Business Report Advanced Statistics

Dhruv Dosad

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?

Answer: Probability of injured player= Player Injured / Total player

P(injured)= 145/235

= 0.617

Probability of chosen player would suffer an injured is 0.617%

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

Answer:

No of players that plays forward : 94

No of players that are winger: 29

Total Players: 235

The probability of a player is forward or a winger: (No of players who are forward + No of players who are winge) / Total Players

P (Forward or Winger) = (94 + 29) / 235 = 123/235

= 0.523

The probability of a player is forward or a winger is 0.523

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

Answer:

No of players that are at striker position & injured : 45

Total No of players that plays in a striker position : 235

Probability of Player plays in a striker position & has a foot injury : No of players at striker position & injured / Total no of players

P(Striker with foot injury) = 45 / 235

= **0.1914**

The Probability that a randomly chosen player plays in a striker position and has a foot injury is 0.1914

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

Answer:

No. of injured Striker: 45

Total Injured Players: 145

Probability of randomly chosen injured player is a striker : No of injured striker / Total injured players

P (Injured Striker) = 45/145

=0.310

The probability that a randomly chosen injured striker player 0.31

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

Answer:

No of injured forward players: 56

No of injured Attacking Midfielder : 24

Total no of Injured players: 145

Probability that a randomly chosen injured player is either a forward or an attacking midfielder= (No of injured forward players + No of injured Attacking Midfielder) / Total no of Injured players

P(Injured Forward or Midfielder) = (56 + 24) / 145

= 0.552

Probability that a randomly chosen injured player is either a forward or an attacking midfielder is 0.552

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?

Answer:

From the above problem statement, the Probabilities given are::

P(Radiation leak | Fire) = 0.2

P(Radiation leak | Mechanical Failure) = 0.5

P(Radiation leak | Human error) = 0.1

P(Radiation leak ∩ Fire) = 0.001

P(Radiation leak ∩ Mechanical Failure) =0.0015

P(Radiation leak ∩ Human error) = 0.0012

The probability of Fire

P(Fire) = P(Radiation leak ∩ Fire) /P(Radiation leak | Fire)

= 0.001/ 0.2

P(Fire)= 0.005

The probability of Mechanical failure

P(Mechanical Failure) = P(Radiation leak ∩ Mechanical Failure) / P(Radiation leak | Mechanical Failure)

=0.0015/0.5

P(Mechanical Failure) = 0.003

The probability of Human Error

P(Human error) = P(Radiation leak ∩ Human)/ P(Radiation leak | Human)

P(Human error) =0.012

2.2 What is the probability of a radiation leak?

Answer:

P(Radiation leak) = P(Radiation leak ∩ Fire) + P(Radiation leak ∩ Mechanical Failure) + P(Radiation leak ∩ Human error)

= 0.001+0.0015+0.0012

P(Radiation leak) = 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.

Answer:

Probability of Radiation leak caused by a Fire.

P(Fire | Radiation) = P(Radiation leak ∩ Fire)/ P(Radiation leak)

= 0.001 / 0.0037

P(Fire | Radiation)= 0.2702

Probability of Radiation leak caused by a Mechanical Failure

P(Mechanical Failure | Radiation) = P(Radiation leak ∩ Mechanical Failure) / P(Radiation leak)

= 0.0015/0.0037

P(Mechanical Failure | Radiation)= 0.4054

Probability of Radiation leak caused by a Human Error.

P(Human Error | Radiation) = P(Radiation leak ∩ Human)/ P(Radiation leak)

= 0.0012/0.0037

P(Human Error || Radiation)= 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. 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?

Answer:

Here as per the problem statement we have

$X = 3.17$, $\mu = 5$, $\sigma = 1.5$

Let us calculate the Z score & p value corresponds to it:

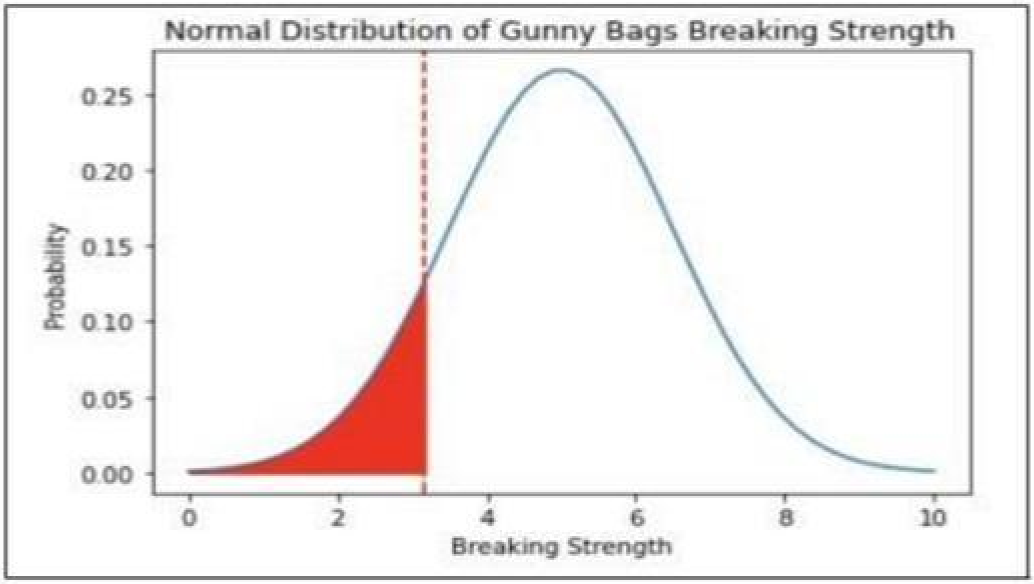
Z score = x - μ/ σ

= (3.17 - 5)/ 1.5 = -1.22

P-value from Z-Table:

P(x<3.17) = 0.11123

Hence as per the p value, 11.12 % of the gunny bags have a breaking strength less than 3.17 kg per sq cm. Visual representation below:



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

Answer:

Here as per the problem statement we have

X = 3.6 , μ = 5, σ = 1.5

Let us calculate the Z score & p value corresponds to it:

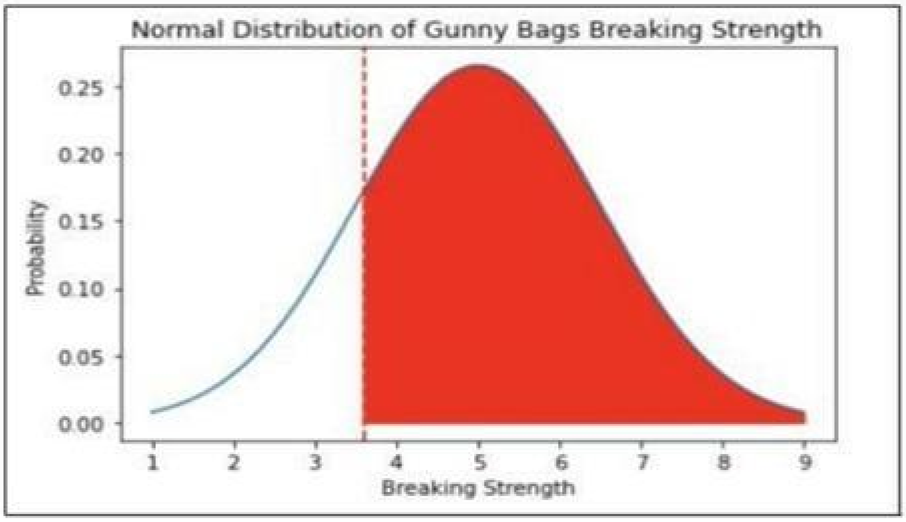
Z score = x - μ/ σ

= (3.6 - 5)/ 1.5

-0.9333
P-value from Z-Table:

P(x>3.6) = 1 - P(x<3.6) = 0.8247

Hence as per the p value, 82.47% of the gunny bags have a breaking strength at least 3.6 kg per sq cm. Visual representation below:



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

Answer:

Here as per the problem statement we have

X = 5.5 , μ = 5, σ = 1.5

Let us calculate the Z score & p value corresponds to it:

Z score = x - μ/ σ

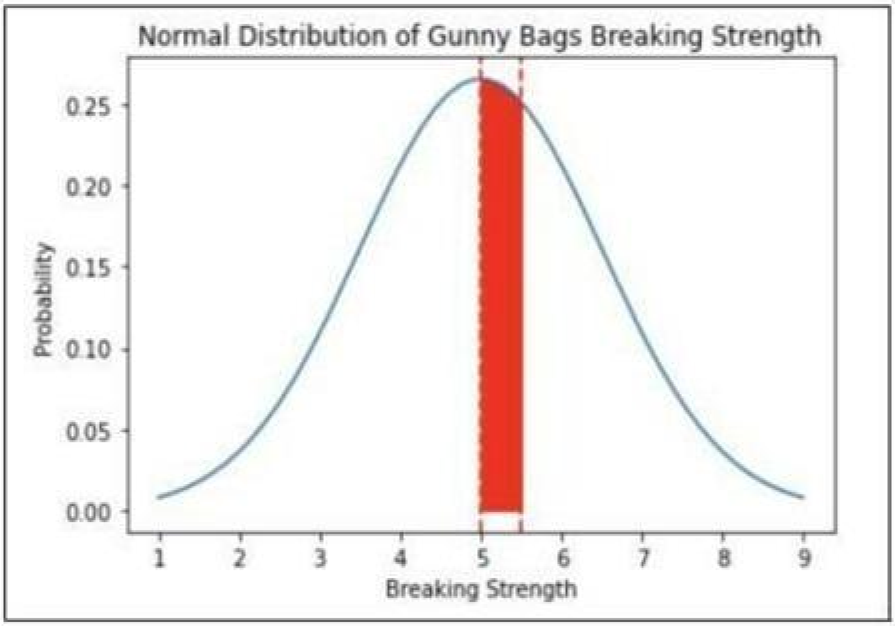
= (5.5 - 5)/ 1.5

= 0.33333

P-value from Z-Table:

P(5<x<5.5) = P(x<5.5) - 0.5 = 0.13056

Hence as per the p value, 13.06% of the gunny bags have a breaking strength between 5 and 5.5 kg per sq cm. Visual representation below:



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

Answer:

Here as per the problem statement we have

X = 3 & 7.5 , $\mu = 5$, $\sigma = 1.5$

Let us calculate the Z score & p value corresponds to it:

Z score when x is 3

Z score = $x - \mu/\sigma$

$=3 - 5 / 1.5$

$= -1.33333$

P-value from Z-Table:

$P(x>3) = 1 - P(x<3) = 0.90879$

Z score when x is 7.5

Z score = $x - \mu/\sigma$

=7.5 - 5 / 1.5

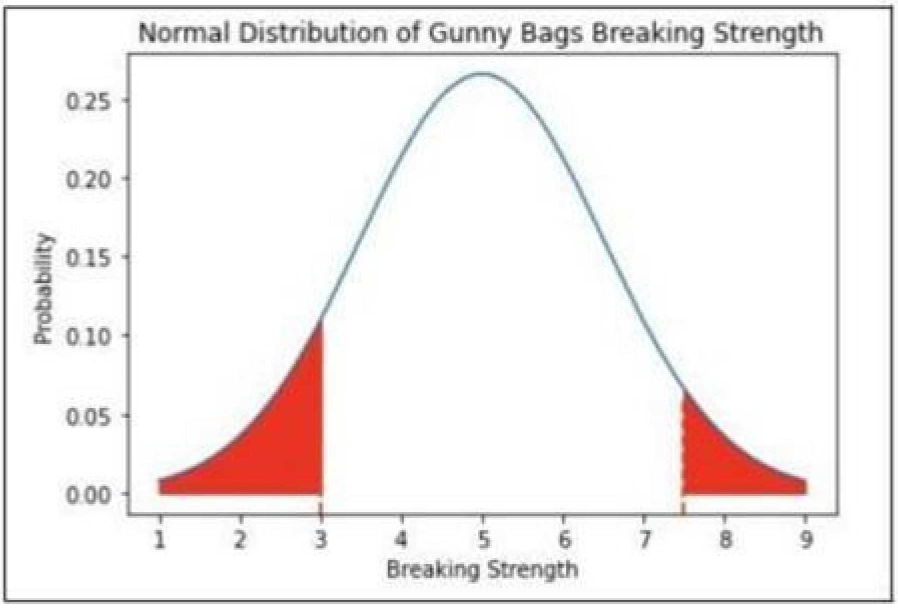
= 1.66667

P-value from Z-Table:

P(x>7.5) = 1 - P(x<7.5) = 0.04779

P value : 0.04779+ 0.091211 = 0.139001

Hence as per the p value, 13.90% of the gunny bags have a breaking strength NOT between 3 and 7.5 kg per sq cm. Visual representation below:



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?

Answer:

Here as per the problem statement we have

$X = 85$, $\mu = 77$, $\sigma = 8.5$

Let us calculate the Z score & p value corresponds to it:

Z score = x - μ /σ

= 85 - 77/8.5

= 0.94118

P-value from Z-Table:

$P(x<85) = 0.82669$

The probability of 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 scores between 65 and 87?

Answer:

Here as per the problem statement we have

$X = 65$ & 87 , $\mu = 77$, $\sigma = 8.5$

Let us calculate the Z score & p value corresponds to it:

For X=65

Z score = x - μ /σ

= 65 - 77/8.5

=-1.41176

For X=87

Z score = x - μ /σ

= 87 - 77/8.5

= 1.17647

Probability between 2 Z scores

P(-1.41176<x<1.17647) = 0.80129

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?

Answer:

Here as per the problem statement we have

X = unknown, μ = 77, σ = 8.5

We will first find the Z score corresponds to 75 percentile which would be 0.67412

(Formula used: Q3 = μ + (. 675)σ = 77+ (. 675)*8.5 =82.7375)

Z score corresponds to the 75th Percentile will be = 0.67412

Let’s us find the X in this case:

Z = x - μ /σ

0.67412 **=** x- 77/8.5

0.67412*8.5 = x-77

5.73002 = x-77

x=82.73002

Hence, the passing cutoff should be 82.73% so that 75% of the students 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 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?

Answer

Assuming the Null & alternate hypothesis basis the problem statement

H0= Zingaro believes now that the unpolished stones may suitable for printing

H1= Zingaro believes now that the unpolished stones may suitable for printing

```
# We will perform one sample t-test
# Hypothesised Mean = 150
t_statistic, p_value = ttest_1samp(Zingaro['Unpolished '], 150)
print('One sample t test \nt statistic: {0} p value: {1} '.format(t_statistic, p_value))
```

```
One sample t test
t statistic: -4.164629601426758 p value: 8.342573994839285e-05
```

After calculating p value for the above, i.e. 4.1712869974196425e-05 which is less than Level of Significance (Alpha - 0.05), hence we reject the Null Hypothesis.

This justifies that Zingaro has reason to believe now that the unpolished stones may not be suitable for printing.

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

Answer:

Assuming the Null & alternate hypothesis basis the problem statement

H0= Mean hardness of the polished and unpolished stones is same

H1= Mean hardness of the polished and unpolished stones is not same

```
t_statistic, p_value  = ttest_ind(data1['Unpolished '],data1['Treated and Polished'])
print('tstat',t_statistic)
print('P Value',p_value)
```

```
tstat -3.242232050141406
P Value 0.001465515019462831
```

After calculating p value for the above, i.e. 0.001465515019462831 which is less than Level of Significance (Alpha - 0.05), hence we reject the Null Hypothesis.

This justifies that the mean hardness of the polished and unpolished stones is the same

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.

Answer:

In order to find out whether the training will make a difference of more than 5, let us assume below hypothesis

Null hypothesis= The training will make a difference of more than 5 push-up

Alternative hypothesis = The training will not make a difference of more than 5 push-ups,

H0: mu(after) - mu(before) = >=5

Ha: mu(after) - mu(before) = <5

```
# paired t-test: when comparing the before and after changes
t_statistic, p_value  =  ttest_rel(ahc['Before'], ahc['After'] - 5)
print('tstat  %1.3f' % t_statistic)
print("p-value for one-tail:", (1-p_value/2))
```

```
tstat  -1.915
p-value for one-tail: 0.9708011278589889
```

By applying the statistical Paired two-sample t-test, the p-value comes as = 0.9708011278589889 i.e. p_value > Alpha. **Hence, We do not have enough evidence to reject the null hypothesis in favor of alternative hypothesis . This conclude that the Aquarius health club Training program is able to make a difference or 5 or more pushups.**

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 favor one method above another and may work better in his/her favorite 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.?

Answer:

The Hypothesis for the One Way ANOVA are:

H0: There is difference among Dentists on implant hardness

Ha: There is no difference among Dentists on implant hardness

After performing one way Anova on each Alloy, we found p value:

For Alloy 1 : 0.116567 , For Alloy 2: 0.718031

Hence at p - value > 0.05, we fail to Reject H0, i.e At 95 % confidence we have statistical evidence to state that there is no difference amount the Dentists on implant hardness

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

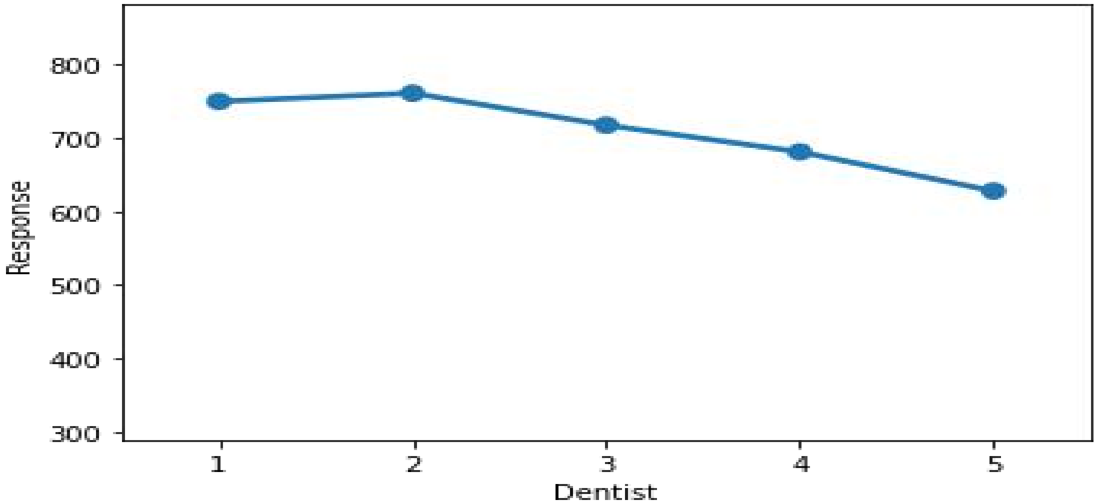
Answer:

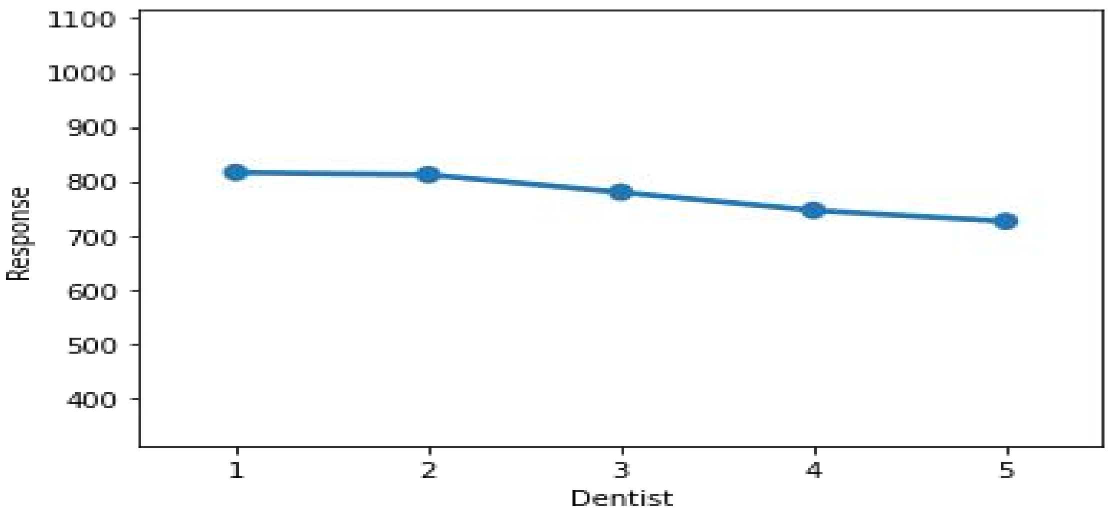
Before the hypothesis may be tested, the assumptions were that there is no difference among dentists on implant hardness. The assumptions were fulfilled in both the Alloy types as the p values were higher than 0.05.

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?

Answer:

We plotted the Dentist & Responses for each Alloy and found no significant difference. The implant hardness is independent of the Dentists,





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?

Answer:

The Hypothesis for the One Way ANOVA are:

H0: There is difference with Method on implant hardness

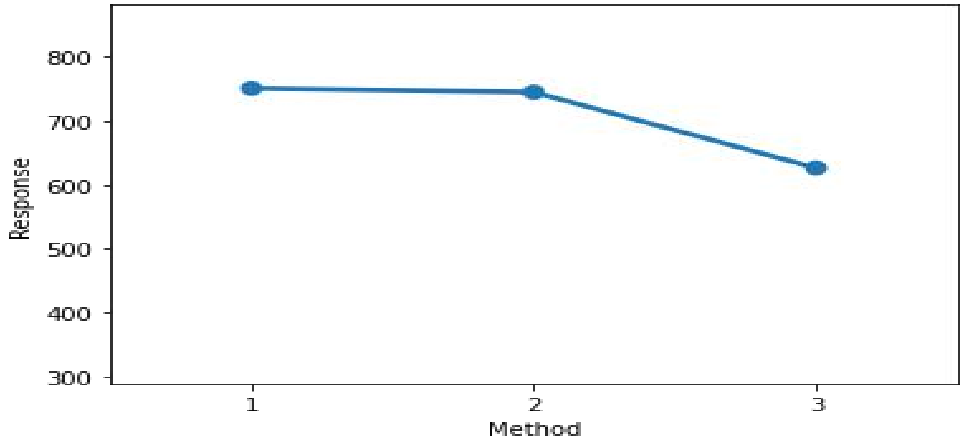
Ha: There is no difference with Method on implant hardness

After performing one way Anova on each Alloy, we found p value:

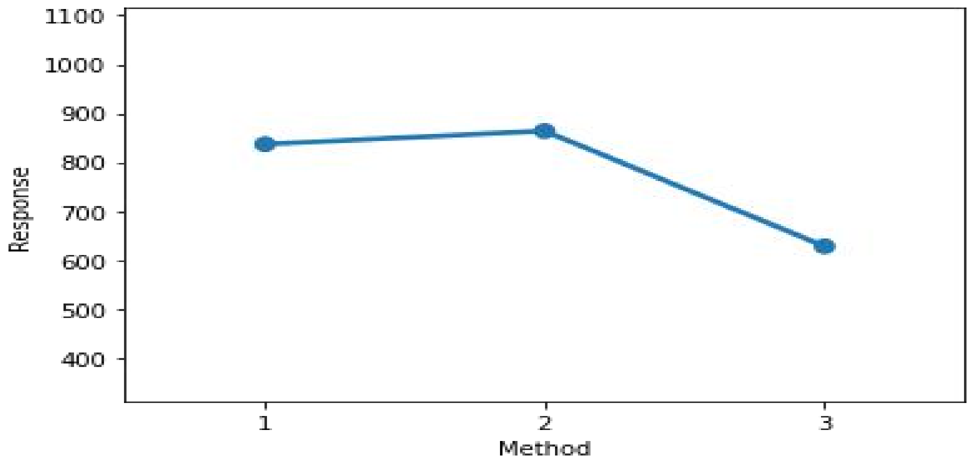
For Alloy 1 : 0.004163 , For Alloy 2: 0.000005

Hence at p - value < 0.05, we accept H0, i.e At 95 % confidence we have statistical evidence to state that there is difference among the method used on implant hardness

Alloy1



Alloy2



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?

Answer:

The Hypothesis for the One Way ANOVA are:

H0: There is difference with Temperature level on implant hardness

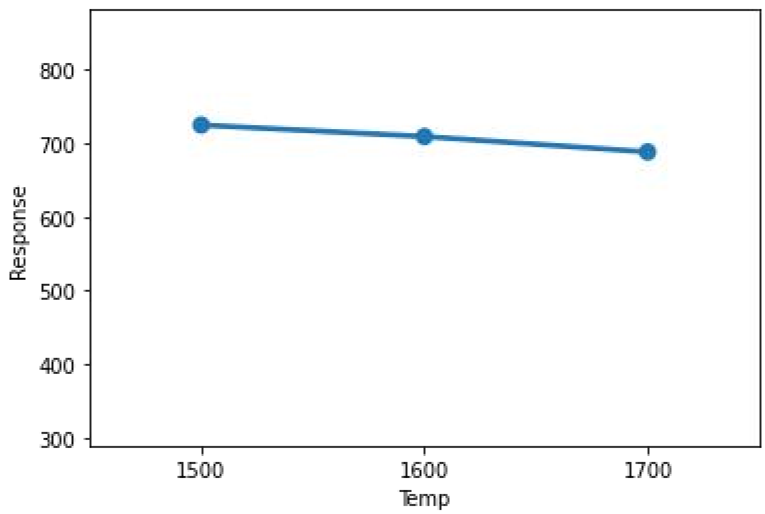
Ha: There is no difference with Temperature level on implant hardness

After performing one way Anova on each Alloy, we found p value:

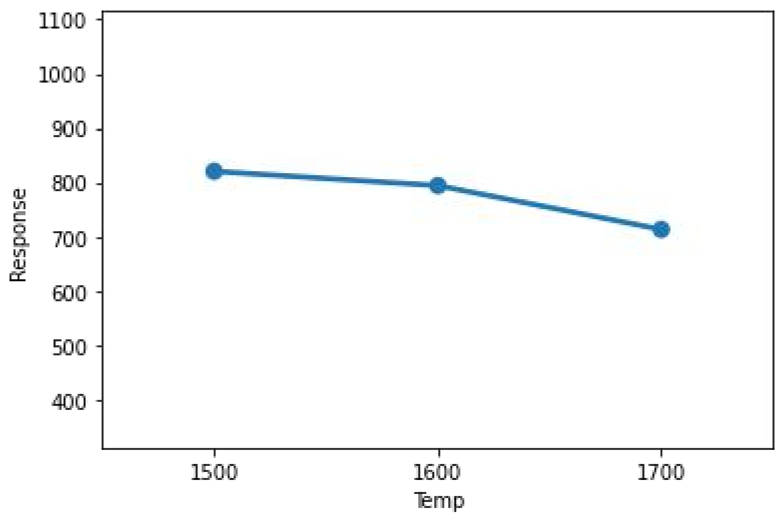
For Alloy 1 : 0.717074 , For Alloy 2: 0.164678

Hence at p - value > 0.05, we fail to reject H0, i.e At 95 % confidence we have statistical evidence to state that there is no difference among the Temperature level on implant hardness

Alloy1



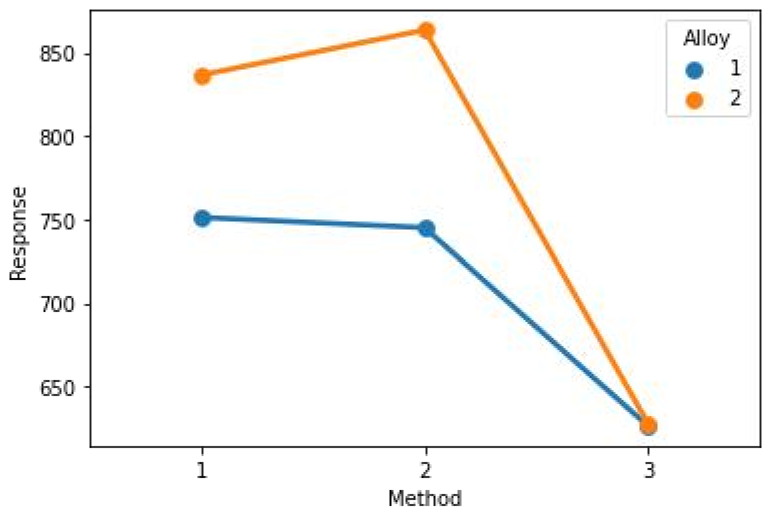
Alloy 2



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

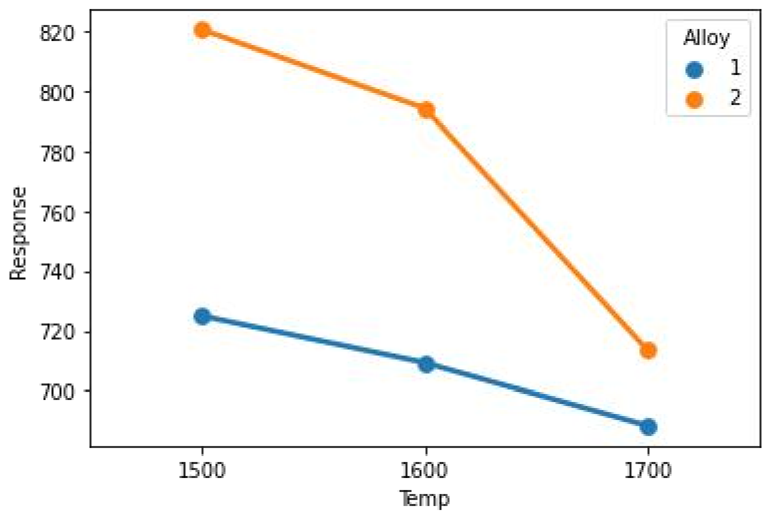
Answer:

Interaction graph between Alloy & Method

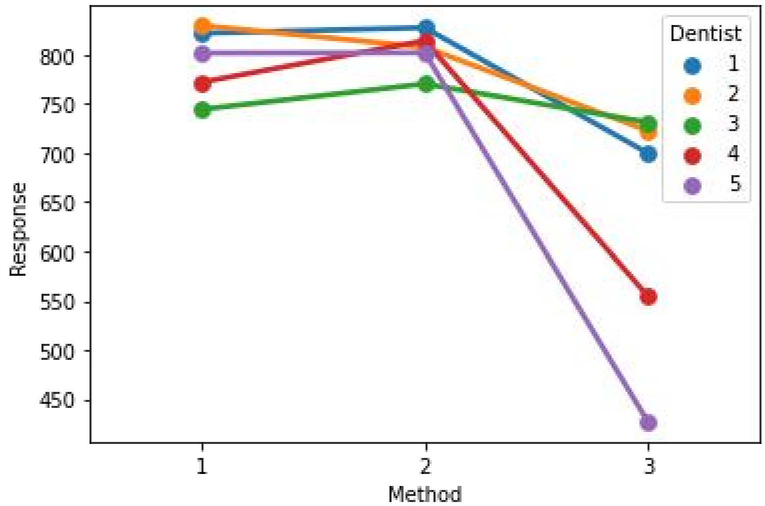


There is no interaction between Method & Alloy

Interaction graph between Temperature & Alloy



Interaction graph between Method & Dentists

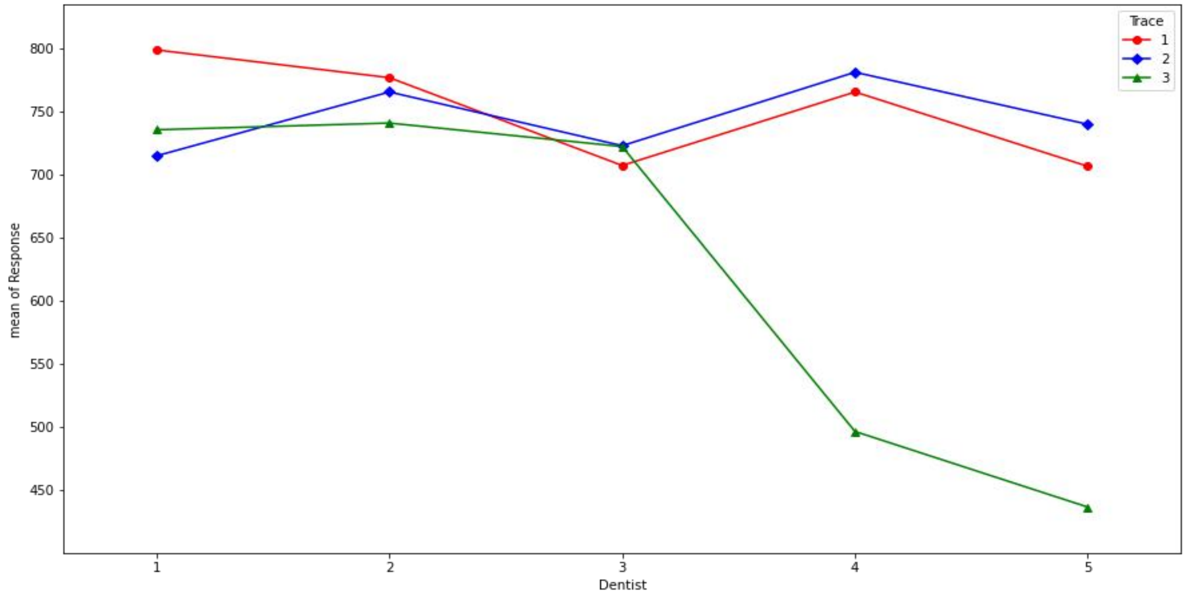


We can see that there is some sort of interaction between the Methods used.

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?

Answer:

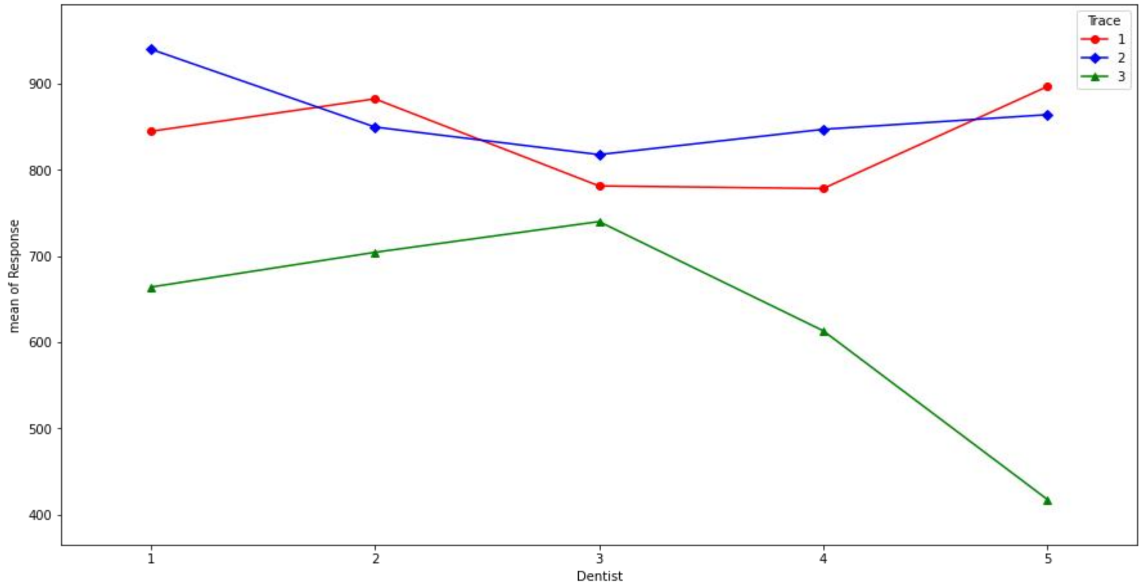
Interaction plot for Alloy1



With above interaction plot, we can conclude that:

- All methods have interaction among Dentists.

Interaction plot for Alloy2



With above interaction plot, we can conclude that :

- Method 1 & Method 2 has interaction among Dentists. While Method 3 has no interactions with others.