



Name: \_\_\_\_\_

PUID \_\_\_\_\_

Instructor (circle one):    Anand Dixit    Timothy Reese    Halin Shin    Khurshid Alam

Class Start Time:    ☐ 11:30 AM    ☐ 12:30 PM    ☐ 1:30 PM    ☐ 2:30 PM    ☐ 3:30 PM    ☐ Online

As a boilermaker pursuing academic excellence, I pledge to be honest and true in all that I do.  
Accountable together - we are Purdue.

**Instructions:**

1. **IMPORTANT** Please write your **name** and **PUID** clearly on every **odd page**.
2. **Write your work in the box. Do not run over into the next question space.**
3. You are expected to uphold the honor code of Purdue University. It is your responsibility to keep your work covered at all times. Anyone cheating on the exam will automatically fail the course and will be reported to the Office of the Dean of Students.
4. It is strictly prohibited to smuggle this exam outside. Your exam will be returned to you on Gradescope after it is graded.
5. One handwritten or type double-sided 8.5 in x 11 in crib sheet is allowed for the exam.
6. The only materials that you are allowed during the exam are your **scientific calculator, writing utensils, erasers, your crib sheet, and your picture ID**. If you bring any other papers into the exam, you will get a **zero** on the exam. Colored scratch paper will be provided if you need more room for your answers. Please write your name at the top of that paper also.
7. Keep your bag closed and cellphone stored away securely at all times during the exam.
8. If you share your calculator or have a cell phone at your desk, you will get a **zero** on the exam.
9. The exam is only 60 minutes long so there will be no breaks (including bathroom breaks) during the exam. If you leave the exam room, you must turn in your exam, and you will not be allowed to come back.
10. You must show **ALL** your work to obtain full credit. An answer without showing any work may result in **zero** credit. If your work is not readable, it will be marked wrong. Remember that work has to be shown for all numbers that are not provided in the problem or no credit will be given for them. All explanations must be in complete English sentences to receive full credit.
11. All numeric answers should have **four decimal places** unless stated otherwise.
12. After you complete the exam, please turn in your exam as well as your table and any scrap paper that you used. Please be prepared to **show your Purdue picture ID**. You will need to **sign a sheet** indicating that you have turned in your exam.

**Your exam is not valid without your signature below. This means that it won't be graded.**

I attest here that I have read and followed the instructions above honestly while taking this exam and that the work submitted is my own, produced without assistance from books, other people (including other students in this class), notes other than my own crib sheet(s), or other aids. In addition, I agree that if I tell any other student in this class anything about the exam BEFORE they take it, I (and the student that I communicate the information to) will fail the course and be reported to the Office of the Dean of Students for Academic Dishonesty.

Signature of Student: \_\_\_\_\_

**1. True/False Questions. Please indicate the correct answer by filling in the circle. If you indicate the correct answer by any other way, you may receive 0 points for the question.**

1.1 Let **A** and **B** be events in a same sample space  $\Omega$ , with  $P(A) > 0$  and  $P(B) > 0$ .

☐ **T** or ☐ **F** If  $P(A \cap B) = P(A|B)P(B)$ , then the events A and B must be independent.

1.2 Let **A** and **B** be events in a same sample space  $\Omega$ . It is known that  $P(A) > 0$  and  $P(B) > 0$ .

☐ **T** or ☐ **F** If **A** and **B** are mutually exclusive, then it must be true that  $P(A) = 1 - P(B)$ .

1.3 A research group is studying a population known to be normally distributed but can only afford a sample of size 5 due to high sampling cost. One of the researchers states that the sampling distribution of the sample mean will be normally distributed. Another colleague in the research group states that this is incorrect as the CLT requires a sample size of at least 30 ( $n \geq 30$ ) for the sampling distribution of the mean to be normally distributed.

☐ **T** or ☐ **F** The colleague is correct in stating that the distribution for the sampling distribution of the mean will not be normally distributed since  $n < 30$ ?

1.4 Suppose **X** follows a normal distribution with mean  $\mu$  and standard deviation  $\sigma$  and  $Z \sim N(0, 1)$ .

☐ **T** or ☐ **F** It follows that  $P(\mu - 2\sigma < X < \mu + 2\sigma) = 2P(Z < 2) - 1$

1.5 Let  $X_1, X_2, \dots, X_n$  be independent random variables from the same identical distribution with mean  $\mu = E[X_1]$  and  $\sigma^2 = \text{Var}(X_1)$ .

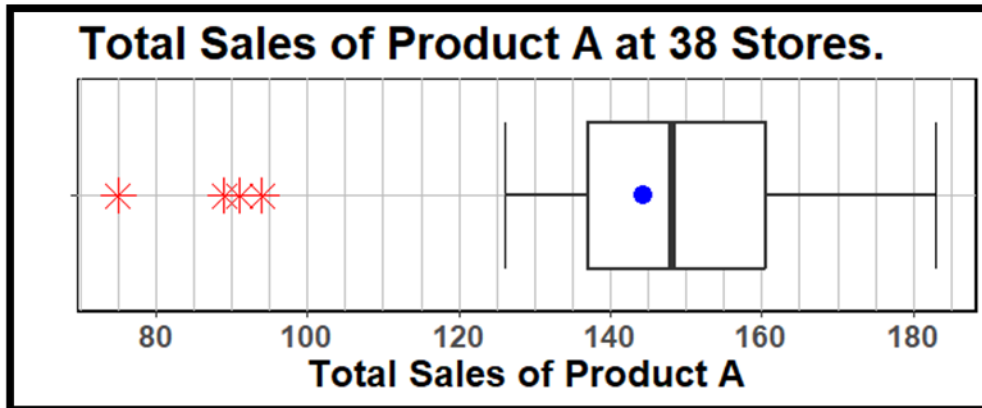
☐ **T** or ☐ **F** If  $Y = \sum_{i=1}^n X_i$  then  $\text{Var}(Y) = \frac{\sigma^2}{n}$ .

1.6 Let **X** satisfy the conditions to be distributed as a Binomial distribution.

☐ **T** or ☐ **F** Since **X** follows a Binomial distribution it follows that  $\text{Var}(X) = \frac{1}{n}E[X](n - E[X])$

**2. Multiple Choice Questions. Please indicate the correct answer by filling in the circle. If you indicate the correct answer by any other way, you will receive 0 points for the question. For each question, there is only one correct option given.**

2.1. Using the modified boxplot below, which of the following is **most accurate** regarding the lower inner fence for the Total Sales of Product A.



- (A)  $IF_L = 127$
- (B)  $IF_L = 100$
- (C)  $IF_L < 95$
- (D)  $95 < IF_L < 110$
- (E)  $125 < IF_L < 130$

2.2. Which of the following is ALWAYS true?

- (A)  $P(A \cup B) = P(A) + P(B)$
- (B)  $P(A \cup B) = P(A) - P(A \cap B)$
- (C)  $P(A \cap B) = P(A) + P(B) - P(A \cup B)$
- (D) There is more than one correct answer.
- (E) None of these are ALWAYS true.

2.3. The measure of central tendency which is the least likely to be influenced by outliers in the dataset is

- (A) sample median
- (B) sample mean
- (C) IQR
- (D) sample standard deviation
- (E) more than one of the above.

2.4. Which of the following is a correct description for  $\bar{x}$ ?

- (A) It is a sample statistic.
- (B) It is a measure of spread
- (C) It is a population average.
- (D) More than one is correct.
- (E) None of the above choices is correct.

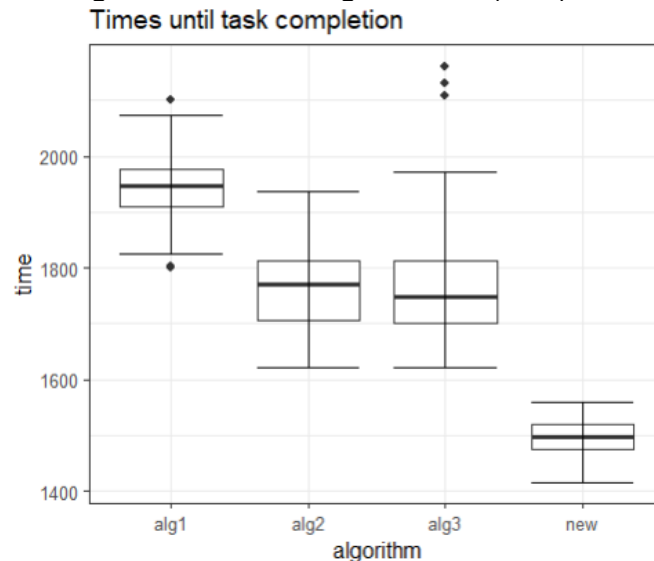
2.5. The salary distribution at a particular company is **extremely right skewed**. The sampling distribution of the sample mean will be

- Ⓐ exactly normal for any sample size.
- Ⓑ approximately normal for any sample size.
- Ⓒ approximately normal for a sample size of 30 provided there are no real outliers.
- Ⓓ approximately normal for sufficiently large sample size and if there are no real outliers.
- Ⓔ never normal it is extremely right skewed as stated in the question.

2.6 Which of the following should be used if we know the form and all parameters of the population level model and we want to gain insight into the possible samples?

- Ⓐ Descriptive statistics
- Ⓑ Data collection
- Ⓒ Probability
- Ⓓ Inferential statistics
- Ⓔ Code

3. A software developer is determining if her algorithm is better than three different commonly used algorithms. For each algorithm, she ran 64 instances of the same problem and recorded the time (in seconds) it took to complete the task in each instance. The boxplots of the outcomes for the four algorithms, including her own (new), are shown below.



- a. Describe the shape of records for Algorithm 1 (alg1). Please explain your answer.

answer: **symmetric**

The median is close to the middle of the box and the upper and lower whiskers are of similar size.

- b. Which algorithms have 'real' outliers? Please explain your answer.

Alg3 only

Reason: The explicit points are from the rest of the data and there is a clear gap between the upper whisker and these points suggesting that this is not simply tail behavior.

Having explicit points using the 1.5 IQR rule is not a valid reason.

- c. Approximately how many times did Algorithm 2 (alg2) run for longer than 1700 seconds?

1700 is approximately the first quartile ( $Q_1$ )

$\frac{3}{4}$  of the points are above  $Q_1$

For each algorithm, she ran 64 instances of the same problem and  $\frac{3}{4}$  are above 1700 so  $\frac{3}{4} \times 64 = 48$ . Therefore, 48 of the runs were longer 1700 seconds.

- d. Why might the new algorithm be advantageous compared to the three existing ones in terms of running time?

The new algorithm takes a shorter time to run on average and is more consistent with regards to the run times. The shorter time is indicated by the lower run times of the new algorithm even in the most extreme case observed. The consistency is indicated by the small box and whiskers.

4. The (PDF) of the random variable **X** has a functional form described by the function below

$$f_X(x) = \begin{cases} k(x+1) & -1 < x < 0 \\ \frac{12}{211}(-x^2 + 2x + 8) & 0 < x < 2 \\ k(3-x) & 2 < x < 3 \\ 0 & \text{otherwise} \end{cases}$$

but the constant **k** that makes this a valid (PDF) is unknown. The corresponding (CDF) is partially known and given below.

$$F_X(x) = \begin{cases} 0 & x < -1 \\ \frac{3}{422}(x+1)^2 & -1 \leq x < 0 \\ \frac{1}{422}(-8x^3 + 24x^2 + 192x + 3) & 0 \leq x < 2 \\ \left[ \quad \quad \quad \mathbf{A} \quad \quad \quad \right] & 2 \leq x < 3 \\ 1 & x \geq 3 \end{cases}$$

- a. Determine the constant **k** that makes the **PDF** valid.  
(You may write the answer as a reduced fraction or use 4 decimal places.)

If we graph the PDF we see it has possible discontinuities at  $x = 0$  and  $x = 2$  but this does not make it invalid. It can still satisfy the requirements of a PDF.

- 1) Non-negativity  $f_X(x) \geq 0$  everywhere.
  - 2) Area under curve is unity:  $\int_{-\infty}^{+\infty} f_X(x)dx = 1$
- Non-negativity is satisfied as long as  $k > 0$

To find the normalizing constant  $k$  we need to figure out what  $k$  makes this integrate to unity.

$$\int_{-\infty}^{+\infty} f_X(x)dx = 1$$

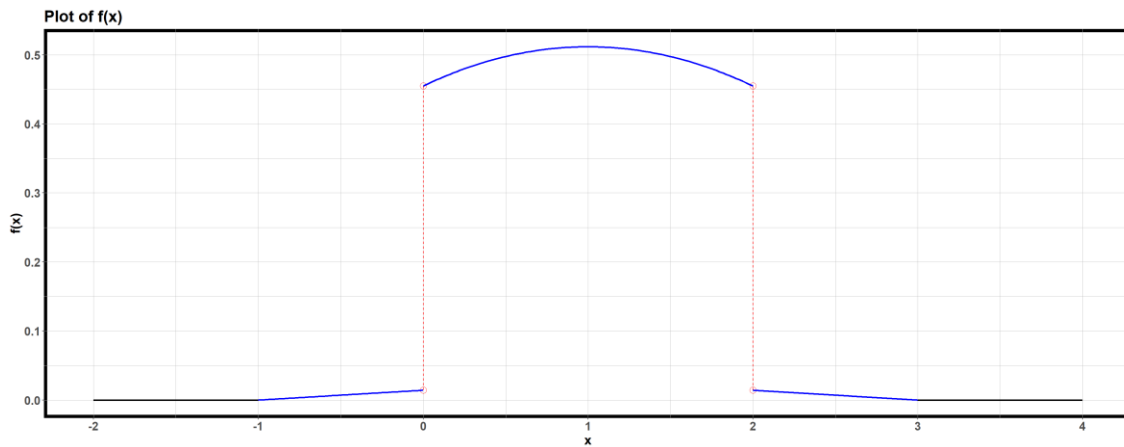
$$k \int_{-1}^0 (x+1)dx + \frac{12}{211} \int_0^2 (-x^2 + 2x + 8)dx + k \int_2^3 (3-x)dx = 1$$

$$k \int_0^1 udu + \frac{12}{211} \left[ -\frac{1}{3}x^3 + x^2 + 8x \right]_0^2 + k \int_0^1 udu = 1$$

$$k + \frac{208}{211} = 1$$

$$k = \frac{3}{211}$$

The pdf has the following form →



b. Determine the missing part of the CDF labeled **A**.

First, the PDF is

$$f_X(x) = \begin{cases} \frac{3}{211}(x+1) & -1 < x < 0 \\ \frac{12}{211}(-x^2 + 2x + 8) & 0 < x < 2 \\ \frac{3}{211}(3-x) & 2 < x < 3 \\ 0 & \text{otherwise} \end{cases}$$

For the part labeled A the CDF has already accumulated the entire area under the curve of the PDF for  $x < 2$  which would have area  $F_X(2) = \frac{1}{422}(-8 * 2^3 + 24 * 2^2 + 192 * 2 + 3) = \frac{419}{422}$ . It then begins to gain a little more area after 2.

$$\begin{aligned} A &= \frac{419}{422} + \int_2^x \frac{3}{211}(3-t)dt = \frac{419}{422} + \frac{3}{211} \left[ -\frac{t^2}{2} + 3t - 4 \right]_2^x \\ &= \frac{1}{422} [-3x^2 + 18x + 395] \end{aligned}$$

- c. Find the probability that the random variables takes on a value **greater than 0.5**. (You may write the answer as a reduced fraction or use 4 decimal places.)

- d. Determine the probability that the random variable **X** would take a value **between 0.5 and 2.5**. (You may write the answer as a reduced fraction or use 4 decimal places.)

First the CDF is now

$$F_X(x) = \begin{cases} 0 & x < -1 \\ \frac{3}{422}(x+1)^2 & -1 \leq x < 0 \\ \frac{1}{422}(-8x^3 + 24x^2 + 192x + 3) & 0 \leq x < 2 \\ \frac{1}{422}[-3x^2 + 18x + 395] & 2 \leq x < 3 \\ 1 & x \geq 3 \end{cases}$$

$$P(0.5 < X < 2.5) = F_X\left(\frac{5}{2}\right) - F_X\left(\frac{1}{2}\right)$$

$$F_X\left(\frac{1}{2}\right) = \frac{52}{211}$$

$$F_X\left(\frac{5}{2}\right) = \frac{1}{422} \left( -3 * \left(\frac{5}{2}\right)^2 + 18 * \left(\frac{5}{2}\right) + 395 \right) = \frac{1685}{1688}$$

$$P(0.5 < X < 2.5) = \frac{1685}{1688} - \frac{52}{211} = \frac{1269}{1688} = 0.7518$$

- e. Determine what value represents the **0.1**( $p = 0.001$ ) **percentile**. (Use 2 decimal places.)

The 0.1 percentile will occur before  $x = 0$  since we accumulate more than  $\frac{3}{422} \approx 0.007$  after 0. We set the CDF in the region  $-1 \leq x < 0$  to 0.001 and solve for  $x$ .

$$\frac{3}{422}(x+1)^2 = 0.001 \rightarrow x^* = \sqrt{\frac{422}{3} \times 0.001} - 1 = -0.6249 \text{ note we use the positive square root because } x = -\sqrt{\frac{422}{3} \times 0.001} - 1 = -1.3751 \text{ which is outside the support of the pdf.}$$

**Therefore the 0.1 percentile is  $x^* = -0.6249$**



- f. Find the standard deviation of the random variable  $X$  using the knowledge below. **(Use 4 decimal places.)**

$$E[X^2] = \frac{2836}{2110}$$

Observing the the pdf is symmetric about  $x = 1$  tells us that  $E[X] = 1$ .

$$\sigma_X = \sqrt{\text{Var}(X)} = \sqrt{E[X^2] - (E[X])^2} = \sqrt{\frac{2836}{2110} - 1} = \sqrt{\frac{363}{1055}} = 0.5866$$

Determine the distribution of the mean of a random sample of 42 observations from this distribution. Since the distribution is symmetric, we do not need a large sample size to ensure the CLT applies. (Hint you need information from part e) to complete this question.)

- g. Using the distribution for the average determine the probability that the average of 42 observations would be greater than 1.35.  
(Use exactly **4 decimal places.**)

Due to the symmetry and the sample of size 42 we can be sure that the random variable  $\bar{X}$  will be normally distributed. Specifically,

$$\bar{X} \sim N\left(\mu_{\bar{X}} = 1, \sigma_{\bar{X}} = \frac{0.5866}{\sqrt{42}}\right)$$

$$P(\bar{X} > 1.35) = 1 - P\left(Z < \frac{1.35 - 1}{\frac{0.5866}{\sqrt{42}}}\right) = 1 - P(Z < 3.87) \approx 0$$

- h. Determine the **91.92<sup>nd</sup>** percentile for the mean of a random sample of 42 observations from this distribution.  
(Use **4 decimal places.**)

91.92 percentile corresponds to a z-score of  $z = 1.4$  Next, we convert this to the distribution of  $\bar{X}$  and get the value under this distribution.  $\bar{x} = \mu + z * \sigma = 1 + 1.4 \times \frac{0.5866}{\sqrt{42}} = 1.1267$ .

**5. (12 points)** An examination for a course has a written component and an oral component. Assume that in each component, a student can either pass or fail and no other grade is possible. The results showed that 75% of the class passed the written component while 85%

of the class passed the oral component. Further it was found that 95% of the class passed either the written component or the oral component.

- a. (5 pts.) What percentage of the class **failed** both components? Please show all your work and include at least 3 decimal places in all work and the answer.

$$\begin{aligned} P(W) &= 0.75 & P(O) &= 0.85 & P(W \cup O) &= 0.95 \\ P(W' \cap O') &= 1 - P(W \cup O) = 1 - 0.95 \end{aligned}$$

- b. A particular student in the class has failed the written component. Given this knowledge, what is the probability that the student has also failed the oral component? Please show all your work and include at-least 3 decimal places in all work and your answer.

$$P(O'|W') = \frac{P(O' \cap W')}{P(W')} = \frac{P(O' \cap W')}{1 - P(W)} = \frac{0.05}{1 - 0.75} = \frac{0.05}{0.25} = \frac{1}{5} = 0.2$$

6. A customer support center for an unnamed company receives a large number of calls per day. An analysis of the service quality has found that for any 12-hour period, an average of 3.5 calls receive an unsatisfactory customer rating. The number of unsatisfactory calls per 12-hour period is known to follow a Poisson distribution. Answer the questions below regarding the number of calls receiving an unsatisfactory customer rating.

- a. What is the average number of unsatisfactory outcomes over a 1-hour period? **(Use 4 decimal places.)**

Since, there are an average of  $\lambda = 3.5$  calls per 12-hour period that receive an unsatisfactory rating therefore during any 1-hour period there is on average

$$\lambda' = \frac{3.5}{12} = \frac{7}{24} \approx 0.2917.$$

- b. Determine the probability of at least one unsatisfactory outcome during any 1-hour period. **(Use 4 decimal places.)**

Let  $X$  be the number of calls during a 1-hour period that receive an unsatisfactory rating. We know that that  $X \sim \text{Pois} \left( \lambda' = \frac{7}{24} \right)$

Using complement rule and the fact that atleast 1 is  $\geq 1$

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{\left(\frac{7}{24}\right)^0 e^{-\frac{7}{24}}}{0!} = 1 - e^{-\frac{7}{24}} \approx 0.2530$$

c. Determine the probability that during any 12-hour period of time the number of unsatisfactory outcomes exceeds the known average number as stated in the introduction. **(Use 4 decimal places.)**

We are now looking at a 12-hour period so we use the original  $\lambda = 3.5$  which is also the known average, let  $Y \sim \text{Pois}\left(\lambda = \frac{7}{2}\right)$ .

$$P(Y > 3.5) = P(Y \geq 4) = 1 - P(Y = 0) - P(Y = 1) - P(Y = 2) - P(Y = 3)$$

$$= 1 - e^{-\frac{7}{2}} - \left(\frac{7}{2}\right)e^{-\frac{7}{2}} - \frac{\left(\frac{7}{2}\right)^2 e^{-\frac{7}{2}}}{2!} - \frac{\left(\frac{7}{2}\right)^3 e^{-\frac{7}{2}}}{3!}$$

$$= 1 - e^{-\frac{7}{2}} \left( 1 + \frac{7}{2} + \frac{\left(\frac{7}{2}\right)^2}{2} + \frac{\left(\frac{7}{2}\right)^3}{6} \right) \approx 0.4634$$

d. Determine the probability that during two separate non-overlapping 1-hour periods there is exactly 1 unsatisfactory outcome during the first period and exactly 1 unsatisfactory outcome during the second period. **(Use 4 decimal places.)**

Since the periods of time do not overlap we can use independence and since both periods are of the same length they are equally distributed as  $X \sim \text{Pois}\left(\lambda' = \frac{7}{24}\right)$ .

$$P(X_1 = 1 \cap X_2 = 1) = P(X = 1)^2 = \left[ \left(\frac{7}{24}\right) e^{-\frac{7}{24}} \right]^2 \approx 0.0475$$

e. Determine the standard deviation for the number of unsatisfactory outcomes during any 1-hour period. **(Use 4 decimal places.)**

Since, we know that for any 1-hour period the distribution is Poisson with  $\lambda' = \frac{7}{24}$  and we know that the standard deviation of a Poisson is the square root of the parameter we have

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

[illegible]