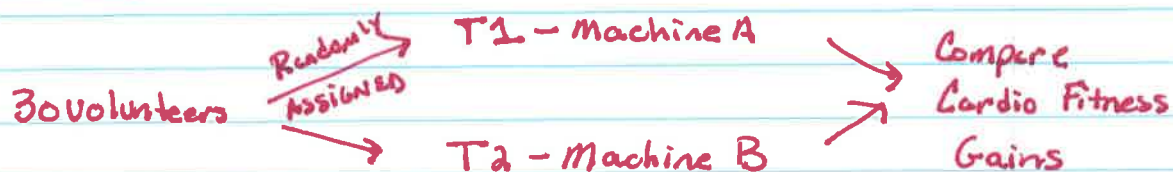


Cumulative AP Practice Test 1 (Chapters 1-4) Solutions

<p>AP1.1 d</p>	<p>AP1.2 e</p> $Z = \frac{X - \mu}{\sigma}$ $1.28 = \frac{40 - 38}{\sigma}$ $\sigma = \frac{2}{1.28}$ $\sigma = 1.5625$	<p>AP 1.3 b</p> <p>$N(60, 18)$ - scored double</p> $\bar{X} = 2(60) - 5 = 115$ $\sigma = 2(18) = 36$
<p>AP 1.4 c</p> <p>impossible to get 2 Females or 2 males</p>	<p>AP 1.5 a</p> <p>X a) For each additional gram, the tail increase 3mm</p> <p>b) $\hat{\text{tail}} = 20 + 3(88) = 134$</p> <p>c) true slope is positive</p> <p>d) true</p> <p>e) $\hat{\text{tail}} = 20 + 3(29) = 107$ residual = $y - \hat{y} = 100 - 107 = -7$</p>	<p>AP 1.6 c</p> $\mu = 225$ $\sigma = 50$ <p>68-95-99.7 rule</p>
<p>AP 1.7 e</p> <p>Correlation does <u>NOT</u> imply Causation!</p>	<p>AP 1.8 e</p> <p>put data in L1</p> <ul style="list-style-type: none"> • 1 var - stat - b is true <p>a) $Q1 - 1.5(Q3 - Q1) = 100$ (96 outlier)</p> <p>Remove 96: 1 var stat</p> <p>c) T 125 > 122</p> <p>d) true - less variation</p> <p>e) X IQR would increase</p>	<p>AP 1.9 d</p> <p>I. X Negative association</p> <p>II true</p> <p>III true</p>
<p>AP 1.10 d</p> <p>Random Sample of COUNTRIES</p> <ul style="list-style-type: none"> → Urban → RA → rural → RA → suburban → RA <p>Homogeneous Groups Samples are strata</p>	<p>AP 1.11 d</p> <p>1 Car - 2 factors</p> <ul style="list-style-type: none"> • Gas - 3 types • Weight - 3 weights <p>9 treatment combinations</p> <p>D is the best answer</p>	<p>AP 1.12 b (10, 4)</p> <ul style="list-style-type: none"> • put x in L1. • put y in L2. • STAT CALC 8 L1NREG • STATPLOT • ScatterPlot \square • XLIST L1 • YLIST \square LIST Resid <p>Corresponds to Pt. 6</p>
<p>AP1.13 b</p> <p>Median is in the 20-30 group and on the high end. Therefore the median must be either 27(a) or 28(b).</p> <p>The distribution is fairly symmetric and therefore the mean and median should be close.</p> <p>Mean is weighted. Calculate by taking the frequency by the waiting time midpoint = $(5*5 + 15*24 + \dots + 65*2) = 4210/140 = 30.07$</p>	<p>AP1.14 a</p> <p>TRACE X=10 Y=3.6</p> <p>The range (min to Max) looks the same</p>	

AP 1.15



- (a) When comparing data sets it is expected that center, spread, shape and unusual patterns are addressed in context of the problem ANSWER:

The median for Machine B (38) is much higher than the median for Machine A (28)

The range for Machine B ($59 - 20 = 29$) is much higher than the range for Machine A ($51 - 41 = 10$)

The distribution for Machine A is roughly symmetric compared to a skewed distribution for Machine B.

- (b) If the company wants to advertise a machine that achieves the highest overall gain in cardio fitness they should choose **MACHINE B** because the median for Machine B (38) is higher than Machine A (28)

- (c) If the company wants to advertise a machine with the most consistent gains in cardio fitness they should choose **MACHINE A** because it has less variation. The range for Machine A is 10 compared to a much larger range for Machine B (29).

AP 1.5 (Cont)

- (d) The company used volunteers for the experiment and these volunteers may not be representative of the population of all people interested in cardio fitness.

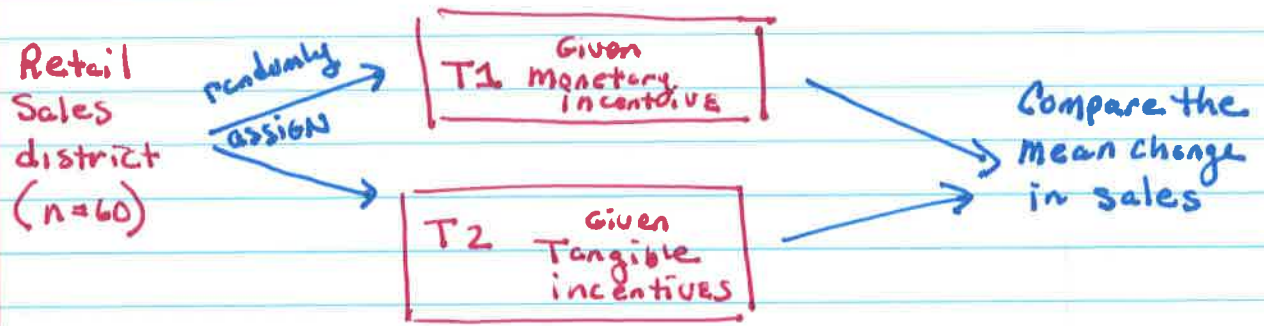
If the company took a random sample from the population of interested people in cardio fitness, then they could use the results to make inferences about this population.

AND/OR

Also, the company only used 1 fitness center for the experiment. Results may vary from center to center and state to state. To improve this experiment, the company could use a stratified random sampling with the fitness centers serving as strata.

AP 1.16

(a) COMPLETELY RANDOMIZED DESIGN (picture is easiest method to describe)



(b) EXECUTE RANDOMIZATION FOR THIS DESIGN
(You must clearly explain so an AP reader can follow your instructions to carry this out)
• Tip use bullets)

- ① Number the 60 ^{districts} from 01 to 60
- ② Use the provided table of random digits and select 2-digit numbers exclude repeats, 00, and 61-99.
- ③ The first 30 numbers will be the ^{districts} that will receive treatment #1 (monetary incentives) and the remaining ^{districts} will receive treatment #2 (tangible incentives)

Select random #'s 07 51 18

Districts numbered 07, 51, 18 will receive the monetary incentive treatment

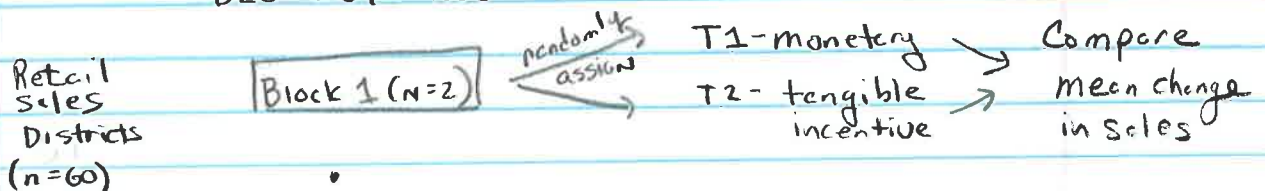
AP 1.16 (CONT)

- (C) **MATCH PAIR DESIGN** is a method of blocking with 2 districts in each. There could be a lot of variation in sales in the 60 districts. To control for this variation, we could match districts based on their sales.

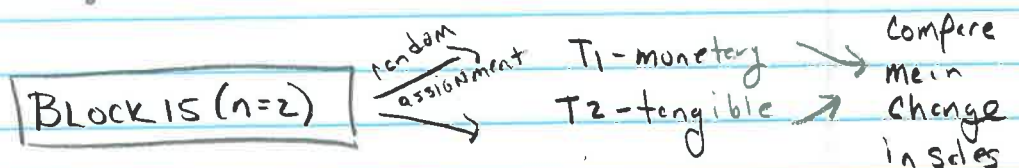
The Design

- ① order the districts from high to low sales
- ② pair the districts. The first pair would be the 2 districts with the highest sales, the next 2 largest, down to the 2 smallest districts. This would result in 15 blocks.
- ③ For EACH OF THE 15 BLOCKS (PAIRS) flip a coin. Heads - receive monetary treatment
Tails - receive tangible incentive treatment
Repeat 15 times for each block
- ④ After a specific amount of time record the change in sales for each district
- ⑤ Compare sales within each pair

BLOCK BY SALES



MUST show all 15 BLOCKS



FOR MATCHED PAIRS, WRITING IN WORDS TAKES LESS TIME THAN A DIAGRAM (BOTH ACCEPTABLE)

AP 1.17

(A) YES. BASED ON THE SCATTER PLOT THERE IS A VERY STRONG, POSITIVE, LINEAR ASSOCIATION BETWEEN SALES AND SHELF LENGTH

(B) 2 methods to write LSR

WRITE WITH WORDS **
(recommended)

$$\widehat{\text{Sales}} = 317.94 + 152.680(\text{SHELF})$$

OR Write with variables

$$\hat{y} = 317.94 + 152.680x$$

where: $x = \text{shelf length (in)}$
 $y = \text{weekly sales (in \$)}$

(C) For 5 linear ft shelf

$$\widehat{\text{sales}} = 317.94 + 152.68(5) = \$1,081.34$$

write in

Context: IF 5 LINEAR FEET OF SHELF WAS ALLOCATED, THE PREDICTED SALES WOULD BE ABOUT \$1,081 per week.

(d) The value $s = 22.9212$ is the standard deviation of the residuals. When using this LSR model using shelf life to predict sales, we will typically be off by about \$23 in our predicted value.

AP 1.17 (Cont)

- ③ The coefficient of determination is $r^2 = .982$.
(IN context) About 98.2% of the variation in weekly sales revenue can be accounted for by this linear model relating sales to shelf life.

remember: Correlation coefficient is $r = \sqrt{.982} = .991$

" r " measures the strength and direction of the linear association (relationship) between sales and shelf length

↑
NO UNITS

- ④ No the manager should Not believe the intercept as a reasonable value, since the data represents sales based on shelf lengths of 3 to 6 feet. Zero feet falls substantially outside the domain.

Therefore, this is an extrapolation beyond the set of data and predictions are not appropriate.