Great Learning

Name-Abhishek Pradhan

Contents

Problem 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**

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

Ans: Probability that a randomly chosen player would suffer an injury = Total Number of Players Injured/ Total Number of Players = 145/235 = 0.62.

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

Probability that a player is a forward or a winger = (Number of Forwards + Number of Wingers) */ Total Number of Players = (94+29)/235 = 0.52.

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

Probability that a randomly chosen player plays in a striker position and has a foot injury = Number of Injured Strikers/ Total Number of Players = 45/235 = 0.19

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

Probability that a randomly chosen injured player is a striker = Number of Injured Strikers/ Total Number of Injured Players = 45/145 = 0.31

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

Probability that a randomly chosen injured player is either a forward or an attacking midfielder = (Number of Injured Forwards + Number of Injured Attacking Midfielders**)/ Total Number of Injured Players = (56+24)/145 = 80/145 = 0.55.

Problem 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?

Fire=0.2

Mechanical_failure=0.5

Human_error=0.1

R_Fire=0.001

R_Mechanical_failure=0.0015

R_Human_error=0.0012

Fire Hazards	0.250
Mechanical Failure	0.625
Human Error	0.125

2.2 What is the probability of a radiation leak?
Probability of a radiation leak is 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.
- A Mechanical Failure.
- A Human Error.

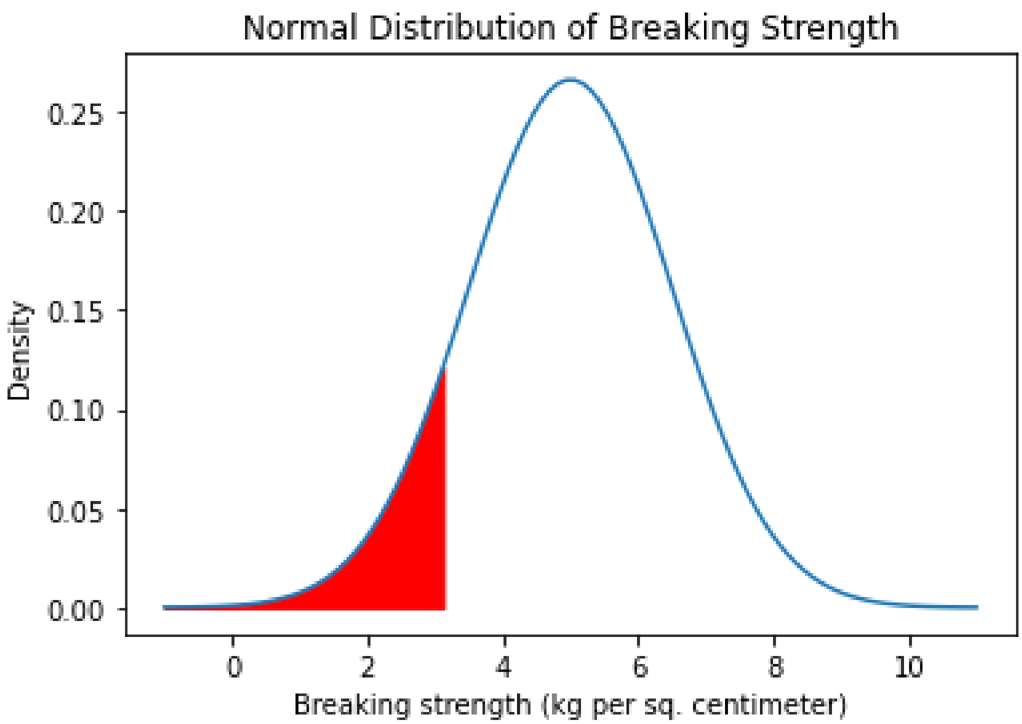
Probability of a radiation leak by Fire: 0.2702702702702703
Probability of a radiation leak by Mechanical Failure: 0.4054054054054054
Probability of a radiation leak by Human Erro: 0.3243243243243243

Problem 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; **(Provide an appropriate visual representation of your answers, without which marks will be deducted)**

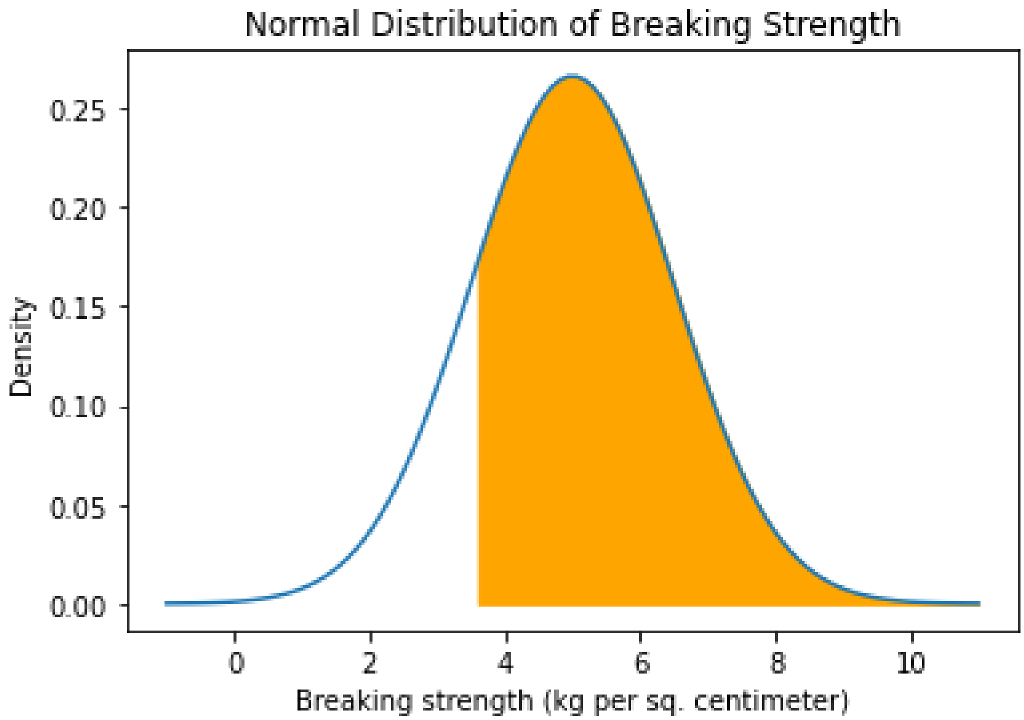
3.1 What proportion of the gunny bags have a breaking strength less than 3.17 kg per sq cm?

Probability Distribution Function of Breaking Strength of Gunny Bags is given as, X~N ($\mu = 5$, $\sigma = 1.5$) 3.1 What proportion of the gunny bags have a breaking strength less than 3.17 kg per sq cm? P ($X < 3.17$) = 0.11

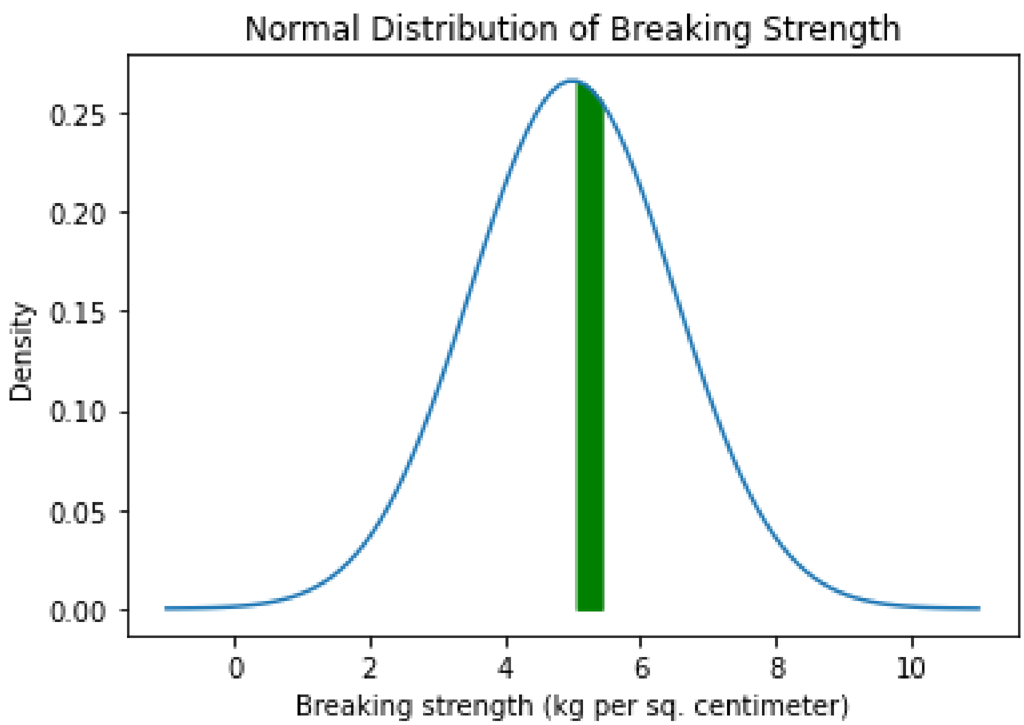


3.2 What proportion of the gunny bags have a breaking strength at least 3.6 kg per sq cm.?

P ($X >= 3.6$) = 1 - P (X



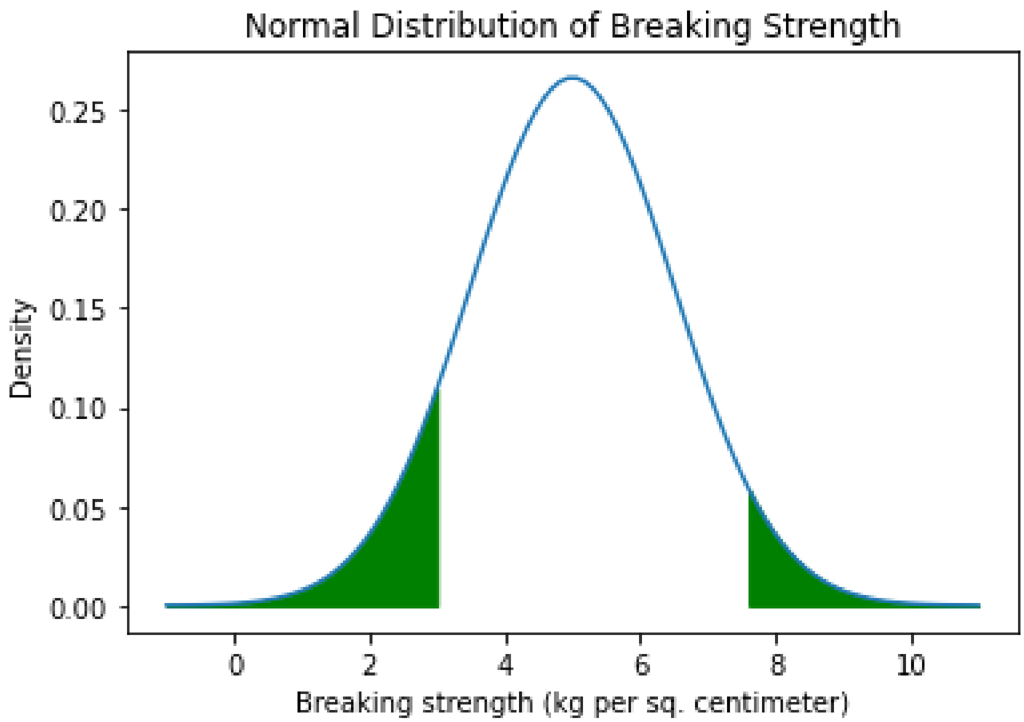
3.3 What proportion of the gunny bags have a breaking strength between 5 and 5.5 kg per sq cm.?



P=0.13

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

P =0.14



Problem 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?
The probability that a randomly chosen student gets a grade below 85 on this exam: 0.8266927837484748

4.2 What is the probability that a randomly selected student scores between 65 and 87?
The probability that a randomly chosen student gets a grade below 85 on this exam: 0.8266927837484748

4.3 What should be the passing cut-off so that 75% of the students clear the exam?
The passing cut-off so that 75% of the students clear the exam: 82.7331628766667

Problem 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?

First, hypothesizing for unpolished stones, we have:

Ho (Null Hypothesis) -> Sample Mean >=150 H1 (Alternate Hypothesis) -> Sample Mean 30 μ = 150 and n=75. Also,

$\bar{X}$unpolished= 134.11, S unpolished = 33.04, $\bar{X}$polished= 147.79, S polished = 15.59.

p-value for unpolished~0 < 0.05 (i.e. α) and hence Null Hypothesis is rejected in that case. So, unpolished stones do not have a Brinell's hardness index of at least 150. Whilst, in case of treated and polished p-value is 0.11 > 0.05 (i.e. α) and hence Null Hypothesis cannot rejected in that case. So, treated, and polished stones have a Brinell's hardness index of at least 150.

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

Hypotheses - Null: mean polished = mean unpolished

Alternate : mean polished $\neq$ mean unpolished

Test: Two-tailed t-test, Also, $\bar{X}$unpolished= 134.11, S unpolished = 33.04, $\bar{X}$polished= 147.79, S polished = 15.59 and n =75. The p-value of the two-tailed test is 0.0007328 and is significantly less than 0.05 (i.e. α), so alternate hypothesis prevails.

Dataset - Link

Problem 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.

Hypotheses - Null: mean difference<= 5 Alternate: mean difference> 5

Given, $\alpha = 0.05$

Sample sizes for both samples are the same. We have two paired samples and we do not know th e population standard deviation. The sample is not a very large sample, n = 100. So, I use the t di stribution and the tSTAT test statistic for two sample paired test. Step 4: Calculating the p-value a nd the t-statistic (Please refer to attached. ipynb file for details) - t stat is -19.323 and Paired two-sample t-test p-value= 2.0039823815285965e-14. We have enough evidence to reject the null hy pothesis in favour of alternative hypothesis

Dataset - Link

Problem 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.

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.?

> Null: meandentist1 = meandentist2 = … = meandentist5 Alternate: meandentist1 ≠ meandentist2 ≠ … ≠ meandentist5

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

> alloy1 and alloy 2 respectively in this case, for hypothesis testing the independence assumption between the independent variables viz., dentist, method, temperature needs to be tested to avoid Multicollinearity problem. Hence, Interaction(V1iV2j) = 0 (wherein V1 and V2 can be Dentist, Method, Alloy, and Temp having 5, 3, 2, 3 values respectively in the dataset).

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?

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?

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?

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

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?