Inferential Statistics (IS)

Coded Project Report

Submitted to

greatlearning
Learning for Life

by

Subhadeep Seal

in Partial Fulfillment of

PGP-DSBA

TEXAS McCombs
The University of Texas at Austin
McCombs School of Business

Table of 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?

Solution: Probability that a randomly chosen player would suffer an injury

= Total Number of Players Injured/ Total Number of Players

= 145/235

= **0.62.**

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

Solution: 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.**

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

Solution: 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.**

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

Solution: Probability that a randomly chosen injured player is a striker

= Number of Injured Strikers/ Total Number of Injured Players

= 45/145

= **0.31.**

Problem 2:

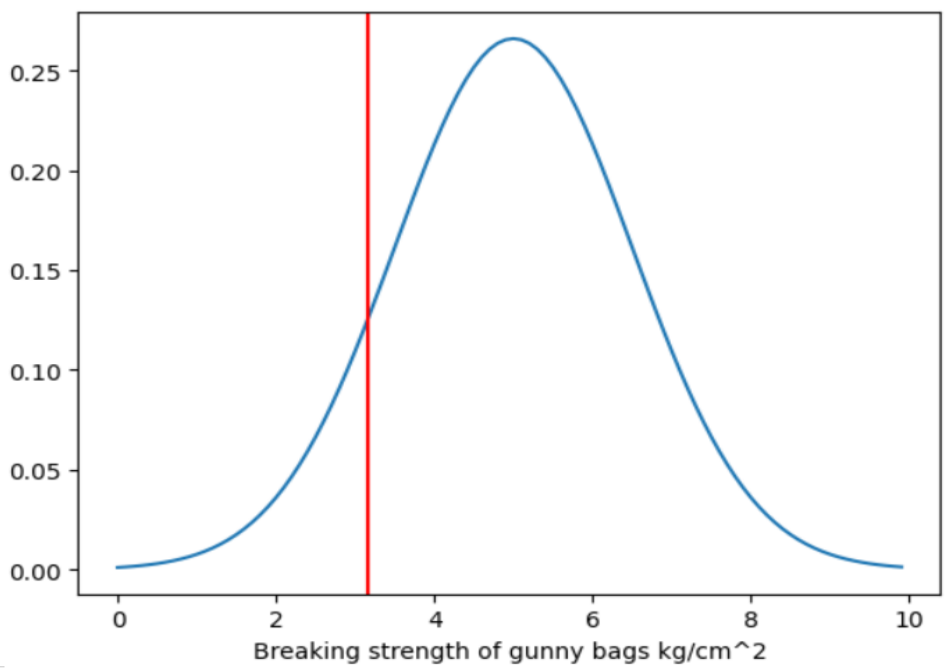
The breaking strength of gunny bags used for packaging cement is normally distributed with a mean of 5 kg per sq. centimeter and a standard deviation of 1.5 kg per sq. centimeter. The quality team of the cement company wants to know the following about the packaging material to better understand wastage 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)

Solution: Probability Distribution Function of Breaking Strength of Gunny Bags is given as, X~N (μ = 5, σ = 1.5)

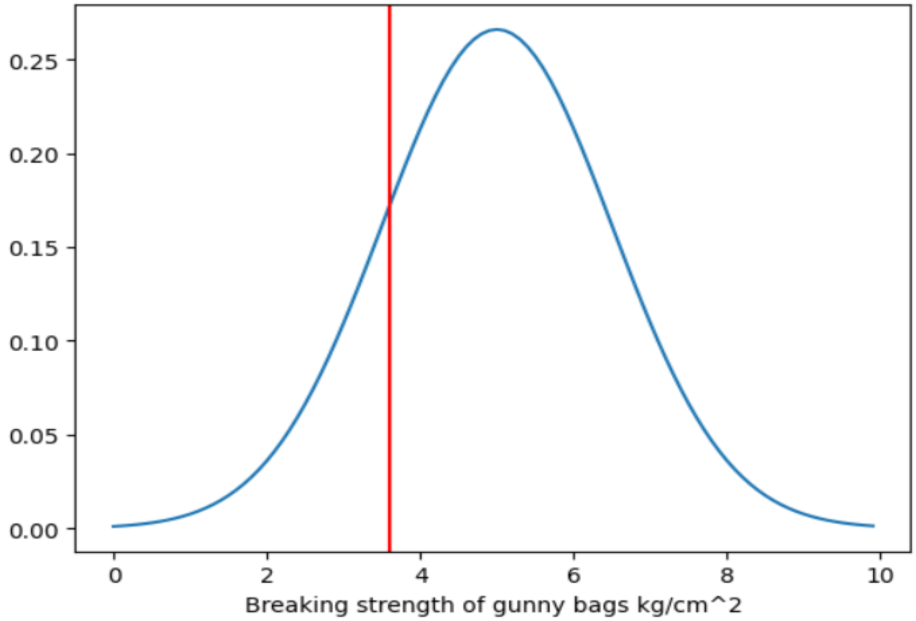
2.1 What proportion of the gunny bags have a breaking strength less than 3.17 kg per sq cm?

P (X < 3.17) = **0.11.**



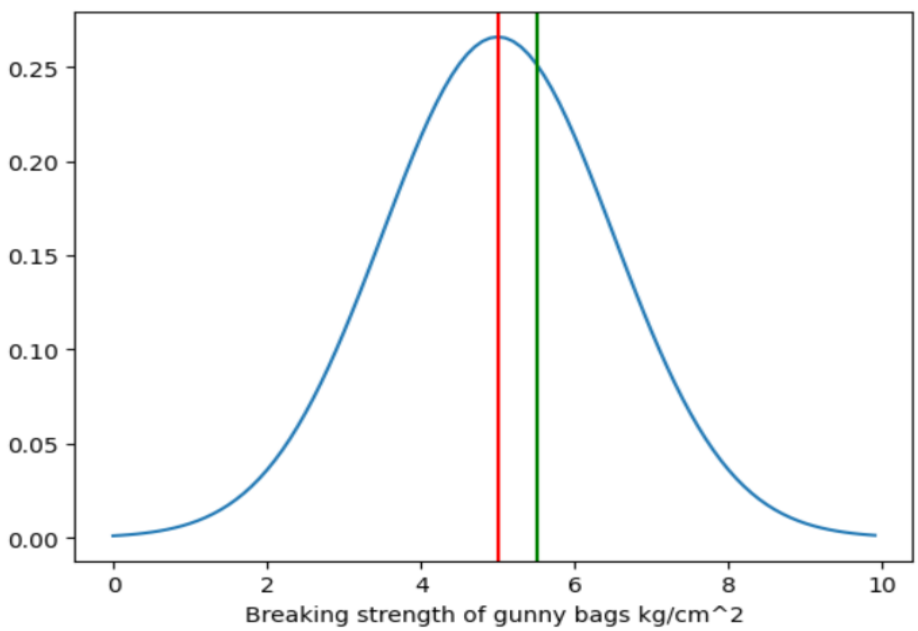
2.2 What proportion of the gunny bags have a breaking strength at least 3.6 kg per sq cm.?

P (X >= 3.6) = 1 - P (X<3.6) = **0.82.**



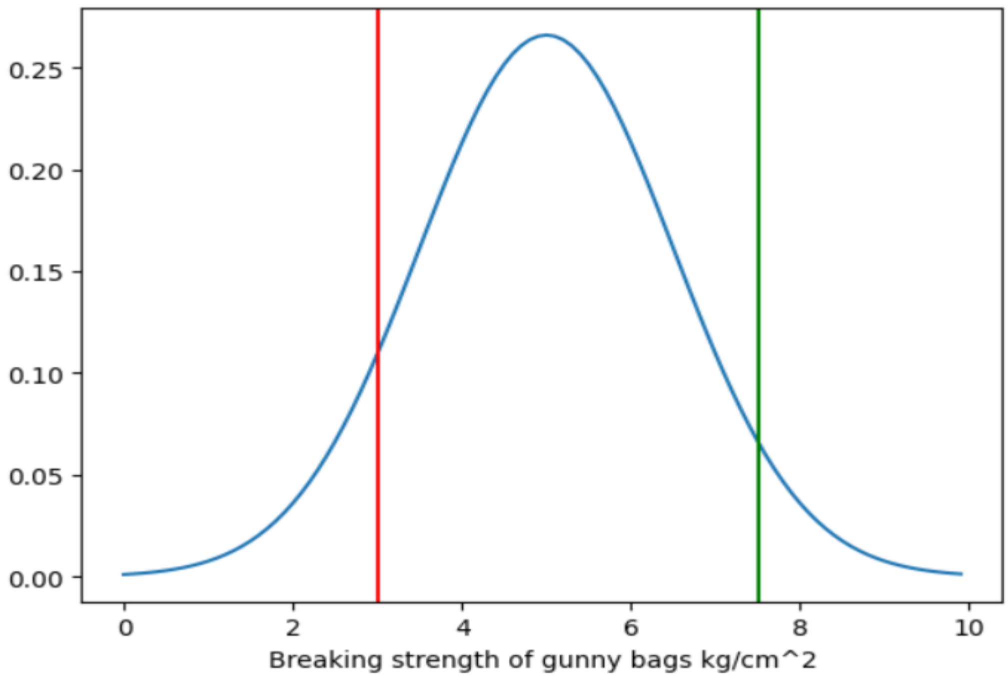
2.3 What proportion of the gunny bags have a breaking strength between 5 and 5.5 kgper sq cm.?

P (5< X <5.5) = P (X<5.5) – P (X<5) = **0.13.**



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

P (X < 3 and X > 7.5) = P (X<3) + 1 - P (X<7.5) = **0.14.**



Problem 3:

Dataset -Zingaro_Company.csv

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);

3.1 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?

Solution: First, hypothesizing for unpolished stones, we have:

Step1: Ho (Null Hypothesis) -> Sample Mean >=150

H1 (Alternate Hypothesis) -> Sample Mean <150

Step2: Given, $\alpha = 0.05$

Step 3: 1 sample z-test as this is a one-sided sample with **sample size is 75>30**

Step 4: $\mu = 150$ and n=75.

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

Type of Stone	Unpolished	Treated and Polished
mean	134.11053	147.7881

Std. dev	33.041804	15.58736
	0.9999844	0.890447
p-value	1.559E-05	0.109553
Two-tailed t-test	0.0007328	

Step 5: 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 be rejected in that case. **So, treated, and polished stones have a Brinell's hardness index of at least 150.**

CONCLUSION 1: *Hence, Zingaro is right in thinking that unpolished stones not suitable for printing unlike the treated and polished stones which have the right fitment.*

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

Solution: Hypotheses - Null: mean $_{polished}$ = mean $_{unpolished}$ | Alternate: mean $_{polished}$ ≠ 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.

CONCLUSION 2: *Hence, the hardness of polishes and unpolished stones are significantly different.*

Problem 4:

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.

4.1 How does the hardness of implants vary depending on dentists?

Solution:

Hypotheses:

- **Null Hypothesis (H0):** There is no significant difference in the mean hardness of dental implants among different dentists.
- **Alternate Hypothesis (H1):** There is a significant difference in the mean hardness of dental implants among different dentists.

Assumptions of the Hypothesis Test

- **Independence:** The hardness measurements for each dentist are independent of each other.

- **Normality:** The hardness measurements for each dentist follow a normal distribution.
- **Homogeneity of Variance:** The variance of hardness measurements is equal across different dentists.

Hardness by Dentists:

We will use ANOVA test to determine if there's a statistically significant difference in hardness based on the dentist.

```
ANOVA Table - Dentist:
                 sum_sq    df         F    PR(>F)
Dentist   1.465472e+05   1.0  7.392259  0.00789
Residual  1.744548e+06  88.0       NaN      NaN
```

Conclusions from the Test Results

If the p-value is less than the chosen significance level (e.g., 0.05), we reject the null hypothesis and conclude that there is a significant difference in the mean hardness of dental implants among different dentists. If the p-value is greater than the significance level, we fail to reject the null hypothesis.

Identifying Pairs with Different Implant Hardness

If the null hypothesis is rejected, post-hoc tests (e.g., Tukey's HSD test) can be conducted to identify which specific pairs of dentists show significant differences in implant hardness.

By following these steps, we can analyse the variation in implant hardness based on dentists and determine if there are significant differences among them.

4.2 How does the hardness of implants vary depending on methods?

Solution:

Hypotheses:

- **Null Hypothesis (H0):** There is no significant difference in the hardness of implants depending on the methods.
- **Alternate Hypothesis (H1):** There is a significant difference in the hardness of implants depending on the method.

Assumptions of the Hypothesis Test:

- **Independence:** The hardness measurements for different methods are independent of each other.
- **Normality:** The hardness measurements for each method follow a normal distribution.
- **Equal Variances:** The variances of hardness measurements for each method are equal.

Hardness by Methods:

Similar to dentists, we will use ANOVA test this analysis checks for differences in hardness based on the method used.

```
ANOVA Table - Method:
                sum_sq    df          F    PR(>F)
Method    4.173336e+05   1.0  24.919463  0.000003
Residual  1.473762e+06  88.0        NaN       NaN
```

Conclusions from the Test Results:

- If the p-value is less than the significance level (e.g., 0.05), reject the null hypothesis.
- If the p-value is greater than the significance level, fail to reject the null hypothesis.

Identifying Differences in Implant Hardness:

- If the null hypothesis is rejected, identify which pairs of methods show significant differences in implant hardness.
- We have to use post-hoc tests (e.g., Tukey's HSD test for equal variances, Games-Howell test for unequal variances) to compare specific pairs of methods.

4.3 What is the interaction effect between the dentist and method on the hardness of dental implants for each type of alloy?

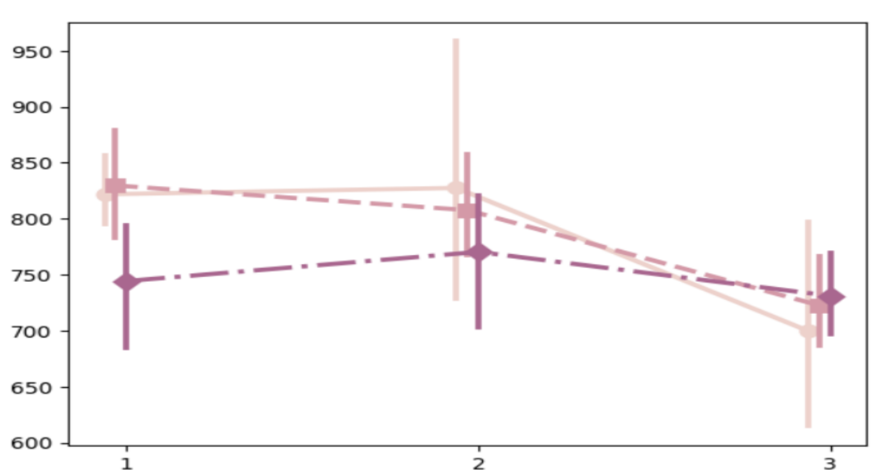
Solution:

Interaction Effect:

This analysis checks if there is an interaction effect between dentist, method, and alloy on hardness. We have used two-way ANOVA Test.

```
ANOVA Table - Interaction (Dentist * Method * Alloy):
                             sum_sq    df          F        PR(>F)
Dentist                1.465472e+05   1.0  11.145162  1.268902e-03
Method                 4.173336e+05   1.0  31.738925  2.412583e-07
Dentist:Method         1.136521e+05   1.0   8.643432  4.263514e-03
Alloy                  1.058155e+05   1.0   8.047448  5.739468e-03
Dentist:Alloy          2.769089e+03   1.0   0.210594  6.475164e-01
Method:Alloy           2.646000e+04   1.0   2.012328  1.598154e-01
Dentist:Method:Alloy   3.040083e+02   1.0   0.023120  8.795180e-01
Residual               1.078214e+06  82.0        NaN           NaN
```

INTERACTION PLOT BETWEEN DENTIST AND METHOD



4.4 How does the hardness of implants vary depending on dentists and methods together?

Solution:

Null hypothesis(Ho): There is no interaction effect between dentists and methods on hardness
Alternate hypothesis (Ha): There is an interaction effect between dentists and methods on hardness

Hardness by Dentists and Methods Together:
This ANOVA checks if there is a combined effect of dentists and methods on hardness.

ANOVA Table - Dentist and Method:

	sum_sq	df	F	PR(>F)
Dentist	1.465472e+05	1.0	10.385174	1.795039e-03
Method	4.173336e+05	1.0	29.574648	4.970469e-07
Dentist:Method	1.136521e+05	1.0	8.054037	5.662740e-03
Residual	1.213563e+06	86.0	NaN	NaN

Conclusions:

- **Individual Effects:** Assess the individual impacts of dentists and methods on hardness using ANOVA.
- **Interaction Effects:** Analyses how the combination of dentists and methods affects hardness using interaction terms in ANOVA or mixed-effects models.
- **Comprehensive Analysis:** Consider both factors together to understand their joint influence on implant hardness.