Sta 101 Spring 2016

Midterm 1 (Version A)

Dr. Çetinkaya-Rundel February 24, 2016

Last Name: _		First Name:						
Section: 8:30	10:05 11:45	1:25	3:05	4:40				Team Name:
=	hat I have not com n, and that all work			or gained	l information	tion in any	way fro	m my classmates
	Signatu	re:						
	violation of Duk Board. All work	_	-		- \ -		eported	to Undergrad-
1. You have 7	5 minutes to complet	e the exa	m.		\			
	our work on the open tions where no work					l credit. No	credit wil	l be given for open
you choose	nswers to the true/fa more than one answ asse questions.							
one $8\frac{1}{2}$ " × 1	owed a calculator, he 11" sheet of notes (cho me and the TA.	owever you neat sheet	u may n) with w	not share a vriting on b	calculator ooth sides,	with anoth pen or a pe	er student ncil, a dic	during the exam, tionary, and to ask
device or he	et allowed a cell phoreadphones, notes (other than myself and the	ner than y	our chea	at sheet), b				
6. Write clear	ly. Short answers are	best!		*				
			Goo	od luck!				
	4. ① ①	!	9. 🛆	B () (E 14	1. AB(
	5. ① ①			B © D		5. AB		
	6. ⊕ ⊕ 7. ⊕ ⊕			B () () () ()		6. AB(7. AB(
	8. ①①			B © 0		3. AB		
					T/F	MC		
		Q 1	Q 2	Q 3	Q 4 - 8	Q 9 - 18	Total	
	Points earned							
	Available points	30	20	10	10	30	100	



1. (30) Apple vs. FBI: Battle over unlocking phone.

Survey USA asked 500 randomly sampled San Franciscans the following question:

"The FBI has a court order demanding Apple help it unlock the iPhone belonging to the San Bernardino shooter. Apple says creating the custom software this would require would set a dangerous precedence and create a back door that, in the wrong hands, could potentially be used to unlock ANY iPhone. Do you think Apple should? Or Should not? Comply with the court order in this particular case?"

The distribution of responses by age group is shown below.¹

Age Group	18-34	35-49	50-64	65+	Total
Should	64	55	88	65	272
Should not	50	51	35	22	158
Not sure	32	12	15	11	70
Total	146	118	138	98	500

- (a) (4) In evaluating the relationship between opinion on whether Apple should comply with the court order or not and age, what is the response variable and what is the explanatory variable? Explain your reasoning in one sentence.
 - explanatory: (1) age group (age also acceptable)
 - response: (1) opinion
 - explanation: (1) More likely that age determines opinion on this issue, not the other way around

(1pt for explanatory and response, 2 pts for explanation)

(b) (1) What are the cases in this study?

(1pt - all or nothing) 500 randomly sampled San Franciscans

- (c) (4) Suppose a statistically significant relationship is found between opinion on whether Apple should comply with the court order or not and age,
 - i. to whom can the results be generalized?

(2pt - all or nothing) All San Franciscans (or San Francisans who are at least 18 years old)

ii. can a causal relationship be inferred? Explain your reasoning.

(2pt - all or nothing, no credit if they just said no, but didn't explain why) No, because this is an observational study.

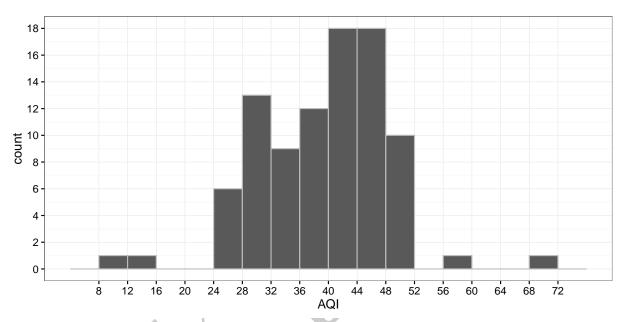
¹Source: Survey USA. Survey conducted on Feb 17, 2016. http://www.surveyusa.com/client/PollReport.aspx?g=3de782b2-2952-4527-8e8a-f922a0aa494e.

(d)	(5) Which of the following are appropriate visua	lizations for these data? Check all that apply.
	\square pie chart	\square scatterplot
	\square segmented bar plot	\square mosaic plot
	\square side-by-side box plots	
	(1pt for each, check as if T/F for each option	<i>ı)</i>
(e)	(8) Does there appear to be a relationship between the court order or not and age? Clearly state any	
	(4 pts for each of the probabilities (could be pts for dependent, 2 pts for some coherent and To answer this question we must calculate the affiliation:	
	ullet P(Should 18-34 year old) = 64 / 146 $pprox$ 0.4	4
	• $P(Should \mid 35-49 \ year \ old) = 55 / 118 \approx 0.4$	7
	• $P(Should \mid 50\text{-}64 \ year \ old) = 88 \mid 138 \approx 0.6$	34
	ullet P(Should 65+ year old) = 65 98 $pprox$ 0.66	
	Since these probabilities are different, there does whether Apple should comply with the court order	
(f)	(2) What is the probability that a randomly select comply with the court order?	ed <u>18-34 year old</u> believes that Apple <u>should not</u>
	(2 - all or nothing) P(Should not 18-34 year	r $old)=50 \ / \ 146 pprox \ 0.34$
(g)	(6) What is the probability that in a random san that Apple should not comply with the court of Using Binomial($n = 5$, $p = 0.34$): (5 - 2 for number of scenarios, 2 for prob of $P(K = 2) = \binom{5}{2}0.34^2 \times (1 - 0.34)^3 = 10 \times 0.34^2 \times (1 - 0.34)$	one scenario, 2 for final answer)

2. (20) Air quality in Durham.

Daily air quality is measured by the air quality index (AQI) reported by the Environmental Protection Agency. This index tells you how clean or polluted your air is, and what associated health effects might be a concern for you. The index is calculated for five major air pollutants regulated by the Clean Air Act, one of which is particulate matter. Values of this index range from 0 to 300 and a higher value indicates lower air quality.

AQI was reported for 90 randomly sampled days in Durham over the past 5 years.² The histogram below shows the distribution of these values. These 90 days represent a reasonably random and representative sample, and you may assume that the AQI levels for these days are independent.



(a) (2) Approximate the median AQI value.

(2 - all or nothing)

The median is in the 40-44 bin, best guess is 42, any value between 40 to 44 acceptable.

(b) (5) Determine whether this distribution has any outliers. If yes, note how many outliers they are and their approximate values.

(5 - 2 for IQR, 1 for fences, 2 for noting the correct outliers) Q1 = 34 (between 32 and 36) Q3 = 46 (between 44 and 48) IQR = 46 - 34 = 12 $Upper fence: 45 + 1.5 * 12 = 64 \rightarrow \text{ one outlier on the upper end, between } 68 \text{ and } 72$ $Lower fence: 34 - 1.5 * 12 = 16 \rightarrow \text{ two outliers on the lower end, one between } 8 \text{ and } 12, \text{ other between } 12 \text{ and } 16$

²Environmental Protection Agency - Air Data, http://aqsdr1.epa.gov/aqsweb/aqstmp/airdata/download_files.html.

(c) (5) We would like to use these data to construct a confidence interval for the average AQI in Durham. Evaluate, in context of the data, whether the conditions for inference are satisfied.

```
(5 - distribution shown below)
```

- Independence: The sample is random (1), and 90 is less than 10% of all days in the past 5 years (1), hence we can assume that the AQI in one sampled day is independent of another (2).
- Sample size / skew: The distribution of the sample is fairly symmetric and $n = 90 \ge 30$ (1), hence we can assume the population distribution is symmetric as well. Therefore, we would expect the sampling distribution of average AQI to be nearly normal (1).
- (d) (4) The average AQI for these 90 days is 39.26, with a standard deviation of 9.13. Construct a 90% confidence interval for the true average AQI in Durham.

(4 - 2 for SE, 1 for Z, 1 for final answer)
$$39.26 \pm 1.65 \frac{9.13}{\sqrt{90}} = 39.26 \pm 1.65 * 0.9624 = 39.26 \pm 1.59 = (37.67, 40.85)$$

(e) (2) Interpret the confidence interval in the previous part.

(2 - all or nothing)

We are 90% confident that the average AQI in Durham is between 37.67 and 40.85.

(f) (2) Below is an excerpt from the Air Quality Guide for Particle Pollution. Based on your results, what can you say about the air quality, with respect to particle pollution, in Durham?

AQI value	Condition	Risk			
0 - 50	Good	Little or no health risk			
51 - 100	Moderate	Moderate health concern for a small number of in			
		dividuals (people who are unusually sensitive may			
		experience respiratory symptoms)			
101 - 150	Unhealthy for	Greater risk for people with heart or lung disease,			
	sensitive groups	older adults, and children, but the general public is			
		unlikely to be affected			
	•••				

(2 - all or nothing, look for some reasonable answer)

While daily values range up to the moderate category, the confidence interval for average AQI value is in the good category. We are 95% confident that the average AQI value poses little or no risk posed to human health due to particle pollution in Durham.

3. (10) Duke or UNC?.

A February 2016 Public Policy Polling survey asked respondents whether they will root for Duke or UNC on the next basketball game between these two teams.³ The responses are summarized below.

Sex	Man	Woman	Total
Duke	188	178	366
UNC	225	287	512
Total	413	465	878

(a) (2) We would like to test if there is a difference between the proportions of men and women who will root for Duke on the next basketball game between these two teams using a randomization test. What are the hypotheses?

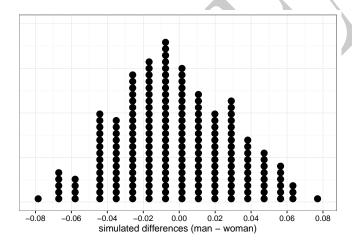
Let p be the proportion who root for Duke.

 $H_0: p_M = p_F$ $H_A: p_M \neq p_F$

(b) (3) Calculate the observed sample statistic.

 $\hat{p}_M = 188/413 = 0.455$ $\hat{p}_W = 178/465 = 0.383$ $\hat{p}_M - \hat{p}_W = 0.455 - 0.383 = 0.072$

(c) (2) The distribution of <u>200 simulated differences</u> in support for Duke $(\hat{p}_M - \hat{p}_F)$ from a randomization test for these hypotheses is shown below. Calculate the p-value.



(2 - only 1 if given one sided, only 1 if divided by 100) 2 / 200 = 0.01

(d) (3) Using $\alpha = 0.05$, interpret the conclusion of this test in context of the data and the hypotheses.

Reject H_0 . The data provide convincing evidence of a difference between the proportions of men and women who will root for Duke on the next basketball game between these two teams.

³Public Policy Polling, Feb 2016, http://www.publicpolicypolling.com/pdf/2015/PPP_Release_NC_21716.pdf.

True / **False** Determine if the following statements are true or false. Fill in the bubbles on the first page of the exam. Each question is worth 2 points.

- 4. (T / F) In order for the sampling distribution of a mean to be nearly normally distributed, the expected numbers of successes and failures must be at least 10. F
- 5. (T / F) Random sampling is useful for finding evidence of a causal relationship. F
- 6. (T / F) If the population distribution is highly skewed, then the central limit theorem will never apply to the sampling distribution of sample means. F
- 7. (T / F) Suppose a researcher sets the significance level at 5% prior to conducting an experiment. After conducting the study, the researcher computes a p-value of 0.04. The researcher can conclude that the null hypothesis is false. F
- 8. (T / F) If a given value (for example, the null hypothesized value of a parameter) is within a 90% confidence interval, it will definitely also be within a 95% confidence interval calculated based on the same sample. T

Multiple Choice 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.

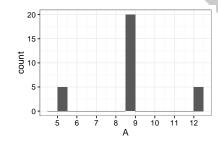
9. A teacher brags that a majority of her students scored above the mean on her last exam. This means

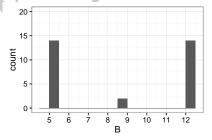
- (a) the teacher is crazy because this is not possible
- (b) the distribution of scores was right-skewed
- (c) the distribution of scores was left-skewed
- (d) the distribution of scores was symmetric
- (e) the distribution of scores was uniform

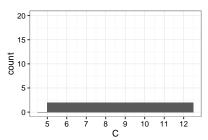
10. A 2015 poll conducted by the Pew Research Center found that 15% of American adults have used online dating sites or mobile dating apps.⁴ What is the probability that in a random sample of 20 adults, at least one has used online dating sites or mobile dating apps?

- (a) 0.0388
- (b) 0.0075
- (c) 0.1368
- (d) *0.9612*
- (e) 1

11. Order histograms A, B, and C from most to least variable.







- (a) $s_C > s_A > s_B$
- (b) $s_B > s_C > s_A$
- (c) $s_A > s_B > s_C$
- (d) $s_C > s_B > s_A$
- (e) $s_A > s_C > s_B$

 $^{^4} Pew \ Research \ Center, \ February, \ 2016, \ ``15\% \ of \ American \ adults \ have \ used \ online \ dating \ sites \ or \ mobile \ dating \ apps\'Ohttp://www.pewinternet.org/files/2016/02/PI_2016.02.11_Online-Dating_FINAL.pdf.$

- 12. Public policy researchers are studying whether a new school lunch program reduces the incidence of obesity amongst elementary school children. The authors compute the p-value for their sample to be 0.10. Which of the following interpretations of the p-value is correct?
 - (a) The probability that the policy is effective.
 - (b) The probability that the policy is **not** effective.
 - (c) The probability of determining the policy is not effective when it actually is.
 - (d) The probability of getting results as extreme or more extreme than the ones in the study if the policy is actually effective.
 - (e) The probability of getting results as extreme or more extreme than the ones in the study if the policy is actually <u>not</u> effective.
- 13. A low-calorie snack company claims that, on average, there are 95 calories in its small bag of crackers. A consumer watchdog group is concerned that the company may be lying and that the population average is not actually 120 calories. An employee of the group samples 45 bags of chips and computes a 95% confidence interval of (92,118). When the employee reports her results, the head of the group says that she wanted to do inference using a 99% confidence interval. Which of the following is the best response the employee can give?
 - (a) It is standard practice to use a 95% confidence interval, so it does not make sense to use a 99% confidence interval.
 - (b) We will reject the null hypothesis at the 99% confidence level.
 - (c) We will fail to reject the null hypothesis at the 99% confidence level.
 - (d) We will accept the null hypothesis at the 99% confidence level.
 - (e) I need to recompute the confidence interval and get back to you.
- 14. The distribution of lengths of adult bass in Cumberland Lake is approximately normal with mean 32" and standard deviation 6".

At the annual Cumberland Lake bass fishing competition, you win a blue ribbon if you catch a bass that is over 38" in length. If you catch a bass over 42" in length you also win a gold medallion.

Assume that an angler (the person fishing) is only allowed to catch the first fish s/he reels in.

Which of the following is the closest to the probability that an angler has not won a gold medallion given that s/he has won a blue ribbon?

- (a) 0.05
- (b) 0.11
- (c) 0.16
- (d) 0.30
- (e) **0.70**

- 15. The Chicago Tribune and the Los Angeles Times conducted separate national polls where they asked full-time employees how many hours they work per week, and reported confidence intervals at the 95% confidence level. The Chicago Tribune surveyed 500 people, and the Los Angeles Times surveyed 300 people. Which paper reported a larger margin of error? Assume the standard deviations of the two samples were equal.
 - (a) The Chicago Tribune
 - (b) The Los Angeles Times
 - (c) The margin of errors are the same
 - (d) The margins of errors will depend on the means of the samples
 - (e) There is not enough information to answer this question

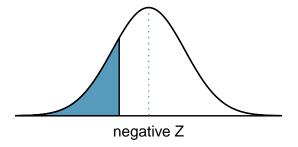
- 16. A professor gives a test to 140 students and determines the median score. After grading the test, she realizes that the 10 students with the highest scores did exceptionally well. She decides to award these 5 students a bonus of 10 more points. The median of the new score distribution will be _____ that of the original score distribution.
 - (a) lower than
 - (b) equal to
 - (c) higher than
 - (d) depending on skewness, higher or lower than
 - (e) depending on modality, higher or lower than

In 2015 the average credit card debt per US household was about \$15,355.⁵ Some households had more debt, and some had less. Assume the distribution of household debt is somewhat right skewed with mean \$15,355 and standard deviation \$10,000.

- 17. What is the probability that a randomly sampled household has more than \$16,000 in credit card debt?
 - (a) 0.0645
 - (b) 0.4743
 - (c) 0.5257
 - (d) 0.9355
 - (e) Cannot be calculated with the information given

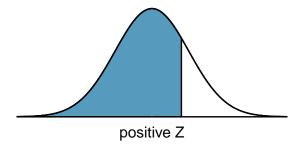
- 18. What is the probability that a random sample of 1,000 households have more than \$16,000 on average in credit card debt?
 - (a) **0.0207**
 - (b) 0.4743
 - (c) 0.5257
 - (d) 0.9793
 - (e) Cannot be calculated with the information given

 $^{^5\}mathrm{NerdWallet},\ \mathtt{http://www.nerdwallet.com/blog/credit-card-data/average-credit-card-debt-household/}.$



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.



Second decimal place of ZZ0.00 0.01 0.02 0.03 0.04 0.050.06 0.070.080.09 0.0 0.5000 0.5040 0.50800.5120 0.5160 0.51990.5239 0.5279 0.5319 0.5359 0.5636 0.5398 0.54380.54780.55170.55570.55960.56750.57140.10.57530.57930.58320.58710.59100.59480.5987 0.6026 0.61030.20.60640.61410.3 0.6179 0.62170.62550.62930.6331 0.6368 0.6406 0.6443 0.64800.65170.40.65540.65910.66280.66640.67000.67360.67720.68080.68440.68790.50.69150.69500.69850.70190.70540.70880.71230.71570.71900.72240.73570.74220.74540.75170.60.72570.72910.73240.73890.74860.75490.75800.76110.76420.76730.77040.77640.77940.78230.78520.70.77340.7881 0.7939 0.7967 0.8106 0.8 0.79100.79950.8023 0.80510.8078 0.8133 0.9 0.81590.81860.82120.82380.82640.82890.83150.83400.83650.83891.0 0.8413 0.8438 0.8461 0.84850.85080.8554 0.85770.85990.8621 0.85310.86861.1 0.86430.86650.8708 0.8729 0.8749 0.8770 0.87900.8810 0.8830 1.2 0.8849 0.8869 0.88880.89070.8925 0.8944 0.8962 0.8980 0.8997 0.9015 0.9099 0.9131 1.3 0.90320.90490.90660.90820.91150.91470.91620.91771.4 0.91920.92070.9222 0.92360.92510.92650.92790.92920.93060.93190.9345 0.93570.9370 0.9406 0.9418 0.9429 1.5 0.9332 0.93820.93940.94410.94520.9463 0.94740.94840.94950.95050.95150.95250.95350.95451.6 0.95540.95640.9573 0.95820.95910.95990.9608 0.96251.7 0.96160.96331.8 0.96410.96490.96560.96640.96710.96780.96860.96930.96990.97060.97130.97190.97320.97440.97500.97560.97611.9 0.97260.97380.97672.0 0.9772 0.9778 0.9783 0.9788 0.9793 0.9798 0.9803 0.9808 0.9812 0.9817 0.9821 0.98340.98420.9854 2.1 0.98260.9830 0.9838 0.98460.98500.9857 0.98712.2 0.98610.98640.98680.98750.98780.98810.98840.98870.98902.3 0.98930.98960.98980.99010.99040.99060.99090.9911 0.99130.99162.4 0.99180.99200.99220.99250.99270.99290.99310.99320.99340.99362.5 0.9940 0.9941 0.9943 0.99450.9948 0.9949 0.9951 0.99380.99460.99522.6 0.99530.99550.9956 0.99570.99590.99600.9961 0.99620.99630.99642.7 0.99650.99660.99670.99680.99690.99700.99710.99720.99730.99742.8 0.99740.99750.9976 0.9977 0.9977 0.99780.9979 0.9979 0.99800.99812.9 0.9981 0.99820.99820.99830.99840.99840.99850.99850.99860.99863.0 0.9987 0.9987 0.9987 0.9988 0.9988 0.99890.9989 0.9989 0.9990 0.9990 3.1 0.99900.9991 0.9991 0.99910.99920.99920.99920.99920.99930.99933.2 0.99930.99930.99940.99940.99940.9994 0.99940.99950.99950.99953.3 0.99950.99960.99960.99950.99950.99960.99960.99960.99960.99970.9997 0.9997 0.9997 0.99970.9997 0.9997 0.99970.9997 0.9997 0.9998

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