

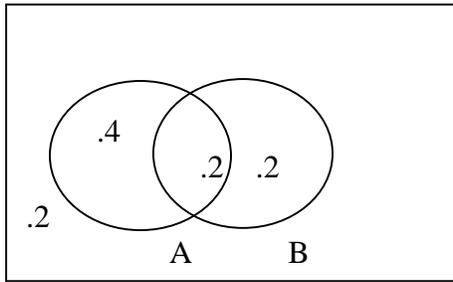
Answers to "Review for AP Exam and Final Exam"

Topic I

1. a.
 - b. center, appears to be in the 50-150 days, clusters in the 50-100 and 500-550 group, gaps between 250-300 and 450-500, spread from 43-541, range is $541-43 = 498$, outliers are probably in the 500's, shape is skewed to the right
 - c. median = 113, mean = 193.05, range = 498, IQR = 129.5, standard deviation = 165.09, variance = 272.71, Q1 = 83, Q2 = 113, Q3 = 212.4, Min = 43, Max = 541, 70th percentile = 178
 - c. z-score of 80 = $-.685$, z-score of 520 = 1.98
 - d. change max days to 100 = max range, mean, standard deviation, variance \rightarrow all increase
Trim the data by 10% = mean (unknown), range (decrease), SD(decrease), Var(decrease),
Min(unknown), Max(unknown)
Divide by 100 = all
2. a. 5# summary for 4th hour (48, 76, 84, 92, 100) , for 5th hour (72, 82, 88, 93, 96) Create box plots
Spread – 4th hour has more spread
Center – are both fairly close
Clusters – none (too little data)
Gaps – none (too little data)
Outliers – 4th hour (48) 5th (none)
Shape – both skewed to the left
3. a. explanatory – text anxiety, response – score on the exam
b. create a scatter plot, appears to be negative association, (7,50) may be an outlier, appears linear, fairly strong association
c. $r = -.7877$, $r^2 = .6205$, $\hat{y} = 68.84 - 1.064x$
d. residual plot and table, appears scattered
e. the residual plot is scattered so the linear model is a good model for the data, the r-value is $-.787$ which shows fairly strong negative association, as test anxiety goes up the test scores go down
f. (5,100) would make a weaker linear relationship, the r-value decreases to $-.73$ and 42 to 53%. This would be an influential point.
4. a. $r = .11$, shows a very weak positive linear relationship between annual raises and teaching evaluation
b. $r^2 = .0121$, about 1.2% of the variation in raises would be explained by the LSRL
5. a. plot the data on the graph, the data points for the years, 92, 93, and 94 don't seem to fit the pattern.
b. most of the ratios are a bit over 1, 1.06 is the average for the 1st 8 data points
b. (2nd b) table of values of years vs logy
c. $\log \hat{y} = -45.865 + .02468x$, $r = .986$, $r^2 = .9726$
Or...using natural logs ($\ln \hat{y} = -105.6 + 0.057x$)
d. $\hat{y} = (1.36 \times 10^{-46})(1.06)^x$
e. 1410.847 (thousands) = 1,410,847
f. 78
6. a. $\log \hat{y} = -.5235 - 2.0126 \log x$, $r = -.9996$, $r^2 = .9992$, $\hat{y} = .2996x^{-2.0126}$
Or...using natural logs ($\ln \hat{y} = -1.205 - 2.01 \ln x$)

- b. the intensity of the light bulb decreases as the distance from the light bulb increases,
7. a. 11,374,000
 b. 51.2%
 c. 60.7%, 22.1%, 68.4%, 11.0%
 d. the 18-21 year old group comprises 60.7% and 68.4% of the 2 yr full or 4 yr full time students which is what we would expect, they are the majority of the students, they make up only 22.1% and 11.05 of the part-time students, part-time students are usually adults who have a family or are also working so don't have time to be a full-time student
8. definitions- get from book
9. a. population is the people who live in Ontario, sample are the 61,239 people who filled out the survey
 b. since the sample is so large, we can be fairly sure that the results of the survey would be a good reflections about the population
 c. This is an observational study because no treatment was imposed upon the subjects
10. Control, Randomization, Replication
11. a. Use the digits 0-9, let 0 and 1 represent that she passes, and 2-9 represent that she fails
 b. simulation results in a proportion of approximately .4 or 40% likelihood of passing
 c. 1st attempt (passing 0, 1), 2nd attempt (passing 0, 1, 2), 3rd attempt (passing 0, 1, 2, 3)
 e. approximately .66 or 66%
12. a. Experimental subjects – physicians, factor – aspirin (pill given) – two levels, response – heart attack?
 b. diagram or explain how to design
13. diagram or explain how to design
14. a. o
 b. 1
 c. 0.01
 d. 0.99
 f. any example
15. a. $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$
 b. $P(A \text{ and } B) = P(A) * P(B)$ or $P(A \text{ and } B) = P(A) * P(B | A)$
 c. $P(B | A) = \frac{P(A \text{ and } B)}{P(A)}$
16. a. .445
 b. .555
 c. .314, .686
 d. not independent because $P(A)*P(B) = .38073$,
17. a. .5264
 b. .4054
 c. not independent

18. a.



b. .8, .2, .4, .2, .2, .4

19. discrete are the counting numbers (whole numbers), continuous are all the real numbers (can include fractions)

20. a. .25

b. .39

c. .74

d. .61

e. .9, .65

f. mean = 4.66, SD = 1.202, Var = 1.444

21. a. mean = .001, SD = .0022369

b. mean = 2.0005, SD = .0011189

22. a. n = 10, p = .25

b. create a chart with 3 columns (x, P(x) using binompdf, CumP(x) using binomcdf) and the 2 histograms, one for the pdf and one of the cdf

c. .2816

d. .5256

e. mean = 2.5, SD = 1.369

23. a. must have 2 categories, "success" and failure, P(success) must the same for each trial, all trials are independent, continue trials until the first success occurs, (no limit to n)

b. create a chart with x, P(x) using geometpdf, cumP(x) using geometcdf and the two histograms for n up to 10

c. .00128

d. .99968

e. mean = 1.25

24. N(0, 1) for standard normal, area under the curve = 1, draw a curve

25. a. .12100

b. .1587

c. .6944

d. 1.00

- e. approx 0
- f. .000005131

- 26.
- a. 80
 - b. .71374
 - c. .00774
 - d. 12.5127

27. as you increase your sample size, the sample statistic will get closer to the true population parameter, when applies to sampling distributions it means that the mean of the sampling distribution can be used to estimate the population mean

28. CLT is a theorem that says if you take many simple random samples of size n from a population, the mean of the samples can be used to estimate the population mean, the sampling distribution will be normally distributed with a mean = to mu and a SD decreased by the sample size, the sampling distribution will be normal no matter what kind of distribution the population has

- 29.
- a. an unbiased statistics is one that is obtained from a SRS, so we want a statistic that is unbiased because we want to make a prediction about the population
 - b. not always, the sample size will have an influences in the final estimating of the population parameter
 - c. mean = 11.967, SD = 1.886, $\hat{p} = .4167$

30. answers are the formulas for the standard deviation of the sampling distribution

32. sample statistic – ME ≤ population parameter ≤ sample statistics + ME

- 33.
- a. $7.78 \leq \mu \leq 8.63$
 - b. if you were to take 100 samples form the same population and every sample was of size 50, then 95 of the sample would have a mean between 7.78 and 8.63
 - c. narrower because the critical value would be smaller, therefore sthe ME would be smaller
 - d. t= .9428, p-value = .3504, so Fail to Reject

- 34.
- a. $.73 \leq \mu \leq .81$
 - b. z = -2.12, p-value = .03400, so Reject

35. use the z-test when the sample size is large and/or the population SD is known, use the t-test when the sample size is small and the population SD is not known

36. ME is the amount of variation above and below the sample statistic where the actual population parameter should be when the samples of size n are taken

- 37.
- a. $.6437 \leq \mu \leq 1.3563$
 - b. t = 7.2424, p-value = $4.566 \cdot 10^{-12}$, Reject

38. a. $t = 2.929$, p-value = .005, so Reject
 b. $t = .1501$, p-value = .88, so Fail to Reject
 c. $t = .6177$, CI = $.23 \leq \mu_2 - \mu_3 \leq 1.47$ ← from calc (.2426, 1.457 with df = 65.262)
 d. Type I means you reject the null and you shouldn't have because it is the acceptable hypothesis, consequences would be that we would say there is a difference in 0 and 2 hours when there really isn't, so may choose to cool the chicken before cooking when you shouldn't
 e. Type II means you fail to reject the null and you should have, we believe that cooling for 8 hours to 24 hours is the same when it isn't, so you would most likely cool only 8 hours to save money when you should have cooled longer
39. a. $t = 211.24$, p-value is approx 0, so Reject
 b. $106.83 \leq \mu_2 - \mu_1 \leq 108.57$, since 50 is not within the CI, Reject the null
40. $z = 6.826$, p-value = 4.389×10^{-12} , Reject
41. $\chi^2 = 11.24$, p-value = .0105, Fail to Reject (barely) ← there was a typo...the significance level should be .01
42. $\chi^2 = 6,8519$, p = .23188, Fail to Reject
43. a. slope = -0.34655, y-int = 4.8615
 b. $\hat{y} = 4.8615 - 0.34655x$
 c. slope models β , y-int models α
 d. $-.47 \leq \beta \leq -.22$
 e. $t = -5.9077$, reject