Statistics in Perspective

118. Every year, the U.S. gives approximately \$ 50 billion in foreign aid. Per-capita, that equals about _______ dollars per U.S. citizen.
Answer: $\frac{50{,}000{,}000{,}000}{300{,}000{,}000} = 167$

119. About 2.3 million people die of AIDS each year. That is about one every _______ seconds.
Answer: $31536000/2300000 = 13.7$

120. In 2013, China had a population 1.35 billion, and a GDP of \$9.19 trillion. What was the Chinese per-capita GDP ?
Answer: $\frac{9{,}190{,}000{,}000{,}000}{1{,}350{,}000{,}000} = 6807$

121. Suppose you drove 100 miles at 50 MPH, then drove 210 miles at 70 MPH. Find your average speed for that trip.
Answer: The first part of the trip takes 2 hours, and the second part takes 3 hours. So you went a total of 310 miles in 5 hours. The average speed is $\frac{310}{5} = 62$ MPH.

122. At a certain college, 70% of female applicants were admitted, and 60% of male applicants were admitted. Is it necessarily true that 65% of all applicants were admitted?
Answer: not unless there were an equal number of male and female applicants

123. In a recent year, Jefferson County residents spent 8.4 million dollars on state lottery tickets. If the per-capita spending was \$165, then what is the population of Jefferson Co ?
Answer: solve $\frac{8{,}400{,}000}{n} = 165$ to get $n = 50909$

Probability and Expected Value

124. Suppose that when a student parks in a faculty parking spot, she has a 23% chance of getting a ticket. What is the probability that she does not get a ticket?
Answer: $1 - .23 = .77 = 77\%$

125. If Carson-Newman's soccer team has a 53% chance of winning and a 29% chance of losing, then what is the probability that the game ends in a tie?
Answer: $1 - .53 - .29 = .18 = 18\%$

126. Let the random variable X be a student's grade on a statistics test. If $P(X \leq 79) = .37$, then what is $P(X > 79)$?
Answer: $1 - .37 = .63$

127. Is this an example of (primarily) empirical, subjective, or theoretical probability?
    (a) There is a 12% chance that the Lady Vols will win the national championship.
    Answer: subjective
    (b) There is a 25% chance of drawing a spade from a well shuffled deck of cards.
    Answer: theoretical
    (c) The stock market has gone up 75 of the last 100 years, so there is a 75% chance it will go up in a given year.
    Answer: empirical
    (d) The stock market has a 95% chance of going up this year because the president will fix the economy.
    Answer: subjective

(e) 60 of 100 patients got better after taking this drug, so there is a 60% chance that a particular patient will benefit from the drug.
Answer: empirical

(f) There is a 1/6 probability of rolling doubles in Monopoly.
Answer: theoretical

(g) There is a 37% chance of finding life on Mars.
Answer: subjective

128. Suppose the probability that the republican candidate wins the election is 59%.

(a) What is the probability that republican does not win?
Answer: .41

(b) Does that necessarily mean that the probability that the democrat wins is 41% ?
Answer: no, there could be a third party candidate running

129. The University of Tennessee graduation rate is about 60%. If you asked a new UT student to give the probability that he/she would eventually graduate, would you expect him/her to answer 60%? Explain.
Answer: no, he/she would give a subjective probablity

130. Suppose that a fan survey shows that 1234 students are rooting for Tennessee to beat Florida, and that 219 are rooting for Florida. If a student is selected at random,

(a) What is the empirical probability that he/she is rooting for Florida?
Answer: $219/1453 = .151$

(b) What is the empirical probability that he/she is rooting for Tennessee?
Answer: $1234/1453 = .849$

131. Here is the cumulative distribution of a college's student body.

class	PDF	CDF
Fr		.31
So		.54
Jr		.74
Sr		.92
Gr		1

(a) Fill out the probability distribution function (PDF).
Answer:

class	PDF	CDF
Fr	.31	.31
So	.23	.54
Jr	.20	.74
Sr	.18	.92
Gr	.08	1

(b) What proportion are sophomores?
Answer: .23

(c) What percentage are undergrads?
Answer: 92%

(d) If a student is selected at random, what is the probability that he/she is either a freshman or senior?
Answer: $.31 + .18 = .49$

132. Here is a stem plot for weight (in grams) for lab mice.

1	46799
2	00134445668
3	011257
4	1

(a) Find the empirical probability that a mouse weighs at least 30 grams.
Answer: $7/23 = .304$

(b) Find the median weight.
Answer: 24

133. Recall the bear cub litter size distribution:

size (X)	frequency	PDF	CDF
1	71	.119	.119
2	206	.346	.465
3	238	.400	.865
4	74	.124	.999
5	6	.010	1
	595		

Let the random variable X be the number of cubs in the litter.

(a) Find the empirical probability that a birth results in triplets.
Answer: $238/595 = .4$

(b) Find the empirical probability of more than 3 cubs.
Answer: $80/595 = .134$

(c) Find the empirical probability of at least 3 cubs.
Answer: $318/595 = .534$

(d) Find the empirical probability of a multiple birth (not just one cub).
Answer: $524/595 = .881$

134. Here is a stem plot for the total points scored in Super Bowls I-XLIII.

2	12379
3	01116777899
4	13445566777
5	002345669
6	115699
7	5

(a) Find the empirical probability that at least 35 points are scored.
Answer: $34/43 = .791$

(b) Find the empirical probability that an odd number of points are scored.
Answer: $28/43 = .651$

(c) Find the empirical probability that fewer than 20 points are scored.
Answer: $0/43 = 0$

135. Two basketball players shot 80% from the free throw line. Jones was 160/200, and Smith was 4/5. Explain in your own words how sample size is vital when assessing empirical probabilities. The "law of ______________" says that empirical probability becomes more accurate as sample size increases.
Answer: Law of Large Numbers - results in small samples are heavily influenced by luck/randomness, so the bigger the sample size the more reliable the empirical probabilities.

136. Here is the empirical probability distribution for the X, the number of regular season wins for an NFL team. Data is taken from the last 25+ years.

wins	PDF	CDF
0	0.00119	0.00119
1	0.01071	0.01190
2	0.02500	0.03690
3	0.03452	0.07143
4	0.07262	0.14405
5	0.07976	0.22381
6	0.09286	0.31667
7	0.11190	0.42857
8	0.12024	0.54881
9	0.11905	0.66786
10	0.11429	0.78214
11	0.08452	0.86667
12	0.06786	0.93452
13	0.03810	0.97262
14	0.02143	0.99405
15	0.00476	0.99881
16	0.00119	1.00000

(a) Find the 5 number summary.
Answer: $0, 6, 8, 10, 16$

(b) $P(X = 6)$
Answer: .09286

(c) $P(X \neq 6)$
Answer: $1 - P(X = 6) = 1 - .09286 = .90714$

(d) $P(X \leq 6)$
Answer: .31667

(e) $P(X < 6)$
Answer: $P(X \leq 5) = .22381$

(f) $P(X > 6)$
Answer: $1 - P(X \leq 6) = 1 - .31667 = .68333$

(g) $P(X \geq 6)$
Answer: $1 - P(X \leq 5) = 1 - .22381 = .77619$

(h) Use your calculator to find the expected number of wins, $\overline{x}$.
Answer: 1-varstats L1,L2 gives $\overline{x} = 8$

(i) Ten wins is the _______ percentile.
Answer: $CDF(10) = .78214$, so the 78th percentile

(j) Describe the shape of the distribution.
Answer: symmetric, bell-shaped

137. Suppose that in a national household survey, the following number of dogs, X, were owned:

dogs	freq
0	175
1	81
2	30
3	12
4	4
5 or more	negligible

(a) What percentage of households own at least one dog?
Answer: $127/302 = .42$

(b) Find $P(X > 1)$
Answer: $46/302 = .152$

(c) Find the expected number of dogs in a randomly picked household.
Answer: mean 0.64 dogs per household

(d) If the mean U.S. household size is 2.6 persons, then estimate the number of domesticated dogs in the U.S.
Answer: $300/2.6 = 115$ million households, at .64 dogs each gives around 74 million dogs

138. Suppose a die is weighted so that the probability distribution for the face-up value is:

x	$P(x)$
1	.12
2	.08
3	.15
4	.19
5	.24
6	.22

Find the expected value of x.
Answer: $\mu = 4.01$

139. Suppose I flip 7 fair coins. Let the random variable X be the number of heads observed. The CDF is given by:

X	CDF
0	.0078
1	.0625
2	.2266
3	.5000
4	.7734
5	.9375
6	.9922
7	1

Find the probabilities:

(a) $P(X = 0)$
Answer: .0078

(b) $P(X = 1)$
Answer: $.0625 - .0078 = .0547$

(c) $P(X \leq 4)$
Answer: .7734

(d) $P(X > 4)$
Answer: $1 - .7734 = .2266$

(e) $P(X < 4)$
Answer: .5

(f) $P(X = 4)$
Answer: $.7734 - .5 = .2734$, or $1 - .2266 - .5 = .2734$

140. Here is part of a probability distribution table:

x	CDF
25	.35
35	.45

(a) The 35th percentile is _______.
Answer: $x = 25$, since 35% are at or below $x = 25$

(b) $x = 35$ is the _______ percentile.
Answer: 45th percentile, since 45% are at or below $x = 35$

141. Consider this information about a distribution:

X	PDF	CDF
⋮	⋮	⋮
8	.18	.54
⋮	⋮	⋮
12	.07	.88
⋮	⋮	⋮

(a) $P(X \leq 8)$
Answer: .54

(b) $P(X < 8)$
Answer: .54 - .18 = .36

(c) $P(8 \leq X \leq 12)$
Answer: $.88 - .36 = .52$

142. Let the random variable X be the departure delay (rounded to the nearest minute) for flights from Knoxville to Atlanta that do not depart on schedule.

Open the Excel file `http://massey.limfinity.com/data/knox-atl2.xlsx` from the class data folder. It contains the empirical probability distribution based on data from the year 2012.

(a) State the sample size.
Answer: $n = 599$

(b) Find the 5 number summary.
Answer: 1,6,20,53,434

(c) $X = 40$ minutes is the _______ percentile.
Answer: $CDF(40) = .67947$, so 40 minutes is about the 68th percentile

(d) The 40th percentile is $X =$ _______.
Answer: 13 minutes

(e) How would you describe the distribution's shape?
Answer: right-skewed

(f) Find the probability that a delayed flight will be less than 15 minutes late, $P(X < 15)$.
Answer: $CDF(14) = .41569$

(g) Find the probability that a delayed flight will no more than 15 minutes late, $P(X \leq 15)$.
Answer: $CDF(15) = .44240$

(h) Find the probability that a delayed flight will at least 15 minutes late, $P(X \geq 15)$.
Answer: $1 - CDF(14) = 1 - .41569 = .58431$

(i) Find the probability that a delayed flight will be at least an hour late, $P(X \geq 60)$.
Answer: $1 - CDF(59) = 1 - .77462 = .22538$

(j) Find $P(X < 30)$.
Answer: $CDF(29) = .61269$

(k) Find $P(15 \leq X < 30)$.
Answer: $CDF(29) - CDF(14) = .61269 - .41569 = .197$

(l) Find $P(60 \leq X \leq 120)$.
Answer: $CDF(120) - CDF(59) = .91653 - .77462 = .14191$

(m) Find the expected length of a delay.
Answer: $\overline{x} = \frac{24047}{599} = 40.1$ minutes

143. Fill out the PDF column, and then answer the questions.

X	PDF	CDF
4		.38
5		.43
6		.77
7		1

Answer:

X	PDF	CDF
4	.38	.38
5	.05	.43
6	.34	.77
7	.23	1

(a) Find the mean.
Answer: $\mu = (4)(.38) + (5)(.05) + (6)(.34) + (7)(.23) = 5.42$

(b) Find the median.
Answer: 6

(c) Find the mode.
Answer: 4

(d) Find the expected value of X.
Answer: 5.42

(e) Does the term "expect" refer to the mean, median, or mode of a distribution ?
Answer: mean

144. An investor has an offer to invest in a struggling company that will most likely go bankrupt. In that case, she receives nothing. But if the company recovers, she will be entitled to 5 million dollars. She estimates there is a 6% chance of recovery. What is the expected payout of this investment ?
Answer: remember, expected value is the mean, not the mode, so $\mu = (.94)(0)+(.06)(5) = .3$ million, or 300 thousand dollars.

145. You have two possible methods to complete a task:

- Method A: 80% chance it takes 3 hours, and 20% chance it takes 4 hours.
- Method B: 50% chance it takes 2 hrs, 40% chance it takes 3 hrs, and 10% chance it takes 6 hrs.

Which method has the lower expected time to completion ?
Answer: $\mu_A = (.8)(3) + (.2)(4) = 3.2$ hours
$\mu_B = (.5)(2) + (.4)(3) + (.1)(6) = 2.8$ hours is lower

146. A movie producer gives this subjective probability distribution for the net revenue (x in millions) on particular movie project:

event	x	$P(x)$
flop	10	.17
mediocre	35	.30
hit	70	.38
blockbuster	120	

(a) What is the probability of a blockbuster?
Answer: $p = .15$

(b) What is the expected revenue?
Answer: $\mu = 56.8$ million

(c) Do you think it is wise to spend \$ 65 million to produce the movie?
Answer: no, since on average, you'd lose money

147. A dairy farmer has cows that produce about 80,000 gallons of milk per year. There is some variation in the market price of milk, and here is the farmer's estimate for the probability distribution of his per-gallon profit, x.

x	$P(x)$
1.25	.05
1.00	.12
0.75	.28
0.50	.22
0.25	.15
0.00	.10
-0.25	.08

Find the expected value of x, and then the farmer's total expected profit.
Answer: $\mu = (1.25)(.05) + \cdots + (-0.25)(.08) = .52$ per gallon. Multiply by 80,000 gallons to get $\$41,600$

148. At a carnival game, you have these possible outcomes:

event	probability	result
lose	85%	-1
small prize	10%	+2
big prize	5%	+B

(a) Find the value of B that makes this game "fair", which means the expected result is zero.
Answer: $(.85)(-1) + (.10)(2) + (.05)(B) = 0$ to get $B = 13$

(b) Suppose the big prize is $B = 20$. Calculate your statistical "edge", which is your postive expected value every time your play this game.
Answer: $(.85)(-1) + (.10)(2) + (.05)(20) = .35$

149. In a hand of poker, the pot stands at 500 chips, and you are the last to play and are deciding whether to call a 100 chip bet. You believe you have a 25% chance of having the winning hand. So the probability distribution for your profit x if you call looks like:

event	x	$P(x)$
you have the winning hand	500	.25
you have a losing hand	-100	.75

Find the expected value of calling.
Answer: $\mu = (.25)(500) + (.75)(-100) = 50$

150. Suppose you have an opportunity to invest ten thousand dollars in a new restaurant. You will be entitled to a 5% share of all future profits. A due diligence study of the restaurant's prospects leads you to estimate the following probability distribution of the current value of the profits (x in thousands):

x (profit)	PDF
0	0.35
100	0.25
200	0.20
500	0.10
1000	0.07
2000	0.02
5000	0.01

(a) At what value of x would you break even?
Answer: $(.05)x = 10$ implies you need $x = 200$ profit to break even

(b) Find the probability that you at least break even.
Answer: $.20 + .10 + .07 + .02 + .01 = .40$

(c) Find the expected value of x; what is your share of that?
Answer: $\mu = (0)(.35) + (100)(.25) \cdots + (2000)(.02) + (3000)(.01) = 275$, and 5% of that is 13.75 thousand

(d) Would you invest in this venture? Explain.
Answer: It's a hard decision. On average, you'd come out ahead - investing 10,000 to make an average of 13,750. However, you would be taking on a lot of risk. You have a high probability of losing money. There's only a 20% chance that the restaurant will be a money-maker, and most of the expectation comes from the small chance that the restaurant is a huge hit.

151. A company is being sued for $ 10 million, and estimates that they have a 30% chance of losing if they go to court. (Assume it would be an all-or-nothing verdict). An offer is on the table to settle out-of-court for $ 2 million. Complete this table:

	expected outcome μ	standard deviation σ
go to court		
settle		

Answer:

	expected outcome	standard deviation σ
go to court	3	4.58
settle	2	0

152. Go to `http://www.random.org/calendar-dates/`. And have it pick 50 random dates between Jan 1 and Dec 31 of this year. Use your simulation to find the empirical probability that the day of the month is larger than the month of the year.

For example, Oct 21 would satisfty the condition since October is the 10th month, and $21 > 10$. But Oct 5 would not satisfy the condition since $5 < 10$.
Answer: count how many had month > day, out of 50

Theoretical Probability

153. Suppose you draw a card from a shuffled standard 52-card deck.

(a) Find the probability that it is an face card and a heart.
Answer: $3/52 = .0577$

(b) Find the probability that it is an face card or a heart.
Answer: $22/52 = .423$

(c) Find the probability that it is a face card or an ace.
Answer: $16/52 = .308$

(d) Find the probability that it is a face card and an ace.
Answer: impossible, disjoint, so 0

154. Two cards are drawn from a deck of cards with replacement, and you observe the suit of each card.

(a) Finish writing the sample space $\Omega = \{SS, SH, SD, SC, HS, \cdots\}$
Answer: Ω ={SS,SH,SD,SC,HS,HH,HD,HC,DS,DH,DD,DC,CS,CH,CD,CC}

(b) Assuming equally likely outcomes, find the probability of drawing at least one spade.
Answer: 7/16

155. Suppose a couple plans to have three children (oldest, middle, youngest).

(a) Finish writing the sample space for the genders of the children, $\Omega = \{FFF, FFM, \cdots\}$.

(b) How many possible outcomes are contained in Ω ?
Answer: 8

(c) Assuming equally likely outcomes, what is the probability of having exactly two girls?
Answer: $3/8 = .375$

(d) Assuming equally likely outcomes, what is the probability of having at least one girl?
Answer: $7/8 = .875$

(e) Assuming equally likely outcomes, what is the probability of having all children of the same gender?
Answer: $2/8 = .25$

156. Suppose you roll two dice. Let the random variable X be the product of the numbers. (e.g. if you rolled a 3 and a 6, the product is $X = 18$).

(a) How many different observations are in Ω ?
Answer: 36

(b) Write all pairs of dice with a product of at least 15.
Answer: 35,36,44,45,46,53,54,55,56,63,64,65,66

(c) Find $P(X \geq 15)$.
Answer: $13/36 = .361$

(d) Find $P(X > 15)$.
Answer: $11/36 = .306$

(e) Find $P(X \leq 15)$.
Answer: $25/36 = .694$

(f) Find $P(15 \leq X \leq 18)$
Answer: $5/36 = .139$

157. This is the sample space if you flip 5 coins:

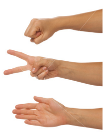
```
HHHHH,HHHHT,HHHTH,HHHTT,
HHTHH,HHTHT,HHTTH,HHTTT,
HTHHH,HTHHT,HTHTH,HTHTT,
HTTHH,HTTHT,HTTTH,HTTTT,
THHHH,THHHT,THHTH,THHTT,
THTHH,THTHT,THTTH,THTTT,
TTHHH,TTHHT,TTHTH,TTHTT,
TTTHH,TTTHT,TTTTH,TTTTT
```

Let the random variable X be the number of switches (from tail to head or from tail to head), e.g. $HTHHT \to 3$ and $HHHTT \to 1$. Fill out the PDF (cross off as you go).

X	PDF
0	2/32
1	
2	
3	
4	2/32

Answer: 8/32, 12/32, 8/32

158. To decide who has to do a chore, three brothers simultaneously throw rock, paper, or scissors.



(a) Finish writing the sample space $\Omega = \{RRR, RRP, RRS, RPR, RPP, RPS, \cdots\}$ (there should be 27 possible outcomes).

(b) Assuming equally likely outcomes, write the distribution for X, the number of distinct symbols thrown. e.g. for SRS there are $X = 2$ distinct symbols (scissors and rock).

X	$P(X)$
1	
2	
3	

Answer:

X	$P(X)$
1	3/27
2	18/27
3	6/27

(c) Find the probability that there is an "odd man out", i.e. $X = 2$.
Answer: 18/27

(d) Find the expected value of X.
Answer: $\mu = \frac{57}{27} = 2.11$

159. Suppose your roll two dice; one is tetrahedral (4-sided, labeled 1-4) and the other is octahedral (8-sided, labeled 1-8). Let the random variable X be the difference between the octahedral and tetrahedral results. e.g. if the 8-sided die shows 5 and the 4-sided die shows 2, then $X = 5 - 2 = 3$.
(note: the difference could be negative)

(a) Above each member of the sample space, write the corresponding value of X.

$$\Omega = \{11, 12, 13, 14, 15, 16, 17, 18$$
$$21, 22, 23, 24, 25, 26, 27, 28$$
$$31, 32, 33, 34, 35, 36, 37, 38$$
$$41, 42, 43, 44, 45, 46, 47, 48\}$$

Answer:
0,1,2,3,4,5,6,7
-1,0,1,2,3,4,5,6
-2,-1,0,1,2,3,4,5
-3,-2,-1,0,1,2,3,4

(b) Assuming equally likely outcomes, write the probability distribution for X.

X	$P(X)$
-3	$1/32$
-2	$2/32$
-1	
0	
1	
2	
3	
4	
5	
6	
7	
	$32/32$

Answer:

X	$P(X)$
-3	$1/32$
-2	$2/32$
-1	$3/32$
0	$4/32$
1	$4/32$
2	$4/32$
3	$4/32$
4	$4/32$
5	$3/32$
6	$2/32$
7	$1/32$

(c) Find the expected value of X.
Answer: $\mu = 2$

(d) Find the probability that X is negative.
Answer: $6/32 = .1875$

(e) Find the probability that X is positive.
Answer: $22/32 = .6875$

(f) Find $P(X < 1 \text{ or } X > 5)$.
Answer: $13/32$

(g) Find $P(X < 5 \text{ or } X > 1)$.
Answer: certain, so 1

(h) Find $P(X < 1 \text{ and } X > 5)$.
Answer: impossible since these are disjoint events, so zero

(i) Find $P(X < 5 \text{ and } X > 1)$.
Answer: 12/32

160. Look up the birthdays of your favorite NFL team's roster, e.g.

`http://www.nfl.com/teams/roster?team=TEN`

There should be 53 players on the active (ACT) roster. Do any two active players have the same birthday (not necessarily the same year)?

Use this webpage to simulate the birthday problem for $k = 53$.

`http://massey.limfinity.com/apps/birthday.php`

Use a sample size of at least 50000 simulations. According to the law of large numbers, what is the approximate probability that two players on a 53-man roster will share a birthday?
Answer: the probability is about .981.

Joint Probability

161. If the probability of C-N mens' soccer winning the SAC championship is 14%, and the probability of C-N womens' soccer winning the SAC is 31%, what is the probability that neither happens? (assume independence)
Answer: $(1 - .14)(1 - .31) = (.86)(.69) = .5934$

162. Suppose the probability of Virginia Tech winning their game is 75%. The probability of Tennessee winning is 40%. (They are not playing each other.)

(a) Find the probability that both teams win.
Answer: $(.75)(.40) = .3$

(b) Find the probability that neither team wins.
Answer: $(.25)(.60) = .15$

(c) Find the probability that one team wins but the other loses.
Answer: $1 - .3 - .15 = .55$

163. Suppose a bag contains 9 orange and 3 blue M&M's. If you select two pieces at random, find the probability that

(a) both are orange
Answer: $(9/12)(8/11) = .5455$

(b) both are blue
Answer: $(3/12)(2/11) = .0455$

(c) they are different colors
Answer: $1 - .5455 - .0455 = .409$

164. Suppose a bag contains 90 orange and 30 blue M&M's. If you select two pieces at random, find the probability that

(a) both are orange
Answer: $(90/120)(89/119) = .5609$

(b) both are blue
Answer: $(30/120)(29/119) = .0609$

(c) they are different colors
Answer: $1 - .5609 - .0609 = .3782$

165. Suppose that

- there is a 12% chance of discovering extraterrestrial life by the year 2015
- there is a 60% chance that a SEC team will win the next college football championship
- there is a 25% chance that the USA or Israel will make an air strike on Iran this year

(a) Find the probability that all 3 of these events take place.
Answer: $(.12)(.60)(.25) = .018$

(b) Find the probability that none of these events take place.
Answer: $(.88)(.40)(.75) = .264$

(c) Find the probability that at least one of these events take place.
Answer: P(not none of them) is $1 - .264 = .736$

166. Is it reasonable to treat the events as independent?

(a) you catch a green light, and they play your favorite song on the radio
Answer: independent

(b) a person is the CEO of a company, and male
Answer: dependent

(c) there is precipitation, and the power goes out
Answer: dependent

(d) Virginia Tech wins a football game, and the stock market goes up
Answer: independent

(e) a wife lives to be 90, and her husband lives to be 90
Answer: dependent

(f) stranger A lives to be 90, and stranger B lives to be 90
Answer: independent

(g) I forget to set my wake-up alarm, and I get a speeding ticket.
Answer: dependent

(h) my house catches on fire, and your house catches on fire
Answer: independent, unless our houses are in close proximity

(i) the stock market goes up, and there is a solar eclipse
Answer: independent

167. Suppose events A and B are independent with $P(A) = .6$ and $P(B) = .8$.

(a) Find $P(A \text{ and } B)$
Answer: $(.6)(.8) = .48$

(b) Find $P(A \text{ or } B)$
Answer: $.6 + .8 - .48 = .92$

168. Suppose that the LA Lakers have a 60% chance of scoring at least 100 points in an NBA game. The Memphis Grizzlies have a 30% chance of scoring at least 100 points. If the Lakers play the Grizzlies, can you say that the probability of both teams scoring at least 100 points in that game is $(.60)(.30) = .18 = 18\%$? Explain.

Answer: No, if the teams play each other, the events are not independent. A fast paced game in which one team scores a lot makes it more likely that the other team also will. Therefore, the second probability in the sequence would have to be adjusted. Since the teams' scores are positively correlated, the answer is going to be higher than 18%.

169. Suppose a free throw shooter makes 75% of his shots, and the outcomes are independent. Find the probability that:

   (a) He makes both of the next two shots.
   Answer: $(.75)(.75) = (.75)^2 = .5625$

   (b) He makes each of the next three shots.
   Answer: $(.75)(.75)(.75) = (.75)^3 = .4219$

   (c) He makes each of the next four shots.
   Answer: $(.75)(.75)(.75)(.75) = (.75)^4 = .3164$

   (d) He makes each of the next ten shots.
   Answer: $(.75)^{10} = .0563$

   (e) He misses both of the next two shots.
   Answer: $(.25)(.25) = (.25)^2 = .0625$

   (f) He makes exactly one of the next two shots.
   Answer: $1 - .5625 - .0625 = .375$

170. The NFL injury report classified players into broad categories based on the probability of playing:

status	chance of playing
out	0 %
doubtful	25 %
questionable	50 %
probable	75 %

Suppose your fantasy football team has two "probables", one "questionable", and one "doubtful" player.

   (a) Find the probability that they all play.
   Answer: $(.75)^2(.5)(.25) = .0703$

   (b) Find the probability that at least one of them plays.
   Answer: $1 - (.25)^2(.5)(.75) = .977$

171. A box of donuts has 8 glazed and 4 chocolate. If you pick two out at random (without replacement), find the probability that you get one of each type.
Answer: $(8/12)(4/11) + (4/12)(8/11) = .485$

172. A bag of candy has 7 Snickers, 5 Reece's, and 2 Almond Joys. Assuming each piece is equally likely to be chosen, if you select three pieces, find the probability that:

   (a) You get at least one Almond Joy.
   Answer: $1 - (12/14)(11/13)(10/12) = .396$

   (b) You get 3 Snickers.
   Answer: $(7/14)(6/13)(5/12) = .096$

   (c) You get 3 Almond Joys.
   Answer: impossible, so the probability is 0

   (d) Find the probability of SRA, (your first piece is Snickers, the second is Reece's, and the third is Almond Joy).
   Answer: $(7/14)(5/13)(2/12) = .032$

(e) You get one of each type of candy.
Answer: Each probability for {SRA, SAR, RSA, RAS, ASR, ARS} is the same, so answer is $6 \times (7/14)(5/13)(2/12) = .192$

173. A baseball team plays a double-header. They have a 85% chance of winning the first game, and a 60% chance of winning the second game. Assuming independent outcomes, write the distribution for the number of wins they will get. Then find the expected number of wins.

wins	probability
0	
1	
2	

Answer:

wins	probability
0	.06
1	.43
2	.51

$\mu = 1.45$

174. If you draw 2 cards from a shuffled standard 52-card deck.

(a) Find the probability that both are clubs.
Answer: $\left(\frac{13}{52}\right)\left(\frac{12}{51}\right) = .0588$

(b) Find the probability that both are the same suit.
Answer: $4\left(\frac{13}{52}\right)\left(\frac{12}{51}\right) = .2353$

175. If you draw 3 cards from a shuffled standard 52-card deck.

(a) What is the probability that you get three hearts?
Answer: $(13/52)(12/51)(11/50) = .0129$

(b) What is the probability that you get all 3 cards of the same suit?
Answer: $4(13/52)(12/51)(11/50) = .0518$

176. A computer chip manufacturing plant tests each chip before shipping to the customer. Over the long-term, 5 percent of the chips are defective and must be discarded.

(a) What is the probability that a chip passes the test (is not defective) ?
Answer: $1 - .05 = .95$

(b) In a batch of 20 chips, find the probability that none of them are defective.
Answer: $(.95)^{20} = .358$

(c) In a batch of 20 chips, find the probability that at least one is defective.
Answer: $1 - (.95)^{20} = .642$

(d) In a batch of 40 chips, find the probability that none of them are defective.
Answer: $(.95)^{40} = .129$

(e) In a batch of 40 chips, find the probability that at least one is defective.
Answer: $1 - (.95)^{40} = .871$

177. Poindexter is trying to get a date. Suppose that each girl responds independently, and each has a 90% chance of rejecting the young suitor.

(a) He has been rejected by 9 girls in a row. Is the tenth girl guaranteed to say yes?
Answer: no, .9 is a long-run probability, and might not be exact for any particular sample

(b) What is the probability that he gets at least one acceptance if he asks out 10 girls?
Answer: the complementary event is that he gets rejected by them all, so $1 - (.9)^{10} = .651$.

(c) What is the probability that he gets exactly one acceptance if he asks out 10 girls?
Answer: There are ten sequences that meet this condition: YNNNNNNNNN, NYNNNNNNNN, NNYNNNNNNN, etc. each one has probability $(.1)(.9)^9 = .0387$, so the answer is .387

178. Imagine a game of Russian roulette with an n chambered gun, so that if you play once, your probability of dying is $\frac{1}{n}$, and your probability of surviving is $\frac{n-1}{n}$.

If $n = 5$ and you are forced to play 5 times, then your probability of surviving is $(4/5)^5 = 0.328$.

(a) Find the probability of surviving if $n = 10$ and you must play 10 times.
Answer: $(.9)^{10} = .349$

(b) Find the probability of surviving if $n = 100$ and you must play 100 times.
Answer: $(.99)^{100} = .366$

(c) Find the probability of surviving if $n = 1000$ and you must play 1000 times.
Answer: $(.999)^{1000} = .368$

(d) Find the probability of surviving if $n = 1000000$ and you must play 1000000 times.
Answer: $(.999999)^{1000000} = .368$

179. The president has a 81% chance of being re-elected. A company has a 30% chance of declaring bankruptcy. There is a 62% chance that an SEC team wins the football national championship.

(a) Find the probability that all three events occur.
Answer: $(.81)(.30)(.62) = .1507$

(b) Find the probability that none of the events occur.
Answer: $(.19)(.70)(.38) = .0505$

(c) Find the probability that exactly one of the events occurs.
Answer: $(.81)(.70)(.38) + (.19)(.30)(.38) + (.19)(.70)(.62) = .3196$

(d) Find the probability that exactly two of the events occurs.
Answer: $(.81)(.30)(.38) + (.81)(.70)(.62) + (.19)(.30)(.62) = .4792$

180. A basketball team plays three games this week, with independent probabilities of winning 90%, 75%, and 50% respectively. Write the distribution for the number of wins they will get. Then find the expected number of wins.

wins	probability
0	
1	
2	
3	

Answer:

wins	probability
0	.0125
1	.1625
2	.4875
3	.3375

$\mu = 2.15$

181. A bag contains 15 green and 10 black marbles. You randomly select three marbles with replacement.

(a) What is the probability that all three are black?
Answer: $(.4)(.4)(.4) = .064$

(b) What is the probability that at least one is black?
Answer: not all green, so $1 - (.6)^3 = .784$

(c) What is the probability that exactly one is black?
Answer: You could get BGG, GBG, or GGB, so there are three ways to do it: $3(.4)(.6)(.6) = .432$.

(d) What is the probability that exactly two are black?
Answer: You could get BBG, GBB, or BGB, so there are three ways to do it: $3(.4)(.4)(.6) = .288$.

182. A class contains 15 girls and 10 boys. You select a three person committee via a simple random sample (without replacement).

(a) What is the probability of selecting 3 boys?
Answer: $(10/25)(9/24)(8/23) = .052$

(b) What is the probability of selecting at least 1 boy?
Answer: not all girls, so $1 - (15/25)(14/24)(13/23) = .802$

(c) What is the probability of selecting exactly 1 boy?
Answer: You could get BGG, GBG, or GGB, so there are three ways to do it: $3(10/25)(15/24)(14/23) = .457$.

(d) What is the probability of selecting exactly 2 boys?
Answer: You could get BBG, GBB, or BGB, so there are three ways to do it: $3(10/25)(9/24)(15/23) = .293$.

183. A class contains 18 males and 7 females. If you select two at random without replacement, find the probability that your pair is mixed gender.
Answer: MF or FM, so $(18/25)(7/24) + (18/25)(7/24) = .42$

184. Suppose you commit $1000 to a two year investment. Each year, it has a 60% chance of doubling in value, and a 40% chance that it will lose half its value.

(a) Find the probability that it doubles both years. What is the final value of your investment in that case?
Answer: $(.6)(.6) = .36$, the investment would be worth $1000(2)(2) = 4000$

(b) Find the probability that it loses half of its value both years. What is the final value of your investment in that case?
Answer: $(.4)(.4) = .16$, the investment would be worth $1000(.5)(.5) = 250$

(c) Find the probability that doubles the first year, then loses half of its value the next year. What is the final value of your investment in that case?
Answer: $(.6)(.4) = .24$, the investment would be worth $1000(2)(.5) = 1000$

(d) Find the probability that loses half the first year, then doubles the next year. What is the final value of your investment in that case?
Answer: $(.4)(.6) = .24$, the investment would be worth $1000(.5)(2) = 1000$

(e) Find the expected value of your investment after the two years have passed.
Answer: $(.36)(4000) + (.16)(250) + (.48)(1000) = 1960$

(f) Use the geometric mean to find the average annual increase that corresponds to the expected outcome.
Answer: $\left(\frac{1960}{1000}\right)^{1/2} = 1.40$ so a 40% annual return

185. If $P(A) = .60$, $P(B) = .48$, and $P(A \text{ and } B) = .20$,

(a) Are A and B independent events?
Answer: no, since $(.60)(.48) = .288 \neq .20$

(b) Find $P(A \text{ or } B)$
Answer: $.60 + .48 - .20 = .88$

186. Among married couples, 21% of husbands die of cancer, and 18% of wives die of cancer. There is a 5% chance that both will die of cancer.

(a) Sketch a Venn diagram, and allocate 100 couples to the different regions.
Answer: 5 in the overlap, 16 other husbands, 13 other wives, and 66 outside the bubbles

(b) Are husband and wife death by cancer independent events ?
Answer: no, because $(.21)(.18) = .0378$, not $.05$

(c) Find the probability that for a random couple, either the husband or the wife dies of cancer.
Answer: $.21 + .18 - .05 = .34$

187. Suppose A and B are independent events with $P(A \text{ and } B) = .465$ and $P(A \text{ or } B) = .905$.
Then we know the following:

- $P(A)P(B) = .465$
- $P(A) + P(B) - .465 = .905$

Solve those equations for $P(A)$ and $P(B)$. (Hint: use the quadratic formula)
Answer: $P(A) = .75$, $P(B) = .62$

188. A 52 card deck is shuffled and dealt to four players.

(a) Find the probability that after one card is dealt to each player, they all have different suits.
Answer: $\left(\frac{39}{51}\right)\left(\frac{26}{50}\right)\left(\frac{13}{49}\right) = .1055$

(b) Find the probability that after two cards are dealt to each player, each player has two cards of the same suit, a different suit for each player.
Answer: multiply the previous answer by $\left(\frac{12}{48}\right)\left(\frac{12}{47}\right)\left(\frac{12}{46}\right)\left(\frac{12}{45}\right)$ to get $.000468 \approx \frac{1}{2135}$

Binomial Distribution

189. A bowler makes a strike on 30% of his frames. If he bowls 10 frames, let $X \sim BI(10, .30)$ be the number of strikes he gets.

(a) Find $P(X = 3)$
Answer: $\mathit{bipdf}(10, .3, 3) = .267$

(b) Find $P(X \leq 3)$
Answer: $\mathit{bicdf}(10, .3, 3) = .650$

(c) Find $P(X \geq 3)$
Answer: $1 - \mathit{bicdf}(10, .3, 2) = .617$

(d) Find $P(X < 3)$
Answer: $\mathit{bicdf}(10, .3, 2) = .383$

(e) Find $P(X > 3)$
Answer: $1 - \mathit{bicdf}(10, .3, 3) = .350$

(f) Find $P(2 \leq X \leq 6)$
Answer: $\mathit{bicdf}(10, .3, 6) - \mathit{bicdf}(10, .3, 1) = .840$

190. Assume that each potential doner has a $\frac{1}{25}$ chance of giving to a cause. A fund raiser solicits 1000 potential doners. Using a binomial model,

(a) What is the expected number of doners?
Answer: $\mu = (1000)(1/25) = 40$

(b) Find the probability that at least 50 people agree to donate.
Answer: $1 - \mathit{bicdf}(1000, 1/25, 49) = .0663$

(c) Find the probability that the number of doners is in the 40's (40-49).
Answer: $\mathit{bicdf}(1000, 1/25, 49) - \mathit{bicdf}(1000, 1/25, 39) = .456$

191. Anna wins 70% of tennis sets against Bianca. Use a binomial model to find:

(a) The probability that Anna wins a majority of $n = 3$ sets versus Bianca.
Answer: $1 - \mathit{bicdf}(3, .7, 1) = .784$

(b) The probability that Anna wins a majority of $n = 5$ sets versus Bianca.
Answer: $1 - \mathit{bicdf}(5, .7, 2) = .837$

(c) The probability that Anna wins a majority of $n = 7$ sets versus Bianca.
Answer: $1 - \mathit{bicdf}(7, .7, 3) = .874$

192. If $X \sim BI(20, .17)$.

(a) Find $P(X = 2)$.
Answer: $\mathit{bipdf}(20, .17, 2) = .192$

(b) Find $P(X \leq 2)$.
Answer: $\mathit{bicdf}(20, .17, 2) = .315$

(c) Find $P(X < 4)$.
Answer: $\mathit{bicdf}(20, .17, 3) = .550$

(d) Find $P(X > 7)$.
Answer: $1 - \mathit{bicdf}(20, .17, 7) = .013$

(e) Find $P(X \geq 7)$.
Answer: $1 - \mathit{bicdf}(20, .17, 6) = .041$

(f) Find $P(2 \leq X \leq 5)$
Answer: $\mathit{bicdf}(20, .17, 5) - \mathit{bicdf}(20, .17, 1) = .768$

193. Use the binomial distribution to find the probabilities:

(a) The probability of a boy is 50%. Find the probability of getting exactly 4 boys in 5 kids.
Answer: $\mathit{bipdf}(5, .5, 4) = .156$

(b) A baseball player is hitting .320. Find the probability of getting exactly 3 hits in his next 10 at-bats.
Answer: $\mathit{bipdf}(10, .32, 3) = .264$

(c) C-N offers honors scholarships to 40 students. If each has a 25% of accepting, find the probability that 10 or fewer accept.
Answer: $\mathit{bicdf}(40, .25, 10) = .584$

(d) A student randomly guesses at 20 True/False questions. Find the probability that he passes (at least 60% correct).
Answer: $1 - \mathit{bicdf}(20, .5, 11) = .252$

(e) A medical procedure is effective 80% of the time. Find the probability that it fails exactly two out of seven times.
Answer: $\mathit{bipdf}(7, .2, 2) = .275$

(f) If 10% of all C-N nursing graduates are male, what is the probability of having at least 5 males in a class of 30?
Answer: $1 - \mathit{bicdf}(30, .1, 4) = .1755$

(g) A computer chip manufacturer finds that 5% of the chips are defective. Find the probability that fewer than 4 chips in a batch of 50 are defective.
Answer: $\mathit{bicdf}(50, .05, 3) = .760$

(h) A sports prognosticator is correct 60% of the time. Find the probability that he is correct on exactly seven of nine predictions.
Answer: $\mathit{bipdf}(9, .6, 7) = .161$

194. Suppose 26 children are registered to be picked up by the church bus on Wed. night. The bus has only 21 seats. Each child independently misses 30% of services.

(a) Let the random variable $X \sim BI(n, p)$ be the number of children that will take the bus on a particular night. Find n and p.
Answer: $n = 26$ and $p = .70$

(b) Find the expected number of children that want to ride the bus.
Answer: $\mu = (26)(.70) = 18.2$

(c) Find the probability that the bus is full, and each child that wanted to go has a seat.
Answer: $P(X = 21) = \mathit{bipdf}(26, .7, 21) = .0893$

(d) Find the probability that at least one child doesn't have a seat.
Answer: $P(X > 21) = 1 - \mathit{bicdf}(26, .7, 21) = .073$

195. Consider the random variables $x \sim BI(5, .2)$, $y \sim BI(5, .5)$, and $z \sim BI(5, .8)$.

(a) Look at the distributions online at `http://limfinity.com/apps/binomial.php`.

(b) Describe the shape of each distribution.
Answer: x is right skewed; y is symmetric; z is left skewed

(c) Find the mean and standard deviation of each distribution.
Answer: for x, $\mu = 1$ and $\sigma = .89$
for y, $\mu = 2.5$ and $\sigma = 1.12$
for z, $\mu = 4$ and $\sigma = .89$

(d) Which one has the highest standard deviation?
Answer: y

196. Compute μ and σ for these binomial random variables.

(a) $x \sim BI(20, .1)$
Answer: $\mu = 2$ and $\sigma = 1.34$

(b) $y \sim BI(4, .5)$
Answer: $\mu = 2$ and $\sigma = 1$

(c) $x \sim BI(10, .6)$
Answer: $\mu = 6$ and $\sigma = 1.549$

(d) $x \sim BI(1000, .6)$
Answer: $\mu = 600$ and $\sigma = 15.49$

197. If $x \sim BI(50, .14)$, then find μ and σ. What is the expected value of x?
Answer: $\mu = 7$, $\sigma = 2.45$; the expected value of x is 7

198. Suppose a college is recruiting 2000 students. The probability of any particular one enrolling is 25%. Model the number that enroll as the binomial random variable $X \sim BI(2000, .25)$.

(a) Find the expected number of students that enroll.
Answer: $\mu = (2000)(.25) = 500$

(b) Find the standard deviation.
Answer: $\sigma = \sqrt{(2000)(.25)(.75)} = 19.36$

(c) Find the probability that more than 510 enroll.
Answer: $P(X > 510) = 1 - \mathit{bicdf}(2000, .25, 510) = .293$

(d) Find the probability that less than 480 enroll.
Answer: $P(X < 480) = \mathit{bicdf}(2000, .25, 479) = .145$

(e) Find the probability that between 480 and 510 enroll.
Answer: $\mathit{bicdf}(2000, .25, 510) - \mathit{bicdf}(2000, .25, 479) = .562$

199. Suppose a large standardized test bank contains thousands of questions, and imagine that you know the answers to roughly five sixths of them. For a particular test, 50 questions are selected.

(a) Model the number of questions that you know as $X \sim BI(n, p)$. Find n and p.
Answer: $n = 50$ and $p = 5/6 = .833$

(b) Find the expected number of questions you will know the answer to.
Answer: $\mu = (50)(.833) = 41.65$

(c) Find the probability that you know exactly 90% of the answers.
Answer: $\mathit{bipdf}(50, .833, 45) = .074$

(d) Find the probability that you know at least 90% of the answers.
Answer: $1 - \mathit{bicdf}(50, .833, 44) = .137$

(e) Find the probability that you know the answer to between 40 and 45 of the questions.
Answer: $\mathit{bicdf}(50, .833, 45) - \mathit{bicdf}(50, .833, 39) = .733$

200. Suppose that on your one-way commute, you have a .5% chance of getting a speeding ticket. If you have to drive to school 180 days a year, then that is 360 trips. Let x be the number of speeding tickets you get in a year.

(a) If $x \sim BI(n, p)$, then what are n and p ?
Answer: $n = 360$, $p = .005$

(b) Find the expected value for the number of tickets you get in a year.
Answer: $np = 1.8$

(c) What is the probability that you get exactly 2 tickets?
Answer: $\mathit{bipdf}(360, .005, 2) = .269$

(d) What is the probability that you get more than 1 ticket?
Answer: $1 - \mathit{bicdf}(360, .005, 1) = .538$

(e) What is the probability that you get no tickets?
Answer: $\mathit{bipdf}(360, .005, 0) = .165$

(f) What is the probability that you go two consecutive years without getting a ticket?
Answer: $(.165)^2 = .027$

201. Suppose a basketball player makes 70% of his foul shots. If his life were on the line, does he have a better chance of making 5 out of 5, or would he have a better chance of making 9 out of 10 ?
Answer: $\mathit{bipdf}(5, .7, 5) = .168$ and $1 - \mathit{bicdf}(10, .7, 8) = .149$, so he has a better chance of making 5 of 5

202. Suppose you invite 30 people to a banquet, and each has an 80% chance of attending. Let X be the number that attend. Assume independence.

(a) Find the expected number of guests.
Answer: $\mu = (30)(.8) = 24$

(b) Find the probability that exactly that expected number of guests attend.
Answer: $\mathit{bipdf}(30, .8, .24) = .179$

(c) Find the probability of at least 27 guests.
Answer: $1 - \mathit{bicdf}(30, .8, 26) = .123$

(d) Suppose you seat the guests five to a table. What is the probability that you need exactly five tables ($21 \leq X \leq 25$)?
Answer: $P(21 \leq X \leq 25) = \mathit{bicdf}(30, .8, 25) - \mathit{bicdf}(30, .8, 20) = .684$

203. Suppose the probability that an elementary school student will contract head lice is 1%. Should you use the binomial distribution to calculate the probability that more than 5 out of 200 students will get lice? Explain.
Answer: no, the events are not independent

204. You attempt seven putts, each with probability p of success. Suppose the probability that you sink a majority of the 7 is 0.8. Use guess-check to find p.
Answer: you want $1 - \mathit{bicdf}(7, p, 3) = .8$, which is true for $p = 0.65$

205. Before each football game, the official tosses a fair coin to determine who gets the ball first. Suppose that the CN football team has correctly called 7 straight coin tosses. Which is true about the probability that they call the 8th toss correctly?

- CN is 'hot'; the probability is higher than 50%
- CN is 'due' to miss one; the probability is lower than 50%
- the coin is fair, and falls independently of what's happened before; the probability is 50%.
Answer: correct

Normal Distribution

206. A golfer's drive distance is modeled by $x \sim N(210, 20)$ yards.

(a) Find the mean.
Answer: $\mu = 210$

(b) Find the median.
Answer: 210 (the distribution is symmetric, so median = mean)

(c) Find the standard deviation.
Answer: $\sigma = 20$

(d) Find the variance.
Answer: $\sigma^2 = 400$

(e) According to the "empirical rule", about 95% of drives will travel between ____ and ____ yards.
Answer: within 2 st. dev of the mean, so between 170 and 250

(f) Out of 50 drives, estimate how many will go between 190 and 230 yards.
Answer: about 68% withing one standard deviation, and $(.68)(50) = 34$

207. Suppose that the annual rainfall in Seattle has distribution $N(52, 8)$.

(a) What is the probability of getting between 52 and 55 inches of rain?
Answer: $\mathit{normcdf}(52, 55, 52, 8) = .146$

(b) What is the probability of getting more than 60 inches?
Answer: $normcdf(60, E99, 52, 8) = .159$

(c) What is the probability of getting exactly 60 inches?
Answer: 0, since this is a continuous variable

(d) What is the probability of getting less than 48 inches?
Answer: $normcdf(-E99, 48, 52, 8) = .309$

(e) Fifty-seven inches of rain in a year corresponds to what percentile?
Answer: $normcdf(-E99, 57, 52, 8) = .734$, so the 73rd percentile

(f) If a certain year was at the 30th percentile, how much rain did they get?
Answer: $invnorm(.30, 52, 8) = 47.8$ inches

(g) In a given year, there is a 70% chance of getting at least ________ inches of rain.
Answer: $invnorm(.30, 52, 8) = 47.8$ inches

208. Suppose that motorist speeds on I-40 are $X \sim N(72, 6)$ MPH.

(a) What percentage of drivers are going between 70 and 75 MPH?
Answer: $normcdf(70, 75, 72, 6) = .322$

(b) What percentage are going between 70 and 80 MPH?
Answer: $normcdf(70, 80, 72, 6) = .539$

(c) What percentage are going under 65 MPH?
Answer: $normcdf(-E99, 65, 72, 6) = .121$

(d) What percentage are going over 85 MPH?
Answer: $normcdf(85, E99, 72, 6) = .015$

(e) What percentage are going exactly 72 MPH?
Answer: 0, since speed is continuous

(f) One third of drivers are going under _______ MPH.
Answer: $invnorm(1/3, 72, 6) = 69.4$

(g) Find the inter-quartile range of driver speeds.
Answer: $invnorm(.75, 72, 6) - invnorm(.25, 72, 6) = 76.047 - 67.953 = 8.09$ MPH

(h) Suppose you know that they won't give you a ticket if you are driving at the 90th percentile. What speed does that correspond to?
Answer: $invnorm(.9, 72, 6) = 79.7$

209. At a certain location, the high temperature on July 4 is modeled by $X \sim N(87, 6.5)$ degrees.

(a) Find the probability that the temperature is exactly 90°, i.e. $P(X = 90)$.
Answer: zero, since temperature is continuous

(b) Find the probability that the temperature is in the 80's, i.e. $P(80 \leq X < 90)$.
Answer: $normcdf(80, 90, 87, 6.5) = .537$

(c) Find the probability that the temperature is more than 95 or less than 85.
Answer: $1 - normcdf(85, 95, 87, 6.5) = .488$

(d) Find the probability that the z-score is less than -1.5.
Answer: first solve $-1.5 = \frac{x-87}{6.5}$ to get $x = 77.25$, then $normcdf(-E99, 77.25, 87, 6.5) = .0668$

(e) According to the model (assume no climate change), approximately how many times since 1776 has the July 4th temperature been at least 100° ?
Answer: $normcdf(100, E99, 87, 6.5) * 241 = 5.48$

210. A Christmas tree farmer planted a field with 3600 trees. After a few years, the heights are distributed $x \sim N(7, 0.8)$ feet. Approximate how many trees in the field are between 7 and 8 feet tall.
Answer: the percentage of trees in that range is $normcdf(7, 8, 7, .8) = .394$, and $(.394)(3600) \approx 1420$

211. A truck driver finds that the time required to complete his route is normally distributed with $\mu = 1000$ and $\sigma = 35$ minutes. Find the probability that the route will take more than 17 hours.
Answer: $x \sim N(1000, 35)$, and 17 hours is 1020 minutes, so $P(x > 1020) = normcdf(1020, E99, 1000, 35) = .284$

212. Suppose that American men have heights $N(70, 3)$ inches, and Whoville men have heights $N(70, 23)$ inches. Which is more likely?

- An American man over 7 feet tall.
- A Whoville man over 12 feet tall.

Answer: $normcdf(84, E99, 70, 3) = 1.53E - 6$, and $normcdf(144, E99, 70, 23) = 6.47E - 4$, so the latter is more likely

213. Suppose the Alabama football team is favored by 20 points, and the standard deviation in a game's outcome is 14 points. Model the point spread outcome as $x \sim N(20, 14)$.

(a) Find the probability that Alabama wins the game, $P(x > 0)$.
Answer: $normcdf(0, E99, 20, 14) = .923$

(b) Suppose the game could be played 10 times. Find the probability that Alabama loses at least once.
Answer: they do not win them all, so $1 - (.923)^{10} = .55$

214. The coach of an underdog basketball team is considering two game plans that would give normally distributed outcomes:

- aggressive strategy: $x \sim N(-8, 20)$
- conservative strategy: $x \sim N(-6, 12)$

Which strategy should he choose to maximize his chance of winning the game?
Answer: chance of winning with agg. strat. is $normcdf(0, E99, -8, 20) = .345$
chance of winning with cons. strat. is $normcdf(0, E99, -6, 12) = .309$
so he should choose the aggressive strategy

215. Suppose $X \sim N(100, 15)$ and $Y \sim N(105, 12)$.

(a) Which is higher: $P(x > 120)$ or $P(y > 120)$?
Answer: $.091 < .106$

(b) Which is higher: $P(x > 140)$ or $P(y > 140)$?
Answer: $.00383 > .00177$

(c) If observations of x and y are made independently, find $P(x > 105 \text{ and } y < 100)$.
Answer: $normcdf(105, E99, 100, 15) * normcdf(-E99, 100, 105, 12) = .125$

216. If $x \sim N(2, 7)$,

(a) Find $P(x = 3)$.
Answer: zero, since x is continuous

(b) Find $P(x < 0 \text{ or } x > 10)$
Answer: $normcdf(-E99, 0, 2, 7) + normcdf(10, E99, 2, 7) = .514$

(c) Find $P(x < 0 \text{ and } x > 10)$
Answer: impossible, so zero

(d) Find $P(x < 0 \text{ or } x < 10)$
Answer: $\mathit{normcdf}(-E99, 10, 2, 7) = .873$

(e) Find $P(x > 0 \text{ or } x < 10)$
Answer: certain so 100%

(f) Find $P(x > 0 \text{ and } x < 10)$
Answer: $\mathit{normcdf}(0, 10, 2, 7) = .486$

217. If x and y are independently drawn from the distribution $N(100, 15)$, find

(a) $P(x > 115 \text{ and } y > 115)$
Answer: $\mathit{normcdf}(115, E99, 100, 15)^2 = .159^2 = .0252$

(b) $P(x > 115 \text{ or } y > 115)$
Answer: neither is above 115 gives $1 - \mathit{normcdf}(-E99, 115, 100, 15)^2 = 1 - .841^2 = .292$

218. Suppose your score on an IQ test is $X \sim N(108, 12)$. Assuming independence, find the probability that you score at least 110 on three consecutive tests.
Answer: $\mathit{normcdf}(110, E99, 108, 12)^3 = (.434)^3 = .0816$

219. Suppose x and y are selected independently from $N(100, 15)$. Find the probability that the maximum (larger value) is less than 110, i.e. $P(x < 110 \text{ and } y < 110)$.
Answer: since they are independent, multiply: $\mathit{normcdf}(0, 110, 100, 15)^2 = .559$

220. For any normal distribution $X \sim N(\mu, \sigma)$, find the z-score of the 33rd percentile.
Answer: $\mathit{invnorm}(.33, 0, 1) = -.44$

221. Suppose $x \sim N(50, \sigma)$, and the inter-quartile range is 12. Find σ (you can use trial-and-error).
Answer: Trial-and-error approach - note that $Q_3 = 56$, so try to get $\mathit{invnorm}(.75, 50, \sigma) = 56$, so $\sigma = 8.9$.
Direct approach: $\sigma \cdot \mathit{invnorm}(0.75) = 6$, so $\sigma = 8.9$

222. Suppose $X \sim N(\mu, \sigma)$, and

- $x = 410$ corresponds to a z-score of 2
- $x = 298$ corresponds to a z-score of -1.5

Find μ and σ.
Answer: solve $\frac{410-\mu}{\sigma} = 2$ and $\frac{298-\mu}{\sigma}$ simultaneously to get $\mu = 346$ and $\sigma = 32$.

223. The population of Canada is about 35 million, and the population of the U.S. is about 315 million. If IQ is distributed $N(100, 15)$ for both countries,

(a) Approximately how many Canadians have IQ above 175.
Answer: $\mathit{normcdf}(175, E99, 100, 15) \times 35,000,000 \approx 10$

(b) Approximately how many Americans have IQ above 175.
Answer: $\mathit{normcdf}(175, E99, 100, 15) \times 35,000,000 \approx 90$

(c) Canada's top million citizens have IQ's of at least ________.
Answer: $\mathit{invnorm}(34/35, 100, 15) = 128.5$

(d) America's top million citizens have IQ's of at least ________.
Answer: $\mathit{invnorm}(314/315, 100, 15) = 140.9$

224. Suppose you have a bell-shaped empirical distribution with $\mu = 1000$ and $\sigma = 200$. From your sample size of $n = 5000$, you see that 90 of them have values over 1500.

(a) If the distribution were normal, how many from your sample would you expect to be over 1500?
Answer: $5000 * normcdf(1500, E99, 1000, 200) = 31$

(b) Do you think the normal distribution is a good model for the right tail of this distribution?
Answer: no, the normal model predicted 31 "outliers", but almost triple that were observed in this "fat tail"

Central Limit Theorem

225. If X is a random variable with $\mu = 50$ and $\sigma = 12$, then find the standard error $\sigma_{\overline{x}}$ if

(a) $n = 4$
Answer: $\mu_{\overline{x}} = 50$ and $\sigma_{\overline{x}} = 6$

(b) $n = 25$
Answer: $\mu_{\overline{x}} = 50$ and $\sigma_{\overline{x}} = 2.4$

(c) $n = 100$
Answer: $\mu_{\overline{x}} = 50$ and $\sigma_{\overline{x}} = 1.2$

(d) $n = 200$
Answer: $\mu_{\overline{x}} = 50$ and $\sigma_{\overline{x}} = .85$

226. Suppose that systolic blood pressure measurements have a standard deviation of $\sigma = 7$.

(a) What is the standard error for average of 2 readings?
Answer: $7/\sqrt{2} = 4.95$

(b) What is the standard error for the average of 5 readings?
Answer: $7/\sqrt{5} = 3.13$

(c) How many measurements should be averaged so that the standard error is less than 2 ?
Answer: solve $7/\sqrt{n} < 2$ to get $n > 12.25$, so you'd need at least 13 readings

227. The Associated Press polls 58 sports writers to rank the top college football teams. Suppose the standard deviation of Virginia Tech's ranking on the individual ballots is $\sigma = 4$.

(a) If you averaged 5 ballots at a time, find the standard error $\sigma_{\overline{x}}$.
Answer: $4/\sqrt{5} = 1.79$

(b) If you averaged 10 ballots at a time, find the standard error $\sigma_{\overline{x}}$.
Answer: $4/\sqrt{10} = 1.26$

(c) If you average all 58 ballots, find the standard error $\sigma_{\overline{x}}$.
Answer: $4/\sqrt{58} = .53$

228. If you quadruple the sample size, then the standard error will decrease by ________ percent.
Answer: $\frac{\sigma}{\sqrt{4n}} = .5\frac{\sigma}{\sqrt{n}}$, so it is 50% lower

229. If $\sigma = 17$, how big must a sample size be so that $\sigma_{\overline{x}} < 3$?
Answer: solve $17/\sqrt{n} < 3$ to get $n > 32.1$ but round up to $n = 33$

230. Suppose you are trying to discover the mean commuting time for Knoxville workers. If $\sigma = 8$ minutes, how many people must you survey so that the standard error for $\overline{x}$ is less than 0.5 minutes ?
Answer: We need $8/\sqrt{n} < .5$, so $16 < \sqrt{n}$, so $n > 256$.

231. Sample four integers between 10 and 20 using the web page:
`http://www.random.org/integers/?num=4&min=10&max=20&col=5&base=10&format=html&rnd=new`

(a) Do it 5 times, and each time record the sample mean $\overline{x}$.
Answer: e.g. if it shows 16,20,13,12, then $\overline{x} = \frac{17+20+12+14}{4} = 15.75$

(b) Find the empirical mean and standard deviation of the sampling distribution.
Answer: put your $\overline{x}$ values into L1 and find the mean and stdev

(c) The theoretical mean is 15. Would you be more surprised to get $\overline{x} = 17$ if $n = 4$, or if $n = 40$? Explain.
Answer: With a small sample, it is conceivable that you might be that far away from the true mean. But with a large sample, it is virtually impossible.

232. Suppose that individual motorist speeds are $X \sim N(72, 6)$, and are independent of each other.

(a) Find the probability that a randomly selected vehicle is going at least 75 MPH.
Answer: $normcdf(75, E99, 72, 6) = .309$

(b) Find the probability that the average speed of a random sample of 4 vehicles is at least 75 MPH.
Answer: $normcdf(75, E99, 72, 6/\sqrt{4}) = .159$

(c) Find the probability that the average speed of a random sample of 9 vehicles is at least 75 MPH.
Answer: $normcdf(75, E99, 72, 6/\sqrt{9}) = .067$

(d) Find the probability that the average speed of a random sample of 25 vehicles is at least 75 MPH.
Answer: $normcdf(75, E99, 72, 6/\sqrt{25}) = .0062$

233. A Jiffy Lube employee does oil changes that take an average of 14 minutes, with a standard deviation of 5 minutes.

(a) His boss wants him to complete 36 oil changes in an 8-hour shift. How many minutes per oil change can he allocate ?
Answer: Eight hours is 480 minutes, so he needs to average $480/36 = 13.33$ minutes per oil change.

(b) According to the CLT, what is the approximate distribution of $\overline{x}$, the average time required per oil change.
Answer: $\overline{x} \sim N(14, 5/\sqrt{36}) = N(14, .833)$

(c) Find the proability that he can complete 36 oil changes in the shift.
Answer: $P(\overline{x} \leq 13.33) = normcdf(-E99, 13.33, 14, 5/\sqrt{36}) = .21$

234. An electric generator is capable of producing 32000 watts of power for an office building with 64 rooms. Peak operating usage by the offices is independent with a mean of 460 watts per room, with a standard deviation of 160 watts.

(a) How much power is available per room?
Answer: $32000/64 = 500$ watts

(b) According to the CLT, what is the approximate distribution of $\overline{x}$, the average power consumption per room.
Answer: $\overline{x} \sim N(460, 160/\sqrt{64}) = N(460, 20)$

(c) Find the $P(\overline{x} \geq 500)$, the probability that the generated power is insufficient for the demand.
Answer: $P(\overline{x} \geq 500) = normcdf(500, E99, 460, 20) = .023$

235. A truck can haul at most 2000 pounds. You need to haul 50 boxes, with weights from a distribution with $\mu = 38$ and $\sigma = 25$. Find the probability that the total weight of the boxes is under the limit.
Hint: $P(\sum x < 2000) = P(\overline{x} < 40)$
Answer: $normcdf(-E99, 40, 38, 25/\sqrt{50}) = .714$

236. A ship bound from Lisbon to the New World has 70 people on board. Assuming the trip will take 3 weeks, that makes $n = (21)(70) = 1470$ days of rations that must be stored. If the amount of water consumed by a person on a day has $\mu = 40$ and $\sigma = 13$ ounces,

(a) Find $\sigma_{\overline{x}}$.
Answer: $\frac{13}{\sqrt{1470}} = .339$

(b) Find the amount of drinking water (in gallons) that should be loaded on the ship to be 99.99% sure of not running out.
Answer: $invnorm(.9999, 40, .339) = 41.26$ oz per person per day, so $(41.26)(1470) = 60653$ oz, or 474 gallons.

237. Suppose a small company's monthly office expenses have a distribution with $\mu = 700$ and $\sigma = 300$ dollars.

(a) Use the CLT to find the distribution of the average monthly expense for a year ($n = 12$).
Answer: $\overline{x} \sim N(700, 86.60)$

(b) Find the probability that the company spends more than \$ 10,000 in a year.
Answer: $10000/12 = 833.33$ per month, so $normcdf(833.33, E99, 700, 86.60) = .062$

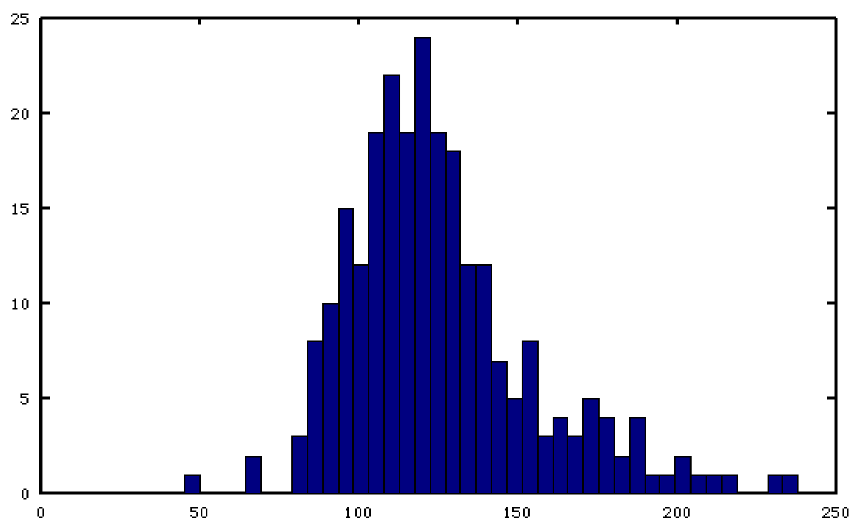
238. Suppose that the annual rainfall in a certain location is $X \sim N(36, 8)$. Assuming year-to-year independence, find the probability of receiving less than 330 inches over a decade.
Answer: $normcdf(-E99, 330/10, 36, 8/\sqrt{10}) = .118$

239. A mutual fund manager calculates that individual company annual stock market returns have a standard deviation of $\sigma = 10$ percent relative to the market.

(a) If she splits the fund's assets among 16 stocks, what is the standard deviation of the return?
Answer: $\sigma_{\overline{x}} = 10/\sqrt{16} = 2.5$

(b) How many stocks should her fund hold to reduce the standard deviation to $\sigma_{\overline{x}} = 1.5$?
Answer: solve $\frac{10}{\sqrt{n}} = 1.5$ to get $n = 44.4$

240. Let the random variable X be a movie's runtime (in minutes). Here is a histogram for a sample of 250 movies retrieved from IMDB.



Assume the population mean is $\mu = 125$ minutes, and the standard deviation is $\sigma = 30$ minutes.

(a) Does the distribution appear to be normal, or skewed in one direction?
Answer: right-skewed

(b) Nevertheless, for a sample of size n, the sampling mean of $\overline{x}$ has approximately what distribution?
Answer: by the CLT, $\overline{x} \sim N(125, 30/\sqrt{n})$.

(c) Suppose you run a movie marathon by picking a random selection of $n = 12$ movies. Find the probability that they can all play in a 24 hour period.
Answer: You have $24 \cdot 60 = 1440$ minutes, so 120 minutes per movie. $N(125, 30/\sqrt{12}) = N(125, 8.66)$, so $normcdf(-E99, 120, 125, 8.66) = .282$

241. A hospital is implementing quality control measures to be sure emergency room patients are treated quickly. Suppose the wait time has mean $\mu = 12$ and $\sigma = 8$ minutes.

(a) Assume the wait time distribution is normal. If a patient comes to the emergency room, what is the probability that her waiting time is over 15 minutes?
Answer: $normcdf(15, E99, 12, 8) = .354$

(b) Do you think the distribution of waiting times is really normal? Explain.
Answer: no, it is right skewed

(c) By the CLT, if the observations are independent, the sampling distribution is approximately normal regardless of the distribution of the underlying variable. If 100 patients come to the emergency room, what is the probability that the mean waiting time is over 15 minutes?
Answer: $normcdf(15, E99, 12, 8/\sqrt{100}) = .0000884$

(d) Do you think the wait times are really independent?
Answer: probably not, since they may get behind and everyone has to wait

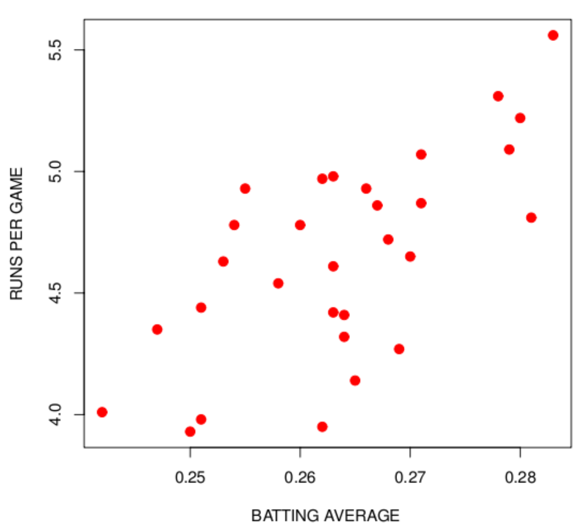
242. A bowler's game scores have a standard deviation of $\sigma = 20$. If he bowls $n = 30$ games, find the probability that $|\overline{x} - \mu| < 5$, i.e. his observed average is within 5 pins of his true ability.
Answer: by the CLT, $\overline{x} \sim N(\mu, 20/\sqrt{30})$, and this will be for any value of μ, so with $\mu = 100$, $normcdf(95, 105, 100, 20/\sqrt{30}) = .829$

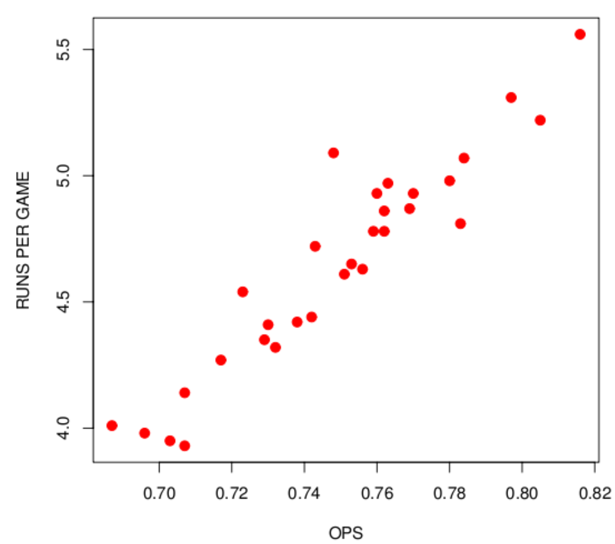
Linear Regression and Correlation

243. Two often quoted baseball statistics are

- batting average (BA)
- on-base plus slugging (OPS)

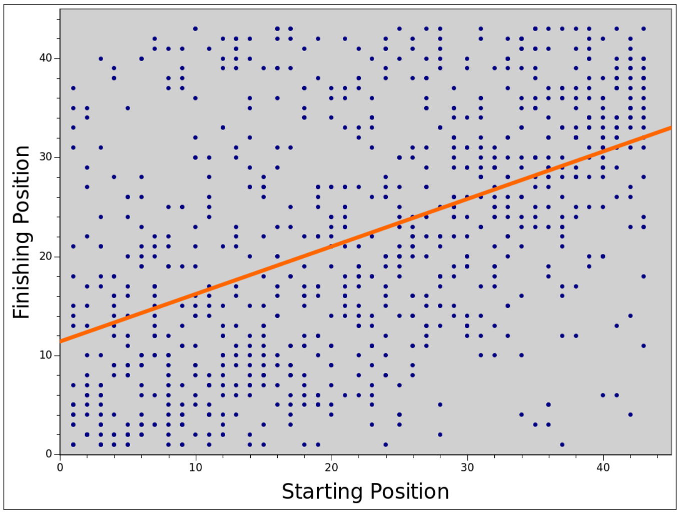
According to these scatterplots:





which statistic has a higher correlation with runs per game for Major League baseball teams? Explain your answer with a complete sentence.
Answer: OPS has a higher correlation to runs since the data points fit a line more tightly.

244. Here is a scatterplot of the finishing versus starting position of drivers in the first twenty races of the 2015 NASCAR season.



(a) Estimate the y-intercept.
Answer: about 11

(b) Estimate the slope. (pick two points and calculate rise over run).
Answer: using points $(0, 11)$ and $(40, 30)$, $\frac{30-11}{40-0} = \frac{19}{40} \approx .475$

(c) Write the equation of the linear regression line $y = a + bx$.
Answer: $y = 11 + .475x$

(d) Predict the finishing position of the driver that starts the race in the 5th position.
Answer: $\hat{y} = 11 + .475(5) = 13.4$

245. Consider this data for a sample of 2011 new car curb weights and highway fuel economy.

make	weight (x)	MPG (y)
BMW 3 Series	3362	28
Cadillac Escalade	5488	18
Chevy Tahoe	5636	21
Ford Explorer	4557	25
Ford Mustang	3401	26
Honda Accord	3217	33
Honda Odyssey	4337	27
Hyundai Elantra	2660	40
Kia Optima	3206	35
Mazda Miata	2480	28
Nissan Versa	2350	36
Subaru Forester	3250	27
Toyota Prius	3042	48
Toyota RAV4	3360	28

(a) Find the correlation.
Answer: linregttest $r = -.70$

(b) Find the linear regression line.
Answer: $y = 49.08 - .005305x$

(c) Interpolate to estimate the MPG of a car that weighs 4000 pounds.
Answer: $y = 49.08 - .005305(4000) = 27.86$ MPG

(d) Extrapolate to estimate the MPG of a car that weighs 2000 pounds.
Answer: $y = 49.08 - .005305(2000) = 38.47$ MPG

246. You are studying the relationship between cricket chirps per minute and the farenheit temperature. Your experiments yield the data:

chirps (x)	temperature (y)
108	67
120	69
130	74
157	75
175	82
196	89

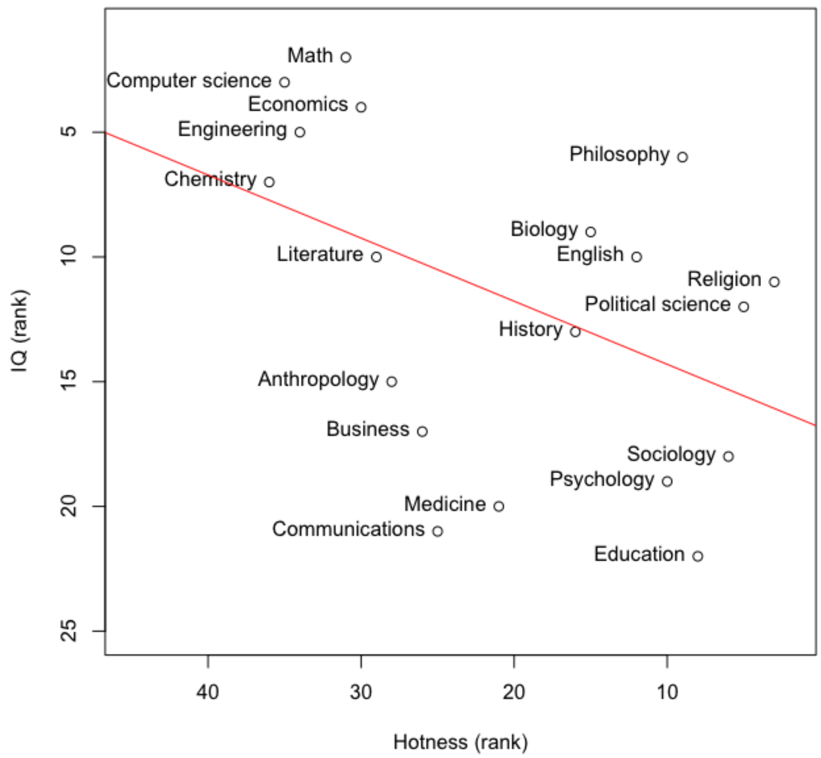
(a) Find the correlation.
Answer: linregttest gives $r = .973$

(b) Find the linear regression line.
Answer: $y = 41.27 + .235x$

(c) If you hear 35 chirps in 15 seconds, estimate the temperature.
Answer: $x = 4(35) = 140$ chirps per minute, so $y = 41.27 + .235(140) = 74.17$ degrees.

(d) If the temperature is 80 degrees, estimate the number of chirps you would hear in a minute.
Answer: solve $80 = 41.27 + .235x$ to get $x = 164.8$

247. This blog post examines the relationship between professor "hotness" and IQ by discipline. Here is the scatterplot:



Given that $r^2 = 0.20$, find the correlation r.
Answer: this is negative correlation so $r = -\sqrt{.20} = -.447$

248. Consider male personal ads on an online dating site. This table lists the man's stated income and the number of responses to his ad.

income ($ 1000)	responses
20	1
30	3
50	5
50	7
60	4
75	8
120	11

(a) Find the correlation between income and responses.
Answer: .925

(b) Find the linear regression line.
Answer: $y = .104 + .094x$

(c) Use the regression line to predict the number of responses for a man that claims to make 100 thousand dollars.
Answer: 9.5

249. Suppose the linear regression line $y = 7 + .43(x - 64)$ predicts a woman's shoe size given her height in inches.

(a) Rewrite the expression in the form $y = a + bx$.
Answer: $y = -20.52 + .43x$

(b) What is the slope?
Answer: .43

(c) If a woman is 5'7" tall, predict her shoe size.
Answer: $x = 67$, and $\hat{y} = 8.29$

250. Suppose a linear regression line is $y = \frac{1}{2}(2 - 3x) + 5$ If $r^2 = .38$, then find the correlation.
Answer: the slope is $-3/2$, so correlation is negative, $r = -\sqrt{.38} = -.616$

251. Here are the average daily temperatures, along with my corresponding electric bills during a six month span.

	degrees	dollars
Nov	51	81
Dec	43	102
Jan	38	123
Feb	47	95
Mar	54	79
Apr	69	65

(a) Sketch a graph and plot the points.

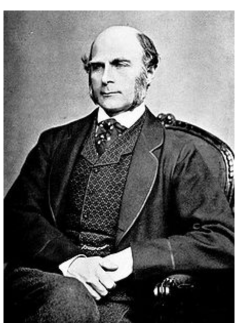
(b) Find the linear regression line.
Answer: $y = 179.86 - 1.77x$

(c) Predict my electric bill if the average temperature in a month were 61 degrees.
Answer: $\hat{y} = 179.86 - 1.77(61) = 71.89$

(d) Predict my electric bill if the average temperature in a month were 91 degrees.
Answer: $\hat{y} = 179.86 - 1.77(91) = 18.79$

(e) Why is your last answer unreasonable?
Answer: summer requires air conditioning, can't extrapolate into summer temps

252. One of the pioneers of linear regression was Francis Galton, who in the 1880's studied the relationship between parent and child heights.



Let x be the average height of the parents. So if the mother was 62 inches tall, and the father was 70 inches tall, then $x = \frac{62+70}{2} = 66$. Let y be the child's height when full grown. Suppose the regression equations are:

- male child: $y = 71 + .57(x - 68)$
- female child: $y = 66 + .57(x - 68)$

(a) State the slope.
Answer: the coefficient of the x is the slope .57

(b) Suppose a father is 6′5″ and the mother is 5′11″. Predict the height of a male child. Also predict the height of a female child.
Answer: $x = \frac{77+71}{2} = 74$
male: $\hat{y} = 71 + .57(74 - 68) = 74.42$ female: $\hat{y} = 66 + .57(74 - 68) = 69.42$
Notice that the children tend to be taller than average, but not as tall as their parents.

253. Suppose the owner of Pal's drive-in must decide what to charge for a Big Pal hamburger. Do you think the correlation between price and sales is positive or negative? Explain.
Answer: negative, since the higher the price, the fewer he will sell

254. Do you think the correlation between ACT score and college retention rate is positive or negative? Explain.
Answer: positive, since weaker students tend to drop out more

255. For each pair of variables, do you expect the correlation to be positive or negative? Explain your answer. Do you think that there is a cause-effect relationship?

(a) literacy & life expectancy
Answer: positive

(b) alcohol consumption & balance
Answer: negative

(c) exercise & weight
Answer: negative

(d) square feet & house price
Answer: positive

(e) latitude & temperature
Answer: negative

(f) intelligence & salary
Answer: positive

(g) beauty & salary
Answer: positive

(h) wealth & health
Answer: positive

256. Suppose that the linear regression line $y = -2 + 1.2x$ models the age of the groom y given the age of the bride x.

(a) Is the correlation between x and y positive or negative?
Answer: positive

(b) If a bride is 22 years old, what is the expected age of her groom?
Answer: $y = -2 + 1.2(22) = 24.4$

(c) If a bride is 35 years old, what is the expected age of her groom?
Answer: $y = -2 + 1.2(35) = 40$

(d) How old would a bride be if her groom was expected to be 30 years old?
Answer: solve $30 = -2 + 1.2x$ to get $x = 26.7$

257. According to this article,

THE IRISH TIMES Wed, May 18, 2016

NEWS SPORT BUSINESS OPINION LIFE & STYLE CULTURE

Social Affairs › Religion & Beliefs

Tree-lined streets ‘can raise children’s intelligence’

Greenery can increase house prices by 20% and make you feel younger, seminar told

Mon, May 16, 2016, 21:20 | Updated: Mon, May 16, 2016, 21:53

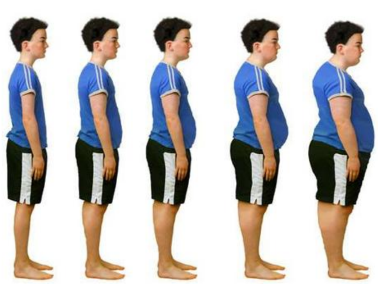
Kitty Holland

Mount Merrion Avenue in Dublin: a seminar, supported by Dublin City Council, the school of geography at UCD and others, heard there was a “gross inequality of beauty” in the capital’s streetscapes. Photograph: Frank Miller

“We all have a right to public beauty, and disadvantaged communities have an equal right to beauty in their neighbourhoods and on their streets,” said Mr Kearns. “Research has shown that living on a street full of trees can raise the value of a home by 20 per cent, is the equivalent of a 9,000 salary raise, can make you feel seven years younger and can raise the IQ of a child,” he said.

Suppose a family moves to a tree-lined street. Do you think the parents will consequentially get a raise, and their kids will do better on tests? Explain how the well-intentioned Mr Kearns has confused correlation with causation.

258. Save the file obesity.xls, and open it with Excel. It contains the obesity rate and median household income for each state. Is there a relationship between the two variables?



Follow these steps to produce a scatterplot with linear regression line:

- Highlight the numerical data (do not highlight the state names).

- From the Excel toolbar, “Insert” a “scatter” plot.
- Select “Chart Layout” and give your graph a title.
- Label the horizontal axis “income (thousands)” and the vertical axis “obesity rate”.
- From the toolbar, select “trendline” and choose a linear trendline.
- From the trendline options, have it display the equation and show the R^2 value.

Once your graph looks good, copy/paste it into Word, and then answer these questions:

(a) Write your linear regression line in the form $y = a + bx$.
Answer: $y = 40.4428 - 0.279758x$

(b) What is the slope of the line?
Answer: -0.279758

(c) Are obesity and income positively or negatively correlated?
Answer: negatively

(d) State the sample correlation r.
Answer: $-\sqrt{.51205} = -.716$

(e) Comment on whether there is a causal relationship between obesity and income. Does gaining weight cause you to make less money? Does making less money cause you to gain weight? How might you explain the observed correlation?
Answer: Probably a general lifestyle difference. Wealthy people more educated, don't eat as much fast food, exercise more, etc.

(f) If you were to extrapolate with the linear regression line, at what income would you predict the obesity rate to be zero? Is that a reasonable prediction?
Answer: Solve $40.4428 - .279758x = 0$ to get $x \approx 144.56$ thousand per year. Not reasonable, since will always be some wealthy people that are overweight.

259. Stock market traders like to understand how a particular company's stock trades relative to the broad market. Let's analyze 1000 trading days of Apple's stock compared to the S&P 500 index.

- Let x be the percentage change in the S&P.
- Let y be the percentage change in Apple.

We will find coefficients α and β for the model:

$$y = \alpha + \beta x$$

In Excel, open “apple.xlsx” from the class data folder. Create a scatter plot for the % change columns, and then add a trendline and have it display the equation and R^2.

(a) Write the linear regression equation.
Answer: $y = .174875 + .844796x$

(b) Find “alpha”, the y-intercept, which represents the over or under performance of Apple compared to the S&P.
Answer: $\alpha = .174875$

(c) Find “beta”, the slope, which represents the sensitivity of Apple to the S&P.
Answer: $\beta = .844796$

(d) Find r, the correlation between Apple and the S&P.
Answer: $r = .6807$

Statistical Models

260. Suppose the length of a MWF English class is a random variable $X \sim N(48, 2.25)$ minutes.

   (a) Find the 15th percentile.
   Answer: $invnorm(.15, 48, 2.25) = 45.67$

   (b) Find the probability that class lets out late ($X > 50$).
   Answer: $normcdf(50, E99, 48, 2.25) = .187$

   (c) Find the probability that class lets out late on at least 3 of the next 9 class periods.
   Answer: $1 - bicdf(9, .187, 2) = .228$

261. If individual blood glucose levels are $x \sim N(110, 30)$, then

   (a) What percentage of people have dangerously high levels (above 150)?
   Answer: $normcdf(150, E99, 110, 30) = .091$

   (b) Fill in the blank: one third of people have blood sugar levels between 100 and ________.
   Answer: draw a picture! $normcdf(-E99, 100, 110, 30) = .37$, so we need $invnorm(.37 + .333, 110, 30) = 126$

   (c) Find Q_1 and Q_3 for this distribution (the values that mark the middle 50%)
   Answer: $Q_1 = invnorm(.25, 110, 30) = 89.77$ and $Q_3 = invnorm(.75, 110, 30) = 130.23$

   (d) In a group of 75 people that ate a certain diet, only 3 of them had high (above 150) levels. Use the binomial distribution to find the probability that 3 or fewer would occur by luck.
   Answer: for each individual, $p = .091$, so we get $bicdf(75, .091, 3) = .081$

262. Suppose annual rainfall is $x \sim N(36, 8)$ inches. A year is considered "dry" if the rainfall is under 30. What is the probability of having more than 3 dry years in a decade?
Answer: For any single year, we get $p = normcdf(-E99, 30, 36, 8) = .227$. For getting more than 3 out of 10, we get $1 - bicdf(10, .227, 3) = .173$

263. Uncle Rico has a powerful arm. The distance he can throw a football is $x \sim N(45, 4)$ yards.

   (a) What is the probability that he throws it between 40 and 50 yards?
   Answer: $normcdf(40, 50, 45, 4) = .789$

   (b) What is the probability that he throws it at least 50 yards?
   Answer: $normcdf(50, E99, 45, 4) = .106$

   (c) What is the probability that he makes three consecutive throws of at least 50 yards?
   Answer: $(.106)^3 = .0012$

(d) What is the probability that he has at least three 50+ yard throws in twenty trys?
Answer: $1 - \mathit{bicdf}(20, .106, 2) = .357$

(e) What distance would a throw go if it is at the 95th percentile?
Answer: $invnorm(.95, 45, 4) = 51.6$ yards

(f) Fill in the blank: 50 percent of his throws are between 40 and _______ yards.
Answer: draw a picture! note the tail to the left of 40 has area $\mathit{normcdf}(-E99, 40, 45, 4) = .106$, so the right endpoint is $invnorm(.606, 45, 4) = 46.08$

264. A medical device uses a battery with lifespan distributed $x \sim N(300, 24)$ days. Out of 5000 such batteries,

(a) How many of them would be expected to last less than 250 days?
Answer: $\mathit{normcdf}(-E99, 250, 300, 24) \cdot 5000 = (.0186)(5000) = 93$

(b) Find the probability that at least 100 of the 5000 batteries will last less than 250 days.
Answer: $1 - \mathit{bicdf}(5000, .0186, 99) = .245$

265. Suppose a basketball team's point total in a game is approximated by $X \sim N(75, 11)$. Using a continuity correction,

(a) find the probability that they score at least 90 points.
Answer: $\mathit{normcdf}(89.5, E99, 75, 11) = .094$

(b) find their 80th percentile score
Answer: $invnorm(.8, 75, 11) = 84$

(c) find the probability that they score exactly 75 points.
Answer: $\mathit{normcdf}(74.5, 75.5, 75, 11) = .036$

266. The number in attendance at church Sunday morning is modeled by $x \sim N(460, 50)$, with continuity correction.

(a) Is the actual attendance discrete or continuous?
Answer: discrete, although the model is continuous, so we will need to round

(b) Find the probability of having at least 500 attendance.
Answer: $\mathit{normcdf}(499.5, E99, 460, 50) = .215$

(c) Find the probability of having between 450 and 490 in attendance.
Answer: $\mathit{normcdf}(449.5, 490.5, 460, 50) = .312$

(d) If only 400 attend, that corresponds to the _______ percentile.
Answer: $\mathit{normcdf}(-E99, 400.5, 460, 50) = .117$ so the 12th percentile

(e) Find the quartiles Q_1 and Q_3.
Answer: $Q_1 = invnorm(.25, 460, 50) = 426$ and $Q_3 = invnorm(.75, 460, 50) = 494$

267. Suppose $X \sim BI(100, 0.8)$.

(a) Find μ and σ.
Answer: $\mu = (100)(.8) = 80$ and $\sigma = \sqrt{(100)(.8)(.2)} = 4$

(b) Describe the shape of the distribution of X.
Answer: left skewed

(c) If you observed a sample of 64 values of X, then the sample mean $\overline{x}$ has mean $\mu_{\overline{x}} =$ _______ and standard deviation $\sigma_{\overline{x}} =$ _______.
Answer: $\mu_{\overline{x}} = 80$, $\sigma_{\overline{x}} = \sigma/\sqrt{64} = 4/8 = .5$

(d) Describe the shape of the distribution of $\overline{x}$.
Answer: nearly bell-shaped

268. Suppose that a civic arena seats 3000 people. Historically about 87% of ticketholders show up to a particular event (e.g. a concert or a game). If you sold 3400 general admission tickets,

(a) How many ticketholders do you expect to attend the event? What is the standard deviation?
Answer: $\mu = np = (3400)(.87) = 2958$ and $\sigma = \sqrt{np(1-p)} = 19.61$

(b) Use bicdf to get the exact probability that there will be enough seats.
Answer: $bicdf(3400, .87, 3000) = .986$

(c) Use normcdf to approximate the probability that there will be enough seats. (do not worry about continuity correction)
Answer: $normcdf(-E99, 3000, 2958, 19.61) = .984$

(d) Fill in the blank: 75% of the time attendance will be fewer than ________
Answer: $invnorm(.75, 2958, 19.61) = 2971$

269. Suppose a baseball player is truely a $p = .300$ hitter. Suppose he has 500 at-bats in a season. Let $x \sim BI(500, .3)$ be the number of hits he gets.

(a) Find μ.
Answer: $\mu = (500)(.3) = 150$

(b) Find σ.
Answer: $\sigma = \sqrt{(500)(.3)(.7)} = 10.25$

(c) Use a normal approximation to find the probability that he bats at least .320 (160 hits). Do not worry about continuity correction.
Answer: $normcdf(160, E99, 150, 10.25) = .165$

270. Suppose $x \sim BI(75, p)$ and $\mu = 21$. Find p.
Answer: solve $np = \mu$ to get $p = .28$

271. Suppose $x \sim BI(n, .8)$ and $\sigma = 10$. Find n.
Answer: solve $\sqrt{np(1-p)} = \sigma$ to get $n = 625$

272. Suppose $x \sim BI(n, p)$ and $\mu = 32$ and $\sigma = 5.185$. Find n and p.
Answer: you know $np = 32$, so $\sqrt{32(1-p)} = 5.185$. Solve to get $p = .16$. Then solve $n(.16) = 32$ to get $n = 200$.

273. If $X \sim N(90, 25)$ and $Y \sim BI(120, .75)$ are independent, find $P(X < 80 \text{ and } Y < 80)$.
Answer: $normcdf(-E99, 80, 90, 25) \times bicdf(120, .75, 79) = .00534$

274. According to the empirical rule, about 99.7% of observations from a normal distribution should be with 3 standard deviations of the mean, that is $|z| < 3$. Suppose you take $n = 7000$ measurements from a distribution that looks bell-shaped.

(a) How many would you "expect" to be in the tails $|z| \geq 3$?
Answer: $(.003)(7000) = 21$

(b) Suppose you actually observed 85 such measurements. Do you think the distribution is normal?
Answer: no, it appears to have fat tails since there are way more outliers than predicted by the normal dist.

275. Suppose the distribution of X is normal with mean 100, and the z-score of $x = 172$ is three. In a sample of 7, what is the probability that a majority are over 95?
Answer: Since $\frac{172-100}{\sigma} = 100$, the standard deviation is 24. The probability of one observation being over 95 is $normcdf(95, E99, 100, 24) = .5825$, so the probability of a majority (at least 4) is $1 - bicdf(7, .5825, 3) = .676$

276. Suppose a case of chips contains 48 bags. If bag weights have a standard deviation of 1.5 oz, find the probability that a case weighs one pound more than usual. (assume independence and use the CLT)
Answer: by the CLT, the average bag's weight is $\overline{x} \sim N(\mu, 1.5/\sqrt{48})$ oz. For the case to be one pound over-weight, the average bag would have to be $16/48 = .333$ oz over-weight. So then $normcdf(.333, E99, 0, 1.5/\sqrt{48}) = .0618$

277. In a group of 15 mothers, 6 of them claimed that their child was in the top quintile of intelligence. What is the probability that in a random sample of 15 children, at least 6 of them would be above the 80th percentile?
Answer: the model is $X \sim BI(15, .2)$, so $P(X \geq 6) = 1 - bicdf(15, .2, 5) = .061$

278. In a sample of $n = 400$ individuals from a normal distribution,

(a) What proportion would you expect to be in the tails: $|z| > 2$?
Answer: $2 * normcdf(2, E99, 0, 1) = .0455$

(b) How many individuals would that be?
Answer: $(.0455)(400) \approx 18$

(c) Find the probability that at least one individual has $|z| > 3$.
Answer: $2 * normcdf(3, E00, 0, 1) = .00270$, then $1 - bicdf(400, .0027, 0) = .661$

Sampling

279. Is it an observational study or a designed experiment?

(a) A sociologist gives a survey about religious beliefs.
Answer: designed experiment

(b) A scientist tests whether sleep deprived mice do poorer in maze tests.
Answer: designed experiment

(c) data is collected about the lifespan of NFL players who have suffered concussions
Answer: obs study

(d) data is collected about whether rats that are given steroids have increased cancer risk
Answer: designed experiment

(e) a farmer tries different types of fertilizer on selected plots of land, and records the harvest yield of each
Answer: designed experiment

(f) data is collected about whether children from one-parent homes are more likely to commit suicide.
Answer: obs study

(g) a bored driver counts the number of men and women drivers that he passes, or pass him, on the interstate
Answer: obs study

280. Is it a simple random sample (SRS), a stratified sample (SS), or a voluntary response sample (VRS)?

(a) An urban charter school is opening and has 100 open spots, so a lottery is held for the 500 applicants. Each potential student has a ball in the hopper.
Answer: SRS

(b) Wal-mart does random drug tests by randomly selecting 10 full-time and 10 part-time employees.
Answer: SS

(c) Prize winners are selected by pulling names out of a (well mixed) hat.
Answer: SRS