Assignment 7: Data Analytics – Histograms & Probability

This assignment includes (2) problems.

Purpose of this assignment

Histograms are graphical representation of data and illustrates the **distribution** of a dataset. In engineering, histograms are used to evaluate products and processes based on specification limits, and to make decisions and predictions. The shape of a histogram provides information about the distribution of the data. In order to be used effectively, histograms must be properly created, formatted, and interpreted.

This assignment develops skills in interpretation and creation of histograms. This assignment also demonstrates the application of **probability** in engineering analysis. You will learn to appropriately represent and analyze data and to use that analysis to evaluate changes in a system or process.

Relevant Course Resources:

Pre-Class Videos	• Data Analytics – Probability • Data Analytics & Excel - Histograms
Lecture Slides	• Data Analytics – Histograms & Probability

Organizing Your Work

Pay attention to how you format and organize your work in Excel.

In general,

1. Complete all of your work on the Excel answer sheet provided **ENGR131_A07_StudentAnswerSheet.xlsx**.
2. Answer all questions with complete sentences by explicitly referring back to your calculations.

Submission Instructions:

1. Re-name your answer sheet as, **ENGR131 _A07_*yourlogin*.xlsx**, where *yourlogin* is your *Purdue Career Account* login.
2. Save your files in a secure location such as your **Purdue Career Account** (This is your Purdue storage space. For more information see https://www.itap.purdue.edu/connections/careeraccount)
3. Submit your work through the designated **Brightspace Assignment Dropbox at** https://purdue.brightspace.com/

Assignment 7: Data Analytics – Histograms & Probability

Problem 1 COMPRESSIVE STRENGTH

Goal Given a set of data (the results of 80 compressive strength tests), your tasks are to:

1. Examine the data provided. Calculate the descriptive statistics for the data, including the minimum, maximum, range, mean, median, variance, and standard deviation. Summarize these in a properly formatted table.
2. Create two histograms. For the first histogram, use **eight** bins. For the second histogram, use **eighteen** bins.
    a. Show your calculations for determining reasonable bin widths given the number of bins, and format your histograms for technical presentation, including an appropriate title, axes titles, and units.
    b. Use an underflow bin and adjust your bin width so your histogram bin limits are whole numbers with no decimal places.
    c. Format your histograms for technical presentation.

After you have created the histograms, answer the following questions:

a. Compare the 8-bin and the 18-bin histograms. Do they look similar? Can we treat the data to be normally distributed? Why or why not?

b. If the required minimum compressive strength for this biomedical application is 106 psi and it is required that no more than 1% of the tests are below the minimum compressive strength, should the engineers recommend this titanium alloy for this application? Compare your calculations using relative probability and using your z-score calculations. Justify your answer referring to your calculations and show evidence of any additional calculations or reasoning you use to support your answer. [Note: the z-score tables are provided at the end of this document]

Background/Technical Content:

Background:

A materials engineer is collaborating with an biomedical engineer to examine the feasibility of using a new alloy (specifically, titanium alloy) to increase the working life of a knee replacement assembly. The alloy must withstand a compressive load in service. To gather data, they apply compressive stresses to **80 specimens** and record the ***Compressive Strength*** of each sample in units of psi (pound per square inch). The engineering team wants to analyze the data to make an evidence-based decision about whether to continue the development of the new alloy.

Figure 1. Artificial knee cross section
https://www.samaterials.com/img/cms/surgical-metals.jpg

Assignment 7: Data Analytics – Histograms & Probability

Technical Content:

Compressive strength is the maximum stress that is applied to a specimen before a fracture occurs (Gere, 2006). **Compressive strength tests** of metals are done using a machine that gradually applies a load on a specimen, typically in the shape of a cube or circular cylinder (See Figure 2). Both the applied load by the machine and the shortening of the specimen can be obtained during such testing.

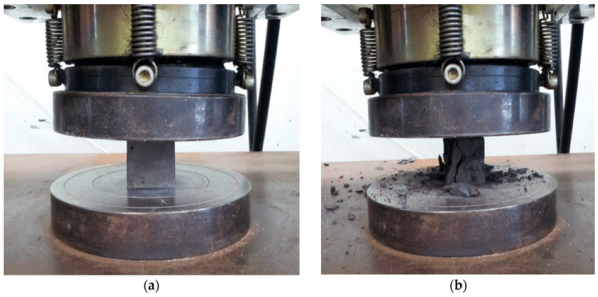


Figure 2. Compressive strength test (before and after)

Koh, Taehoon & Moon, Sung-Woo & Jung, Hyuksang & Jeong, Yeonung & Pyo, Sukhoon. (2018). A Feasibility Study on the Application of Basic Oxygen Furnace (BOF) Steel Slag for Railway Ballast Material. Sustainability. 10. 287. 10.3390/su10020287.

Organizing Your Work:

Inputs:	**Calculations**:		**Outputs:**	
Table 1: Compressive load to failure (psi)	Table 2: Descriptive Statistics	Calculations for histogram creation (e.g., bin widths) and z-score calculations	Histograms	**1a)** **1b)**
	Calculate the minimum, maximum, range, mean, median, variance, and standard deviation and summarize in a table.	Perform the calculations needed to create a properly formatted histogram.	Create the two histograms with different bin numbers.	Answer all questions with complete sentences (in 2-3 sentences) AND by explicitly referencing data

Assignment 7: Data Analytics – Histograms & Probability

Problem 2 EVALUATING IMPROVEMENTS TO QUALITY CONTROL PROCESS

Goal: Given three years of process data in an Excel worksheet, your tasks are to:

1. Examine the data provided. Do the following:

a. Create a new column that calculates the total number of Skittles in each bag.

b. Calculate the descriptive statistics for the total number of Skittles in a bag for each year. Use Excel built-in functions and summarize your results in a properly formatted table:

i. Mean

ii. Standard deviation

iii. Count

iv. Max

v. Min

vi. Range

2. For each year, under the Histogram section, insert a histogram of the total Skittles count for that year. You are creating three different histograms. Use the descriptive statistics calculated in Step 1b to build your histograms:

a. Are there any outliers in the data? If there are, provide an explanation of how you will handle the outlier(s).

i. If you discard any outliers, re-calculate the descriptive statistics and build a new histogram representing your adjusted data. Copy and paste the data to a new sheet to perform the new calculations.

b. Use an underflow bin to make your bin edges whole numbers.

c. If necessary, adjust your bin width. *Hint: try a bin width of 1.*

d. Format your histograms for technical presentation.

After you have created the histograms, answer the following questions:

a. The standard weight for a bag of original Skittles is 61.5 grams. Each individual Skittle weighs an average of 1.062 grams. Using a z-score, for each year, calculate the probability that a bag of Skittles manufactured in that year weighs more than the goal of 61.5 grams. Explain why you can use the z-score to calculate these probabilities.

b. The goals of the industrial engineer's project were to reduce variability in the packaging and reduce the number of bags produced over the target weight of 61.5 grams (see Background information below). Did she achieve these goals? Reference data in your answer. [Note: the z-score tables are provided at the end of this document]

Assignment 7: Data Analytics – Histograms & Probability

Background/Technical Content:

The Wrigley Company manufactures over 200 million Skittles candies each day at their six manufacturing plants, including plants in Texas, Illinois, and China. Skittles Originals come in five flavors: orange, lemon, lime, grape, and strawberry (see Figure 1). Each flavor is made on separate manufacturing equipment, and then the flavors are combined for packaging. The goal is to have 20% of each color in every bag and a consistent total number of Skittles per bag.

Figure 1. Skittles Original.

Skittles production begins at a mixing station, where five large mixers develop the five signature flavors. The Skittles are tested for flavor and color, and fed into a mixing conveyor, where they are further tested by x-ray and ultimately packaged.

Figure 2: Quality Control Testing

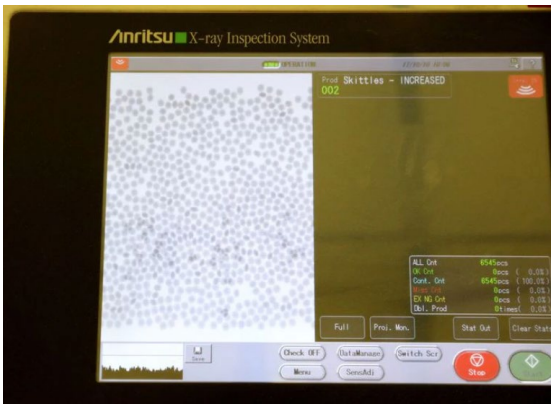


Figure 3: X-ray of Skittles on conveyor

Assignment 7: Data Analytics – Histograms & Probability

Figure 4: Skittles rolling down the conveyor belt towards the packaging step.

In 2017, an industrial engineer began leading a five-year project to **optimize** the production of Skittles at Mars, Incorporated. With output rates of 200,000,000 Skittles per day, even small increases in productivity can have large financial impacts. The engineer has been concentrating on **reducing variability in the packaging** of the Skittles to provide customers with a more consistent product. She is also working to **reduce the cost/bag of Skittles** by improving the packaging process to reduce the number of bags that leave the factory out of specification – over the target weight of 61.5 grams/bag of Skittles.

When products like Skittles are mass-produced, it is impossible to quality test every single bag. To address this, a company will statistically sample their product. A small number of units from the population are chosen at random. The bags are opened, and the contents are evaluated against the standards for count, color distribution, and shape. If the **statistical sample** passes, the quality control engineer can be confident the entire batch is good. This is a destructive test, meaning the Skittles cannot be repackaged and sold. The engineer has compiled sample data from 2017, 2018, and 2019 in a table. This table is in your Excel document.

Organizing your work:

Inputs:	**Calculations:**			**Outputs:**
Table 1: All data	Table 2: Descriptive Statistics	Histograms	Table 4: Calculation of Z-score	Answers to questions a and b.
Calculate the total number of Skittles per bag	Use Excel **built-in functions** to calculate descriptive statistics. Create Table 3 if any outliers are removed.	Create three histograms to display the data from each of the three years.	Use Excel **built-in functions** to perform calculations	Answer all questions with complete sentences (in 2-3 sentences) AND by explicitly referencing data

References

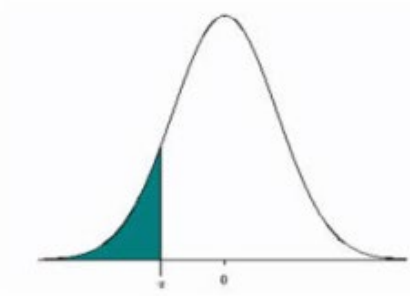
Skittles Original candies [online image]. (n.d.). Retrieved from http://www.amazon.com/Skittles-Original-Candy-14-Ounce-Packages/dp/B0029K81ZU.

Bomkamp, S. (2018, December 15). Check out how Skittles are made at a candy factory in Yorkville. Retrieved from https://www.chicagotribune.com/business/ct-biz-skittles-factory-photo-20170922-story.html

Z-score tables [online image]. (n.d.). Retrieved from http://www.ttable.org/z-score-table.html

Assignment 7: Data Analytics – Histograms & Probability

Table of Standard Normal Probabilities for Negative Z-scores

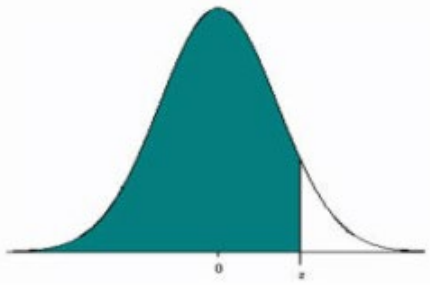


z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Table 1: Negative z-score table

Assignment 7: Data Analytics – Histograms & Probability

Table of Standard Normal Probabilities for Positive Z-scores



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Table 2: Positive z-score table

Assignment 7: Data Analytics – Histograms & Probability

Problem 1*: Aerospace Material Compressive Strength Data*	
Learning Objectives Your work will be graded on demonstration of proficiency of the following learning objectives:	**Did you address this?**
DV01 - Efficient use of engineering tools for basic statistics. Use of automated solutions, such as cell referencing and built in functions.	
DV04 – Prepare a table for technical presentation with proper formatting with title, row labels, column labels, units and correct decimal places.	
DV06 - Describe the central tendency of data using descriptive statistics (mean, median, mode).	
DV07 – Describe the variability of data using statistical methods (standard deviation, variance).	
DV10 – Create a histogram with meaningful number of bins, bin widths, and upper and lower limits	
DV05 – Prepare a chart for technical presentation with proper formatting, including title, axes labels, appropriately scaled axes, units and appropriate markers.	
SQ01 – Use accurate scientific, mathematical and/or technical concepts, units and/or data in solution.	
DV17 – Describe the relationship between z-score and probability	
EB03 – Clearly articulate reasons for answers with explicit reference to data to justify decisions or to evaluate alternative solutions.	

Problem 2*: Evaluating Improvements to Quality Control Process*	
Learning Objectives Your work will be graded on demonstration of proficiency of the following learning objectives:	**Did you address this?**
DV01 - Efficient use of engineering tools for basic statistics. Use of automated solutions, such as cell referencing and built in functions.	
DV04 – Prepare a table for technical presentation with proper formatting with title, row labels, column labels, units and correct decimal places.	
DV06 - Describe the central tendency of data using descriptive statistics (mean, median, mode).	
DV07 – Describe the variability of data using statistical methods (standard deviation, variance).	
DV08 – Make accurate comparisons across groups with explicit reference to data	
DV10 – Create a histogram with meaningful number of bins, bin widths, and upper and lower limits	
DV05 – Prepare a chart for technical presentation with proper formatting, including title, axes labels, appropriately scaled axes, units and appropriate markers.	
DV17 – Describe the relationship between z-score and probability	
EB03 – Clearly articulate reasons for answers with explicit reference to data to justify decisions or to evaluate alternative solutions.	