \mu = 5 \quad vs \quad H_1: \mu > 5$at, $\alpha = 0.05$ then we

$$Z = -1.75 \notin R.R = (1.645, \infty)$$

Then we accept H_0.

$$P - value = P(Z > -1.75) = 1 - P(Z < -1.75) = 1 - 0.0401 = 0.9599 > \alpha$$

Question 2:

An electrical firm manufactures light bulbs that have a length of life that is normally distributed with a standard deviation of 30 hours. A sample of 50 bulbs were selected randomly and found to have an average of 750 hours. Let μ be the population mean of life of all bulbs manufactured by this firm.

Test $H_0: \mu = 740 \ \ vs \ \ H_1: \mu < 740$? Use a 0.05 level of significance.

$$\sigma = 30 \ \ , \ \ n = 50 \ \ , \ \ \bar{X} = 750$$

- *Hypotheses:*

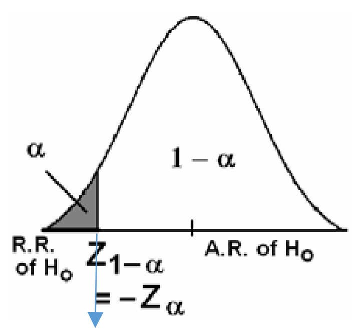
$$H_0: \mu = 740 \quad vs \qquad H_1: \mu < 740$$

- *Test Statistic (T.S):*

$$Z = \frac{\bar{X} - \mu_o}{\sigma/\sqrt{n}} = \frac{750 - 740}{30/\sqrt{50}} = 2.37$$

- *Rejection Region (R.R):*

$$Z_\alpha = Z_{0.05} = 1.645$$



- *Decision:* *-1.645*

$$Z = 2.37 \notin R.R, then\ we\ accept\ H_0$$

$$\boxed{P - value = P(Z < 2.37) = 0.9911 > \alpha}$$

Question 3:

A random sample of size $n = 36$ from a normal quantitative population produced a mean $\bar{X} = 15.2$ and a variance $S^2 = 9$.

Test $H_0: \mu = 15$ vs $H_1: \mu \neq 15$, use $\alpha = 0.05$.

$$s = 3 \quad , \quad n = 36 \quad , \quad \bar{X} = 15.2 \quad , \quad \alpha = 0.05$$

- *Hypotheses:*

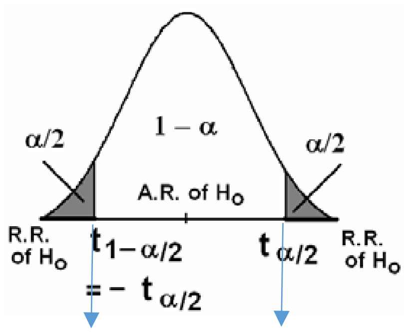
$$H_0: \mu = 15 \quad vs \quad H_1: \mu \neq 15$$

- *Test statistic (T.S):*

$$t = \frac{\bar{X} - \mu_o}{S/\sqrt{n}} = \frac{15.2 - 15}{3/\sqrt{36}} = 0.4$$

- *Rejection Region (R.R):*

$$t_{\frac{\alpha}{2},n-1} = t_{0.025\,,35} = 2.032$$

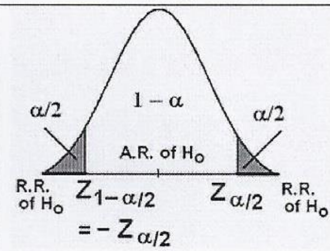
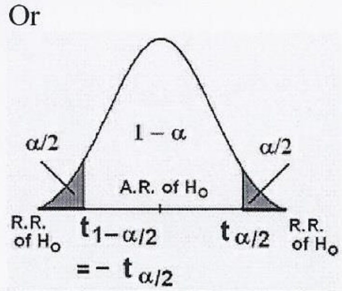
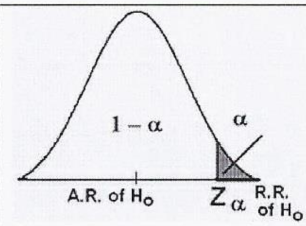
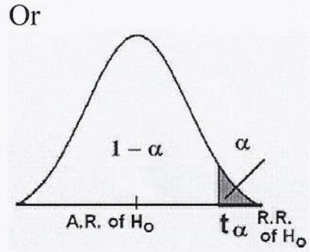
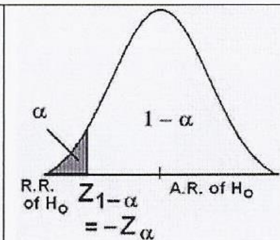
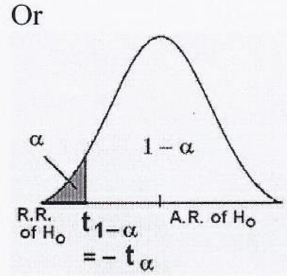


-2.032 *2.032*

- *Decision:*

$$t = 0.4 \notin R.R, \text{then we accept } H_0$$

2-Two Means:

Hypotheses	$H_o: \mu_1 - \mu_2 = d$ $H_1: \mu_1 - \mu_2 \neq d$	$H_o: \mu_1 - \mu_2 = d$ $H_1: \mu_1 - \mu_2 > d$	$H_o: \mu_1 - \mu_2 = d$ $H_1: \mu_1 - \mu_2 < d$
Test Statistic (T.S.)	$Z = \dfrac{(\bar{X}_1 - \bar{X}_2) - d}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim N(0,1)$ {if σ_1^2 and σ_2^2 are known} or $T = \dfrac{(\bar{X}_1 - \bar{X}_2) - d}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \sim t(n_1+n_2-2)$ {if $\sigma_1^2 = \sigma_2^2 = \sigma^2$ is unknown}		
R.R. and A.R. of H_o	 Or 	 Or 	 Or 
Decision:	Reject H_o (and accept H_1) at the significance level α if:		
	T.S. $\in$ R.R. Two-Sided Test	T.S. $\in$ R.R. One-Sided Test	T.S. $\in$ R.R. One-Sided Test

$$Sp^2 = \frac{S_1^2(n_1 - 1) + S_2^2(n_2 - 1)}{n_1 + n_2 - 2}$$

Question 1:

Two random samples were independently selected from two normal populations with equal variances. The results are summarized as follows:

	First sample	***Second sample***
Sample size (n)	***12***	***14***
Sample mean $(\bar{X})$	***10.5***	***10***
Sample variance (S^2)	***4***	***5***

Let μ_1 ***and*** μ_2 ***be the true means of the first and second populations, respectively. Test*** $H_0: \mu_1 = \mu_2 \;\; vs \;\; H_1: \mu_1 \neq \mu_2$ ***.(use*** $\alpha = 0.05$ ***)***

- *Hypotheses:*

$$H_0: \mu_1 = \mu_2 \qquad or \qquad H_0: \mu_1 - \mu_2 = 0$$
$$H_1: \mu_1 \neq \mu_2 \qquad\qquad H_1: \mu_1 - \mu_2 \neq 0$$

- *Test statistic (T.S):*

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - d}{Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$= \frac{(10.5 - 10) - 0}{2.13\sqrt{\frac{1}{12} + \frac{1}{14}}} = 0.597$$

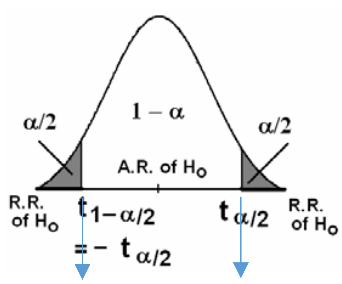
$$Sp^2 = \frac{S_1^2(n_1 - 1) + S_2^2(n_2 - 1)}{n_1 + n_2 - 2}$$

$$= \frac{4(11) + 5(13)}{24} = 4.54$$

$$Sp^2 = 4.54 \Rightarrow Sp = 2.13$$

- *Rejection Region (R.R):*

$$\alpha = 0.05 \rightarrow \quad t_{\frac{\alpha}{2}, n_1 + n_2 - 2} = t_{0.025, 24} = 2.064$$



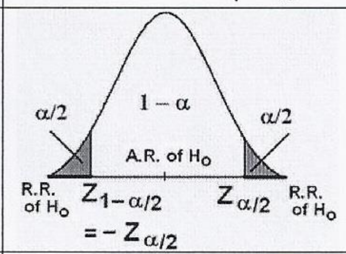
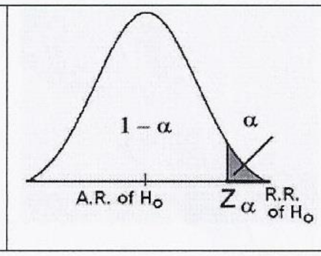
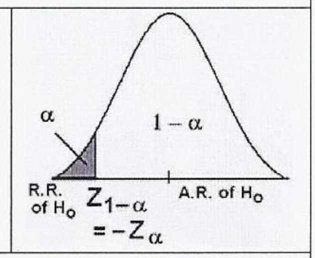
-2.064 2.064

- *Decision:*

$$t = 0.597 \notin R.R \text{ Then we accept } H_0.$$

$$\boxed{P - value = 2 \times P(t_{24} < 0.597) \approx 2 \times 0.3 = 0.6 > \alpha}$$

<u>3-Single Proportion:</u>

Hypotheses	$H_o: p = p_o$ $H_1: p \neq p_o$	$H_o: p = p_o$ $H_1: p > p_o$	$H_o: p = p_o$ $H_1: p < p_o$
Test Statistic (T.S.)	$Z = \dfrac{\hat{p} - p_o}{\sqrt{\dfrac{p_o q_o}{n}}} = \dfrac{X - np_o}{\sqrt{np_o q_o}} \sim N(0,1) \quad (q_o = 1 - p_o)$		
R.R. and A.R. of H_o			
Decision:	Reject H_o (and accept H_1) at the significance level α if:		
	$Z > Z_{\alpha/2}$ or $Z < -Z_{\alpha/2}$ Two-Sided Test	$Z > Z_{\alpha}$ One-Sided Test	$Z < -Z_{\alpha}$ One-Sided Test

Question 1:

A researcher was interested in making some statistical inferences about the proportion of smokers (p) among the students of a certain university. A random sample of 500 students showed that 150 students smoke.

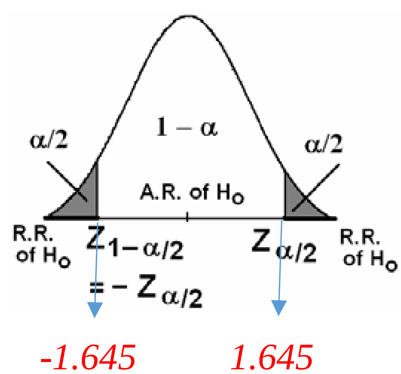
$$n = 500 \ \ \& \ \ x = 150 \ \rightarrow \hat{p} = \frac{150}{500} = 0.3$$

(1): If we want to test $H_0: p = 0.25$ vs $H_1: p \neq 0.25$ then the test statistic:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.3 - 0.25}{\sqrt{\frac{0.25 \times 0.75}{500}}} = 2.58$$

(2): If we want to test $H_0: p = 0.25$ vs $H_1: p \neq 0.25$ at $\alpha = 0.1$, then the Acceptance Region of H_0 is

$$\alpha = 0.1 \rightarrow Z_{\frac{\alpha}{2}} = Z_{0.05} = 1.645 \rightarrow A.R = (-1.645, 1.645).$$



(3): If we want to test $H_0: p = 0.25$ vs $H_1: p \neq 0.25$ at $\alpha = 0.1$, then we:

$Z = 2.58 \notin A.R$, *then we reject* H_0.

$$\boxed{P - value = 2 \times P(Z > 2.58) = 2 \times 0.0049 = 0.0098 < \alpha}$$

__*Question 2:*__

In a random sample of 500 homes in a certain city, it is found that 114 are heated by oil. Let p be the proportion of homes in this city that are heated by oil. A builder claims that less than 20% of the homes in this city are heated by oil. Would you agree with this claim? Use a 0.02 level of significance.

$$n = 500 \;\;\&\;\; x = 114 \;\rightarrow \hat{p} = \frac{114}{500} = 0.23$$

- *Hypotheses:*

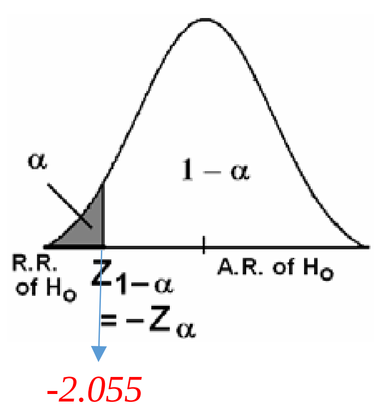
$$H_0: p = 0.2 \quad vs \quad H_1: p < 0.2$$

- *Test statistic (T.S):*

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.23 - 0.2}{\sqrt{\frac{0.2 \times 0.8}{500}}} = 1.67$$

- *Rejection Region (R.R):*

$$\alpha = 0.02 \;\rightarrow Z_\alpha = Z_{0.02} = 2.055$$



- *Decision:*

$$Z = 1.67 \;\notin R.R, \textit{then we accept } H_0.$$

$$P - value = P(Z > 1.67) = 1 - P(Z < 1.67) = 1 - 0.9525 = 0.0475 > \alpha$$

- ***Two Samples Test for Paired Observation***

Q1.The following contains the calcium levels of eleven test subjects at zero hours and three hours after taking a multi-vitamin containing calcium.

Pair	0 hour (X_i)	3 hours (Y_i)	Difference $D_i = X_i - Y_i$
1	17.0	17.0	0.0
2	13.2	12.9	0.3
3	35.3	35.4	-0.1
4	13.6	13.2	0.4
5	32.7	32.5	0.2
6	18.4	18.1	0.3
7	22.5	22.5	0.0
8	26.8	26.7	0.1
9	15.1	15.0	0.1

The sample mean and sample standard deviation of the differences D are 0.144 and 0.167, respectively. To test whether the data provide sufficient evidence to indicate a difference in mean calcium levels ($H_0: \mu_1 = \mu_2$ against$H_1: \mu_1 \neq \mu_2$) with $\alpha = 0.10$ we have: $\boxed{\bar{D} = 0.144 \ , \quad S_d = 0.167 \quad , \quad n = 9}$

[1]. the reliability coefficient (the tabulated value) is:

$$t_{\frac{\alpha}{2}, n-1} = t_{\frac{0.1}{2}, 9-1} = t_{0.05, 8} = \boxed{1.860}$$

[2]. the value of the test statistic is:

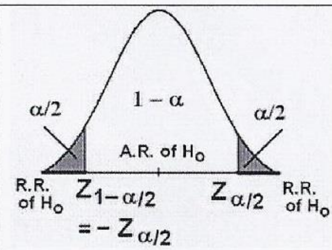
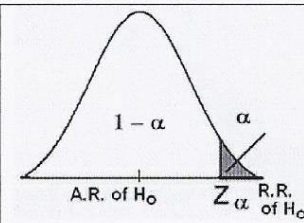
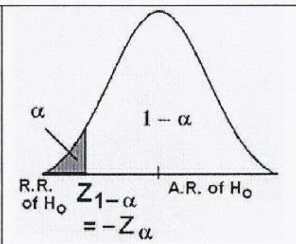
$$\boxed{\begin{array}{l} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{array}} \Longrightarrow \boxed{\begin{array}{l} H_0: \mu_1 - \mu_2 = 0 \\ H_1: \mu_1 - \mu_2 \neq 0 \end{array}} \Longrightarrow \boxed{\begin{array}{l} H_0: \mu_D = 0 \\ H_1: \mu_D \neq 0 \end{array}}$$

$$T = \frac{\bar{D} - \mu_D}{S_d / \sqrt{n}} = \frac{0.144 - 0}{0.167/\sqrt{9}} = \boxed{2.5868}$$

[3]. the decision is:

$$T = 2.5868 \in R.R, then\ we\ \boxed{Reject\ H_0}$$

4-Two Proportions:

Hypotheses	$H_o: p_1 - p_2 = 0$ $H_1: p_1 - p_2 \neq 0$	$H_o: p_1 - p_2 = 0$ $H_1: p_1 - p_2 > 0$	$H_o: p_1 - p_2 = 0$ $H_1: p_1 - p_2 < 0$
Test Statistic (T.S.)	$Z = \dfrac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}\hat{q}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \sim N(0,1)$		
R.R. and A.R. of H_o			
Decision:	Reject H_o (and accept H_1) at the significance level α if:		

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

<u>***Question 1:***</u>

A random sample of 100 students from school "A" showed that 15 students smoke. Another independent random sample of 200 students from school "B" showed that 20 students smoke. Let p_1 be the proportion of smokers in school "A" and p_2 is the proportion of smokers in school "B". Test $H_0: p_1 = p_2$ vs $H_1: p_1 > p_2$. (use α=0.05)

$$n_1 = 100 \ \ \& \ \ x_1 = 15 \ \rightarrow \ \hat{p}_1 = \frac{x_1}{n_1} = \frac{15}{100} = 0.15$$

$$n_2 = 200 \ \ \& \ \ x_2 = 20 \ \rightarrow \ \hat{p}_2 = \frac{x_2}{n_2} = \frac{20}{200} = 0.10$$

- *Hypotheses:*

$$H_0: p_1 = p_2 \qquad or \qquad H_0: p_1 - p_2 = 0$$

$$H_1: p_1 > p_2 \qquad\qquad H_1: p_1 - p_2 > 0$$

- *Test statistic (T.S):*

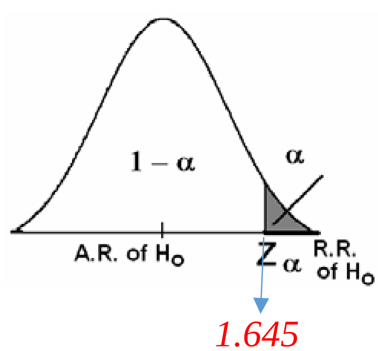
$$Z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \qquad\qquad \hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

$$= \frac{(0.15-0.10)}{\sqrt{(0.12)(0.88)(\frac{1}{100}+\frac{1}{200})}} = 1.26 \qquad\qquad = \frac{15+20}{100+200}$$

$$= 0.12$$

- *Rejection Region (R.R):*

$$\alpha = 0.05 \rightarrow Z_\alpha = Z_{0.05} = 1.645$$



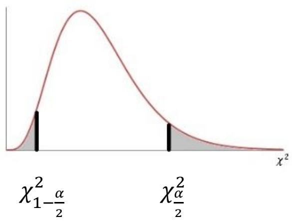
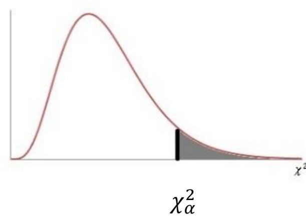
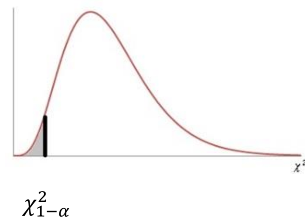
- *Decision:*

$$Z = 1.26 \notin R.R = (\ 1.645, \infty), \text{ Then we accept } H_0.$$

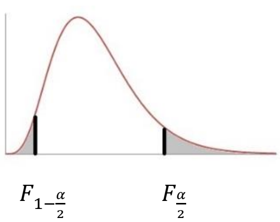
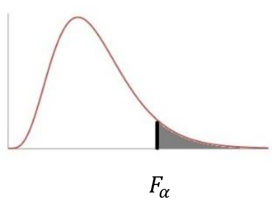
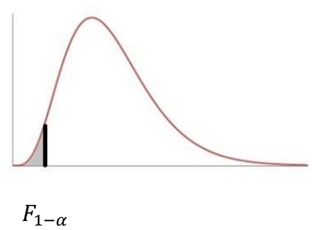
$$\boxed{P - value = P(Z > 1.26) = 1 - P(Z < 1.26) = 1 - 0.8962 = 0.1038 > \alpha}$$

Testing the Population Variance

- *One sample variance:*

Hypotheses	$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 \neq \sigma_0^2$	$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 > \sigma_0^2$	$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 < \sigma_0^2$
Test Statistics (T.S)	$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{n-1}$		
R.R and A.R of H_0			
Decision	*Reject* H_0 *at the significance level* α		

- *Two-sample variances:*

Hypotheses	$H_0: \sigma_1^2 = \sigma_2^2$ $H_1: \sigma_1^2 \neq \sigma_2^2$	$H_0: \sigma_1^2 = \sigma_2^2$ $H_1: \sigma_1^2 > \sigma_2^2$	$H_0: \sigma_1^2 = \sigma_2^2$ $H_1: \sigma_1^2 < \sigma_2^2$
Test Statistics (T.S)	$F = \frac{S_1^2}{S_2^2} \sim F_{n_1-1, n_2-1}$		
R.R and A.R of H_0			
Decision	*Reject* H_0 *at the significance level* α		

<u>Question 1:</u>

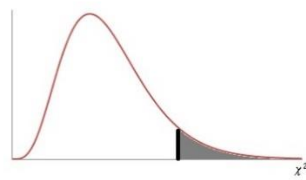
We are interested in the content of a soft-drink dispensing machine. A random sample of 25 drinks gave a variance of 2.03 deciliters^2. Assume that the contents are approximately normally distributed. The soft-drink machine is said to be out of control if the variance exceeds 1.15 deciliters^2. To test
$H_0: \sigma^2 = 1.15 \;\; vs \;\; H_1: \sigma^2 > 1.15$ ***with 0.05 level of significance, then:***

(5) ***The test statistic used is:***

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{(25-1)2.03}{1.15} = 42.365$$

(A)	*7.500*	*(B)*	*0.833*	*(C)*	*42.365*	*(D)*	*10.823*

(6) ***The critical region is:***



$$\chi^2_{\alpha,n-1} = \chi^2_{0.05,24} = 36.415$$

(A)	$(36.415, \infty)$	*(B)*	$(33.196, \infty)$	*(C)*	$(1.96, \infty)$	*(D)*	$(1.645, \infty)$

(7) ***The decision is:***

(A)	*Not reject* H_0	*(B)*	*Reject* H_0

Question 2:

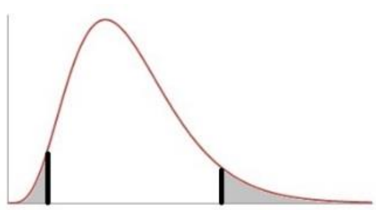
In a series of experiments to determine the absorption rate of certain pesticides into skin, measured amounts of two pesticides were applied to several skin specimens. For pesticide A, the variance of the amounts absorbed in $n_1 = 6$ ***specimens was*** $s_1^2 = 2.3$***, while for pesticide B, the variance of the amounts absorbed in*** $n_2 = 10$ ***specimens was*** $s_2^2 = 0.6$***. Assume that for each pesticide, the amounts absorbed are a simple random sample from a normal population. To test the claim that the variance in the amount absorbed is greater for pesticide A than for pesticide B, that is*** $H_0 : \sigma_1^2 = \sigma_2^2 \ vr \ \ H_1 : \sigma_1^2 \neq \sigma_2^2$ ***with 0.10 level of significance, then:***

1. ***The value of test statistic is:***

$$F = \frac{S_1^2}{S_2^2} = \frac{2.3}{0.6} = 3.833$$

(A)	3.833	(B)	0.1	(C)	0.057	(D)	2.35

2. ***The non-rejection region is:***



$$F_{1-\frac{\alpha}{2}, n_1-1, n_2-1} = F_{0.95,5,9} = \frac{1}{F_{0.05,9,5}} = \frac{1}{4.77} = 0.21$$
$$F_{\frac{\alpha}{2}, n_1-1, n_2-1} = F_{0.05,5,9} = 3.48$$

(A)	$(-\infty, -1.96)$	(B)	$(-1.96, \infty)$	(C)	$(-1.96, 1.96)$	(D)	$(0.21, 3.48)$

3. ***The decision is:***

(A)	Not reject H_0	(B)	Reject H_0

<u>Question 3</u>:

The following data represent the sizes and variances of two normal samples:

Sample1	$n_1 = 10$	$s_1^2 = 1.8$
Sample2	$n_2 = 13$	$s_2^2 = 1.5$

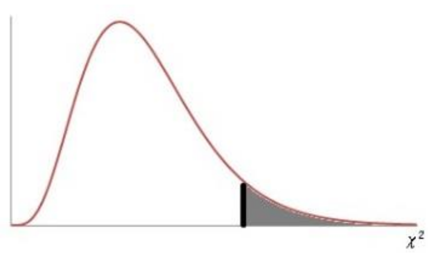
- ***To test*** $H_0 : \sigma_1^2 = 1.5 \;\; vr \;\; H_1 : \sigma_1^2 > 1.5$ ***with level 0.05, then:***

a. The test statistic used is:

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{(10-1)1.8}{1.5} = 10.8$$

(A)	*7.5*	*(B)*	*0.833*	*(C)*	*1.2*	*(D)*	*10.8*

b. The rejection region is:



$$\chi^2_{\alpha,n-1} = \chi^2_{0.05,9} = 16.919$$

(A)	$(-\infty, -16.92)$	*(B)*	$(-16.92, \infty)$	*<u>(C)</u>*	$(16.92, \infty)$	*(D)*	$(1.645, \infty)$

c. The decision is:

(A)	*Accept* H_0	*(B)*	*Reject* H_0

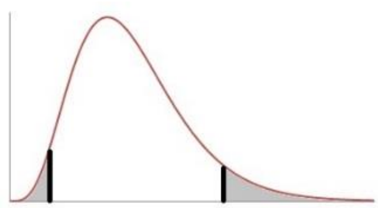
- ***To test claim with level 0.10 that both variances <u>are equal</u>, then:***

d. The test statistic used is:

$$F = \frac{S_1^2}{S_2^2} = \frac{1.8}{1.5} = 1.2$$

(A)	*7.5*	*(B)*	*0.833*	*(C)*	*1.2*	*(D)*	*10.8*

e. The acceptance region is:



$$F_{1-\frac{\alpha}{2},n_1-1,n_2-1} = F_{0.95,9,12} = \frac{1}{F_{0.05,12,9}} = \frac{1}{3.07} = 0.326$$

$$F_{\frac{\alpha}{2},n_1-1,n_2-1} = F_{0.05,9,12} = 2.8$$

(A)	$(0.326\,, 2.8)$	*(B)*	$(-2.8, 2.8)$	*(C)*	$(2.8, \infty)$	*(D)*	$(1.645, \infty)$

f. The decision is:

(A)	*Accept* H_0	*(B)*	*Reject* H_0

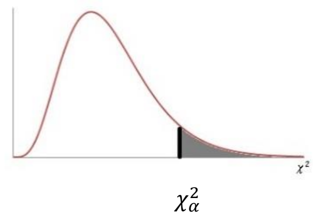
Testing		
single Population	Mean	$Z = \frac{\bar{X} - \mu_o}{\sigma/\sqrt{n}}$ σ known
		$t = \frac{\bar{X} - \mu_o}{S/\sqrt{n}}$ σ unknown
	Proportion	$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$
	Variance	$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$
Two Populations	Means	$Z = \frac{(\bar{X}_1 - \bar{X}_2) - d}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ σ_1 & σ_2 known
		$t = \frac{(\bar{X}_1 - \bar{X}_2) - d}{Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ σ_1 & σ_2 unknown
	Proportions	$Z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ $\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$
	Variances	$F = \frac{S_1^2}{S_2^2}$

Two samples T test for paired observation $T = \frac{\bar{D} - \mu_D}{S_d/\sqrt{n}}$

CHAPTER 6

- *Goodness of Fit Test.*
- *Test for Independency and Homogeneity.*

Goodness of Fit Test

Hypotheses	$H_0: The\ data\ follow\ (\ldots)\ distribution$ $H_1: The\ data\ do\ not\ follow\ (\ldots)\ distribution$
Test Statistics (T.S)	$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$
R.R and A.R of H_0	
Decision	*If* $\chi^2 > \chi^2_\alpha \implies$ *Reject* H_0

***Q1**: We investigate the number of errors in electrical circles panel printers. A random sample of 200 panel printers is selected and we recorded X= Number of errors in each printer. The table is the observed and the expected of each value of X in the sample:*

Values of X	$x_1 = 0$	$x_2 = 1$	$x_3 = 2$	$x_4 = 3$	$x_5 = 4$
O_i	$O_1 = 75$	$O_2 = 66$	$O_3 = 42$	$O_4 = 13$	$O_5 = 4$
E_i	$E_1 = 71.76$	$E_2 =$	$E_3 = 37.7$	$E_4 =$	$E_5 = 3.3$

Test at** $\underline{\alpha = 0.10}$ **the inspector claim that X follows Poisson distribution

1) ***The estimate of Poisson parameter** λ **is:***

$$\lambda = \frac{(0\times75)+(1\times66)+(2\times42)+(3\times13)+(4\times4)}{200} = \boxed{1.025}$$

2) ***The expected number of errors** E_2 **is:***

$$E_2 = \frac{\lambda^{x_2} e^{-\lambda}}{x_2!} \times 200 = \frac{(1.025)^1 e^{-(1.025)}}{1!} \times 200 = \boxed{73.55}$$

3) ***The expected number of errors** E_4 **is:***

$$E_4 = \frac{\lambda^{x_4} e^{-\lambda}}{x_4!} \times 200 = \frac{(1.025)^3 e^{-(1.025)}}{3!} \times 200 = \boxed{12.88}$$

4) ***The degree of freedom of the** χ^2 **test statistic is:***

$$v = 5 - 1 = \boxed{4}$$

5) ***The value of the** χ^2 **test statistic is:***

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = \frac{(75-71.76)^2}{71.76} + \frac{(66-73.55)^2}{73.55} + \cdots + \frac{(4-3.3)^2}{3.3} = \boxed{1.56}$$

6) ***The rejection region R.R is***

$$\chi^2_{\alpha,n-1} = \chi^2_{0.1,4} = 7.779 \implies R.R \in \boxed{(7.779\,,\infty)}$$

7) ***The conclusion is***

$$1.56 \notin (7.779\,,\infty) \implies \boxed{Accept\ the\ claim}$$

Q2. 1000 bags of oranges, each containing 10 oranges. Some of the oranges are rotten. The observed numbers of rotten oranges per bag are tabulated as follows:

No. of rotten oranges	0	1	2	3	4	5	6
Observed counts (O_i)	334	369	191	63	22	12	9
Expected counts (E_i)	E_1= 297.4	E_2=383.4	E_3= ...	E_4=...	E_5=17.3	E_6=2.7	E_7= 0.3

Suppose that we want to test at level of significance 0.05, that this data suggest that the distribution of rotten oranges in the individual bags follow a binomial distribution with parameters n=10 and unknown p (Bin(10, p)). Then:

(1) The null hypothesis H_0 is:

A	B
The counts of rotten oranges follow Poisson distribution	*The counts of rotten oranges do not follow a binomial distribution (Bin(10, p) for some p)*
C	**D**
The counts of rotten oranges follow a binomial distribution (Bin(10, p) for some p)	*The counts of rotten oranges follow a binomial distribution (Bin(n, 10)*

(2) The estimate of the unknown parameter p is:

$$E(X) = n \times p = \frac{(0\times334)+(1\times369)+\cdots+(6\times9)}{1000} = 1.142$$

$$n \times p = 1.142 \Longrightarrow 10 \times p = 1.142 \Longrightarrow \boxed{p = 0.1142}$$

(3) The expected number of errors E_3 is:

$$p(X=2) = \binom{10}{2}(0.1142)^2(1-0.1142)^8 = 0.2224$$

$$E_3 = 1000 \times p(X=2) = 1000 \times 0.2224 = \boxed{222.4}$$

(4) The expected number of errors E_4 is:

$$p(X=3) = \binom{10}{3}(0.1142)^3(1-0.1142)^7 = 0.0765$$

$$E_4 = 1000 \times p(X=3) = 1000 \times 0.0765 = \boxed{76.5}$$

(5) The degree of freedom of the χ^2 test statistic is:

$$v = 7 - 1 = \boxed{6}$$

(6) The value of the χ^2 test statistic is:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = \frac{(334-297.4)^2}{297.4} + \frac{(369-383.4)^2}{383.4} + \cdots + \frac{(9-0.3)^2}{0.3} = \boxed{297.47}$$

(7) The rejection region (R.R) of H_0 is:

$$\chi^2_{\alpha,n-1} = \chi^2_{0.05,6} = 12.592 \implies R.R \in \boxed{(12.592\,,\infty)}$$

(8) The decision is:

$$297.47 \in (12.592\,,\infty) \implies \boxed{Reject\ H_0}$$

<u>Q3</u>. A doctor believes that the proportions of births in this country on each day of the week are equal. A simple random sample of 700 births from a recent year is selected, and the results are below.

Day	*Sunday*	*Monday*	*Tuesday*	*Wednesday*	*Thursday*	*Friday*	*Saturday*
Frequency	*65*	*103*	*114*	*116*	*115*	*112*	*75*
Expected (E_i)	$E_1 = 100$	$E_2 = \cdots$	$E_3 = \cdots$	$E_4 = 100$	$E_5 = 100$	$E_6 = 100$	$E_7 = 100$

At a significance level of α =0.01, we want to test the hypothesis if there is enough evidence to support the doctor's claim.

1. ***The expected frequency E_2 is***

$$E_2 = 100$$

2. ***The degree of freedom of the χ^2 test statistic is***

$$v = 7 - 1 = \boxed{6}$$

3. ***The value of the χ^2 test statistic is***

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = \frac{(65-100)^2}{100} + \frac{(103-100)^2}{100} + \cdots + \frac{(75-100)^2}{100} = \boxed{26.8}$$

4. ***The critical value is***

$$\chi^2_{\alpha, n-1} = \chi^2_{0.01,6} = \boxed{16.812}$$

5. ***The decision about the doctor's claim is***

$$26.8 \in (16.812\,, \infty) \Longrightarrow \boxed{Reject\ H_0}$$

Q4: ***We test the balance of a dice. If the results of tossing the dice 180 times are:***

X	1	2	3	4	5	6	
Frequency	15	40	36	25	45	19	180
E_i	30	30	30	30	30	30	

1. The expected frequency of each face is

(A)	40	(B)	50	(C)	30	(D)	60

2. The test statistic used to perfume this test is distributed as:

(A)	T	(B)	F	(C)	Normal	(D)	Chi squares

2. The degree of freedom of the value of the test statistic is:

(A)	6	(B)	5	(C)	7	(D)	8

3. The value of the test statistic is:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = \frac{(15-30)^2}{30} + \frac{(40-30)^2}{30} + \cdots + \frac{(19-30)^2}{30} = \boxed{24.4}$$

(A)	8.94	(B)	22.4	(C)	24.4	(D)	44.7

4. The table value, at the 0.05 significance level, is

$$\chi^2_{\alpha, n-1} = \chi^2_{0.05,5} = \boxed{11.070}$$

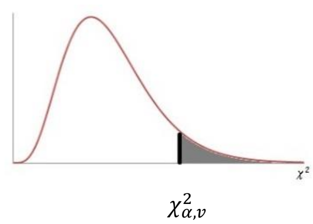
(A)	22.5	(B)	17.6	(C)	18.45	(D)	11.070

5. The conclusion is

$$24.4 \in (11.070\,, \infty) \Longrightarrow \boxed{Reject\ H_0}$$

(A)	Accept the balance	(B)	Reject the balance	(C)	No Decision

Test for Independence (categorical data)

Hypotheses	H_0: *The two random variable are independent* H_1: *The two random variable are not independent*
Test Statistics (T.S)	$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$ $E_{ij} = \frac{(j^{th}\ column\ total) \times (i^{th}\ row\ total)}{grand\ total}$
R.R and A.R of H_0	 $v = (c - 1) \times (r - 1)$
Decision	*If* $\chi^2 > \chi^2_\alpha \implies$ *Reject* H_0

Q1: ***We test the hypothesis at level of significance 0.05, that the size of a family is independent of the level of education attained by the father. A random sample of 400 married men, all retired, were classified according to education and number of children. The table shows the observed O_{ij} and the expected E_{ij} values:***

			Number of Children			
			0-1	2-3	4-5	Over 5
Education	Elementary	**O**	25	48	44	50
		E	32.57	53.86	38.83	41.75
	Secondary		30	53	28	32
			27.89	E_{22}	E_{23}	35.75
	College		23	28	21	18
			17.55	29.03	20.93	22.5

1) ***The test statistic used to perfume this test is distributed as:***

(A)	T	(B)	F	(C)	Normal	(D)	Chi squares

2) ***The mathematic expression of the test statistic is:***

(A)	$\sum_{1}^{C}\frac{(O_j - E_j)^2}{O_j}$	(B)	$\sum_{1}^{C}\frac{(O_j - E_j)^2}{E_j}$
(C)	$\sum_{i=1}^{r}\sum_{j=1}^{C}\frac{(O_{ij} - E_{ij})^2}{O_{ij}}$	(D)	$\sum_{i=1}^{r}\sum_{j=1}^{C}\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$

3) ***The expectation E_{23} is:***

$$E_{23} = \frac{143\times 93}{400} = \boxed{33.25}$$

4) ***The expectation E_{22} is:***

$$E_{22} = \frac{143\times 129}{400} = \boxed{46.12}$$

5) ***The value of the test statistic is:***

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} = \frac{(25-32.57)^2}{32.57} + \frac{(48-53.86)^2}{53.86} + \cdots + \frac{(18-22.5)^2}{22.5} = \boxed{9.75}$$

6) ***The degree of freedom of the test statistic is:***

$$v = (c-1)\times(r-1) = (4-1)\times(3-1) = 3\times 2 = \boxed{6}$$

7) ***The rejection region R.R is:***

$$\chi^2_{0.05,6} = 12.592 \Rightarrow \boxed{(12.592\,,\infty)}$$

8) ***The decision about the independence is***

$$9.75 \notin (12.592\,,\infty) \Rightarrow \boxed{Accept\ H_0}$$

***Q2:* In an experiment to study the dependence of hypertension on smoking habits, the following data were taken on 180 individuals:**

	Non-smokers (NS)	Moderate Smokers (MS)	Heavy Smokers (HS)	Total
Hypertension	21 (E_{11}=33.35)	36 (E_{12}=29.97)	30 (E_{13}=23.68)	
No hypertension	48 (E_{21}=35.65)	26 (E_{22}=32.03)	19 (E_{23}=25.32)	93
Total		62	49	180

Suppose that we want to test the hypothesis at level of significance 0.05, that the presence or absence of hypertension and the smoking habits are independent.

The test statistic used to perfume this test is distributed as:

A	B	C	D
T	F	Normal	Chi squares

1. ***The mathematical expression of the test statistic is:***

A	B	C	D
$\sum_{1}^{C}\frac{(O_j-E_j)^2}{O_j}$	$\sum_{1}^{C}\frac{(O_j-E_j)^2}{E_j}$	$\sum_{i=1}^{r}\sum_{j=1}^{C}\frac{(O_{ij}-E_{ij})^2}{O_{ij}}$	$\sum_{i=1}^{r}\sum_{j=1}^{C}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$

2. ***The value of the expectation E_{22} is:***

$$E_{22}=\frac{93\times 62}{180}=\boxed{32.03}$$

3. ***The value of the expectation E_{23} is:***

$$E_{23}=\frac{93\times 49}{180}=\boxed{25.32}$$

4. ***The value of the test statistic is:***

$$\chi^2=\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}=\frac{(21-33.35)^2}{33.35}+\frac{(36-29.97)^2}{29.97}+\cdots+\frac{(19-25.32)^2}{22.5}=\boxed{14.46}$$

5. ***The degree of freedom of the test statistic is:***

$$v=(c-1)\times(r-1)=(3-1)\times(2-1)=2\times 1=\boxed{2}$$

6. ***The rejection region (R.R) of H_0 is:***

$$\chi^2_{0.05,2}=5.991\Longrightarrow\boxed{(5.991\,,\infty)}$$

7. ***The decision about the independence is:***

$$14.46\in(5.991\,,\infty)\Longrightarrow\boxed{Reject\ H_0}$$

***Q3*. Results of a random sample of children with pain from musculoskeletal injuries treated with acetaminophen, ibuprofen, or codeine are shown in the table. At α = 0.10, we want to test the hypothesis that the treatment and result are independent**

	Acetaminophen	*Ibuprofen*	*Codeine*	*Total*
Significant Improvement	*58 (E_{11}=66.7)*	*81 (E_{12}=...)*	*61 (E_{13}=66.6)*	*200*
Slight Improvement	*42 (E_{21}=...)*	*19 (E_{22}=33.3)*	*39 (E_{23}=33.4)*	*100*
Total	*100*	*100*		*300*

1. ***The distribution of the test statistic is***

(A) t	(B) Binomial	(C) Chi squares	(D) Normal

2. ***The value of the expectation E_{12} is***

$$E_{12} = \frac{200\times100}{300} = \boxed{66.7}$$

3. ***The value of the expectation E_{21} is***

$$E_{21} = \frac{100\times100}{300} = \boxed{33.3}$$

4. ***The mathematical expression of the test statistic is***

(A)	(B)	(C)	(D)
$\sum_{1}^{c}\frac{(O_j - E_j)^2}{E_j}$	$\sum_{1}^{c}\frac{(O_j - E_j)^2}{O_j}$	$\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$	$\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij} - E_{ij})^2}{O_{ij}}$

5. ***The value of the χ^2 test statistic is***

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} = \frac{(58-66.7)^2}{66.7} + \frac{(81-66.7)^2}{66.7} + \cdots + \frac{(39-33.4)^2}{33.4} = \boxed{14.02}$$

6. The critical value is

$$v = (c-1)\times(r-1) = (3-1)\times(2-1) = 2\times1 = 2$$

$$\chi^2_{0.10,2} = \boxed{4.605}$$

7. The decision about the independence is

$$14.02 \in (4.605\,,\infty) \Longrightarrow \boxed{Reject\ H_0}$$

CHAPTER 7

- ***Correlation:***

$$Corr(X,Y) = r_{XY} = \frac{S_{XY}}{\sqrt{S_{XX}\ \ S_{YY}}}$$

$$S_{XX} = \sum(X_i - \bar{X})^2 = \sum X_i^2 - n\bar{X}^2$$

$$S_{YY} = \sum(Y_i - \bar{Y})^2 = \sum Y_i^2 - n\bar{Y}^2$$

$$S_{XY} = \sum(X_i - \bar{X})(Y_i - \bar{Y}) = \sum X_i Y_i - n\bar{X}\bar{Y}$$

$$\boxed{Coefficient\ of\ determination: R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = r_{XY}^2}$$

$$SSR = \Sigma(\hat{y}_i - \bar{y})^2 = \frac{S_{XY}^2}{S_{XX}}$$

$$SSE = \Sigma(y_i - \hat{y}_i)^2 = S_{YY} - \frac{S_{XY}^2}{S_{XX}} = \sum e_i^2 = \Sigma y_i^2 - b_o\Sigma y_i - b_1\Sigma x_i y_i$$

$$SST = \Sigma(y_i - \bar{y})^2 = S_{YY}$$

$$\boxed{SST = SSE + SSR}$$

- ***Simple Linear Regression:***

$$\hat{\boldsymbol{y}} = \boldsymbol{b_0} + \boldsymbol{b_1}\boldsymbol{X}$$

$$b_0 = \bar{y} - b\bar{x} \qquad b_1 = \frac{S_{XY}}{S_{XX}}$$

- ***Testing β_1:***

$$t = \frac{b_1 - \beta_{10}}{S(b_1)}$$

- ***Estimation for β_1:***

$$\beta_1 \in \left(b_1 \pm t_{1-\frac{\alpha}{2},n-2}\, S(b_1)\right)$$

$$S(b_1) = \frac{\hat{\sigma}}{\sqrt{S_{XX}}} \quad , \quad \hat{\sigma}^2 = \frac{SSE}{n-2} = \frac{\Sigma(y_i - \hat{y}_i)^2}{n-2}$$

***Q1:* The grades of a class of 9 students on a midterm report (X) and on the final examination (Y) are as follows:**

X	82	66	78	80	85	85	99	99	68	$\sum_i x_i = 742$, $\sum_i y_i = 707$, $\sum_i x_i y_i = 59648$
Y	77	50	71	72	81	94	96	99	67	$\sum_i x_i^2 = 62240$, $\sum_i y_i^2 = 57557$

The value of Pearson Correlation coefficient is:

$$S_{XX} = \sum(X_i - \bar{X})^2 = \sum X_i^2 - n\bar{X}^2 = 62240 - 9 \times \left(\frac{742}{9}\right)^2 = \boxed{1066.222}$$

$$S_{YY} = \sum(Y_i - \bar{Y})^2 = \sum Y_i^2 - n\bar{Y}^2 = 57557 - 9 \times \left(\frac{707}{9}\right)^2 = \boxed{2018.222}$$

$$S_{XY} = \sum(X_i - \bar{X})(Y_i - \bar{Y}) = \sum X_i Y_i - n\bar{X}\bar{Y} = (59648) - 9 \times \left(\frac{742}{9}\right)\left(\frac{707}{9}\right) = \boxed{1359.778}$$

$$\boxed{r = \frac{S_{XY}}{\sqrt{S_{XX}\ S_{YY}}} = \frac{1359.778}{\sqrt{1066.222 \times 2018.222}} = 0.972}$$

If the estimate of the linear regression line is $\hat{y} = b_0 + b_1 X$, then

1. ***The value of b_1 is:***

$$b_1 = \frac{S_{XY}}{S_{XX}} = \frac{1359.222}{1066.222} = 1.275$$

3. ***The value of b_0 is:***

$$b_0 = \bar{y} - b_1\bar{x}$$
$$= \left(\frac{707}{9}\right) - (1.275) \times \left(\frac{742}{9}\right) = -26.56$$

4. ***A student got 85 on the midterm, then the estimate of the final grade is:***

$$\hat{y} = b_0 + b_1 X = -26.56 + 1.275X$$
$$= -26.56 + 1.275 \times (85) = 81.82$$

Q2. ***A study was made by a retail merchant to determine the relation between weekly advertising expenditures and sales.***

costs \$ (X)	40	20	25	20	30	50	40	20	50	40	25	50
Sales \$ (Y)	385	400	395	365	475	440	490	420	560	525	480	510

$$\sum x_i = 410, \sum y_i = 5445, \sum x_i^2 = 15650, \sum y_i^2 = 2512925, \sum x_i y_i = 191325,$$

1. ***The Pearson correlation coefficient of sales and advertising costs is:***

$$S_{XX} = \sum (X_i - \bar{X})^2 = \sum X_i^2 - n\bar{X}^2 = 15650 - 12 \times \left(\frac{410}{12}\right)^2 = \boxed{1641.67}$$

$$S_{YY} = \sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - n\bar{Y}^2 = 2512925 - 12 \times \left(\frac{5445}{12}\right)^2 = \boxed{42256.25}$$

$$S_{XY} = \sum (X_i - \bar{X})(Y_i - \bar{Y}) = \sum X_i Y_i - n\bar{X}\bar{Y} = (191325) - 12 \times \left(\frac{410}{12}\right)\left(\frac{5445}{12}\right) = \boxed{5287.5}$$

$$r = \frac{S_{XY}}{\sqrt{S_{XX} \; S_{YY}}} = \frac{5287.5}{\sqrt{1641.67 \times 42256.25}} = 0.63$$

- If the estimate of the linear regression line is $\hat{y} = b_0 + b_1 X$, then:

2. ***The value of b_1 is:***

$$b_1 = \frac{S_{XY}}{S_{XX}} = \frac{5287.5}{1641.67} = 3.22$$

3. ***The value of b_0 is:***

$$b_0 = \bar{y} - b_1\bar{x}$$
$$= \left(\frac{5445}{12}\right) - (3.22) \times \left(\frac{410}{12}\right) = 343.71$$

4. ***If the advertising costs is \$30, then the weekly sales is:***

$$\hat{y} = b_0 + b_1 X = 343.71 + 3.22X$$
$$= 343.71 + 3.22 \times (30) = 440.31$$

- We want to test the hypothesis that $\boldsymbol{\beta_1 = 0}$ against the alternative that $\boldsymbol{\beta_1 \neq 0}$ at the 0.05 level of significance.

1. ***The residual sum of squares SSR is:***

$$SSR = \frac{S_{XY}^2}{S_{XX}} = \frac{(5287.5)^2}{1641.67} = 17030.04$$

2. ***The total sum of squares SST is:***

$$SST = S_{YY} = 42256.25$$

3. ***The value of the statistic test is:***

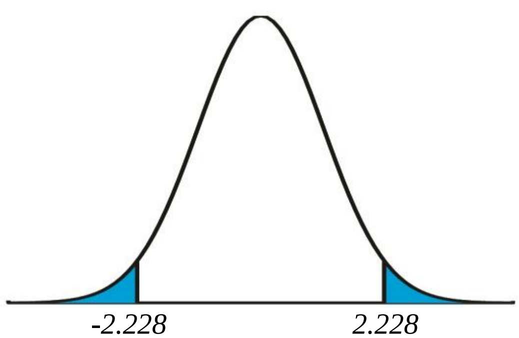
$$t = \frac{b_1 - \beta_{10}}{S(b_1)} = \frac{3.22 - 0}{1.24} = \boxed{2.6}$$

$$\hat{\sigma}^2 = \frac{SSE}{n-2} = \frac{SST-SSR}{n-2} = \frac{42256.25-17030.04}{12-2} = 2522.625 \implies \hat{\sigma} = 50.2257$$

$$S(b_1) = \frac{\hat{\sigma}}{\sqrt{S_{XX}}} = \frac{50.2257}{\sqrt{1641.67}} = 1.24$$

4. ***The decision is:***

$$t_{\frac{\alpha}{2}, n-2} = t_{0.025,10} = 2.228$$



$$t = 2.6 \notin (-2.228, 2.228) \implies Reject\ H_0$$

Q3. ***The shear resistance of soil, Y , is determined by measurements as a function of the normal stress, X. We assume that the errors ε_i are normally distributed. The data are as shown below:***

x_i	10	11	12	13	14	15	16	17	18	19	20	21
y_i	14.08	15.57	16.94	17.68	18.49	19.55	20.68	21.72	22.8	23.84	24.79	25.67

We have $\sum_i x_i = 186$, $\sum_i y_i = 241.81$, $\sum_i x_i^2 = 3026$, $\sum_i y_i^2 = 5025.399$, $\sum_i x_i y_i = 3895.65$

1. ***The coefficient S_{XX} is***

$$S_{XX} = \sum (X_i - \bar{X})^2 = \sum X_i^2 - n\bar{X}^2 = 3026 - 12 \times \left(\frac{186}{12}\right)^2 = \boxed{143}$$

2. ***The coefficient S_{YY} is***

$$S_{YY} = \sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - n\bar{Y}^2 = 5025.399 - 12 \times \left(\frac{241.81}{12}\right)^2 = \boxed{152.726}$$

3. ***The coefficient S_{XY} is***

$$S_{XY} = \sum (X_i - \bar{X})(Y_i - \bar{Y}) = \sum X_i Y_i - n\bar{X}\bar{Y} = (3895.65) - 12 \times \left(\frac{186}{12}\right)\left(\frac{241.81}{12}\right) = \boxed{147.595}$$

4. ***The sample linear correlation coefficient r is***

$$r = \frac{S_{XY}}{\sqrt{S_{XX} \; S_{YY}}} = \frac{147.595}{\sqrt{143 \times 152.726}} = 0.9987$$

- If the estimate of the linear regression line is $\hat{y} = b_0 + b_1 X$, then:

5. ***The value of b_1 is:***

$$b_1 = \frac{S_{XY}}{S_{XX}} = \frac{147.595}{143} = 1.032$$

6. ***The value of b_0 is:***

$$b_0 = \bar{y} - b_1\bar{x}$$
$$= \left(\frac{241.81}{12}\right) - (1.032) \times \left(\frac{186}{12}\right) = 4.15$$

-We want to test the hypothesis that $\beta_1 = 1$ against the alternative that $\beta_1 > 1$ at the 0.05 level of significance. The residuals e_i are

-0.394 0.064 0.402 0.109 -0.113 -0.085 0.013 0.021 0.069 0.077 -0.005 -0.158

7. Deduce that the value of SSE is

$$SSE = \sum e_i^2 = (-0.394)^2 + (0.064)^2 + (0.402)^2 + \cdots + (-0.158)^2 = \boxed{0.389}$$

8. The unbiased estimate of σ^2 is

$$\hat{\sigma}^2 = \frac{SSE}{n-2} = \frac{0.389}{12-2} = \boxed{0.0389} \implies \hat{\sigma} = 0.197$$

9. The value of the test statistic is

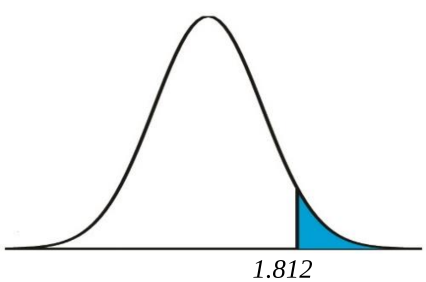
$$t = \frac{b_1 - \beta_{10}}{S(b_1)} = \frac{1.032 - 1}{0.0165} = \boxed{1.94}$$

$$S(b_1) = \frac{\hat{\sigma}}{\sqrt{S_{XX}}} = \frac{0.197}{\sqrt{143}} = 0.0165$$

10. The critical value is

$$t_{\alpha,n-2} = t_{0.05,10} = \boxed{1.812}$$

11. The decision is



$$t = 1.94 > 1.812 \implies \boxed{Reject\ H_0}$$

12. The coefficient of determination R^2 is

$$R^2 = \frac{SSR}{SST} = r_{XY}^2 = (0.9987)^2 = 0.9974$$

13. Determine the 90% confidence interval for the parameter β_1 (2 marks).

$$\beta_1 \in \left(b_1 \pm t_{\frac{\alpha}{2},n-2} \times S(b_1)\right)$$

$$\beta_1 \in \left(1.032 \pm t_{0.05,10} \times (0.0165)\right)$$

$$\beta_1 \in \left(1.032 \pm (1.812) \times (0.0165)\right)$$

$$\beta_1 \in (1.002 \ , 1.062)$$

CHAPTER 8

- *On way ANOVA*

$H_0: \mu_1 = \mu_2 = \cdots = \mu_k \quad vs \quad H_1: al\ least\ two\ of\ the\ means\ are\ not\ eqaul.$

Source of Variation	Sum Square	Degrees of freedom	Mean Square	F
Treatment	$SSA = SS_{trt} = n_i \sum_{i=1}^{k} (\bar{y}_{i.} - \bar{y}_{..})^2$	$df_{trt} = k - 1$	$MSA = MS_{trt} = \frac{SSA}{k-1}$	$F = \frac{MSA}{MSE}$
Error	$SSE = SS_{er} = \sum_{i=1}^{k} \sum_{j=1}^{n} \left(y_{ij} - \bar{y}_{i.}\right)^2$	$df_{er} = k(n-1) = kn - k$	$MSE = MS_{er} = \frac{SSE}{k(n-1)}$	
Total	$SST = SS_{tot} = \sum_{i=1}^{k} \sum_{j=1}^{n} \left(y_{ij} - \bar{y}_{..}\right)^2$ $\boxed{SST = SSA + SSE}$	$df_{tot} = kn - 1$		

***Q1.** The statistics classroom is divided into three rows: front, middle, and back. The instructor noticed that the further the students were from him, the more likely they were to miss class or use an instant messenger during class. He wanted to see at level of significance 0.05: Are the students further away did worse on the exams? For this end, a random sample of the students in each row was taken. The score for those students on the second exam was recorded:*

- *Front:* 82, 83, 97, 93, 55, 67, 53
- *Middle:* 83, 78, 68, 61, 77, 54, 69, 51, 63
- *Back:* 38, 59, 55, 66, 45, 52, 52, 61

Let the one way ANOVA tabulated as follows:

Source of variation	***Sum of squares***	***Degrees of freedom***	***Mean Squares***	***Test Statistics***
Treatments	SS_{trt}	df_{trt}	MS_{trt}	F_0
Errors	SS_{er}	df_{er}	MS_{er}	
Total	$SS_{tot} = 5287.83$	df_{tot}		

										n_i	Sum	Mean	
Front	82	83	97	93	55	67	53			7	530	75.71	
Middle	83	78	68	61	77	54	69	51	63	9	604	67.11	
Back	38	59	55	66	45	52	52	61		8	428	53.50	
										24	1562	65.08	Total

1. **The value of SS_{trt} is:**

$$SS_{trt} = n_i \sum_{i=1}^{k} (\bar{y}_{i.} - \bar{y}_{..})^2$$

$$= 7 \times (75.71 - 65.08)^2 + 9 \times (67.11 - 65.08)^2 + 8 \times (53.50 - 65.08)^2 = 1901.52$$

2. ***The value of SS_{er} is:***

$$SS_{er} = SS_{tot} - SS_{trt} = 5287.83 - 1901.52 = 3386.32$$

3. ***The degrees of freedom of the treatments (df_{trt}) is:***

$$k - 1 = 3 - 1 = 2$$

4. ***The degrees of freedom of the errors (df_{er}) is:***

$$k(n-1) = kn - k = 24 - 3 = 21$$

5. ***The degrees of freedom of the total (df_{tot}) is:***

$$kn - 1 = 3 \times 8 - 1 = 23$$

6. ***The Mean Squares of the treatments (MS_{trt}) is:***

$$MS_{trt} = \frac{SS_{trt}}{k-1} = \frac{1901.52}{3-1} = 950.76$$

7. ***The Mean Squares of the errors (MS_{er}) is:***

$$MS_{er} = \frac{SS_{er}}{kn-k} = \frac{3386.32}{24-3} = 161.25$$

8. ***The value of the test statistic F_0 is:***

$$F = \frac{MS_{trt}}{MS_{er}} = \frac{950.76}{161.25} = 5.896$$

9. ***The rejection region (R.R) of H_0 is:***

$$F_{\alpha,(k-1),(kn-k)} = F_{0.05,2,21} = 3.47 \Longrightarrow \boxed{RR \in (3.47, \infty)}$$

10. ***The decision is:***

$$F = 5.896 > 3.47 \Longrightarrow \boxed{Reject\ H_0}$$

<u>Q2.</u> Three types of medium sized cars assembled in New Zealand have been test driven by a motoring magazine and compared on a variety of criteria. In the area of fuel efficiency performance, five cars of each brand were each test driven 1000 km; the km per liter data are obtained as follows:

Kilometres per liter						Total	Mean
Brand A	7.6	8.4	8	7.6	8.4	40	8
Brand B	7.8	8	9.1	8.5	9.6	43	8.6
Brand C	9.6	10.4	9.2	9.7	10.6	49.5	9.9

Let the one way ANOVA tabulated as follows:

Source of variation	**Sum of squares**	**Degrees of freedom**	**Mean Squares**	**Test Statistics**
Treatments	SSA	df_{trt}	MSA	
Errors	SSE	df_{er}	MSE	f
Total	SST	df_{tot}		

At a significance level of α =0.05, we want to compare the means of the three groups.

1. ***Write the hypotheses H_0 and H_1. Explain (2 marks).***

$$H_0: \mu_1 = \mu_2 = \mu_3 \quad vs \quad H_1: al\ least\ two\ of\ the\ means\ are\ not\ eqaul.$$

Are the three bands have the same fuel consumption or not?

2. ***The grand mean $\bar{y}_{..}$ is***

(A) 40+43+49.5)/3	(B) (40+43+49.5)/5	(C) (40+43+49.5)/15=8.83333

3. ***The value of SSA is***

$$SS_{trt} = n_i \sum_{i=1}^{k} (\bar{y}_{i.} - \bar{y}_{..})^2$$
$$= 5 \times (8 - 8.83)^2 + 5 \times (8.6 - 8.83)^2 + 5 \times (9.9 - 8.83)^2 = \boxed{9.43}$$

4. $\sum\sum\left(y_{ij}-\bar{y}_{..}\right)^2 = \sum\sum y_{ij}^2 - 15\bar{y}_{..}^2$ *and* $\sum\sum y_{ij}^2 = 1184.11$. *Then SST is:*

$$\sum\sum y_{ij}^2 - 15\bar{y}_{..}^2 = 1184.11 - 15(8.8333)^2 = 13.69$$

5. *The value of SSE is*

$$SSE = SST - SSA = 13.69 - 9.43 = \boxed{4.26}$$

6. *The degrees of freedom of the treatments (df_{trt}) is*

$$k-1 = 3-1 = 2$$

7. *The degrees of freedom of the error (df_{er}) is*

$$k(n-1) = 3(5-1) = 12$$

8. *The degrees of freedom of the total (df_{tot}) is*

$$kn-1 = 3\times 5-1 = 14$$

9. *The Mean Squares of the treatments (MSA) is*

$$MSA = \frac{SSA}{k-1} = \frac{9.43}{3-1} = 4.72$$

10. *The Mean Squares of the errors (MSE) is*

$$MSE = \frac{SSE}{kn-k} = \frac{4.26}{15-3} = 0.355$$

11. *The value of the test statistic F is*

$$F = \frac{MSA}{MSE} = \frac{4.72}{0.355} = 13.296$$

12. *The rejection region (R.R) of H_0 is*

$$F_{\alpha,(k-1),(kn-k)} = F_{0.05,2,14} = 3.89 \Rightarrow \boxed{RR \in (3.89,\infty)}$$

13. *The decision is*

$$F = 13.296 > 3.89 \Rightarrow \boxed{Reject\ H_0}$$

CHAPTER 9

- ***Rank Correlation Coefficient (Spearman Correlation Coefficient):***

$$r_S = 1 - \frac{6 \times \sum d_i^2}{n(n^2 - 1)}$$

- ***Find the Rank correlation coefficient for X and Y:***

X	*Y*
50	*1.80*
175	*1.20*
270	*2.00*
375	*1.00*
425	*1.00*
580	*1.20*
710	*0.80*
790	*0.60*
890	*1.00*
980	*0.85*

<u>Solution:</u>

X	*Rank X*	*Y*	*Rank Y*	$d_i^2 = (Rank\ X - Rank\ Y)^2$
50	*1*	*1.80*	*9*	*64*
175	*2*	*1.20*	*7.5*	*30.25*
270	*3*	*2.00*	*10*	*49*
375	*4*	*1.00*	*5*	*1*
425	*5*	*1.00*	*5*	*0*
580	*6*	*1.20*	*7.5*	*2.25*
710	*7*	*0.80*	*2*	*25*
790	*8*	*0.60*	*1*	*49*
890	*9*	*1.00*	*5*	*16*
980	*10*	*0.85*	*3*	*49*
			Total	*285.5*

$$r_S = 1 - \frac{6 \times \sum d_i^2}{n(n^2-1)} = 1 - \frac{6 \times 285.5}{10((10)^2-1)} = \boxed{-0.73}$$