III . Anticipating Patterns: Exploring random phenomena using probability and simulation (20%–30%): Probability is the tool used for anticipating what the distribution of data should look like under a given model.

A . Probability

- **1 . Interpreting probability, including long-run relative frequency interpretation**
- **2 . "Law of Large Numbers" concept**
- **3 . Addition rule, multiplication rule, conditional probability and independence**
- **4 . Discrete random variables and their probability distributions, including binomial and geometric**
- **5 . Simulation of random behavior and probability distributions**
- **6 . Mean (expected value) and standard deviation of a random variable, and linear transformation of a random variable**

B . Combining independent random variables

- **1 . Notion of independence versus dependence**
- **2 . Mean and standard deviation for sums and differences of independent random variables**

C . The normal distribution

- **1 . Properties of the normal distribution**
- **2 . Using tables of the normal distribution**
- **3 . The normal distribution as a model for measurements**

D . Sampling distributions

- **1 . Sampling distribution of a sample proportion**
- **2 . Sampling distribution of a sample mean**
- **3 . Central Limit Theorem**
- **4 . Sampling distribution of a difference between two independent sample proportions**
- **5 . Sampling distribution of a difference between two independent sample means**
- **6 . Simulation of sampling distributions**
- **7 . t-distribution**
- **8 . Chi-square distribution**

Checking for Independence in Samples

An advertising agency in a large city is conducting a survey of adults to investigate whether there is an association between highest level of educational achievement and primary source for news. The company takes a random sample of 2,500 adults in the city. The results are shown in the table below.

Highest Level of Educational Achievement

Primary Source For News	Not a high school graduate	High school graduate but not a college graduate	College graduate	Total
Newspaper	49	205	188	442
Local Television	90	170	75	335
Cable Television	113	496	147	756
Internet	41	401	245	687
None	77	165	38	280
Total	370	1437	693	2500

Know how to find a probability, a union probability (A or B), and a conditional probability using a two way table, (11, 2a; 10b, 5ab; 03b, 2ab; M7, 18)

If an adult is selected at random from this sample, what is the probability that the selected adult obtains news primarily from the internet?

$$P(Internet) = \frac{687}{2500} = 0.2748$$

If an adult is selected at random from this sample, what is the probability that the selected adult is a college graduate or obtains news primarily from the internet?

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(College \cup Internet) = P(College) + P(Internet) - P(College \cap Internet)$$

$$= \frac{693}{2500} + \frac{687}{2500} - \frac{245}{2500} = \frac{1135}{2500} = 0.454$$

If an adult who is a college graduate is to be selected at random from this sample, what is the probability that the selected adult obtains news primarily from the internet?

We are only looking at college graduates

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(Internent|College) = \frac{P(Internet \cap College)}{P(College)} = \frac{\frac{245}{2500}}{\frac{693}{2500}} = \frac{245}{693} = 0.3535$$

Know the two ways to check for independence (11, 2b; 10b, 5c; 03b, 2c)

When selecting adults at random from the sample of 2500 adults are the events “is a college graduate” and “obtains news primarily from the internet” independent?

Method 1

$$P(A|B) = P(A) \rightarrow P(Internet|College) = P(Internet) \rightarrow 0.3535 \neq 0.2748$$

Since these are not equal, they are not independent. It appears that level of education does impact the likelihood of using the internet for news

Method 2

$$P(A \cap B) = P(A) \cdot P(B) \rightarrow P(Intenet \cap College) = P(Internet) \cdot P(College) \rightarrow \frac{245}{2500} \neq \frac{687}{2500} \cdot \frac{693}{2500}$$

Since these are not equal, they are not independent. It appears that level of education does impact the likelihood of using the internet for news

<u>OTHER STUFF</u>

Know how to find union probabilities of disjoint events, (09b, 2bc;08, 3cd; 04, 4a)

The ELISA tests whether a patient has contracted HIV. The ELISA is said to be positive if it indicates that HIV is present in a blood sample, and the ELISA is said to be negative if it does not indicate that HIV is present in a blood sample. Because of variability in antibody levels among human patients, the ELISA does not always indicate the correct result.

As part of a training program, the staff applied the ELISA to 500 blood samples known to not contain HIV. The ELISA was positive for 37 of those blood samples and negative for the other 463 samples.

When a blood sample yields a positive ELISA result, two more ELISAs are performed on the same blood sample. If at least one of the two additional ELISAs is positive, the blood sample is subjected to a more expensive and more accurate test to make a definitive determination of whether HIV is present in the sample. Under the assumption of independence and using the data above, what is the probability that a new blood sample that comes into the lab will be subjected to the more expensive test if that sample does not contain HIV?

$P(positive) = \frac{37}{500} = 0.074$

We will have to run the more expensive test for one of the two situations listed below

1) PP (2 positive tests in a row)
2) PNP (First positive, then a negative, but third test positive)

Since we are assuming independence, a positive result will always have 0.074 probability.

The probability of option 1, positive on first test and then another positive test is

P(positive then positive) = (0.074)(0.074) = 0.005476

The probability of option 2, positive on the first test, negative on second test, positive on third test is

P(positive then negative then positive) = (0.074)(0.926)(0.074) = 0.005070776

So, the probability of having to run the more expensive test if you are HIV negative is

= 0.005476 + 0.005070776
= 0.010546776

Know how to find a joint probability (A and B) of two or more independent events, (08, 3b; 04, 3b; 04, 4a; 03b, 5a; M7, 38; M2, 36)

In the problem above, each test is independent

$P(Positive \cap Postivie) = 0.074 \cdot 0.074$ is the joint probability of two independent events

Know how to interpret a probability in the context of a problem and deem if it is likely to occur or not, (04, 3c)

If I have a 0.0001% chance of winning a game, you would determine that this probably would not happen randomly. You would call me a cheater if I won.

If I have a 25% chance of winning a game, you would determine that this could happen randomly. You would not have strong evidence to call me a cheater if I won.

Know what it means for events to be mutually exclusive and how that affects their joint probability and their union probability, (M7, 36; M2, 23)

If two events are mutually exclusive, they cannot happen at the same time. For example, I cannot be in ninth grade and tenth grade at the same time. If two events are mutually exclusive, $P(A \cup B) = P(A) + P(B)$. You do not have to worry about $P(A \cap B)$ because $P(A \cap B) = 0$ for disjoint events. They cannot occur simultaneously.

Know how to set up a simulation and perform it using a table of random digits, (M2, 4)

If 80% of population is female and 20% male, you could use a RNG and make

0,1→Male 2, 3, 4, 5, 6, 7, 8, 9→Female

Probability Distributions

Know how to find the expected outcome of a discrete random variable, (12, 2b; 08, 3a; 05, 2a; 05b, 2a; 04, 4b; 03b, 5b; M2, 5)

Let the random variable X represent the number of caffeinated drinks that a student drinks on a daily basis. The probability distribution of X is shown in the table below.

X	0	1	2	3	4	5	6
P(X)	0.15	0.30	0.25	0.15	0.10	0.04	0.01

Calculate the expected value (the mean) of X.

$$E(X) = \mu_x = 0(0.15) + 1(0.30) + 2(0.25) + 3(0.15) + 4(0.10) + 5(0.04) + 6(0.01) = 1.91$$

Know what the law of large numbers is and how to describe in the context of a problem, (05, 2b)

Using a sample of 30 students, there was an average of 1.75. A group plans to take another sample of 1000 students and compute the average from those students. How do you expect this average to compare to that of the smaller sample? Justify your response.

We would expect the average to be larger than 1.75, because by the Law of Large numbers our average should be approaching 1.91 (the expected count). The more observations we take, we should approach 1.91

Know how to find the median, quartiles, and interquartile range by looking at a probability distribution for a discrete random variable, (05, 2c; M7, 12; M7, 24)

Using the data about caffeinated drinks, it is obvious that the minimum is 0 and the maximum is 6. Since the 15th percentile is the last 0 and the 45th percentile is the last 1, we can conclude that the first quartile (25th percentile) is also 1, because it would be all 1's between the 15th and 45th percentile. Using that same logic, we can say that the median (50th percentile) is 2 and the third quartile (75th percentile) is 3.

Know how to find the standard deviation of a discrete random variable, (05b, 2a)

$$\sigma^2(x) = \sum (x_i - \mu_x)^2 p_i = (0 - 1.91)^2(0.15) + (1 - 1.91)^2(0.30) + (2 - 1.91)^2(0.25) + \cdots (6 - 1.91)^2(0.01)$$

If you have two normal distributions that are both normal and you add or subtract them, what is the shape of the new distribution? How do you find the average of the new distribution? How do you find the standard deviation of the new distribution? (08b, 5a; 05b, 2b; M7, 26)

If Mr. Merlo's bowling scores follow N(120, 30), and Mr. Furutani's bowling scores follow N(100, 40). How often should Mr. Merlo beat Mr. Furutani if their scores are independent.

Let's look at the difference between their scores

$\mu_M - \mu_F = 120 - 100 = 20$. Mr. Merlo should win by 20 on average.

$$\sigma^2_{M-F} = \sigma^2_M + \mu^2_F \rightarrow \sigma^2_{M-F} = 900 + 1600 \rightarrow \sigma_{M-F} = 50$$

So, now we know that the difference in the scores follows $N_{M-F}(20{,}50)$, because Mr. Merlo should win by 20 on average, but there is a standard deviation of 50, because he isn't always going to win by exactly 50.

$$P(Merlo\ Wins) = P(M - F > 0) = P\left(z > \frac{0 - 20}{50}\right) = P(z > -0.4) = 0.6554$$

Merlo should win 65.54% of the time.

More Probability Distributions

Know how to find the standard deviation when you average two variables, (08, 4c)

Using the information from above, what is the average and standard deviation of Merlo and Furutani's average bowling scores? Essentially, if they were to average ($\frac{M+F}{2}$) their scores every time they played, what would the distribution of those numbers be.

$$\mu_{\frac{M+F}{2}} = \frac{\mu_M + \mu_F}{2} = \frac{120 + 100}{2} = 110$$

$$\sigma^2_{\frac{M+F}{2}} = \sigma^2_{\frac{1}{2}M+\frac{1}{2}F} = \sigma^2_{\frac{1}{2}M} + \sigma^2_{\frac{1}{2}F} = 15^2 + 20^2 \quad \rightarrow \quad \sigma_{\frac{1}{2}M+\frac{1}{2}F} = 25$$

Know how to find the average and standard deviation when you combine distributions and change the units(μ_{2a+3b} or σ_{2a+3b}), (12, 6bc; 05b, 2c)

Know that if you add a constant to every number, the average goes up by that constant, but the standard deviation remains the same. If you multiply a constant by every number, both the average and the standard deviation are multiplied by that number.

Multiplying by a constant multiplies your standard deviation→ $\sigma_{3x} = 3\sigma_x$

Adding a constant doesn't change the standard deviation→ $\sigma_{x+10} = \sigma_x$

If you do both, only the multiplication affects the standard deviation→ $\sigma_{3x+10} = 3\sigma_x$

Binomial and Geometric Distributions

A local lottery has been created to help the elementary school in the area. The cost of one ticket is $10, and if your ticket is chosen, then you win $20. The people running the lottery guarantee that each ticket has a 40% chance of winning.

If you believe that you are doing a binomial question, state so immediately in your answer, B(n, p), (11b, 2b; 10b, 3a;09, 2b; 07b, 1b; 06b, 6c; 03, 3c). Know how to solve a basic binomial problem (P(X=1)), (14, 3c; 11b, 2b; 07b, 2b; 06b, 6c; 03, 3c; M2, 32)

Using the lottery situation described above, find the probability of having exactly 7 winning lottery tickets if you purchases 10 total tickets.

State distribution --- B(n,p) → B(10, 0.4)

P(X = 7) = $\binom{10}{7}(0.4)^7(0.6)^3 = 0.0425$

Or

Binompdf(n,p,X) = Binompdf(10, 0.4, 7) = 0.0425 (Note: make sure you note that n =10, p =0.4, and X =7)

Can you set up and solve a binomial problem that is cumulative using a calculator($PX \leq 30$)=binomcdf(n, p, 30)? (2010, 4b;10b, 3c;09, 2b). Know how to solve a binomial problem that is the complement of a cumulative problem ($P(X \geq 3) = 1 - P(X \leq 2) = 1 - binomcdf(n, p, 2)$), (11b, 3c; 06, 3b; 06b, 3b)

Using the lottery situation described above, find the probability that you bought at least 7 winning tickets if you purchased 10 total tickets.

State distribution --- B(n,p) → B(10,0.4)

P($X \geq$ 7) = 1 – P(X ≤ 6) = 1 – Binomcdf(10, 0.4, 6) = 0.0548

Where n =10, p = 0.4, and X = 6 (Note: make sure you note what the parts of the computer input represent)

Can you find the expected outcome and standard deviation of a binomial distribution? (10, 4a;10b, 3b)—525*

Using the lottery situation described above, how many winning tickets would you expect to hold if you purchased 10 tickets? What would be the standard deviation of the winning number or tickets if you purchased 10 tickets.

$\mu_x = np = 10(0.4) = 4$ tickets

$\sigma_x = \sqrt{np(1-p)} = \sqrt{10(0.4)(0.6)} = 1.549$ tickets

If people purchased 10 tickets, the average number of winning tickets would be 4 tickets with a standard deviation of 1.549 tickets.

Know the conditions for a binomial problem and when a problem is not binomial, (04, 3a; M7, 7; M7, 11)

So far this year Matt Kemp gets a hit in 40% of his at bats. During the season he will probably have 500 at bats, and he gets a bonus if he has more than 200 hits. If X is the number of hits that he gets, does X have a binomial distribution? Explain why or why not

There are four things necessary to create a binomial distribution

1) Set number of observations
2) Success/Failure
3) Probability is constant
4) Observations are independent

In this case, 1 and 2 have been met. He will have 500 observations in which he either gets a hit (success) or does not get a hit (failure). In this scenario, though, 3 and 4 have not been met. It is probably not true that Matt Kemp's probability of success is always 40%. It will most likely be a lot harder to get a hit if he is facing the best pitcher in the league, or a lot easier to get a hit if he is facing the worst pitcher in the league. At bats are not independent, each at bat is affected by several factors, including pitcher, weather, batter health, time or game, etc. So, using the binomial formula would not be appropriate.

Know how to solve a geometric problem where you need to find the probability of taking more than a certain number of tries (Find the probability of taking more than 5 tries to succed), (11b, 3a)

A local lottery has been created to help the elementary school in the area. The cost of one ticket is $10, and if your ticket is chosen, then you win $20. The people running the lottery guarantee that each ticket has a 40% chance of winning. What is the likelihood that you would have to purchase more than 10 tickets in order to get at least one winning ticket.

State distribution → G(0.4) because you are essentially playing until you win

P(X > 10) = P(X = 11) + P(X = 12) + P(X = 13) + P(X =14) +…

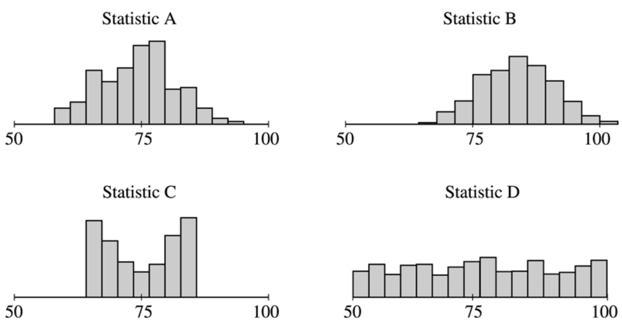
Since it is infinite, that would be hard to calculate all those individual probabilities. It's easier to think of it this way: Taking longer than 10 tries implies that you MUST lose your first ten tries.

So, P(losing first ten tries) = $0.6^{10} = 0.00605$

Sampling Distributions

What does it mean to be considered a biased or unbiased estimator? (08b, 2; M7, 33)

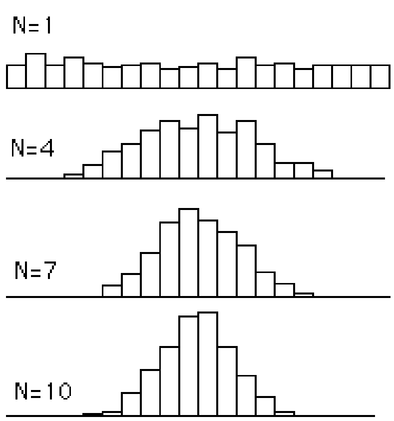
Four different statistics have been proposed as estimators of a population parameter. To investigate the behavior of these estimators, 500 random samples are selected from a known population and each statistic is calculated for each sample. The true value of the population parameter is 75. The graphs below show the distribution of values for each statistic.



Which of the statistics appear to be unbiased estimators of the population parameter?How can you tell?

Statistics A, C, and D appear to be unbiased. This is indicated by the fact that the mean of the estimated sampling distribution for each of these statistics is about 75, the value of the true population parameter.

Know that when you are referring to a standard deviation of a sampling distribution of the sample mean, you do not have to divide by rad n, it has all ready been done, (07, 3b)



Definitions

Sampling Distribution: This is the distribution of the original data taken individually (n =1). This is what would be created if you collected every person's weight in school.

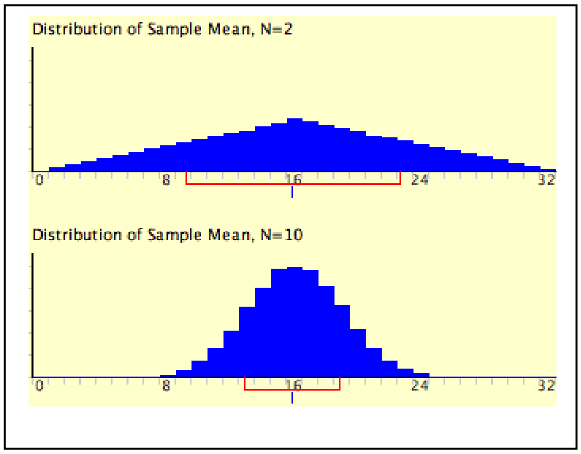
The parameters for a sampling distribution would be μ and σ

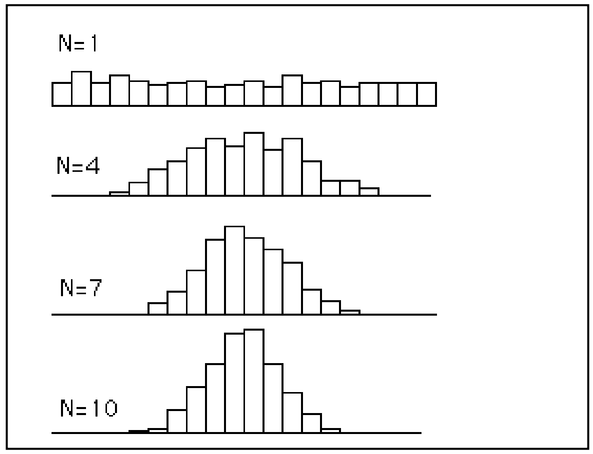
Sampling Distrubution of sample means: This is the distribution of average taken from samples of different sizes (n=4, n=7, or n=10). This is what would be created if 100 people surveyed 4 people randomly and collected their weight. We would then take those 100 averages and create a graph, as is done in N = 4 above.

The parameters for a sampling distribution of sample means would be $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

How is a sampling distribution of a sample mean different than a distribution of individual observations? (10, 2a; M7, 23). Know that when you have a large sample size, even the most nonnormal distributions will have a sampling distribution of the sample mean close to a normal distribution by the Central Limit Theorem, (07, 3c; 07b, 2c)

Sampling Distributions of Sample Means





As your sample size become larger, these are things that you you notice:

1) It becomes more normal (But will never be normal, only approximately normal)—(This is the Central Limit Theorem)
   a. If something is already symmetric, then you only have to make a few observations for the sampling distribution of the sample means to become approximately normal (notice the picture on the right)
   b. If something is not at all symmetric, then you need about 30 observations for the sampling distribution of the sample means to become approximately normal
2) The average remains the same (If the overall $\mu = 16$ then $\mu_{\bar{x}} = 16$ no matter how large the sample size)
3) It becomes less spread (If the overall $\sigma = 7$ then $\sigma_{\bar{x}} = \frac{7}{\sqrt{n}}$ no matter what n is)

This is true for all independent observations, even if the sampling distribution of the sample means is not normal. This means that these formulas apply to all of the distributions pictured in the figure on the right. It is true for n =1, n = 4, n = 7, n= 10.

Know how sampling distributions of sample means change as your sample size gets larger, and how that affects the likelihood of extreme events occurring, (07, 3a; M2, 9)

Big Town Fisheries recently stocked a new lake in a city park with 2,000 fish of various sizes. The distribution of the lengths of these fish is approximately normal.

(a)Big Town Fisheries claims that the mean length of the fish is 8 inches. If the claim is true, which of the following would be more likely? Justify your answer.
• A random sample of 15 fish having a mean length that is greater than 10 inches or
• A random sample of 50 fish having a mean length that is greater than 10 inches

We would expect it to be more difficult to get a random sample of 50 fish to be greater than 10, because of the difference in variability. Whether your sample size is n =15 or n =50, they both should be centered around 8 inches, because they came from the same population (this is true whether it was approximately normal or not). Even though they have the same center, though, we will expect a difference in variability. It is more difficult to have a group of 50 be far away from average than to have a group of 15 be far away from the average. Mathematically speaking, no matter what the standard deviation is, $\frac{\sigma}{\sqrt{50}} < \frac{\sigma}{\sqrt{15}}$. So, there is more variability in a group of 15 than a group of 50.

Know how to find the probability of averaging a certain amount in a Normal distribution ($P(\bar{x} > 10)$? (10, 2b; 09, 2c; 07, 3b; 06, 3c; 04b, 3c). Know the difference between finding the probability that a group of 10 will average more than 150 lbs and a group of ten will all weight more than 150 lbs, (05b, 6c). Know how to find the average and standard deviation of a sampling distribution of a sample mean if you know the average and standard deviation of the original data, (11b, 6c; 07b, 2c; M2, 30)

A group of 10,000 people have a normal distribution of weights with an average of 145 and a standard deviation of 20.

1) What is the probability that a randomly chosen person will weigh more than 150lb?

N(145, 20)

$$P(X > 150)$$
$$= P\left(Z > \frac{150 - 145}{20}\right)$$
$$= P(Z > 0.25)$$
$$= 0.4013$$

2) What is the probability of 10 random people averaging more than 150lb?

Using the information, we know that the average and standard deviation of the sampling distribution of sample means follows N(145, $\frac{20}{\sqrt{10}}$)

$$P(\bar{X} > 150)$$
$$= P\left(Z > \frac{150 - 145}{\frac{20}{\sqrt{10}}}\right)$$
$$= P(Z > 0.79)$$
$$= 0.2148$$

3) What is the probability that 10 random people will all weigh more than 150?

Since the probability that one individual will weigh more than 150 is 0.4013, the probability that 10 individuals will all weigh more than 150 is…

$$(0.4013)^{10}$$
$$= 0.000108$$

Know that when you have a large sample size, even the most nonnormal distributions will have a sampling distribution of the sample mean close to a normal distribution by the Central Limit Theorem, (07, 3c; 07b, 2c)

Suppose the distribution of weights in the previous problem was was nonnormal but had the same mean and standard deviation. Would it still be appropriate to use the normal distribution to compare the probability in question (2)? Justify your answer.

In this case, it would not be appropriate, because a sample size of 10 would not guarantee that the distribution of sample means would be approximately normal. If, though, we took a sample great than 30, we would be able to say that the distribution of sample means is approximately normal by the Central Limit Theorem.

Note: If you said normal, it would be partially correct. You must say APPROXIMATELY normal.

Practice Problems #1

1) In a particular game, a fair die is tossed. If the number of spots showing is either a 4 or 5, you win $1; if the number of spots showing is a 6, you win $4; and if the number of spots showing is 1, 2, or 3, you win nothing. Let X be the amount that you win on a single play of the game. What is the expected value of X?
   a. $0
   b. $1
   c. $1.33
   d. $2.50
   e. $4
2) In a particular game, a fair die is tossed. If the number of spots showing is either a 4 or 5, you win $1; if the number of spots showing is a 6, you win $4; and if the number of spots showing is 1, 2, or 3, you win nothing. Let X be the amount that you win on a single play of the game. What is the variance of X?
   a. 1
   b. 1.414
   c. 3/2
   d. 2
   e. 13/6
3) The weight of medium-sized tomatoes selected at random from a bin at the local supermarket is a random normal distribution with mean $\mu = 10$ ounces and standard deviation $\sigma = 1$ ounce. Suppose we pick two tomatoes at random from the bin, so the weights of the tomatoes are independent. What is the probability that the sum of the weights of the two tomatoes exceeds 22 ounces is closest to
   a. 0.017
   b. 0.068
   c. 0.079
   d. 0.159
   e. 0.921

4) The admissions policy at a certain university requires that incoming students score in the upper 20% on a standardized test. If the mean score on the test is 510 and the standard deviation of the scores is 80, what is the minimum score that students can earn on the test to meet the admissions requirement? Scores on the test are normally distributed and are reported in intervals of 10.
   a. 520
   b. 580
   c. 590
   d. 600
   e. 620
5) Which of the following is true concerning a normal distribution curve
   I) The curve is symmetric about its mean
   II) The curve is always symmetric about 0
   III) The area under the curve lying within one standard deviation of the mean contains approximately 50% of the scores
      a. I only
      b. II only
      c. III only
      d. I and III only
      e. I, II, and III
6) The scores on the real estate licensing examination given in a particular state are normally distributed with a standard deviation of 70. What is the mean test score if 25% of the applicants score above a 475?
   a. 416
   b. 428
   c. 468
   d. 522
   e. There is not enough information

7) A normal population has a mean of 100 and standard deviation of 16. If a sample of size 64 is drawn from the population, what proportion of the sample should be between 91 and 103?
   a. 0.2875
   b. 0.4987
   c. 0.9050
   d. 0.9319
   e. 0.9332
8) The mean score on a nationally used standardized reading exam was 75 with a standard deviation of four. What would be the z-score for a child who scored an 81?
   a. -1.50
   b. -1.03
   c. 1.03
   d. 1.05
   e. 1.50
9) Evan and Amy are two seniors in a local high school. The grades on a Calculus final which Evan took had an average of 75 and a standard deviation of 7. The grades on a Psychology final which Amy took had an average of 85 with a standard deviation of 4. Even scored an 89 on the Calculus exam and Amy scored a 91 on the Psychology exam. Relative to the other students in their classes, who did better?
   a. Evan
   b. Amy
   c. Both were the same
   d. You cannot compare because we don't know if the distributions were normal
   e. There is not enough information

Solutions

1) B. E(X) = 0(3/6) + 1(2/6) + 4(1/6) = 6/6 = 1
2) D. Using the handy dandy notebook, we can find that the variance of a probability distribution is Var(X) = $(0-1)^2\left(\frac{3}{6}\right) + (1-1)^2\left(\frac{2}{6}\right) + (4-1)^2\left(\frac{1}{6}\right) = \frac{12}{6} = 2$. If you wanted to find the standard deviation, you would square root the variance, $\sqrt{2}$.
3) C. We can think of choosing two tomatoes as two distributions. In other words, the first event is picking a tomato from a N(10,1) distribution, and the second event is picking a tomato from an equal N(10,1) distribution. In that case
   $\mu_1 + \mu_2 = 10 + 10 = 20$. If we choose two tomatoes, we expect the sum to be 20 ounces on average.
   $\sigma_{1+2}^2 = \sigma_1^2 + \mu_2^2 \rightarrow \sigma_{1+2}^2 = 1 + 1 \rightarrow \sigma_{1+2} = \sqrt{2}$. The sum should be twenty give or take 1.414 ounces.

   Now, we have a new distribution. If everybody in the world chose two of these tomatoes and added their weights together, the distribution would be N(20,1.414).

$$P(Sum > 22) = P\left(z > \frac{22-20}{1.414}\right) = P(z > 1.414) = 0.0786$$

4) B. N(510, 80). The top 20% has 80% below it, which produces a z = 0.84 in a normal distribution. $0.84 = \frac{x-510}{80}$. This produces x = 577.2. 570 won't put you in the top 20% but 580 will. You could also use invNorm(0.8, 510, 80) = 577.3 on your calculator to do percentiles.
5) A. I—A normal curve always has the average at the middle. II-The curve is always symmetric around a z = 0, but that doesn't mean that the average is always 0. III-By the Empirical Rule (68-95-99.7 Rule) 68% of the data lies within one standard deviation of the average, not 50%.
6) B. N(μ, 70). We don't know the average, but we do know that the top 25% has a z = 0.68. So, $0.68 = \frac{475-\mu}{70}$. $\mu = 427.4$
7) A. First, we do not need to divide by $\sqrt{n}$, because we are not asking for the likelihood of the 64 AVERAGING a certain amount. In this case it doesn't matter if we pull 64 or 64,000. The proportion that fulfill our qualifications will be the same. So, N(100,16). P(91 < X < 103) = P(-0.56 < z < 0.19) = 0.5714 – 0.2877 = 0.2837. You can also use normalcdf(91,103,100,16) = 0.2875
8) E. $z = \frac{81-75}{4} = 1.50$

A. Evan → $z = \frac{89-75}{7} = 2.00$. Amy → $z = \frac{91-85}{4} = 1.50$. Since Evan was two standard deviations above average and Amy was only 1.50 standard deviations above average, Evan did butter even though his score was lower

Practice Problems Number 2

1) A survey reveals that 60% of patients who receive prescription medications from their doctors prefer the brand name drug as opposed to the generic brand. In a random sample of 40 patients, what is the probability that at least 25 of them prefer the brand name?
a) 0.1255
b) 0.3745
c) 0.4402
d) 0.6250
e) 0.9500

2) A survey indicates that approximately 40% of college students own a cellular phone. In a random sample of 15 college students, what is the probability that at least five own a cell phone?
a) 0.010
b) 0.186
c) 0.217
d) 0.333
e) 0.783

3) A report from the Main Department of Inland Fisheries and Wildlife indicates that there occurs on average 1 fatality per 100 collisions between cars and deer. In 300 collisions between a car and a deer, what is the expected number of fatalities and the standard deviation?
a) Mean: 0.33 standard deviation: 0.01
b) Mean: 1 standard deviation: 0.01
c) Mean: 3 standard deviation: 0.172
d) Mean: 3 standard deviation: 2.97
e) Mean: 30 standard deviation: 3.0

4) Which of the following are conditions for a binomial experiment?
i. Number of trials n is a fixed number
ii. The n trials are independent
iii. The probability of success p is equal to the probability of failure q
a) i. only
b) ii. only
c) iii only
d) i. and ii. only
e) i. ii. and iii.

5) The all-time leader in career batting average among major league baseball players is Ty Cobb with career average of 0.366. This means he got a hit in 36.6% of his official at-bats. What was Cobb's probability of getting at least one hit in 4 official at-bats?
a) 0.092
b) 0.134
c) 0.162
d) 0.366
e) 0.838

6) Mr. and Mrs. Jones have 5 children. The 5 births were independent with no multiple births. Assuming that the chances of having a boy or a girl are the same, what is the probability that the family has at least 2 girls?
a) 0.1875
b) 0.3125
c) 0.4000
d) 0.5000
e) 0.8125

7) According to a CBS/ New York Times poll taken in 1992, 15% of the public have responded to a telephone call-in poll. In a random group of five people, what is the probability that exactly two have responded to a call-in poll?
A. 0.138
B. 0.165
C. 0.300
D. 0.835
E. 0.973

8) In a 1974 “Dear Abby” letter a woman lamented that she had just given birth to her 8th child, and all were girls! Her doctor had assured her that the chance of the 8th child being a girl was only 1 in 100. What was the real probability that the 8th child would be a girl?
A. 0.01
B. 0.5
C. (0.5)^7
D. (0.5)^8
E. (0.5)^7+(0.5)^8

9) The yearly mortality rate for American men from prostate cancer has been constant for decades at about 25 of every 100,000 men. In a group of 100 American men, what is the probability that at least 1 will die from prostate cancer in a given year?

A. 0.00025
B. 0.0247
C. 0.025
D. 0.9753
E. 0.99975

10) 65% of all divorce cases cite incompatibility as the underlying reason. If four couples file for a divorce, what is the probability that exactly 2 will state incompatibility as the reason?
A. 0.104
B. 0.207
C. 0.254
D. 0.311
E. 0.423

Answers:

1) c) 0.4402
$P(X \geq 25)$ = 1 - binomcdf (40, .6, 24) = 1 - .5598 = 0.4402

2) e) 0.783
n= 15 p= .4 $x \geq 5$

1- (Binomcdf(15, .4, 4)
1- (.2173)
= 0.7827

3) c)
(300) (1/100) = 300 (0.01) = 3
Expected = 3
Sd= $\sqrt{np(1-p)}$
Sd= $\sqrt{(330(0.01))(1- 0.01)}$
Sd= 1.72

4) d) i and ii only
i and ii are both true but iii states the probability of success *p* is equal to the probability of failure *q* which is wrong because the probability of failure would equal 1-p.

5) e) 0.838
n= 4 p=.366 x= 0
1- (binomcdf(4, .366, 0))
= 0.838

6) e) 0.8125
n= 5 p= .5 x=1
1- (binomcdf (5, .5, 1))
1- (.1875)
= 0.8125

7) A: binompdf(5,0.15,2)= 0.138. 5 is the sample taken. 0.15 is the proportion. 2 is the x

8) B: the probability of the next child being a girl is independent of the gender of the previous children.

9) B: 1-(0.99975)^100 = 0.0247

10) Binompdf(4, 0.65, 2)=0.311

Practice Problems Number 3

1) A consequence of the Central Limit Theorem is that for nu sufficiently large (n > 30), if all samples of size n are taken, the mean of the sample means $\mu_{\bar{x}}$ is equal to the population μ. Since the mean of the sampling distribution is equal to the population mean, $\bar{x}$ is referred to as
   a. A biased estimator
   b. An unbiased estimator
   c. A random estimator
   d. A controlled variable
   e. A parameter
2) Billie decides to participate in one of two games of chance involving dice. In the first game each player gets a chance to throw the dice 100 times and wins if the dice show a sum of 7 between 15% and 20% of the time. In the second game, he gets to toss the dice 125 times and wins if his dice show a sum of 6 or 8 between 30% and 35% of the time. Billie decides to play the first game because he has a better chance of winning in that game, according to
   a. The law of large numbers
   b. The law of averages
   c. The central limit theorem
   d. Chebyshev's theorem
   e. Bayes' theorem
3) A normal population has a mean of 100 and a standard deviation of 16. If samples of size 64 are drawn from the population, what proportion of the samples should have mean values between 91 and 103?
   a. 0.0668
   b. 0.4987
   c. 0.9050
   d. 0.9319
   e. 0.9332
4) The variance of a random variable is 100 and a sample size of 25 is taken. The standard error of the sample means is
   a. 2
   b. 4
   c. 5
   d. 10
   e. 100
5) The Central Limit Theorem of statistics states which of the following for a sample of size n, (n > 1)
   a. The shape of the sampling distribution of sample means approaches normal as the n gets larger
   b. The mean of the set of sample means is always less than the mean of the population
   c. The standard deviation of the set of sample means is equal to the standard deviation of the population
   d. The mean of the set of sample means is smaller than the mean of the population when n is large
   e. The standard deviation of the set of sample means is greater than the standard deviation of the population

1) B
2) A
3) E. normalcdf(91, 103, 100, $\frac{16}{\sqrt{64}}$) $= 0.933189$
4) A. If the variance $\sigma^2 = 100$, then $\sigma = 10$. So the standard deviation of the sample means would be $\frac{10}{\sqrt{25}} = 2$.
5) A