Cumulative AP Practice Test 1 Solutions

Page 276

AP1.1 d.

AP1.2 e.

AP 1.3 b.

AP 1.4 c.

AP 1.5 a.

AP 1.6 c.

AP 1.7 e.

AP 1.8 e.

AP 1.9 d.

AP 1.10 d.

AP 1.11 d.

AP 1.12 b.

AP1.13 b.

AP1.14 a.

AP1.15 (a) When comparing data sets it is expected that center, spread, and shape will be addressed. The median for the distribution of cardiovascular fitness measures for machine A (28) is lower than the median of the distribution of cardiovascular fitness measures for machine B (38). The range for the distribution of cardiovascular fitness measures for machine A is substantially smaller than the range of the distribution of cardiovascular fitness measures for machine B (32 < 57). (Alternatively, the *IQR* for machine A (37 – 22 = 15) is lower than the *IQR* for machine B (47 – 25 = 22).) The distribution of cardiovascular fitness measures for machine A is reasonably symmetric in shape, while the distribution of cardiovascular fitness measures for machine B is skewed to the left (toward the lower values). In general, the cardiovascular fitness measures for machine B tend to be higher than those for machine A. (b) The company should choose machine B if they want to advertise it as achieving the highest overall gain in cardiovascular fitness. The median for machine B is higher than it is for machine A, as is the mean ($\bar{x}_B = 35.4$ *versus* $\bar{x}_A = 28.9$). (c) The company should choose machine A if they want to advertise it as achieving the most consistent gains in cardiovascular fitness. Machine A exhibits less variation in gains than does machine B. The *IQR* for machine A is 15, while the *IQR* for machine B is 22. Additionally, the standard deviation for machine A is 9.38, while the standard deviation for machine B is 16.19. (d) Volunteers were used for the experiment and these volunteers may be different in some way from the general population of those who are interested in cardiovascular fitness. Another reason is that the experiment was conducted at only <u>one</u> fitness center. Results may vary at other fitness centers in this city and in other cities.
If the company had taken a random sample of <u>all</u> people who were interested in cardiovascular fitness it could infer to this much larger population. If the company had randomly selected a number of fitness centers from across the country and then randomly selected members of each fitness center to participate

in the experiment, the company would be using a stratified random sampling method with the fitness centers serving as the strata.

AP1.16 (a) Assign each retail sales district a number from 1-60 using a random number generator. Order the sales districts numerically. The first 30 are assigned to the monetary incentives group and the remaining 30 to the intangibles incentive group. After a specified period of time compare the mean sales for each of the treatment groups to see if there is a difference. (b) Assign each retail sales district a number between 01 and 60. Go through the random number table taking two digits at a time. The first 30 two-digit numbers between 01 and 60 to come up are assigned to the monetary incentives group. The reminder will be assigned to the intangible incentives group.

07511 88915 41267 16853 84569 79367 32337 03316

The districts labeled 07, 51, and 18 are the first three to be assigned to the monetary incentives group. (c) It would be better to use a matched pairs design. There could be a large variation among the sales figures for the various districts due to the various sizes of those districts across the different regions of the United States. Matching the districts based on their size reduces the effect of variation among the experimental units due to their size on the response variable—sales volume. Pair the two largest districts in size, the next two largest, down to the two smallest districts. For each pair, pick one of the districts and flip a coin if the flip is "heads" this district is assigned to the monetary incentives group. If it is "tails," this district is assigned to the intangible incentives group. The other district in the pair is assigned to the other group. After a specified period of time compare the mean sales for each of the treatment groups to see if there is a difference.

AP1.17 (a) Yes. A linear model is reasonable since the scatterplot shows a strong positive linear association between shelf length and weekly sales (in dollars). (b) $\hat{y} = 317.94 + 152.68x$, where $\hat{y} =$ weekly sales (in dollars) and $x =$ shelf length (in feet). (c) $\hat{y} = 317.94 + 152.68(5) = 1081.34$. A shelf length of 5 feet would, on average, yield weekly sales of \$1081.34. (d) The value $s = 22.9212$ represents the standard deviation of the residuals. On average, there is a \$22.92 difference between the predicted weekly sales and the actual weekly sales. (e) $r^2 = 0.982$. 98.2% of the variation in weekly sales revenue can be explained by the linear regression using shelf length allocated to the house brand as the predictor. (f) It would be inappropriate to interpret the intercept, since the data represent sales based on shelf lengths of 3 to 6 feet and 0 feet falls substantially outside that domain. We would be trying to extrapolate beyond the set of data.