Sta 101 Fall 2016

Midterm 2 (Version A)

Dr. Mukherjee November 10, 2016

Last N	Name: _						Fi	rst Name:		
Sectio	n: 8:30	10:05	11:45	1:25	3:05	7	Геат N	ame:		
	y state t this exan					th or ga	$iined\ info$	ormation in	any wa	y from my classmates
			Signa	$ture: _$						
	— — — otential Conduct			_	-				– – be repo	 orted to Undergrad-
1. Y	ou have 7	5 minutes	s to comp	lete the	exam.					
	how <u>all</u> yo nded ques								. No cred	lit will be given for open
t										elow. If you choose more t will be given for these
О	_	11" sheet	of notes							tudent during the exam, a dictionary, and to ask
d		eadphone	s, notes (other th	an your c	heat shee				checking the time, music and to communicate with
6. V	Vrite clear	ly. Short	answers a	are best!						
					G	lood lu	.ck!			
			4. (a (B	 ©®		9.		 D	
			5. (A B	\odot		10.		_ D	
			6. (A B	\bigcirc		11.		D	
			7. <	\mathbb{A}^{\oplus}	\bigcirc		12. 🤇		D	
			8. (© (D)		13. (D	
		-]
					0.1	0.9	0.2	MC A 12	To4-1	
			Points on	rnod	Q 1	Q 2	Q 3	Q 4 - 13	Total	

25

20

25

30

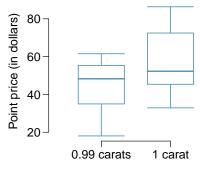
100

Available points

1. (20) Diamond Prices.

Prices of diamonds are determined by what is known as the 4 Cs: cut, clarity, color, and carat weight. The prices of diamonds go up as the carat weight increases, but the increase is not smooth. For example, the difference between the size of a 0.99 carat diamond and a 1 carat diamond is undetectable to the naked human eye, but the price of a 1 carat diamond tends to be much higher than the price of a 0.99 diamond. In this question we use two random samples of diamonds, 0.99 carats and 1 carat, each sample of size 23, and compare the average prices of the diamonds. In order to be able to compare equivalent units, we first divide the price for each diamond by 100 times its weight in carats. That is, for a 0.99 carat diamond, we divide the price by 99. For a 1 carat diamond, we divide the price by 100. The distributions and some sample statistics are shown below.

	0.99 carats	1 carat
Mean	\$ 44.51	\$ 56.81
SD	\$ 13.32	\$ 16.13
\mathbf{n}	23	23



(a) (2) Suppose we want to determine if there is a difference between the average standardized prices of 0.99 and 1 carat diamonds. State the null and alternative hypotheses clearly.

(b) (6) What are the necessary conditions for test? Circle all that apply.

☐ Independence

- ☐ Constant variance
- ☐ 5 expected counts for each group
- \square Approximate normality

 \square $np \ge 10$ and $n(1-p) \ge 10$

 $\square \ \ n \geq 30$ for all groups.

(c)	(8)	Conduct the hypothesis test.
	i.	Compute the test statistic value
	ii	Compute the p -value for this test.
	11.	compare the p value for this test.
		State your decision and conclusion for this test using a 0.05
	111.	State your decision and conclusion for this test using $\alpha = 0.05$.
(4)	(4)	Suppose the 05% confidence interval for u.s. u. is (21.36 - 3.23) FILL IN and CIDCLE
(u)		Suppose the 95% confidence interval for $\mu_{.99} - \mu_1$ is $(-21.36, -3.23)$, FILL IN and CIRCLE correct responses.

We are 95% confident that the average price standardized price of .99 carat diamonds

is _____ dollars more / less to _____ dollars more / less $than \ the$

average standardized price of 1 carat diamonds.

2. (25) Action on Aquatic Environment.

In a study that investigates the impact of the Danish Action Plan for the Aquatic Environment, which addresses pollution of the Danish water resources, the concentration of nitrogen (measured in g/m^3) was measured in a particular river in 1998, 2003, and 2011. Six measurements were randomly taken in each year. A summary of the nitrogen concentration is provided. We want to evaluate the relationship between the two variables using an ANOVA at the 5% significance level.

	\bar{x}	s	\overline{n}
1998	5.55	0.486	6
2003	5.19	0.362	6
2011	4.05	0.673	6

(a) (3) What are the hypotheses for evaluating the relationship between the two variables?

(b) (6) What are you necessary conditions for ANOVA? Circle all that apply.

☐ Independence

☐ Constant variance

 \square 5 expected counts for each group

☐ Approximate normality

 \square $np \ge 10$ and $n(1-p) \ge 10$

 \square $n \ge 30$ for all groups.

(c) (7) Complete the following ANOVA table. Show any work in the space provided below, and insert final values in the table.

	degrees of freedom	Sum Sq	Mean Sq	F value	<i>p</i> -value
Year		7.39			0.0004436
Residuals					
Total		11.49			

(d)	(3)	State the decision and conclusion of the hypothesis test in context.
(e)	(2)	What percent of total variability in nitrogen concentration is explained by year?
(f)	(4) cone	If we were to conduct a pairwise post-hoc test for a difference between the average nitrogen centrations between 1998 and 2003, what values should be used for the following quantities?
		(2) degrees of freedom:
	-	(2) significance level:

3. (25) Perils of Living Dangerously in the Slasher Horror Film. The slasher horror film has been deplored based on claims that it depicts eroticized violence against predominately female characters as punishment for sexual activities. To test this assertion, a quantitative content analysis was conducted to examine the extent to which gender differences are evident in the association between character survival and engagement in sexual activities. Information pertaining to gender, engagement in sexual activities, and survival was coded for film characters from a simple random sample of 50 English-language, North American slasher films released between 1960 and 2009.¹

Gender	Sexual activity	Outcome of physi	cal aggression	n
		Survival		
Female				
	Present	13.3% (<i>n</i> =11)	86.7% (n=72)	83
	Absent	28.1% (<i>n</i> =39)	71.9% (n=100)	139
Male				
	Present	9.5% (n=7)	90.5% (n=67)	74
	Absent	14.8% (<i>n</i> =28)	85.2% (n=161)	189

- (a) (20) Suppose we want to conduct a hypothesis test to evaluate whether the survival rates of female characters who engage in sexual activity is different from female characters who do not.
 - i. (2) State the null and alternative hypotheses for this test

ii. (8) Calculate the test statistic for this hypothesis test.

iii. (2) Compute the p-value for this test

¹Welsh, Andrew. "On the perils of living dangerously in the slasher horror film: Gender differences in the association between sexual activity and survival." Sex Roles 62.11-12 (2010): 762-773.

iv.	(2)	Using your resu	ılt from	(iii), is the	survival	of female	characters	in slasher	films	associa	ted
	with	sexual activity	? Yes	s. No.	(Cire	cle one, a	assume $\alpha =$	0.05)			

v. (6) Could this same test be used for testing association between sexual activity and survival for male characters? Explain your answer by checking appropriate conditions. If the answer is NO, what alternative testing method should be used?

(b) (3) Compute the standard error used for computing the 95% confidence interval for $p_{pres} - p_{abs}$ for females.

Multiple Choice - 30 points (3 points each) Choose the best answer for the following questions. Fill in the bubbles on the first page of the exam. Each question is worth 3 points.

4. (3) Suppose in a population 20% of adults do not have a savings account. What is the expected shape of the sampling distribution of proportions of adults without a savings account in random samples of 60 adults from this population?

(a) right-skewed

(c) symmetric

(b) left-skewed

(d) uniform

- 5. (3) A November 2015 Gallup poll reported that 45% of lesbian, gay, bisexual or transgender (LGBT) Americans living with a same-sex partner are married. The poll also reported that the "margin of error" for this poll was 2%. What does the margin of error of 2% indicate?
 - (a) The true percent of LGBT Americans living with a same-sex partner who are married is estimated to be between 43% and 47%.
 - (b) The true percent of LGBT Americans living with a same-sex partner who are married is probably higher than 45% and closer to 47%.
 - (c) There is a 2% chance that the estimate of 45% is wrong.
 - (d) The estimate of 45% can be at most 2% off of the true percent of LGBT Americans living with a same-sex partner who are married.
- 6. (3) In a test of the effects of sleep deprivation, college student volunteers were randomly assigned to two groups. The treatment group was kept awake for 24 hours, but then were allowed to sleep as much as they wanted. The control group was allowed to sleep as much as they wanted, whenever they wanted, during the study. At the start of the study, it was determined that both groups of students had roughly the same mean blood pressure on average. Three days after the start of the study, blood pressure was measured again.

Let μ_T represent the mean blood pressure at the end of the study of all students who might stay up all night, and μ_C represent the mean blood pressure of all students under usual sleeping conditions. The researchers' theory predicts that the sleep-deprivation will result in higher blood pressure, even three days later. To test this hypothesis, they compute a 95% confidence interval for $\mu_T - \mu_C$. This turns out to be (3.5, 16.8). Which of the following is **true** based on this study?

- (a) The data provide no evidence that sleep deprivation raises blood pressure.
- (b) The confidence interval is too wide for a valid comparison.
- (c) Based on this study we can conclude a causal relationship between sleep deprivation and blood pressure, as well as generalize our conclusions to all college students.
- (d) The researchers should conclude that sleep deprivation raises blood pressure.

- 7. (3) In a warehouse, employees have asked management to play music to relieve the boredom of the job. The manager wants to know whether efficiency is affected by the music. A random sample of 15 workers were selected. Their average efficiency score was 30.47 before the installation of the music system and 38 after its installation. We want to evaluate whether average efficiency score is affected by the music. What other information is required?
 - (a) Standard deviation of efficiency scores after the installation of the music system: s_{after}
 - (b) Standard deviation of efficiency scores before the installation of the music system: s_{before}
 - (c) Standard deviation of efficiency scores before and after the installation of the music system: s_{after} and s_{before}
 - (d) Standard deviation of differences in efficiency scores before and after: s_{diff}
- 8. A variety of studies suggest that 10% of the world population is left-handed. It is also claimed that artists are more likely to be left-handed. In order to test this claim we take a random sample of 40 art students at a college and find that 6 of them (15%) are left handed. Which of the following is the correct set-up for calculating the p-value for this test?
 - (a) Randomly sample 40 non-art students, and record the number of left-handed students in the sample. Repeat this many times and calculate the proportion of samples where at least 15% of the students are left-handed.
 - (b) Roll a 10-sided die 40 times and record the proportion of times you get a 1. Repeat this many times, and calculate the number of simulations where the sample proportion is 10% or more.
 - (c) Roll a 10-sided die 40 times and record the proportion of times you get a 1. Repeat this many times, and calculate the number of simulations where the sample proportion is 15% or more.
 - (d) In a bag place 40 chips, 6 red and 34 blue. Randomly sample 40 chips, with replacement, and record the proportion of red chips in the sample. Repeat this many times, and calculate the proportion of samples where at least 10% of the chips are red.
- 9. Which of the following is **false** about bootstrapping?
 - (a) Bootstrap distributions that are extremely skewed or have isolated clumps of values may yield unreliable confidence intervals.
 - (b) Bootstrap distributions are constructed by sampling with replacement from the original sample, while sampling distributions are constructed by sampling with replacement from the population.
 - (c) A bootstrap confidence interval constructed based on a biased sample will yield an unbiased estimate for the population parameter of interest.
 - (d) The endpoints of a 95% bootstrap confidence interval are the cutoff values for the top and bottom 2.5% of the bootstrap distribution.

- 10. (3) A random sample of 600 24-35 year-old unemployed Americans yielded an average unemployment of 13 weeks. In order to construct a bootstrap confidence interval based on this sample a statistician took 100 bootstrap samples and recorded their means. The standard deviation of these means was found to be 0.5. Which of the following is the correct 95% bootstrap interval?
 - (a) $13 \pm 1.96 \times 0.5$

(c) $13 \pm 1.96 \times \frac{0.5}{\sqrt{100}}$

(b) $13 \pm 1.96 \times \frac{0.5}{\sqrt{600}}$

(d) $13 \pm 1.96 \times \sqrt{\frac{0.5 \times 0.5}{600}}$

Answer questions 11 to 13 based on the information below.

Hepatitis C causes about 10,000 deaths each year in the US, but often lies undetected for year after infection. A study from University of Texas Southwestern Medical Center examined whether the risk of Hepatitis C was related to whether people had tattoos and to where they got their tattoos. The data from this study can be summarized in a two-way table, as follows:

	Hepatitis C	No Hepatitis C	Total
Tattoo, parlor	17	35	52
Tattoo, elsewhere	12	53	65
No tattoo	22	491	513
Total	51	579	630

- 11. (3) If in fact having Hepatitis C is independent of having a tattoo (and where one got their tattoo), how many people with no tattoos would you expect to have Hepatitis C? Choose the closest answer.
 - (a) 22

(c) 47

(b) 42

- (d) 491
- 12. (3) Which of the following is the appropriate test for evaluating the relationship between Hepatitis C and having a tattoo (and where one got their tattoo)?
 - (a) Z-test

(c) chi-square test of independence

(b) T-test

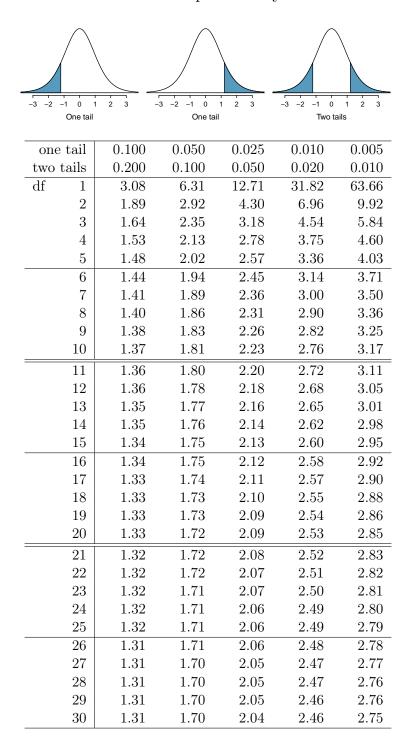
- (d) chi-square test of goodness of fit
- 13. (3) Based on your decision in question # 12, what is the degrees of freedom for the above test?
 - (a) 2

(c) 4

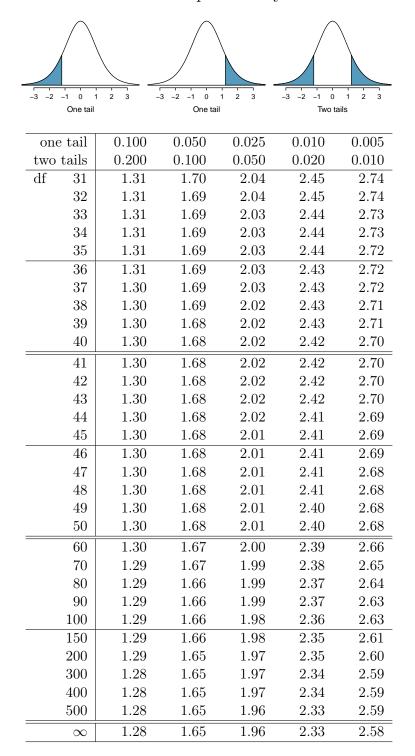
(b) 3

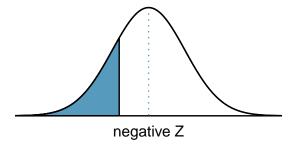
(d) none of the above.

t distribution probability table



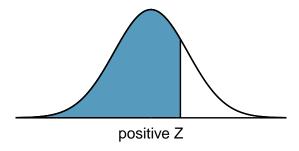
t distribution probability table





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

*For $Z \leq -3.50$, the probability is less than or equal to 0.0002.



				Seco	ond decim	al place	of Z			
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

^{*}For $Z \ge 3.50$, the probability is greater than or equal to 0.9998.