AP Stats Exam Review Answers

1a) Min: -34.04255
Q1: -2.95
Med: 3.4691
Q3: 8.4511
Max: 58.67769

1b) The data is roughly symmetric and approximately normal with the median at 3.47 and IQR of 11.4. There are several outliers to the left and right.

1c) IQR = 11.401; Yes, there are outliers and it appears that Minitab does use the formula.

1d) "Resistant" means it is not strongly influenced by outliers. Some examples of resistant statistics are median, mode and IQR.

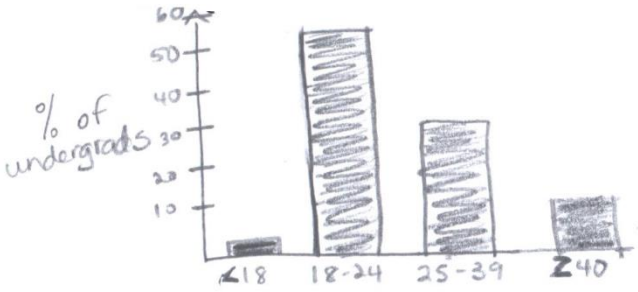
2) a) The data is slightly skewed right with the center about 4 and a range of 9. There are no apparent outliers.
b) Median and IQR. C) About 4 hurricanes per year

3) 55.05%

4) 0.0605

5) It appears the majority of college students are between 18 and 24 years old followed by 25-29 year olds. Very few college students are under 18.

	Counts	%
Under 18	286	2%
18 to 24	7771	55%
25 to 39	4388	31%
40 and up	1672	12%
Total	14,116	



6) The distribution for Hank Aaron appears to be slightly skew left but the distribution for Barry Bonds is skew right. The median of 38 for Hank Aaron is higher than the median of 35 for Barry Bonds. Hank Aaron's IQR if 13 is less than Barry Bond's IQR of 16. Barry Bonds has a higher maximum at 73 than Hank Aaron's at 48. (Note: these numbers may vary slightly)

7) 68%

8) 0.1151

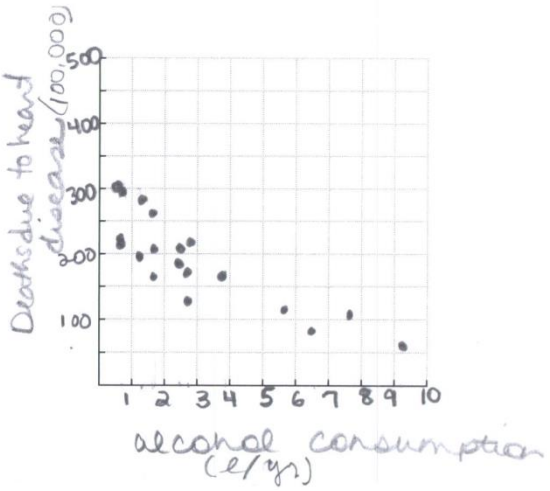
9) 142

10) 168

11) -0.2. This means that the student is just below the mean by 0.2 standard deviations.

12) 0.895

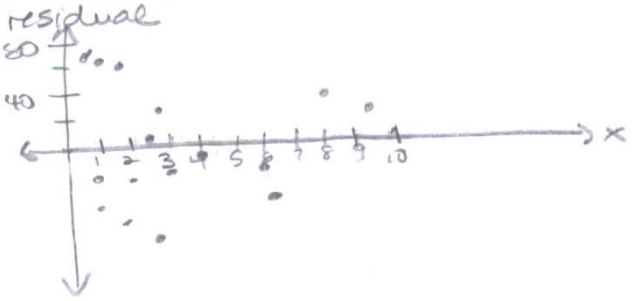
13) There appears to be a strong negative linear association between alcohol consumption and death rate due to heart disease.



14) $y = 260.6 - 22.97x$

15) $r = -0.8428; r^2 = 0.713$; The correlation coefficient is negative which indicates that there is a fairly strong negative linear association between alcohol consumption and death rate due to heart disease. 71% of the variation in the death rate due to heart disease is accounted for by the LSRL.

16) a) 2,860,300 deaths from heart disease per year. I am not confident in this answer because 10.1 is outside the domain of this problem. This is called extrapolation. b) There is a random scatter with no apparent pattern but it appears that France may be influential since it is an outlier in the x-direction.



17) a) $y = -0.12608 + 0.060782x$ where x is the measure of social distress and y is the measure of brain activity.

b) $r = 0.878$ which indicates that there is a strong positive linear association between social distress and brain activity.

18) $y = 22.73 + 0.187x$

19)

a) explanatory – type of surgery; response – length of survival after treatment
b) This is an observational study
c) Yes, for example the overall health and age of the patients is a confounding variable

20)

a) The experimental units are the 60 one-day old chicks and the response variable is weight gain.
b) There are 2 factors (corn variety and protein level) and 6 treatments. 60 experimental units are required.
c) Number the chicks from 01 to 60 and use a random digit table to assign 10 chicks to each category. Ignore repeats and numbers not in this range.
d) You could use blocking and block by gender.

21) This is an experiment in which neither the subjects nor the person administering the treatment know which is Pepsi and which is Coke. This design is a matched pairs design.

22) 0.5769

23) 0.6035

24) 0.699

25) No, these events are not independent, since $P(\text{married}) \neq P(\text{married} | \text{age } 30-64)$.

26) No, these events are not disjoint, since a woman can be in both categories.

27) The Law of Large Numbers

28) a) 0.978 b) Let the digits 00-71 represent landing on the buttered side and the digits 72-99 not landing on the buttered side. Use the random digit table to select 2-digit numbers and determine what percent of those numbers are from 00-71.

29) 0.719 is the P(A) given surgery. Therefore, he should have surgery since this is slightly higher that 0.7 with medical management.

30) $\frac{1}{8}$

31) $\frac{1}{8}$

32) $\frac{3}{4}$

33) 0.5

34) 0.2
35) 7.9; 4.1581
36) 31.6; 8.3162
37) 9.5; 8.8752
38) 31.05; 4.65
39) 0.0016384
40) 0.193536
41) 0.096256
42) 4.2; 1.2961
43) No, np is not ≥10
44) 0.321
45) 0.068
46) 0.687
47) 199 hits
48) 3.115
49) a) p = 0.68 is a parameter; p = 0.73 is a statistic b)0.68; 0.0381 c)0.0947
50)
    a) $\bar{x}$ = 27.4
    b) 0.5271
51) Central Limit Theorem
52) (a) We are told this is an SRS. There are more than 3000 voters in Troy. $np = 300(0.54) = 162 \geq 10$ $n(1-p) = 300(0.46) = 138 \geq 10$

    (b) $p \pm z^* \sqrt{\frac{p(1-p)}{n}} = 0.54 \pm 1.96 \sqrt{\frac{0.54(0.46)}{300}} = (0.4836, 0.5964)$ We are 95% confident that the true proportion of voters on that would vote in favor of the millage is between 0.4836 and 0.5964. Since part of this interval is less than 0.5, the city should NOT be too confident that it will pass.

    $\text{ME} \leq 0.04$

    c) $1.96\sqrt{\frac{0.54(0.46)}{n}} \leq 0.04$ They should have a sample size of 597 (which is still less than 10% of voters).

    $n \geq 596.4$
53) (a) $\bar{x} \pm t^* \left(\frac{s_x}{\sqrt{n}} \right) = 157.92 \pm 3.106 \left(\frac{16.037}{\sqrt{12}} \right) = (143.54, 172.3)$ We are 99% confident that the true mean score for this bowler is between 143.54 and 172.3.

    (b) The random sample condition and 10% condition are met. However, the sample size is small $(n \leq 30)$ and we do not know if the population distribution is normal. A normal probability plot of the data shows a roughly linear pattern and t-intervals are robust so I am fairly confident that my interval is accurate.