ADVANCE STATISTICS PROJECT REPORT

Prepared By: Barkha Agarwal

PROBLEM STATEMENT-1

A physiotherapist with a male football team is interested in studying the relationship between foot injuries and the positions at which the players play from the data collected

	Striker	Forward	Attacking Midfielder	Winger	Total
Players Injured	45	56	24	20	145
Players Not Injured	32	38	11	9	90
Total	77	94	35	29	235

1.1 What is the probability that a randomly chosen player would suffer an injury?
1.2 What is the probability that a player is a forward or a winger?
1.3 What is the probability that a randomly chosen player plays in a striker position and has a foot injury?
1.4 What is the probability that a randomly chosen injured player is a striker?
1.5 What is the probability that a randomly chosen injured player is either a forward or an attacking midfielder?

INTRODUCTION TO PROBLEM STATEMENT-1

The purpose of this business report is to understand the relationship between the foot injuries and the position at which the players played. Here the physiotherapist wants to get certain probabilities of the data given to him of the male team. So, here we will be calculating certain probabilities on the basis of data provided.

1.1 What is the probability that a randomly chosen player would suffer an injury?

As we can see from the table provided that the players injured include all the four positions i.e; Striker, Forward, Attacking Midfielder, Winger. Here we will use the formula of addition.

Probability that the random chosen player would suffer an injury is. is **61.7%**

1.2 What is the probability that a player is a forward or a winger?

The total no. of forward players and the total no. of winger are taken here for calculating the probability. Here we will calculate the probability by taking the total no. of players.

Probability that the player is forward or winger is **52.3%**

1.3 What is the probability that a randomly chosen player plays in a striker position and has a foot injury?

The total no. of players in striker position and is injured is taken here and also the
The total no. of players is taken into account.

Probability that the randomly chosen player plays in a striker position and has a foot injury is **19.1%**

1.4 What is the probability that a randomly chosen injured player is a striker?

The total no. of players in striker position is injured and the total no. of players injured is taken into account.

Probability that the randomly chosen injured player is striker is **31.0%**

1.5 What is the probability that a randomly chosen injured player is either a forward or an attacking midfielder?

The total no. of players in forward position and is injured and the total no. of positions in attacking midfielder and is injured are taken into account. Also, the total no. of players injured are taken here.

Probability that the randomly chosen injured player is either forward or attacking midfielder is **55.2%**

PROBLEM STATEMENT-2

An independent research organization is trying to estimate the probability that an accident at a nuclear power plant will result in radiation leakage. The types of accidents possible at the plant are, fire hazards, mechanical failure, or human error. The research organization also knows that two or more types of accidents cannot occur simultaneously.

According to the studies carried out by the organization, the probability of a radiation leak in case of a fire is 20%, the probability of a radiation leak in case of a mechanical 50%, and the probability of a radiation leak in case of a human error is 10%. The studies also showed the following;

The probability of a radiation leak occurring simultaneously with a fire is 0.1%.
The probability of a radiation leak occurring simultaneously with a mechanical failure is 0.15%.
The probability of a radiation leak occurring simultaneously with a human error is 0.12%.

On the basis of the information available, answer the questions below:

2.1 What are the probabilities of a fire, a mechanical failure, and a human error respectively?
2.2 What is the probability of a radiation leak?
2.3 Suppose there has been a radiation leak in the reactor for which the definite cause is not known. What is the probability that it has been caused by:

A Fire.
A Mechanical Failure.
A Human Error.

INTRODUCTION TO PROBLEM STATEMENT-2

In this problem statement we are trying to figure out the probability of an accident at a nuclear plant resulting in radiation leak. The types of accidents possible are:

Probability of fire=0.20
Probability of mechanical failure=0.50
Probability of human error=0.10
Probability radiation leak simultaneously with fire=0.001
Probability radiation leak simultaneously with mechanical failure=0.0015
Probability radiation leak simultaneously with human error=0.0012

2.1 What are the probabilities of a fire, a mechanical failure, and a human error respectively?

P(F)= (0.001/0.2) = 0.005

P(M)= (0.0015/0.50) = 0.003

P(H)= (0.0012/0.1) = 0.012

2.2 What is the probability of a radiation leak?

The probability of radiation leak P(R) = (0.001+0.0015+0.0012+0) =0.0037

2.3 Suppose there has been a radiation leak in the reactor for which the definite cause is not known. What is the probability that it has been caused by:

A Fire.

P(F|R) = (0.001/0.0037) = 0.027027027

A Mechanical Failure.

P(M|R) = (0.0015/0.0037) = 0.40540541

A Human Error.

P(H|R) = (0.0012/0.0037) = 0.32432432

PROBLEM STATEMENT-3

The breaking strength of gunny bags used for packaging cement is normally distributed with a mean of 5 kg per sq. centimeter and a standard deviation of 1.5 kg per sq. centimeter. The quality team of the cement company wants to know the following about the packaging material to better understand wastage or pilferage within the supply chain; Answer the questions below based on the given information

3.1 What proportion of the gunny bags have a breaking strength less than 3.17 kg per sq cm?
3.2 What proportion of the gunny bags have a breaking strength at least 3.6 kg per sq cm.?
3.3 What proportion of the gunny bags have a breaking strength between 5 and 5.5 kg per sq cm.?
3.4 What proportion of the gunny bags have a breaking strength NOT between 3 and 7.5 kg per sq cm.?

INTRODUCTION TO PROBLEM STATEMENT-3

We can see that the problem statement is on normal distribution and here the mean and the standard deviation is mentioned. In order to solve for a normal distribution, we can never calculate the probability of x at a given value or at a particular point. Hence, the questions asked here are either less than, at least, between, not between.

3.1 What proportion of the gunny bags have a breaking strength less than 3.17 kg per sq cm?

Here we will calculate the z score of 3.17 kg with respect to mean of 5 kg having a standard deviation of 1.5 kg.

So, for a z score of -1.22, what is the area to the left of the curve is we need to find out.

Therefore, 0.1112 proportion of the gunny bags have a breaking strength less than 3.17 kg per sq cm.

3.2 What proportion of the gunny bags have a breaking strength at least 3.6 kg per sq cm.?

Firstly, we will calculate the z score which is -0.933 here and then we will calculate the proportion of gunny bags having a breaking strength of atleast 3.6 kg per sq cm.

Therefore, 0.8246 proportion of the gunny bags have a breaking strength at least 3.6 kg per sq cm**.**

3.3 What proportion of the gunny bags have a breaking strength between 5 and 5.5 kg per sq cm.?

Here, we will calculate the z scores for both 5 and 5.5 kg per sq cm and then substract the same from each other to arrive at the chance between the two.

Therefore, 0.13055 proportion of the gunny bags have a breaking strength between 5 and 5.5 kg pe sq cm**.**

3.4 What proportion of the gunny bags have a breaking strength NOT between 3 and 7.5 kg per sq cm.?

Therefore, proportion of the gunny bags have a breaking strength NOT between 3 and 7.5 kg per sq cm is 0.13900

PROBLEM STATEMENT-4

Grades of the final examination in a training course are found to be normally distributed, with a mean of 77 and a standard deviation of 8.5. Based on the given information answer the questions below.

4.1 What is the probability that a randomly chosen student gets a grade below 85 on this exam?
4.2 What is the probability that a randomly selected student scores between 65 and 87?
4.3 What should be the passing cut-off so that 75% of the students clear the exam?

INTRODUCTION TO PROBLEM STATEMENT-4

In this problem statement we can see that the mean and the standard deviation is given and the data given is normally distributed. This problem is to find the probabilities of the grades of the final examination in a training course.

4.1 What is the probability that a randomly chosen student gets a grade below 85 on this exam?

In this question we will use the cdf function to calculate the probability.

The probability that a randomly chosen student gets a grade below 85 on this exam is **82.67%**

4.2 What is the probability that a randomly selected student score between 65 and 87?

Here, we will use the cdf for both 65 and 87. Then subtract the same from each other to arrive at the chance between the two.

The probability that a randomly selected student scores between 65 and 87 is **80.13%**

4.3 What should be the passing cut-off so that 75% of the students clear the exam?

Here we are using the percent point function(ppf) to solve this problem. The passing cut off so that 75% of the students clear the exam is **71.27**

PROBLEM STATEMENT-5

Zingaro stone printing is a company that specializes in printing images or patterns on polished or unpolished stones. However, for the optimum level of printing of the image the stone surface has to have a Brinell's hardness index of at least 150. Recently, Zingaro has received a batch of polished and unpolished stones from its clients. Use the data provided to answer the following (assuming a 5% significance level)

5.1 Earlier experience of Zingaro with this particular client is favorable as the stone surface was found to be of adequate hardness. However, Zingaro has reason to believe now that the unpolished stones may not be suitable for printing. Do you think Zingaro is justified in thinking so?
5.2 Is the mean hardness of the polished and unpolished stones the same?

INTRODUCTION TO PROBLEM STATEMENT-5

In this problem statement Zingaro the stone printing company specializes in printing images or patterns on polished or unpolished stones. The hardness index of the stone surface for optimum level of printing is said to be 150 which we can say is the mean here.
The level of significance(alpha) is said to be 5% i.e, 0.05.

5.1 Earlier experience of Zingaro with this particular client is favorable as the stone surface was found to be of adequate hardness. However, Zingaro has reason to believe now that the unpolished stones may not be suitable for printing. Do you think Zingaro is justified in thinking so?

Defining Null and Alternate hypothesis

Null hypothesis(H0): mean Brinell's hardness index of unpolished stones is >= 150

Alternate hypothesis(H1): mean Brinell's hardness index of unpolished stones is < 150

Level of significance(alpha)= 0.05
Sample size=75
We don't know the standard deviation here

Since the p value is less than 0.05, we reject the null hypothesis. Therefore, the batch of unpolished stone is not suitable for printing.

5.2 Is the mean hardness of the polished and unpolished stones the same?

H(0): Average hardness of polished stone= Average hardness of unpolished stone

Vs

H(1): Hardness are not equal
Here we perform two tailed test

Here p value is less than 0.05 therefore we reject the null hypothesis and the mean hardness of polished and unpolished are not the same

PROBLEM STATEMENT-6

Aquarius health club, one of the largest and most popular cross-fit gyms in the country has been advertising a rigorous program for body conditioning. The program is considered successful if the candidate is able to do more than 5 push-ups, as compared to when he/she enrolled in the program. Using the sample data provided can you conclude whether the program is successful? (Consider the level of Significance as 5%)
Note that this is a problem of the paired-t-test. Since the claim is that the training will make a difference of more than 5, the null and alternative hypotheses must be formed accordingly.

INTRODUCTION TO PROBLEM STATEMENT-6

In this problem statement we have to perform paired-t-test and form null and alternate hypothesis. The level of significance(alpha) here is 5% i.e, 0.05 and here it is mentioned that the body conditioning will only be considered successful if candidates are able to do more than 5 push ups, as compared to when he/she enrolled for the program.

This is a problem of the paired-t-test. Since the claim is that the training will make a difference of more than 5, the null and alternative hypotheses must be formed accordingly

Paired test is best performed when we want to get a before and after sort of result. The test statistic of paired-t-test is:

$$z = x_{diff} / (s_{diff}/\sqrt{n})$$

where:
xdiff: sample mean of the differences
s: sample standard deviation of the differences
n: sample size (i.e. number of pairs)
degrees of freedom=n-1

Null hypothesis(H0)=5 push ups

Alternate hypothesis(H1)>5 push ups
Level of significance(alpha)=0.05
Sample size=100

We have two samples and we do not know the standard deviation.
Sample sizes for both samples are same.
The sample is large sample, n > 30.

We can conclude by performing the test that the p-value is less than alpha i.e; mean push ups is more than 5. We have enough evidence to reject the null hypothesis in favour of alternative hypothesis.

PROBLEM STATEMENT-7

Dental implant data: The hardness of metal implant in dental cavities depends on multiple factors, such as the method of implant, the temperature at which the metal is treated, the alloy used as well as on the dentists who may favour one method above another and may work better in his/her favourite method. The response is the variable of interest.

1.Test whether there is any difference among the dentists on the implant hardness. State the null and alternative hypotheses. Note that both types of alloys cannot be considered together. You must state the null and alternative hypotheses separately for the two types of alloys.?

2.Before the hypotheses may be tested, state the required assumptions. Are the assumptions fulfilled? Comment separately on both alloy types.?

3.Irrespective of your conclusion in 2, we will continue with the testing procedure. What do you conclude regarding whether implant hardness depends on dentists? Clearly state your conclusion. If the null hypothesis is rejected, is it possible to identify which pairs of dentists differ?

4.Now test whether there is any difference among the methods on the hardness of dental implant, separately for the two types of alloys. What are your conclusions? If the null hypothesis is rejected, is it possible to identify which pairs of methods differ?

5.Now test whether there is any difference among the temperature levels on the hardness of dental implant, separately for the two types of alloys. What are your conclusions? If the null hypothesis is rejected, is it possible to identify which levels of temperatures differ?

6.Consider the interaction effect of dentist and method and comment on the interaction plot, separately for the two types of alloys?

7.Now consider the effect of both factors, dentist, and method, separately on each alloy. What do you conclude? Is it possible to identify which dentists are different, which methods are different, and which interaction levels are different?

INTRODUCTION TO PROBLEM STATEMENT-7

In this problem statement we will perform two sample ANOVA as there are multiple factors which we have to look upon in the dental cavities.

1.Test whether there is any difference among the dentists on the implant hardness. State the null and alternative hypotheses. Note that both types of alloys cannot be considered together. You must state the null and alternative hypotheses separately for the two types of alloys.?

Hypothesis:

A factor of the form H0: 1 =.. = a

The means of the variable A categories are identical.

Factor B: H0:.. = b

The means of the variable B categories are identical. H0: Interaction(AiBj) = 0 (where I = 1 for a and j = 1 for b). The effects of variable A on the means of the cells do not rely on the effects of variable B, and vice versa, for all the cells. This shows that there is no interaction between variables A and B.