Statistics Individual

AoPS Mu Alpha Theta

February 23 - March 9, 2019

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	**0.50**	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
two-tails	**1.00**	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

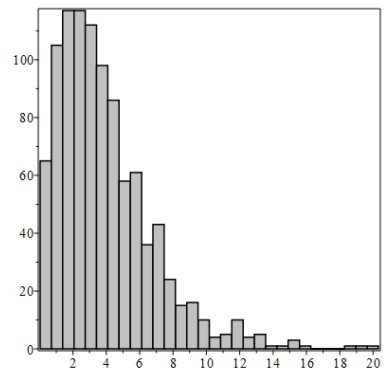
The answer choice **(E)** NOTA denotes that "none of these answers" are correct. **Round all decimal answers to the number of decimal places suggested by the answer choices.** A t-table is attached for your convenience; please use it to calculate z and t values. Tests are scored such that each correct answer is worth 4 points, each question left blank is worth 0 points, and each incorrect answer is -1 point. All answers must be exact unless otherwise specified. Good luck, and more importantly, have fun!

---

1. Which of the following is a discrete quantitative variable?

(A) Height **(B)** Weight **(C)** ZIP code **(D)** Shoe size **(E)** NOTA

For the following three problems, consider a set S, constructed from a random sample of 1000 points from a χ^2 distribution with 4 degrees of freedom. A histogram is provided to the right.



2. Which of the following is closest to the mean of S?

(A) 1 **(B)** 4 **(C)** 7
(D) 10 **(E)** NOTA

3. Which of the following sample statistics would be most fitting to report the center and spread of S? Let $\overline{x}$, s_x, M, and IQR represent the sample mean, sample standard deviation, sample median, and sample interquartile range, respectively.

(A) $\overline{x}$ and s_x **(B)** $\overline{x}$ and IQR **(C)** M and s_x **(D)** M and IQR **(E)** NOTA

4. Which of the following is most likely to be true about distributions similar to that of S?

(A) $\overline{x} > M$ **(B)** $\overline{x} < M$ **(C)** $\overline{x} \approx M$ **(D)** $\overline{x} = M$ **(E)** NOTA

5. The function $p(x) = 2x$ is a continuous probability distribution for $x \in [0, 1]$. What is the median of this probability distribution?

(A) $\frac{1}{2}$ **(B)** $\frac{2}{3}$ **(C)** $\frac{\sqrt{2}}{2}$ **(D)** $\frac{3}{4}$ **(E)** NOTA

6. What is the mean of the probability distribution described in the previous question?

(A) $\frac{1}{2}$ **(B)** $\frac{2}{3}$ **(C)** $\frac{\sqrt{2}}{2}$ **(D)** $\frac{3}{4}$ **(E)** NOTA

7. Mr. Sofy gives a very hard test which had a mean score of 60 with a standard deviation of 20. He curves the test so that all students' z-scores are preserved, and the new mean and standard deviation are 80 and 10 respectively. Joanne initially had an 80 on the test. What is her score after the curve?

(A) 85 **(B)** 90 **(C)** 95 **(D)** 100 **(E)** NOTA

8. A normal probability plot of a set of data resembles the graph of $y = x^2$ from $x = 1$ to $x = 3$. Which of the following describes the distribution of this set of data?

(A) Skewed left **(B)** Approximately Normal **(C)** Skewed Right
(D) Approximately Uniform **(E)** NOTA

For the following two problems, there appears to be a linear relationship between the number of students on the American Heritage campus and the number of ducks on the American Heritage campus at any given time. From a random sample of 30 days, the mean and standard deviation of the number of students on campus was 1000 and 100, respectively. From the same random sample, the mean and standard deviation of the number of ducks on campus was 50 and 10, respectively. The observed correlation coefficient is -0.8.

9. What is the equation of the least-squares regression line if the number of ducks is the explanatory variable and the number of students is the response variable?

(A) $\hat{y} = 10x + 500$ **(B)** $\hat{y} = -10x + 1500$ **(C)** $\hat{y} = 8x + 600$
(D) $\hat{y} = -8x + 1400$ **(E)** NOTA

10. One day, 20 ducks and 1200 students were observed on campus. What is the residual of this point?

(A) -100 **(B)** -60 **(C)** 0 **(D)** 60 **(E)** NOTA

11. Which of the following statements is always true regarding linear regression and correlation?

(A) A correlation of 0 implies that there is no relation between the two variables.
(B) A correlation close to ± 1 implies there is a linear relation between the two variables
(C) Switching the explanatory and response variables changes the sign of the correlation coefficient.
(D) If a set of data initially has a correlation close to 1, the transformation $(x, y) \mapsto (x, y^2)$ has a correlation much closer to 0.
(E) NOTA

12. A public interest group wishes to obtain data about people of different nationalities and their opinions on a particular issue. To avoid undercoverage, which sampling method should the group use?

(A) Simple Random Sample **(B)** Stratified Random Sample **(C)** Cluster Sample
(D) Systematic Sample **(E)** NOTA

13. A medical company ran a clinical trial of a new drug using a double blind, placebo-controlled experiment which yielded statistically significant results. The subjects consisted of 50 adult volunteers who were randomly allocated either the actual treatment or a placebo. Which generalization(s) can the company make about their findings about cause-and-effect and connection to the population of all adults?

(A) They can establish cause-and-effect and generalize to the population.
(B) They can establish cause-and-effect, but cannot generalize to the population.
(C) They cannot establish cause-and-effect, but they can generalize to the population.
(D) They can neither establish cause-and-effect nor generalize to the population.
(E) NOTA

14. A botanist wishes to determine which combination of room temperature and daily water intake maximizes plant growth. She tests two different temperatures and three different volumes of water. Find the number of treatments in this experiment.

(A) 2 **(B)** 3 **(C)** 5 **(D)** 6 **(E)** NOTA

15. Consider the χ^2 distribution with 4 degrees of freedom from earlier in this test. If many samples of 40 randomly selected points in this distribution were selected and their sample means were recorded, what would the shape of the distribution of the sample means be?

(A) Approximately Normal **(B)** Approximately Uniform **(C)** Skewed Right
(D) Skewed Left **(E)** NOTA

I don't know about you, but this test feels kind of boring at the moment. Let's make this more fun! Remember, use the given t-table to calculate z and t statistics.

16. Vincent and Jae play, on average, 2 and 3 hours of League of Legends every day, respectively. Their respective percentiles in playing time relative to the rest of the Alpha division are 80th and 90th. Find the sum of the mean and standard deviation of the mean playing times of all students in the Alpha division, assuming that the population is normally distributed. (Assume that, somehow, negative playing time is possible. Don't ask how; I'm not sure).

(A) 0 **(B)** 0.1 **(C)** 2.3 **(D)** 2.4 **(E)** NOTA

17. Logan has been practicing his free throws, and has increased his success rate to an all-time high of 1.2%. What is the probability that, in his first 10 throws, he makes exactly one of them?

(A) 0.10 **(B)** 0.11 **(C)** 0.12 **(D)** 0.13 **(E)** NOTA

18. What is the expected number of free throws that Logan needs to make a basket? Round your answer to the nearest integer.

(A) 83 **(B)** 84 **(C)** 85 **(D)** 86 **(E)** NOTA

19. Jin wishes to figure out what proportion of AoPS users listen to K-pop music. She randomly samples 100 AoPS users and finds that 18 of them listen to K-pop music. What is the 95% confidence interval for the proportion of AoPS users that listen to K-pop music?

(A) (0.105, 0.255) **(B)** (0.104, 0.256) **(C)** (0.082, 0.278)
(D) (0.081, 0.279) **(E)** NOTA

20. Jackson is interested in estimating the mean number of times Dr. Santos says "Um" during a calculus lecture. He tracks the number of "Um"s said for 10 lectures and obtains the following values.

64	62	61	59	63	60	61	58	62	66

What is the 90% confidence interval for the true mean number of "Um"s Dr. Santos says during a calculus lecture? Assume all conditions for inference are met. (To help you out, $s_x = 2.366$)

(A) (60.9, 62.3) **(B)** (60.4, 62.8) **(C)** (59.9, 63.3) **(D)** (59.4, 63.8) **(E)** NOTA

21. Daniel randomly surveyed 101 American Heritage students and asked them to rate how strongly they agree with the statement "Lemonade is a juice." Students responded with a number from 1-100. He received a sample mean of 62 and a sample standard deviation of 8.502. What is the P-value of the appropriate test to see if the mean rating is different from his initial guess of 60?

(A) 0.01 **(B)** 0.02 **(C)** 0.03 **(D)** 0.04 **(E)** NOTA

22. If the power of the test described in the previous question is 0.80, and the alpha level was 0.05, which of the following is the closest to the true mean rating of American Heritage students of their agreement with the statement that lemonade is a juice?

(A) 59 **(B)** 61 **(C)** 63 **(D)** 65 **(E)** NOTA

23. Which of the following could **not** have occurred during the test referred to in the previous two problems?

(A) Type-I error **(B)** Type-II error **(C)** Undercoverage
(D) Response bias **(E)** NOTA

24. Connor opened a large bag of M&Ms and recorded the frequencies of the different colors. To see if the proportions of each color match what the company claims, he conducts a χ^2 goodness-of-fit test and obtains a test statistic of 10. If he then doubles the sample size, obtains the same proportions as his original sample, and conducts another test, what will the new test statistic be?

(A) 5 **(B)** 10 **(C)** 20 **(D)** 40 **(E)** NOTA

25. Ria conducted a significance test to determine if listening to music while studying causes a student to score lower on a test than they normally would. She obtains a P-value of 0.03. What is the correct interpretation of the P-value?

(A) Students who listen to music while studying perform 3% worse, on average.
(B) Students who listen to music while studying earn scores that are 3 points lower, on average.
(C) Students who listen to music while studying have a 3% chance to score lower.
(D) Students who do not listen to music have a 3% chance to score higher.
(E) NOTA

26. Oleksii, the stats god, takes three stats tests at different times, on which he earns three different scores. What is the probability that his scores improve over time, assuming all three tests are independent, of the same difficulty, etc?

(A) $\frac{1}{8}$ **(B)** $\frac{1}{6}$ **(C)** $\frac{1}{4}$ **(D)** $\frac{1}{3}$ **(E)** NOTA

27. Kev and Adam are walking on coordinate grid. Kev starts at $(0, 0)$ and is only able to move along grid lines up and to the right. Adam starts at $(7, 3)$ and is only able to move along grid lines down and to the left. If they walk at the same speed, the probability that they meet each other can be expressed in simplest form as $\frac{p}{q}$. What is $p + q$?

(A) 5 **(B)** 31 **(C)** 143 **(D)** 625 **(E)** NOTA

28. Surya has a box containing three coins. Two of the coins are fair, but one of the coins is weighted such that it lands heads with probability 0.6. He selects a coin from the box at random and flips it twice, and it lands heads both times. If he flips the coin a third time, the probability it lands heads can be expressed in simplest form as $\frac{p}{q}$. What is $p + q$?

(A) 3 **(B)** 61 **(C)** 67 **(D)** 663 **(E)** NOTA

29. Michelle and Carol are having a dice rolling competition. They take turns rolling a pair of dice until one of them gets a pair of ones. Whoever gets this pair first wins. If Michelle goes first, the probability that she wins can be expressed in simplest form as $\frac{p}{q}$. What is $p + q$?

(A) 105 **(B)** 106 **(C)** 107 **(D)** 108 **(E)** NOTA

30. Arnav has an unfair coin such that the expected number of flips needed to get 2 consecutive heads is the same as the expected number of flips to get 2019 (not necessarily consecutive) tails. If the square of the probability of the coin landing heads can be expressed in simplest form as $\frac{p}{q}$, what is $p + q$?

(A) 2018 **(B)** 2019 **(C)** 2020 **(D)** 2021 **(E)** NOTA