



10/16/2022

Advanced Statistics Project

# Advanced Statistics



**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** |

**1.1** What is the probability that a randomly chosen player would suffer an injury?...........................................4

**1.2** What is the probability that a player is a forward or a winger?.....................................................................4

**1.3** What is the probability that a randomly chosen player plays in a striker position and has a foot injury? .4

**1.4** What is the probability that a randomly chosen injured player is a striker?..................................................4

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

midfielder?...........................................................................................................................................................4

**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?..........................5

**2.2** What is the probability of a radiation leak?.....................................................................................................6

**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: ……………………………………………………………………………………….…………………...6

* A Fire.
* A Mechanical Failure.
* A Human Error.

**Problem 3:**

The breaking strength of gunny bags used for packaging cement is normally distributed with a mean of 5 kg per sq. centimetre and a standard deviation of 1.5 kg per sq. centimetre. 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?.......................6

**3.2** What proportion of the gunny bags have a breaking strength at least 3.6 kg per sq.cm.?............................6

**3.3** What proportion of the gunny bags have a breaking strength between 5 and 5.5 kg per sq. cm.?...............6

**3.4** What proportion of the gunny bags have a breaking strength NOT between 3 and 7.5 kg per sq. cm.?.......6

**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?....................7

**4.2** What is the probability that a randomly selected student scores between 65 and 87?.................................7

**4.3** What should be the passing cut-off so that 75% of the students clear the exam?.........................................7

**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 favourable 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?....................................................................8

**5.2** Is the mean hardness of the polished and unpolished stones the same?.........................................................8

**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………………………………………………………......9

**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.

* 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?......................................................................................................................10

**Solutions:**

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

**Ans.: P(**Injured**) =** 145/235 = 0.617

* 1. **What is the probability that a player is a forward or a winger?**

**Ans.:** The probability that a player is a forward or a winger is **P(**Forward or Winger**)**

**= P(**Forward**) + P(**Winger**)**

**= (**94/235 + 29/235**) =** 123/235

= 0.523

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

**Ans.:** The probability that a randomly chosen player plays in a striker position and has a foot injury is **P(**Striker ∩ Foot injury**)**

**=** 45/235 **=** 0.191

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

**Ans.:** The probability that a randomly chosen injured player is a striker is **P(**Striker | Injured**)**

**= P(**Striker ∩ Foot injury**) / P(**Injured**)**

**= (**45/235**) / (**145/235**)**

**= (**45/145**)** = 0.310

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

**Ans.:** Total Injured forward player and injured attacking midfielder, **(**56+25**)** = 80, and altogether we have 145 injured players.

So, the probability that a randomly chosen injured player is either a forward or an attacking midfielder is: **(**80/145**)** = 0.551.

**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:

**Ans.:** Here, from the problem statement we can define the events as:

‘**F**’ stands for fire hazards.

‘**M**’ stands for mechanical failure.

‘**H**’ Stands for human error.

‘**R**’ Stands for radiation leak.

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

**Ans.:** According to the studies carried out by the organization,

* the probability of a radiation leak in case of a fire is 20%, we can formulate this as, P(R|F) = 0.2.
* the probability of a radiation leak in case of a mechanical 50%, we can formulate this as, P(R|M) = 0.5.
* and the probability of a radiation leak in case of a human error is 10%, we can formulate this as, P(R|H) = 0.1.

Again,

* The probability of a radiation leak occurring simultaneously with a fire is 0.1%, we can formulate this as, P(R∩F) = 0.001.
* The probability of a radiation leak occurring simultaneously with a mechanical failure is 0.15%, we can formulate this as, P(R∩M) = 0.0015.
* The probability of a radiation leak occurring simultaneously with a human error is 0.12%, we can formulate this as, P(R∩H) = 0.0012.

So, the **probabilities of a fire** P(F): P(R∩F)/ P(R|F) = 0.001/0.2 = 0.005.

the **probabilities of a mechanical failure** P(M): P(R∩M)/ P(R|M) = 0.0015/0.5 = 0.003

the **probabilities of a human error** P(H): P(R∩H)/ P(R|H) = 0.0012/0.1 = 0.012

**2.2 What is the probability of a radiation leak?**

**Ans.:** The types of accidents possible at the plant are, fire hazards, mechanical failure, or human error, so using Addition Rule we can calculate **probability of a radiation leak** which is as follows:

* **P(R) = P(R∩F) + P(R∩M) + P(R∩H)** = 0.001 + 0.0015 + 0.0012 = 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.

**Ans.:** Using **Bayes' theorem** we can calculate,

* The probability of **a radiation leak in the reactor due to a Fire P(F|R)**: P(R∩F)/P(R) = 0.001/0.0037 = 0.2702
* The probability of **a radiation leak in the reactor due to a Mechanical Failure P(M|R)**: P(R∩M)/P(R) = 0.0015/0.0037 = 0.4054
* The probability of **a radiation leak in the reactor due to a** **Human Error** **P(H|R)**: P(R∩H)/P(R) = 0.0012/0.0037 = 0.3243

**Problem 3:**

The breaking strength of gunny bags used for packaging cement is normally distributed with a mean of 5 kg per sq. centimetre and a standard deviation of 1.5 kg per sq. centimetre. 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?**

**Ans.:** The proportion of the gunny bags have a breaking strength less than 3.17 kg. /sq. cm. is: 0.11123243744783456

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

**Ans.:** The proportion of the gunny bags have a breaking strength at least 3.6 kg./sq. cm. is : 0.8246760551477705

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

**Ans.:** The proportion of the gunny bags have a breaking strength between 5 and 5.5 kg./sq. cm. is : 0.13055865981823633

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

**Ans.:** The proportion of the gunny bags have a breaking strength NOT between 3 and 7.5 kg./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?**

**Ans.:** The probability that a randomly chosen student gets a grade below 85 on this exam is : 0.8266927837484748

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

**Ans.:** The probability that a randomly selected student scores between 65 and 87 is : 0.8012869336779058

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

**Ans.:** The passing cut-off should be 82.7331628766667 so that 75% of the students could clear the exam.

**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 favourable 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?**

##### **Ans.:** *From earlier experience Zingaro assumes that the unpolished stone is not suitable for printing. that means For printing in stone Brinell's hardness index should be at least 150.*[*¶*](http://localhost:8888/notebooks/Desktop/AS_PGP-DSBA_July_2022/ARNAB_CHOWDHURY_AS_16-Oct-2022.ipynb##From-earlier-experience-Zingaro-assumes-that-the-unpolished-stone-is-not-suitable-for-printing.-that-means-For-printing-in-stone-Brinell's-hardness-index-should-be-atleast-150.)

Here we need to test the hardness of unpolished stone. As per Zingano’s believe unpolished stones may not be suitable for printing, so we can formulate the hypothesis as, null hypothesis states that the mean hardness of 'Unpolished stones' is more than 150 or same, and the alternative hypothesis states that the mean hardness of 'Unpolished stones' is less than 150.

Null hypothesis(H0): μ >= 150

Alternative Hypothesis (HA): μ < 150

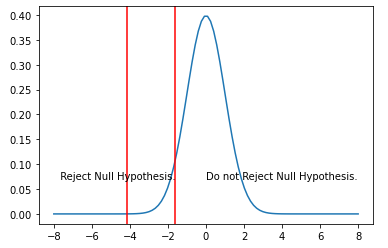
Here we have a sample size(n) of: 75

#### As we have a data set having sample size is more than '30' we can proceed further with 'Z-test', again this test only concerned about a single population we should go for 'One tailed Z-test '.

#### ##This will be a 'Left-tailed Z-test', as we need to conduct the test to the for the values which is less than Brinell's hardness index.

#### As Zingaro didn’t specify the significance value (α), so we can consider α = 0.05

The Critical value is: -1.6448536269514729



#Null hypothesis should be rejected and alternative hypothesis should be accepted. So, Zingaro is justified to think that unpolished stone is not suitable for printing.

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

**Ans.:** #Compearing the mean hardness of the polished and unpolished stones

Here we need to test the hardness of unpolished stone with polished stone. As per Zingaros believe unpolished stones may not be suitable for printing, so we can formulate the hypothesis as, null hypothesis states that the mean hardness of 'Unpolished stones' and 'polished stones' are same, and the alternative hypothesis states that the mean hardness of 'Unpolished stones' and 'polished stones' are different.

Null hypothesis(H0): μp = μu

Alternative Hypothesis (HA : μp ≠ μu

#Here we need to conduct a Two-tail test as we are compering mean hardness for 'Unpolished stones' and 'polished stones'.

t\_statistic: -3.242232050141406

P\_value: 0.001465515019462831

**## As we can see the p\_value is less than the level of significance, so we can reject the null hypothesis and hence can conclude that the mean hardness of 'Unpolished stones' and 'polished stones' are different.**

**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.

##### **Ans.:** *Let's assume the candidates of Aquarius health club can do more than 5 push-ups and we can denote that as "μ”.*

##### *A paired t-test is used when we are interested in the difference between two variables for the same subject.*

Here we need to test the assumption where null hypothesis states that the candidates of Aquarius health club can do more than 5 push-ups. The alternative hypothesis states that the candidates of Aquarius health club fail to do more than 5 push-ups.

#### Hence, we can frame the hypothesis of mentioned paired t-test as:

#### Null hypothesis (H0): μ >5

#### Alternative Hypothesis H1: μ <=5

\*t\_statistic: -19.322619811082458

\*\*p-value for one-tail: 1.1460209626255983e-35

#### From the test output, we can say the p-value is less than 0.05\*\*. So, we reject the null hypothesis at 5% level of significance.

#### #**Conclusion:** We have enough evidence to reject the null hypothesis in favor of alternative hypothesis, so we can conclude that the claim of Aquarius health club training program is unsuccessful, that the candidates of Aquarius health club fail to do more than 5 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.

* 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?