Statistics Exam
2009 University of Houston Math Contest

Name: __________________________________

School: __________________________________

Please read the questions carefully and give a clear indication of your answer on each question.

There is no penalty for guessing.

Judges will use written comments and/or calculations to settle ties.
Good luck.

Statistics Exam

NAME:____

SCHOOL:____

You may use a calculator and the attached tables and formula sheets.

Part 1: Multiple Choice. 6 points each.

1. The heights of American males aged 18 to 24 years are normally distributed with a mean of 68 inches and a standard deviation of 2.5 inches. 10% of all young American men have heights greater than or equal to

   (a) 71.2 inches

   (b) 64.8 inches

   (c) 73.0 inches

   (d) The 10^{th} percentile of heights

   (e) The coefficient of variation of heights.

2. Which of the following is an example of a binomial random variable?

   (a) The number of traffic accidents reported in a 24 hour period.

   (b) The number of throws of a pair of dice before a seven is rolled.

   (c) The number of students who get their own test papers when 20 tests are scrambled and randomly returned to the students in a class.

   (d) The number of pet owners in a random sample, without replacement, of students in a first-grade class.

   (e) The number of AP Calculus students in a random sample, with replacement, of 100 students in a high school.

NAME:____

SCHOOL:____

3. In estimating a population mean or proportion, an advantage of stratified sampling over simple random sampling is

   (a) Inferences based on the assumption of a normal distribution are more accurate.

   (b) For a given total sample size, the variance of the estimator is smaller if allocation is done properly.

   (c) It is less expensive.

   (d) It removes any bias in the estimate.

   (e) It is suitable for any population, whereas simple random sampling is effective only for very large populations.

4. Students in a math course took a prerequisite test at the beginning of the semester. At the end of the semester the distributions of test scores for students earning each letter grade were obtained. Boxplots (box and whisker diagrams) of prerequisite test scores for each letter grade are given below. Which of the following statements is true?

   (a) There are extreme outliers in test scores for F students.

   (b) The distribution of test scores for B students is symmetric.

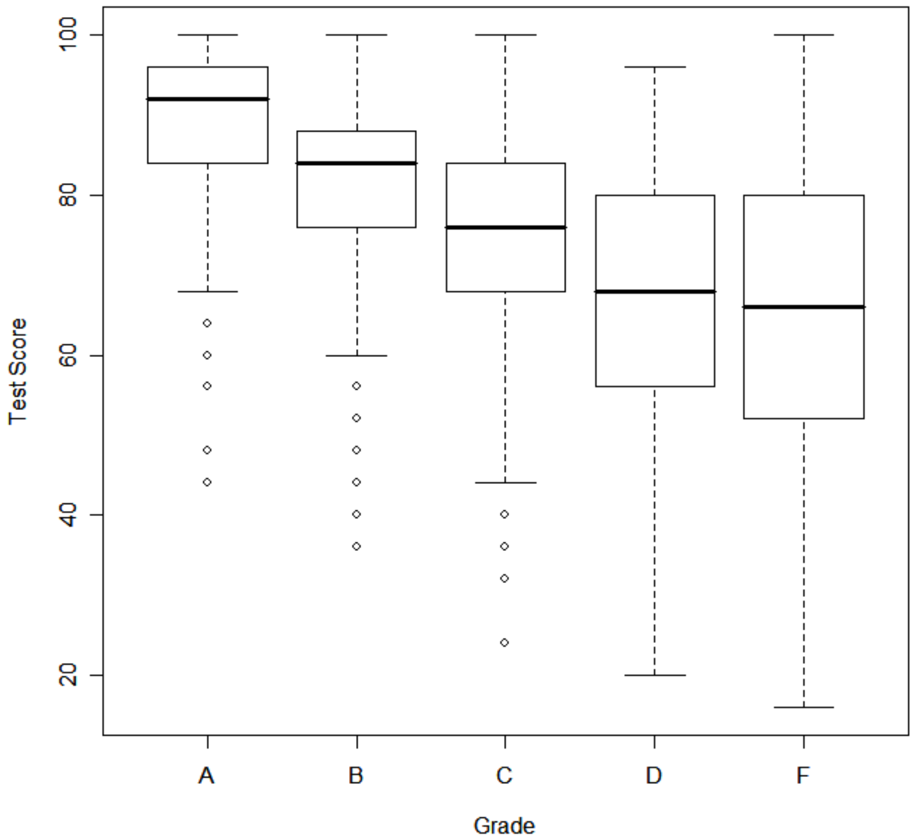
   (c) 75% of the A students did as well as or better than half of the B students.

   (d) 75% or more of all D students scored below 60.

   (e) The interquartile range for A students is the same as the interquartile range for C students.

NAME:____

SCHOOL:____



5. The figure below is a histogram of 999 measurements. The median of the data is closest to:

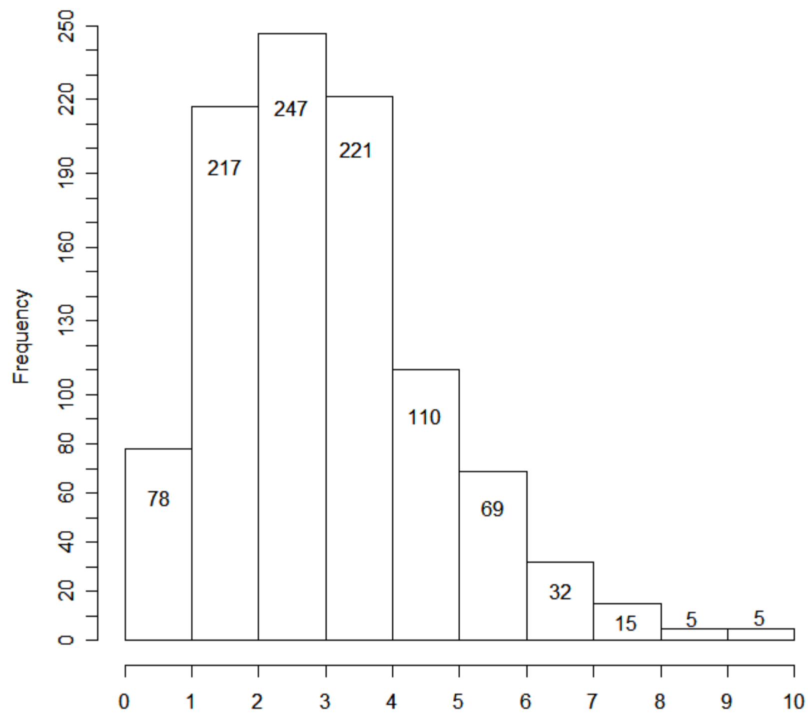
(a) 4.5

(b) 4.0

(c) 247

(d) 2.5

(e) Since the data is not given, there is no way to estimate the median.



6. An interplanetary space probe is designed to set down at a certain point on Mars. Because of random perturbations it will miss its planned impact point. The square of the distance from the target point has a chi-squared distribution with two degrees of freedom. Which of the following is closest to the probability that it will land within 2.45 units of distance from its target?

(a) 0.10

(b) 0.05

(c) 0.95

(d) 0.50

(e) 0.99

NAME:____

SCHOOL:____

7. Scores on a nationwide professional qualifying exam are normally distributed with a population mean of 800 and a population standard deviation of 50. An examinee scored at the 60th percentile nationally. To the nearest whole number, what was the examinee's numerical score?

   (a) 945

   (b) 800

   (c) 688

   (d) 858

   (e) 813

8. Concerned officials know that the number of hours students spend weekly on MySpace is normally distributed with a standard deviation of 3 hours. They would like to monitor a sample of students to estimate the mean amount of time spent. They would like to have an error no greater than 20 minutes with 90% confidence. The minimum number of students needed for their sample is

   (a) 81

   (b) 1162

   (c) 380

   (d) 220

   (e) 149

NAME:____

SCHOOL:____

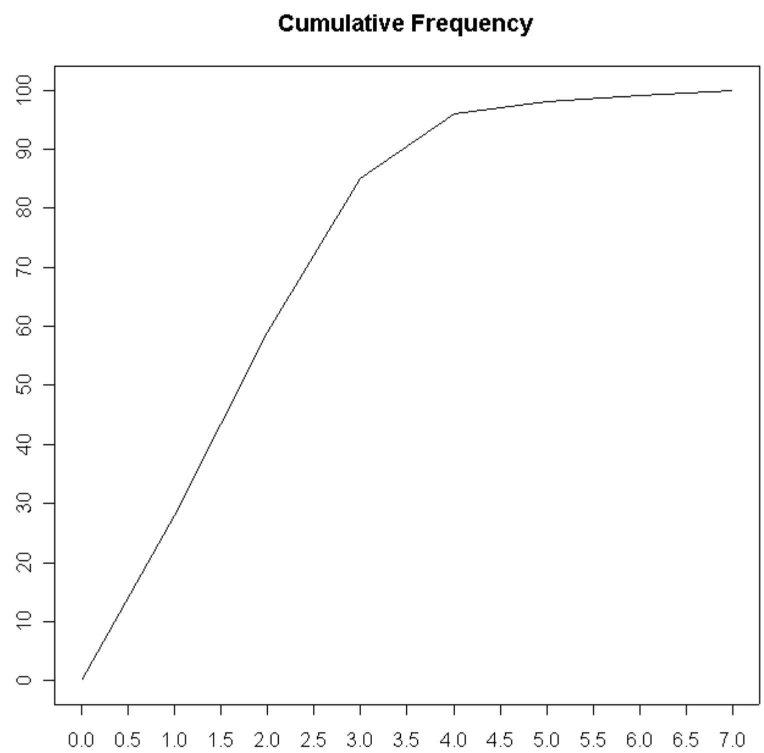
Part 2: Free Response. 12 points each. Answer in the space provided. Show your work. Three place accuracy is sufficient for decimal answers.

9. The incidence of a disease in a population is 3%. A diagnostic test gives false positive results 20% of the time and false negative results 2% of the time. Given that the test is positive, what is the probability that the person tested actually has the disease?

10. The diagram on the next page is a relative frequency polygon of 100 measurements. Estimate the median, the quartiles, and the interquartile range. An outlier is defined as an observation that is more than 1.5 times the interquartile range from the nearest quartile. Are there any outliers in this data?

NAME:____

SCHOOL:____

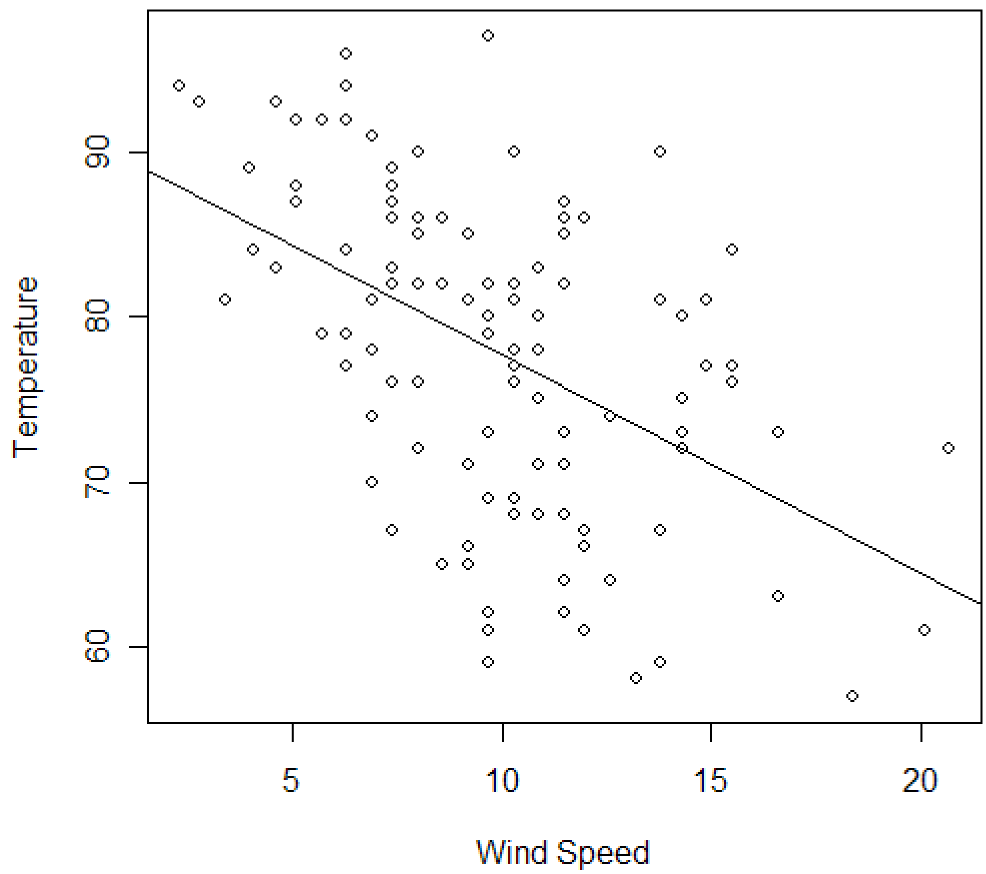


11. Inspection of 50 randomly selected houses in Houston revealed that 8 of them had termite damage. Find a 95% confidence interval for the proportion of all houses in Houston that have termite damage.

NAME:____

SCHOOL:____

12. The two sets of software output below are from R and Excel and both apply to the data in the scatter diagram. Use either one to answer the following questions.



(a) What is the equation of the fitted least squares line? Let x stand for Wind Speed and y for Temperature.

(b) What is the predicted temperature when the wind speed is 15?

(c) What fraction of the total variation in temperature is attributable to the linear relationship and the variation in wind speed?

(d) What is the estimated standard deviation of random error?

R output

```
> summary(temp.lm)

Call:
lm(formula = Temp ~ Wind, data = Air)

Residuals:
  Min       1Q        Median     3Q      Max
-19.112   -5.646      1.014    6.254   18.888

Coefficients:
             Estimate   Std. Error   t value   Pr(>|t|)
(Intercept)  91.0305    2.3489       38.754    < 2e-16 ***
Wind         -1.3318    0.2226       -5.983    2.84e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1  ' ' 1

Residual standard error: 8.306 on 109 degrees of freedom
Multiple R-squared: 0.2472,     Adjusted R-squared: 0.2403
F-statistic: 35.79 on 1 and 109 DF,   p-value: 2.842e-08
```

Excel output (continued on the next page)

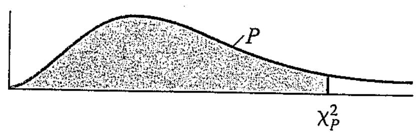
Regression Statistics						
Multiple R	0.49718972					
R Square	0.24719761					
Adjusted R Square	0.24029117					
Standard Error	8.30644282					
Observations	111					

ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	2469.562064	2469.6	35.79	2.84197E-08
Residual	109	7520.67217	68.997		
Total	110	9990.234234			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	91.030517	2.348945166	38.754	1E-65	86.3749842	95.6860498
Wind	-1.3318113	0.222611603	-5.983	3E-08	-1.77302023	-0.8906023

TABLE 3

Percentiles of the χ^2 Distribution—Values of χ^2_P Corresponding to P



df	$\chi^2_{.005}$	$\chi^2_{.01}$	$\chi^2_{.025}$	$\chi^2_{.05}$	$\chi^2_{.10}$	$\chi^2_{.90}$	$\chi^2_{.95}$	$\chi^2_{.975}$	$\chi^2_{.99}$	$\chi^2_{.995}$
1	.000039	.00016	.00098	.0039	.0158	2.71	3.84	5.02	6.63	7.88
2	.0100	.0201	.0506	.1026	.2107	4.61	5.99	7.38	9.21	10.60
3	.0717	.115	.216	.352	.584	6.25	7.81	9.35	11.34	12.84
4	.207	.297	.484	.711	1.064	7.78	9.49	11.14	13.28	14.86
5	.412	.554	.831	1.15	1.61	9.24	11.07	12.83	15.09	16.75
6	.676	.872	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	.989	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
24	9.89	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	51.81	55.76	59.34	63.69	66.77
60	35.53	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38	91.95
120	83.85	86.92	91.58	95.70	100.62	140.23	146.57	152.21	158.95	163.64

For large degrees of freedom,

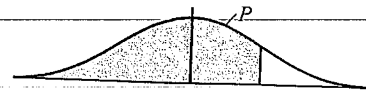
$$\chi^2_P = \tfrac{1}{2}(z_P + \sqrt{2v - 1})^2 \text{ approximately,}$$

where v = degrees of freedom and z_P is given in Table 2.

.99
.000
.000
.000
.000
.000
.000
.000
.000
.000
.000
.000
.000
.000
.000
.000
.000
.000
.000
.000
.000
.000
.000
.002
.026
.222

TABLE 2

Cumulative Normal Distribution—Values of P Corresponding to z_p for the Normal Curve



z is the standard normal variable. The value of P for $-z_p$ equals 1 minus the value of P for $+z_p$; for example, the P for -1.62 equals $1 - .9474 = .0526$.

z_p	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Formulas

Note: As far as possible, these expressions use notation that is common in many statistics textbooks, but in almost every case there are some books that use different notation. Also, the same symbol might have different meanings in different expressions. Most of these will not be needed on the test.

1. $\chi^2 = \sum_{i=1}^{m} \frac{(O_i - E_i)^2}{E_i} = \sum_{i=1}^{m} \frac{(O_i - np_i)^2}{np_i}$ is distributed as $\chi^2(m-1)$.

2. $\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

3. $\bar{x} - \bar{y} \pm t_{\alpha/2} s_p \sqrt{\frac{1}{m} + \frac{1}{n}}$

4. $\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

5. $s_p^2 = \frac{(m-1)s_x^2 + (n-1)s_y^2}{m+n-2}$

6. $s_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$

7. $s^2 = \frac{SS(resid)}{n-2}$

8. $z = \frac{x-\mu}{\sigma}$

9. $z = \frac{\sqrt{n}(\bar{x}-\mu)}{\sigma}$

10. $\hat{\mu}(x) = \bar{y} + \hat{\beta}(x - \bar{x})$

11. $\hat{\beta} = \frac{S_{xy}}{S_{xx}}$

12. $S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$

13. $S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$

14. $S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2$

15. $SS(resid) = \sum_{i=1}^{n} (y_i - \hat{\mu}(x_i))^2 = S_{yy} - \hat{\beta}^2 S_{xx}$

16. $\hat{\mu}(x) \pm t_{\alpha/2}(n-2)\, s\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}$

17. Reject $H_0 : \sigma^2 = \sigma_0^2$ in favor of $H_1 : \sigma^2 < \sigma_0^2$ if $s^2 < \sigma_0^2 \frac{\chi^2_{1-\alpha}(n-1)}{n-1}$.

18. $n > z^2_{\alpha/2} \frac{1}{4\varepsilon^2}$

19. $n > z^2_{\alpha/2} \frac{\sigma^2}{\varepsilon^2}$

20. Reject $H_0 : \mu = \mu_0$ in favor of $H_1 : \mu > \mu_0$ if $\bar{x} > \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$.