

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ

V PRAZE

FAKULTA STROJNÍ

ÚSTAV TECHNOLOGIE OBRÁBĚNÍ,

PROJEKTOVÁNÍ A METROLOGIE



BAKALÁŘSKÁ PRÁCE

MĚŘENÁ DATA A JEJICH
STATISTICKÉ ZPRACOVÁNÍ

ONDŘEJ NOVÁK

PRAHA 2024

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Novák** Jméno: **Ondřej** Osobní číslo: **502356**
Fakulta/ústav: **Fakulta strojní**
Zadávací katedra/ústav: **Ústav technologie obrábění, projektování a metrologie**
Studijní program: **Výroba a ekonomika ve strojírenství**
Studijní obor: **Technologie, materiály a ekonomika strojírenství**

II. ÚDAJE K BAKALÁŘSKÉ PRÁCI

Název bakalářské práce:

Měřená data a jejich statistické zpracování

Název bakalářské práce anglicky:

Measured data and their statistical processing

Pokyny pro vypracování:

1. Význam statistiky pro zpracování dat.
2. Možnosti testování naměřených dat.
3. Získání a zpracování naměřených dat.
4. Zhodnocení statistického zpracování naměřených dat.

Seznam doporučené literatury:

Jméno a pracoviště vedoucí(ho) bakalářské práce:

Ing. BcA. Jan Podaný, Ph.D. ústav technologie obrábění, projektování a metrologie FS

Jméno a pracoviště druhé(ho) vedoucí(ho) nebo konzultanta(ky) bakalářské práce:

Datum zadání bakalářské práce: **12.10.2023**

Termín odevzdání bakalářské práce: **02.01.2024**

Platnost zadání bakalářské práce: _____

Ing. BcA. Jan Podaný, Ph.D.
podpis vedoucí(ho) práce

Ing. Libor Beránek, Ph.D.
podpis vedoucí(ho) ústavu/katedry

doc. Ing. Miroslav Španiel, CSc.
podpis děkana(ky)

III. PŘEVZETÍ ZADÁNÍ

Student bere na vědomí, že je povinen vypracovat bakalářskou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací. Seznam použité literatury, jiných pramenů a jmen konzultantů je třeba uvést v bakalářské práci.

Datum převzetí zadání

Podpis studenta

Čestné prohlášení

Prohlašuji, že jsem bakalářskou práci na téma Měřená data a jejich statistické zpracování vypracoval samostatně a použil pouze podklady uvedené v příloženém seznamu.

V Praze dne

.....

Ondřej Novák

Poděkování

Tímto bych rád poděkoval vedoucímu mé bakalářské práce panu Ing. BcA. Janu Podanému, Ph.D. za cenné rady, odbornou pomoc, a především ochotu řešit úskalí, která v průběhu vznikla.

ANOTACE

Tato bakalářská práce se zabývá problematikou statistického zpracování naměřených dat. Cílem této práce bylo vytvořit ucelený soubor hlavních metod, které k tomuto účelu můžeme využít. Aby to bylo možné, jsou nejdříve definovány základní pojmy z oblasti metrologie a statistiky. Dále se práce zabývá rozděleními pravděpodobnosti. Zvláště je kladen důraz na normální a Weibullovo rozdělení. Normální rozdělení je předpokladem pro množství testů a modelů. Bakalářská práce se proto zabývá tím, jak vhodně otestovat a zdali data, se kterými pracujeme, jsou z normálního rozdělení. Náplň praktické části je testování tří skupin dat na normalitu grafickými i početními metodami. Dvě sady dat jsou generovány náhodně, přičemž jedna je z normálního rozdělení a druhá je z Weibullova rozdělení, které normální rozdělení přibližně aproximuje a je sledováno, jakým způsobem ho testy zachytí. Poslední skupina dat je získána měřením čepů ze sériové výroby. Funkčnost testů se zde demonstruje na praktickém příkladu.

Klíčová slova

Zpracování dat, statistické testy, normální rozdělení, testy normality

ANNOTATION

This bachelor's thesis deals with the issues of statistical processing of measured data. The aim of this thesis was to create a comprehensive set of main methods that can be used for this purpose. To achieve this, fundamental concepts from the fields of metrology and statistics are first defined. The thesis then delves into probability distributions, with particular emphasis on the normal and Weibull distributions. The normal distribution is a prerequisite for many tests and models. Therefore, the bachelor's thesis examines how to appropriately test and determine whether the data being worked with follow a normal distribution. The practical part focuses on testing three groups of data for normality using graphical and numerical methods. Two sets of data are randomly generated, where one is from a normal distribution and the other is from a Weibull distribution, which approximates the normal distribution. The aim is to observe how these tests capture the distribution. The last group of data is obtained by measuring pins from a serial production line. The functionality of the tests is demonstrated here with a practical example.

Keywords

Data processing, statistical tests, normal distribution, tests of normality

OBSAH

Úvod.....	12
1 Metrologie	13
1.1 Vývoj Metrologie	13
1.2 Chyby měření.....	14
1.2.1 Systematická chyba	15
1.2.2 Hrubá chyba	15
1.2.3 Náhodná chyba.....	15
2 Statistika.....	16
2.1 Základní pojmy	17
2.1.1 Aritmetický průměr.....	17
2.1.2 Směrodatná odchylka	18
2.1.3 Kvantily.....	18
2.2 Statistický software	18
2.2.1 Excel.....	19
2.2.2 Minitab.....	20
2.2.3 R – programovací jazyk.....	21
3 Rozdělení pravděpodobnosti.....	23
3.1 Náhodný pokus a náhodný jev.....	23
3.2 Náhodná veličina a její popis	24
3.3 Diskrétní rozdělení pravděpodobnosti.....	25
3.3.1 Alternativní rozdělení.....	26
3.3.2 Binomické rozdělení	27
3.3.3 Poissonovo rozdělení.....	28
3.3.4 Geometrické rozdělení.....	29
3.4 Spojité rozdělení pravděpodobnosti.....	30

3.4.1	Rovnoměrné rozdělení	31
3.4.2	Exponenciální rozdělení	32
3.4.3	Weibullovo rozdělení	33
3.5	Normální rozdělení	36
3.5.1	Normované normální rozdělení	37
4	Testování statistických hypotéz	39
4.1	Jednovýběrový Studentův t-test	40
4.2	Testování odlehlých hodnot	42
4.2.1	Dean-Dixonův test	42
4.2.2	Grubbsův test	44
5	Testy normality	46
5.1	Grafické metody	46
5.1.1	Histogram	46
5.1.2	Q-Q graf	48
5.1.3	P-P graf	49
5.2	Ověření normality výpočtem	49
5.2.1	Kolmogorovův-Smirnovův test	49
6	Zpracování získaných dat	52
6.1	Získání dat	52
6.1.1	Průběh měření	52
6.2	Testy normality veličiny X	55
6.2.1	Graficky	55
6.2.2	Početně	57
6.3	Testy normality veličiny Y	59
6.3.1	Graficky	59
6.3.2	Početně	61

6.4	Testy normality veličiny Z	63
6.5	Další testy veličiny Z.....	66
6.5.1	Grubbsův test	66
6.5.2	Studentův t-test.....	68
	Závěr.....	69
	Citovaná literatura	71
	Seznam obrázků.....	77
	Seznam tabulek	78
	Přílohy.....	79

Seznam symbolů a zkratek

\bar{x}	aritmetický průměr
n	počet členů výběru
μ	střední hodnota rozdělení
i	pořadí veličiny
s, σ	směrodatná odchylka, výběrová směrodatná odchylka
Ω	prostor elementárních jevů
X, Y, Z	náhodné veličiny
p	pravděpodobnost
$P(x)$	pravděpodobnostní funkce
$f(x)$	hustota pravděpodobnosti
$F(x)$	distribuční funkce
$E(x)$	střední hodnota
$D(x)$	rozptyl
$A(p)$	alternativní rozdělení
$Po(\lambda)$	Poissonovo rozdělení
$Bi(n, p)$	binomické rozdělení
$Rs(a, b)$	rovnoměrné rozdělení
$Ex(\lambda)$	exponenciální rozdělení
$W(\eta, \beta, \gamma)$	Weibullovo rozdělení
$N(\mu, \sigma^2)$	normální rozdělení
$\Gamma(x)$	gamma funkce
R	rozpětí souboru
Q_{\min}, Q_{\max}	testovací kritéria Dean-Dixonova testu
T_{\min}, T_{\max}	testovací kritéria Grubbsova testu

k počet intervalů

h šířka intervalu

ÚVOD

Měření a zpracování dat hraje klíčovou roli v mnoha odvětvích výzkumu i praxe. S rostoucí rychlostí jejich sběru vzniká potřeba je i nějakým způsobem vhodně interpretovat. Jsou hledány metody, jak co nejlépe vytěžit informace z dostupných zdrojů. Tato práce se zabývá právě tímto tématem, tedy statistickým zpracováním dat. Zvláštní pozornost je v ní věnována rozdělení pravděpodobností, zejména pak normálnímu a Weibullovu. Zjištění, z jakého rozdělení pravděpodobnosti data pocházejí, by měl být jeden z prvních kroků při jejich analýze.

V úvodu práce jsou nastíněny hlavní principy metrologie a statistiky, které spolu s teorií pravděpodobnosti a testy normality tvoří základ pro praktickou část práce.

Cíl praktické části je sledování chování testů normality a porovnání jejich citlivosti. Za tímto účelem jsou zpracovávány tři soubory dat. Jeden soubor dat je záměrně vytvořen tak, že nepochází z normálního rozdělení, ale blízce se mu podobá a bude pozorováno, jestli ho testy normality zachytí. Aby bylo docíleno požadovaných vlastností jsou první dva soubory generovány náhodně.

Třetí testovaný soubor jsou výsledky měření provedené na sériové výrobě, kde také často nalézají uplatnění testy normality. Ověření normality zde využijeme jako základ před provedením dalších testů, které ji vyžadují.

1 METROLOGIE

Metrologie se jako vědní a technická disciplína zabývá poznatky a činnostmi, které se týkají měření. Pomáhá nám systematicky stanovovat hodnoty fyzikálních veličin, což je pro mnoho odvětví vědy, průmyslu, státní správy, a dokonce ochrany zdraví a životního prostředí nezbytné. Správné měření je neodmyslitelně spojeno s jakoukoli efektivní výrobou. [1]

Metrologie nám pomáhá plnit tyto tři hlavní úkoly:

- 1) Definice základních jednotek měření, které jsou mezinárodně uznané. Takovou jednotkou může být například metr.
- 2) Realizace těchto jednotek měření vědeckými metodami. Například měření metru je realizováno za pomoci laserových paprsků.
- 3) Určení řetězců návaznosti cestou stanovení a zdokumentování této hodnoty a přesnosti měření a přenosu daných údajů. Příkladem v provozu přesného strojírenství je dokumentovaný vztah mezi mikrometrickým šroubem a primární laboratoří metrologie optické délky. [2]

1.1 Vývoj Metrologie

Od počátků známé civilizace měli lidé potřebu porovnávat věci měřením. Důvodem bylo zajištění spravedlivé směny a obchodu. Části těla jako dlaň, loket nebo stopa se nabízejí jako základní a dostupné jednotky míry. V záznamu knihy Genesis nalézáme údaj, že Noemova archa byla 300 loktů dlouhá, 50 loktů široká a 30 loktů vysoká. [3] Není známo, proč byl tento údaj v Bibli uvedený, ale jedná se pouze o jeden příklad z mnoha, které se zde nalézají. Z toho je patrné, jakou roli hrála metrologie v symbolice ve starověké kultuře a umění. Není to však jediný doklad měření ve starověku. V Egyptě, tři tisíce let před naším letopočtem při stavbě prvních pyramid a chrámů, měl architekt povinnost kalibrovat¹

¹ „činnost, která za specifikovaných podmínek v prvním kroku stanoví vztah mezi hodnotami veličiny s nejistotami měření poskytnutými etalony a odpovídajícími indikacemi s přidruženými nejistotami měření a ve druhém kroku použije tyto informace ke stanovení vztahu pro získání výsledku měření z indikace“ [2]

své měřidlo délky o každém úplňku. Hrozící trest smrti za nedodržení tohoto nařízení naznačuje, s jakou vážností na něj bylo nahlíženo. Základní jednotkou míry byl první královský loket definovaný vzdáleností od natažené špičky prsteníčku k lokti zrovna vládnoucího faraona. Tato míra byla následně vytesána do žulového kamene. Z toho byly vytvořeny kopie rovněž ze žuly nebo ze dřeva a za jejich udržování zodpovídali architekti.

Na našem území se o rozvoj měření zasloužil Přemysl Otakar II. V roce 1268 ustanovil jednotnou královskou míru, čímž se pokusil o sjednocení měř a vah. Jednou ze základních jednotek byl pražský loket, který se dále dělil na tři pídě a ty se rovnaly deseti prstům umístěným vedle sebe. [4] Přes všechny snahy o sjednocení v průběhu století vznikala v Evropě spousta měrných systémů. Raketový rozvoj vědy, techniky a výroby na konci 18. století však byl hnacím motorem pro pokusy o jejich unifikaci. Jednotky vycházející z rozměrů lidského těla nebo jinak spojené s měřeným objektem již přestával plnit svoji funkci. V roce 1790 vznikla ve Francii komise, která tento problém měla řešit. Nové měrné jednotky měly být srozumitelné a hlavně trvalé. Již o rok později vznikla metrická soustava. Ta se zakládá na dvou pilířích, a to je metr, podle kterého se nazývá, a dekadickém počítání jeho násobků. Metr byl definován jako desetimiliontý díl jedné čtvrtiny zemského poledníku procházející Paříží. [5] Mezinárodní rozšíření metrického systému nastalo v roce 1875 na tzv. Metrické konvenci v Paříži, kdy ho oficiálně přijalo 18 zemí včetně Rakouska-Uherska. Tím byl položen kámen moderní metrologie, jak ji známe dnes [4]

1.2 Chyby měření

Provedeme-li opakovaná měření jedné fyzikální veličiny za stejných podmínek, pravděpodobně získáme různé hodnoty. Jelikož příslušné veličině musí náležet pouze jedna správná hodnota, odchylkám od skutečné hodnoty říkáme všeobecná chyba měření. [6]

Rozlišujeme troje chyby měření podle vzniku: systematická, náhodná a hrubá chyba.

1.2.1 Systematická chyba

Systematické chyby se opakují pravidelně a nemění se opakovanými měřeními. Příčinu jejich vzniku můžeme většinou dobře definovat. Může to být například zaokrouhlování, vlastnosti konstrukce měřidla, okolní vliv nebo lidský vliv. Výhodou je, že lze odstranit pomocí korekcí, provedeme-li měření přesnějším přístrojem a výsledky porovnáme. [7]

1.2.2 Hrubá chyba

Bude-li se nějaká hodnota nápadně lišit od zbylých hodnot ze stejného souboru, může se jednat o hrubou chybu. Většinou hrubé chyby vznikají nedodržením pravidel měření nebo chybou pracovníka, ať už nepozorností nebo omylem. [7]

Tyto chyby je nutné vyloučit ze souboru měření. Existují metody, které nám pomohou určit, jestli se jedná o hrubou chybu (Grubbsův test, Dean-Dixonův test)

1.2.3 Náhodná chyba

Náhodné chyby vznikají kvůli nekontrolovatelným vnějším vlivům. Vzájemně na sebe působící vlivy prostředí, měřícího přístroje a obsluhy a s určitou pravděpodobností snižují nebo zvyšují měřenou veličinu. Odchytky v měření jsou tedy vysvětlovány působením náhodných chyb.

2 STATISTIKA

Ani si to nemusíme uvědomit, ale se statistikou se setkáváme téměř každý den. Může to být hned z rána při pohledu na počasí. Čeká nás zamyšlení nad tím, jaká je pravděpodobnost, že bude pršet a zdali se vyplatí vzít si deštník. Při otevření novin nepotrvá dlouho, než narazíme na graf nebo tabulku vytvořenou pomocí statistického zpracování dat.

Všude, kde jsme schopni změřit data nebo se snažíme o interpretaci různých jevů, přichází na scénu statistika. Takže nás asi nepřekvapí, že je propletena napříč mnoha obory. V průmyslu pomáhá při vývoji výrobků a jejich produkci, ale to není zdaleka všechno. Je jí dále využito při rozhodování, analýze výrobních cen a plánování cen konečných. Kromě toho může podnik na základě makroekonomických statistických dat (inlace, nezaměstnanost, růst reálné mzdy atd.) sledovat vývoj na trhu, hlavně pak poptávky a tím předpovídat chování zákazníků. Statistika se rovněž využívá ve státních úřadech a institucích. Zde se podílejí na rozhodovacích procesech a vzniku strategií. Vytvářejí se podpůrná data k vládním materiálům, kdy se z velkého množství nepřehledných čísel stávají důležité ukazatele pro ekonomiku, sociální politiku, ale i průmysl a obchod. Rovněž se statistika hojně používá ve zdravotnictví, školství, zemědělství, marketingu, PR, farmakologii a mnoha dalších oborech. Jistě by šlo vypsát mnoho dalších oborů, kde se statistiky využívá, ale i tento značně neúplný výčet ukazuje na její rozšíření a široké možnosti využití. [8]

Slovo statistika vychází z latinského slova „status“, což by bylo možné přeložit jako „stav“ nebo přeneseně „stát“. Toto slovo popularizoval Gottfried Achenwall (1719-1772), který ho používal pro označení vědy určující sílu tehdejších států. Jednalo se o slovní popis území státu, armády, zemědělství atd. a to bylo doprovázeno číselnými údaji. Postupem času se obor vyvíjel a začalo se ustupovat od slovních popisů směrem k číselnému vyjádření. Tím vznikla také potřeba data nějak přehledně vyjádřit a začaly se používat první grafické metody. [9] Dnes statistiku definujeme jako vědeckou disciplínu zabývající se získáváním, analýzou a následným vyhodnocením a prezentací empirických dat. Motivací v dnešní době stejně jako v minulosti je získání informací, které nejsou

na první pohled patrné při zkoumání dat z určité oblasti, čímž nám pomáhá k jejímu lepšímu pochopení.

Samotnou statistiku můžeme dále dělit na jednotlivé dílčí disciplíny. Základní a nestarší je **popisná (deskriptivní) statistika**. Ta se zabývá vývojem nebo popisem stavu hromadných jevů. Nejdříve je nutné vymezit soubor prvků na kterém budeme jev zkoumat. Následně se soubor vyšetří z hlediska studovaného jevu. Výsledky se nejčastěji vyjádří jako číselný popis. Popisná statistika často tvoří první krok analýzy dat, protože nám poskytuje jejich základní popis. Novější disciplína statistiky je **matematická statistika**. Ta na rozdíl od popisné statistiky pracuje s pravděpodobností. Ze základního souboru prvků provádíme výběr a na takto získané údaje pohlížíme jako na výsledek náhodného pokusu. [10]

2.1 Základní pojmy

Tato kapitola se bude věnovat základním statistickým pojmům. Ačkoliv je jich celá řada, budou popsány jenom ty základní, se kterými se dále bude pracovat.

2.1.1 Aritmetický průměr

Aritmetický průměr je součet všech hodnot x_n podělený jejich počtem n . Značí se \bar{x} nebo v některých případech μ . Pořadí členu je značeno n . Můžeme ho vyjádřit rovnicí:

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n) = \frac{1}{n} \sum_{i=1}^n x_i \quad (2.1)$$

2.1.2 Směrodatná odchylka

Směrodatná odchylka určuje, jak moc jsou hodnoty rozptýleny od průměru hodnot. Značí se s nebo σ a vypočteme jí ze vztahu:

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (2.2)$$

V matematické statistice často používáme místo směrodatné odchylky výběrovou směrodatnou odchylku ve tvaru:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (2.3)$$

2.1.3 Kvantily

Chceme-li sledovat rozdělení hodnot v souboru, využíváme k tomu kvantily. Značí se $x_{\frac{i}{100}}$ pro horní polovinu souboru a $x_{\frac{1-i}{100}}$ pro tu spodní, kde i označuje procento hodnot menších nebo rovných $x_{\frac{i}{100}}$ a zároveň také procento hodnot větších nebo rovných $x_{\frac{1-i}{100}}$.

2.2 Statistický software

Je obecně přijímáno, že největším hybatelem změn ve statistice za posledních 50 let je rozvoj technologií. Technologii využíváme nejenom ke zpracování dat, ale také k jejich sběru. Satelitní snímkování, monitoring různých procesů, snímkování ve zdravotnictví, výzkum genomu nebo třeba počítačové průzkumy. Všechny tyto oblasti produkují nepředstavitelné množství dat. Tyto nové formy sběru dat vyžadují vývoj nových metod a nástrojů, k tomu, jak je zpracovávat. [11]

Statistický software nám umožňuje sbírat data, pomocí statistických metod je analyzovat a provést jejich prezentaci. S nástupem prvních počítačů v 50. letech 20. století se rovněž začaly objevovat první statistické programy. Další vývoj nastal v 70. a 80. letech, kdy vznikají první komerční programy jako

SAS (Statistical Analysis System), SPSS (Statistical Package for the Social Sciences) a MINITAB, které nabízejí široké spektrum statistických funkcí. Postupně se začínají rozšiřovat osobní počítače a se zlepšením uživatelského rozhraní a grafického prostředí. To vede k rozšiřování dostupnosti statistického softwaru. [12]

V dnešní době roste popularita open-source statistických nástrojů, které jsou zdarma k dispozici a nabízejí širokou škálu funkcí a balíčků. Dále se začínají ve větším měřítku uplatňovat cloudová řešení, které nám umožňují online ke statistickým nástrojům a softwaru. S rostoucími objemy dat se začínají objevovat nástroje na jejich zpracování a strojové učení jako je například softwarová knihovna TensorFlow. [13]

Jelikož je statistického softwaru nepřeborné množství, byl vybrán pouze nepatrný vzorek, který bude představen.

2.2.1 Excel

Ačkoliv Excel jako tabulkový procesor není primárně určen pro statistické účely, často je tak využíván. Jeho oblíbenost je zapříčiněna několika faktory. Tím že je běžně dostupný jako součást balíčku Microsoft Office ho má mnoho lidí k dispozici bez dalších nákladů na specializovaný statistický software. K Excelu existuje zároveň mnoho volně dostupných alternativ s podobnými parametry, které tento typ softwaru dělají ještě dostupnější. Znalost tohoto programu je také považována za běžnou dovednost v mnoha pracovních odvětvích, což mu dává výhodu nad ostatními počítačovými programy. Mnoha uživatelům také imponuje jeho uživatelská přívětivost, protože umožňuje pracovat s daty a provádět základní statistické operace bez hlubších znalostí statistických metod. Všechny tyto vlastnosti z něj paradoxně dělají jeden z nejrozšířenějších nástrojů pro statistické zpracování dat. [14]

Jako nástroj Excel dobře poslouží pro zjištění základních statistických ukazatelů jako je průměr, směrodatná odchylka, medián, kvartily a podobně. Umožňuje nám také provádět základní statistické testy jako t-test, analýzu rozptylu nebo korelační analýzu. Snadno s ním jdou také vytvořit základní typy

grafů, například ve statistice často používaný histogram. Chceme-li ale provádět podrobnější analýzu dat, začne Excel rychle narážet na svoje limity. Chybí mu široká škála statistických funkcí a metod, které jsou ve specializovaném softwaru běžně dostupné. Narazíme také na problémy při práci s velkými datovými soubory, chybějícími daty a pokročilou vizualizací dat. [15]

Excel je tedy vhodný jako základní nástroj, jak si rychle a jednoduše udělat analýzu dat, ale pro podrobnější zpracování je potřeba sáhnout po specializovaném statistickém softwaru.

2.2.2 Minitab

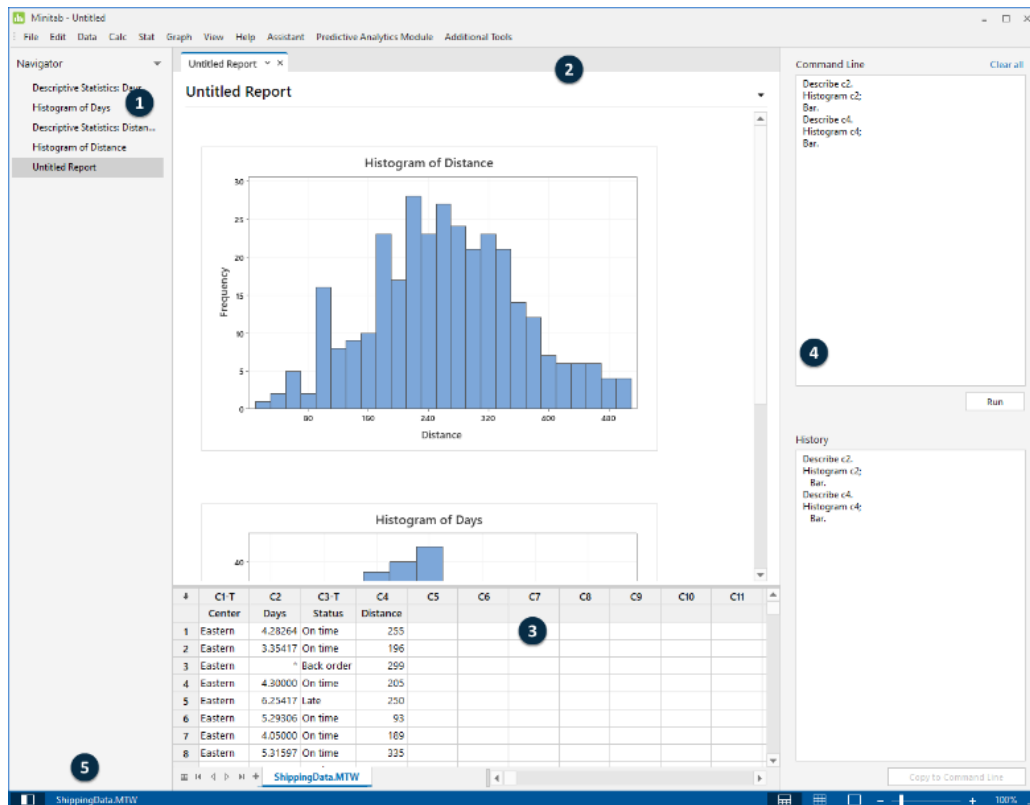
Minitab je statistický software, který poskytuje širokou škálu základních i pokročilých funkcí pro analýzu dat. Rozbalovací seznamy s funkcemi a dialogová okna ho dělají velmi intuitivní. Další výhodou je podpora přenášení dat mezi Minitabem a tabulkovými procesory. Minitab má jak desktopovou, ale i cloudovou verzi, což umožňuje spustit program z jakéhokoliv zařízení s připojením k internetu, čímž se stává dostupnější a umožňuje lepší sdílení dat. [16]

Často je používán na školách právě díky svojí uživatelské přívětivosti a díky tomu, že je vhodný pro použití uživateli na všech úrovních znalostí statistiky. Další odvětví, ve kterém je velmi populární je řízení kvality, protože je společně s implementací metody Six Sigma (snaží se o odstranění příčiny defektů a chyb v procesech výroby) a jiných statisticky založených metod využíván ke zlepšení procesů. [16] [17]

Na obrázku 1 je vidět náhled uživatelské prostředí programu Minitab. Na první pohled připomíná tabulkové procesory. Panel nástrojů a list s buňkami je podobný. Obsahuje však ještě další prvky:

1. Navigační podokno – Obsahuje názvy výstupů a zpráv z projektu. Výstupy jsou seřazeny od nejstarší po nejnovější, ty se nacházejí na spodu seznamu.
2. Karta výstupů – Zobrazuje grafy, tabulky a další výstupy po provedení analýzy.

3. Pracovní list- Místo pro zadání a vložení dat.
4. Příkazový řádek a historie – Místo kam lze vkládat příkazy pro provedení analýzy, z historie lze příkazy zpětně vyvolat a popřípadě je upravit před dalším použitím.
5. Stavový řádek – Ve stavovém řádku lze nelézt a upravovat zobrazení a skrytí karet a panelů. [18]



Obrázek 1 Prostředí softwaru Minitab [18]

2.2.3 R – programovací jazyk

R je programovací jazyk a prostředí pro statistiku a tvorbu grafických výstupů. Jelikož se jedná o volně šiřitelný software, tedy je zdarma, tak si přitáhl celou řadu uživatelů. R je zapisováno v příkazovém řádku, ale existují dostupná grafická uživatelská rozhraní, která práci s ním ulehčují. Oproti ostatnímu statistickému softwaru má relativně pomalou křivku učení, ale po naučení se z R stává nepřekonatelný statistický nástroj.

R nabízí širokou škálu statistických a grafických technik, které je možno rozšířit o dodatečné knihovny, někdy označovány jako balíčky. Ty jsou vytvořeny

samotnými uživateli a díky nim jsou k dispozici velmi specifické statistické funkce. Další velmi silnou stránkou jsou grafické výstupy na profesionální úrovni, do kterých je v případě potřeby snadno možné vložit vzorce a matematické symboly. Grafy mají nastaveno standardní vykreslování, ale uživateli je ponechána volnost, jak si výsledný vzhled upravit dle svého. R nabízí možnost úpravy nejenom grafických výstupů, ale i celé řady dalších parametrů k potřebám uživatele. [19]

3 ROZDĚLENÍ PRAVDĚPODOBNOSTI

Vyšetřujeme-li výskyt různých jevů nebo hodnoty veličin, zjistíme že jsou doprovázeny určitou nejistotou. Tyto činitele nemůžeme ovlivnit, označíme je tedy jako náhodné. K práci s nimi využíváme pravděpodobnosti a statistické metody. Vyznačením námi získané náhodné proměnné podél číselné osy získáme **rozdělení pravděpodobnosti**. To také můžeme chápat jako zobrazení, které každému elementárnímu jevu přiřazuje reálné číslo vyjadřující pravděpodobnost již zmíněného jevu.

3.1 Náhodný pokus a náhodný jev

Úplným základem statistického pozorování je **náhodný pokus**. To je proces, který když provedeme opakovaně, tak nám může dát rozdílný výsledek, aniž bychom změnili vstupní podmínky. Před provedením pokusu není výsledek známý, ale je možné určit množinu možných výsledků. Například při hodu klasickou kostkou nevíme, které číslo padne, ale můžeme určit, že výsledek se bude pohybovat v rozmezí jedné a šesti. Náhodným pokusem se však neoznačuje jenom hod kostkou nebo mincí, může to být i spočtení zmetků mezi dobrými kusy ve výrobě nebo venkovní měření teploty v určitou hodinu atd. [20]

Vedle náhodných pokusů můžeme ještě provádět pokusy deterministické. Ty při opakování za stejných podmínek dávají i stejný výsledek. Jedná se například o fyzikální pokusy nebo řízené chemické reakce. Je však zřejmé, že tyto pokusy nejsou z hlediska teorie pravděpodobnosti zajímavé.

Než samotný náhodný pokus je pro nás důležitější jeho výsledek. Ten označíme jako **náhodný jev**. Po provedení pokusu již můžeme říct, jestli určitý jev nastal, či nenastal. Jelikož jevů většinou může nastat velké množství, značí se pro přehlednost velkými písmeny abecedy. Budeme-li pracovat s nejjednoduššími výsledky náhodného pokusu, jedná se **elementární jevy**. Tedy takové, které již nejsou složené z dalších jiných jevů. Při hodu kostkou je elementárním jevem určení pouze jednoho čísla. Když budeme uvažovat, že jev A nastane při hodu čísla jedna, použijeme značení $A = \{1\}$. Jestli na kostce

bude po hodu skutečně číslo jedna, říkáme, že nastal jev A. Všechny možné výsledky elementárních jevů shrnuje **prostor elementárních jevů** Ω . Pro hod kostkou bude prostor elementárních jevů $\Omega = \{1, 2, 3, 4, 5, 6\}$. [20]

3.2 Náhodná veličina a její popis

Náhodná veličina je jednoznačně určena výsledkem náhodného pokusu. V podstatě přiřazuje číselnou hodnotu elementárním náhodným jevům. Například pro snazší zpracování výsledku hodu mincí přiřadíme lícové straně nulu a rubové straně jedničku. V určitých případech číselné přiřazení ani nemusíme provádět, protože hodnotou veličiny je přímo výsledek pokusu. Může se jednat o hod kostkou, kdy jednoduše spočítáme počet ok na vrchní straně kostky nebo výsledky různých měření (délka, počet vadných výrobků, hmotnost, počet přijatých telefonátů) a další. Nejčastější značení náhodné veličiny je velkými písmeny od konce abecedy X, Y, Z, ... nebo X_1, X_2, X_3, \dots a výsledné hodnoty, kterých nabývají x, y, z, \dots , respektive x_1, x_2, x_3, \dots . [21]

Na základě hodnot, kterých náhodné veličiny nabývají, je můžeme rozdělit na dva typy:

- 1) Diskrétní (nespojité) náhodná veličina – může být tvořena pouze celými čísly. Interval výsledných hodnot může být konečný i nekonečný, ale může se měnit pouze po skocích.
- 2) Spojitá náhodná veličina – je taková, která nabývá libovolné hodnoty z konečného i nekonečného intervalu. Bude se měnit spojitě bez skoků.

[22]

Budeme-li se snažit blíže popsat chování náhodné veličiny, využijeme **distribuční funkci**. Ta udává pravděpodobnost pro každé reálné číslo, že hodnota nabytá náhodnou veličinou bude menší nebo rovna než toto číslo. To můžeme vyjádřit následující rovnicí:

$$F(x) = P(X \leq x) \quad (3.1)$$

Distribuční funkce se standardně značí $F(x)$, X vyjadřuje náhodnou proměnou z určitého rozdělení a x je libovolné reálné číslo. Pravděpodobnost toho, že náhodná veličina X bude menší než x , je udána distribuční funkcí. [23]

Každá distribuční funkce musí splňovat určité podmínky, aby byla splněna její platnost a to:

- Hodnoty, kterých nabývá, leží mezi 0 a 1.

$$0 \leq F(x) \leq 1 \quad (3.2)$$

- Jedná se o funkci neklesající, takže pro $x_2 > x_1$ bude platit, že:

$$F(x_2) \geq F(x_1) \quad (3.3)$$

- $\lim_{x \rightarrow -\infty} F(x) = 0$ a $\lim_{x \rightarrow \infty} F(x) = 1$ (3.4),(3.5)

- Je spojitá zprava v libovolném bodě $x \in (-\infty; \infty)$

- Počet bodů nespojitosti je konečné číslo

3.3 Diskrétní rozdělení pravděpodobnosti

Jak již bylo řečeno, diskrétní náhodná veličina může nabývat pouze celých čísel z intervalu, který je konečný i nekonečný, ale může se měnit jenom po skocích. Může se například jednat o počet zmetků ve výrobě, čekající zákazník ve frontě nebo počet zásahů do terče.

Pravděpodobnost, s jakou se bude chovat diskrétní náhodná veličina popisujeme pravděpodobnostní funkcí.

$$P(X = x_i) = p_i \quad (3.6)$$

Tato funkce popisuje s jakou pravděpodobností p_i bude náhodná veličina X nabývat hodnoty x_i . Suma všech pravděpodobností musí být rovna jedné a tedy: [21]

$$\sum_i p_i = 1 \quad (3.7)$$

3.3.1 Alternativní rozdělení

Alternativní rozdělení se používá při sledování výskytu náhodné události. Náhodná veličina může nabýt dvou hodnot. Těmi je 0 a 1. Jestliže nastane sledovaný jev, přiřadíme mu číslo 1 a pravděpodobnost výskytu p . V druhém případě jev nenastane, bude mít hodnotu 0 a vznikne s pravděpodobností $1-p$. Pravděpodobnost p je jediným parametrem alternativního rozdělení a zcela ho popisuje [24]

Pravděpodobnostní funkce má tvar:

$$P(x) \begin{cases} 1-p & \text{pro } x = 0 \\ p & \text{pro } x = 1 \\ 0 & \text{ostatní} \end{cases} \quad (3.8)$$

Distribuční funkce:

$$F(x) \begin{cases} 0 & \text{pro } x < 0 \\ 1-p & \text{pro } 0 \leq x < 1 \\ 1 & \text{pro } x \geq 1 \end{cases} \quad (3.9)$$

Dále vztahy pro střední hodnotu a rozptyl:

$$E(X) = p \quad (3.10)$$

$$D(X) = p * (1 - p) \quad (3.11)$$

Využití nalezne vždy, když jsou možné pouze dva výsledky. Příkladem může být vybavenost domácnosti určitým spotřebičem (ne = 0, ano = 1), pohlaví dítěte při narození (dívka = 0, chlapec = 1), shodnost výrob, jestli výrobek splňuje požadavky (shodný = 0, neshodný = 1) a mnoho dalších. Jestliže má veličina X alternativní rozdělení zapisujeme ji jako $X \sim A(p)$.

3.3.2 Binomické rozdělení

Binomické rozdělení určuje četnost výskytu náhodného jevu s pravděpodobností p při n nezávislých pokusech. Pravděpodobnost výskytu určitého jevu není závislá na předchozích pokusech a ani se nemění. Jednotlivé pokusy jsou shodné s alternativním rozdělením, kdy dostáváme jeden ze dvou možných výsledků. Značí se $Bi(n,p)$, kde n je počet pokusů a p je pravděpodobnost výskytu sledovaného jevu.

Pravděpodobnostní funkce binomického rozdělení:

$$P(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{pro } x = 0, 1, \dots, n, \\ 0 & \text{jinak} \end{cases}, \quad (3.12)$$

kde $\binom{n}{x}$ je kombinační číslo, počet výskytu jevu je označen x při n pokusech a pravděpodobnosti p .

Distribuční funkce:

$$F(x) = \sum_{x_i \leq x} P(X = x_i) \quad (3.13)$$

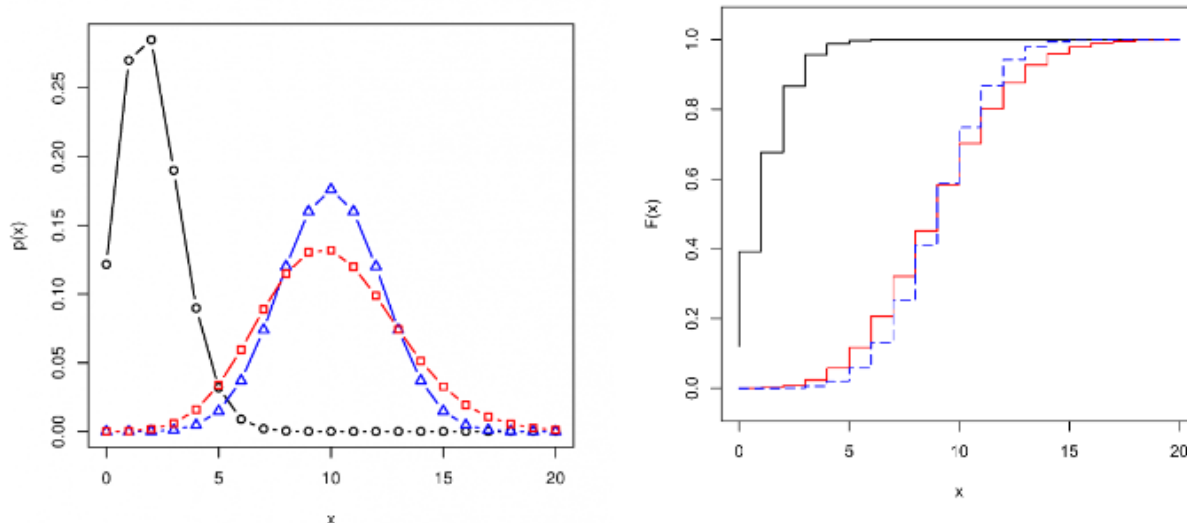
Střední hodnota:

$$E(X) = n * p \quad (3.14)$$

Rozptyl:

$$D(X) = n * p * (1 - p) \quad (3.15)$$

Na obrázku 2 je vidět jak parametry p a x ovlivňují tvar pravděpodobnostní (vlevo) a distribuční (vpravo) funkce. Černá kolečka jsou pro $n = 100$ a $p = 0,1$, červené čtverečky pro $n = 20$ a $p = 0,1$ a modré trojúhelníčky pro $n = 20$ a $p = 0,5$. [25] Je-li pravděpodobnost $0,5$, vykazuje pravděpodobnostní funkce symetrii kolem střední hodnoty. V případě nižší pravděpodobnosti než $0,5$ jsou hodnoty asymetricky rozloženy směrem k levé straně. Je to způsobeno poklesem úspěšnosti daného jevu.



Obrázek 2 Pravděpodobnostní (vlevo) a distribuční (vpravo) funkce binomického rozdělení [25]

3.3.3 Poissonovo rozdělení

Poissonovo rozdělení definujeme jedním parametrem, a to je parametr λ . Toto rozdělení bude mít veličina, u které hledáme počet výskytů za danou časovou jednotku. Události ve sledovaném období nastávají náhodně a jsou na sobě nezávislé. Parametr λ nám udává střední počet událostí, které se v dané časové jednotce vyskytly. Náhodnou veličinu, která nabývá Poissonovo rozdělení označíme $X \sim \text{Po}(\lambda)$.

Jelikož Poissonovo rozdělení udává počet výskytů při velkém počtu nezávislých pokusů, přičemž pravděpodobnost výskytu v jednom pokusu je mizivá, můžeme jím dobře aproximovat binomické rozdělení s parametry $p \leq 0,1$ a $n > 30$. Parametr λ jednoduše vypočteme ze vztahu $\lambda = n \cdot p$. Chyba, jež se touto aproximací dopouštíme, je menší než 10^{-2} , čímž je více než přijatelná vzhledem k uvážení zjednodušení výpočtů.

Pravděpodobnostní funkce:

$$P(x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda} & \text{pro } x = 0, 1, 2, \dots \\ 0 & \text{jinak} \end{cases} \quad (3.16)$$

Distribuční funkce:

$$F(x) = \sum_{x_i \leq x} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \quad (3.16)$$

Střední hodnota a rozptyl:

$$E(X) = D(X) = \lambda \quad (3.17)$$

Poissonovo rozdělení úzce souvisí s teorií hromadné obsluhy. Zde očekáváme, že máme nějakou obsluhu a rádi bychom co nejlépe využili její kapacitu, ale zase chceme zabránit vzniku front. V případě obchodu by to byl ideální počet prodavačů, který zvládne obsloužit přicházející zákazníky. Jelikož je počet poruch nebo počet vadných výrobků také modelován Poissonovým rozdělením, nalézá široké uplatnění i v technické praxi. [26]

3.3.4 Geometrické rozdělení

Popisuje-li náhodná veličina X počet úspěšných pokusů, než nastane první neúspěch, tak se jedná o geometrické rozdělení a značíme ho $X \sim \text{Ge}(p)$. Z označení geometrického rozdělení je patrné, že jediným parametrem, který ho popisuje, je pravděpodobnost úspěchu pokusu p . Pravděpodobnost zůstává po všechny pokusy stejná. Výsledkem pokusu vždy může být jenom úspěch nebo neúspěch. Hledaná pravděpodobnost až do prvního úspěšného pokusu (včetně) při uskutečnění x pokusů můžeme zjistit z pravděpodobnostní funkce: [27]

$$P(x) = \begin{cases} p * (1 - p)^x & \text{pro } x = 0, 1, 2, \dots \\ 0 & \text{jinak} \end{cases} \quad (3.18)$$

Distribuční funkce:

$$F(x) = \sum_{x_i \leq x} (1 - p)^{x_i} \quad (3.19)$$

Střední hodnota:

$$E(X) = \frac{1-p}{p} \quad (3.20)$$

Rozptyl:

$$D(X) = \frac{1-p}{p^2} \quad (3.21)$$

3.4 Spojité rozdělení pravděpodobnosti

Oproti diskrétní náhodné veličině může spojitá náhodná veličina nabývat hodnot všech čísel z intervalu ať už je konečný nebo nekonečný. Spojité hodnoty často dostáváme měřeními, ale můžou to být i životnosti strojů a podobně.

Distribuční funkci pro spojitě náhodné veličiny definujeme jako:

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt \quad (3.22)$$

V tomto integrálu značí $f(t)$ funkci hustoty rozdělení pravděpodobnosti, což je obdoba pravděpodobností funkce u diskrétního náhodného rozdělení. Obsah plochy mezi křivkou a osou x je v libovolném intervalu pravděpodobnost, že náhodná veličina X nabude určité hodnoty v tomto intervalu. Také z toho plyne, že nejsme schopni určit pravděpodobnost jedné dané hodnoty, protože plocha pod křivkou by byla rovna nule, ale vždy musíme volit interval.

Každá hustota pravděpodobnosti má takovou vlastnost, že celá plocha ležící pod křivkou je rovna jedné. Je to obdobné diskrétní náhodné veličině, kdy suma všech pravděpodobností je rovněž jedna. Matematický zápis vypadá následovně:

$$\int_{-\infty}^{\infty} f(x) dx = 1 \quad (3.23)$$

Z hustoty pravděpodobnosti $f(x)$ lze získat vztah pro distribuční funkci $F(x)$ a naopak pokud platí, že existuje derivace distribuční funkce v daném bodě x . Za splnění těchto podmínek bude platit vztah: [28]

$$f(x) = \frac{dF(x)}{dx} \quad (3.24)$$

3.4.1 Rovnoměrné rozdělení

Nejjednodušší spojité rozdělení pravděpodobnosti je rovnoměrné rozdělení. To se vyznačuje konstantní hustotou pravděpodobnosti na intervalu (a, b) , jinými slovy je každá hodnota z tohoto intervalu stejně pravděpodobná. Náhodnou veličinu v tomto rozdělení značíme $X \sim R_s(a, b)$. Hustota pravděpodobnosti má tvar:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{pro } x \in (a, b) \\ 0 & \text{jinak} \end{cases} \quad (3.25)$$

Distribuční funkci dostaneme integrací hustoty pravděpodobnosti a bude platit:

$$F(x) = \begin{cases} 0 & \text{pro } x \leq a \\ \frac{x-a}{b-a} & \text{pro } x \in (a, b) \\ 1 & \text{pro } x \geq b \end{cases} \quad (3.26)$$

Střední hodnota:

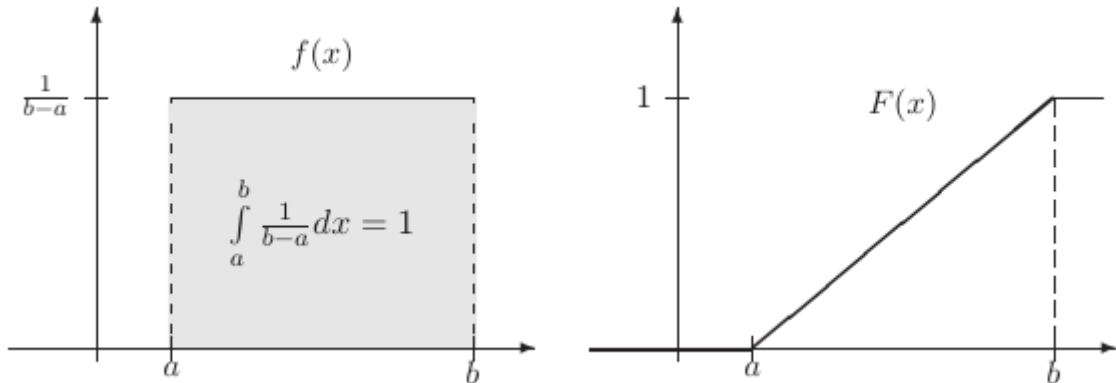
$$E(X) = \frac{a+b}{2} \quad (3.27)$$

Rozptyl:

$$D(X) = \frac{(b-a)^2}{12} \quad (3.28)$$

Rovnoměrné rozdělení nalézá využití v případech, kdy očekáváme, že hodnoty sledované veličiny nastávají v daném intervalu se stejnou pravděpodobností. Toto chování je patrné z grafu funkce pro hustotu pravděpodobnosti (obrázek 3). Třeba při čekání na tramvaj můžeme využít

rovnorné rozdělení a spočítat si pravděpodobnost, že spoj přijede do určité doby. Má však i využití v technické praxi, protože se jím řídí chyby vzniklé při zaokrouhlování a při odečítání z lineárních měřících přístrojů. [29]



Obrázek 3 Hustota pravděpodobností a distribuční funkce rovnoměrného rozdělení

3.4.2 Exponenciální rozdělení

Náhodná veličina $X \sim \text{Ex}(\lambda)$ z exponenciálního rozdělení určuje dobu čekání, než nastane nějaká událost, která může nastat stále se stejnou pravděpodobností nehlédě na již uplynutou dobu. Rozdělení je dáno jedním parametrem λ , $\lambda > 0$ a pak má hustota pravděpodobnosti tvar: [30]

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{pro } x \in (0, \infty) \\ 0 & \text{jinak} \end{cases} \quad (3.29)$$

Distribuční funkce:

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{pro } x \in (0, \infty) \\ 0 & \text{jinak} \end{cases} \quad (3.30)$$

Střední hodnota:

$$E(X) = \frac{1}{\lambda} \quad (3.31)$$

Rozptyl:

$$D(X) = \frac{1}{\lambda^2} \quad (3.32)$$

Exponenciální rozdělení vykazuje určitou podobnost s Poissonovým rozdělením. Obě pracují s výskytem náhodného jevu a uplynulým časem. Rozdíl je v nich ale takový, že zatímco Poissonovo rozdělení pracuje s počtem výskytů dané události v časovém intervalu, exponenciální rozdělení modeluje, jaká doba uběhne, než se daná událost vyskytne. Příkladem může být použití Poissonova rozdělení k modelování počtu poruch na stroji za daný časový interval, zatímco čas od jedné poruchy k druhé budeme modelovat rozdělením exponenciálním. [31]

3.4.3 Weibullovo rozdělení

Až na normální rozdělení bychom těžko našli nějaké další rozdělení, které má tak širokou oblast použití jako to Weibullovo. O jeho rozvoj se hlavně zasloužil švédský stavební inženýr Waloddi Weibull, když v roce 1939 publikoval práci na toto téma a pak na jeho rozvoji soustavně nadále pracoval. Ačkoliv není jeho původním autorem, toto rozdělení je dnes pojmenováno díky jeho přínosu právě po něm. S příspěvky dalších autorů je dnes Weibullovo rozdělení nástrojem pro modelování širokého spektra náhodných veličin.

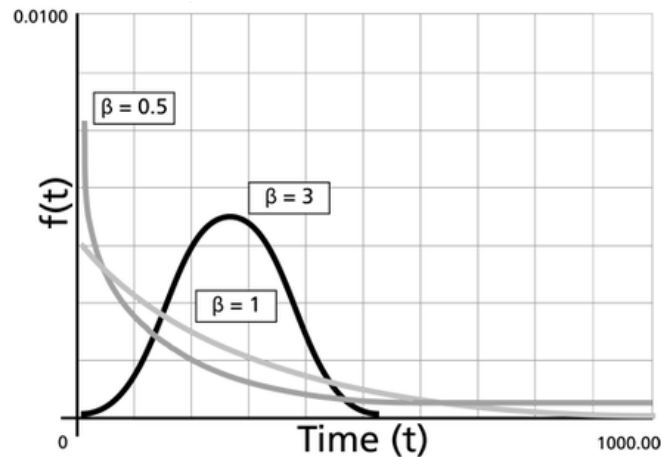
Původně je odvozeno pro tři parametry η , β a γ . Náhodná veličina z tohoto rozdělení bude mít označení $X \sim W(\eta, \beta, \gamma)$ a funkce hustoty pravděpodobnosti bude ve tvaru:

$$f(x) = \begin{cases} \frac{\beta}{\eta} \left(\frac{x-\gamma}{\eta}\right)^{\beta-1} e^{-\left(\frac{x-\gamma}{\eta}\right)^\beta} & \text{pro } x \in (\gamma, \infty) \\ 0 & \text{jinak} \end{cases} \quad (3.33)$$

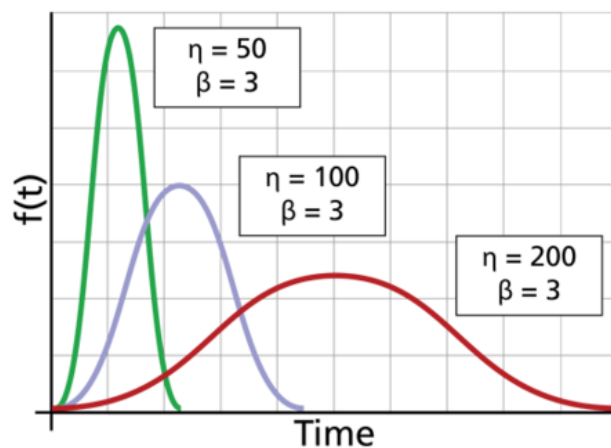
Parametry mají na tvar Weibullova rozdělení zásadní vliv. Označujeme je jako:

- η – parametr měřítka, $\eta > 0$, díky tomu parametru může rozdělení nabývat celou řadu tvarů, z obrázku 5 je patrné, jak markantní tato změna je.

- β – parametr tvaru, $\beta > 0$, změny tohoto tvaru ovlivňují jak bude funkce „roztažená“, s rostoucím parametrem tvaru se funkce roztahuje doprava a klesá její výška, jak je vidět na obrázku 4. [32]
- γ – parametr polohy, ovlivňuje posunutí funkce na ose x



Obrázek 4 Vliv parametru β na tvar Weibullova rozdělení [32]



Obrázek 5 Vliv parametru η na tvar Weibullova rozdělení [32]

Často rozdělení zjednodušíme položením parametru $\gamma = 0$. Vznikne nám tím dvouparametrové Weibullovo rozdělení se značením $W(\eta, \beta)$. To je v praxi používáno nejčastěji a tato práce bude pracovat s ním. Výsledná hustota pravděpodobnosti bude mít tvar:

$$f(x) = \begin{cases} \frac{\beta}{\eta} \left(\frac{x}{\eta}\right)^{\beta-1} e^{-\left(\frac{x}{\eta}\right)^{\beta}} & \text{pro } x \in (0, \infty) \\ 0 & \text{jinak} \end{cases} \quad (3.34)$$

Distribuční funkce:

$$F(x) = \begin{cases} 1 - e^{-\left(\frac{x}{\eta}\right)^\beta} & \text{pro } x \in (0, \infty) \\ 0 & \text{jinak} \end{cases} \quad (3.35)$$

Střední hodnota:

$$E(X) = \eta \Gamma\left(\frac{1}{\beta} + 1\right) \quad (3.36)$$

Rozptyl:

$$D(X) = \eta^2 \left[\Gamma\left(\frac{2}{\beta} + 1\right) - \Gamma^2\left(\frac{1}{\beta} + 1\right) \right] \quad (3.37)$$

Kde $\Gamma(x)$ značí gamma funkci, která je definována pro každé kladné x jako:

$$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt \quad (3.38)$$

Velkou výhodou Weibullova rozdělení je jeho univerzálnost, protože pomocí vhodné volby parametrů dokážeme vyjádřit nebo aproximovat jiná rozdělení. Tato vlastnost se hodí, protože nám stačí rovnice:

- Exponenciální rozdělení – pro $\beta = 1$ se Weibullovo rozdělení zjednoduší na exponenciální
- Rayleighovo rozdělení – pro $\beta = 2$ nám vznikne Rayleighovo rozdělení. To se například využívá pro zjištění životnosti součástí v závislosti na stáří nebo radiolokačních aplikacích.
- Logaritmicko-normální rozdělení – pro $\beta = 2,5$ získáme aproximaci logaritmicko-normálního rozdělení
- Normální rozdělení – pro $\beta = 3,6$ dostaneme přibližnou aproximaci normálního rozdělení [33]

První využití si Weibullovo rozdělení, i když se tehdy tak ještě nejmenovalo, našlo při popisu částic vzniklých drcením a mletím. V dnešní době

se díky jeho variabilitě využívá ve značném počtu odvětví. Může to být strojírenství, elektrotechnika, meteorologie, hydrologii, pojišťovnictví, lékařství a mnoho dalších. Hraje významnou roli při modelování životnosti výrobků, pro které nevyhovuje exponenciální rozdělení, zvláště tam kde se projevuje únava materiálu a opotřebení. [34]

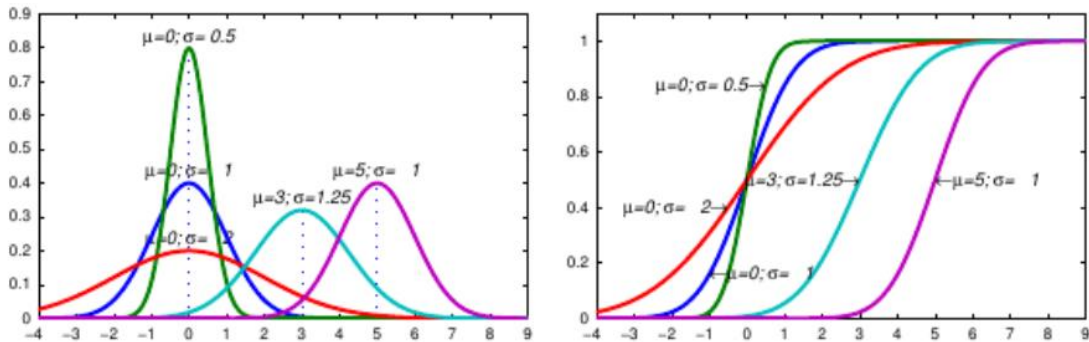
3.5 Normální rozdělení

Normální rozdělení (někdy také nazývané Gaussovo) je považováno za nejdůležitější rozdělení v teorii pravděpodobnosti a statistice. Velké množství náhodných jevů může být popsáno právě tímto rozdělením. Jenom v lidské populaci je nepočítaně příkladů, které mu odpovídají, ať už je to výška, hmotnost, inteligence, reakční časy v psychologických experimentech, rozličné schopnosti a nespočet dalších. Samozřejmě se nevyskytuje jenom u člověka, najde svoje zastoupení i v ekonomii, technice a přírodních vědách. Typickým příkladem v technice je vznik náhodných chyb při měření, které jsou způsobeny nekontrolovatelnými vlivy (změna teploty, vibrace, vlhkost, tlak atd.)

Shlukující se hodnoty kolem středu vytvářejí velmi známý zvonovitý tvar hustoty pravděpodobnosti, jemuž se někdy také říká Gaussova křivka. Tento tvar vzniká jako výsledek kolísání náhodné veličiny, na kterou působí velký počet na sobě nezávislých jevů. [28] Že tomuto rozdělení říkáme normální, neznamená, že by bylo „normálnější“ než rozdělení jiná. Statistci si pouze všimli, že ten samý tvar se vyskytuje znovu a znovu, a z toho vznikl název normální rozdělení. [35]

Normální rozdělení pravděpodobnosti jsme schopni zcela popsat dvěma parametry, které standardně značíme μ a σ^2 . První označuje střední hodnotu a druhé rozptyl. To, že veličina X bude pocházet z normální rozdělení,

zapisujeme $X \sim N(\mu, \sigma^2)$. Obrázek 6 ukazuje, jak se mění graf hustoty pravděpodobnosti a distribuční funkce pro různé parametry.



Obrázek 6 Hustota pravděpodobnosti a distribuční funkce pro různé parametry [36]

Hustota pravděpodobnosti:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} \quad \text{pro } x \in (-\infty, \infty) \quad (3.39)$$

Střední hodnota:

$$E(X) = \mu \quad (3.40)$$

Rozptyl:

$$D(X) = \sigma^2 \quad (3.41)$$

3.5.1 Normované normální rozdělení

Jelikož pro normální rozdělení ve tvaru $N(\mu, \sigma^2)$ nejsme schopni analytickou cestou zjistit hodnoty z distribuční funkce a výpočet musí být proveden numericky vznikla otázka jak tento výpočet zjednodušit. Jedna možnost je pro libovolné parametry hodnoty tabelovat, což by nebylo příliš efektivní. Lepší je využít vlastnosti zachování normality při změně měřítka osy, kam vynášíme hodnoty náhodné veličiny X . Tomuto procesu se říká transformace a převedeme při něm náhodnou veličinu X s rozdělením $N(\mu, \sigma^2)$ na náhodnou veličinu Z s rozdělením $N(0, 1)$. Pro transformaci platí:

$$X \sim N(\mu, \sigma^2) \rightarrow Z = \frac{X - \mu}{\sqrt{\sigma^2}} \sim N(0, 1) \quad (3.42)$$

Tomuto novému rozdělení se říká normované (nebo standardizované) normální rozdělení. Jeho výhodou je, že jsou pro něj tabelovány hodnoty distribuční funkce, které jsou zároveň dostupné ve většině statistických softwarů. Nová hustota pravděpodobnosti má tvar:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}} \quad \text{pro } x \in (-\infty, \infty) \quad (3.43)$$

4 TESTOVÁNÍ STATISTICKÝCH HYPOTÉZ

Provádíme-li statistické testování, nesnažíme se určit pouze velikosti statistických ukazatelů. Občas očekáváme výsledek nějakého procesu a potřebuje si ověřit, jestli se předpokládaný výsledek potvrdí. Proto zavádíme nový termín statistická hypotéza

Statistická hypotéza je jakékoliv tvrzení, které provedeme o rozdělení náhodné veličiny. Například můžeme mít hypotézu, že rozdělení náhodné veličiny, kterou zkoumáme, je z normálního rozdělení. Jindy můžeme předpokládat, že dva náhodné výběry pocházejí ze stejného rozdělení.

Správná formulace hypotéz je základ všech statistických testů. Buď můžeme pracovat s **nulovou hypotézou** H_0 nebo použít **alternativní hypotézu** H_1 . Nulová hypotéza se často vyjadřuje rovností, třeba vyjadřuje tvrzení, že data pocházejí z určitého rozdělení. Alternativní hypotéza je opak nulové hypotézy, třeba tvrzení, že data z určitého rozdělení nepocházejí.

Testování statistických hypotéz není nikdy na 100% spolehlivé. Vždycky existuje malá pravděpodobnost, že odmítneme hypotézu i když byla ve skutečnosti správná a ani jsme se nedopustili chyby v metodice testování. Této chybě se říká **chyba I. druhu** a nastane zamítneme-li testem hypotézu H_0 , která ve skutečnosti platí. Opakem je **chyba II. druhu**, která vznikne přijetím neplatné hypotézy.

Jak velké riziko chyby I. druhu je pro nás přípustné si určíme z takzvané **hladiny významnosti** α . Jedná se pravděpodobnost, že dojde k chybě I. druhu. I když je volba libovolná, obvykle se volí hodnota 5 %. Hladinu významnosti není vhodné volit příliš nízko, protože se tím bude snižovat síla testu $(1-\beta)$, kde β je **pravděpodobnost vzniku chyby II. druhu**. Oba tyto parametry jsou spolu provázány, takže určením jednoho získáme i druhý.

4.1 Jednovýběrový Studentův t-test

Cílem jednovýběrového t-testu je testovat hypotézu, jestli data náhodného výběru pochází z rozdělení se stejnou střední hodnotou.

Testovaná statistika je dána vzorcem:

$$t = \frac{\bar{x} - \mu_0}{s} \sqrt{n}, \quad (4.1)$$

kde μ_0 je teoretická střední hodnota, n je rozsah náhodného výběru, s je odhad rozptylu základního souboru (pro odhad použit výběrový rozptyl), \bar{x} je průměr z testovaného vzorku.

Jelikož statistika t nepracuje s normálním rozdělením, ale (Studentovým) t -rozdělením stačí nám pouze jeden parametr k jeho popsání, tím jsou stupně volnosti ν a pro jejich výpočet platí vztah $\nu = n-1$. To využijeme ke stanovení kritické hodnoty T_k pro příslušný kvartil Studentova rozdělení o $n-1$ stupních volnosti. Tyto hodnoty jsou k nalezení ve statistických tabulkách (tabulka 1).

Tabulka 1 Kvantily Studentova t rozdělení [37]

f	P						
	0,75	0,875	0,95	0,975	0,9875	0,995	0,9975
1	1,00000	2,4142	6,3138	12,706	25,452	63,657	127,32
2	0,81650	1,6036	2,9200	4,3027	6,2053	9,9248	14,089
3	0,76489	1,4226	2,3534	3,1825	4,1765	5,8409	7,4533
4	0,74070	1,3444	2,1318	2,7764	3,4954	4,6041	5,5976
5	0,72669	1,3009	2,0150	2,5706	3,1634	4,0321	4,7733
6	0,71756	1,2733	1,9432	2,4469	2,9687	3,7074	4,4168
7	0,71114	1,2543	1,8946	2,3646	2,8412	3,4995	4,0193
8	0,70639	1,2403	1,8595	2,3060	2,7515	3,3554	3,8325
9	0,70272	1,2297	1,8331	2,2622	2,6850	3,2498	3,6897
10	0,69981	1,2213	1,8125	2,2281	2,6338	3,1693	3,5814
11	0,69745	1,2145	1,7959	2,2010	2,5931	3,1058	3,4966
12	0,69548	1,2089	1,7823	2,1788	2,5600	3,0545	3,4284
13	0,69384	1,2041	1,7709	2,1604	2,5326	3,0123	3,3725
14	0,69242	1,2001	1,7613	2,1448	2,5096	2,9768	3,3257
15	0,69120	1,1967	1,7530	2,1315	2,4899	2,9467	3,2860
16	0,69013	1,1937	1,7459	2,1199	2,4729	2,9208	3,2520
17	0,68919	1,1910	1,7396	2,1098	2,4581	2,8982	3,2225
18	0,68837	1,1887	1,7341	2,1009	2,4450	2,8784	3,1966
19	0,68763	1,1866	1,7291	2,0930	2,4334	2,8609	3,1737
20	0,68696	1,1848	1,7247	2,0860	2,4231	2,8453	3,1534
21	0,68635	1,1831	1,7207	2,0796	2,4138	2,8314	3,1352
22	0,68580	1,1816	1,7171	2,0739	2,4055	2,8188	3,1188
23	0,68531	1,1802	1,7139	2,0687	2,3979	2,8073	3,1040
24	0,68485	1,1789	1,7109	2,0639	2,3910	2,7969	3,0905
25	0,68443	1,1777	1,7081	2,0595	2,3846	2,7874	3,0782
26	0,68405	1,1766	1,7056	2,0555	2,3788	2,7787	3,0669
27	0,68370	1,1757	1,7033	2,0518	2,3734	2,7707	3,0565
28	0,68335	1,1748	1,7011	2,0484	2,3685	2,7633	3,0469
29	0,68204	1,1739	1,6991	2,0452	2,3638	2,7564	3,0480
30	0,68276	1,1731	1,6973	2,0423	2,3596	2,7500	3,0298
40	0,68066	1,1673	1,6839	2,0211	2,3289	2,7045	2,9712
60	0,67862	1,1616	1,6707	2,0003	2,2991	2,6603	2,9146
120	0,67656	1,1559	1,6577	1,9799	2,2699	2,6174	2,8599
∞	0,67449	1,1503	1,6449	1,9600	2,2414	2,5758	2,8070

Při řešení postupuje následovně:

1. Formulujeme hypotézu H_0 , kdy předpokládáme, že střední hodnota základního souboru je stejná se střední hodnotou pokusného výběrového souboru
2. Je vypočteno testovací kritérium t ze vzorce 4.1
3. Pro testovací kritérium se určí počet stupňů volnosti $v = n-1$
4. Zvolíme hladinu významnosti α a s pomocí stupňů volnosti odečteme kritickou hodnotu, kterou zapisujeme $T_k(1-\alpha; n-1)$
5. Nakonec porovnáme t a T_k . Pokud je $t \leq T_k$, tak přijímáme nulovou hypotézu H_0 a můžeme říct, že rozdíl $|\bar{x} - \mu_0|$ je způsoben variabilitou a výsledek je správný, v opačném případě hypotézu zamítáme, protože je výsledek zatížen soustavnou chybou. [38]

4.2 Testování odlehlých hodnot

Někdy se v jinak stejnorodém souboru objeví hodnota, které se od ostatních silně odlišuje. Ta může být způsobena nedodržením metodiky, měřením nebo jiným faktorem, tuto hodnotu pak nazýváme hrubou chybou. Testy odlehlých hodnot nám pak pomáhají rozlišit, jestli se skutečně jedná o hrubou chybu nebo chybu náhodnou přirozeně vzniklou variabilitou procesu. Když jsme si naprosto jistí, že hodnota nepochází ze stejného základního souboru jako ostatní hodnoty výběrového souboru, můžeme ji vyloučit. [39]

4.2.1 Dean-Dixonův test

Dean-Dixonův test je univerzální nástroj, jak ověřit rozdělení na odlehlé hodnoty. Je neparametrický, což znamená, že nemusíme vědět z jakého rozdělení výběr pochází.

Test probíhá následovně:

- 1) Zjištěné hodnoty seřadíme podle velikosti od nejmenší po největší hodnotu a tím nám vznikne posloupnost $x_1 < x_2 \dots < x_n$
- 2) Stanovení nulové hypotézy, že krajní hodnoty nejsou odlehlé.
 - H_0 : hodnota x_1 není odlehlá.
 - H_0 : hodnota x_n není odlehlá.
- 3) Vypočteme testové kritérium Q pro dolní odlehlou hodnotu:

$$Q_{min} = \frac{x_2 - x_1}{x_n - x_1} = \frac{x_2 - x_1}{R} \quad (4.2)$$

Podobně postupujeme s výpočtem horní odlehlé hodnoty:

$$Q_{max} = \frac{x_n - x_{n-1}}{R} \quad (4.3)$$

R označuje variační rozpětí souboru.

- 4) Ze statistických tabulek odečteme (tabulka 2) kritickou hodnotu Q_k pro počet hodnot n, které testujeme a požadovanou hladinu významnosti α .

5) Hodnoty Q_{\min} a Q_{\max} porovnáme s kritickou hodnotou Dean-Dixonova rozdělení $Q_k(\alpha; n)$.

- Nulovou hypotézu přijímáme v případě $Q_{\min} \leq Q_k$ nebo $Q_{\max} \leq Q_k$ a hodnota tím pádem není odlehlá.
- Nastane-li případ $Q_{\min} \geq Q_k$ nebo $Q_{\max} \geq Q_k$, nulovou hypotézu zamítneme. Testovanou hodnotu považujeme za odlehlou a vyloučíme jí ze souboru dat. [40]

Tabulka 2 Kritické hodnoty pro Dean-Dixonův test [37]

n	α	
	0,05	0,01
3	0,941	0,988
4	0,765	0,889
5	0,642	0,780
6	0,560	0,698
7	0,507	0,637
8	0,468	0,590
9	0,437	0,555
10	0,412	0,527
11	0,392	0,502
12	0,376	0,482
13	0,361	0,465
14	0,349	0,450
15	0,338	0,438
16	0,329	0,426
17	0,320	0,416
18	0,313	0,407
19	0,306	0,398
20	0,300	0,391
21	0,295	0,384
22	0,290	0,378
23	0,285	0,372
24	0,281	0,367
25	0,277	0,362
26	0,273	0,357
27	0,269	0,353
28	0,266	0,349
29	0,263	0,345
30	0,260	0,341

4.2.2 Grubbsův test

Víme-li, že testovaná náhodná veličina pochází z normálního rozdělení je vhodné použít Grubbsův test místo Dean-Dixonova testu. Grubbsův test je parametrický a dává nám spolehlivější výsledky. Míru odlehlosti testujeme jako vzdálenost od aritmetického průměru podělenou směrodatnou odchylkou. Postup se velmi podobá Dean-Dixonově testu a vypadá následovně:

- 1) Zjištěné výsledky seřadíme podle velikosti od nejmenší po největší hodnotu a tím nám vznikne posloupnost $x_1 < x_2 \dots < x_n$.
- 2) Stanovíme nulové hypotézy, že krajní hodnoty nejsou odlehlé.
 - H_0 : hodnota x_1 není odlehlá.
 - H_0 : hodnota x_n není odlehlá.
- 3) Vypočteme testové kritérium T pro dolní odlehlou hodnotu:

$$T_{min} = \frac{\bar{x} - x_1}{S_n} \quad (4.4)$$

Podobně postupujeme s výpočtem horní odlehlé hodnoty:

$$T_{max} = \frac{x_n - \bar{x}}{S_n} \quad (4.5)$$

S_n značí směrodatnou odchylku. Je důležité jí nezaměnit za výběrovou směrodatnou odchylku, které se často používá v rozdělení pravděpodobnosti.

- 4) Ze statistických tabulek odečteme (tabulka 3) kritickou hodnotu T_k pro počet hodnot n které testujeme a požadovanou hladinu významnosti α .
- 5) Hodnoty T_{min} a T_{max} porovnáme s kritickou hodnotou Grubbsova rozdělení $T_k(\alpha; n)$.
 - Nulovou hypotézu přijímáme v případě $T_{min} \leq T_k$ nebo $T_{max} \leq T_k$ a hodnota tím pádem není odlehlá.
 - Nastane-li případ $T_{min} \geq T_k$ nebo $T_{max} \geq T_k$, nulovou hypotézu zamítneme. Testovanou hodnotu považujeme za odlehlou a vyloučíme jí ze souboru dat. [37]

Tabulka 3 Kritické hodnoty pro Grubbsův test [37]

$n \backslash \alpha$	0,01	0,05	0,10
3	1,414	1,412	1,406
4	1,723	1,689	1,645
5	1,955	1,869	1,791
6	2,130	1,996	1,894
7	2,265	2,093	1,974
8	2,374	2,172	2,041
9	2,464	2,237	2,097
10	2,540	2,294	2,146
11	2,606	2,343	2,190
12	2,663	2,387	2,229
13	2,714	2,426	2,264
14	2,759	2,461	2,297
15	2,800	2,493	2,326
16	2,837	2,523	2,354
17	2,871	2,551	2,380
18	2,903	2,577	2,404
19	2,932	2,600	2,426
20	2,959	2,623	2,447
21	2,984	2,644	2,467
22	3,008	2,664	2,486
23	3,030	2,683	2,504
24	3,051	2,701	2,520
25	3,071	2,717	2,537

5 TESTY NORMALITY

Statistika využívá celou řadu metod ke sběru, analýze, zpracování a interpretaci dat. Ovšem ty nejsou univerzálně využitelné. Jedno ze základních dělení metod je na neparametrické a parametrické. **Neparametrické metody** nevyžadují znalost rozdělení pravděpodobnosti, ze kterého pochází náhodná veličina. Na druhou stranu jsou zde **metody parametrické**, které tuto znalost vyžadují. Může vzniknout otázka, proč se tedy nezaobírat pouze metodami neparametrickými? Odpovědí je, že použití parametrických metod je většinou přesnější a také jednodušší, takže se nám vyplatí vynaložit další úsilí a snažit se zjistit z jakého rozdělení náhodná veličina pochází.

Rozhodneme-li se tedy využít parametrických metod, určení správného rozdělení pravděpodobnosti je nesmírně důležité. Když budeme předpokládat, že veličina má určité rozdělení, ale ve skutečnosti tento předpoklad není splněn, získané výsledky budou zavádějící a můžeme dojít k chybným závěrům. Proto se tato kapitola věnuje testům, jak určit, zda náhodná veličina pochází z normálního rozdělení. Testy mohou být početní i grafické. [41]

5.1 Grafické metody

Grafické testy nám umožňují vizuálně porovnat tvar našeho rozdělení s tím teoretickým normálním. Je z nich však téměř nemožné určit statisticky významný závěr, proto je vhodné provést i testy početní. Na druhou stranu jsou na nich dobře vidět poruchy tvaru nebo odlehlé hodnoty.

5.1.1 Histogram

Histogram graficky znázorňuje četnosti s použitím sloupcového grafu. Každý sloupec vyjadřuje šířku zvoleného intervalu. Výška sloupce nám zobrazuje četnost sledované veličiny daného souboru ve svém intervalu. Historicky byl hojně používán kvůli snadné konstrukci bez výpočetní techniky, ale ani dnes neztrácí na oblibě. To je zapříčiněno snadnou čitelností a pochopitelností pro osoby, které nemají statistické vzdělání.

Klíčové pro konstrukci histogramu je správné určení počtu intervalů. Pochybení v tomto kroku může zapříčinit ztrátu informačního obsahu histogramu. Přesto však neexistuje žádný pevně daný postup pro jejich určení. Pro posouzení normality také budeme konstruovat histogram s pevně danou šířkou intervalu, ačkoliv existují i histogramy s nekonstantní šířkou. [42] Často používaný způsob určení počtu intervalů je Sturgesovo pravidlo:

$$k = 1 + 3,3 * \log(n) \quad (5.1)$$

kdy k je počet intervalů a n značí počet pozorování.

Jedna z možných konstrukcí histogramu může probíhat dle postupu:

- 1) Určení minima x_{min} a maxima x_{max} souboru a z nich vypočtení variačního rozpětí R :

$$R = x_{max} - x_{min} \quad (5.2)$$

- 2) Pomocí Sturgesova pravidla vypočítat počet intervalů k a výsledek zaokrouhlit na celé číslo.
- 3) Vypočítat šířku intervalu h :

$$h = \frac{R}{k} \quad (5.3)$$

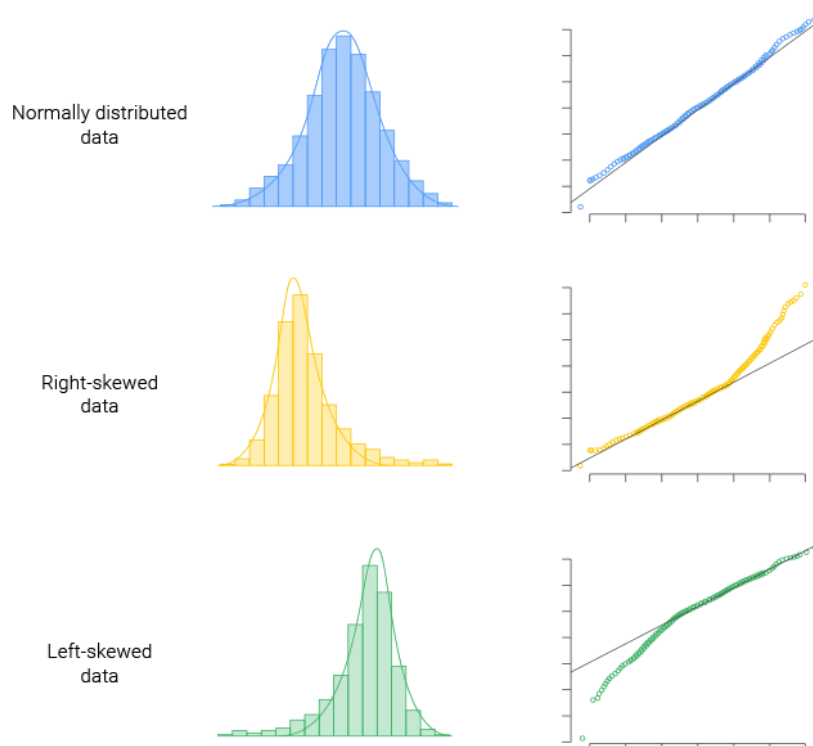
- 4) Určení dolní hranice intervalu x_o , tak aby obsahoval hodnotu x_{min} . Následně kreslíme jednotlivé sloupce, kdy jejich výška je určena četností změřené veličiny v daném intervalu. S vykreslováním končíme při dosažení počtu intervalů k

Snažíme-li se histogramem odhadnout normalitu, tak by data měla být rozdělena přibližně symetricky a kumulovat se kolem střední hodnoty. Celková podoba by měla připomínat zvonovitý tvar, jaký má normální rozdělení. Dále je vidět, jestli jsou obsažené odlehlé hodnoty. Nevýhodou histogramu je, že při malém počtu pozorování dostaneme rovněž malý počet intervalů a jakákoliv pozorování mohou být problematická.

5.1.2 Q-Q graf

Q-Q graf (kvantil-kvantil, z anglického quantile-quantile) je založený na porovnání naměřených kvantilů vůči kvantilům z teoretického normálního rozdělení. Předpoklad je, že pokud bude rozdělení, které testujeme normální budou výsledné body ležet zhruba na přímce. Odhad této přímky je možné do grafu přímo vložit pro lepší čitelnost. Když budou body vzdáleny od této přímky, tak to značí, že data nepocházejí z normálního rozdělení a je vhodné zvážit, že se jedná o jiné rozdělení. [42]

Obrázek 7 zobrazuje tři rozdělení, přičemž pouze u modrého by šlo předpokládat, že pochází z normálního rozdělení pravděpodobnosti. U žlutého a zeleného se v Q-Q grafu vykreslila asymetričnost, přičemž žluté je kladně zešikmeno a zelené záporně. [43]



Obrázek 7 Vliv tvaru rozdělení na Q-Q graf [43]

5.1.3 P-P graf

Jeho název je zkratkou z anglického probability-probability a znamená pravděpodobnost-pravděpodobnost. Jeho konstrukce je z velké části podobná jako v grafu Q-Q, ale místo kvantilů se vynáší na osy výsledného grafu pravděpodobnost. Z toho vyplývá, že obě osy budou mít rozsah mezi 0 a 1. Také se bude lépe konstruovat referenční úsečka s krajními body [0;0] a [1;1], která nebude muset být odhadována. Odhady normality se provádí stejným způsobem jako v Q-Q grafu, tedy pokud budou body ležet na jedné přímce, můžeme rozdělení dat považovat za normální. [44]

5.2 Ověření normality výpočtem

Jak již bylo řečeno, grafické metody ověření normality mohou být neprůkazné, takže je vhodné ještě provést početní ověření.

5.2.1 Kolmogorovův-Smirnovův test

Kolmogorovův-Smirnovův test je metoda, kterou používáme k testování shody dvou jednorozměrných náhodných veličin. Existují dvě varianty tohoto testu, těmi jsou jednovýběrový a dvouvýběrový test.

Jednovýběrovým testem ověřujeme, jakým způsobem se liší rozdělení náhodné veličiny od určitého teoretického rozdělení. Autorem tohoto testu je Andrej Nikolajevič Kolmogorov a přišel s ním v roce 1933. O pár let později přišel Vladimir Ivanovič Smirnov s dvouvýběrovým testem. Ten nepracuje s tím, že jedno rozdělení pravděpodobnosti známe. Dvouvýběrový test srovnává rozdělení dvou náhodných veličin a hledá, jestli jsou ze stejného rozdělení. Nedokáže však určit o jaká rozdělení se jedná.

Pro potřeby testů normality je zapotřebí využít jednovýběrového testu. Rozdělení pravděpodobnosti, se kterým budeme porovnávat náhodnou veličinu je normální rozdělení. Také je nutné znát distribuční funkce rozdělení jak hledaného, tak rozdělení normálního. Jeden z možných postupů řešení Kolmogorova-Smirnovova testu může vypadat následovně:

1. Uvažujeme nulovou hypotézu H_0 , že výběry jsou ze stejného rozdělení.

2. Seřadíme hodnoty údajů ve výběru od nejmenší po největší, tedy $x_1 \leq x_2 \leq \dots \leq x_n$.

3. Určíme hodnoty výběrové distribuční funkce ve všech bodech dle vztahu:

$$F_E(i) = \frac{i}{n+1}, \quad (5.4)$$

kde i je pořadí prvku a n je počet prvků výběru.

4. Dále určíme hodnoty teoretické distribuční funkce v bodech x_1 až x_n

5. Pro každý bod se zjistí absolutní hodnota rozdílu mezi teoretickou distribuční funkcí a výběrovou funkcí tedy:

$$|F_T(x) - F_E(x)| \quad (5.5)$$

6. Testovací kritérium $D(n)$ je největší rozdíl mezi teoretickou a výběrovou distribuční funkcí:

$$D(n) = \max |F_T(x) - F_E(x)| \quad (5.6)$$

7. Stanovíme požadovanou hladinu významnosti a pro odpovídající velikost souboru vybereme ve statistických tabulkách kritickou hodnotu t_{krit} (tabulka 4)

8. Porovnáme testovací kritérium s kritickou hodnotou a jestli platí rovnost $D(n) < t_{krit}$, tak přijímáme nulovou hypotézu, což znamená, že testované rozdělení je normální. [45]

Tabulka 4 Kritických hodnot Kolmogorova-Smirnova testu

$1-\alpha/2$ n	0,975	0,995	$1-\alpha/2$ n	0,975	0,995
1	0,975	0,995	21	0,287	0,344
2	0,842	0,929	22	0,281	0,337
3	0,708	0,829	23	0,275	0,330
4	0,624	0,734	24	0,269	0,333
5	0,563	0,669	25	0,264	0,317
6	0,519	0,617	26	0,259	0,311
7	0,483	0,576	27	0,254	0,305
8	0,454	0,542	28	0,250	0,300
9	0,430	0,513	29	0,246	0,295
10	0,409	0,489	30	0,242	0,290
11	0,391	0,468	31	0,238	0,285
12	0,375	0,449	32	0,234	0,281
13	0,361	0,432	33	0,231	0,277
14	0,349	0,418	34	0,227	0,273
15	0,338	0,404	35	0,224	0,269
16	0,327	0,392	36	0,221	0,265
17	0,318	0,381	37	0,218	0,262
18	0,309	0,371	38	0,215	0,258
19	0,301	0,361	39	0,213	0,255
20	0,294	0,352	40	0,210	0,252
			nad 40	$\frac{1,36}{\sqrt{n}}$	$\frac{1,63}{\sqrt{n}}$

6 ZPRACOVÁNÍ ZÍSKANÝCH DAT

Tato část se bude zaměřovat na uplatnění znalostí popsaných v teoretické části. K tomu byly získány tři soubory dat. Ty budou nejdříve otestovány na normalitu, jak graficky, tak početně a výsledky budou mezi sebou porovnány. Na posledním souboru dat budou provedeny ještě další statistické testy.

6.1 Získání dat

Jak již bylo zmíněno získali jsme tři soubory dat, které budeme nazývat X, Y a Z. Soubory X a Y byly vzhledem k většímu množství vygenerovány ve statistickém softwaru Minitab 16. Soubor neboli náhodná veličina X je generována dle normálního rozdělení $N(\mu, \sigma^2)$. K vytvoření je potřeba zadat parametry μ a σ^2 . Střední hodnota byla zvolena $\mu = 105$ a směrodatná odchylka $\sigma^2 = 15$. Velikost souboru $n = 150$. Vznikne veličina, kterou budeme zapisovat $X \sim N(105, 15)$.

Druhá náhodně generovaná veličina byla veličiny Y. Ta pochází z Weibullova rozdělení. Při určitých parametrech tímto rozdělením můžeme aproximovat normální rozdělení. Přesněji Weibullovo rozdělení s parametrem tvaru $\beta = 3,6$ se bude nejvíce blížit normálnímu rozdělení. Tento parametr však není znám a my normalitu musíme otestovat bez jeho znalosti. Víme tedy, že veličina Y pochází z dvouparametrového Weibullova rozdělení o neznámých parametrech $Y \sim W(\eta, \beta)$. Velikost souboru je taktéž $n = 150$.

Třetí soubor Z byl získán měřením, jehož průběh je popsán v následující podkapitole.

6.1.1 Průběh měření

Poslední skupina dat Z byla získána měřením čepů. Byly měřeny sériově vyráběné čepy s průměrem $7,9 \pm 0,04$ mm (obrázek 8). Změřeno bylo prvních 25 čepů, protože je předpoklad použití Grubbsova testu, který nabízí kritické hodnoty jenom do tohoto čísla.



Obrázek 8 Deska s měřnými čepy

Jako měřidlo byl zvolen digitální třmenový mikrometr Micromar 40 EWV, který je zobrazen na obrázku 9. Pro každý čep byly provedeny 3 měření, ze kterých byl vypočten průměr, aby se zamezilo náhodné chybě, protože se dále budeme zabývat variabilitou výrobního procesu a jestli je veličina Z odpovídá normálnímu rozdělení. Výsledky měření jsou zapsané v tabulce 5.

Tabulka 5 Naměřené hodnoty průměrů čepů

Číslo měření	1	2	3	Průměr [mm]
1	7,938	7,938	7,937	7,938
2	7,927	7,925	7,926	7,926
3	7,922	7,924	7,925	7,924
4	7,888	7,884	7,886	7,886
5	7,88	7,882	7,884	7,882
6	7,881	7,878	7,88	7,88
7	7,889	7,891	7,897	7,892
8	8,004	7,999	7,997	8
9	7,936	7,932	7,93	7,933
10	7,904	7,901	7,902	7,902
11	7,934	7,931	7,933	7,933
12	7,929	7,933	7,933	7,932
13	7,997	8,005	8,005	8,002
14	7,928	7,926	7,927	7,927
15	7,886	7,886	7,885	7,886
16	7,966	7,967	7,967	7,967
17	7,895	7,893	7,898	7,895
18	7,918	7,916	7,915	7,916
19	7,893	7,894	7,896	7,894
20	7,984	7,985	7,985	7,985
21	7,935	7,935	7,938	7,936
22	7,886	7,886	7,887	7,886
23	7,885	7,887	7,887	7,886
24	7,924	7,922	7,924	7,923
25	7,937	7,936	7,938	7,937



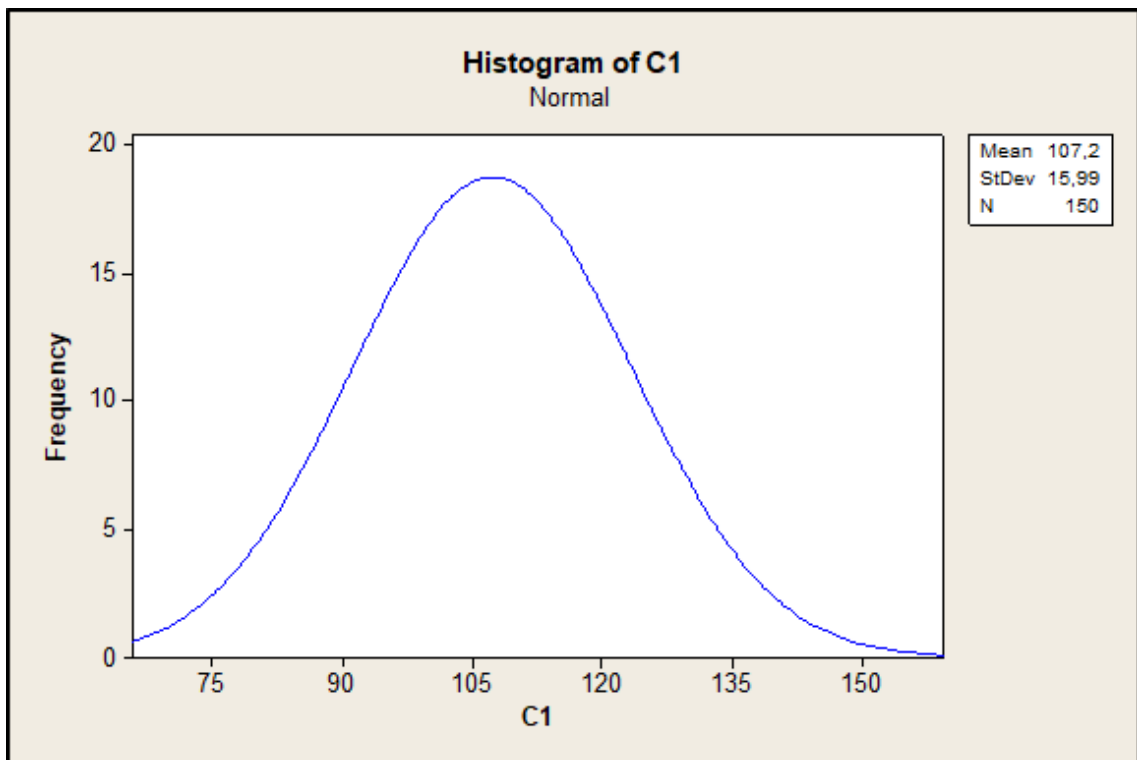
Obrázek 9 Mikrometr Mahr upevněný ve stojanu

6.2 Testy normality veličiny X

První je na normalitu testována veličina X, Nejdříve je vyvozen závěr z grafických metod a pak následně potvrzen výpočtem.

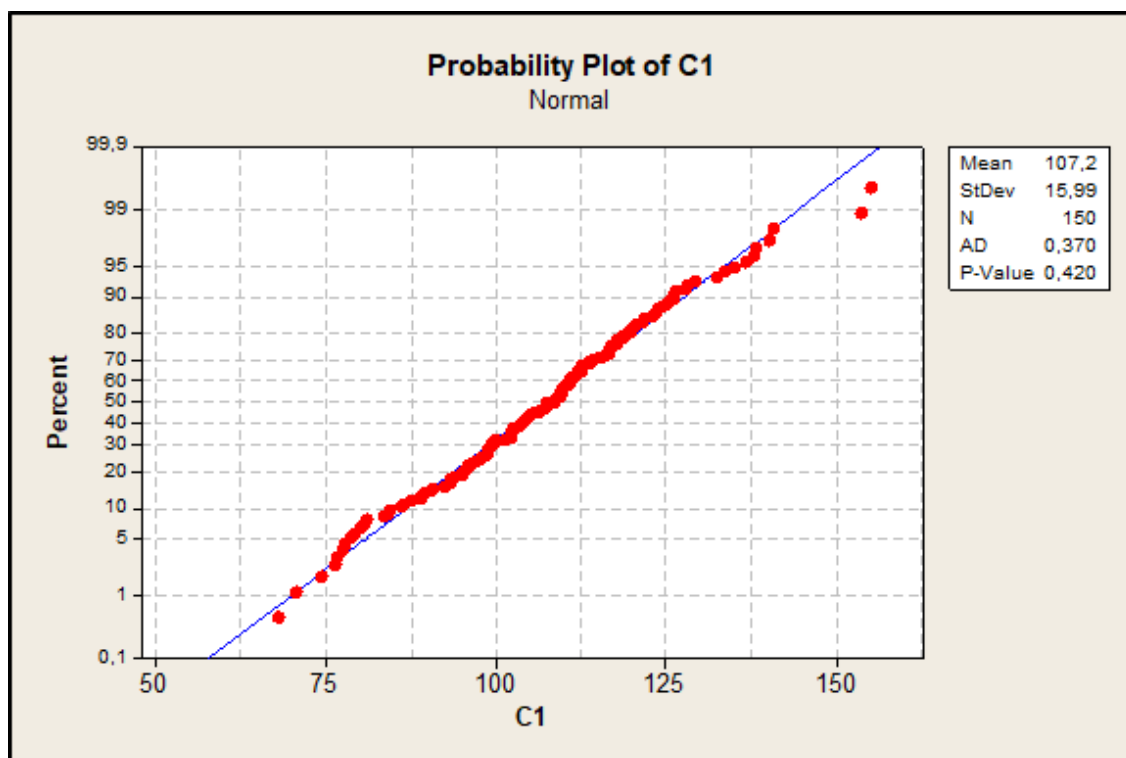
6.2.1 Graficky

Podíváme-li se na **histogram** (obrázek 10) vytvořený v softwaru Minitab, je na první pohled zřejmé, že připomíná zvonovitou křivku normálního rozdělení. Jediná anomálie, které je si možná všimnou, je „useknutí“ v levé části grafu. Z tohoto zobrazení nepoznáme, proč nastalo, bude tedy nutné se podívat, jestli z dalších grafický metod zjistíme více.



Obrázek 10 Histogram veličiny X

Další z grafických metod jsou Q-Q a P-P grafy. Na obrázku 11 je zobrazen P-P graf veličiny X. Ve střední části referenční přímky jednotlivé body blízce kopírují její průběh. Směrem ke krajům rozdělení hodnoty lehce kolísají, ale ne tolik, že by byl důvod domnívat se, že rozdělení není normální. Jediná větší odchylka jsou dva rozptýlené body vpravo nahoře.



Obrázek 11 P-P graf veličiny X

Grafické metody ukázaly, že se s největší pravděpodobností jedná o normální rozdělení. Rozdělení téměř v celé délce kopíruje přímku toho teoretického. Pouze dva body pravého okraje tvoří určitou anomálii rozdělení, kvůli svému rozptýlení. Tento jev by mohl vysvětlovat proč je střed rozdělení $\mu = 107,2$ i rozptyl $\sigma^2 = 15,99$ větší než původní rozdělení $N(105, 15)$, ze kterého byly data generovány.

6.2.2 Početně

I když není důvod domnívat se, že se nejedná o normální rozdělení, vzhledem k výsledku grafických metod, přesto nyní bude výsledek zkontrolován poččetně.

K počtetnímu testu normality je využit Kolmogorovův-Smirnovův test. V tabulce 6 je prvních deset hodnot z náhodně generovaných dat. Zbytek souboru je v příloze A.

Tabulka 6 Výčet prvních 10 hodnot veličiny X

Pořadí prvku (i)	Veličina X	F _E (x)	F _T (x)	F _T (x)-F _E (x)
1	67,99	0,0066	0,0069	0,0003
2	70,616	0,0132	0,0108	0,0024
3	74,381	0,0199	0,0197	0,0002
4	76,33	0,0265	0,0263	0,0002
5	76,733	0,0331	0,0279	0,0052
6	77,702	0,0397	0,0320	0,0077
7	77,732	0,0464	0,0321	0,0142
8	78,797	0,0530	0,0372	0,0157
9	79,422	0,0596	0,0405	0,0191
10	80,099	0,0662	0,0444	0,0218

Kolmogorovův-Smirnovův test probíhal podle postupu rozebraného v teoretické části a to následovně:

1. Pracujeme s nulovou hypotézou H_0 , že se jedná o normální rozdělení, a to také použijeme jako rozdělení referenční.
2. Při přenosu dat z Minitabu do Excelu jsou data řazena podle velikosti. To je patrné z tabulky 6, kde je zobrazených deset nejmenších hodnot.
3. Určení hodnot výběrové funkce dle zadaného vzorce, první člen bude vypadat následovně:

$$F_E(i) = \frac{i}{n+1} \rightarrow F_E(1) = \frac{1}{150+1} = 0,0066 \quad (6.1)$$

4. Hodnoty teoretické distribuční $F_T(x)$ funkce se zjistí pomocí funkce NORM.DIST zabudované v Excelu. Do funkce NORM.DIST se zadávají tři argumenty, je to střední hodnota výběru, směrodatná odchylka a hodnota veličiny X, pro kterou chceme zjistit hodnotu rozdělení. Střední hodnota rozdělení $\mu = 107,2$ a směrodatná odchylka $\sigma = 15,99$.
5. Pro každý člen se zjistí rozdíl mezi teoretickou a výběrovou funkcí:

$$|F_T(x) - F_E(x)| \quad (6.2)$$

6. Najdeme největší rozdíl mezi rozdíly mezi teoretickou a výběrovou funkcí $D(n) = \max |F_T(x) - F_E(x)|$ (6.3) a dostáváme $D(150) = 0,0403$

Největší rozdíl je hledán pomocí excelové funkce MAX.

7. Hladinu významnosti α volíme standardně používanou, tedy $\alpha = 0,05$. Pro naši hladinu významnosti a velikost souboru $n = 150$ odečteme z tabulky 4 kritickou hodnotu t_{krit} . Pro naše parametry je zadán následující vzorec výpočtu kritické hodnoty:

$$t_{krit} = \frac{1,36}{\sqrt{n}} = \frac{1,36}{\sqrt{150}} = 0,1110 \quad (6.4)$$

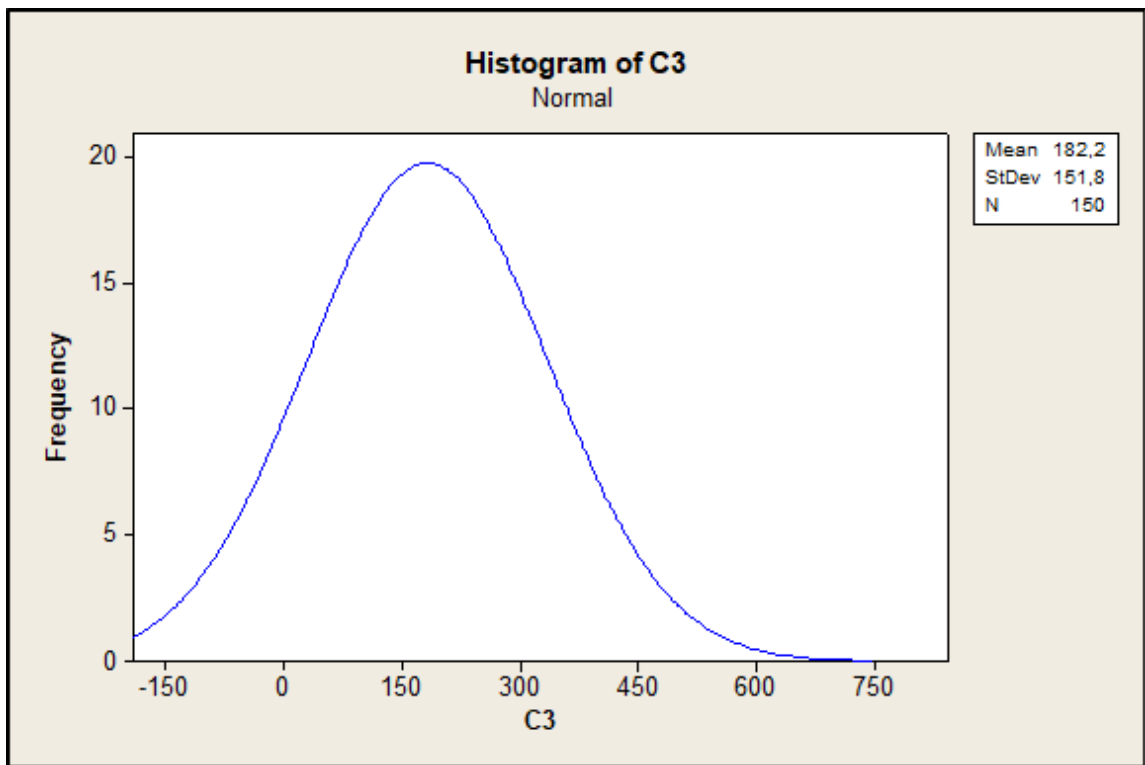
8. Porovnáním zjišťujeme, jestli platí $D(n) < t_{krit}$ (6.4). Protože $D(150) = 0,0403 < t_{krit} = 0,1110$, přijímáme nulovou hypotézu a říkáme, že rozdělení veličiny X pochází z normálního rozdělení.

6.3 Testy normality veličiny Y

Na rozdíl od veličiny X , zde nutně nepředpokládáme, že rozdělení bude dostatečně dobře aproximovat to normálové.

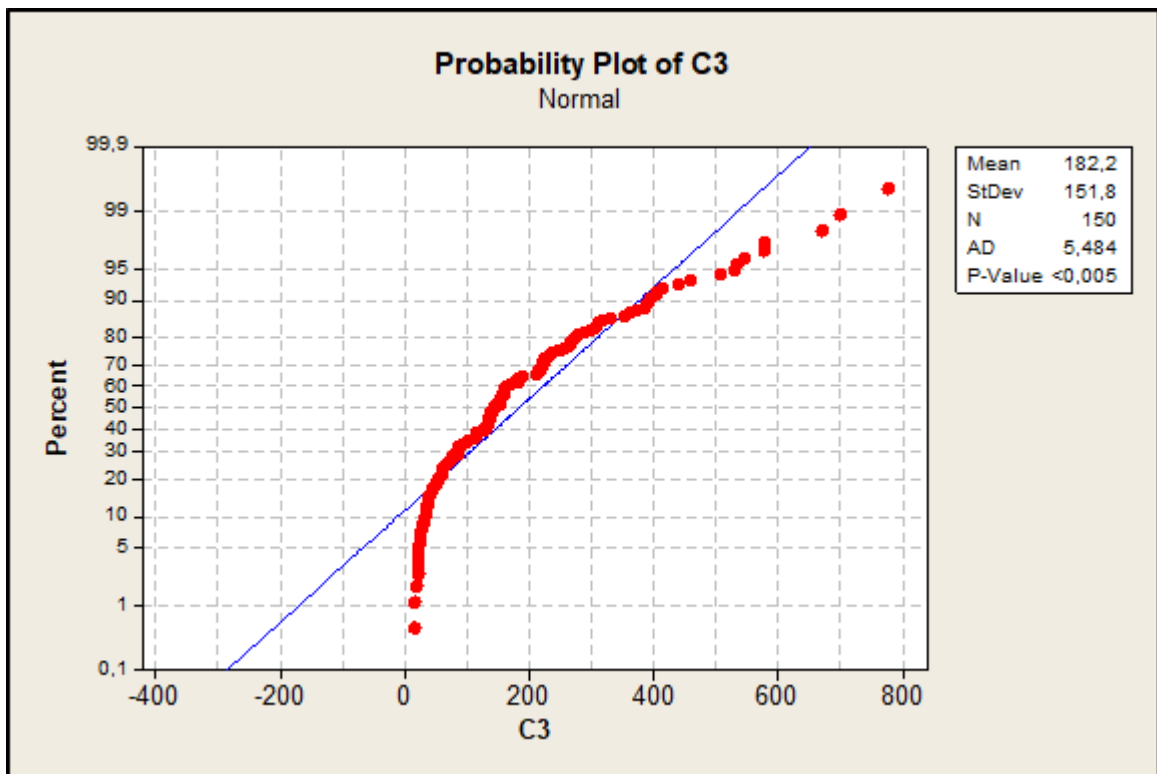
6.3.1 Graficky

Grafické podklady pro veličinu Y byly rovněž vytvořeny ve statistickém softwaru Minitab. Na obrázku 12 je zobrazen histogram. Ten připomíná zvonovitou křivku, ale je v něm vidět lehká nesymetričnost podle střední hodnoty, zvláště porovnáme-li ho s histogramem veličiny X .



Obrázek 12 Histogram veličiny Y

Nesymetričnost histogramu, kterou by šlo lehce přehlédnout se plně projeví zobrazením v P-P grafu (obrázek 13). Je zde vidět, jak hodnoty málo kopírují referenční přímku. Při takto velké odchylce je vysoce pravděpodobné, že veličina Y není z normálního rozdělení.



Obrázek 13 P-Pgraf veličiny Y

Ačkoliv by bylo z histogramu možné usuzovat, že veličina Y je dostatečnou aproximací normálního rozdělení, P-P graf této hypotéze nevyhovuje. Je zde vidět, že člověk, zvláště bez větších zkušeností, nemusí být schopný přesně pohledem určit, jak dobře odpovídá tvar histogramu normálnímu rozdělení. Také se zde projevuje, jak může být histogram nespolehlivý ukazatel normality.

6.3.2 Početně

Grafické metody naznačují, že se nejedná o normální rozdělení, tato hypotéza bude ověřena ještě početně. V tabulce 7 je pro ilustraci uvedeno prvních deset naměřených hodnot a jejich zpracování a zbytek tabulky je v příloze B.

Tabulka 7 Výčet prvních 10 hodnot veličiny Y

Pořadí prvku	Veličina Y	F _E (x)	F _T (x)	F _T (x)-F _E (x)
1	15,146	0,0066	0,1347	0,1281
2	15,988	0,0132	0,1359	0,1227
3	19,248	0,0199	0,1407	0,1208
4	21,023	0,0265	0,1433	0,1168
5	21,432	0,0331	0,1439	0,1108
6	22,246	0,0397	0,1451	0,1054
7	22,286	0,0464	0,1452	0,0988
8	22,74	0,0530	0,1459	0,0929
9	25,159	0,0596	0,1496	0,0900
10	25,31	0,0662	0,1498	0,0836

Stejně jako u veličiny X je i zde pro výpočet normality použit Kolmogorův-Smirnovův test a výpočet probíhá následovně:

1. Pracujeme s nulovou hypotézou H_0 , že se jedná o normální rozdělení, a to také použijeme jako rozdělení referenční.
2. Při přenosu dat z Minitabu do Excelu jsou data řazena podle velikosti. To je patrné z tabulky 7, kde je zobrazených deset nejmenších hodnot.
3. Určení hodnot výběrové funkce dle zadaného vzorce, první člen bude vypadat následovně:

$$F_E(i) = \frac{i}{n+1} \rightarrow F_E(1) = \frac{1}{150+1} = 0,0066 \quad (6.5)$$

4. Hodnoty teoretické distribuční $F_T(x)$ funkce se zjistí pomocí funkce NORM.DIST zabudované v Excelu. Do funkce NORM.DIST se zadávají tři argumenty, je to střední hodnota výběru, směrodatná odchylka a hodnota veličiny X, pro kterou chceme zjistit hodnotu rozdělení. Střední hodnota rozdělení $\mu = 182,2$ a směrodatná odchylka $\sigma = 151,8$.
5. Pro každý člen se zjistí rozdíl mezi teoretickou a výběrovou funkcí:

$$|F_T(x) - F_E(x)| \quad (6.6)$$

6. Najdeme největší rozdíl mezi rozdíl mezi teoretickou a výběrovou funkcí $D(n) = \max |F_T(x) - F_E(x)|$ (6.7) a dostáváme $D(150) = 0,1517$

Největší rozdíl je hledán pomocí excelové funkce Max.

7. Hladinu významnosti α volíme standardně používanou, tedy $\alpha = 0,05$. Pro naši hladinu významnosti a velikost souboru $n = 150$ odečteme z tabulky 4 kritickou hodnotu t_{krit} . Pro naše parametry je zadán následující vzorec výpočtu kritické hodnoty:

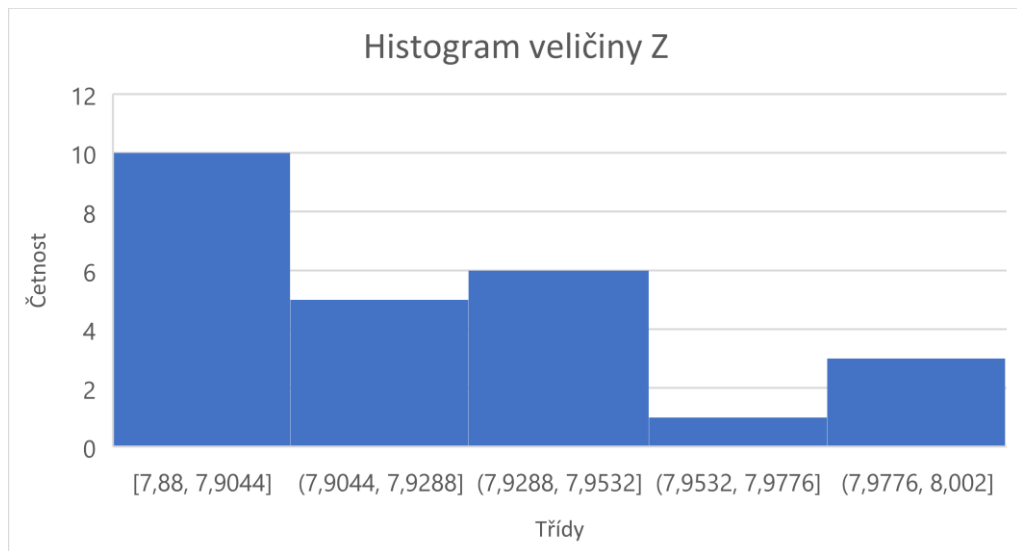
$$t_{krit} = \frac{1,36}{\sqrt{n}} = \frac{1,36}{\sqrt{150}} = 0,1110 \quad (6.7)$$

8. Porovnáním zjišťujeme, jestli platí $D(n) < t_{krit}$ (6.8). Protože $D(150) = 0,1517 > t_{krit} = 0,1110$, nulovou hypotézu zamítáme a říkáme, že veličina nepochází z normálního rozdělení.

6.4 Testy normality veličiny Z

Na rozdíl od veličin X a Y, budeme na veličině Z provádět ještě další statistické testy. Některé statistické testy jsou parametrické, což znamená, že musíme znát z jakého rozdělení náhodná veličina pochází. Například Grubbsův test předpokládá normální rozdělení, a proto jej musíme u veličiny Z ověřit.

Naměřený soubor má 25 hodnot. Pokud bychom chtěli testovat normalitu graficky na souboru této velikosti, výsledky budou s největší pravděpodobností neprůkazné. To je vidět i z histogramu veličiny Z (obrázek 14), který na první pohled nepřipomíná normální rozdělení. Co ale můžeme vidět, jsou možné odlehle hodnoty na pravé straně, na které bude soubor otestován.



Obrázek 14 Histogram veličiny Z

Abychom si byly jistí rozdělením, bude proveden Kolmogorovův-Smirnovův test normality výpočtem stejným způsobem jako u předchozích veličin:

1. Uvažujme nultou hypotézu H_0 , že veličina pochází z normálního rozdělení
2. Po změření a zapsání jsou data seřazena podle velikosti, jak je patrné z tabulky 8.
3. Určení hodnot výběrové funkce dle zadaného vzorce, první člen bude vypadat následovně:

$$F_E(i) = \frac{i}{n+1} \rightarrow F_E(1) = \frac{1}{25+1} = 0,0385 \quad (6.9)$$

4. Hodnoty teoretické distribuční $F_T(x)$ funkce se zjistí pomocí funkce NORM.DIST zabudované v Excelu. Do funkce NORM.DIST se zadávají tři argumenty, je to střední hodnota výběru, směrodatná odchylka a hodnota veličiny X , pro kterou chceme zjistit hodnotu rozdělení. Střední hodnota rozdělení $\mu = 7,923$ a směrodatná odchylka $\sigma = 0,0359$.

5. Pro každý člen se zjistí rozdíl mezi teoretickou a výběrovou funkcí:

$$|F_T(x) - F_E(x)| \quad (6.10)$$

6. Najdeme největší rozdíl mezi rozdíl mezi teoretickou a výběrovou funkcí $D(n) = \max |F_T(x) - F_E(x)|$ (6.11) a dostáváme $D(25) = 0,1429$

Největší rozdíl je hledán pomocí funkce MAX.

7. Hladinu významnosti α volíme standardně používanou, tedy $\alpha = 0,05$. Pro naši hladinu významnosti a velikost souboru $n = 25$ odečteme z tabulky 4 kritickou hodnotu t_{krit} . Pro naše parametry je hodnota $t_{krit} = 0,264$
8. Porovnáním zjišťujeme, jestli platí $D(n) < t_{krit}$ (6.12) . Protože $D(25) = 0,1429 < t_{krit} = 0,264$, přijímáme nulovou hypotézu a říkáme, že rozdělení veličiny Z pochází z normálního rozdělení.

Tabulka 8 Hodnoty veličiny Z

Pořadí prvku	F_E(x)	Veličina Z	F_T(x)	F_T(x)-F_E(X)
1	0,0385	7,8800	0,1170	0,0770
2	0,0769	7,8820	0,1283	0,0514
3	0,1154	7,8860	0,1532	0,0378
4	0,1538	7,8860	0,1532	0,0007
5	0,1923	7,8860	0,1532	0,0391
6	0,2308	7,8860	0,1532	0,0776
7	0,2692	7,8920	0,1961	0,0732
8	0,3077	7,8940	0,2119	0,0958
9	0,3462	7,8950	0,2200	0,1261
10	0,3846	7,9020	0,2819	0,1027
11	0,4231	7,9160	0,4258	0,0027
12	0,4615	7,9230	0,5031	0,0416
13	0,5000	7,9240	0,5142	0,0142
14	0,5385	7,9260	0,5364	0,0021
15	0,5769	7,9270	0,5474	0,0295
16	0,6154	7,9320	0,6020	0,0134
17	0,6538	7,9330	0,6127	0,0412
18	0,6923	7,9330	0,6127	0,0796
19	0,7308	7,9360	0,6443	0,0865
20	0,7692	7,9370	0,6546	0,1146
21	0,8077	7,9380	0,6648	0,1429
22	0,8462	7,9670	0,8913	0,0451
23	0,8846	7,9850	0,9586	0,0740
24	0,9231	8,0000	0,9843	0,0612
25	0,9615	8,0020	0,9864	0,0249

6.5 Další testy veličiny Z

Jelikož nyní víme, že veličina Z má normální rozdělení pravděpodobnosti, můžeme jí testovat pomocí parametrických testů pro toto rozdělení.

6.5.1 Grubbsův test

Z histogramu (obrázek 14) bylo patrné pár rozptýlených hodnot na straně výběrů. Bude proto proveden Grubbsův test, který rozhodne, jestli vznikly variabilitou procesu nebo hrubou chybou. Jestli vznikly hrubou chybou, můžeme

je ze souboru vyloučit. Test je proveden podle postupu popsaném v teoretické části:

1. Zjištěné výsledky seřadíme podle velikosti od nejmenší po největší hodnotu a tím nám vznikne posloupnost $x_1 < x_2 \dots < x_{25}$, tento krok jsme již provedli při testu normality a seřazený soubor najdeme tam (tabulka 8).
2. Stanovíme nulové hypotézy, že krajní hodnoty nejsou odlehlé.
 - H_0 : hodnota x_1 není odlehlá.
 - H_0 : hodnota x_n není odlehlá.
3. Vypočteme testové kritérium T pro dolní odlehlou hodnotu:

$$T_{min} = \frac{\bar{x} - x_1}{S} = \frac{7,923 - 7,880}{0,0352} = 1,221 \quad (6.13)$$

Podobně postupujeme s výpočtem horní odlehlé hodnoty:

$$T_{max} = \frac{x_{25} - \bar{x}}{S} = \frac{8,002 - 7,923}{0,0352} = 2,244 \quad (6.13)$$

S značí směrodatnou odchylku. Je důležité jí nezaměnit za výběrovou směrodatnou odchylku, které se často používá v rozdělení pravděpodobnosti.

4. Ze statistických tabulek odečteme (tabulka 3) odečteme kritickou hodnotu T_k pro počet hodnot $n = 25$, které testujeme a požadovanou hladinu významnosti $\alpha = 0,05$, která je běžně používána. Pro tyto parametry dostaneme kritickou hodnotu $T_k = 2,717$
5. Hodnoty T_{min} a T_{max} porovnáme s kritickou hodnotou Grubbsova rozdělení $T_k(\alpha; n)$.
 - Nulovou hypotézu přijímáme, protože $T_{min} = 1,221 \leq T_k = 2,717$ (6.14) a zároveň $T_{max} = 2,244 \leq T_k = 2,717$ (6.15) a hodnota tím pádem není odlehlá.

Nulová hypotéza nebyla zamítnuta, což znamená, že žádná z hodnot ve výběrovém souboru není odlehlá a nemůžeme jí vyloučit. Taky z toho můžeme usuzovat, že jsme se nedopustili hrubé chyby při měření.

6.5.2 Studentův t-test

Zabýváme-li se sériovou výrobou, v našem případě výrobou čepů, tak vždy počítáme s určitou zmetkovitostí. Je přirozené, že se jí snažíme držet na co nejnižší úrovni, i vyvarovat se jí zcela nemůžeme. Jak jsme předpokládali a následně ověřily výpočtem, naše změřená data průměrů čepů odpovídají normálnímu rozdělení, přičemž četnosti výskytu jevů se shlukují u střední hodnoty a směrem od ní klesají. Nejpřesnější výroby tedy dosáhneme, když nastavíme střední hodnotu výsledného rozdělení na střed tolerančního pole. K ověření, že tomu tak skutečně je, použijeme Studentův t-test. Budeme postupovat následovně:

1. Pro náš měřený čep $\varnothing 7,90 \pm 0,04$ určíme střed tolerančního pole což je 7,90, tuto hodnotu považujeme za střední hodnotu μ_0 statistického souboru a určíme nulovou hypotézu H_0 , že výběrová střední hodnota se jí bude blížit.

Vypočteme testovací kritérium:

$$t = \frac{\bar{x} - \mu_0}{s} \sqrt{n} = \frac{7,923 - 7,90}{0,03589} \sqrt{25} = 1,4329 \quad (6.16)$$

2. Určíme počet stupňů volnosti $\nu = n - 1 = 25 - 1 = 24$ (6.17)
3. Hladinu významnosti α volíme 5 % a pro stupně volnosti odečteme z tabulky 1 kritickou hodnotu $T_k = 2,7969$
4. Vyhodnocením $t = 1,4329 < T_k = 2,7969$ (6.18) přijímáme nulovou hypotézu a můžeme říct, že rozdíl mezi průměrem \bar{x} a směrodatnou odchylkou

Díky tomuto testu můžeme říci, že rozdíl mezi průměrem \bar{x} a směrodatnou odchylkou μ_0 je způsoben variabilitou procesu a nemusíme do něj zasahovat.

ZÁVĚR

Ze začátku práce byly shrnuty nutné poznatky z oblasti metrologie a statistiky, se kterými se dále pracovalo. Pak byla rozebrána problematika rozdělení pravděpodobnosti. V této kapitole byla věnována hlavní pozornost normálnímu rozdělení, protože nám umožňuje popsat nespočet jevů kolem nás. Dále bylo zdůrazněno Weibullovo rozdělení, se kterým se dále pracuje v praktické části.

Cílem práce bylo vhodně popsat metody, jakými mohu testovat normalitu. Testování normality je jedním ze základních kamenů statistického zpracování dat. Jak se jednotlivé metody chovají bylo vyzkoušeno v praktické části. K tomuto účelu byly náhodně vygenerovány dva soubory dat, přičemž jeden byl generován dle normálního rozdělení a druhý podle rozdělení Weibullova. Weibullovo rozdělení bylo použito proto, protože jeho pomocí mohu při určitých parametrech normální rozdělení aproximovat.

Na datech byly prováděny testy normality. Nejdříve grafickými metodami, kdy byly vytvořeny určité předpoklady o datech, které byly potvrzeny výpočtem. První test byl pokus odečtu dat normality z histogramu. Oba histogramy měly zvonovitý tvar, avšak histogram vytvořený z Weibullova rozdělení vykazoval při delším zkoumání nepatrnou asymetričnost, která by byla lehce přehlédnuta. Dále byl proveden odečet z P-P grafu, kdy se data vygenerovaná z Weibullova rozdělení odchylovala od referenční přímky, což signalizuje odchylky od tvaru teoretického normálního rozdělení. To, že data nejsou z normálního rozdělení, bylo na zadané hladině významnosti prokázáno až početně pomocí Kolmogorovova-Smirnovova testu. Data generována z normálního rozdělení prošla nepřekvapivě všemi testy v pořádku.

V tomto pokusu byla ukázána citlivost různých testů normality. Kdyby byl použit jenom histogram k hrubému odhadu normality, snadno by se stalo, že by data z Weibullova rozdělení prošla jako normální a závěry z jejich dalšího zpracování by byly chybné. Odečet P-P grafu již ukázal na jisté odchylky, ale výsledné potvrzení muselo být provedené výpočtem.

Zároveň byla testována jedna sada dat skutečně naměřených. Jednalo se o čepy ze sériové výroby. Z histogramu se zdálo, že data nejsou z normálního rozdělení, což odporuje teoretickým předpokladům. Byla proto ještě otestována pomocí Kolmogorovova-Smirnovova testu, kde se ukázalo, že odpovídají normálnímu rozdělení a nedopustil jsem se tedy hrubé chyby při měření. Zároveň bylo pomocí Grubbsova testu zjištěno, že soubor neobsahuje odlehlé hodnoty a pomocí Studentova t-testu, že je výroba vystředěna na střed tolerančního pole.

CITOVANÁ LITERATURA

- [1] PETŘKOVSKÁ, Lenka a Lenka ČEPOVÁ. *Strojírenská metrologie: studijní opora "Strojírenská metrologie"*. Ostrava: Vysoká škola báňská - Technická univerzita Ostrava, 2011. ISBN ISBN978-80-248-2723-0.
- [2] POSPÍŠIL, Miroslav a Vladimír LUDVÍK. *Terminologie z oblasti metrologie*. Praha: Bořivoj Kleník, PhDr. - Q-art, Praha, 2008.
- [3] *Bible: překlad 21. století: překlad 21. století*. Praha: Biblion, 2009. ISBN 978-808-7282-014.
- [4] TESARČÍK, Bohumil. *Fyzika a klasická energetika: rubriky* [online]. 2017 [cit. 2023-11-19]. Dostupné z: www.3pol.cz/cz/rubriky/fyzika-a-klasicka-energetika/1964-z-historie-mereni
- [5] *Metrický systém byl u nás zaveden před 140 lety* [online]. 2011 [cit. 2023-11-16]. Dostupné z: <https://elektrika.cz/data/clanky/metricky-system-byl-u-nas-zaveden-pred-140-lety>
- [6] PETŘKOVSKÁ, Lenka a Lenka ČEPOVÁ. *Strojírenská metrologie: studijní opora "Strojírenská metrologie"*. Ostrava: Vysoká škola báňská - Technická univerzita Ostrava, 2011. ISBN ISBN978-80-248-2723-0.
- [7] RADA, Václav. *Teorie měření a regulace* [online]. 160 [cit. 2023-11-16]. Dostupné z: www.fce.vutbr.cz/tst/rada.v/TMaR/t-mar-2015-p2p-tm2-CHYB1.pdf
- [8] ČESKÝ STATISTICKÝ ÚŘAD. *Studovat matematiku a statistiku se vyplatí* [online]. aktualizováno 20.12.2014 [cit. 2024-11-23]. Dostupné z: <https://www.czso.cz/csu/czso/0b0050b02f>
- [9] *Acta Universitatis Palackianae Olomucensis*. Olomouc: Univerzita Palackého, 2003. ISSN 1214-3251.

- [10] HOMOLA, Vladimír. *Úvod do statistiky* [online]. 2002, revize 11/2014 [cit. 2023-11-19]. Dostupné z: <https://homel.vsb.cz/~hom50/SLBSTATS/UST/GS02.HTM>
- [11] DE LEEUW, Jan. Statistical Software: An Overview. In: LOVRIC, Miodrag, ed., Miodrag LOVRIC. *International Encyclopedia of Statistical Science* [online]. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, s. 1470-1473 [cit. 2024-01-02]. ISBN 978-3-642-04897-5. Dostupné z: doi:10.1007/978-3-642-04898-2_553
- [12] The evolution of statistical computing: a captivating journey through history. In: *Vsni.co.uk* [online]. 2021 [cit. 2023-11-28]. Dostupné z: <https://vsni.co.uk/blogs/evolution-of-statistical-computing>
- [13] TENSORFLOW. Probability. In: *Tensorflow.com* [online]. 2023 [cit. 2023-12-02]. Dostupné z: <https://www.tensorflow.org/probability>
- [14] GOLDWATER, Eva. Excel. In: *People.umass.edu* [online]. 2011 [cit. 2023-11-04]. Dostupné z: <https://people.umass.edu/evagold/excel.html>
- [15] Excel functions (alphabetical). In: *Support.microsoft.com/* [online]. 2023 [cit. 2023-12-16]. Dostupné z: <https://support.microsoft.com/en-us/office/excel-functions-alphabetical-b3944572-255d-4efb-bb96-c6d90033e188>
- [16] Minitab. In: *Minitab.com* [online]. 2023 [cit. 2023-12-16]. Dostupné z: <https://www.minitab.com/en-us/products/minitab/>
- [17] Introduction to MiniTab: Data analysis, stats & process improvement tools. In: *Simplilearn.com* [online]. 2022 [cit. 2023-12-16]. Dostupné z: https://www.simplilearn.com/what-is-minitab-data-analysis-tool-article#minitab_in_six_sigma
- [18] *Getting Started with Minitab Statistical Software* [online]. 2023 [cit. 2023-12-16]. Dostupné z:

www.minitab.com/content/dam/www/en/uploadedfiles/documents/getting-started/MinitabGettingStarted_EN.pdf

- [19] Co je R?. In: *R-project.cz* [online]. [cit. 2023-12-16]. Dostupné z: <https://www.r-project.cz/about.html>
- [20] *Náhodné jevy a jejich pravděpodobnost* [online]. 2020 [cit. 2023-12-17]. Dostupné z: http://fse.ujep.cz/~moc/studmat/pdf/pstat/Stat_2020_05.pdf
- [21] Náhodná veličina a její popis. In: *Physics.ujep.cz* [online]. [cit. 2023-12-19]. Dostupné z: https://physics.ujep.cz/~mmaly/vyuka/poc_fyz_1/zdroje/NahodnaVelicina_zaklady.pdf
- [22] Náhodná veličina. In: *Cit.vfu.cz* [online]. 2015 [cit. 2024-01-02]. Dostupné z: <https://cit.vfu.cz/statpotr/POTR/Teorie/Predn1/nahvelic.htm>
- [23] GLÜCKSMANNOVÁ, Radka. *PRAVDĚPODOBNOSTNÍ MODELY KOLEM NÁS* [online]. České Budějovice, 2007 [cit. 2024-01-02]. Dostupné z: <http://home.ef.jcu.cz/~mrkvicka/vyuka/Pmkn.pdf>. Bakalářská práce. Jihočeská univerzita v Českých Budějovicích. Vedoucí práce RNDr. Tomáš Mrkvicka, Ph.D.
- [24] Alternativní rozdělení. In: *Is.muni.cz* [online]. 2023 [cit. 2024-01-02]. Dostupné z: <https://is.muni.cz/do/rect/el/estud/prif/ps15/statistika/web/pages/alternativni-rozd.html>
- [25] Binomické a Poissonovo rozdělení pravděpodobnosti. In: *Portal.matematickabiologie.cz* [online]. 2020 [cit. 2023-12-21]. Dostupné z: <https://portal.matematickabiologie.cz/index.php?pg=analiza-a-modelovani-dynamicky-biologicky-dat--vybrane-kapitoly-z-matematickeho-modelovani--zakon-malych-cisel--binomicke-a-poissonovo-rozdeleni-pravdepodobnosti>

- [26] Poissonovo rozdělení. In: *Simulace.info* [online]. 2013 [cit. 2023-12-20]. Dostupné z: https://www.simulace.info/index.php/Poisson_distribution/cs
- [27] LITSCHMANNOVÁ, Martina. *Vybrané kapitoly z pravděpodobnosti* [online]. Ostrava: VŠB – TU Ostrava, Fakulta elektrotechniky a informatiky, 2011 [cit. 2023-12-21]. Dostupné z: https://mi21.vsb.cz/sites/mi21.vsb.cz/files/unit/vybrane_kapitoly_pravdepodobnost.pdf
- [28] DEVORE, Jay a Kenneth BERK. *Modern Mathematical Statistics with Applications* [online]. Second Edition. Springer, 2018 [cit. 2024-01-02]. ISBN 978-1-4614-0390-6. Dostupné z: doi:10.1007/978-1-4614-0391-3
- [29] Rovnoměrné rozdělení náhodné veličiny. In: *Books.fs.vsb.cz* [online]. 2013 [cit. 2023-12-22]. Dostupné z: <http://books.fs.vsb.cz/SystAnal/texty/09.htm>
- [30] Exponenciální rozdělení. In: *Is.muni.cz* [online]. 2020 [cit. 2023-12-22]. Dostupné z: <https://is.muni.cz/do/rect/el/estud/prif/ps15/statistika/web/pages/exponencialni.html>
- [31] ŘAZANKOVÁ, Hana, Luboš MAREK a Michal VRABEC. *INTERAKTIVNÍ UČEBNICE STATISTIKY* [online]. 2006 [cit. 2024-01-02]. Dostupné z: <https://iastat.vse.cz/>
- [32] Weibull Distribution Characteristics. In: *Reliawiki.com* [online]. 2023 [cit. 2023-12-22]. Dostupné z: https://www.reliawiki.com/index.php/Weibull_Distribution_Characteristics
- [33] RINNE, Horst. *The Weibull Distribution A Handbook*. Giessen: CRC Press, 2008. ISBN 978-1-4200-8744-4.
- [34] Weibullovo rozdělení. In: *Is.muni.cz* [online]. 2023 [cit. 2023-12-22]. Dostupné z:

https://is.muni.cz/do/rect/el/estud/prif/ps15/statistika/web/pages/weibul_lovo.html

- [35] Normal distributions review. In: *Khanacademy.org* [online]. 2016 [cit. 2023-12-22]. Dostupné z: <https://www.khanacademy.org/math/statistics-probability/modeling-distributions-of-data/normal-distributions-library/a/normal-distributions-review>
- [36] Náhodné veličiny a náhodné vektory. In: *Is.muni.cz* [online]. 2017 [cit. 2024-01-01]. Dostupné z: https://is.muni.cz/do/rect/el/estud/prif/ps13/prav_stat/web_1/pages/03_06-priklady-spoj-rozdel.html
- [37] KOŽÍŠEK, Jan. *Statistické tabulky a jejich použití*. Vyd. 3. Praha: České vysoké učení technické, 1996. ISBN 80-010-1469-X.
- [38] Parametrické testy - Studentův t-test. In: *Cit.vfu.cz* [online]. 2010 [cit. 2023-12-26]. Dostupné z: <https://cit.vfu.cz/statpotr/POTR/Teorie/Predn3/ttest.htm>
- [39] Vylučování extrémních hodnot souboru. In: *Cit.vfu.cz* [online]. 2015 [cit. 2023-12-27]. Dostupné z: <https://cit.vfu.cz/statpotr/POTR/Teorie/Predn2/extremy.htm>
- [40] Základy zpracování geologických dat. In: *Is.muni.cz* [online]. 2011 [cit. 2023-12-27]. Dostupné z: https://is.muni.cz/el/1431/podzim2011/G3101/um/prednaska_6_finalni.pdf?lang=cs
- [41] Parametrické a neparametrické odhady. In: *Portal.matematickabiologie.cz* [online]. 2020 [cit. 2023-12-28]. Dostupné z: <https://portal.matematickabiologie.cz/index.php?pg=aplikovana-analyza-klinickyh-a-biologickyh-dat--aplikovana-analyza-preziti--neparametricke-odhady--parametricke-a-neparametricke-odhady>

- [42] KUPKA, Karel. *Statistické řízení jakosti: interaktivní analýza a interpretace dat pro řízení jakosti a ekonomiku*. Pardubice: TriloByte, 1997. ISBN 80-238-1818-X.
- [43] Quantile-Quantile (QQ) Plot – Base Graph. In: *Learnbyexample.org* [online]. 2020 [cit. 2023-12-29]. Dostupné z: <https://www.learnbyexample.org/r-quantile-quantile-qq-plot-base-graph/>
- [44] THODE, Henry C. *Testing for normality*. Marcell Dekker, 2002. ISBN 0-8247-9613-6.
- [45] HEBÁK, Petr. *Statistické myšlení a nástroje analýzy dat*. Praha: Informatorium, 2013. ISBN 978-80-7333-105-4.

SEZNAM OBRÁZKŮ

Obrázek 1 Prostředí softwaru Minitab [18]	21
Obrázek 2 Pravděpodobnostní (vlevo) a distribuční (vpravo) funkce binomického rozdělení [25]	28
Obrázek 3 Hustota pravděpodobností a distribuční funkce rovnoměrného rozdělení	32
Obrázek 4 Vliv parametru β na tvar Weibullova rozdělení [32]	34
Obrázek 5 Vliv parametru η na tvar Weibullova rozdělení [32]	34
Obrázek 6 Hustota pravděpodobnosti a distribuční funkce pro různé parametry [36]	37
Obrázek 7 Vliv tvaru rozdělení na Q-Q graf [43]	48
Obrázek 8 Deska s měřenými čepy	53
Obrázek 9 Mikrometr Mahr upevněný ve stojanu	55
Obrázek 10 Histogram veličiny X.....	56
Obrázek 11 P-P graf veličiny X.....	57
Obrázek 12 Histogram veličiny Y.....	60
Obrázek 13 P-Pgraf veličiny Y.....	61
Obrázek 14 Histogram veličiny Z.....	64

SEZNAM TABULEK

Tabulka 1 Kvantily Studentova t rozdělení [37]	41
Tabulka 2 Kritické hodnoty pro Dean-Dixonův test [37]	43
Tabulka 3 Kritické hodnoty pro Grubbsův test [37]	45
Tabulka 4 Kritických hodnot Kolmogorova-Smirnovova testu	51
Tabulka 5 Naměřené hodnoty průměrů čepů.....	54
Tabulka 6 Výčet prvních 10 hodnot veličiny X.....	58
Tabulka 7 Výčet prvních 10 hodnot veličiny Y	62
Tabulka 8 Hodnoty veličiny Z	66

PŘÍLOHY

Příloha A – Hodnoty veličiny X

Pořadí prvku	Veličina X	F_E(x)	F_T(x)	F_T(x)-F_E(X)
1	67,99	0,0066	0,0069	0,0003
2	70,616	0,0132	0,0108	0,0024
3	74,381	0,0199	0,0197	0,0002
4	76,33	0,0265	0,0263	0,0002
5	76,733	0,0331	0,0279	0,0052
6	77,702	0,0397	0,0320	0,0077
7	77,732	0,0464	0,0321	0,0142
8	78,797	0,0530	0,0372	0,0157
9	79,422	0,0596	0,0405	0,0191
10	80,099	0,0662	0,0444	0,0218
11	80,875	0,0728	0,0491	0,0237
12	80,909	0,0795	0,0494	0,0301
13	83,482	0,0861	0,0682	0,0179
14	84,286	0,0927	0,0751	0,0177
15	84,402	0,0993	0,0761	0,0232
16	86,222	0,1060	0,0938	0,0121
17	86,444	0,1126	0,0962	0,0164
18	87,515	0,1192	0,1081	0,0111
19	88,962	0,1258	0,1260	0,0001
20	89,026	0,1325	0,1268	0,0057
21	89,644	0,1391	0,1350	0,0040
22	90,575	0,1457	0,1481	0,0024
23	90,672	0,1523	0,1496	0,0028
24	92,51	0,1589	0,1780	0,0191
25	93,242	0,1656	0,1902	0,0247
26	93,253	0,1722	0,1904	0,0182
27	93,464	0,1788	0,1940	0,0152
28	93,986	0,1854	0,2032	0,0177
29	95,072	0,1921	0,2230	0,0309
30	95,147	0,1987	0,2244	0,0257
31	95,517	0,2053	0,2314	0,0261
32	95,571	0,2119	0,2324	0,0205
33	95,884	0,2185	0,2385	0,0199
34	95,94	0,2252	0,2396	0,0144
35	96,548	0,2318	0,2516	0,0198
36	97,029	0,2384	0,2613	0,0229
37	97,7	0,2450	0,2752	0,0301
38	97,777	0,2517	0,2768	0,0251
39	98,006	0,2583	0,2816	0,0233

40	98,591	0,2649	0,2941	0,0292
41	98,732	0,2715	0,2972	0,0257
42	98,824	0,2781	0,2992	0,0210
43	98,926	0,2848	0,3014	0,0166
44	98,979	0,2914	0,3026	0,0112
45	99,187	0,2980	0,3071	0,0091
46	99,349	0,3046	0,3107	0,0061
47	99,391	0,3113	0,3117	0,0004
48	99,845	0,3179	0,3218	0,0039
49	100,063	0,3245	0,3267	0,0022
50	101,412	0,3311	0,3578	0,0267
51	102,147	0,3377	0,3752	0,0374
52	102,284	0,3444	0,3784	0,0341
53	102,307	0,3510	0,3790	0,0280
54	102,31	0,3576	0,3791	0,0214
55	102,372	0,3642	0,3805	0,0163
56	102,546	0,3709	0,3847	0,0138
57	102,593	0,3775	0,3858	0,0083
58	102,772	0,3841	0,3901	0,0060
59	103,383	0,3907	0,4049	0,0142
60	103,613	0,3974	0,4105	0,0132
61	103,907	0,4040	0,4177	0,0137
62	104,036	0,4106	0,4209	0,0103
63	104,258	0,4172	0,4263	0,0091
64	104,578	0,4238	0,4342	0,0104
65	104,769	0,4305	0,4389	0,0085
66	104,788	0,4371	0,4394	0,0023
67	105,111	0,4437	0,4474	0,0037
68	105,836	0,4503	0,4655	0,0151
69	106,316	0,4570	0,4775	0,0205
70	106,512	0,4636	0,4824	0,0188
71	107,215	0,4702	0,4999	0,0298
72	107,313	0,4768	0,5024	0,0256
73	107,347	0,4834	0,5033	0,0198
74	107,463	0,4901	0,5062	0,0161
75	107,521	0,4967	0,5076	0,0109
76	108,532	0,5033	0,5329	0,0296
77	108,591	0,5099	0,5344	0,0244
78	108,934	0,5166	0,5429	0,0263
79	108,943	0,5232	0,5431	0,0200
80	109,364	0,5298	0,5536	0,0238
81	109,446	0,5364	0,5556	0,0192
82	109,477	0,5430	0,5564	0,0133
83	109,742	0,5497	0,5630	0,0133
84	109,819	0,5563	0,5649	0,0086

85	109,864	0,5629	0,5660	0,0031
86	109,872	0,5695	0,5662	0,0034
87	110,093	0,5762	0,5716	0,0045
88	110,564	0,5828	0,5832	0,0004
89	110,613	0,5894	0,5844	0,0050
90	110,741	0,5960	0,5875	0,0085
91	110,828	0,6026	0,5896	0,0130
92	110,946	0,6093	0,5925	0,0168
93	111,16	0,6159	0,5977	0,0182
94	111,397	0,6225	0,6035	0,0190
95	111,78	0,6291	0,6127	0,0164
96	111,96	0,6358	0,6170	0,0187
97	112,053	0,6424	0,6192	0,0231
98	112,539	0,6490	0,6308	0,0182
99	112,669	0,6556	0,6339	0,0217
100	112,694	0,6623	0,6345	0,0278
101	112,713	0,6689	0,6349	0,0340
102	112,723	0,6755	0,6352	0,0403
103	113,771	0,6821	0,6596	0,0225
104	113,815	0,6887	0,6606	0,0281
105	114,192	0,6954	0,6692	0,0262
106	115,08	0,7020	0,6892	0,0128
107	115,591	0,7086	0,7004	0,0082
108	115,997	0,7152	0,7092	0,0060
109	116,592	0,7219	0,7219	0,0000
110	116,616	0,7285	0,7224	0,0061
111	116,621	0,7351	0,7225	0,0126
112	116,867	0,7417	0,7276	0,0141
113	116,942	0,7483	0,7292	0,0192
114	117,273	0,7550	0,7360	0,0190
115	117,648	0,7616	0,7436	0,0179
116	117,729	0,7682	0,7453	0,0229
117	117,887	0,7748	0,7485	0,0264
118	118,629	0,7815	0,7631	0,0184
119	118,72	0,7881	0,7648	0,0233
120	119,211	0,7947	0,7742	0,0205
121	119,718	0,8013	0,7836	0,0177
122	119,815	0,8079	0,7854	0,0225
123	119,958	0,8146	0,7880	0,0265
124	120,412	0,8212	0,7962	0,0250
125	120,699	0,8278	0,8013	0,0266
126	121,58	0,8344	0,8163	0,0181
127	121,851	0,8411	0,8208	0,0203
128	122,876	0,8477	0,8371	0,0105
129	123,361	0,8543	0,8445	0,0098

130	123,626	0,8609	0,8485	0,0125
131	123,826	0,8675	0,8514	0,0162
132	124,515	0,8742	0,8612	0,0130
133	124,947	0,8808	0,8671	0,0137
134	125,605	0,8874	0,8758	0,0117
135	125,964	0,8940	0,8803	0,0137
136	126,162	0,9007	0,8828	0,0179
137	126,342	0,9073	0,8850	0,0223
138	127,693	0,9139	0,9006	0,0133
139	128,185	0,9205	0,9059	0,0146
140	129,153	0,9272	0,9157	0,0115
141	132,314	0,9338	0,9424	0,0086
142	133,556	0,9404	0,9508	0,0104
143	134,916	0,9470	0,9589	0,0119
144	136,616	0,9536	0,9675	0,0138
145	137,879	0,9603	0,9728	0,0126
146	138,117	0,9669	0,9738	0,0069
147	140,066	0,9735	0,9804	0,0069
148	140,692	0,9801	0,9822	0,0020
149	153,708	0,9868	0,9982	0,0115
150	155,107	0,9934	0,9987	0,0053

Příloha B – hodnoty veličiny Y

Pořadí prvku	Veličina Y	F_E(x)	F_T(x)	F_T(x)-F_E(x)
1	15,146	0,0066	0,1347	0,1281
2	15,988	0,0132	0,1359	0,1227
3	19,248	0,0199	0,1407	0,1208
4	21,023	0,0265	0,1433	0,1168
5	21,432	0,0331	0,1439	0,1108
6	22,246	0,0397	0,1451	0,1054
7	22,286	0,0464	0,1452	0,0988
8	22,74	0,0530	0,1459	0,0929
9	25,159	0,0596	0,1496	0,0900
10	25,31	0,0662	0,1498	0,0836
11	25,467	0,0728	0,1501	0,0772
12	28,171	0,0795	0,1543	0,0748
13	29,488	0,0861	0,1563	0,0702
14	30,252	0,0927	0,1575	0,0648
15	32,279	0,0993	0,1608	0,0615
16	32,852	0,1060	0,1617	0,0558
17	33,158	0,1126	0,1622	0,0496
18	33,167	0,1192	0,1622	0,0430
19	35,046	0,1258	0,1653	0,0395

20	36,875	0,1325	0,1683	0,0359
21	37,564	0,1391	0,1695	0,0304
22	38,291	0,1457	0,1707	0,0250
23	38,739	0,1523	0,1714	0,0191
24	40,131	0,1589	0,1738	0,0149
25	42,468	0,1656	0,1778	0,0122
26	43,691	0,1722	0,1799	0,0077
27	48,727	0,1788	0,1888	0,0100
28	51,407	0,1854	0,1936	0,0082
29	51,68	0,1921	0,1941	0,0020
30	52,085	0,1987	0,1948	0,0039
31	52,855	0,2053	0,1962	0,0091
32	55,387	0,2119	0,2009	0,0110
33	58,947	0,2185	0,2076	0,0110
34	59,489	0,2252	0,2086	0,0166
35	59,526	0,2318	0,2087	0,0231
36	60,51	0,2384	0,2105	0,0279
37	62,895	0,2450	0,2151	0,0299
38	64,824	0,2517	0,2189	0,0328
39	65,609	0,2583	0,2204	0,0379
40	70,322	0,2649	0,2297	0,0352
41	73,522	0,2715	0,2362	0,0353
42	75,739	0,2781	0,2407	0,0374
43	77,188	0,2848	0,2437	0,0410
44	80,556	0,2914	0,2508	0,0406
45	84,239	0,2980	0,2586	0,0394
46	84,743	0,3046	0,2597	0,0450
47	85,377	0,3113	0,2610	0,0502
48	85,926	0,3179	0,2622	0,0557
49	86,114	0,3245	0,2626	0,0619
50	86,137	0,3311	0,2627	0,0685
51	91,858	0,3377	0,2751	0,0626
52	97,035	0,3444	0,2867	0,0577
53	100,851	0,3510	0,2953	0,0557
54	103,326	0,3576	0,3010	0,0566
55	110,628	0,3642	0,3180	0,0462
56	113,034	0,3709	0,3237	0,0472
57	114,417	0,3775	0,3270	0,0505
58	114,618	0,3841	0,3275	0,0566
59	115,926	0,3907	0,3306	0,0601
60	126,356	0,3974	0,3560	0,0414
61	128,764	0,4040	0,3619	0,0421
62	129,612	0,4106	0,3640	0,0466
63	130,052	0,4172	0,3651	0,0521
64	132,844	0,4238	0,3721	0,0518

65	133,063	0,4305	0,3726	0,0579
66	133,111	0,4371	0,3727	0,0644
67	133,128	0,4437	0,3728	0,0709
68	135,288	0,4503	0,3782	0,0721
69	136,278	0,4570	0,3807	0,0763
70	136,776	0,4636	0,3819	0,0816
71	137,578	0,4702	0,3840	0,0862
72	137,637	0,4768	0,3841	0,0927
73	139,672	0,4834	0,3892	0,0942
74	139,698	0,4901	0,3893	0,1008
75	143,925	0,4967	0,4001	0,0966
76	144,923	0,5033	0,4026	0,1007
77	144,973	0,5099	0,4027	0,1072
78	148,275	0,5166	0,4112	0,1053
79	151,888	0,5232	0,4205	0,1026
80	151,917	0,5298	0,4206	0,1092
81	152,024	0,5364	0,4209	0,1155
82	153,225	0,5430	0,4240	0,1191
83	154,03	0,5497	0,4261	0,1236
84	157,659	0,5563	0,4355	0,1208
85	157,796	0,5629	0,4359	0,1271
86	159,56	0,5695	0,4405	0,1291
87	159,799	0,5762	0,4411	0,1351
88	160,81	0,5828	0,4437	0,1391
89	161,009	0,5894	0,4442	0,1452
90	161,04	0,5960	0,4443	0,1517
91	163,615	0,6026	0,4511	0,1516
92	170,617	0,6093	0,4694	0,1398
93	176,594	0,6159	0,4852	0,1307
94	180,896	0,6225	0,4965	0,1260
95	181,888	0,6291	0,4991	0,1300
96	182,675	0,6358	0,5012	0,1346
97	187,424	0,6424	0,5137	0,1287
98	188,032	0,6490	0,5153	0,1337
99	210,029	0,6556	0,5729	0,0827
100	213,503	0,6623	0,5819	0,0803
101	215,213	0,6689	0,5863	0,0826
102	215,896	0,6755	0,5881	0,0874
103	218,015	0,6821	0,5935	0,0886
104	219,739	0,6887	0,5979	0,0908
105	220,571	0,6954	0,6001	0,0953
106	221,931	0,7020	0,6035	0,0985
107	222,573	0,7086	0,6052	0,1035
108	223,446	0,7152	0,6074	0,1079
109	225,139	0,7219	0,6117	0,1102

110	231,182	0,7285	0,6269	0,1016
111	232,917	0,7351	0,6312	0,1039
112	237,609	0,7417	0,6429	0,0989
113	247,48	0,7483	0,6669	0,0814
114	250,728	0,7550	0,6747	0,0803
115	255,776	0,7616	0,6866	0,0750
116	262,925	0,7682	0,7031	0,0651
117	264,591	0,7748	0,7069	0,0679
118	265,765	0,7815	0,7096	0,0719
119	267,403	0,7881	0,7133	0,0748
120	271,197	0,7947	0,7218	0,0729
121	275,094	0,8013	0,7304	0,0710
122	279,422	0,8079	0,7397	0,0682
123	286,712	0,8146	0,7551	0,0594
124	297,586	0,8212	0,7771	0,0440
125	304,257	0,8278	0,7901	0,0377
126	308,194	0,8344	0,7975	0,0369
127	311,833	0,8411	0,8042	0,0368
128	315,746	0,8477	0,8113	0,0364
129	329,033	0,8543	0,8341	0,0202
130	353,313	0,8609	0,8710	0,0100
131	360,262	0,8675	0,8804	0,0128
132	370,971	0,8742	0,8939	0,0198
133	383,287	0,8808	0,9081	0,0273
134	388,419	0,8874	0,9136	0,0261
135	389,571	0,8940	0,9148	0,0207
136	394,616	0,9007	0,9198	0,0192
137	401,788	0,9073	0,9267	0,0194
138	403,707	0,9139	0,9284	0,0145
139	414,266	0,9205	0,9375	0,0169
140	438,709	0,9272	0,9550	0,0279
141	458,716	0,9338	0,9662	0,0324
142	505,48	0,9404	0,9837	0,0433
143	528,996	0,9470	0,9891	0,0420
144	531,9	0,9536	0,9896	0,0360
145	543,302	0,9603	0,9915	0,0312
146	575,631	0,9669	0,9953	0,0285
147	576,893	0,9735	0,9955	0,0219
148	668,858	0,9801	0,9994	0,0192
149	699,089	0,9868	0,9997	0,0129
150	776,064	0,9934	1,0000	0,0066