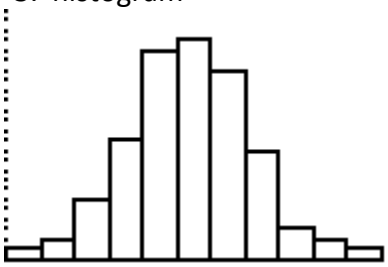
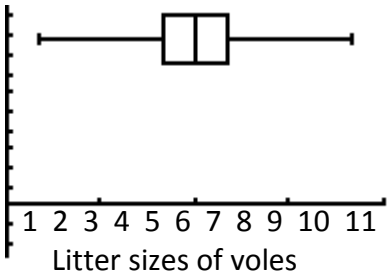
AP STATISTICS: MIDTERM REVIEW SOLUTIONS

-1. **categorical** – color of the first born in each litter
quantitative – number of baby moles in each litter
-2. bar chart
-3. histogram



-4. dotplot
-5. No apparent or visual outliers (but 1 and 11 are if you check the 1.5IQR rule)
-6. shape: approximately symmetric
-7. x-bar: 5.87; The mean is NOT resistant to outliers
-8. Median = 6; The median is resistant to outliers
-9. range: max – min = 11 – 1 = 10
-10. 5 number summary: min = 1, Q1 = 5, med = 6, Q3 = 7, max = 11
-11. IQR = Q3 – Q1 = 7 – 5 = 2
-12. boxplot:



-13. $S_x = 1.813$
-14. 1
-15. Median
-16. Mean
-17. Mean, median
-18. $\bar{x}$ is the mean of a sample, μ is the mean of a population (or model)
-19. S is the standard deviation of a sample, σ is the standard deviation of a population or model
-20. Where the curve changes from falling more steeply to less steeply. Find the point where the curvature changes.
-21. graph of normal distribution
-22. 68-95-99.7 rule: 68% of the data are within 1 standard deviation of the mean; 95% of the data are within 2 standard deviations of the mean, 99.7% of the data are within 3 standard deviations of the mean.
-23. mean + 1 standard deviation = 266 + 16 = 282 days or greater
-24. middle 99.7% (mean +/- 3 stand deviations) = 218 to 314 days
-25. shortest 2.5%: mean – 2* standard deviations = 266 – 32 = 234 days or less
-26. Percentile: .13%
-27. Percentile: 97.7%
-28. 50% percentile
-29. z = (257 – 266)/16 = -.5625

-30. P(x < 257) = P(z < -.5625) = .286 (Use Normalcdf(-1E99, -.52625)

-31. P(x > 280) = P(z > (280 – 266)/16) = P(z > .875) = .191 (Use Normalcdf(.191, 1E99)

-32. P(260 < x < 270) = P((260-266)/16 < z < (270-266)/16) = .2449

-33. Continuous: measurement of time is continuous (or discrete if considered the number of days)

-34. 287 days or longer (Use INVNORM(.90)) convert z value to number of days

-35. about 255 days or less. Use INVNORM(.25)

-36. between 262 and 270 days. Use INVNORM(.6) and INVNORM(.4)

-37. Use the Normal Probability Plot and check for a straight line

-38. back-to-back stemplot

4^{th} graders		7^{th} graders
	0	1
2	1	2
85	1	588
200	2	0334
9865	2	578
21	3	001333
976555	3	56
20	4	

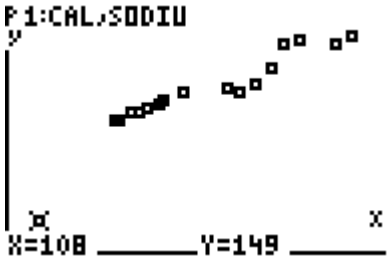
-39. The centers and spread are about the same. 4^{th} graders have a higher maximum reading score than the 7^{th} graders.

-40. mode: 4^{th} graders = 35, 7^{th} graders = 33

-41. Using the 1.5IQR rule, 1 is not an outlier. IQR = 32 – 19 = 13; 1.5*13 = 19.5; Q1 – 19.5 = 19 – 19.5 = -.5. 1 is not less than -.5 so 1 is not an outlier

-42. Modified boxplots show outliers

-43. scatterplot



-44. either

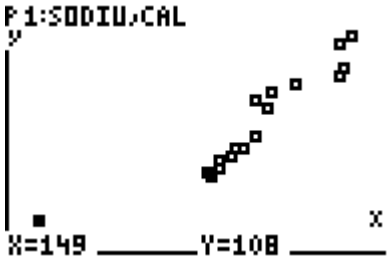
-45. either

-46. positive

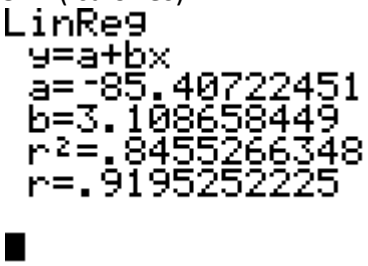
-47. linear

-48. strong (tightly clustered)

-49. outlier: (108, 149) has residual of –101.3



-50. is influential, when removed the regression line changes
-51. r = .92
-52. r = .95
-53. strength and direction of a linear association
-54. r remains the same
-55. r = 1
-56. r = -1
-57. linear
-58. Correlation is NOT resistant to outliers
-59. no (Strong correlation does not imply causation)
-60. scatterplot
-61. scatterplot
-62. Predicted sodium = -85.4 + 3.12(calories)



-63. slope = 3.12 sodium/calorie; For each increase of 1 calorie, the average predicted amount of sodium increases by 3.12 mg, according to this model
-64. predicted sodium = -85.4 + 3.12(345) = 398.2 mg
-65. predicted sodium = -85.4 + 3.12(345) = 991 mg
-66. extrapolation
-67. residual = y – y-hat = 500 – 476.2 = 23.8
-68. x-bar: 156.4, y-bar: 400.8: 400.8 should be about –85.4 + 3.12(345)
-69. $S_{calories}$ = 25.64
-70. S_{sodium} = 86.68
-71. r^2 = .84
-72. 84% of the variation of sodium is explained by the linear relationship between sodium and calories
-73. Look for random scatter in the residual plot to see is the error is randomly scattered showing that a linear model is an appropriate fit
-74. **extrapolation**: the use of a regression line for predication <u>outside</u> the domain of values of the explanatory variable, x, that you used to obtain the model. These predictions cannot be trusted.
-75. **lurking variable**: A variable that has an important effect on the response variable, but is not included in the variables being studied.
-76. **Causation** is when changes in explanatory variable cause the changes in the response variable. Examples vary…but smoking causes cancer is one.
-77. **Common Response** is when two variables are actually both caused by a third variable. Ex: Ice cream sales and drowning deaths in Santa Monica.
-78. **Confounding** is when the effects of a supposed explanatory variable cannot be separated from the effects of another "confounding" variable. Ex: exercise levels and weight (people who exercise more may also eat healthier)

-79. Marginal distributions

Education	
Did not complete high school	214
Completed high school	225
1 to 3 years of college	200
4 or more years of college	186

Never Smoked	Smoked, but Quit	Smokes
357	156	312

-80. percent of people who smoke = 312/825 = .38
-81. percent complete hs, given never smoked = 97/357 = .27
-82. P(quit smoking, given 4+ yrs college) = 63/186 = 34%
-83. P(didn't finish hs, given smoke) = 113/312 = 36%
-84. Smoking and level of education does not appear to be independent. Appears that people with more education are less likely to smoke:
P(smoke| no hs) = .53, P(smoke|hs) = .46, P(smoke| 1 –3 yrs college) = .3 and P(smoke|4 > yrs college) = .2
-85. **Observational study** – experiments observe what is already happening.
Experiment – A treatment is imposed on the subjects
-86. A **voluntary response sample** consists of people who choose themselves to be in the sample by responding to a general appeal. Also called a self-selected sample.
-87. A sample is a part of the population that we can actually examine in order to gather information about the population.
-88. A **convenience sample** does not represent the entire population – it favors some parts of the population over others. It is a sample that is convenient or easy to obtain.
-89. **Simple Random Sample** – consists of individuals chosen so that every set of the same sample size has the same chance of being selected.
-90. Label the towns from 01 to 12. Using a table of random digits, read 2 digits at a time. Ignore 00, 13 – 99 and any repeats. Stop after four towns are selected.
-91. Divide the population into groups that are similar (homogeneous), called strata. Then choose an SRS within each stratum.
-92. Divide population into groups that are heterogeneous, called clusters. Randomly select 1 cluster and survey everyone within that cluster.
-93. **Undercoverage** occurs when some parts or groups in a population of systematically left out of the sample.
-94. **Nonresponse** is when an individual chosen for the sample cannot be contacted or refuses to cooperate.

-95. **Response bias** is when respondents may lie or the interviewer influences the response either by appearance or wording of the question.
-96. Confusing or misleading questions can introduce bias. Ex: Knowing that the subject of Statistics is the most useful subject a person can know, do you think that every high school student should be required to study statistics?
-97. When humans are experimental units, they are called subjects.
-98. factors
-99. levels

-100. A physical response to a dummy treatment

-101. to control the effects of lurking variables on the outcome

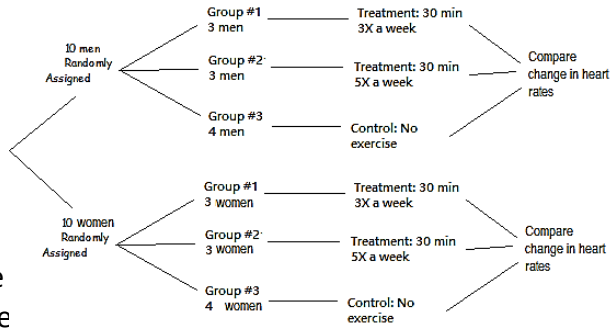
-102. 1 subject (ex: pre and post tests) or match 2 subjects with similar characteristics

-103. control (effects of lurking variables), randomization, replication (to reduce chance variation)

-104. Double-blind: neither the subject nor the people who evaluate the subjects know which treatment the subject has received.

-105. Block design: Break subjects into groups, called blocks, according to a variable that has an effect on the response variable. Then carry out the experiment separately within each block.

-106. experiment chart



-107. Step 1: Clearly describe
Step 2: State independe
Step 3: Define trial/count
Step 4: Simulate many repetitions. (Use table)
Step 5: State your conclusion

-108. One possible solution: Read 1 digit at a time. Let the odd digits represent a son and the even digits represent a daughter. Continue reading digits until you get 2 odd numbers. Count the number of digits read – this will be the number of children. Repeat this 10 times.

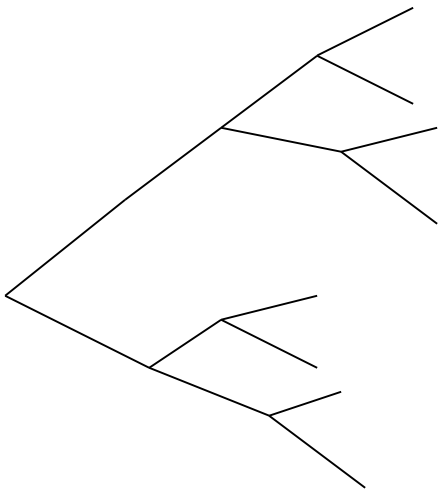
-109. Probability of 1 event occurring does not effect the probability of the other event occurring.

-110. {H, T}

-111. {0, 1, 2, 3} (either heads or tails)

-112. {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

-113. tree-diagram



-114. 12

-115. 1000

-116. 5040

-117. An event is a set of outcomes of an experiment (a subset of the sample space) to which a probability is assigned.

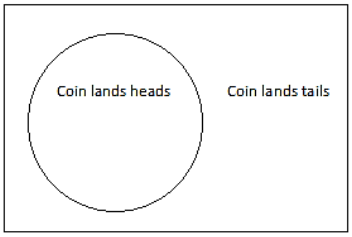
-118. 0, 1

-119. 1

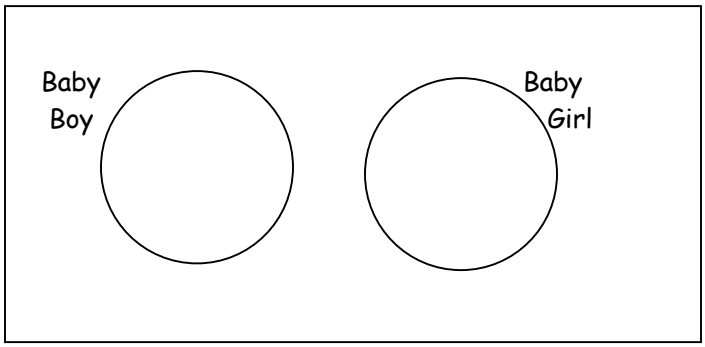
-120. P(S) = 1

-121. Compliment: P(Event does not occur) = 1 – P(Event Occurs)

Venn diagram:



-122. Disjoint events have no outcomes in common.



-123. P(Blue M&M) = .1

-124. P(red or green) = .3

-125. P(yellow and orange) = 0

-126. assume independence: .09 * .92 = .0828

-127. Independence….probably not.

-128. P(sum of 7 or 11) = 8/36

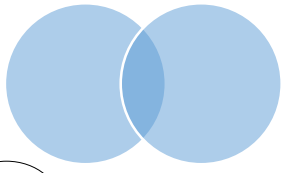
-129. P(roll doubles) = 6/36

-130. P(sum of 7 or 11) and P(doubles) = (8/36)*(6/36) = .037

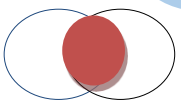
-131. independence, yes

-132. 0

-133. OR



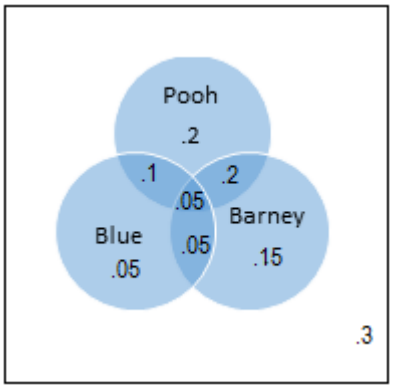
-134. AND



-135. P(A) = P(A|B)

-136. P(Smoking) does not equal P(Smoking|NO high school)

12/825 does not equal 111/214



-137.

-138. P = (area of circle)/(area of square) = $3.14(10)^2/24^2$ = .545

-151. Not disjoint: We can have a King of Hearts

-152. yes independent: P(A) = 13/52 = ¼

P(A|B) = P(King of Hearts)/P(KIng) = ¼

-153. yes

-154. no

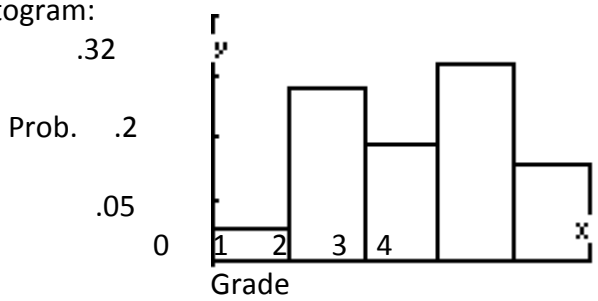
-155. ⋃ union OR

-156. ∩ Intersection AND

-157. A count (No. of students)

-158. A measurement (Distance from home to school)

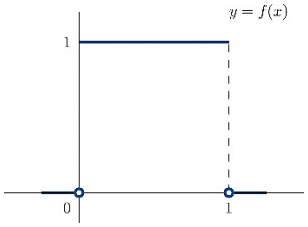
-159. Histogram:



-160. **P(X > 2)** = .32 + .16 = .48

-161. **P(X ≧ 2)** = .19 + .32 + .16 = .67

-162. A distribution that has constant probability



-163. .4

-164. .4

-165. For continuous random variables < and ≤ are the same (P that x = a particular number is 0. For discrete random variables, they are not.

-166. continuous

-167. mean

-168. **E(X)** = 0(.05) + 1(.28) + 2(.19) + 3(.32) + 4(.16) = 2.26

-169. σ^2=1.372

-170. σ=1.17

-171. The Law of Large Numbers states that when a trial is conducted a large number of times, the relative frequency of the event being measured in the trial tends to get closer to the actual probability of the event.

-172. X = no. of books, Y = no. of CDs, E(X + Y) = 5 + 7 = 12

-173. **E(2X + 1.5Y)** = 2E(X) + 1.5E(Y) = 20.5

-174. independent

-175. X_1 = test 1 grades, X_2 = test 2 grades, E(X1 + X2) = 150

-176. assuming independence, $Var(X1 + X2) = (10)^2 + (12)^2 = 244$

sd(X1 + X2) = 15.62

-177. **E(X1 – X2)** = 10

-178. **sd(X1 – X2)** = $\sqrt{(10^2 + 12^2)}$ = 15.62 (Variances always add)

-179. **E(.5X2 + 50)** = .5E(X2) + 50 = .5(70) + 50 = 85

-180. New sd for #162 **sd(.5X2 + 50)** = .5(12) = 6

-181. **sd(X1 + 7)** = sd(X1) = 10 (Adding/subtracting a constant to each value does not change the standard deviation

-182. **E(2X1 – 80)** = 2E(X1) – 80 = 2(80) – 80 = 80

-183. **sd(2X1 – 80)** = 2sd(X1) = 2(10) = 20

-184. ***1:*** *The number of observations n is fixed.* ***2:*** *Each observation is independent.* ***3:*** *Each observation represents one of two outcomes ("success" or "failure").* ***4:*** *The probability of "success" p is the same for each outcome.*

-185. ***1:*** *Keep going until you get your first success.* ***2:*** *Each observation is independent.* ***3:*** *Each observation represents one of two outcomes ("success" or "failure").* ***4:*** *The probability of "success" p is the same for each outcome.*

-186. Binomial, n = 4, p = .25 (2 outcomes, fixed no. of trials)

-187. **P(X = 2)** = $_4C_2(.25)^2(.75)^2$ = binompdf(4, .25, 2) = .211

-188. **P(X<3)** = P(X≤2) = P(X=0) + P(X=1) + P(X = 2) = $_4C_0(.25)^0(.75)^4 + {}_4C_1(.25)^1(.75)^3 + {}_4C_2(.25)^2(.75)^2 =$ binomcdf(4, .25, 2) = .949

-189. **P(X≥1)** = 1 – P(X ≤ 0) = 1 - $_4C_0(.25)^0(.75)^4$ = 1 – binompdf(4, .25, 0) = .6836

-190. **P(1≤X≤3)** = P(X=1) + P(X=2) + P(X = 3) = $_4C_1(.25)^1(.75)^3 + {}_4C_2(.25)^2(.75)^2 + {}_4C_3(.25)^3(.75)^1$ = .679

-191. **P(2<X<4)** = P(X = 3) = $_4C_3(.25)^3(.75)^1$ = .047

-192. **E(X)** = np = (4)(.25) = 1

-193. σ = $\sqrt{npq} = \sqrt{4(.25)(.75)}$ = .866

-194. geometric

-195. **P(X = 1)** = $(1 - .25)^0$ (.25) = .25

-196. **P(X≤2)** = P(X=1) + P(X=2) = $(.75)^0(.25) + (.75)^1(.25)$ = geometcdf(.25,2) = .4375

-197. **P(X>5)** = $(1 - .25)^5$ = .2373 (1^{st} success occurs after the 5^{th} trial)

-198. **P(2≤X<4)** = P(X=2) + P(X=3) = $(.75)^1(.25) + (.75)^2(.25)$ = .328

-199. **P(2<X≤5)** = P(X=3) + P(X=4) + P(X=5) =$(.75)^2(.25) + (.75)^3(.25) + (.75)^4(.25)$ = .3252

-200. μ = 1/.25 = 4