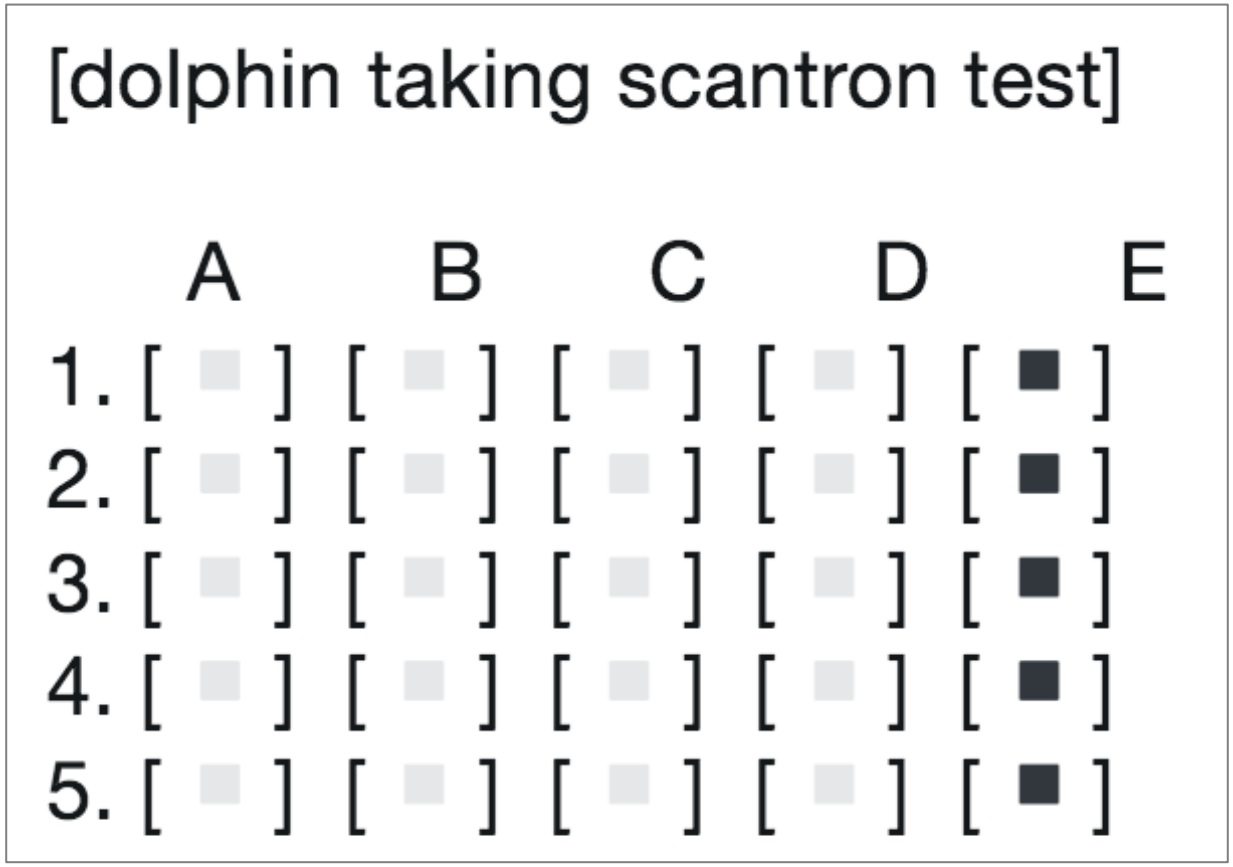
Learning Objectives

- **Review** inferential statistics, *z*-distribution, and models of randomness
- **Describe** 5 conditions required for using the *binomial distribution*
- **Visualize** probability in the binomial distribution
- **Solve** for expected probability using the binomial table
- **Describe and use** *normal approximation* (use of the *z*-distribution) to solve binomial problems

Another example: Multiple choice



5 Rules

1. Series of *N* trials

each question = separate observation, N = # questions

2. Only 2 outcomes

correct/incorrect

3. Outcomes are *mutually exclusive*

yes: either correct or incorrect

4. Outcomes are *independent*

somewhat, but not entirely; questions can relate to each other, clues, same person completing Qs

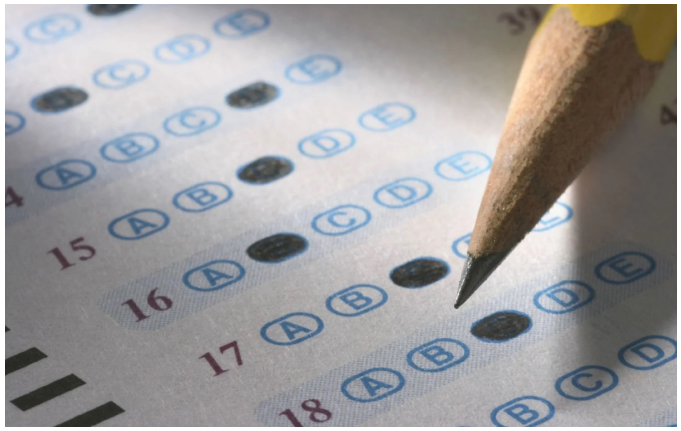
5. Probability of *P* remains consistent

Text

Does a multiple choice test meet these?

Not usually! We'll need to make a few assumptions

Will they pass?



Student needs at least 50% score

p(>4.5 correct) **no partial = must be ≥5**

– *N* = 9 questions

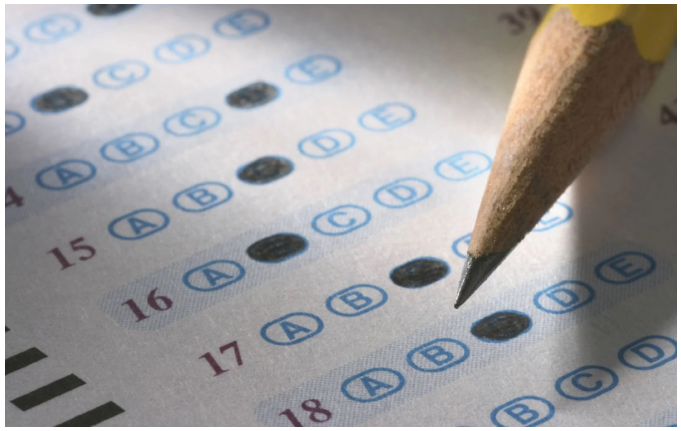
– *P*(correct) = .20 **randomly guessing**

- Choices a, b, c, d, & e

table B Binomial distribution—*cont'd*

N	No. of *P* or *Q* Events	*P* or *Q* .05	.10	.15	.20	.25	.30	.35	.40	.45	.50
9	0	.6302	.3874	.2316	.1342	.0751	.0404	.0277	.0101	.0046	.0020
	1	.2985	.3874	.3679	.3020	.2253	.1556	.1004	.0605	.0339	.0176
	2	.0629	.1722	.2597	.3020	.3003	.2668	.2162	.1612	.1110	.0703
	3	.0077	.0446	.1069	.1762	.2336	.2668	.2716	.2508	.2119	.1641
	4	.0006	.0074	.0283	.0661	.1168	.1715	.2194	.2508	.2600	.2461
	5	.0000	.0008	.0050	.0165	.0389	.0735	.1181	.1672	.2128	.2461
	6	.0000	.0001	.0006	.0028	.0087	.0210	.0424	.0743	.1160	.1641
	7	.0000	.0000	.0000	.0003	.0012	.0039	.0098	.0212	.0407	.0703
	8	.0000	.0000	.0000	.0000	.0001	.0004	.0013	.0035	.0083	.0176
	9	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0008	.0020

Will they pass?



Student needs at least 50% score

– $N = 9$, $P = .20$, $p(\geq 5)$

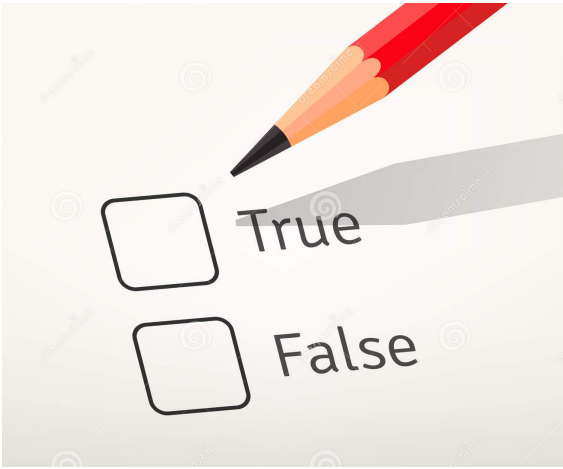
– Answer: $p(5) + p(6) + p(7) + p(8) + p(9) =$

$.0165 + .0028 + .0003 + .0000 + .0000 = .0196$

table B Binomial distribution—*cont'd*

N	No. of *P* or *Q* Events	*P* or *Q* .05	.10	.15	.20	.25	.30	.35	.40	.45	.50
9	0	.6302	.3874	.2316	.1342	.0751	.0404	.0277	.0101	.0046	.0020
	1	.2985	.3874	.3679	.3020	.2253	.1556	.1004	.0605	.0339	.0176
	2	.0629	.1722	.2597	.3020	.3003	.2668	.2162	.1612	.1110	.0703
	3	.0077	.0446	.1069	.1762	.2336	.2668	.2716	.2508	.2119	.1641
	4	.0006	.0074	.0283	.0661	.1168	.1715	.2194	.2508	.2600	.2461
	5	.0000	.0008	.0050	.0165	.0389	.0735	.1181	.1672	.2128	.2461
	6	.0000	.0001	.0006	.0028	.0087	.0210	.0424	.0743	.1160	.1641
	7	.0000	.0000	.0000	.0003	.0012	.0039	.0098	.0212	.0407	.0703
	8	.0000	.0000	.0000	.0000	.0001	.0004	.0013	.0035	.0083	.0176
	9	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0008	.0020

Will they pass?



- Student needs at least 50% score
  - *N* = 20 questions, *P*(T/F) = .50, *p*(≥10)
  - Answer: *p*(10) + … *p*(20) = .5881

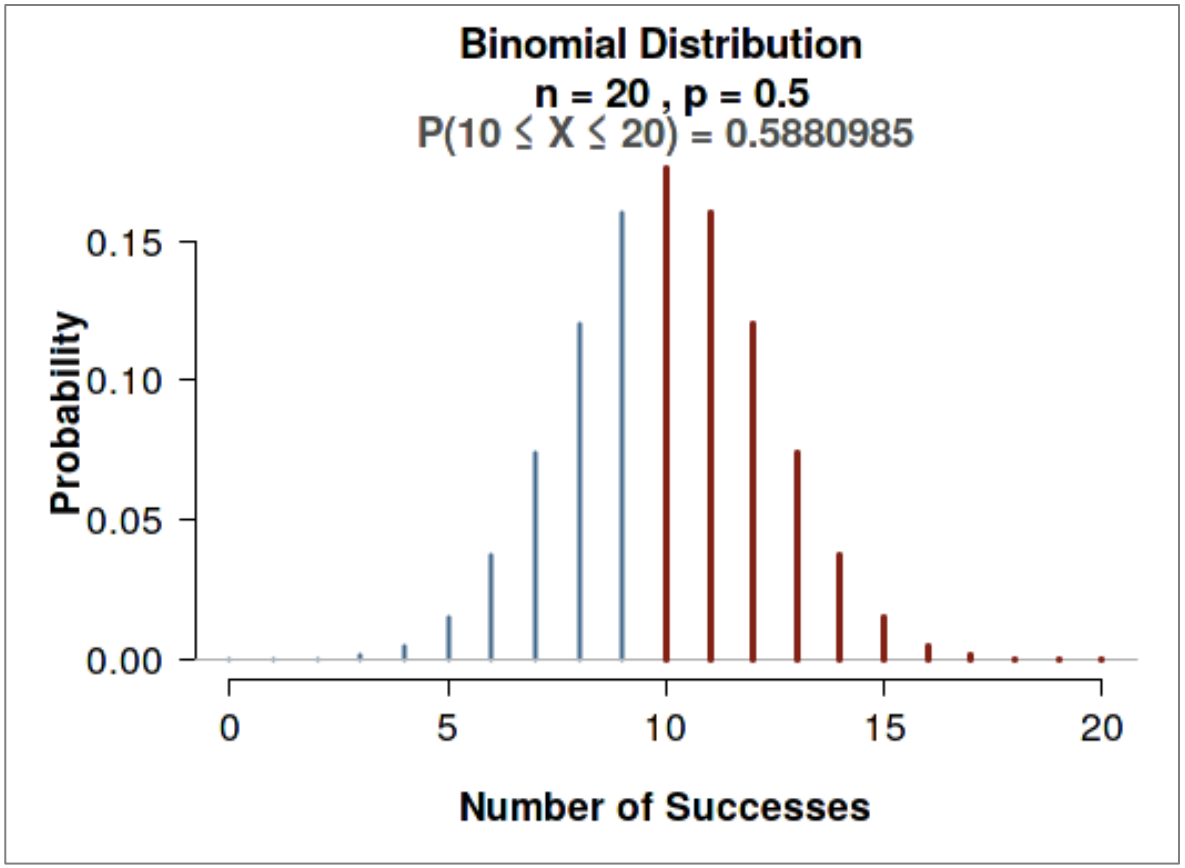


table B Binomial distribution—*cont'd*

N	No. of *P* or *Q* Events	*P* or *Q* .05	.10	.15	.20	.25	.30	.35	.40	.45	.50
20	0	.3585	.1216	.0388	.0115	.0032	.0008	.0002	.0000	.0000	.0000
	1	.3774	.2702	.1368	.0576	.0211	.0068	.0020	.0005	.0001	.0000
	2	.1887	.2852	.2293	.1369	.0669	.0278	.0100	.0031	.0008	.0002
	3	.0596	.1901	.2428	.2054	.1339	.0716	.0323	.0123	.0040	.0011
	4	.0133	.0898	.1821	.2182	.1897	.1304	.0738	.0350	.0139	.0046
	5	.0022	.0319	.1028	.1746	.2023	.1789	.1272	.0746	.0365	.0148
	6	.0003	.0089	.0454	.1091	.1686	.1916	.1712	.1244	.0746	.0370
	7	.0000	.0020	.0160	.0545	.1124	.1643	.1844	.1659	.1221	.0739
	8	.0000	.0004	.0046	.0222	.0609	.1144	.1614	.1797	.1623	.1201
	9	.0000	.0001	.0011	.0074	.0271	.0654	.1158	.1597	.1771	.1602
	10	.0000	.0000	.0002	.0020	.0099	.0308	.0686	.1171	.1593	.1762
	11	.0000	.0000	.0000	.0005	.0030	.0120	.0336	.0710	.1185	.1602
	12	.0000	.0000	.0000	.0001	.0008	.0039	.0136	.0355	.0727	.1201
	13	.0000	.0000	.0000	.0000	.0002	.0010	.0045	.0146	.0366	.0739
	14	.0000	.0000	.0000	.0000	.0000	.0002	.0012	.0049	.0150	.0370
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0003	.0013	.0049	.0148
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0003	.0013	.0046
	17	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0011
	18	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002
	19	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	20	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000

Should I be suspicious?

- Student scores 20% on midterm exam 1
- Student brags they have not studied at all for midterm exam 2

- Midterm Exam 2:
  - N = 20 questions
  - ***Student scores 90%!!!***

  - **Hypothesis 1**: Assuming they have not studied, student performs no better than exam 1
    - $P_{expectation}$ = .20
  - **Hypothesis 2**: The student has not studied, but cheats!
    - $P_{expectation}$ = higher than .20, but no clue how much higher

- We start by assuming hypothesis 1 (because it can be quantitatively modeled)
  - Then assess *P*robability of observed result (90%) or a result even less consistent with Hypothesis 1

Should I be suspicious?

- N = 20 questions
- $P_{expectation}$ = .20
  - Student, on average, should get 4 correct
- $P_{observed}$ = .90
  - Student got 18 correct
  - But, 19 and 20 correct would be even stronger evidence against Hypothesis 1
- Thus, compute $p(18)$ and add to $p(19)$ & $p(20)$

Referred to as "tail" of the distribution

table B Binomial distribution—*cont'd*

N	No. of *P* or *Q* Events	*P* or *Q* .05	.10	.15	.20	.25	.30	.35	.40	.45	.50
20	0	.3585	.1216	.0388	.0115	.0032	.0008	.0002	.0000	.0000	.0000
	1	.3774	.2702	.1368	.0576	.0211	.0068	.0020	.0005	.0001	.0000
	2	.1887	.2852	.2293	.1369	.0669	.0278	.0100	.0031	.0008	.0002
	3	.0596	.1901	.2428	.2054	.1339	.0716	.0323	.0123	.0040	.0011
	4	.0133	.0898	.1821	.2182	.1897	.1304	.0738	.0350	.0139	.0046
	5	.0022	.0319	.1028	.1746	.2023	.1789	.1272	.0746	.0365	.0148
	6	.0003	.0089	.0454	.1091	.1686	.1916	.1712	.1244	.0746	.0370
	7	.0000	.0020	.0160	.0545	.1124	.1643	.1844	.1659	.1221	.0739
	8	.0000	.0004	.0046	.0222	.0609	.1144	.1614	.1797	.1623	.1201
	9	.0000	.0001	.0011	.0074	.0271	.0654	.1158	.1597	.1771	.1602
	10	.0000	.0000	.0002	.0020	.0099	.0308	.0686	.1171	.1593	.1762
	11	.0000	.0000	.0000	.0005	.0030	.0120	.0336	.0710	.1185	.1602
	12	.0000	.0000	.0000	.0001	.0008	.0039	.0136	.0355	.0727	.1201
	13	.0000	.0000	.0000	.0000	.0002	.0010	.0045	.0146	.0366	.0739
	14	.0000	.0000	.0000	.0000	.0000	.0002	.0012	.0049	.0150	.0370
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0003	.0013	.0049	.0148
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0003	.0013	.0046
	17	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0011
	18	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002
	19	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	20	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000

Will they get an ‘A’?

- Student needs at least 80% score
  - *N* = 50 questions
  - *P* = .70
- Table doesn’t go that high?
  - We need the normal approximation!
- Remember, binomial distribution approximates *z*-distribution as *N* increases!…

Normal approximation

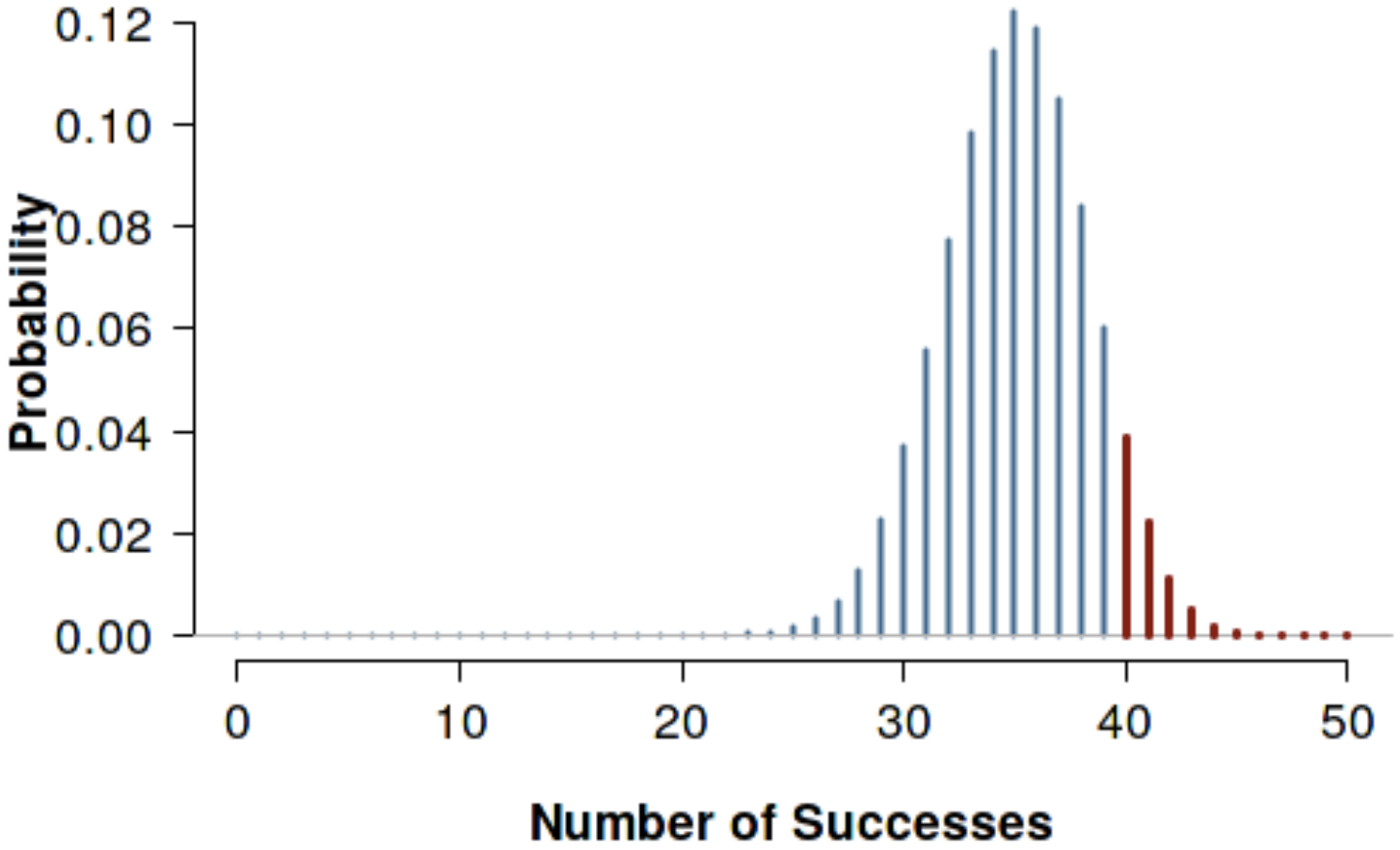
- Student needs at least 80% score
  - *N* = 50 questions
  - *P* = .70
  - What *z*-score corresponds to 80% score?

$$z = \frac{X - \mu}{\sigma}$$

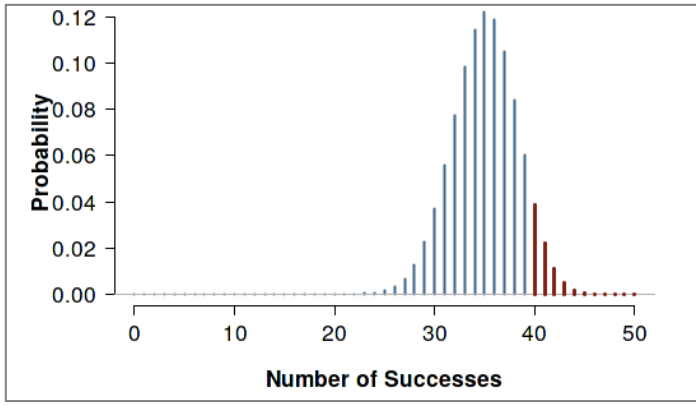
Where, $\mu = NP$

and, $\sigma = \sqrt{NPQ}$

Normal Approximation



Conceptual Steps in using Normal Approximation



1. Want to know $p(\geq 80\%)$
2. But N=50, tables won't work
3. Binomial approximates z-distribution
4. What z-score corresponds to 80%?
5. What is area under the curve for score that is $z_{80\%}$ or higher?