Predictive Model of Rhomboidity in Billets.

Report by: Rohan Joydhar
VT Number: VT20211672
Guided by: Swapnil N Dhakate, E.Z. Chacko
Dated on: July 26, 2021
TATA STEEL LIMITED,JAMSHEDPUR

Acknowledgement

At this momentous occasion of completion of my project binding, I would like to acknowledge the contribution of all those benevolent people I have been blessed to associate with. Behind every project, there stand a myriad of people whose help and contribution makes things successful.

I am deeply indebted to my guide Swapnil N Dhakate,for giving me an opportunity to work with him on this project. His advice and guidance throughout this work has been of immeasurable help not only in research work, but also in my development as a whole. I would like to specially thank Mr. E.Z. Chacko, Sr. Manager, TGGWL for his continual input and help in analysis.

Last but not the least, I would like to express my deepest gratitude to the almighty without his blessings, this project would not have been even half as much interesting or fun as it has been.

Executive Summary

Rhomboidity in billets was investigated in this industrial study on CC2 billet caster at LD#1. This study focuses on events leading to variability in severity of rhomboidity. The effect of casting speed, superheat and mould design on severity of rhomboidity was evaluated. Based on the knowledge of the rhomboidity generated in this study, recommendations are proposed to minimize the problem.

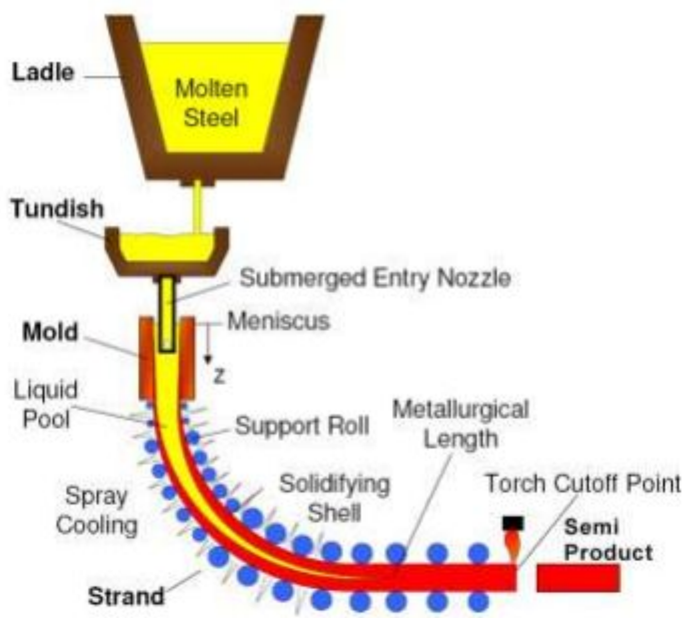
The study was only focussed on TMT grade of steel.The variables that are considered in this study are the phosphorus content, manganese content, sulphur content, Mn/S ratio of steel, supreheat of steel, mould level, mould cooling water flow rate, mould cooling water delta temp, casting speed, secondary cooling zone flow rate, M factor and heat extraction value. By using advanced data analysis and machine learning linear regression algorithm the model is being produced.

Billet casting process.

Introduction

Continuous casting, also called strand casting, is the process whereby molten metal is solidified into a "semi finished" billet, bloom, or slab for subsequent rolling in the finishing mills. Prior to the introduction of continuous casting in the 1950s, steel was poured into stationary moulds to form ingots. Since then, "continuous casting" has evolved to achieve improved yield, quality, productivity and cost efficiency. It allows lower-cost production of metal sections with better quality, due to the inherently lower costs of continuous, standardized production of a product, as well as providing increased control over the process through automation.

Molten metal (known as hot metal) is tapped into the ladle from furnaces. After undergoing any ladle treatments, such as alloying and degassing, and arriving at the correct temperature, the ladle is transported to the top of the casting machine. Usually, the ladle sits in a slot on a rotating turret at the casting machine; one ladle is 'on cast' (feeding the casting machine) while the other is made ready, and is switched to the casting position once the first ladle is empty.



From the ladle, the hot metal is transferred via a refractory shroud to a holding bath called a tundish. The tundish allows a reservoir of metal to feed the casting machine while ladles are switched, thus acting as a buffer of hot metal, as well as smoothing out flow, regulating metal feed to the moulds and cleaning the metal.

Metal is drained from the tundish through another shroud into the top of an open-base copper mould. The depth of the mould can range from 0.5 to 2 meters, depending on the casting speed and section size. The mould is water-cooled to solidify the hot metal directly in contact with it; this is the primary cooling process.
In the mould, a thin shell of metal next to the mould walls solidifies before the middle section, now called a strand, exits the base of the mould into a spray-chamber; the bulk of metal within the walls of the strand is still molten.

The strand is immediately supported by closely spaced, water cooled rollers, these acts to support the walls of the strand against the ferrostatic pressure of the still-solidifying liquid within the strand. To increase the rate of solidification, the strand is also sprayed with large amounts of water as it passes through the spray-chamber; this is the secondary cooling process. Final solidification of the strand may take place after the strand has exited the spray-chamber.

Billet rhomboidity (or off-squareness) which is defined as the difference between the lengths of the two diagonals, has been a major quality problem since the inception of continuous casting. The absolute value of the difference is a measure of severity of rhomboidity while sign (positive or negative) associated with the measured value is indicative of the orientation of rhomboidity. It is also common for the magnitude and orientation of rhomboidity to change with time in the heat and sometimes “twisting” of billets on cooling bed is observed. With respect to severity of problem many mills consider a difference in diagonals of greater than ~8 mm unacceptable because the rhomboid billet causes processing difficulties in the reheat furnace and subsequent hot rolling operations. Furthermore, billets with excessively large rhomboidity can crack along the diagonal or the corners.

A large rhomboidity obviously suggests lack of control and the presence of non-uniform cooling conditions in the mould and/or sprays. In the mould, heat transfer is related to thermomechanical behaviour of the mould and the nature of mould-strand interaction. The behaviour of meniscus which is affected by metal level fluctuation is another critical variable that can influence heat transfer and progress of solidification, leading to non-uniform shell thickness and oscillation marks of variable depth. Thus rhomboidity is affected by a number of mould designs, operating variables and process upsets.

The information presented in literature provides useful insights on generation of rhomboidity. For example, the effect of steel carbon content on the severity of rhomboidity is quite important; it was reported that steel grades with carbon content in the ~0.17 to 0.45 percent range are more sensitive to the problem than others. Another important parameter is the mould taper at the meniscus level. The parabolically and double tapered moulds generally have steeper taper (in excess of ~2.0%/m) in the meniscus region and low heat transfer while single tapered moulds have shallow tapers (~0.8%/m) and high heat transfer. Even flow of lubricating oil plays a role in rhomboidity generation. In earlier studies on rhomboidity in billets, asynchronous boiling in the cooling water channel was identified as an important contributor. The region of the mould that is most affected by boiling is the area close to meniscus. It was shown mathematically that operating with low water velocity results in intermittent, asynchronous boiling on the four cold faces.

Other shape defects are namely, high section and bulging. High section is the increase in the billet cross sectional dimension occurred due to mould wear at the mould exit. Bulging mainly occurs due to low shell thickness at the mould exit, occurrence of re-entrant corners and excess pressure at withdrawal unit.

Rhomboidity in billet casting.

Rhomboidity in Billets

Billet rhomboidity (or off-squareness) which is defined as the difference between the lengths of the two diagonals, has been a major quality problem since the inception of continuous casting. The absolute value of the difference is a measure of severity of rhomboidity while sign (positive or negative) associated with the measured value is indicative of the orientation of rhomboidity. It is also common for the magnitude and orientation of rhomboidity to change with time in the heat and sometimes "twisting" of billets on cooling bed is observed. With respect to severity of problem many mills consider a difference in diagonals of greater than ~8 mm unacceptable because the rhomboid billet causes processing difficulties in the reheat furnace and subsequent hot rolling operations. Furthermore, billets with excessively large rhomboidity can crack along the diagonal or the corners.

A large rhomboidity obviously suggests lack of control and the presence of non-uniform cooling conditions in the mould and/or sprays. In the mould, heat transfer is related to thermomechanical behaviour of the mould and the nature of mould-strand interaction. The behaviour of meniscus which is affected by metal level fluctuation is another critical variable that can influence heat transfer and progress of solidification, leading to non-uniform shell thickness and oscillation marks of variable depth. Thus rhomboidity is affected by a number of mould designs, operating variables and process upsets.

The information presented in literature provides useful insights on generation of rhomboidity. For example, the effect of steel carbon content on the severity of rhomboidity is quite important; it was reported that steel grades with carbon content in the ~0.17 to 0.45 percent range are more sensitive to the problem than others. Another important parameter is the mould taper at the meniscus level. The parabolically and double tapered moulds generally have steeper taper (in excess of ~2.0%/m) in the meniscus region and low heat transfer while single tapered moulds have shallow tapers (~0.8%/m) and high heat transfer. Even flow of lubricating oil plays a role in rhomboidity generation.

In earlier studies on rhomboidity in billets, asynchronous boiling in the cooling water channel was identified as an important contributor. The region of the mould that is most affected by boiling is the area close to meniscus. It was shown mathematically that operating with low water velocity results in intermittent, asynchronous boiling on the four cold faces. When boiling occurs, the cold face temperature of the mould strongly influences the instantaneous rate of heat removal by cooling water such that the mould wall temperature on a given face can fluctuate out of control. When the water velocity is reduced, boiling on the four cold faces becomes more vigorous and less intermittent such that the cooling around the billet periphery becomes more uniform.

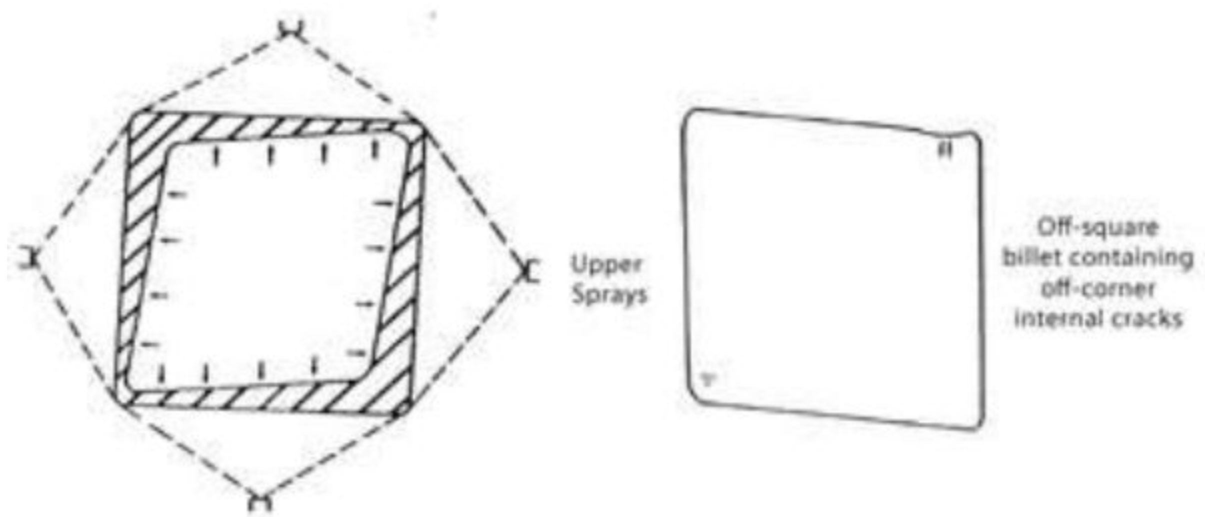
Another mechanism proposed to explain generation of rhomboidity based on oscillation mark formation and non-uniform heat transfer in mould and the sprays.

The problem begins with the formation of deep and non-uniform oscillation marks around the billet periphery. In the vicinity of a deep oscillation mark, the rate of heat removal is low due to mould-strand gap. On the other hand, regions of the billets having shallow oscillation marks experience higher rates of heat extraction. Thus, the presence of non-uniform oscillation marks on the billet surface gives rise to markedly difference in heat extraction leading to non-uniform solid shell. Thus, the billet exiting mould, although reasonably square, has non-uniform solid shell, as shown in Figure 2.12 which is a schematic representation of this concept. In the sprays, the colder portions of the strand, having thicker solid shell, tend to cool faster than the hotter regions because of the effect of unstable boiling; the result is non-uniform shrinkage of the billet and rhomboidity. During casting process following are observed:

- The obtuse-angle corners of rhomboid billet have been found to have deepest oscillation marks .

- The billets emerging from the mould, when observed through peep-hole, showed that, of the two corners in view, one was cold (dark) and the other was hot (bright). Subsequent inspection of billets on cooling bed indicated that acute-angle corners in billet correspond to the colder corners whereas the hot corners formed the obtuse angle of the billet.

Occasionally, billets on the cooling bed are found to be twisted. This indicates that magnitude and orientation of rhomboidity varies along the length of the billet and this suggests that the event generating rhomboidity changes with time. It is possible that the root cause may lay in the mould such metal level fluctuation.



Another interesting aspect of rhomboidity is the effect of mould taper and steel grade (carbon content) on its severity as shown in Figure. With respect to the severity of rhomboidity, the graph indicates that parabolically tapered moulds are most effective followed by double tapered moulds while single tapered moulds are the wors. Furthermore, steel grades with carbon content in the range ~0.17 to ~0.45% are worse than other grades.

Description of LD#1.

PLANT INDUCTION

Overview of LD#1

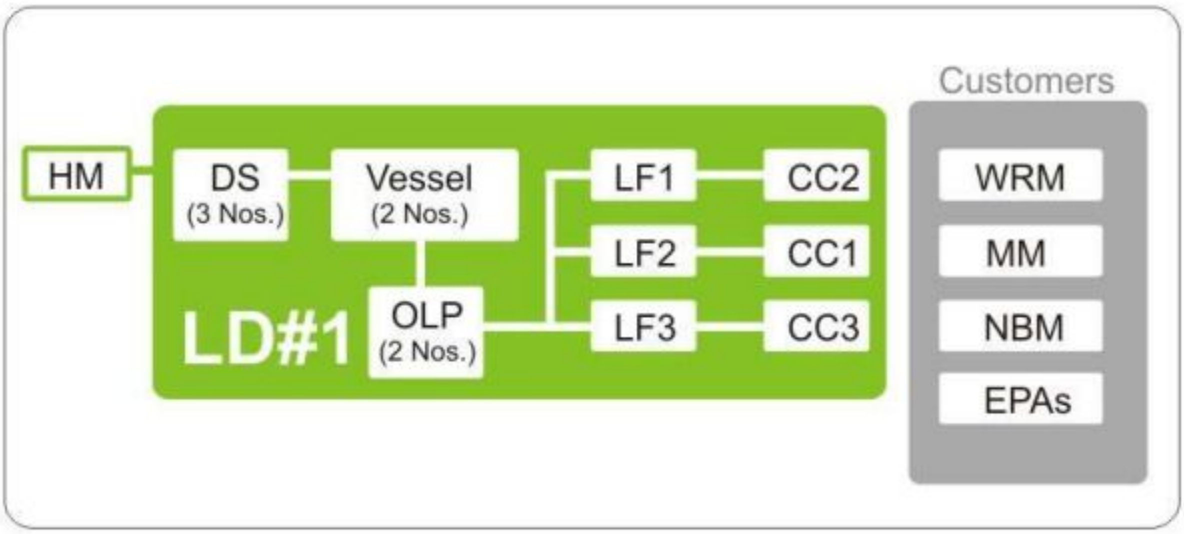
During modernization of SMS2 (steel melting shop) Tata Steel has established its first BOF (basic oxygen furnace) to meet the steel demand and for technological improvement in steel making. The Modernization Phase 1 Program of Rs. 220 crores, which basically updated the steel making facilities, marks an important milestone in the history of the Plant. The 29-month program was formally launched on Dec. 8, 1980, and it was completed in March 1983 with the commissioning of the 1.1 million tonnes per annum Basic Oxygen Furnace, i.e. L.D. Shop#1.

It has the following main features:

- Two numbers of 155 ton basic O2 converters, with flux charging system, gas cleaning and gas recovery plant. Top blowing with six lance hole and bottom stirring with TBM (thyssen niederrhein blowing metallurgy). Supplier of Main Equipments: M/s Davy McKee, U.K. Electrical Equipment Supplier: M/s Bharat Heavy Electricals Ltd. (BHEL)
- Basic refractories for Ladle
- One 130 tonne Vacuum Arc Degassing and Refining (VADR) unit supplied by M/s Standard Messo, Duisberg and Westerwerke Project Limited for making special quality engineering steels such as CHQ,CSQ. Bbp teeming facilities for VADR heats
- Three numbers of Ladle furnaces supplied by M/s SMS Demag.
- Calcium injection treatment of Liquid steel.
- Two Hot Metal Desulphurisation Units with facility of mono-injection.
- Ladle metallurgy stations for each caster equipped with top purging and Ca wire feeding
- Basic [Dry vibratable mass] & deeper tundishes.
- Ladle to Tundish shroud facility.
- One six - strand continuous casting machine (Radius: 6 m) supplied by M/s Concast AG and Concast India with SEN/EMS facility. Cast Section: 130 mm^2 . Equipped with Hard Cooling & Mould Stirrer established in 1983.
- One six strand continuous casting machine (Radius: 9 m) supplied by M/s VAI Pomini. Cast Sections: 130 & 150 mm^2 . Equipped with Hard Cooling & Mould Stirrer established in 1997.
- One another six strand continuous casting machine (Radius: 9 m) supplied by M/s VAI Pomini. Cast Sections: 130 & 150 mm^2 . Equipped with Hard Cooling, Mould Stirrer & SEN facility established in 2008.

Supporting Facilities at LD#1

- Two 250 tonnes per day Oxygen Plants for supplying 99.5% pure O2 for blowing in the converters. This plant has been supplied by M/s Cryoplants, U.K., and their Indian associates, Indian Oxygen Ltd.
- Two 300 tonnes per day Lime Calcining Plants to provide quality flux for steel making. Engineering Projects India Ltd has supplied the limekilns of Maerz design.
- One 18,000 tonne per annum Tar-Dolo Block lant to supply special refractories for lining the converters, along with a Dolo-sintering kiln. The equipment for the Dolo Block Plant has been imported from M/s Laeiswerke, West Germany.
- The basic engineering and know-how for the Calcining of superior grade limestone for the production of high reactivity lime and sintering of high purity dolomite essential for the L.D. lining has been supplied by M/s Dolomitwerke, West Germany, the consultants for these two schemes.



DS: Desulphurising Unit　LF: Ladle Furnace　OLP: Online Purging
CC: Continuous Caster　MM: Merchant Mill　WRM: Wire Rod Mill
NBM: New Bar Mill　EPAs: External Processing Agencies

Details of the Billet Casters

- LD#1 has three 6-strand billet casters.

Item		UOM	Plant		
			Jamshedpur – Longs		
			CC1	CC2	CC3
Machine	Make		Concast	Concast/VAIP	VAIP
	Start-up	year	1983	1997	3-Oct-08
	Upgradation (if any)	year	2002/2005	2005	-
	Caster type		Curved	Curved	Curved
	Metallurgical length	m	15.9	25.4	26.4
	Radius	m	6/11	9	9
	Section size	mm	130	130, 150	130, 150
	Strands	nos	6	6	6
	Interstrand distance	mm	1100	1100	1250
	Turret	Type	Butterfly - Fixed- TGS	Independent arm - Danieli	Butterfly - Fixed- TGS
	Tundish Car	Type	Semi-Cantilever	Semi-Cantilever	Full Canteilever
	Dummy Bar	Type	Rigid	Rigid	Rigid
	Cutoff	Type	Torch	Torch	Torch
	Length Measuring Device	Type	Encoder in Top roll of withdrawal	Encoder in Top roll of withdrawal	Encoder in Top roll of withdrawal
	Make of TCM	Type	Concast	Alba	Alba
	Pusher	Type	Single level	Double level	Double level
	Cooling Bed	Type	Hydraulic	Electr- Mechanical	Hydraulic
	Billet Marking	Type	Manual	Manual	Numtec – Stamping
	Max Casting Speed (130 mm sq)	M/min	3.4	3.8	4.4
	Present Capacity	million tons	0.8	1	1.1
	Capacity planned under next revamp	million tons	1 (FY '11)	1.2 (FY '11)	
Ladle	Heat size	t	150		
	Vortex breaker	Y/N	No		
	Slag detector	Y/N	No		
	Slide Gate	Make	LS-70		
	Slide gate control	Type	Manual		
	Lining	Type	Tar Dolo with Mag C slag zone		
	Porous Plug	Make	Vesuvius - IPV 2		
	Nozzle size	mm	65		
Tundish	Size	t	24	32	~ 34
	Shape	Type	Delta	Delta -T	Delta – T
	Covering type		Basic + Rice husk		
	Anti-vortex	Type	Impact pot flow modifier – Turbostop		
	Lining	Type	MgO dry powder		
	Casting Modes	Type	Open & Submerged	Open	Open & Submerged
	Nozzle Changer	Type	CNC- Vesuvius		
	Stopper Control	Type	Electro-Mech	None	Hydraulic

	Preheater	Type	CEBA	None	CEBA
	SEN	Type	Alumina C with zirconia band	NA	
Mould	Type	**Type**	Tube	Tube with slot (at region of radioactive source)	Tube with slot (at region of radioactive source)
	Mould Tubes Supplier	Make	Concast, Europa Metalli	Concast, Europa Metalli	Concast, Europa Metalli
	Mould Taper	**Type**	Parabolic, Convex	Parabolic, Convex	Parabolic, Convex
	Mould Material	Type	Convex - Cu Ag, EM -		
	Coating Material	Type	Chromium		
	Length	mm	900	1000	1000
	Life	Hours	250	150mm2 - 500, 130mm2 - 250	300
	Wall thickness	mm	13	150mm2 - 15 (11 at slot area), 130mm2 - 13 (9 at slot area)	150mm2 - 15 (11 at slot area), 130mm2 - 13 (9 at slot area)
	Water Channel gap	mm	3.5	4	4
	Inner jacket	Type	Self aligning with ribs	Fixed	Self aligning with buttons
	Max water flow	lpm	2200	2400	2400
	Operating flows	lpm	2000	1800 (130), 2100 (150)	2000 (130), 2200(150)
	Back Pressure	Bar			
	Foot Rolls	No.s	Single row	Single row	Single row
	Level Control	Type	Co60 – Berthold	Co60 - Berthold	Co60 – Berthold
	Machine start	Type	Automatic	Automatic	Automatic
	Lubrication	Type	Oil / Powder	Oil	Oil/Powder
	Type of Oil	Type	Rapeseed	Synthetic	Rapeseed
	Supplier of Casting Powder	Make	Stollberg	-	Stollberg
	Flux feeding	Type	Manual	NA	Manual
	Oscillation	Type	Electro-Mech	Hydraulic - Dynaflex	Hydraulic – Dynaflex
	Oscillation stoke	mm	3.4 - 4.1-5.2-6.4-7.6-8.8-9.8-10.8-11.5-12.5-12.7	Variable - 4 to 14 mm linked to casting speed	Variable - 4 to 14 mm linked to casting speed
	Oscillation Frequency	1/min	60 - 250 cpm linked to casting speed	30-300 cpm	30-300 cpm
	Oscillation type	Type	Fixed stroke and variable frequency	Variable stroke and frequency - Inverse oscillation	Variable stroke and frequency - Inverse oscillation
	Negative strip %	%	Fixed, -9	Variable	Variable
	Negative strip time	sec	Variable -0.08-0.12	0.12	0.09 - 0.11
Withdrawal Unit	Withdrawal	Type	Top & Bottom roll driven	Top & Bottom roll driven	Top & Bottom roll driven
	Straightener	Type	No drive	Top roll driven	Top roll driven
	Aux. Withdrawal	Y/N	Top roll driven	-	-
	Withdrawal Pressure (dummy bar/ strand)	bar			
	Drive	type	Cardon shaft	Chainless	Cardon Shaft
Spray Cooling	Type	Spray/mist	Water spray	Water spray	Water spray
	Dynamic / Manual	Type	Dynamic	Dynamic	Dynamic
	Zones	No.	3	4	4
EMS	Type	Mould/ strand / final	Mould EMS	Mould EMS	Mould & Final EMS
	Make	Make	Concast	DG	ABB
	Max current	Amp	400	400	400
	Frequency	Hz	3 – 7	3.5 - 7	3.5 - 5.5
	Location (zone of max gauss)	mm, from top of copper tube	300 (up position) 450 (down position)	600	700
	Poles	nos	6	6	6

Data analysis, predictive model building.

What is data analysis?

Data analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making.Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, and is used in different business, science, and social science domains.

What is data cleaning?

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.

When combining multiple data sources, there are many opportunities for data to be duplicated or mislabeled. If data is incorrect, outcomes and algorithms are unreliable, even though they may look correct. There is no one absolute way to prescribe the exact steps in the data cleaning process because the processes will vary from dataset to dataset. But it is crucial to establish a template for your data cleaning process so you know you are doing it the right way every time.

How do you clean data?

While the techniques used for data cleaning may vary according to the types of data your company stores, you can follow these basic steps to map out a framework for your organization.

Step 1: Remove duplicate or irrelevant observations

Remove unwanted observations from your dataset, including duplicate observations or irrelevant observations. Duplicate observations will happen most often during data collection. When you combine data sets from multiple places, scrape data, or receive data from clients or multiple departments, there are opportunities to create duplicate data. De-duplication is one of the largest areas to be considered in this process.

Irrelevant observations are when you notice observations that do not fit into the specific problem you are trying to analyze. For example, if you want to analyze data regarding millennial customers, but your dataset includes older generations, you might remove those irrelevant observations. This can make analysis more efficient and minimize distraction from your primary target—as well as creating a more manageable and more performant dataset.

Step 2: Fix structural errors

Structural errors are when you measure or transfer data and notice strange naming conventions, typos, or incorrect capitalization. These inconsistencies can cause mislabeled categories or classes. For example, you may find “N/A” and “Not Applicable” both appear, but they should be analyzed as the same category.

Step 3: Filter unwanted outliers

Often, there will be one-off observations where, at a glance, they do not appear to fit within the data you are analyzing. If you have a legitimate reason to remove an outlier, like improper data-entry, doing so will help the performance of the data you are working with. However, sometimes it is the appearance of an outlier that will prove a theory you are working on.

Remember: just because an outlier exists, doesn't mean it is incorrect. This step is needed to determine the validity of that number. If an outlier proves to be irrelevant for analysis or is a mistake, consider removing it.

Step 4: Handle missing data

You can't ignore missing data because many algorithms will not accept missing values. There are a couple of ways to deal with missing data. Neither is optimal, but both can be considered.

As a first option, you can drop observations that have missing values, but doing this will drop or lose information, so be mindful of this before you remove it.
As a second option, you can input missing values based on other observations; again, there is an opportunity to lose integrity of the data because you may be operating from assumptions and not actual observations.
As a third option, you might alter the way the data is used to effectively navigate null values.

Step 5: Validate and QA

At the end of the data cleaning process, you should be able to answer these questions as a part of basic validation:

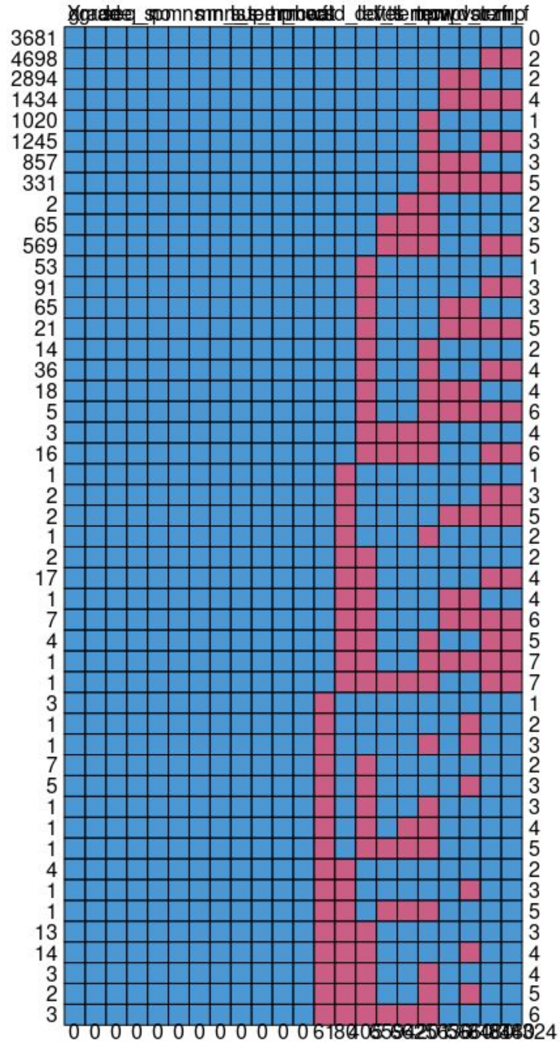
- Does the data make sense?
- Does the data follow the appropriate rules for its field?
- Does it prove or disprove your working theory, or bring any insight to light?
- Can you find trends in the data to help you form your next theory?
- If not, is that because of a data quality issue?

False conclusions because of incorrect or "dirty" data can inform poor business strategy and decision-making. False conclusions can lead to an embarrassing moment in a reporting meeting when you realize your data doesn't stand up to scrutiny.

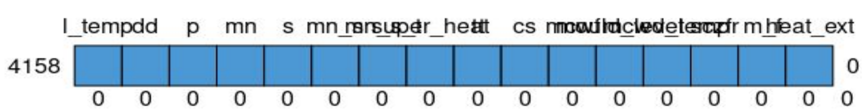
Before you get there, it is important to create a culture of quality data in your organization. To do this, you should document the tools you might use to create this culture and what data quality means to you.

Data cleaning tools and software for efficiency:

Data cleaning tools like spreadsheets sql and R is use to clean the data and one-hot encoding of variables.



Before data cleaning there was lots of vulnerabilities in data and data was looked like full of “NA” values shown in above diagram where pink boxes are missing data. After data cleaning data is ready for standardization and normalization.



After data cleaning the dataset is shown like above with no missing and duplicate values.

About the data:

The data contains the key features to predict the rhomboidity or diagonal difference of billets by linear regression model using R programming language. The initial data contains the data from CC#2, with around twenty thousands inputs and each of them contains forty six columns with several “NA” and “0” values. So we have to clean the data and prepare it for data transformation. After cleaning the data we finally got our data with three thousand six hundred data points. After that we have to standardize and normalize the data to prepare it for model building.

grade_t	grade_s	seq_no	l_temp	dd	p	mn	s	mn_s	mn_s_t	super_heat	tt	cs	mcwf	mould_level	mcwd_temp	sczfr	m_f	heat_ext
F	980	S5	1508	6	0.023	0.712	0.027	26.37037037	23.61991192	43	1551	2.88	2242.53	58.82	7.91	273.44	94.94444444	25868517.94
F	980	S10	1508	2	0.025	0.751	0.028	26.82142857	22.94811947	47	1555	2.71	2251.06	58.94	7.08	257.56	95.04059041	24700191.94
F	980	S15	1508	6	0.019	0.75	0.027	27.77777778	23.61991192	37	1545	2.9	2252.73	59.12	6.9	274.88	94.7862069	22511763.93
F	980	S25	1508	10	0.025	0.757	0.023	32.91304348	26.82423957	21	1529	3.09	2248.91	59.11	7.52	293.42	94.9579288	22986916.97
F	980	S30	1508	7	0.019	0.857	0.028	30.60714286	22.94811947	47	1555	2.76	2241.4	59.08	7.56	262.32	95.04347826	25785845.22
F	980	M	1508	15	0.024	0.726	0.021	34.57142857	28.83191546	65	1573	2.34	2238.96	58.5	7.46	223.75	95.61965812	29979100.31
F	976	S14	1510	4	0.018	1.075	0.026	41.34615385	24.33786233	29	1539	2.61	2234.79	59.14	6.68	247.64	94.88122605	24022708.14
F	980	S19	1508	11	0.022	0.719	0.031	23.19354839	21.1678094	47	1555	2.87	2239.75	58.94	7.92	274.12	95.51219512	25959248.78
F	980	S23	1508	5	0.024	0.747	0.028	26.67857143	22.94811947	-8	1500	2.8	2240.76	58.93	7.51	265.61	94.86071429	25242161.4
F	980	S29	1508	8	0.021	0.721	0.017	42.41176471	34.09448306	37	1545	3.09	2243.29	59.1	7.89	292.64	94.70550162	24057651.79
F	980	S5	1508	7	0.019	0.711	0.026	27.34615385	24.33786233	45	1553	2.94	2242.28	59.1	7.91	277.86	94.51020408	25337764
F	980	S10	1508	8	0.02	0.74	0.023	32.17391304	26.82423957	55	1563	2.74	2240.92	58.86	7.82	261.92	95.59124088	26861597.26
F	980	S15	1508	11	0.02	0.757	0.029	26.10344828	22.31802259	36	1544	2.5	2242.51	58.95	6.98	239.19	95.676	26296569.26
F	980	S20	1508	10	0.012	0.737	0.02	36.85	29.96988564	37	1545	2.44	2243.04	59.09	7.07	231.05	94.69262295	27297061.38
F	980	S25	1508	10	0.023	0.77	0.027	28.51851852	23.61991192	47	1555	2.88	2245.47	58.57	7.22	275.14	95.53472222	23642927.88
F	980	S30	1508	10	0.013	0.724	0.023	31.47826087	26.82423957	46	1554	2.88	2243.2	59.12	7.59	273.34	94.90972222	24829420
F	980	M	1508	6	0.022	0.742	0.021	35.33333333	28.83191546	48	1556	2.27	2229.37	58.53	7.07	217.05	95.61674009	29162516.64
F	980	S4	1508	7	0.025	0.74	0.03	24.66666667	21.72572514	44	1552	2.88	2233.43	58.85	7.9	273.56	94.98611111	25730974.79
F	980	S14	1508	13	0.018	0.8	0.029	27.5862069	22.31802259	45	1553	2.88	2235.88	58.84	7.68	273.86	95.09027778	25041856
F	980	M	1508	11	0.027	0.725	0.019	38.15789474	31.21470086	55	1563	2.31	2239.89	58.6	7.12	219.45	95	28996394.18
F	980	S5	1508	11	0.015	0.745	0.019	39.21052632	31.21470086	45	1553	2.72	2241.29	59.07	7.38	258.7	95.11029412	25540817.96
F	980	S10	1508	8	0.043	0.728	0.032	22.75	20.64126415	37	1545	2.31	2239.67	59.14	6.78	218.84	94.73593074	27609022.91
F	980	S15	1508	6	0.02	0.774	0.024	32.25	25.93359266	42	1550	3.15	2235.06	58.94	7.84	299.64	95.12380952	23363827.2
F	980	S25	1508	7	0.014	0.74	0.026	28.46153846	24.33786233	47	1555	3.03	2237.22	58.9	7.66	289.04	95.39273927	23754403.25
F	980	M	1508	12	0.021	0.742	0.02	37.1	29.96988564	66	1574	2.21	2235.3	58.48	7.29	211.06	95.50226244	30968513.76
F	980	S5	1508	9	0.031	0.842	0.021	40.0952381	28.83191546	40	1548	3.09	2238.96	59.01	7.9	291.92	94.47249191	24041648.16
F	980	S10	1508	8	0.029	0.733	0.023	31.86956522	26.82423957	48	1556	2.83	2238.73	58.86	7.67	270.15	95.45936396	25483550.61
F	980	S15	1508	6	0.021	0.755	0.018	41.94444444	32.58285073	57	1565	2.72	2239.5	58.89	7.33	258.91	95.1875	25347517.28
F	980	S20	1508	10	0.022	0.77	0.028	27.5	22.94811947	50	1558	2.66	2235.54	58.88	7	253.56	95.32330827	24708600
F	980	M	1508	12	0.019	0.72	0.028	25.71428571	22.94811947	62	1570	2.26	2240.86	58.76	7.48	214.98	95.12389381	31149937.06
F	980	S2	1508	8	0.029	1.146	0.027	42.44444444	23.61991192	41	1549	3.09	2237.24	59.14	8.27	292.1	94.53074434	25148315.26
F	980	S5	1508	15	0.017	0.754	0.03	25.13333333	21.72572514	47	1555	3.06	2237.71	59.22	7.86	287.89	94.08169935	24140942
F	980	S20	1508	11	0.02	0.754	0.025	30.16	25.10710746	57	1565	3.08	2239.17	58.96	7.89	294.04	95.46753247	24091433.59

What is standardization?

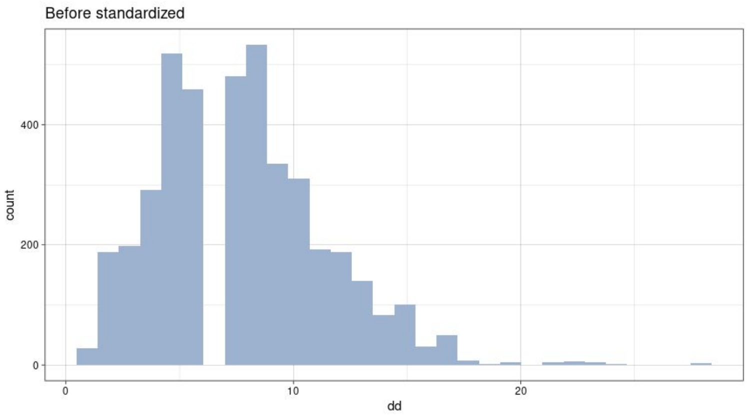
It’s the process of making something, usually information, consistent.To understand how standardized data is the key to scaling analytics, it’s important to understand how it works. Standardizing data focuses on transforming raw data into usable information before it’s analyzed. Raw data can contain variations in entries that are meant to be the same that could later affect data analysis. As part of data prep, the data that needs to be standardized will be changed to be consistent across all entries. Values will all be in the same format, and variables will be consistent. Standardizing data can help make regression, patterns, and deviations easier to pick out of a dataset. Once the information in the dataset is consistent and standardized, it will be significantly easier to analyze and use. The key is to find a solution for quickly standardizing data.

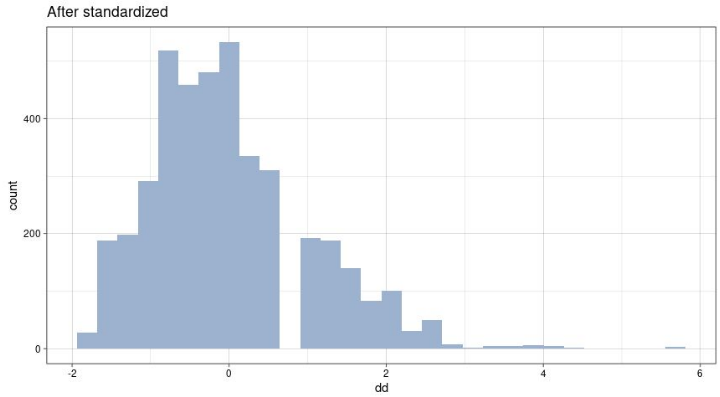
What is Normalization?

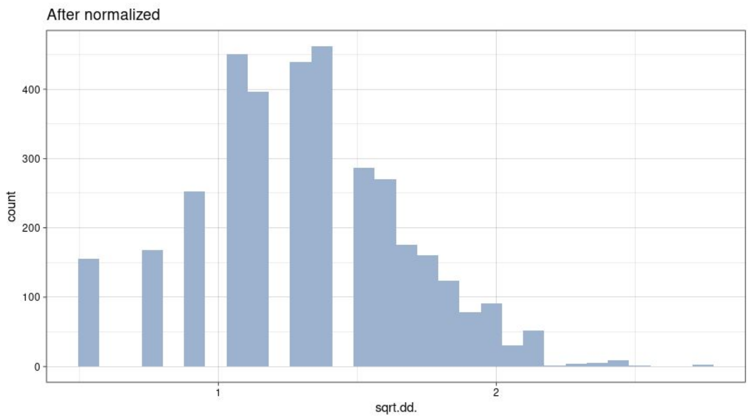
In statistics and applications of statistics, normalization can have a range of meanings.In the simplest cases, normalization of ratings means adjusting values measured on different scales to a notionally common scale, often prior to averaging. In more complicated cases, normalization may refer to more sophisticated adjustments where the intention is to bring the entire probability distributions of adjusted values into alignment. In the case of normalization of scores in educational assessment, there may be an intention to align distributions to a normal distribution. A different approach to normalization of probability distributions is quantile normalization, where the quantiles of the different measures are brought into alignment.

After standardization and normalization of data now our data is ready for model building. Let’s take a look into the data.

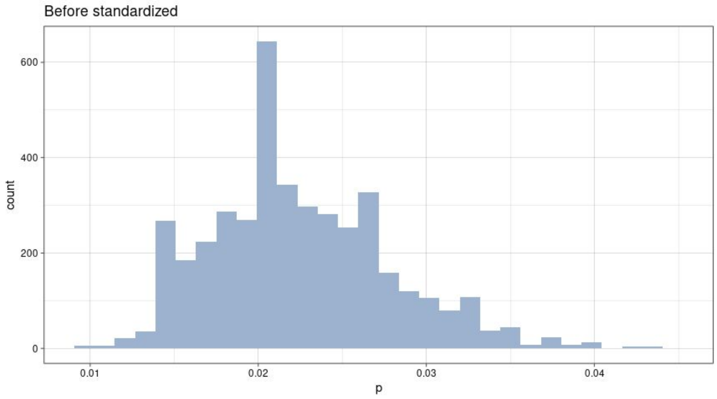
Diagonal difference:

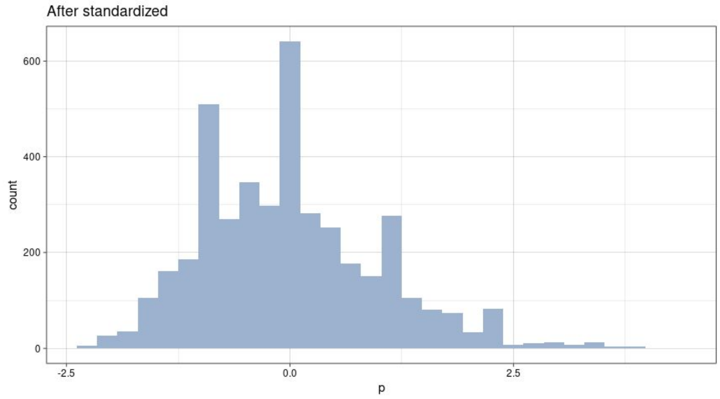


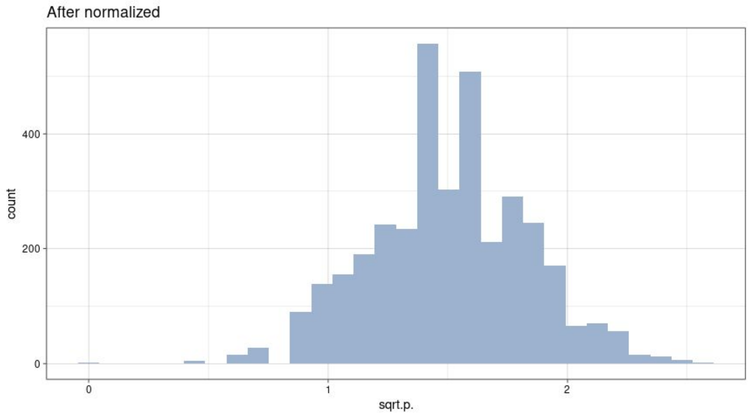




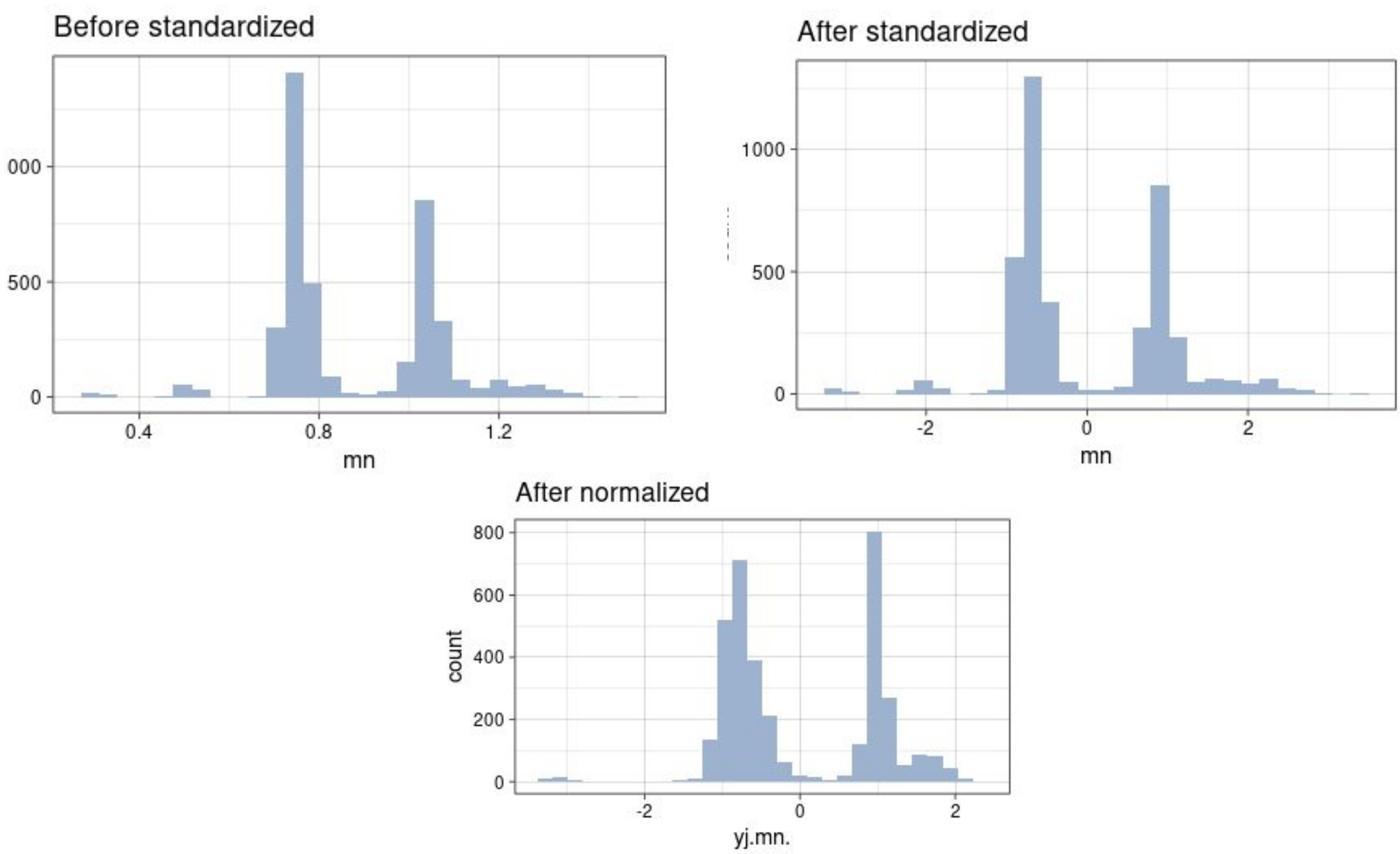
Phosphorus content:



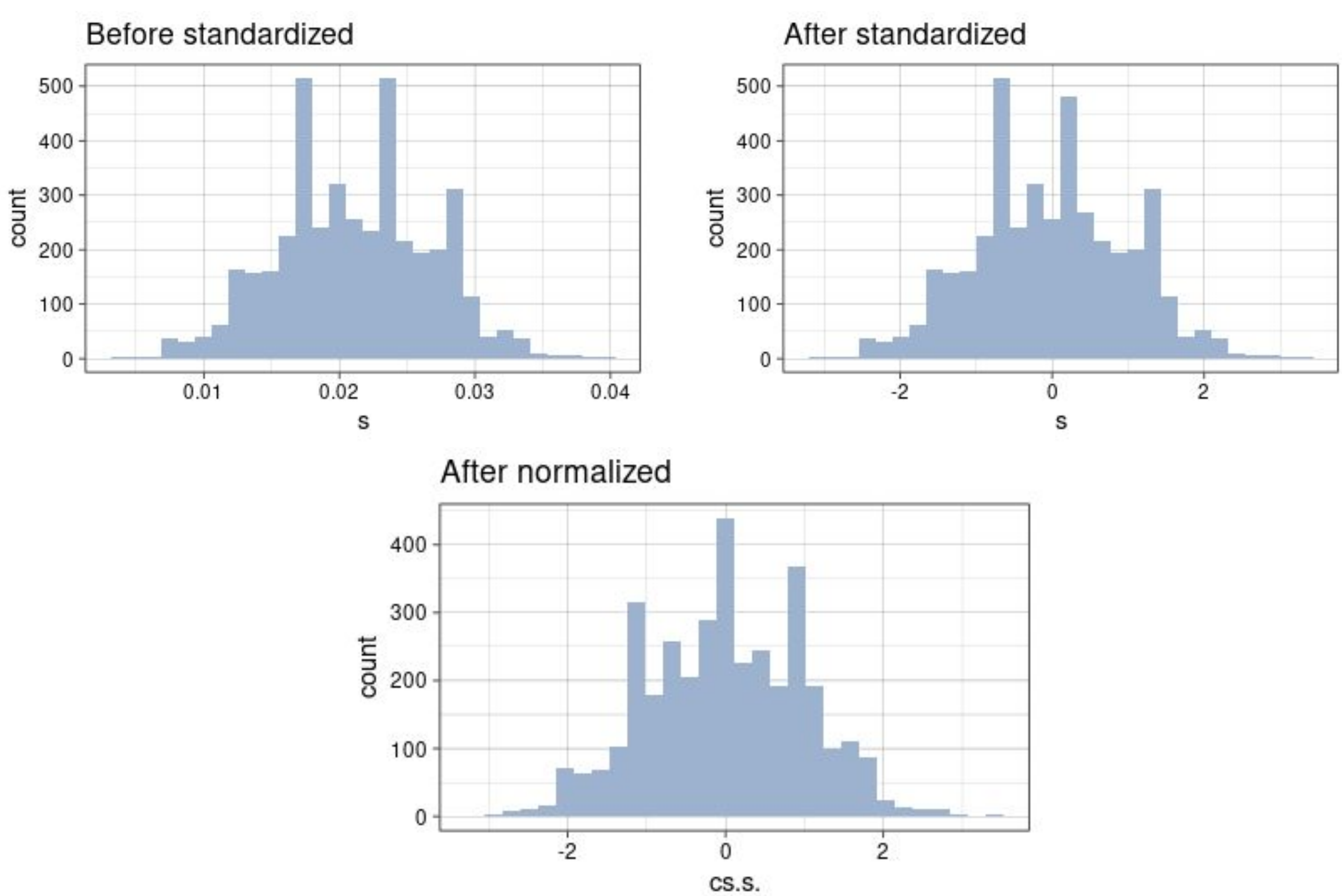




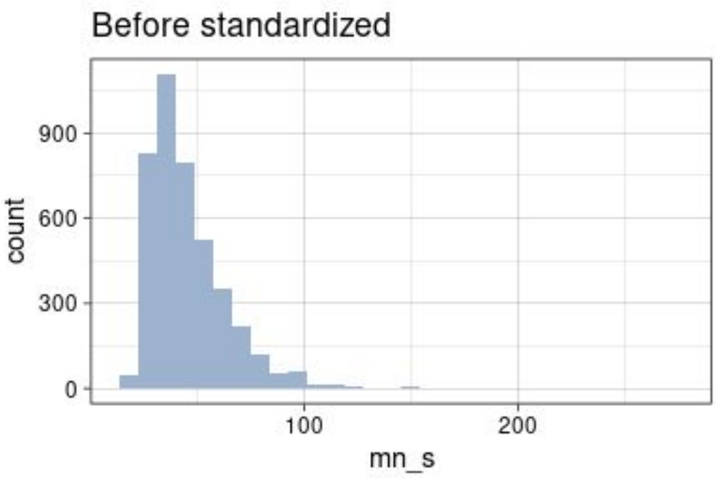
Manganese content:

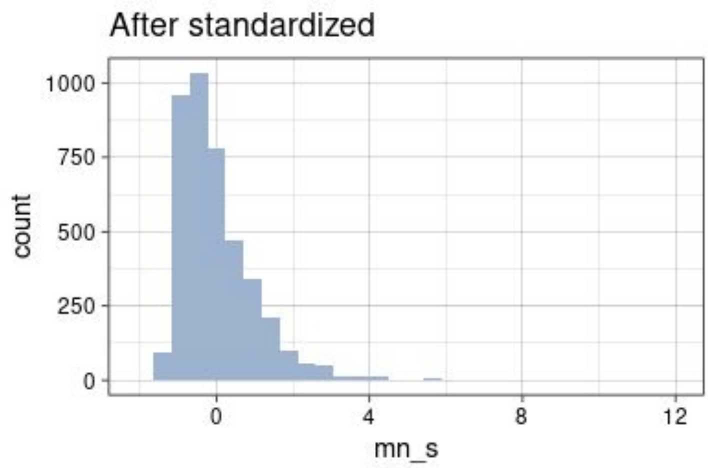


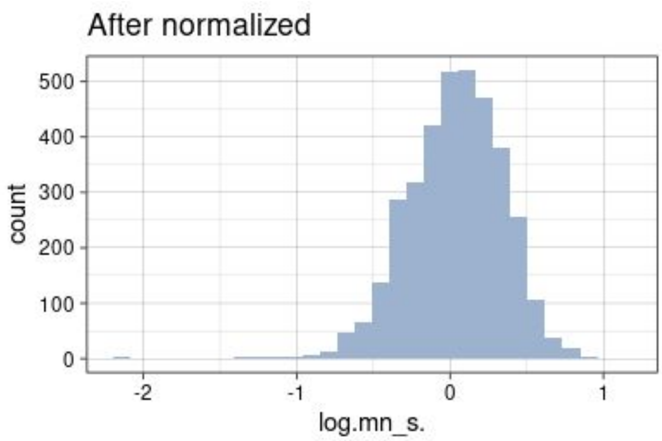
Sulphur content:



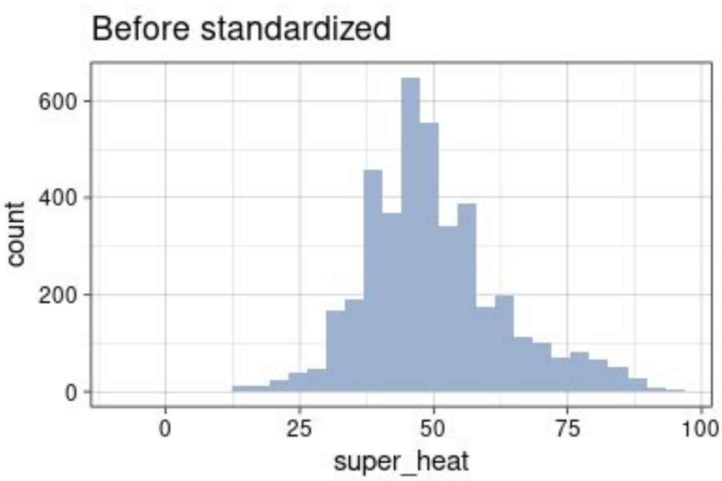
Mn/S ratio:

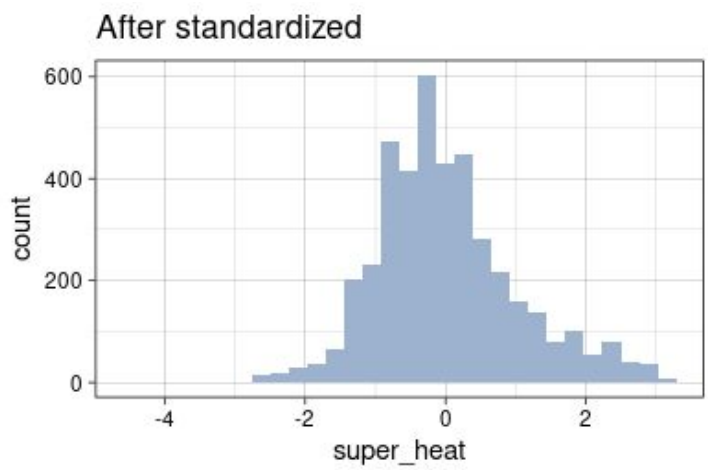


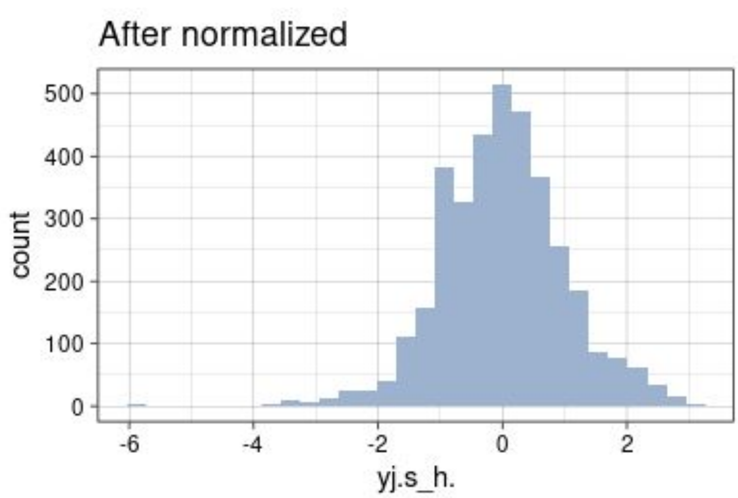




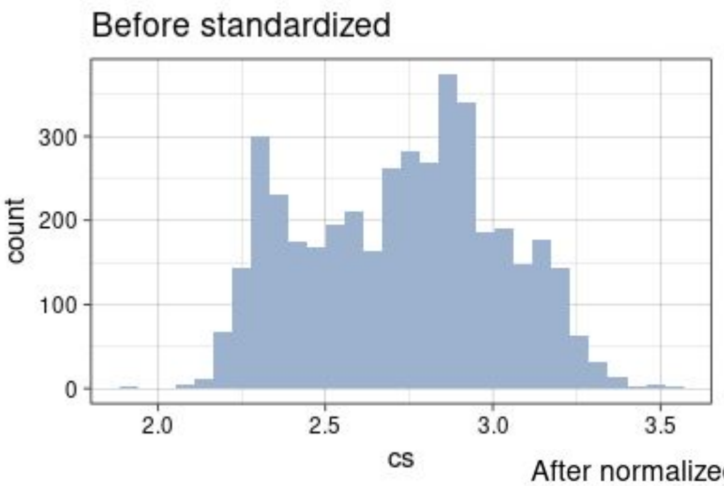
Superheat:

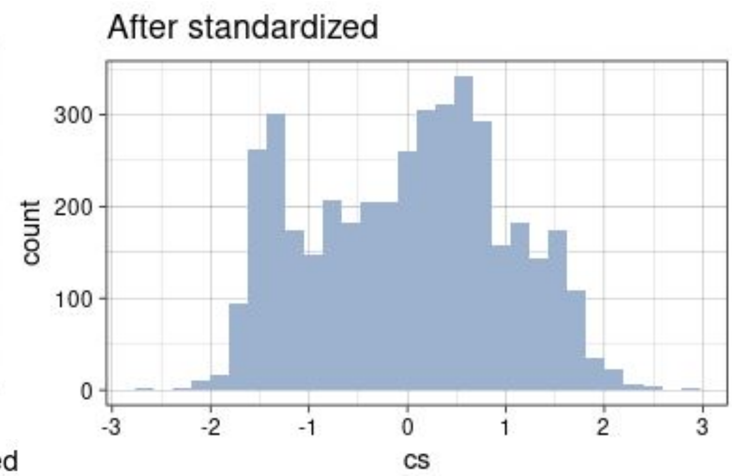


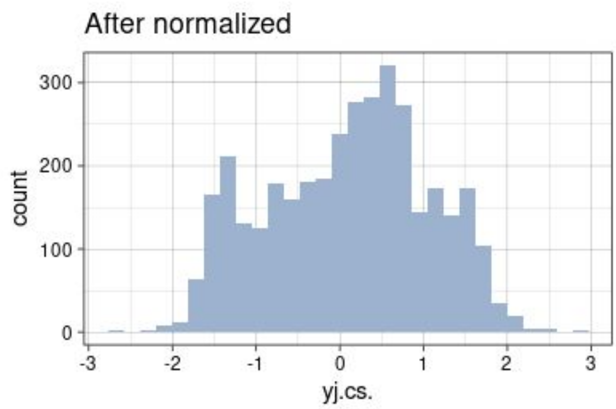




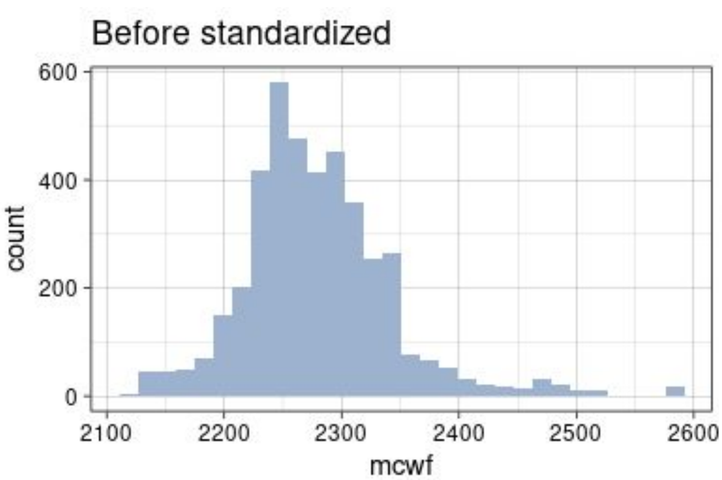
Casting speed:

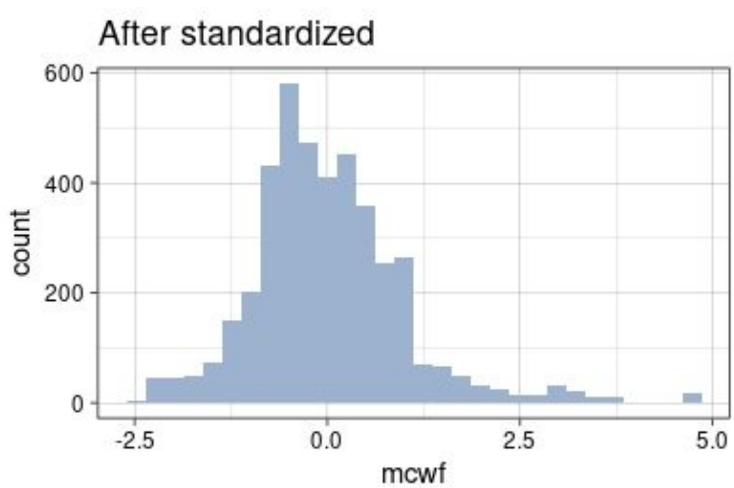


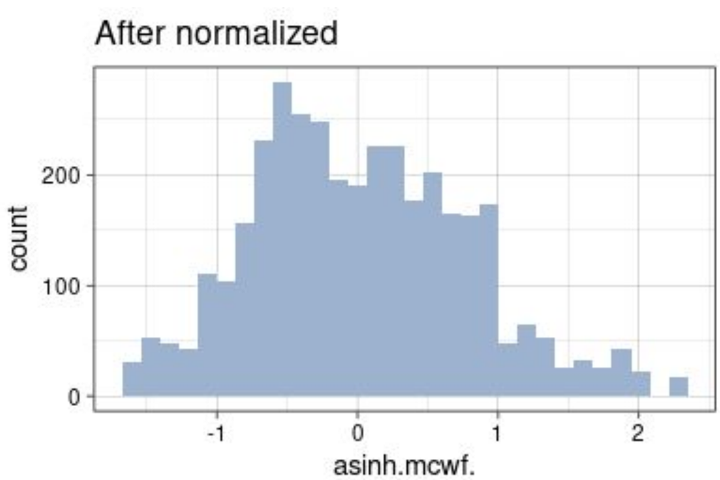




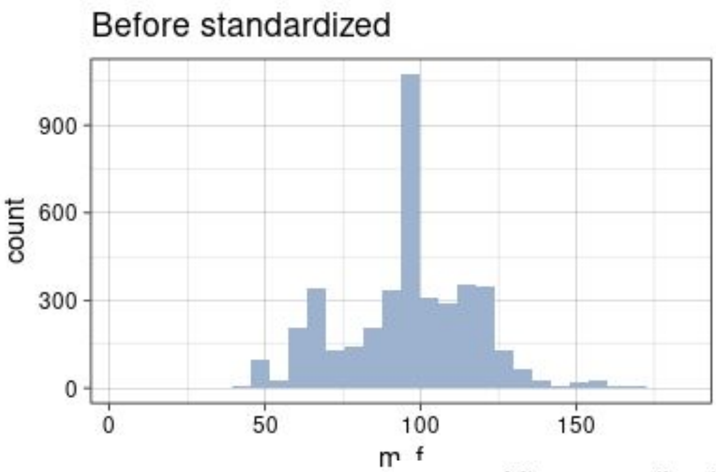
Mould cooling water flow rate:

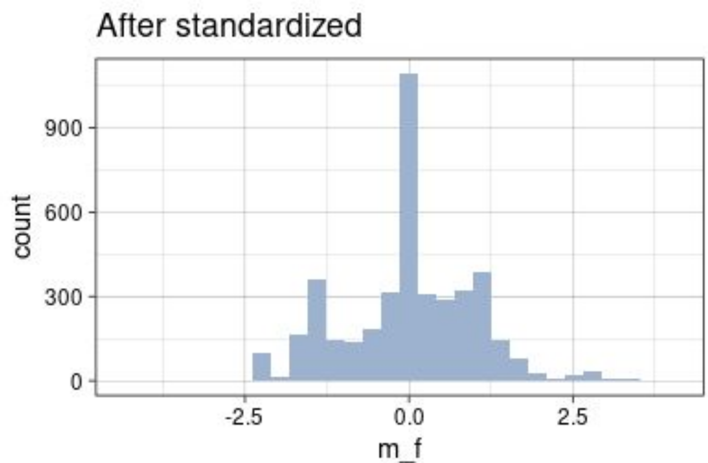


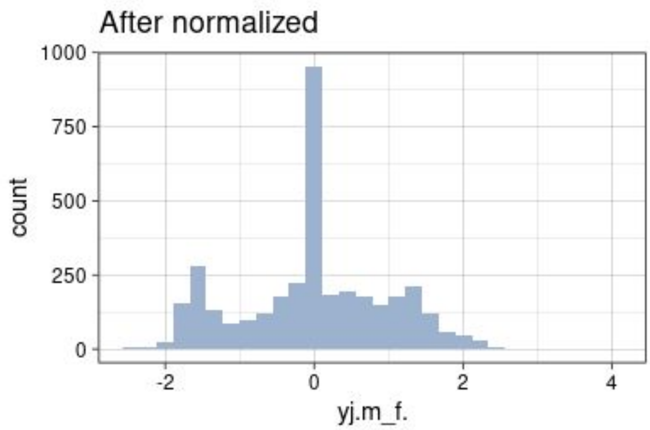




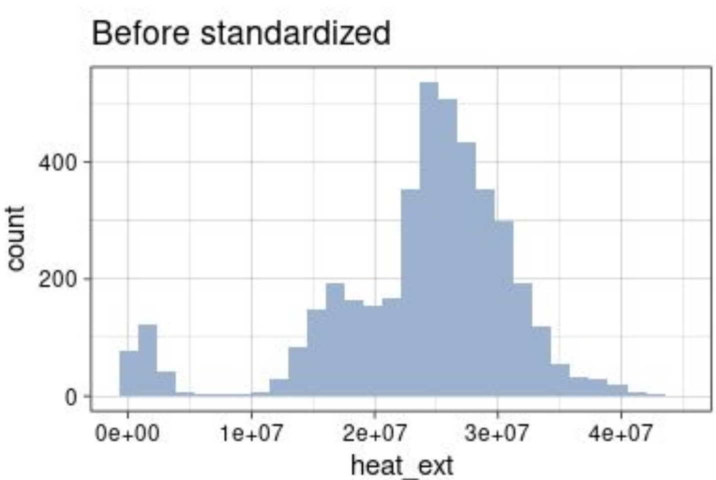
M factor:

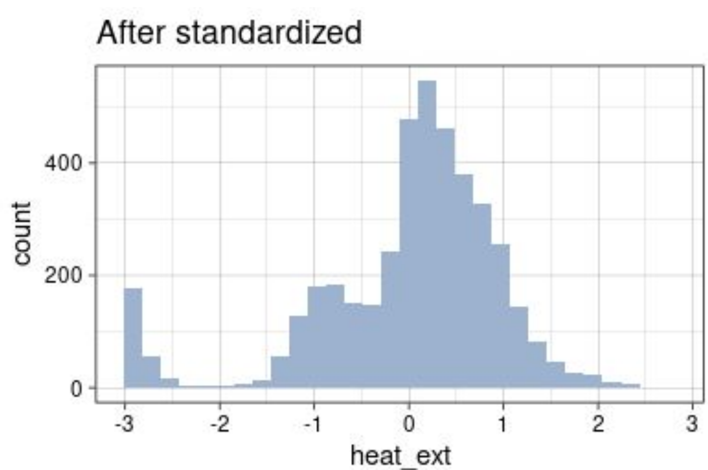


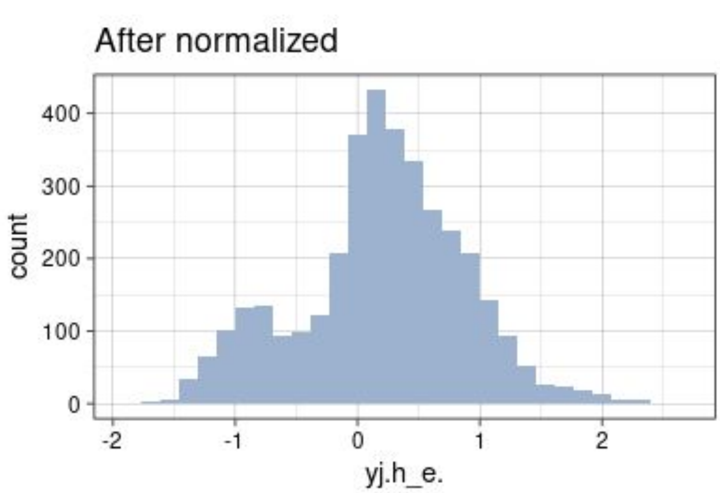




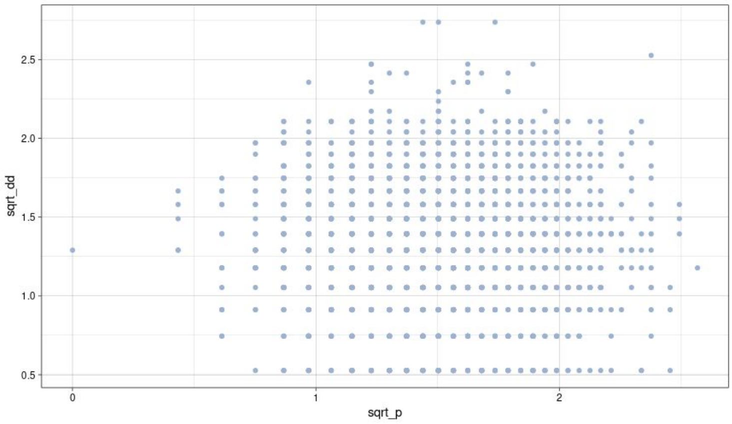
Heat extracted:

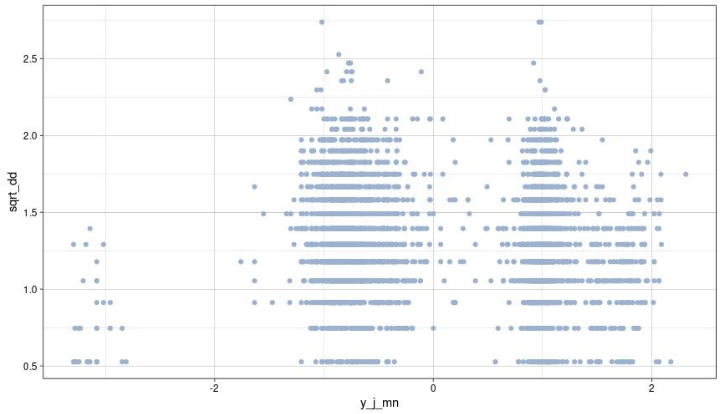


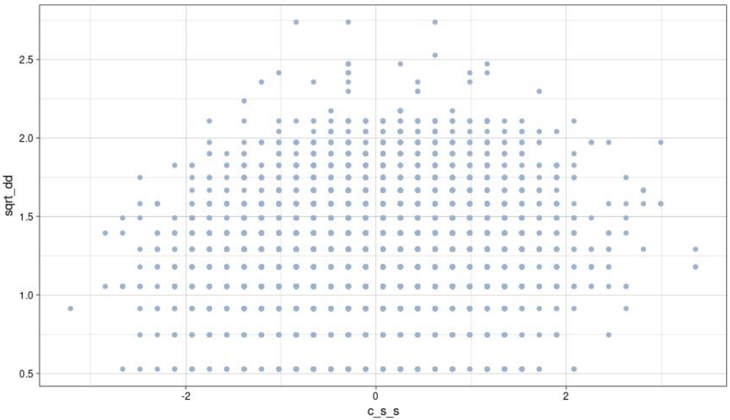


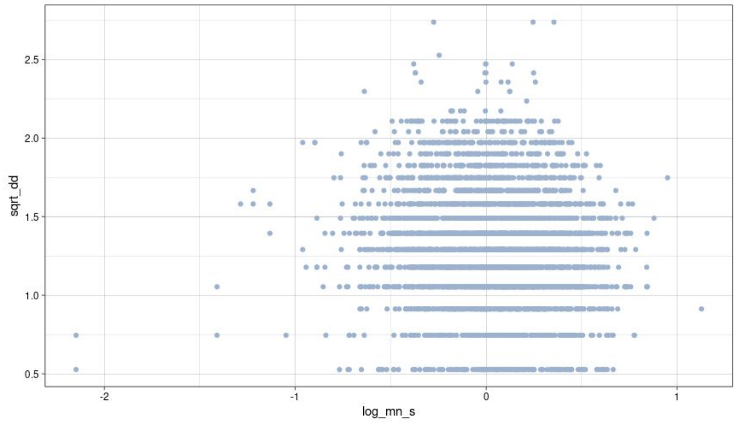


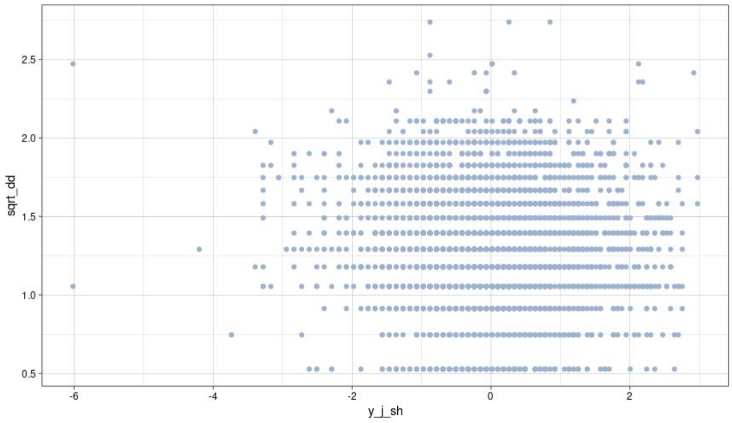
Scatter plot with variables and diagonal difference after normalization:

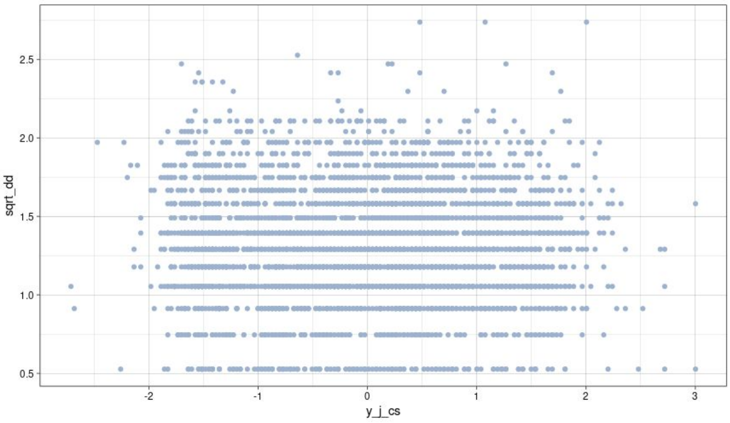


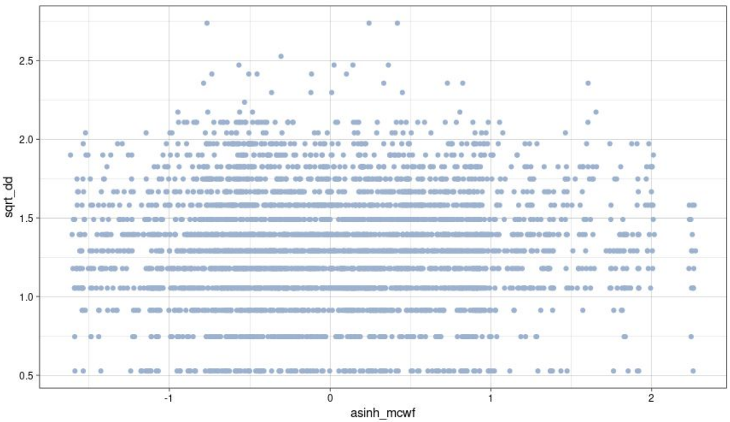


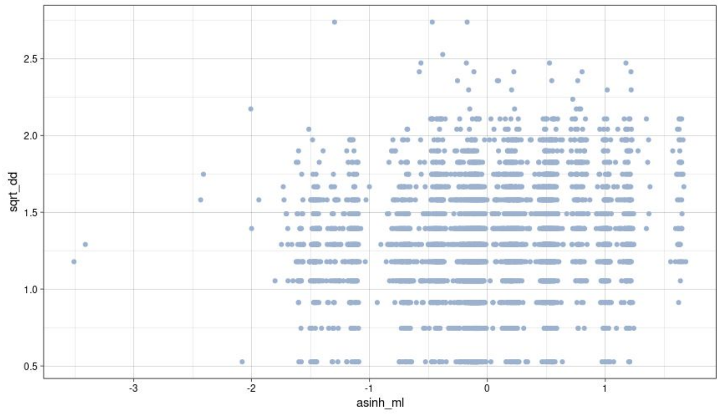


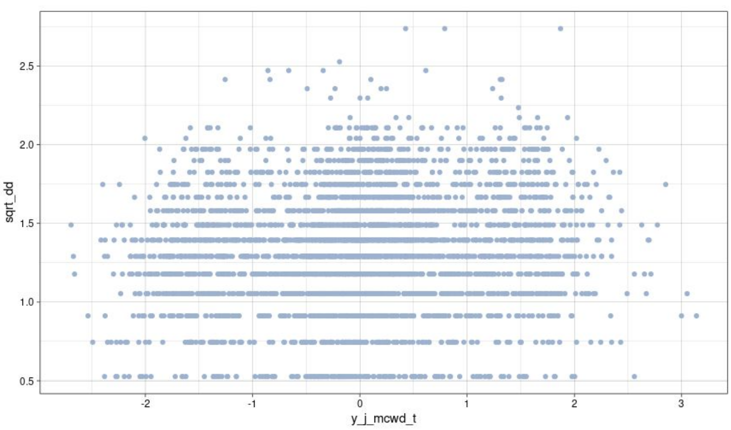


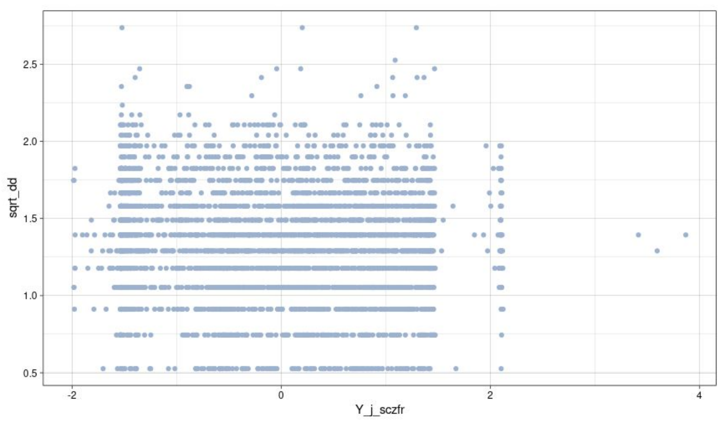


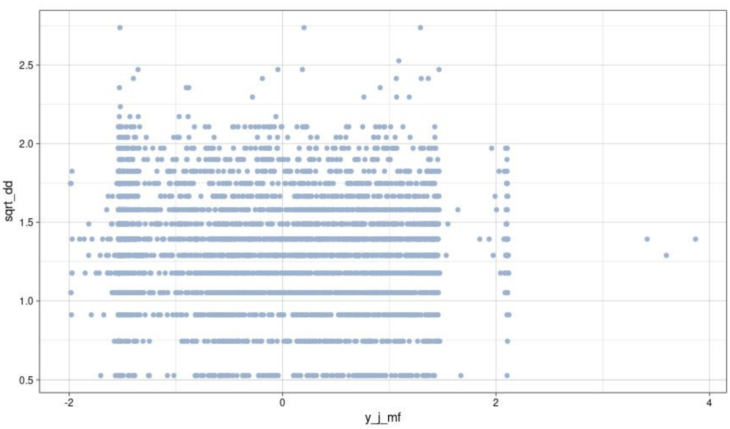


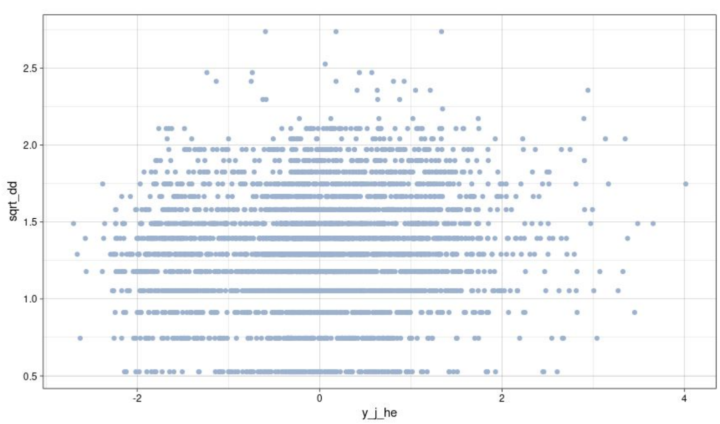

sqrt_dd
2.5
2.0
1.5
1.0
0.5
-1
0
1
2
asinh_mcwf


sqrt_dd
2.5
2.0
1.5
1.0
0.5
-3
-2
-1
0
1
asinh_ml


sqrt_dd
2.5
2.0
1.5
1.0
0.5
-2
-1
0
1
2
3
y_j_mcwd_t


sqrt_dd
2.5
2.0
1.5
1.0
0.5
-2
0
2
4
Y_j_sczfr


sqrt_dd
2.5
2.0
1.5
1.0
0.5
-2
0
2
4
y_j_mf

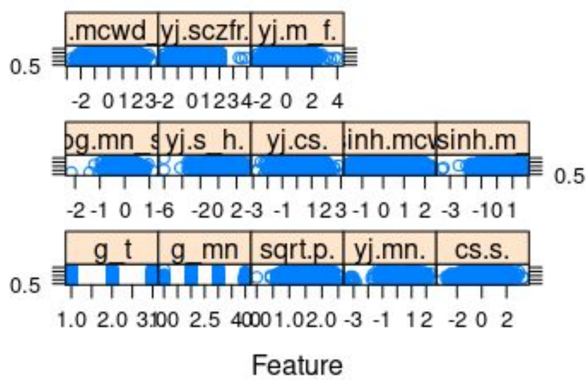

sqrt_dd
2.5
2.0
1.5
1.0
0.5
-2
0
2
4
y_j_he

Data normalization process:

Data normalization is done by R programming language using “best Normalize” package. This package helps and suggests the normalization method after analyzing the data. The data points are normalized by :

- Diagonal difference : sqrt(dd)
- P content : sqrt(p)
- Mn content : yeo johnsons(mn)
- S content : center scale(s)
- Mn/S ratio : $\log_{10}$(mn/s)
- Superheat : yeo johnsons(sh)
- Casting speed : yeo johnsons(cs)
- Mould level : asinh(ml)
- Mound cooling water flow rate : asinh(mcwf)
- M factor : yeo johnsons(mf)
- Heat extracted : yeo johnsons(he)

After normalization data is ready for model building and training of model. 70% of normalized data will be used as training and rest of 30% data will be used as validation data.



Building the actual model:

In statistics, linear regression is a linear approach to modelling the relationship between a scalar response and one or more explanatory variables (also known as dependent and independent variables). The case of one explanatory variable is called simple linear regression; for more than one, the process is called multiple linear regression. This term is distinct from multivariate linear regression, where multiple correlated dependent variables are predicted, rather than a single scalar variable.

In linear regression, the relationships are modeled using linear predictor functions whose unknown model parameters are estimated from the data. Such models are called linear models. Most commonly, the conditional mean of the response given the values of the explanatory variables (or predictors) is assumed to be an affine function of those values; less commonly, the conditional median or some other quantile is used. Like all forms of regression analysis, linear regression focuses on the conditional probability distribution of the response given the values of the predictors, rather than on the joint probability distribution of all of these variables, which is the domain of multivariate analysis.

In a linear regression model we use linear model to predict Y form the value X. If a data set is given as $\{Y_i, X_1, X_2, \ldots\ldots\ldots, X_n\}$.Thus the model takes the form:

$$Y_i = C + (M_1X_1 + M_2X_2 + \ldots\ldots\ldots + M_nX_n)$$

Where Y_i is the value we want to predict C is the intercept and $X_1, X_2, \ldots, X_n$ are the variables from which we can predict the Y_i.

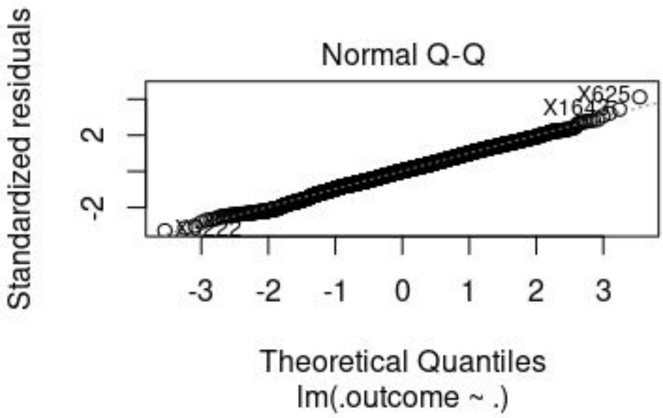
In our case Y is the diagonal difference and X values are the rest of other variables in final dataset. We used R to build our model. We used “Caret” as our model building library.

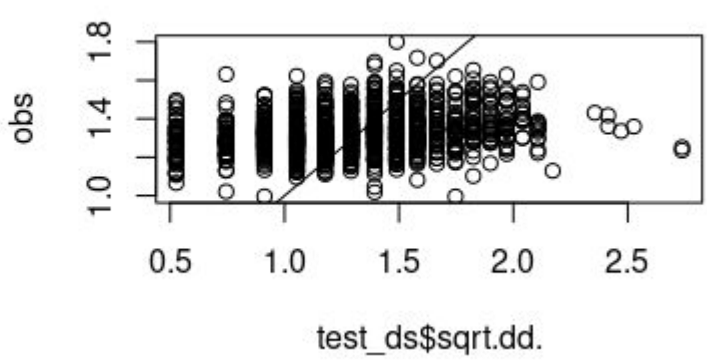
Introduction to Caret:

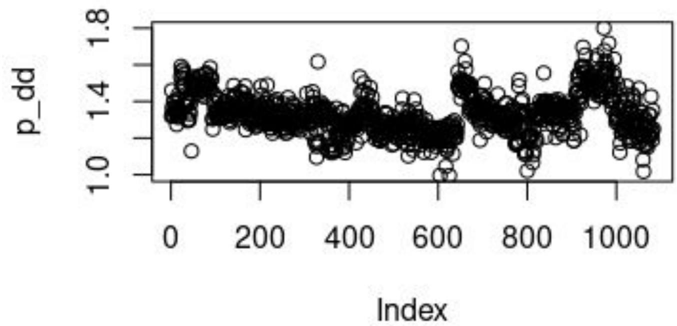
The caret package (short for Classification And REgression Training) is a set of functions that attempt to streamline the process for creating predictive models. The package contains tools for:

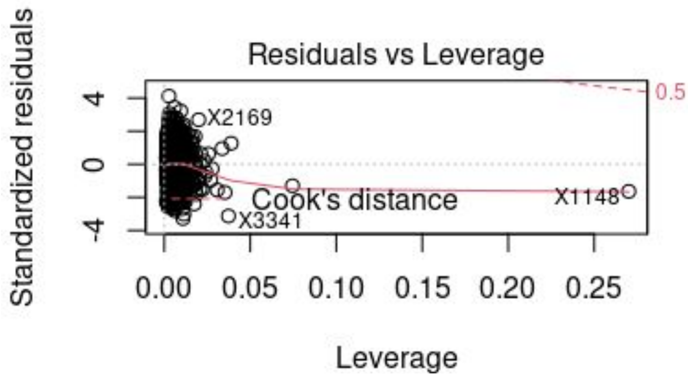
- data splitting
- pre-processing
- feature selection
- model tuning using resampling
- variable importance estimation

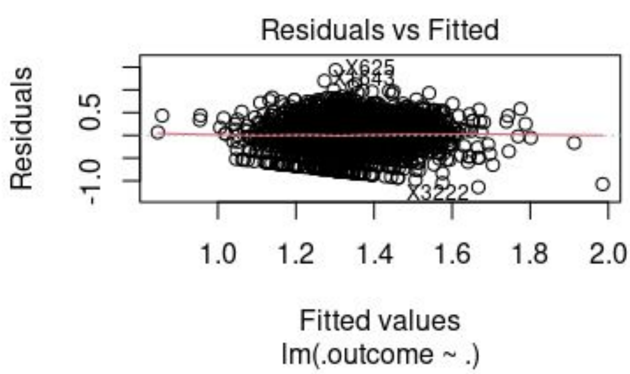
Now when our data is fully prepared and splited into training and test data, we are ready to use it. So we used R studio to compile our R code. So we loaded the csv file and start train the model.
After training and validating the model our predictive model is ready.

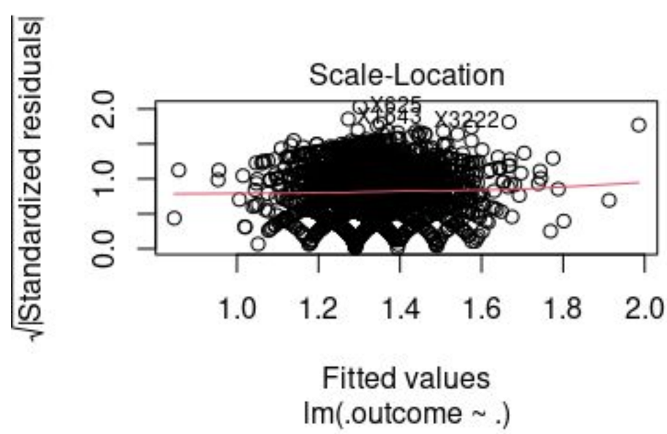












The final model is following:

Sqrt_dd = 1.733767 + (-0.004348sqrt(p)) + (0.137070yj(mn)) + (-0.016241cs(s)) + (-0.095583log(mn/s)) + (-0.007528yj(sh)) + (0.107360y(cs)) + (-0.022141asinh(mcwf)) + (0.126433asinh(ml)) + (-0.155385yj(mcwd_t)) + (-0.002183yj(sczfr)) + (0.017456yj(mf)) + (0.395928yj(he))

CONCLUSIONS

- Some of our variables shows some correlation with rhomboidity of billet casting.
- The correlation between manganese content in steel and the mould level of continuous caster have some effect on rhomboidity.
- The mould cooling water flow rate plays a significant role in rhomboidity.
- The M factor and heat extraction are also have effect on rhomboidity.