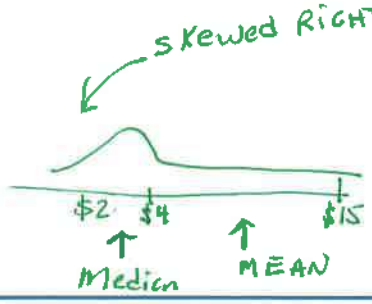
Cumulative AP Practice Test 1 (Chapters1-4) Solutions

AP1.1 d



AP1.2 e

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

$1.28 = \frac{40-38}{\sigma}$

$\sigma = \frac{2}{1.28}$

$\sigma = 1.56$ in

area = .9

$\mu = 38$ 40

$Z = \text{invNorm}(.9,0,1)$

$z = 1.28$

AP 1.3 b

$N(60, 18)$ – scored doubled

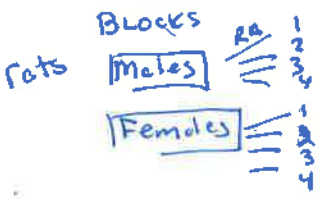
-5

$\bar{x} = 2(60) - 5 = 115$

$\sigma = 2(18) = 36$

AP 1.4 c

impossible to get 2 Females or 2 males



AP 1.5 a

X a) For each additional gram, the tail increases 3mm

b) $\hat{\text{tail}} = 20 + 3(88) = 134$ ✓

c) true slope is positive

d) true

e) $\hat{\text{tail}} = 20 + 3(29) = 107$

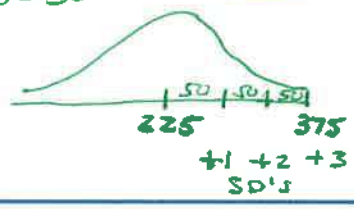
residual $= y - \hat{y} = 100 - 107 = -7$

AP 1.6 c

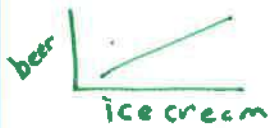
$\mu = 225$

$\sigma = 50$

68-95-99.7 rule



AP 1.7 e



Correlation does NOT imply Causation!

AP 1.8 e

put data in L1

• 1 Var-Stat – b is true

a) $Q1 - 1.5(Q3 - Q1) = 100$ (96 outlier)

Remove 96 : 1 Var Stat

c) T $125 > 122$

d) true – less variation

e) X IQR would increase

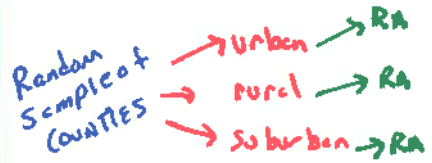
AP 1.9 d

I. X Negative association

II true

III true

AP 1.10 d



Homogeneous Groups Samples are strata

AP 1.11 d

1 Car – 2 factors

• Gas – 3 types

• Weight – 3 weights

9 treatment combinations

D is the best answer

AP 1.12 b (10,4)

• put x in L1.

• put y in L2.

• STAT CALC 8 LINREG

• STATPLOT

  • Scatterplot

  • XList L1

  • YList 2ND LIST Resid

TRACE X = 10 Y = 3.6

Corresponds to pt up

AP1.13 b

Median is in the 20-30 group and on the high end. Therefore the median must be either 27(a) or 28(b).

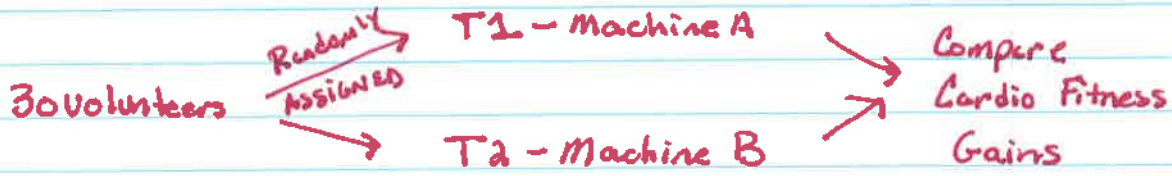
The distribution is fairly symmetric and therefore the mean and median should be close.

Mean is weighted. Calculate by taking the frequency by the waiting time midpoint = (5*5 + 15*24 + … +65*2) = 4210/140 = 30.07

AP1.14 a

The range (Min to Max) looks the same

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 large range for Machine B (28).

AP 1.5 (Cont)

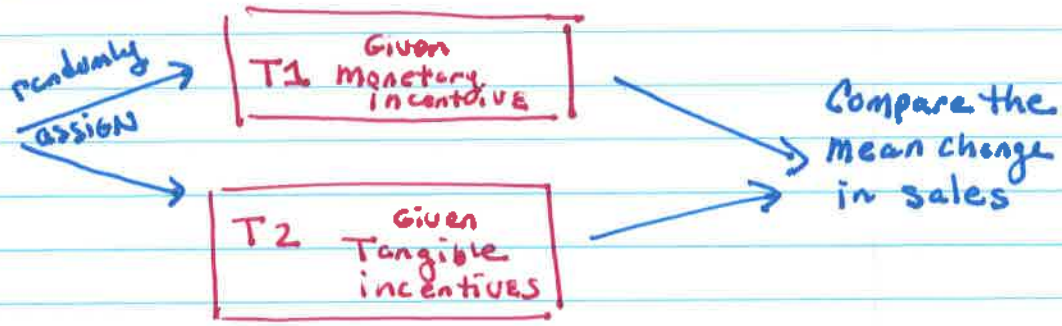
(d) The company used volunteers for the experiment and these volunteers may not be representetive of the population of all people interested in cardio fitness. If the company took a random sample from the population of interested people in cardio fitness, then they could use the results to make inferences 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 alot of variation in sales in the 60 districts. To control for this variations, 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

Retail sales Districts (n=60)

Block 1 (n=2) → random assign → T1 - monetary → Compare mean change in sales
Block 1 (n=2) → random assign → T2 - tangible incentive → Compare mean change in sales

MUST show all 15 BLOCKS

Block 15 (n=2) → random assignment → T1 - monetary → Compare mean change in sales
Block 15 (n=2) → random assignment → T2 - tangible → Compare mean change in sales

FOR MATCHED PAIRS, WRITING IN WORDS takes less time than a diagram (both acceptable)

AP 1.17

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

(B) 2 methods to write LSRL

WRITE WITH words ** (recommended)

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

OR Write with variables

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

where: x = Shelf length (in)

y = Weekly sales ($'s

(C) FOR 5 linear ft shelf

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

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 LSRL model using shelf life to predict sales, we will typically be off by about $23 in our predicted value.

AP 1.17 (Cont)

(e) 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 $r = \sqrt{.982} = .991$ ← NOONITE

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

(f) No the manager should Not believe the intercept as a reasoneble value, since the data represents sales based on shelf lengths of 3 to 6 feet. Zero feet falls substantially out side the domain.

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