AP Statistics Semester I Final Exam Review Solutions

1. One-variable data analysis

Player	Mean	Median	Mode
Jerry	5	5	7
George	5.33	4	3

Since George has an outlier on Hole 9, his mean is greatly affected. Using the median would be a better comparison when outliers are present. Other than Hole 9, George did better.

2. Analysis of histograms – center and spread

Era	Center	Spread
1901-1930	.375	.300-.430
1931-1960	.360	.300-.410
1961-1990	.340	.300-.400

Over time, batting averages are decreasing in their measure of center and the spread is decreasing a little on the high end.

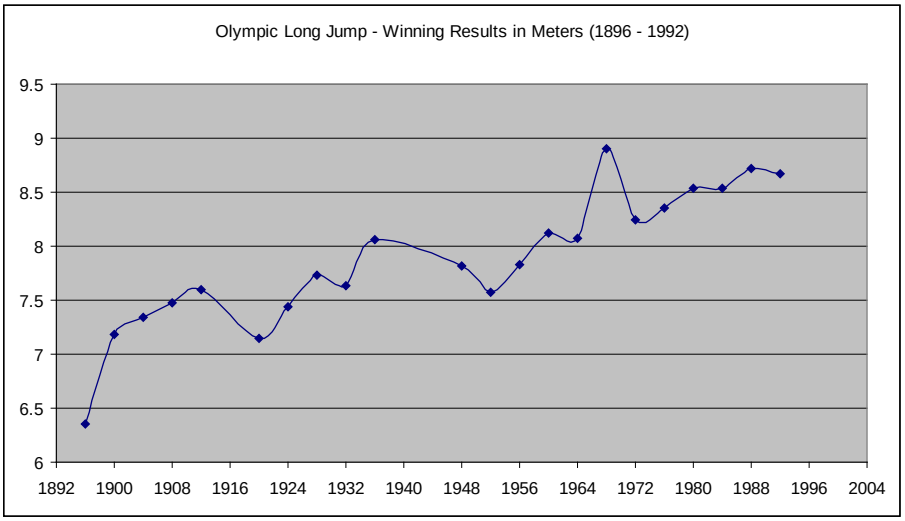
3. Stem and leaf plot, time plot

a)

6	4
6	
7	2 2 3 4
7	5 6 6 6 7 8 8
8	1 1 1 2 4
8	5 5 7 7 9

Results are rounded to the nearest tenth of a meter
6 | 4 = 6.4 meters (rounded)

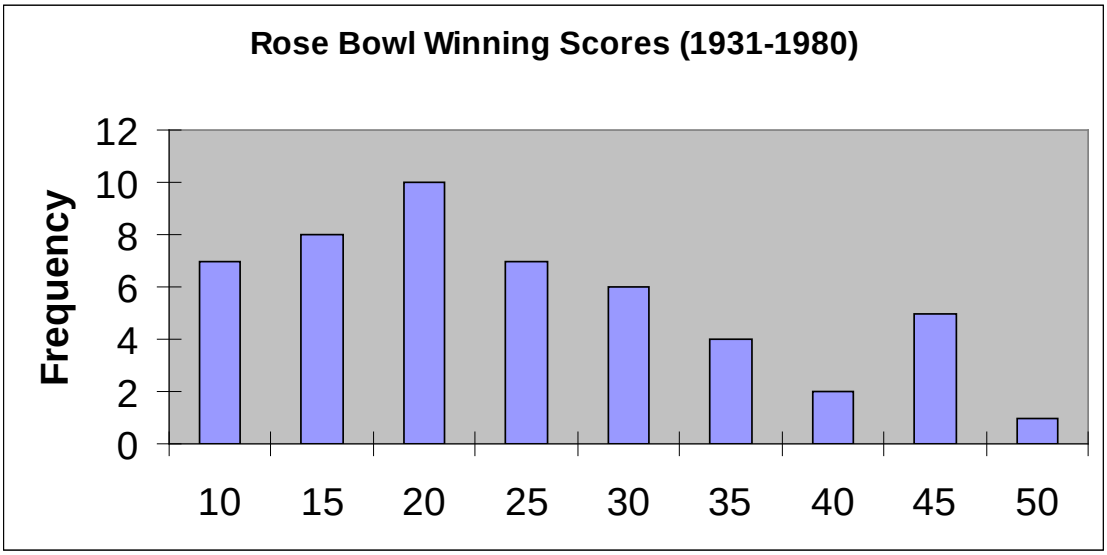
b)



c) The time plot allows us to see the progression of the long jump winning length over time. It shows a gradual climb with occasional dips. There is one major peak in 1968, when the record was increased by a great amount. The stem plot allows us to see the overall distribution better. We can see that the center is around 7.8 meters. The 6.4 is a bit of an outlier. The other values have a fairly uniform shape from 7.0 to 8.9.

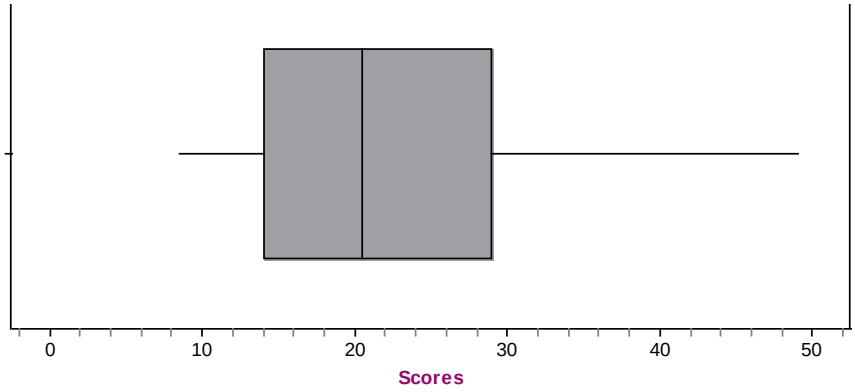
4. Histogram of Rose Bowl winning scores

a)



b)

MIN	Q1	MEDIAN	Q3	MAX
7	14	20.5	29	49



c) Skewed right. If a "whisker" of a box and whisker plot is stretch longer than the other, there is an outlier in that direction, causing a skew.

d) If there is an outlier on the high side, it would have to be greater than $Q_3 + 1.5 \bullet IQR$

Here that would be: $29 + 1.5(29 - 14) = 51.5$. So the highest value (49) is <u>not</u> an outlier.

5. Analysis of normal density curves

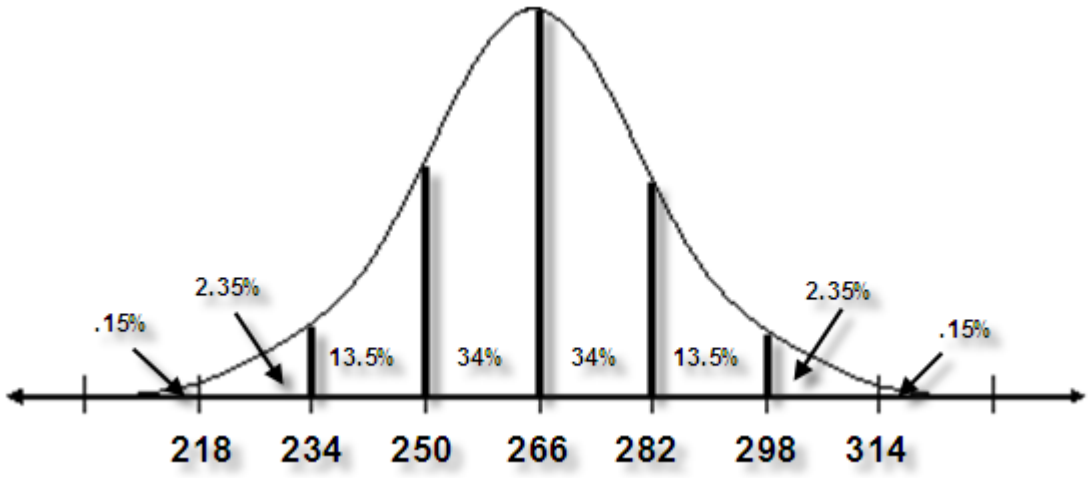
a) C = new, B = 5 years, A = 10 years. Weights will be very similar at beginning, then will vary greatly over time.

b) Average weight decreases, on average.

c) The standard deviation increases, becomes more variable.

6. 68-95-99.7 Rule

a)



b) 68% c) 81.5% d) 47.5%

e) z = (310 – 266) / 16 = 2.75. From Table A, the answer is .003 or .3%

7. Linear Regression Analysis

a) y = 50.49 – 5.74x, r = -.469, r^2 = .22 or 22%

b) Fairly weak linear relationship, r close to -.5, r^2 only 22%.

c)

Weight (x)	Actual MPG (y)	Predicted MPG (y)	Residual
3.0	43	33.255	9.745
4.0	30	27.511	2.489
5.0	15	21.766	-6.766

d) Residual Plot – use calculator

e) Residual plot shows a curved pattern, so a line is NOT a good model for this data.

8. Assessment of statements concerning correlation and regression

a)	regression	e)	positive	i)	response variable
b)	strength of	f)	TRUE	j)	TRUE
c)	+1 or -1	g)	TRUE	k)	TRUE
d)	TRUE	h)	-1 and +1	l)	below

9. Surveys and samples

a) Students at the university
b) Grade level, major, gender (there are many others)
c) Using ID numbers, select a simple random sample of the students using some form of random number generation. It might also be useful to stratify by age or gender.
d) Convenience sample
e) Systematic sample
f) Voluntary response sample

10. Experimental Design

a) A treatment is imposed
b) Matched pairs
c) Control – eliminate lurking variables by ensuring that the type of play is similar
d) Randomization – randomize which child wears which clothing
e) Use many sets of twins to collect a large amount of data

11. Discrete Random Variables

a) See table, top right
b) See table, bottom right
c) .5 d) .333 e) 4.5

(11a)

		2nd Die					
		2	**2**	**2**	**3**	**3**	**3**
1st Die	**1**	3	3	3	4	4	4
	1	3	3	3	4	4	4
	2	4	4	4	5	5	5
	2	4	4	4	5	5	5
	3	5	5	5	6	6	6
	3	5	5	5	6	6	6

12. Geometric Probability

a) The variable of interest is how long until the first success. ($p = .6$)
b) $P(X = 4) = (.4)^3(.6) = geometpdf(.6, 4) = .0384$
c) $P(X \le 4) = geometcdf(.6, 4) = .9744$
d) $P(X > 3) = 1 - geometcdf(.6, 3) = .064$
e) $\mu = \frac{1}{p} = \frac{1}{.6} = 1.\overline{6}$

(11b) sum	**3**	**4**	**5**	**6**
Prob(sum)	.167	.333	.333	.167

13. Binomial Probability

a) The variable of interest is how many successes in a fixed number of trials. ($p = .2$, n = 8)
b) $P(X = 4) = \frac{8!}{4!4!}(.2)^4(.8)^4 = binompdf(8, .2, 4) = .0459$
c) $P(X \le 2) = binomcdf(8, .2, 2) = .7969$
d) $P(X > 4) = 1 - binomcdf(8, .2, 4) = .0104$
e) $\mu = np = 8(.2) = 1.6$