Statistics and Probability Sample Questions set2

Books Legends:
WM = Probability and Statistics for Engineers and Scientists (Ninth Edition) – Walpole, Myers, Myers and Ye.

PSM = Introductory Statistics – Prem S. Mann and Christopher Jay Lacke
SCH = Schaum's Outlines Statistics – Murray R. Spiegel and Larry J. Stephens
SPG = Advanced Practical Statistics – S. P. Gupta
BS = Business Statistics – S. P. Gupta and M. P. Gupta

PMS138/4.1
Experiment, Outcomes, and Sample Space

Definition:
An experiment is a process that, when performed, results in one and only one of many observations. These observations are called the outcomes of the experiment. The collection of all outcomes for an experiment is called a sample space.

[SCH 127/Chap 6: Probability]

Sample space, Events, Probability and Random Variables

1. [Ref: WM 36/Def 2.1, Ex 2.1, 2.2, 2.2 and WM 39/Def 2.2,Ex 2.4]

What is sample space and event in statistics? Explain with examples.

Def 2.1: The set of all possible outcomes of a statistical experiment is called the sample space and is represented by the symbol S.

Def 2.2: An event is a subset of a sample space.

2. [WM9Ed 53/Ex 2.24]

A coin is tossed twice. What is the probability that at least 1 head occurs?

Solution : The sample space for this experiment is S = {HH, HT, TH, TT}.
If A represents the event of at least 1 head occurring, then A = {HH, HT, TH} , then
Number of elements or sample points in S is 4 ; Number of elements or sample points in A is 3
and thus probability of occuring at least 1 head is P(A) = 3/4.

PSM172/ 4.9.2 Addition Rule,
PSM 155/4.6 Independent Versus Dependent Events
PSM 161/4.8 Intersection of Events and the Multiplication Rule

3. Write down the definition of probability and the sum and product rules of probabilities of two events A and B.

Definition of Probability: Probability is a numerical measure of the likelihood that a specific event will occur.

PSM145/Definition

Using Relative Frequency as an Approximation of Probability: If an experiment is repeated n times and an event A is observed f times, then, according to the relative frequency concept of probability, $P(A) = \frac{f}{n}$

SCH139/ Definition of Probability

If P(success) = P(E) = p, P(failure) = P(not E) = q, Then p + q =1.

4. A container has 2 defective and 3 non defective items. If a sample of 3 items are drawn one after another without replacement what is the probability that; (i) the sample will have no defective items, (ii) the sample will have two defective items, (iii) the sample will have at least one defective items. (iv) the sample will have at most one defective item.

5. The probability that a shooter can hit the target is 0.9. If the shooter shoots three times one after another, what is the probability that;

   (a) The shooter will have 3 successes ?
   (b) The shooter will have 1 success ?
   (c) The shooter will have no success ?

6. [SCH 4ed 147/Ex 14]

Write down the sample space of tossing a pair of true dice and observing the “up” faces. Consider the following two events

A: the sum of the “up” faces is 7

B: the figure on the first die is equal to or greater than that on the second die.

Write down the events A and B as a subset of the sample space then find the probabilities P(A) and P(B).

WM50/Theorem 2.5, Ex 2.21

The number of ways of partitioning a set of n objects into r cells with n_1 elements in the first cell, n_2 elements in the second, and so forth, is

$$\binom{n}{n_1, n_2, \cdots n_r} = \frac{n!}{n_1!n_2!\cdots n_r!}$$

Where $n_1 + n_2 + \cdots + n_r = n$.

Ex 2.21: In how many ways can 7 graduate students be assigned to 1 triple and 2 double hotel rooms during a conference?

Solution: The total number of possible partitions would be

$$\binom{7}{3, 2, 2} = \frac{7!}{3!\ 2!\ 2!} = 210.$$

Combinations, $\binom{n}{r, n-r}$, is usually shortened to $\binom{n}{r}$; since the number of elements in the second cell must be n - r.

7. [WM 67/Ex 2.38]

A small town has one fire engine and one ambulance available for emergencies. The probability that the fire engine is available when needed is 0.98, and the probability that the ambulance is available when called is 0.92. In the event of a resultant of an injury resulting from a burning building, find the probability that both the ambulance and fire engine will be available.

Solution:

Let the events are

A = the fire engine available

B = the ambulance available

Then the probability of availability of both is

$$P(A \cap B) = P(AB) = P(A)\,P(B) = (0.98)(0.92) = 0.9016 \quad \text{(Ans)}$$

8. [SCH 136/Ex 6.8]

One bag contains 4 white balls and 2 black balls; another bag contains 3 white balls and 5 black balls. If one ball is drawn from each ,bag find the probability that (a) both are white, (b) both are black and (c) one is white and one is black.

[SCH 4ed 141/Ex 6-7]
See book for examples of sum rules of probability

9. [WM 65/Ex 2.37]

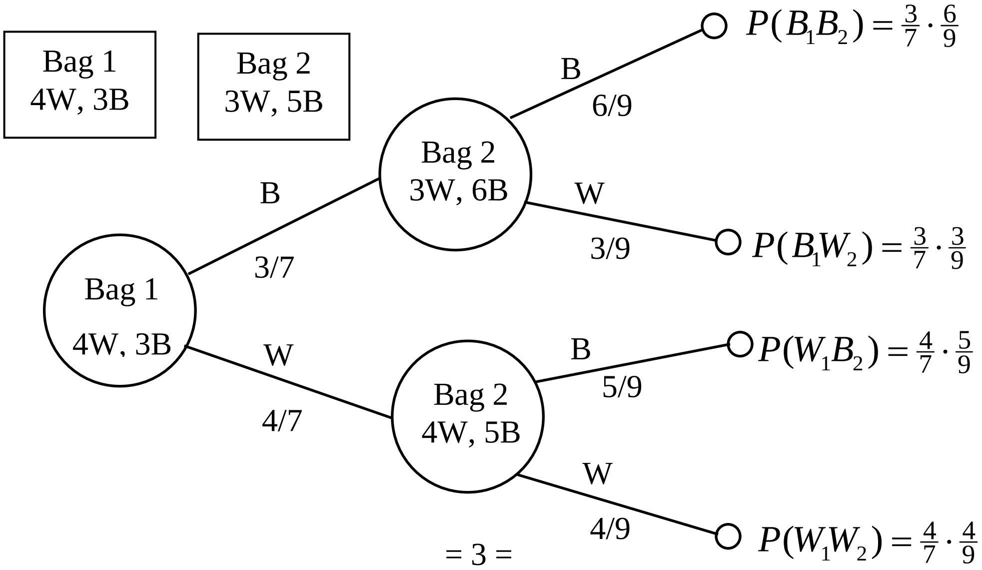
One bag contains 4 white balls and 3 black balls, and a second bag contains 3 white balls and 5 black balls. One ball is drawn from the first bag and placed unseen in the second bag. What is the probability that a ball now drawn from the second bag is (i) black? (ii) white?

Solution:

Let B_1 = Black ball from bag1 ; B_2 = Black ball from bag2

W_1 = White ball from bag1 ; W_2 = White ball from bag2

Given :



Now, the required probability is

(i) $P(B_1B_2 \text{ or } W_1B_2) = P(B_1B_2) + P(W_1B_2)$

$= P(B_1)\,P(B_2 / B_1) + P(W_1)\,P(B_2 / W_1)$

$= \left(\frac{3}{7}\right)\left(\frac{6}{9}\right) + \left(\frac{4}{7}\right)\left(\frac{5}{9}\right) = \frac{38}{63}$ (Ans)

(ii) (Do it yourselves; Ans. 25/63)

10. [SPG 566/Ex 12]

A bag contains 8 green and 10 white balls. Two drawings of 4 balls are made such that (a) the balls are replaced before the second trial (b) the balls are not replaced before the second trial. Find the probability that the first drawing will give 4 green and the second 4 white balls jn each case.

Solution. (a) When balls are replaced before the second trial.

Total no. of balls in the bag = 8 + 10 = 18.

4 balls can be drawn from 18 in ${}^{18}C_4$ ways

The no. of ways in which 4 green balls can be drawn = ${}^{8}C_4$.

The no. of ways in which 4 white balls can be drawn = ${}^{10}C_4$

The probability of drawing 4 green at first trial is $\frac{{}^{8}C_4}{{}^{18}C_4} = \frac{7}{306}$

and the probability of drawing 4 white at the second trial is $\frac{{}^{10}C_4}{{}^{18}C_4} = \frac{7}{102}$

The chance of compound event is $\frac{7}{306} \times \frac{7}{102} = \frac{49}{31212}$ (Ans)

(b) When balls are not replaced.

The probability of 4 green balls at first trial is $\frac{{}^{8}C_4}{{}^{18}C_4} = \frac{7}{306}$ (as above)

When 4 green balls have been drawn and removed, the bag contains 4 green and 10 white balls.

At the second trial, 4 balls can be drawn in ${}^{14}C_4$ ways

and 4 white balls can be drawn in ${}^{10}C_4$ ways

Therefore the chance of white at the second trial is $\frac{{}^{10}C_4}{{}^{14}C_4} = \frac{30}{143}$

The chance of compound event is $\frac{7}{306} \times \frac{30}{143} = \frac{35}{7293}$ (Ans)

11. [SPG 568/Ex 15]

A speaks out truth in 80% cases and B in 85% of the cases. In what percentage of cases are they likely to contradict each other in stating the same fact ?

Solution. The probability that A speaks the truth and B a lie is

$$\frac{80}{100} \times \frac{15}{100} = \frac{3}{25}$$

The probability that B speaks the truth and A a lie is

$$\frac{85}{100} \times \frac{20}{100} = \frac{17}{100}$$

The total probability = $\frac{3}{25} + \frac{17}{100} = \frac{29}{100}$

$\therefore$ The percentage of cases in which they contradict each other is 29 percent.

PSM 150/ 4.4 Marginal and Conditional Probabilities

12. [PSM 169/Ex 4.83]

Two thousand randomly selected adults were asked whether or not they have ever shopped on the Internet. The following table gives a two-way classification of the responses obtained.

	Have Shopped	Have Never Shopped
Male	500	700
Female	300	500

a. Suppose one adult is selected at random from these 2000 adults. Find the following probabilities.
i. P(has never shopped on the Internet and is a male)
ii. P(has shopped on the Internet and is a female)
b. Mention what other joint probabilities you can calculate for this table and then find those. You may draw a tree diagram to find these probabilities.
Solution.

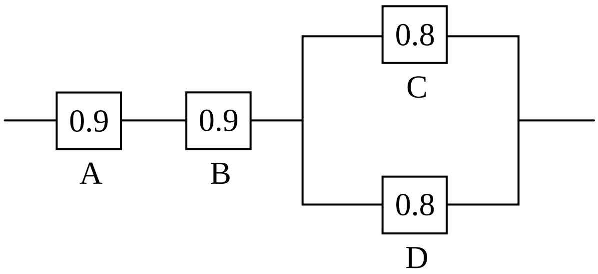
	S	$\bar{S}$	
M	500	700	1200
F	300	500	800
	800	1200	2000

(i) P(has never shopped on the Internet and is a male) = $P(M\bar{S}) = \frac{700}{2000} = \frac{7}{20}$

(ii) P(has shopped on the Internet and is a female) = $P(FS) = \frac{300}{2000} = \frac{3}{20}$

13. [WM 67/Ex 2.39]

An electrical system consists of four components as illustrated in the figure below. The system works if component A and B work and if either of the components C or D work. The reliability (probability of working) of each component is also shown in figure. Find the probability that (a) the entire system works, and (b) the component C does not work, given that the entire system works. Assume that four component works independently.



Solution:

In this configuration of the system, A, B, and the subsystem C and D constitute a serial circuit system, whereas the subsystem C and D itself is parallel circuit system.

(a) Clearly the probability that the entire system works can be calculated as the following:

$$\begin{aligned} P((A\cap B)\cap(C\cup D)) &= P(A)\,P(B)\,P(C\cup D)\,, \quad \text{[As the events are independent]} \\ &= P(A)\,P(B)[1-P(C'\cap D')] \\ &= P(A)\,P(B)[1-P(C')\,P(D')] \\ &= (0.9)(0.9)[1-(1-0.8)(1-0.8)] \\ &= (0.9)(0.9)[1-(0.2)(0.2)] = 0.7776 \qquad \text{(Ans)} \end{aligned}$$

(b) To calculate the conditional probability in this case, we have

$$\begin{aligned} P &= \frac{P(\text{the system works but C does not work})}{P(\text{the system works})} \\ &= \frac{P(A\cap B\cap C'\cap D)}{P(\text{the system works})} \\ &= \frac{(0.9)(0.9)(1-0.8)(0.8)}{0.7776} = 0.1667 \qquad \text{(Ans)} \end{aligned}$$

Formula for Conditional probability:

$P(AB) = P(A \cap B) = P(A)\, P(B/A) \quad \cdots\cdots \quad (1)$

or, $P(B/A) = P(AB)/P(A) = P(BA)/P(A) \quad \cdots\cdots \quad (2)$

14. [WM 9Ed 75/Theorem 2.14]

Bayes' Rule:

If the events B_1 , B_2, . . . , B_k constitute a partition of the sample space S such that $P(B_i) \neq 0$ for i = 1, 2, . . . , k, then for any event A in S such that $P(A) \neq 0$,

$$P(B_r / A) = \frac{P(B_r \cap A)}{\sum_{i=1}^{k} P(B_i \cap A)} = \frac{P(B_r)P(A/B_r)}{\sum_{i=1}^{k} P(B_i)P(A/B_i)} \quad , \quad \text{for } i = 1, 2, \ldots, k \; .$$

Proof:

Conditional probability:

$P(AB) = P(A \cap B) = P(A)\, P(B/A) \quad \cdots\cdots \quad (1)$

or, $P(B/A) = P(AB)/P(A) = P(BA)/P(A) \quad \cdots\cdots \quad (2)$

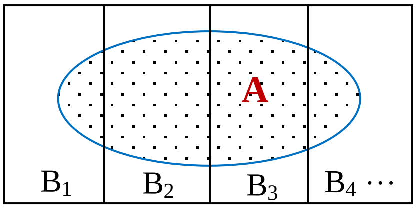


Figure: Partitioning the sample space S.

$A = B_1A \cup B_2A \cup \ldots \cup B_k A$

$$\therefore \; P(A) = P(B_1A) + P(B_2A) + \cdots + P(B_k A) = \sum_{i=1}^{k} P(B_iA) = \sum_{i=1}^{k} P(B_i \cap A) \quad \cdots\cdots \quad (3)$$

From (2),

$$P(B_r / A) = \frac{P(B_rA)}{P(A)} = \frac{P(B_r \cap A)}{\sum_{i=1}^{k} P(B_i \cap A)} = \frac{P(B_r)P(A/B_r)}{\sum_{i=1}^{k} P(B_i)P(A/B_i)} \quad , \quad \text{for } i = 1, 2, \ldots, k$$

which completes the proof.

PSM162/Ex 4–23

PSM163/Ex 4–24

A box contains 20 DVDs, 4 of which are defective. If two DVDs are selected at random (without replacement) from this box, what is the probability that both are defective?

[**Hints.** $P(D_1 \text{ and } D_2) = P(D_1)\, P(D_2/D_1) = 4/20 \,.\, (3/19) = \mathbf{0.0316}$
Use tree diagram to get the same result]

PSM165/Ex 4–25

The probability that a randomly selected student from a college is a senior is .20, and the joint probability that the student is a computer science major and a senior is .03. Find the conditional probability that a student selected at random is a computer science major given that the student is a senior.

PSM166/Ex 4 -27
The probability that a patient is allergic to penicillin is .20. Suppose this drug is administered to three patients.
(a) Find the probability that all three of them are allergic to it.
(b) Find the probability that at least one of them is not allergic to it.
[**Hints.** use tree diagram]

PSM175/Ex 4–33
Consider the experiment of rolling a die twice. Find the probability that the sum of the numbers obtained on two rolls is 5, 7, or 10.

The events that give the sum of two numbers equal to 5 or 7 or 10 are shaded in the table. As we can observe, the three events "the sum is 5," "the sum is 7," and "the sum is 10" are mutually exclusive. Four outcomes give a sum of 5, six give a sum of 7, and three outcomes give a sum of 10. Thus,

P(sum is 5 or 7 or 10) = P(sum is 5) + P(sum is 7) + P(sum is 10)

$$= 4/36+6/36+3/36 = 13/36 = \mathbf{0.3611}$$

PSM175/Ex 4–34
The probability that a person is in favor of genetic engineering is .55 and that a person is against it is .45. Two persons are randomly selected, and it is observed whether they favor or oppose genetic engineering.
(a) Draw a tree diagram for this experiment.
(b) Find the probability that at least one of the two persons favors genetic engineering.

15. [WM 81/Def 3.1, Ex 3.1]

What is random variable? Explain with example.

Random variable:

A random variable is a function that associates a real number with each element in the sample space.

Example:

Experiment, E: Toss a true coin three times and observe the sequence of heads and tails.
Sample Space, S: {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

If X is the number of heads showing, then
X (HHH) =3
X (HHT) =2
X (HTH) =2
X (HTT) =1
X (THH) =2
X(THT)= 1
X (TTH)=1
X (TTT)= 0

The range space R is {x: x = 0, 1, 2, 3}.
Here x is a random variable.

16. [WM 84, 85, 89, 90/Def 3.4, 3.5, 3.6, 3.7, Ex 3.8, 3.10, 3.11]

(a) Write the definition of probability distribution and probability density function with examples.

(b) Write down the definition of cumulative probability distribution function or distribution function with example.

Def 3.4: **Probability distribution**
The set of ordered pairs $(x, f(x))$ is a probability function, probability mass function, or probability distribution of the discrete random variable X if, for each possible outcome x,

1. $f(x) \geq 0$,
2. $\sum_x f(x) = 1$,
3. $P(X = x) = f(x)$

Def 3.6: **Probability density function**
The function f(x) is a probability density function (pdf) for the continuous random variable X, defined over the set of real numbers **R**, if

1. $f(x) \geq 0$, for all $x \in \mathbf{R}$.
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. $P(a < X < b) = \int_a^b f(x)dx$

Def 3.5: **Cumulative distribution function** (for discrete random variable)
The cumulative distribution function $F(x)$ of a discrete random variable X, with probability distribution $f(x)$ is

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t), \text{ for } -\infty < x < \infty .$$

Def 3.7: **Cumulative distribution function**
The cumulative distribution function $F(x)$ of a continuous random variable X, with probability density function $f(x)$ is

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)dx, \text{ for } -\infty < x < \infty .$$

Then $\dfrac{dF(x)}{dx} = f(x)$, if the derivative exists, and $P(a < X < b) = F(b) - F(a)$

17. [WM9Ed 84/Ex 3.8]

A shipment of 20 similar laptop computers to a retail outlet contains 3 that are defective. If a school makes a random purchase of 2 of these computers, find the probability distribution for the number of defectives.

Solution: Let X be a random variable whose values x are the possible numbers of defective computers purchased by the school. Then x can only take the numbers 0, 1, and 2.

Now

$f(0) = P(X=0) = \dfrac{{}^{3}C_0 \, {}^{17}C_2}{{}^{20}C_2} = \dfrac{68}{95}$, $f(1) = P(X=1) = \dfrac{{}^{3}C_1 \, {}^{17}C_1}{{}^{20}C_2} = \dfrac{51}{190}$

$f(2) = P(X=2) = \dfrac{{}^{3}C_2 \, {}^{17}C_0}{{}^{20}C_2} = \dfrac{3}{190}$

Thus, the probability distribution of X is

x	0	1	2
$f(x)$	68/95	51/190	3/190

(Ans)

18. [WM 89-90/Ex 3.11, 3.12]

Suppose that the error in the reaction temperature in ^{o}C, for a controlled laboratory experiment is a continuous random variable X having the probability density function f(x) given by

$$f(x) = \begin{cases} \dfrac{x^2}{3} & , \quad -1 < x < 2 \\ 0 & , \quad \text{elsewhere} \end{cases} .$$

(a) Verify that $\int_{-\infty}^{\infty} f(x)dx = 1$ (b) Find the cumulative distribution function F(x), and use it to evaluate $P(0 < x \leq 1)$

Solution:

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{-1} f(x)dx + \int_{-1}^{2} f(x)dx + \int_{2}^{\infty} f(x)dx$$

$$= 0 + \int_{-1}^{2} \frac{x^2}{3} dx + 0$$

$$= \left[\frac{x^3}{9} \right]_{-1}^{2} = \frac{2^3 - (-1)^3}{9} = 1$$

The Cumulative distribution function is F(x)

For, $-1 \leq x < 2$

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

$$= 0 + \int_{-1}^{x} \frac{x^2}{3} dx = \left[\frac{x^3}{9} \right]_{-1}^{x} = \frac{x^3+1}{9}$$

So,

$$F(x)=\begin{cases} 0 \quad , \quad \text{for } x<-1 \\ \dfrac{x^3+1}{9} \quad , \quad \text{for } -1\le x<2 \\ 1 \quad , \quad \text{for } x\ge 2 \end{cases}$$

Now

$$P(0<x\le 1)=\int_0^1 f(x)dx=F(1)-F(0)=\frac{2}{9}-\frac{1}{9}=\frac{1}{9}$$

19. [WM 92/ Ex 3.7]

The total number of hours, measured in units of 100 hours, that a family runs a vacuum cleaner over a period of one year is a continuous random variable X that has the density function

$$f(x)=\begin{cases} x \quad , \quad 0\le x<1 \\ 2-x \quad , \quad 1\le x<2 \\ 0 \quad , \quad \text{elsewhere} \end{cases} .$$

Find the probability that over a period of one year, a family runs a vacuum cleaner (a) less than 120 hours; (b) between 50 and 100 hours. [Ans. (a)0.68; (b) 0.375]

20. Question:

A random variable *x* has the probability density function *f*(*x*) is given by

$$f(x)=\begin{cases} x \quad , \quad 0\le x<1 \\ 2-x \quad , \quad 1\le x<2 \\ 0 \quad , \quad \text{elsewhere} \end{cases} .$$

Show graphically and find out the probabilities; (a) $P(-1<x<1/2)$, (b) $P(x\le 3/2)$ (c) $P(X\le 3)$, (d) $P(1/4<X<3/2)$.

Solution: (a)

$$P(-1<x<1/2)=\int_{-1}^{0} f(x)dx+\int_0^{1/2} f(x)dx$$

$$=0+\int_0^{1/2} xdx=\left[\frac{x^2}{2}\right]_0^{1/2}=\frac{1}{2}\left(\frac{1}{4}-0\right)=\frac{1}{8} \quad \text{(Ans)}$$

[Note: Draw the related graphs yourselves, ask the teacher if required]

[Note: Solve other questions yourselves]

21. Question:

Let X be the density function

$$f(x)=\begin{cases} \dfrac{3}{4}\left(1-x^2\right) \quad , \quad -1\le x\le 1 \\ 0 \quad , \quad \text{otherwise} \end{cases}$$

a. Find the distribution function.
b. Find $P(1/4 \le x \le 2)$
c. Find X such that $P(X \le x) = 0.95$

22. [WM9Ed 93/ Exr 3.21]

Consider the density function $f(x) = \begin{cases} k\sqrt{x}, & 0 < x < 1 \\ 0, & \text{elsewhere} \end{cases}$

(a) Evaluate k.
(b) Find F(x) and use it to evaluate P(0.3 < X < 0.6).

Mathematical Expectation and Probability Distributions

23. [SCH 130, 139/ Ex 6.15]

(i) Define Mathematical Expectation. Find (a) E(X), (b) E(X^2), and (c) $E[(X - \bar{X})^2]$ for the following probability distribution

X	8	12	16	20	24
P(X)	1/8	1/6	3/8	1/4	1/12

[Ref: WM101/Def 3.1, WM102/Ex1, WM132/Def 4.1, WM115/Ex4.5, WM123/Ex 4.12]

(ii) What is random variable? Explain with example. Define expected value E(X) of a random variable X. Let X be a random variable with density function

$$f(x) = \begin{cases} \dfrac{x^2}{3}, & -1 < x < 2 \\ 0, & \text{elsewhere} \end{cases}$$

(a) Find the expected value of $g(x) = 4x + 3$. and (b) Find variance of $g(x)$.

24. [WM 171/ Th^{m} 5.2 or BS 488, SCH 163/ Ex 7.10-7.11]

Show that the mean and variance of the binomial distribution $b(x;n,p)$ are μ = np and σ^2 = npq.

25. [BS 492 or WM 187/ Th^{m} 5.5]

Show that the mean and variance of the poisson distribution p(x;m) are μ = m and σ^2 = m.
[**Note:** If the poisson distribution is written as p(x;λt) that is m = λt, then μ = λt and σ^2 = λt.]

26. [WM 199/ No.5]

Show that for the normal distribution N(μ,σ), mean = μ and variance = σ^2.

27. [BS 488/ Ex 2]

The mean of a binomial distribution is 40 and standard deviation 6. Calculate n, p, q.

Solution: np = 40, σ^2 = npq = $(6)^2$ = 36,

$\therefore q = 36/40 = 0.9$, $p = 1- q = 0.1$ and $n = 40/p = 40/0.1= 400$

28. [SCH 163/ Ex 7.11-7.12]

If a variable is binomially distributed, show that its mean $\mu = np$ and variance $\sigma^2 = npq$. Use the above results to find the mean and standard deviation of defective bolts in a total of 400 when the probability of a defective bolt is 0.1.

Solution: $\mu = np = 400(0.1) = 40$, $\sigma^2 = npq = 400(0.1)(0.9) = 36$, so, $\sigma = 6$ Ans.

Binomial Distribution

Binomial Distribution is $f(x) = b(x;n,p) = {}^nC_x\, p^x\, q^{n-x}$, $x = 0, 1, 2, \ldots, n$.

With Mean, $\mu = np$ and variance $\sigma^2 = npq$.

Poisson Distribution

The Poisson distribution is given by

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

Where m is called the parameter of the distribution. Probability of x is $P(x) = f(x)$

$$\text{Mean}, \mu = E(x) = \sum x f(x) = \sum_{x=0}^{\infty} x \frac{e^{-m} m^x}{\lfloor x}$$

$$= 0 + me^{-m} + m^2 e^{-m} + 3\frac{e^{-m} m^3}{\lfloor 3} + \cdots\cdots$$

$$= me^{-m}\left(1 + m + \frac{m^2}{\lfloor 2} + \cdots\cdots\right)$$

$$= me^{-m} . e^{m} = m$$

Thus the mean of Poisson distribution is m

$$E(x^2) = E[x(x-1) + x] = E[x(x-1)] + E(x) = E[x(x-1)] + m$$

$$E[x(x-1)] = \sum_{x=0}^{\infty} [x(x-1)\frac{e^{-m} m^x}{\lfloor x} + \sum_{x=2}^{\infty} \frac{e^{-m} m^x}{\lfloor x-2}$$

Let $x–2 = r$, then

$$E[x(x-1)] = \sum_{r=0}^{\infty} \frac{e^{-m} m^{r+2}}{\lfloor r} = m^2 e^{-m} \sum_{r=0}^{\infty} \frac{m^r}{\lfloor r} = m^2 e^{-m} e^{m} = m^2$$

$\therefore$ $E(x^2) = E[x(x-1)] + m = m^2 + m$

Hence the Variance, $\sigma^2 = E(x^2) – (E(x))^2 = (m^2 + m) – m^2 = m$ (Proved)

Normal Distribution

One of the most important examples of a continuous probability distribution is normal distribution.

In probability theory, the normal (or Gaussian) distribution is a continuous probability distribution, defined on the entire real line, that has a bell-shaped probability density function, known as the Gaussian function or informally as the bell curve:

It is defined by the equation

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} = N(\mu,\sigma)$$

Where μ = mean and σ = standard deviation. The total area under the normal curve and above the x-axis is 1.

When the variable x is expressed in terms of standard units $z = \dfrac{x-\mu}{\sigma}$ then the above equation can be expressed in standard form as

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} = N(0,1)$$

Note: x~ N(μ,σ) means x is normally distributed with mean μ and standard deviation σ.
x~ N(0,1) means x is normally distributed with mean 0 and standard deviation 1.

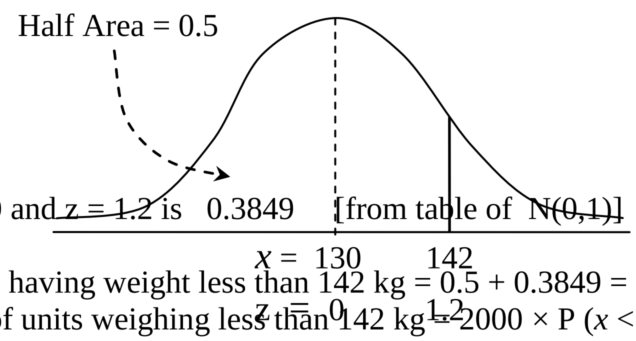
The distribution with μ= 0 and σ^2 = 1 is called the **standard normal distribution** or the **unit normal distribution**.

29. [Ref: BS 515, 14Ed444/ Ex 26]

A workshop produces 2000 units per day. The average weight of units is 130 kg with standard deviation of 10 kg. Assuming normal distribution, how many units are expected to weigh less than 142 kg?

Solution: We have μ = 130, σ = 10, N = 2000, x = 142

$$\therefore z = \frac{x-\mu}{\sigma} = \frac{142 - 130}{10} = 1.2$$



Area between z = 0 and z = 1.2 is 0.3849 [from table of N(0,1)]

Probability of units having weight less than 142 kg = 0.5 + 0.3849 = 0.8849

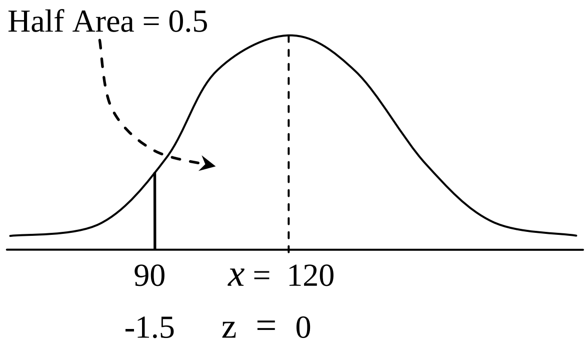
Expected number of units weighing less than 142 kg = 2000 × P (x <142) = 2000 × 0.8849
= 1769.8 or 1770 (Ans)

30. [Ref: BS 14Ed445/ Ex 28]

1000 tube lights with mean life of 120 days are installed in a new factory, their length of life is normally distributed with standard deviation 20 days. (i) How many tube lights will expire in less than 90 days? (ii) If it is decided to replace all the tube lights together, what interval should be allowed between replacements if not more than 10 percent should expire before replacement?

Solution: (i) We have $\mu = 120$, $\sigma = 20$, $N = 1000$, $x = 90$

$$\therefore z = \frac{x-\mu}{\sigma} = \frac{90 - 120}{20} = -1.5$$



Area under the curve between z = 0 and z = -1.5 is 0.4332 [from table of N(0,1)]

Area to the left of z = -1.5 is (0.5 – 0.4332) = 0.0668 .
Number of tube lights expected to expire in less than 90 days = 1000 × P (x <90) = 1000 × 0.0668
= 66.8 or 67 (Ans)

(ii) The value of standard normal variate corresponds to an area (0.5 – 0.1) = 0.4 is 1.28 .

Now, $\frac{x-120}{20} = -1.28$ or, x = 94.4 or 94.

Hence, 10 percent of the tube lights will have to be replaced after 94 days (Ans)

PSM341/ 8.1 Estimation: An Introduction

Estimation is a procedure by which a numerical value or values are assigned to a population parameter based on the information collected from a sample.

Definition

Estimation: The assignment of value(s) to a population parameter based on a value of the corresponding sample statistic is called estimation.

In inferential statistics, is called the true population mean and p is called the true population proportion. There are many other population parameters, such as the median, mode, variance, and standard deviation.

Definition

Estimate and Estimator: The value(s) assigned to a population parameter based on the value of a sample statistic is called an estimate. The sample statistic used to estimate a population parameter is called an estimator.

The estimation procedure involves the following steps.
1. Select a sample.
2. Collect the required information from the members of the sample.
3. Calculate the value of the sample statistic.
4. Assign value(s) to the corresponding population parameter.

Remember, the procedures to be learned in this chapter assume that the sample taken is a **simple random sample**.

PSM342/ 8.2 Point and Interval Estimates
An estimate may be a point estimate or an interval estimate. These two types of estimates are described in this section.

PSM342/ 8.2.1 A Point Estimate
If we select a sample and compute the value of the sample statistic for this sample, then this value gives the point estimate of the corresponding population parameter.

Definition
Point Estimate: The value of a sample statistic that is used to estimate a population parameter is called a point estimate.

Thus, the value computed for the sample mean, $\bar{x}$, from a sample is a point estimate of the corresponding population mean, μ.

For the example, suppose the Census Bureau takes a sample of 10,000 households and determines that the mean housing expenditure per month, $\bar{x}$, for this sample is \$1970. Then, using $\bar{x}$ as a point estimate of μ, the Bureau can state that the mean housing expenditure per month, μ, for all households is about \$1970. Thus,
Point estimate of a population parameter = Value of the corresponding sample statistic

PSM342/ 8.2.2 An Interval Estimate
In the case of interval estimation, instead of assigning a single value to a population parameter, an interval is constructed around the point estimate, and then a probabilistic statement that this interval contains the corresponding population parameter is made.

Definition
Interval Estimation: In interval estimation, an interval is constructed around the point estimate, and it is stated that this interval is likely to contain the corresponding population parameter.

PSM343/Definition
Confidence Level and Confidence Interval: Each interval is constructed with regard to a given confidence level and is called a *confidence interval*. The confidence interval is given as

Point estimate ± Margin of error

The confidence level associated with a confidence interval states how much confidence we have that this interval contains the true population parameter. The confidence level is denoted by $(1 - \alpha)100\%$.

When expressed as probability, it is called the confidence coefficient and is denoted by $1 - \alpha$. In passing, note that α is called the significance level.

PSM346/ Ex8-1
A publishing company has just published a new college textbook. Before the company decides the price at which to sell this textbook, it wants to know the average price of all such textbooks in the market. The research department at the company took a sample of 25 comparable textbooks and collected information on their prices. This information produced a mean price of \$145 for this

sample. It is known that the standard deviation of the prices of all such textbooks is $35 and the population of such prices is normal.

(a) What is the point estimate of the mean price of all such college textbooks?
(b) Construct a 90% confidence interval for the mean price of all such college textbooks.

Test of Hypothesis, Chi square test and Goodness of Fit

31. [SCH 216-218/ Defs]

What do you mean by Test of Hypothesis? Explain about the (i) Null hypothesis and Alternative hypothesis (ii) Type I and type II errors (iii) Level of significance, (iv) One tailed and two tailed tests.

Ans.
A hypothesis is an assumption to be tested. The statistical testing of hypothesis is most important technique in statistical inference. Hypothesis tests are widely used in business, engineering and in industry for making decisions.

A statistical hypothesis test is a method of making decisions using data, whether from a controlled experiment or an observational study (not controlled). Hypothesis testing is sometimes called confirmatory data analysis. In statistics, a result is called statistically significant if it is unlikely to have occurred by chance alone, according to a pre-determined threshold probability, the significance level. The phrase "test of significance" was coined by Ronald Fisher: "Critical tests of this kind may be called tests of significance, and when such tests are available we may discover whether a second sample is or is not significantly different from the first.

Null Hypothesis (H_0)

Null hypothesis is denoted by H_0 . The null hypothesis asserts that there is no true difference in the sample statistic and population parameter under consideration. Hence the word "Null" means invalid, void or amounting to nothing, and the difference found is accidental arising out of fluctuations of sampling.

Alternative Hypothesis (H_1)

Alternative hypothesis is denoted by H_1. The hypothesis that is different from the null hypothesis is the hypothesis is the alternative hypothesis. If the sample information leads us to reject H_0 , then we will accept the alternative hypothesis H_1. Thus the two hypothesis areconstructed so that, if one is true, the other is false.

Type I and type II errors

When a statistical hypothesis tested, there are four possible situations:

	Conditions	
Decision	**H_0 is True**	**H_0 is False**
Accept H_0	Correct Decision	Type II Error
Reject H_0	Type I Error	Correct Decision

The probability of committing a type I error is designated as " α " and is called the level of significance.

So, α = Pr [Type I Error]
= Pr [Rejecting H_0 / H_0 is true]

Complement of α is
$(1 - \alpha)$ = Pr [Accepting H_0 / H_0 is true]
This probability $(1 - \alpha)$ corresponds to the concept of $100(1 - \alpha)$ % confidence interval.

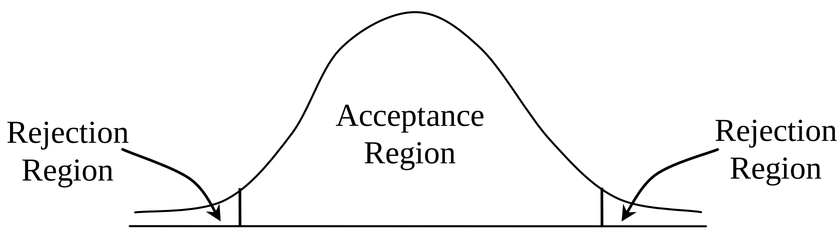
Similarly, the probability of committing a type II error is designated by "β " and is called the level of significance.

So, β = Pr [Type II Error]
= Pr [Accepting H_0 / H_0 is false]

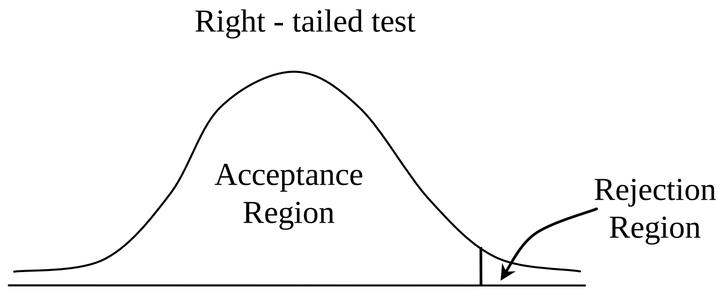
Complement of β is
$(1 - \beta)$ = Pr [Rejecting H_0 / H_0 is false]
This probability $(1 - \beta)$ is known as the power of a statistical test.

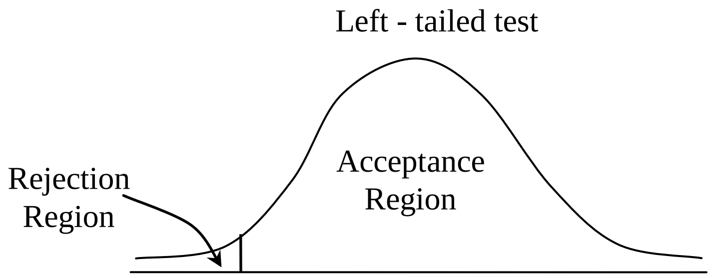
One tailed and two tailed tests

Two tailed tests



One tailed tests





Test of Hypothesis

Testing Population mean $\bar{x}$

H_0 : $\mu = \mu_0$.

$\bar{x} \sim \mathrm{N}(\mu, \sigma_{\bar{x}})$

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}, \qquad \text{where, } \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$$= \frac{S}{\sqrt{n}} \quad (\text{If } \sigma \text{ is unknown for large samples})$$

Testing the difference between two means

$$z = \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}}, \qquad \text{where, } \theta = \mu_1 - \mu_2$$

$$\hat{\theta} = \text{ estimate of } \theta.$$

Since the best unbiased estimator of $\theta = \mu_1 - \mu_2$ is $\bar{x}_1 - \bar{x}_2$

$\sigma_{\hat{\theta}}$, is the standard deviation of the sampling distribution of $\bar{x}_1 - \bar{x}_2$ is given by

$$\sigma_{\hat{\theta}}^2 = \sigma_{\bar{x}_1-\bar{x}_2}^2 = \text{var}(\bar{x}_1) + \text{var}(\bar{x}_2)$$

$$= \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$\therefore$ The z statistic is given by

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

The null hypothesis is H_0 : $\mu_1 - \mu_2 = 0$. Then the z statistic is reduced to

$$z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

At 5 % level of significance, z = $\pm$ 1.96 for two tailed test.

32. [BS 585/ Ex 1 ; BS 14Ed2012 505/ Ex 1]

The mean lifetime of a sample of 100 light tube produced by a company is found to be 1580 hours with standard deviation of 90 hours. Test the hypothesis that the mean lifetime of the tubes produced by the company is 1600 hours.

[Use 5 % level of significance, for which the tabulated value of z = $\pm$ 1.96 for two tailed test]

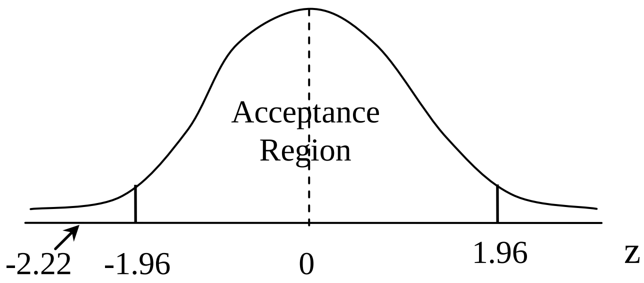
Ans. The null hypothesis is that there is no significant difference between the sample mean and the hypothetical population mean.

That is
H_0 : $\mu = \mu_0$.
H_1 : $\mu \neq \mu_0$.

$$z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}}, \qquad \text{where, } \sigma_{\bar{x}} = \frac{S}{\sqrt{n}} \quad [\text{Since } \sigma \text{ is unknown for large samples}]$$

Here, $\bar{x} = \mu$ = 1580, μ_0 = 1600, sample size, n = 100
Standard deviation for sample is s = 90

$$\therefore z = \frac{1580 - 1600}{90/\sqrt{100}} = -2.22$$



The critical value of $z = \pm 1.96$ for two tailed test at 5% level of significance. Since the computed value of $z = -2.22$ falls in the rejection region, we reject the null hypothesis.

Hence the mean lifetime of the tubes produced by the company may not be 1600 hours.

33. [BS 585/ Ex 2; BS 14Ed2012 506/ Ex 2]

You are working as a purchase manager of a company. The following information has been supplied to you by two manufacturers of electric bulbs:

	Company A	**Company B**
Mean life (in hours):	1300	1288
Standard deviation (in hours):	82	93
Sample size:	100	100

Which brand of bulbs are you going to purchase if your desire is to take a risk of 5% ?

Ans. Let us take the hypothesis that there is no significant difference in the quality of the two brands of bulbs produced by the company do not differ significantly.

$$z = \frac{(\overline{x}_1 - \overline{x}_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$, [Since σ_1 and σ_2 are not known and can be replaced by S_1 and S_2]

$$\text{or,} \quad z = \frac{1300 - 1288}{\sqrt{\frac{(82)^2}{100} + \frac{(93)^2}{100}}} = \frac{12}{\sqrt{62.24 + 86.49}} = \frac{12}{12.399} = 0.968$$

Since the computed value of $z = 0.968$ is less than the critical value of $z = 1.96$ at 5% level of significance for two tailed test. We accept the null hypothesis.

Hence the quality of two brands of bulbs do not differ significantly and the purchase manager can purchase any of the two brands of bulbs.

34. [WM 364/ Ex 10.3, 9Ed 338]

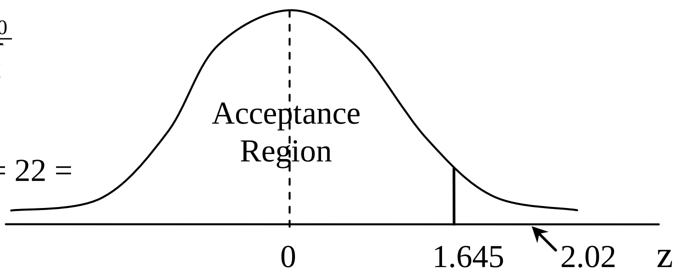
A random sample of 100 recorded deaths in United States during the past year showed an average life span of 71.8 years. Assuming a population standard deviation of 8.9 years, does this seem to indicate that the mean life span today is greater than 70 years? Use a 0.05 level of significance.

Solution Hints. The null and alternative hypotheses are

$H_0: \mu = 70.$
$H_1: \mu > 70.$

Level of significance $\alpha = 0.05$

Critical region : $z > 1.645$ where $z = \dfrac{\overline{x} - \mu_0}{\sigma / \sqrt{n}}$



Calculation

$$z = \frac{71.8 - 70}{8.9/\sqrt{100}} = 2.02$$

Since the calculated value of z = 2.02 is greater than the critical value of z = 1.645 at 5% level of significance for one tailed test. We reject the null hypothesis.

Hence we conclude that the mean lifespan today is greater than 70 years.

35. [WM 364/ Ex 10.4, 9Ed 338]

A manufacturer of sports equipment has developed a new synthetic fishing line that he claims has a mean breaking strength of 8 kilograms with a standard deviation of 0.5 kilogram. Test the hypothesis that $\mu = 8$ kilograms against the alternative hypothesis that $\mu \neq 8$ kilograms if a random sample of 50 lines is tested and found to have a mean breaking strength of 7.8 kilograms. Use a 0.01 level of significance.

Ans. The null hypothesis is that there is no significant difference between the sample mean and the hypothetical population mean.

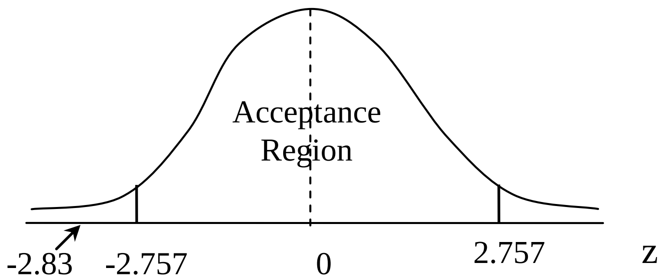
That is
H_0 : $\mu = 8$.
H_1 : $\mu \neq 8$.

Level of significance $\alpha = 0.01$

Critical region : $z < -2.575$ and $z > 2.575$, where $z = \dfrac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$

Calculation

$$z = \frac{7.8 - 8}{0.5/\sqrt{50}} = -2.83$$



The critical value of z = ± 2.575 for two tailed test at 1% level of significance. Since the computed value of z = -2.83 falls in the rejection region, we reject the null hypothesis.

Hence the breaking strength is not equal to 8 but is, in fact, less than 8 kilograms.

36. [SCH 226/ Ex 10.9]

The breaking strengths of cables produced by manufacturer have a mean of 1800 pounds (lb) and a standard deviation of 100 lb. By a new technique in the manufacturing process, it is claimed that the breaking strength can be increased. To test this claim, a sample of 50 cables is tested and it is found that the mean breaking strength is 1850 lb. Can we support the claim at the 0.01 significance level ?

37. [SCH 233/ Ex 10.18]

The mean height of 50 male students who showed above-average participation in college athletics was 68.2 inches (in) with a standard deviation of 2.5 in, while 50 male students who showed no interest in such participation had a mean height of 67.5 in with a standard deviation of 2.8 in. Test the hypothesis that male students who participate in college athletics are taller than other male students.

Chi square (χ^2) test and Goodness of Fit

$$\chi^2 = \sum \frac{(O-E)^2}{E} \quad ; \quad \text{Where O = Observed value, and E = Expected value}$$

38. [SPG 696/ Ex 1]

With the help of the following data find out whether there is any relationship between smoking and drinking.

	Drinking	Not-Drinking
Smoking	74	26
Not- Smoking	10	30

[Tabulated value: For $\nu = 1$, the value of χ^2 is 3.84 at 5 % level of significance]

Solution. Let us take the hypothesis that smoking and drinking are not associated.

Applying χ^2 test

<u>Table of Observed Frequency</u>

	D	$\bar{D}$	
S	74	26	100
$\bar{S}$	10	30	40
	84	56	140

Expected value of DS is

$$\text{Expectation}(DS) = \frac{D \times S}{N} = \frac{84 \times 100}{140} = 60$$

<u>Table of Expected Frequency</u>

	D	$\bar{D}$	
S	60	40	100
$\bar{S}$	24	16	40
	84	56	140

O	E	$(O-E)^2$	$(O-E)^2/E$
74	60	196	3.27
10	24	196	8.17
26	40	196	4.90
30	16	196	12.25
		$\Sigma\,(O-E)^2/E$	= 28.59

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 28.59$$

Degrees of freedom, $\nu = (r-1)(c-1) = (2-1)(2-1) = 1$,
[where r = number of rows, c = number of columns]

For $\nu = 1$, $\chi^2_{0.05} = 3.84$ (from table)

The calculated value of χ^2 is 28.59 which is greater than the table value. Hence the hypothesis is rejected. We therefore, conclude that smoking and drinking are associated.

39. [SPG 698/ Ex 3]

The following data relate to the sales in a time of trade depression of a certain article in wide demand. Do the data suggest that the sales are significantly affected by the depression.

District where sales are	District not hit by depression	District hit by depression
Satisfactory	250	80
Not- Satisfactory	140	30

or,

District where sales are	District not hit by depression	District hit by depression	Total
Satisfactory	250	80	330
Not- Satisfactory	140	30	170
Total	390	110	500

[Tabulated value: For $\nu = 1$, the value of χ^2 is 3.84 at 5 % level of significance]

Solution Hints: Take the hypothesis H_0 : The sales are not affected by depression.

Calculated value of χ^2 = 2.844 which is less than the table value and hence the null hypothesis holds true. We therefore conclude that the sales are not significantly affected by depression.

Solution. Let us take the hypothesis that the sales are not affected by depression.

Applying χ^2 test

Table of Observed Frequency

	D	$\bar{D}$	
S	250	80	330
$\bar{S}$	140	30	170
	390	110	500

D = District not hit by depression, $\bar{D}$ = District hit by depression
S = Satisfactory, $\bar{S}$ = Not- Satisfactory

Expected value of DS is

$$\text{Expectation}(DS) = \frac{D \times S}{N} = \frac{390 \times 330}{500} = 257.4$$

Table of Expected Frequency

	D	$\bar{D}$	
S	257.4	72.6	330
$\bar{S}$	132.6	37.4	170
	390	110	500

O	E	$(O-E)^2$	$(O-E)^2/E$
250	257.4	54.76	0.213
140	132.6	54.76	0.413
80	72.6	54.76	0.754
30	37.4	54.76	1.464
		$\Sigma\,(O-E)^2/E$	= 2.844

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 2.844$$

Degrees of freedom, $\nu = (r-1)(c-1) = (2-1)(2-1) = 1$,

[where r = number of rows, c = number of columns]

For $\nu = 1$, $\chi^2_{0.05} = 3.84$ (from table)

The calculated value of χ^2 is 2.844 which is less than the table value and hence the hypothesis holds true. We therefore conclude that the sales are not significantly affected by depression.

40. [SCH 228/ Ex 10.13]

A Company manufactures rope whose breaking strengths have a mean of 300 pounds (lb) and a standard deviation of 24 lb. It is believed that by a newly developed process the mean breaking strength can be increased.

(a) Design a decision rule for rejecting the old process at the 0.01 significance level if it is agreed to test 64 ropes.

(b) Under the decision rule adopted in part (a), what is the probability of accepting the old process when in fact the new process has increased the mean breaking strength to 310 lb? Assume that the standard deviation is still 24 lb.

41. [SCH 251/10.2]

To test the hypothesis that a coin is fair, adopt the following decision rule:

Accept the hypothesis if the number of heads in a single sample of 100 tosses is between 40 and 60 inclusive. Reject the hypothesis otherwise.

(a) Find the probability of rejecting the hypothesis when it is actually correct.

(b) Graph the decision rule and the result of part (a).

(c) What conclusions would you draw if the sample of 100 tosses yielded 53 heads? And if yielded 60 heads.

(d) Could you be wrong in your conclusion about part (c)? Explain.

42. [PSM 516/11-7]

A researcher wanted to study the relationship between gender and owning cell phones. She took a sample of 2000 adults and obtained the information given in the following table.

	Own Cell Phones	**Do Not Own Cell Phones**
Men	74	26
Women	10	30

At the 5% level of significance, can you conclude that gender and owning a cell phone are related for all adults?

43. [Extra questions]

SCH 160/ Ex 7.3, SCH 162/7.7-7.8, 163/7.10, 165/7.17, 166/7.18, 167/7.20, 168/7.21

WM115/Ex3.36, 146/4.50, 206/6.8, 208/6.11, 6.12

Table 1: Area under the Standard Normal curve from 0 to z

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	**0.3849**	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	**0.4495**	**0.4505**	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	**0.4750**	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.3	0.4893	0.4896	0.4898	**0.4901**	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	**0.4949**	**0.4951**	0.4952
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

Calculation: For 5% level two tailed test; The critical value is $z_{0.025}$
The area from the centre to the z-score = 0.5000 – 0.025 = 0.4750
Looking in the table we observe that which matches a z-score of 1.96.

(In this case each tail contains: $\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$ (2.5%) of the total area under the curve.)

For 1% level, two tailed test; (0.4949+ 0.4951)/2 = 0.4950 ; 0.5 – 0.4950 = 0.005 = half of 0.01;
and the z-score is 2.575.
For 5% level one tailed test; (0.4495 + 0.4505)/2 = 0.45 ; 0.50 – 0.45 = 0.05; and the z-score is 1.645.
For 1% level one tailed test; ; The critical value is $z_{0.01}$; and the z-score is 2.33.
The area from the centre to the z-score = 0.5000 – 0.01 = 0.4900
Looking in the normal curve areas table, the area closest to 0.4900 is 0.4901
which matches a z-score of 2.33.

Common critical scores found in confidence interval and hypothesis test calculations are:
$z_{0.005} = 2.57$, $z_{0.01} = 2.33$, $z_{0.025} = 1.96$, $z_{0.05} = 1.64$.

<u>For a 90% confidence level and two tailed test</u> : The critical value is $z_{0.05}$
The area from the centre to the z-score = 0.5000 – 0.05 = 0.4500
Looking in the table the closest area to 0.4500 is 0.4495 which matches a z-score of 1.64 or we can find the best match which is equal to 1.645.

Table 2: The Chi-Square Distribution

Degrees *of* freedom	**α**									
	0.995	**0.99**	**0.975**	**0.95**	**0.90**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**
1	---	---	0.001	0.004	0.016	2.706	**3.841**	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997

Problems on Binomial Distribution

Binomial Distribution is $f(x) = b(x;n,p) = {}^{n}C_{x}\, p^{x}\, q^{n-x}$, $x = 0, 1, 2, \ldots, n$.

44. [SCH 4Ed2008,P178 / Ex 7.3]

Find the probability that in five tosses of a fair die a 3 appears (a) at no time, (b) once, (c) twice, (d) three times, (e) four times, (f) five times.

Ans. The probability of 3 in a single toss is, p = 1/6 , and the probability of no 3 in a single toss is q = 1 – p = 5/6; [Here P(s) + P(*f*) = p + q = 1. P(3) in a single toss is p = 1/6, so, q = 5/6]

thus:

(a) n = 5 , *x* = 0

Pr(3 occurs zero times) = $^{n}C_{x}\, p^{x}\, q^{n-x} = {}^{5}C_{0}\, p^{0}\, q^{5} = (1)(1)\left(\frac{5}{6}\right)^{5} = \frac{3125}{7776}$

(b) n = 5 , *x* = 1

Pr(3 occurs 1 time) = $^{5}C_{1}\, p^{1}\, q^{4} = (5)\left(\frac{1}{6}\right)^{1}\left(\frac{5}{6}\right)^{4} = \frac{3125}{7776}$

(c) Pr(3 occurs 2 times) = $^{5}C_{2}\, p^{2}\, q^{3} = (10)\left(\frac{1}{6}\right)^{2}\left(\frac{5}{6}\right)^{3} = (10)\left(\frac{1}{36}\right)\left(\frac{125}{216}\right) = \frac{625}{3888}$

(d) Pr(3 occurs 3 times) = $^{5}C_{3}\, p^{3}\, q^{2} = (10)\left(\frac{1}{6}\right)^{3}\left(\frac{5}{6}\right)^{2} = (10)\left(\frac{1}{216}\right)\left(\frac{25}{36}\right) = \frac{125}{3888}$

(e) Pr(3 occurs 4 times) = $^{5}C_{4}\, p^{4}\, q^{1} = (5)\left(\frac{1}{6}\right)^{4}\left(\frac{5}{6}\right)^{1} = (5)\left(\frac{1}{1296}\right)\left(\frac{5}{6}\right) = \frac{25}{7776}$

(f) Pr(3 occurs 5 times) = $^{5}C_{5}\, p^{5}\, q^{0} = (1)\left(\frac{1}{6}\right)^{5}\left(\frac{5}{6}\right)^{0} = (1)\left(\frac{1}{7776}\right)(1) = \frac{1}{7776}$

45. [Lecs Note prob 4; WM/ Ex]

The probability that a shooter can hit the target is 0.9. If the shooter shoots three times one after another.

Ans. The probability of success in a single shoot is, p = P(success) = P(S) = 0.9 , and the probability of failure in a single shoot is, q = P(failure) = P(F) = 1 – p = 0.1.

(a) Pr(3 successes) = P(SSS) = (0.9) (0.9) (0.9) = 0.729 .
(b) Pr(1 successes); the cases are SFF, FFF, FFS

Pr(1 successes) = P(SFF) + P(FSF) + P(FFS) = (0.9) (0.1) (0.1) + (0.1) (0.9) (0.1) + (0.1)(0.1) (0.9) = 3(0.009) = 0.027.

(c) Pr(no successes) = P(FFF) = (0.1) (0.1) (0.1) = 0.001 .

Calculation Using Binomial Distribution

(a) Pr(3 successes in three shoots) ; here, n = 3 , x = 3

Pr(3 successes) = ${}^{n}C_{x}\ p^{x}\ q^{n-x} = {}^{3}C_{3}\ p^{3}\ q^{0} = (1)(0.9)^{3}\ (0.1)^{1} =$ 0.729 .

(b) Pr(1 successes);) ; here, n = 3 , x = 1

Pr(1 successes) = ${}^{3}C_{1}\ p^{1}\ q^{2} = (3)(0.9)^{1}\ (0.1)^{2} = (3)(0.009) = 0.027$.

(c) Pr(no successes) = ${}^{3}C_{0}\ p^{0}\ q^{3} = (1)(0.9)^{0}\ (0.1)^{3} = (1)(1)\ (0.1)^{3} = 0.001$.

The results are same as above.

Extra Problems to Practice

Mann241/8.1 , Mann346/Ex 8-1 , WM266/9.3, WM99/3.18, WM100/Ex3.19,

WM93/Exr3.21, WM121/Ex4.10, 4.11,

WM94/3.4 ,