Advanced Statistics Graded Project

-Chetan R Deshpande

Contents Page No.

Problem 1

	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?

The probability that a randomly chosen player would suffer an injury is **0.6170212765957447.** Therefore, 61.7% of players have suffered from injury.

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

The probability that a player is a forward or a winger is **0.5234042553191489.**

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

The probability that a player plays in a striker position and has a foot injury is **0.19148936170212766.**

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

The probability that a randomly chosen injured player is a striker is **0.3103448275862069.**

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

The probability that a randomly chosen injured player is either a forward or an attacking midfielder is **0.5517241379310345.**

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?

a) the probabilities of a fire is **0.25**
b) the probabilities of a mechanical failure is **0.625**
c) the probabilities of a human error is **0.125**

2.2 What is the probability of a radiation leak?

Probability radiation leak is **0.37**

(We arrive at probability of leak by summing up the probabilities of fire, mechanical failure and human error)

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.

a) the probability that it has been caused by a leak with fire is **0.27**
b) the probability that it has been caused by a leak with mechanical failure is **0.41**
c) the probability that it has been caused by a leak with human error is **0.32**

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.

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

Here,

Mu = 5

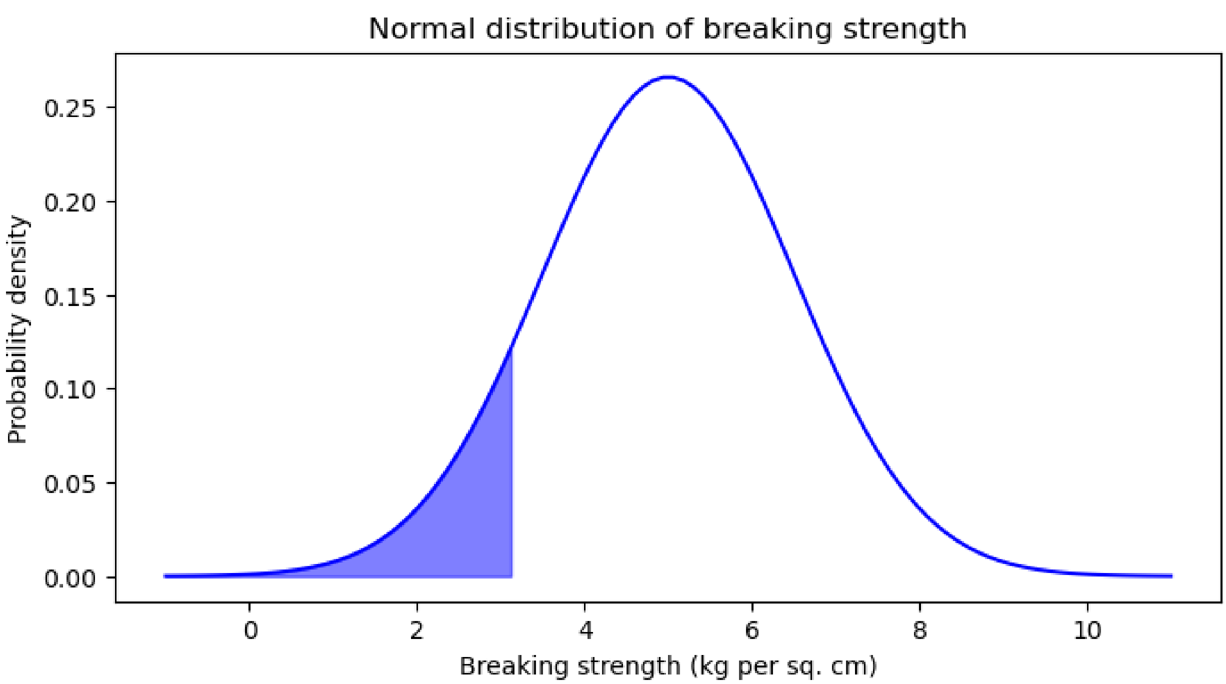
Sigma = 1.5

X = 3.17

We get z as -1.22 further to which we have obtained the cumulative probability for **z = -1.22** as **0.11**

Therefore, the proportion of the gunny bags that have a breaking strength less than 3.17 kg per sq cm is 0.11.

[formula and function used here are z=(X-Mu/Sigma) and stats.norm.cdf(X,Mu,Sigma) respectively].



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

Here,

Mu = 5

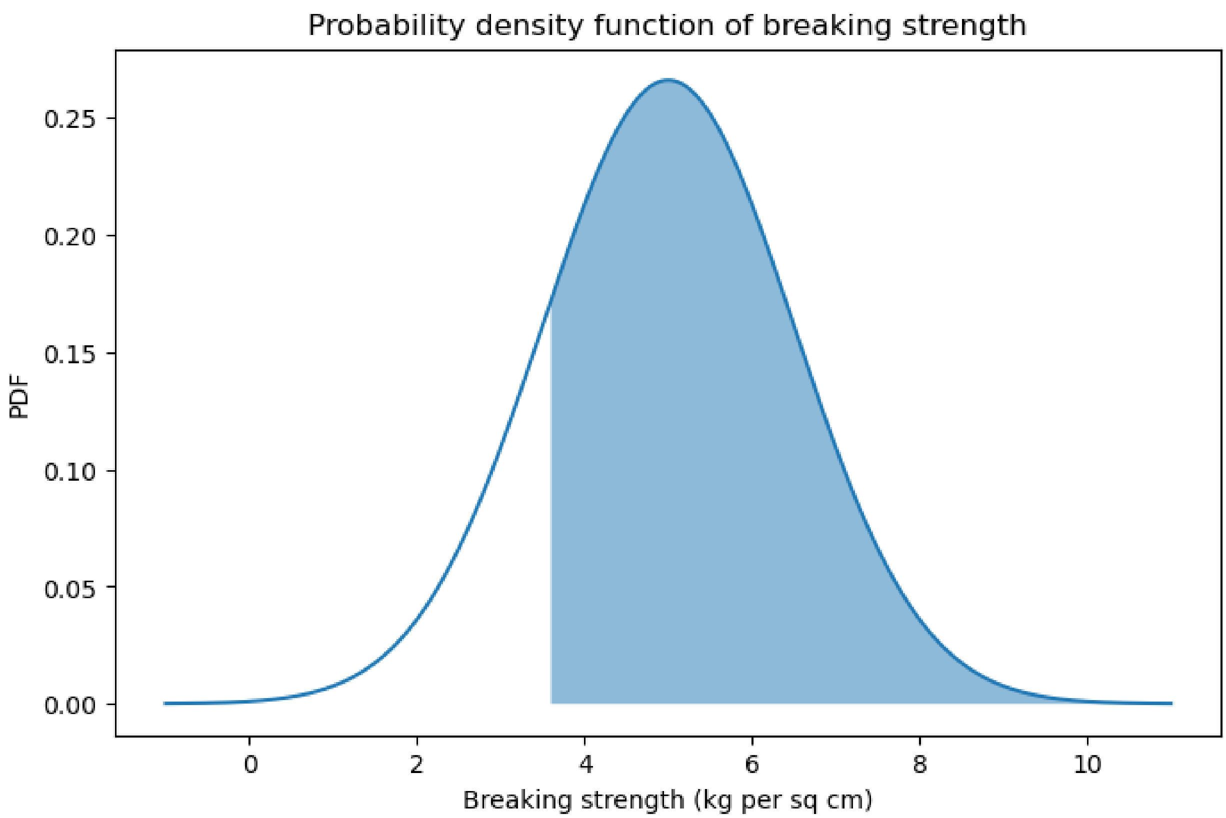
Sigma = 1.5

X = 3.6

We get **z = -0.93** and the cumulative probability for z = -0.93 is: **0.82.**

This proves that the proportion of the gunny bags that have a breaking strength at least 3.6 kg per sq cm is 0.82.

[formula and function used here are z=(X-Mu/Sigma) and 1-stats.norm.cdf(X,Mu,Sigma) respectively].

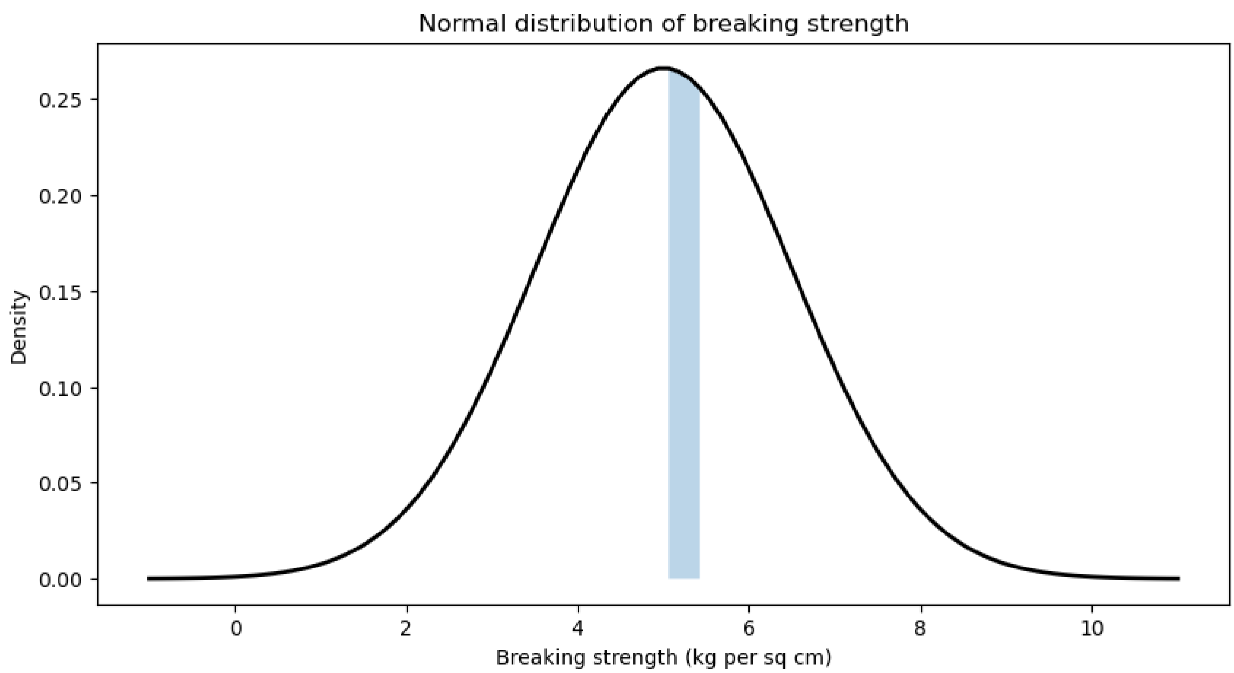


3.3 What proportion of the gunny bags have a breaking strength between 5 and 5.5 kg per sq cm.?

Here, we have got the 2 z scores, z1 and z2 as **0.0 & 0.33** respectively.

Then using scipy.stats.norm, we have gotten the answer as **0.13.**

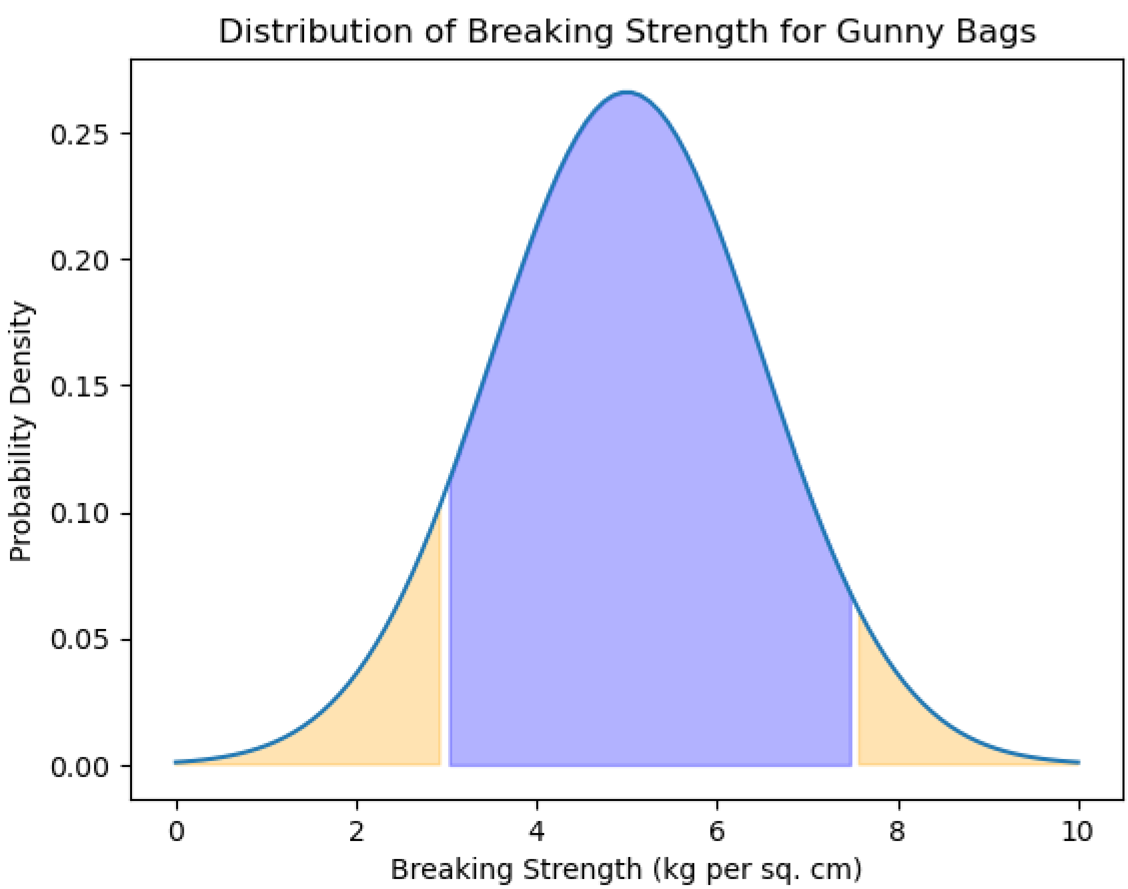
Hence, the proportion of gunny bags that have a breaking strength between 5 and 5.5 kg per sq cm is 0.13



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

Here, first we find out what is the proportion between 3 and 7.5 kg per sq cm and then subtract that answer by 1 to arrive at a proportion not between 3 and 7.5 kg per sq cm.

Therefore, the proportion of bags not between 3 and 7.5 kg per sq cm is 0.13900157199868257



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?

From the formula z=X-Mu/Sigma we get a z score of **0.94.** Further with the help of stats.norm.cdf, we get **0.8267.**

So this means the probability that a randomly chosen student gets a grade below 85 on this exam is 0.8267 or 82.67%.

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

From the formula z=X-Mu/Sigma we get a z score of z1 and z2 as -1.41 and 1.17 respectively. Further with the help of stats.norm.cdf, we get **0.079 and 0.88**.

After subtracting z2 with z1 we get **0.801.**

So this means the probability that a randomly selected student scores between 65 and 87 is 0.801 or 80.1%.

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

By using norm.ppf, we have found the z score to be **0.674**.

Further by using **passing_score = mean + z_score * std_dev,** we have arrived at **82.7331628766667.**

This means, the passing cut-off so that 75% of the students clear the exam is 82.729.

Problem 6:

Aquarius health club, one of the largest and most popular crossfit 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.

Here,

Ho : the mean difference in push-ups before and after the program is less than or equal to 5

Ha : the mean difference is greater than 5

Alpha = 0.05

Further, we use t_statistic = (x_bar - 5) / (s / np.sqrt(n)) to find the t statistic value and 1 - t.cdf(t_statistic, df=n-1) to find the p value.

From calculating using the above formulae we get

t-statistic: **1.9148542155126758**

p-value: **0.029198872141011245**

As p<Alpha, we reject the null hypothesis and conclude that the program is successful in increasing the number of push-ups.

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.

7.1 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.?

Stating the null and alternative hypotheses for each type of alloy:

For Alloy 1:

- **Null hypothesis: There is no significant difference in the implant hardness among the dentists who use Alloy 1.**
- **Alternative hypothesis: There is a significant difference in the implant hardness among the dentists who use Alloy 1.**

For Alloy 2:

- **Null hypothesis: There is no significant difference in the implant hardness among the dentists who use Alloy 2.**
- **Alternative hypothesis: There is a significant difference in the implant hardness among the dentists who use Alloy 2.**

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

The assumptions for conducting an independent samples t-test are:

1. **Independence:** The observations should be independent of each other.
2. **Normality:** The data should be normally distributed.
3. **Homogeneity of variances:** The variances of the two groups should be equal.

For Alloy 1 & Alloy 2:

1. **Independence:** It can be assumed that the observations are independent.
2. **Normality:** If the histogram or Q-Q plot appears to be roughly bell-shaped, then the normality assumption is likely to be met.
3. **Homogeneity of variances:** If the p-value of the test is greater than the significance level (alpha 0.05), then the assumption of homogeneity of variances is met.