3.4 Binomial distribution

Fundamental Counting Principle

The **Fundamental Counting Principle** can be used determine the number of possible outcomes when there are two or more characteristics .

The **Fundamental Counting Principle** states that if an event has m possible outcomes and another independent event has n possible outcomes, then there are $m * n$ possible outcomes for the two events together.

Fundamental Counting Principle

Lets start with a simple example.

A student is to roll a die and flip a coin. How many possible outcomes will there be?

1H	2H	3H	4H	5H	6H
1T	2T	3T	4T	5T	6T

6*2 = 12 outcomes

12 outcomes

Fundamental Counting Principle

For a college interview, Robert has to choose what to wear from the following: 4 slacks, 3 shirts, 2 shoes and 5 ties. How many possible outfits does he have to choose from?

4*3*2*5 = 120 outfits

Permutations

A **Permutation** is an arrangement of items in a particular order.

Notice, ORDER MATTERS!

To find the number of Permutations of n items, we can use the Fundamental Counting Principle or factorial notation.

Permutations

The number of ways to arrange the letters ABC: ___ ___ ___

Number of choices for first blank? 3 ___ ___

Number of choices for second blank? 3 2 ___

Number of choices for third blank? 3 2 1

3*2*1 = 6   3! = 3*2*1 = 6

ABC ACB BAC BCA CAB CBA

Permutations

To find the number of Permutations of n items chosen r at a time, you can use the formula

$$_nP_r = \frac{n!}{(n-r)!} \quad \text{where } 0 \le r \le n.$$

$$_5P_3 = \frac{\mathbf{5!}}{\mathbf{(5-3)!}} = \frac{5!}{2!} = \mathbf{5*4*3 = 60}$$

Permutations

Practice:

A combination lock will open when the right choice of three numbers (from 1 to 30, inclusive) is selected. How many different lock combinations are possible assuming no number is repeated?

$$_{30}P_3 = \frac{\mathbf{30!}}{\mathbf{(30-3)!}} = \frac{30!}{27!} = \mathbf{30*29*28 = 24360}$$

Permutations

Practice:

From a club of 24 members, a President, Vice President, Secretary, Treasurer and Historian are to be elected. In how many ways can the offices be filled?

$$_{24}p_5 = \frac{24!}{(24-5)!} = \frac{24!}{19!} =$$

$$24 * 23 * 22 * 21 * 20 = 5{,}100{,}480$$

Combinations

A **Combination** is an arrangement of items in which order does not matter.

ORDER DOES NOT MATTER!

Since the order does not matter in combinations, there are fewer combinations than permutations. The combinations are a "subset" of the permutations.

Combinations

To find the number of Combinations of n items chosen r at a time, you can use the formula

$$_nC_r = \frac{n!}{r!(n-r)!} \quad \text{where } 0 \le r \le n.$$

Combinations

To find the number of Combinations of n items chosen r at a time, you can use the formula

$$_nC_r = \frac{n!}{r!(n-r)!} \quad \text{where } 0 \le r \le n.$$

$$_5C_3 = \frac{5!}{3!(5-3)!} = \frac{5!}{3!2!} =$$

$$\frac{5*4*3*2*1}{3*2*1*2*1} = \frac{5*4}{2*1} = \frac{20}{2} = 10$$

Combinations

Practice: To play a particular card game, each player is dealt five cards from a standard deck of 52 cards. How many different hands are possible?

$$ {}_{52}C_5 = \frac{52!}{5!(52-5)!} = \frac{52!}{5!47!} = $$

$$ \frac{52*51*50*49*48}{5*4*3*2*1} = 2{,}598{,}960 $$

Combinations

Practice: A student must answer 3 out of 5 essay questions on a test. In how many different ways can the student select the questions?

$$ {}_5C_3 = \frac{5!}{3!(5-3)!} = \frac{5!}{3!2!} = \frac{5*4}{2*1} = 10 $$

Combinations

Practice: A basketball team consists of two centers, five forwards, and four guards. In how many ways can the coach select a starting line up of one center, two forwards, and two guards?

Center:

$$ {}_2C_1 = \frac{2!}{1!1!} = 2 $$

Forwards:

$$ {}_5C_2 = \frac{5!}{2!3!} = \frac{5*4}{2*1} = 10 $$

Guards:

$$ {}_4C_2 = \frac{4!}{2!2!} = \frac{4*3}{2*1} = 6 $$

$$ {}_2C_1 * {}_5C_2 * {}_4C_2 $$

Thus, the number of ways to select the starting line up is 2*10*6 = 120.

Suppose we randomly select four individuals to participate in the Milgram experiment. What is the probability that exactly 1 of them will refuse to administer the shock?

Let's call these people Allen (A), Brittany (B), Caroline (C), and Damian (D). Each one of the four scenarios below will satisfy the condition of "exactly 1 of them refuses to administer the shock":

Scenario 1: $\underset{\text{(A) refuse}}{0.35} \times \underset{\text{(B) shock}}{0.65} \times \underset{\text{(C) shock}}{0.65} \times \underset{\text{(D) shock}}{0.65} = 0.0961$

Scenario 2: $\underset{\text{(A) shock}}{0.65} \times \underset{\text{(B) refuse}}{0.35} \times \underset{\text{(C) shock}}{0.65} \times \underset{\text{(D) shock}}{0.65} = 0.0961$

Scenario 3: $\underset{\text{(A) shock}}{0.65} \times \underset{\text{(B) shock}}{0.65} \times \underset{\text{(C) refuse}}{0.35} \times \underset{\text{(D) shock}}{0.65} = 0.0961$

Scenario 4: $\underset{\text{(A) shock}}{0.65} \times \underset{\text{(B) shock}}{0.65} \times \underset{\text{(C) shock}}{0.65} \times \underset{\text{(D) refuse}}{0.35} = 0.0961$

The probability of exactly one 1 of 4 people refusing to administer the shock is the sum of all of these probabilities.

$$ 0.0961 + 0.0961 + 0.0961 + 0.0961 = 4 \times 0.0961 = 0.3844 $$

62

Binomial distribution

The question from the prior slide asked for the probability of given number of successes, k, in a given number of trials, n, (k = 1 success in n = 4 trials), and we calculated this probability as

# of scenarios x P(single scenario)

# of scenarios: there is a less tedious way to figure this out, we'll get to that shortly...

P(single scenario) = $p^k(1-p)^{n-k}$
where p is the probability of success to the power of number of successes, probability of failure to the power of number of failures

The *Binomial distribution* describes the probability of having exactly k successes in n independent Bernoulli trials with probability of success p.

63

Computing the # of scenarios

Earlier we wrote out all possible scenarios that fit the condition of exactly one person refusing to administer the shock. If n was larger and/or k was different than 1, for example, n = 9 and k = 2:

*RR*SSSSSSS
S*RR*SSSSSS
SS*RR*SSSSS
...
SS*R*SS*R*SSS
...
SSSSSSS*RR*

writing out all possible scenarios would be incredibly tedious and prone to errors.

64

Computing the # of scenarios

Choose function, or # combinations

The *choose function* is useful for calculating the number of ways to choose k successes in n trials.

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

$$k = 1, n = 4: \binom{4}{1} = \frac{4!}{1!(4-1)!} = \frac{4\times3\times2\times1}{1\times(3\times2\times1)} = 4$$

$$k = 2, n = 9: \binom{9}{2} = \frac{9!}{2!(9-1)!} = \frac{9\times8\times7!}{2\times1\times7!} = \frac{72}{2} = 36$$

Note: You can also use R for these calculations:

```
> choose(9,2)
[1] 36
```

65

Practice

Which of the following is false?

(a) There are n ways of getting 1 success in n trials, $\binom{n}{1} = n$.

(b) There is only 1 way of getting n successes in n trials, $\binom{n}{n} = 1$.

(c) There is only 1 way of getting n failures in n trials, $\binom{n}{0} = 1$.

(d) There are $n-1$ ways of getting $n-1$ successes in n trials, $\binom{n}{n-1} = n-1$.

66

Practice

Which of the following is false?

(a) There are n ways of getting 1 success in n trials, $\binom{n}{1} = n$.

(b) There is only 1 way of getting n successes in n trials, $\binom{n}{n} = 1$.

(c) There is only 1 way of getting n failures in n trials, $\binom{n}{0} = 1$.

(d) *There are $n-1$ ways of getting $n-1$ successes in n trials, $\binom{n}{n-1} = n-1$.*

67

Binomial distribution (cont.)

Binomial probabilities

If p represents probability of success, (1-p) represents probability of failure, n represents number of independent trials, and k represents number of successes

$$P(k \text{ successes in } n \text{ trials}) = \binom{n}{k} p^k (1-p)^{(n-k)}$$

68

Practice

Which of the following is not a condition that needs to be met for the binomial distribution to be applicable?

A. the trials must be independent
B. the number of trials, *n*, must be fixed
C. each trial outcome must be classified as a *success* or a *failure*
D. the number of desired successes, *k*, must be greater than the number of trials
E. the probability of success, *p*, must be the same for each trial

69

Practice

Which of the following is not a condition that needs to be met for the binomial distribution to be applicable?

A. the trials must be independent
B. the number of trials, *n*, must be fixed
C. each trial outcome must be classified as a *success* or a *failure*
D. *the number of desired successes, k, must be greater than the number of trials*
E. the probability of success, *p*, must be the same for each trial

70

Practice

A 2012 Gallup survey suggests that 26.2% of Americans are obese. Among a random sample of 10 Americans, what is the probability that exactly 8 are obese?

A. pretty high
B. pretty low

Gallup: http://www.gallup.com/poll/160061/obesity-rate-stable-2012.aspx, January 23, 2013.

71

Practice

A 2012 Gallup survey suggests that 26.2% of Americans are obese. Among a random sample of 10 Americans, what is the probability that exactly 8 are obese?

A. pretty high
B. pretty low

Gallup: http://www.gallup.com/poll/160061/obesity-rate-stable-2012.aspx, January 23, 2013.

72

Practice

A 2012 Gallup survey suggests that 26.2% of Americans are obese. Among a random sample of 10 Americans, what is the probability that exactly 8 are obese?

(a) $0.262^8 \times 0.738^2$
(b) $\binom{8}{10} \times 0.262^8 \times 0.738^2$
(c) $\binom{10}{8} \times 0.262^8 \times 0.738^2$
(d) $\binom{10}{8} \times 0.262^2 \times 0.738^8$

73

Practice

A 2012 Gallup survey suggests that 26.2% of Americans are obese. Among a random sample of 10 Americans, what is the probability that exactly 8 are obese?

(a) $0.262^8 \times 0.738^2$
(b) $\binom{8}{10} \times 0.262^8 \times 0.738^2$
(c) $\binom{10}{8} \times 0.262^8 \times 0.738^2 = 45 \times 0.262^8 \times 0.738^2 = 0.0005$
(d) $\binom{10}{8} \times 0.262^2 \times 0.738^8$

74

The birthday problem

What is the probability that 2 randomly chosen people share a birthday?

Pretty low, 1 / 365 ≈ 0.0027

What is the probability that at least 2 people out of 366 people share a birthday?

Exactly 1! (Excluding the possibility of a leap year birthday.)

75

The birthday problem (cont.)

What is the probability that at least 2 people (1 match) out of 121 people share a birthday?

Somewhat complicated to calculate, but we can think of it as the complement of the probability that there are no matches in 121 people.

$$
\begin{aligned}
P(\textit{no matches}) &= 1 \times \left(1 - \frac{1}{365}\right) \times \left(1 - \frac{2}{365}\right) \times \cdots \times \left(1 - \frac{120}{365}\right) \\
&= \frac{365 \times 364 \times \cdots \times 245}{365^{121}} \\
&= \frac{365!}{365^{121} \times (365 - 121)!} \\
&= \frac{121! \times \binom{365}{121}}{365^{121}} \approx 0 \\
P(\textit{at least 1 match}) &\approx 1
\end{aligned}
$$

76

Expected value

A 2012 Gallup survey suggests that 26.2% of Americans are obese.

Among a random sample of 100 Americans, how many would you expect to be obese?

- Easy enough, 100 x 0.262 = 26.2.
- Or more formally, $\mu = np$ = 100 x 0.262 = 26.2.
- But this doesn't mean in every random sample of 100 people exactly 26.2 will be obese. In fact, that's not even possible. In some samples this value will be less, and in others more. How much would we expect this value to vary?

77

Expected value and its variability

Mean and standard deviation of binomial distribution

$$\mu = np \qquad \sigma = \sqrt{np(1-p)}$$

Going back to the obesity rate:

$$\sigma = \sqrt{np(1-p)} = \sqrt{100 \times 0.262 \times 0.738} \approx 4.4$$

We would expect 26.2 out of 100 randomly sampled Americans to be obese, with a standard deviation of 4.4.

Note: Mean and standard deviation of a binomial might not always be whole numbers, and that is alright, these values represent what we would expect to see on average.

78

Unusual observations

Using the notion that *observations that are more than 2 standard deviations away from the mean are considered unusual* and the mean and the standard deviation we just computed, we can calculate a range for the plausible number of obese Americans in random samples of 100.

26.2 ± (2 x 4.4) → (17.4, 35.0)

79

Practice

An August 2012 Gallup poll suggests that 13% of Americans think home schooling provides an excellent education for children. Would a random sample of 1,000 Americans where only 100 share this opinion be considered unusual?

A. Yes B. No

	Excellent	Good	Only fair	Poor	Total excellent/ good
	%	%	%	%	%
Independent private school	31	47	13	2	78
Parochial or church-related schools	21	48	18	5	69
Charter schools	17	43	23	5	60
Home schooling	13	33	30	14	46
Public schools	5	32	42	19	37

Gallup, Aug. 9-12, 2012

http://www.gallup.com/poll/156974/private-schools-top-marks-educating-children.aspx

80

Practice

An August 2012 Gallup poll suggests that 13% of Americans think home schooling provides an excellent education for children. Would a random sample of 1,000 Americans where only 100 share this opinion be considered unusual?

A. Yes B. No

$$\mu = np = 1,000 \times 0.13 = 130$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{1,000 \times 0.13 \times 0.87} \approx 10.6$$

Method 1: *Range of usual observations:* $130 \pm 2 \times 10.6 = (108.8, 151.2)$
100 is outside this range, so would be considered unusual.

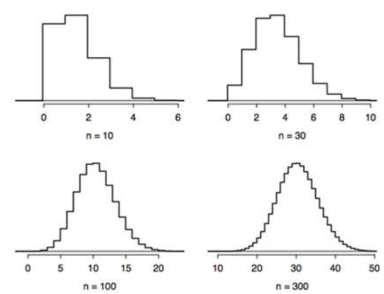
Method 2: *Z-score of observation:* $Z = \frac{x - mean}{SD} = \frac{100-130}{10.6} = -2.83$
100 is more than 2 SD below the mean, so would be considered unusual.

http://www.gallup.com/poll/156974/private-schools-top-marks-educating-children.aspx

81

Distributions of number of successes

Hollow histograms of samples from the binomial model where $p = 0.10$ and $n = 10, 30, 100$, and 300. What happens as n increases?



82

An analysis of Facebook users

A recent study found that "Facebook users get more than they give". For example:

1. 40% of Facebook users in our sample made a friend request, but 63% received at least one request
2. Users in our sample pressed the like button next to friends' content an average of 14 times, but had their content ``liked" an average of 20 times
3. Users sent 9 personal messages, but received 12
4. 12% of users tagged a friend in a photo, but 35% were themselves tagged in a photo

Any guesses for how this pattern can be explained?

Power users contribute much more content than the typical user.

http://www.pewinternet.org/Reports/2012/Facebook-users/Summary.aspx

83

Practice

This study also found that approximately 25% of Facebook users are considered power users. The same study found that the average Facebook user has 245 friends. What is the probability that the average Facebook user with 245 friends has 70 or more friends who would be considered power users? Note any assumptions you must make.

We are given that $n = 245$, $p = 0.25$, and we are asked for the probability $P(K \geq 70)$. To proceed, we need independence, which we'll assume but could check if we had access to more Facebook data.

$P(X \geq 70) = P(K = 70 \text{ or } K = 71 \text{ or } K = 72 \text{ or } \dots \text{ or } K = 245)$
$= P(K = 70) + P(K = 71) + P(K = 72) + \dots + P(K = 245)$

This seems like an awful lot of work...

84

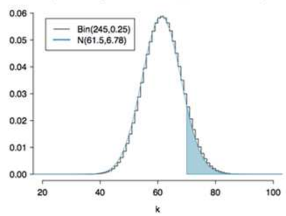
Normal approximation to the binomial

When the sample size is large enough, the binomial distribution with parameters n and p can be approximated by the normal model with parameters $\mu = np$ and $\sigma = \sqrt{np(1-p)}$.

- In the case of the Facebook power users, $n = 245$ and $p = 0.25$.

  $\mu = 245 \times 0.25 = 61.25 \qquad \sigma = \sqrt{245 \times 0.25 \times 0.75} = 6.78$

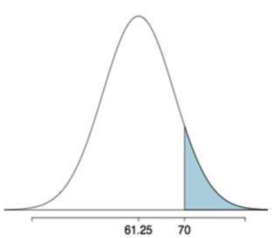
- $Bin(n = 245, p = 0.25) \approx N(\mu = 61.25, \sigma = 6.78)$.



85

Practice

What is the probability that the average Facebook user with 245 friends has 70 or more friends who would be considered power users?



$$Z = \frac{obs - mean}{SD} = \frac{70 - 61.25}{6.78} = 1.29$$

	Second decimal place of Z				
Z	0.05	0.06	0.07	0.08	0.09
1.0	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8944	0.8962	0.8980	0.8997	0.9015

$$P(Z > 1.29) = 1 - 0.9015 = 0.0985$$

86

Low large is large enough?

The sample size is considered large enough if the expected number of successes and failures are both at least 10.

$np \geq 10$ and $n(1 - p) \geq 10$

87

Practice

Below are four pairs of Binomial distribution parameters. Which distribution can be approximated by the normal distribution?

A. $n = 100, p = 0.95$
B. $n = 25, p = 0.45$
C. $n = 150, p = 0.05$
D. $n = 500, p = 0.015$

88

Practice

Below are four pairs of Binomial distribution parameters. Which distribution can be approximated by the normal distribution?

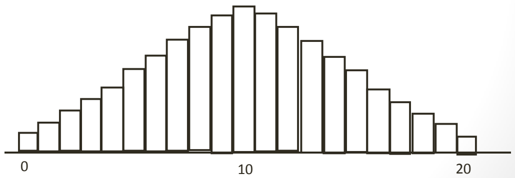
A. $n = 100, p = 0.95$
B. $n = 25, p = 0.45 \rightarrow 25 \times 0.45 = 11.25, 25 \times 0.55 = 13.75$
C. $n = 150, p = 0.05$
D. $n = 500, p = 0.015$

89

Experiment: tossing a coin 20 times

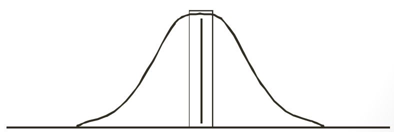
Problem: Find the probability of getting exactly 10 heads.

Distribution of the number of heads appearing should look like:



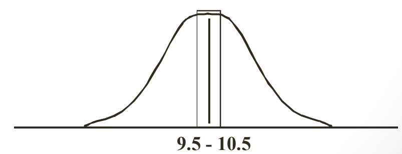
The Continuity Correction

Continuity Correction is needed because we are approximating a discrete probability distribution with a continuous distribution.



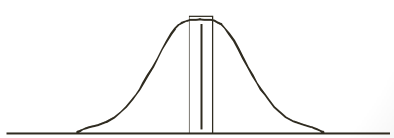
The Continuity Correction

We are using the area under the curve to approximate the area of the rectangle.



The Continuity Correction

<u>Continuity Correction</u>: to compute the probability of getting <u>exactly</u> 10 heads, find the probability of getting between 9.5 and 10.5 heads.



Using the Normal Distribution

$P(9.5 \leq x \leq 10.5) = ?$

for $x = 9.5$: $z = -0.22$

$P(z < -0.22) = .4129$

Using the Normal Distribution

for x = 10.5: z = = 0.22

P(z < .22) = .5871

P(9.5 $\leq$ x $\leq$ 10.5) = .5871 - .4129 = .1742

Application of Normal Distribution

If 22% of all patients with high blood pressure have side effects from a certain medication, and 100 patients are treated, find the probability that at least 30 of them will have side effects.

Using the Binomial Probability Formula we would need to compute:

$$P(30) + P(31) + \ldots + P(100) \text{ or } 1 - P(x \leq 29)$$

Using the Normal Approximation to the Binomial Distribution

Is it appropriate to use the normal distribution?

Check: $n\,p = 22$

$n\,q = 78$

Both are greater than 10

Find the mean and standard deviation

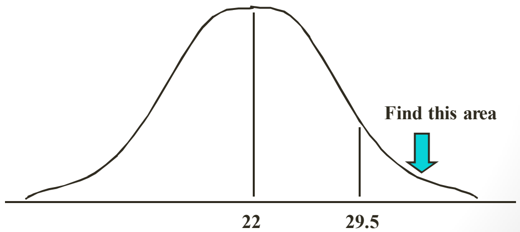
$$\mu = 100(.22) = 22$$

and $\sigma = \sqrt{100(.22)(.78)} =$

$$\sqrt{17.16} = 4.14$$

Applying the Normal Distribution

To find the probability that at least 30 of them will have side effects, find $P(x \geq 29.5)$

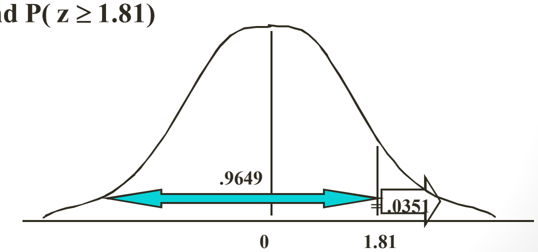


Applying the Normal Distribution

$$z = \frac{29.5 - 22}{4.14} = 1.81$$

Find $P(z \geq 1.81)$

The probability that at least 30 of the patients will have side effects is 0.0351.



Reminders:

Use the normal distribution to approximate the binomial only if both np and nq are greater than 10 (some texts say 5).

Always use the continuity correction when approximating the binomial distribution.

101