OpenIntro Statistics

Chapter 4 Exercises

Jai Jeffryes

4/11/2020 - 4/28/2020"

Notes

Make a table of mean and SD formulas for the distributions.

Distribution	Mean $= \mu$	Standard Deviation $= \sigma$
Normal	μ	σ
Bernoulli random variable	p	$\sqrt{p(1-p)}$
Geometric	$\frac{1}{p}$	$\sqrt{\frac{1-p}{p^2}}$
Binomial	np	$\sqrt{np(1-p} = \sqrt{npq}$
Approximate normal	np	$\sqrt{np(1-p} = \sqrt{npq}$
Poisson	λ	$\sqrt{\lambda}$

Tables	Are	Cool
col 1 is	left-aligned	\$1600
col 2 is	centered	\$12
col 3 is	right-aligned	\$1

Normal distribution

- Z-score. Number of standard deviations away from the mean. $Z = \frac{x-\mu}{\sigma}$.
- Finding tail areas.
  - Draw a picture.
  - Area under the curve corresponds to a cumulative distribution function. Same as a percentile. Use `pnorm()`.
- 68 - 95 - 99.7 Rule.

Geometric distribution

- Bernoulli random variable. Takes value 1 with probability p, value 0 with probability $1-p$.
- Geometric describes number of trials until a success.
- Probability of observing 1st success on n^{th} trial: $(1-p)^{n-1}p$
- `dgeom()`. 1st parameter is number of failures.
- The geometric distribution is right-skewed and never can be approximated by the normal distribution.

Binomial distribution

- Describes the number of successes in a fixed number of trials. Probability of having k successes in n independent Bernoulli trials with probability of success p.

- One scenario: $p^k(1-p)^{n-k}$
- Number of ways to arrange the scenarios: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$
- $\binom{n}{k}p^k(1-p)^{n-k} = \frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$
- Four conditions. Really check this 1st one.

  1. The trials are independent.
  2. The number of trials, n, is fixed.
  3. Each trial outcome can be classified as a *success* or *failure*.
  4. The probability of success, p, is the same in each trial.

- Computing steps.

  1. Check conditions.
  2. Identify n, p, and k.
  3. Use software or the formulas to determine the probability.
  4. Interpret results.

Normal approximation to the binomial distribution

- Conditions:

  1. $np > 10$
  2. $n(1-p) > 10$

- Binomial is cumbersome when n is large.
- Why bother when you have `pbinom()` in `R`? From Dr. Fulton.

  When N is super large, then the number of combinations available exceeds the ability of computers to process it. This was more of an issue in the days of old.

  Try, for example, `=COMBIN(10000000,200)` In Excel. That combination is too large to calculate.

- The normal approximation may be used when computing the range of many possible successes.
- I should use the approach of finding the Z-score first. That would help me get over the agony of the boundary point.
- Continuity correction. Left side of interval reduce by 0.5. Right side, increase by 0.5. Apply *before* the Z-score conversion.

Negative binomial distribution

- While the geometric distribution describes the probability of observing the first success on the n^{th} trial, the negative binomial distribution is more general, describing the probability of observing the kth success on the n^{th} trial.
- Four conditions. First 3 are common to the binomial.

  1. The trials are independent.
  2. Each trial outcome can be classified as a *success* or *failure*.
  3. The probability of success, p, is the same in each trial.
  4. The last trial must be a success.

- Compute. First, break into pieces.

  - Number of possible sequences times probability of a single sequence.
  - $P(Single\ sequence = P(n-k) failures\ and\ k\ successes = (1-p)^{n-k}p^k$
  - Fix last trial as a success. What are the number of combinations for the other observations? $\binom{n-1}{k-1} = \frac{(n-1)!}{(k-1)!(n-k)!}$

- All together: $P(\textit{the } k^{th} \textit{ success on the } n^{th} \textit{ trial}) = \binom{n-1}{k-1}(1-p)^{n-k}p^k = \frac{(n-1)!}{(k-1)!(n-k)!}(1-p)^{n-k}p^k$

- Binomial: Fixed trials and consider number of successes. Negative binomial: Fix successes and consider how many trials to make last observation a success.

Poisson distribution

- Estimates number of events in a population over time.
- Rate = λ, the average number (expected number) of occurrences in a unit of time.
- $P(\textit{observe } k \textit{ events}) = \frac{\lambda^k e^{-k}}{k!}$
- Modeling rates for a Poisson distribution against a second variable, like day of the week, falls under **general linear models**, discussed later.

Supplemental exercises

Binary distribution

- The clearest illustration I've found for comparing the binary distribution and normal approximation of it is CUNY CSI MATH 113 - Lab 5.
- Had a lot of help from this video and
- This Binomial Probability Calculator and
- Cillian McNeill's notes
- A Guide to dbinom, pbinom, qbinom, and rbinom in R

Summary function

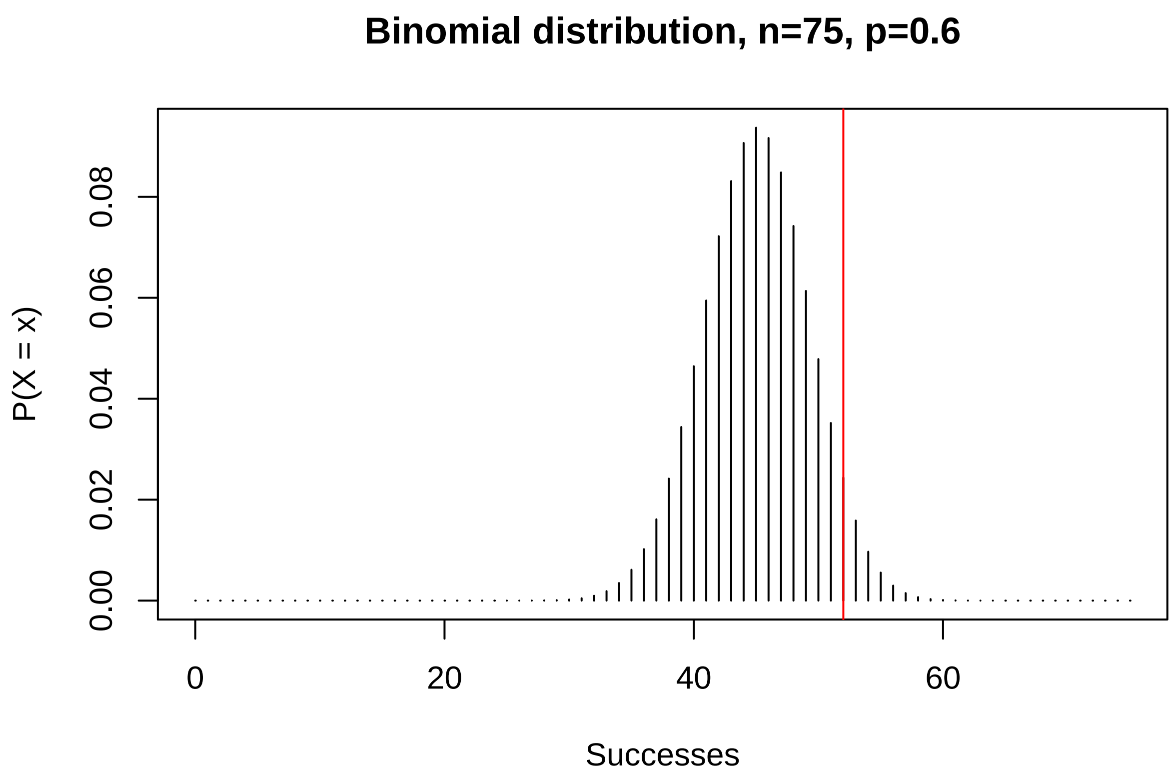
I wrote this function to plot the distribution and compute all the intervals, both exact and approximation to the normal, related to a given number of trials and successes.

- Plots the distribution.
- Computes each interval.
- If n and p satisfy conditions, it computes approximations with and without continuity correction.
- Validated with Binomial Probability Calculator and video

```
binom_summary <- function(successes, trials, prob,
                          main = NULL, xlab = "Successes", ylab = "P(X = x)") {
  # Annotate plot.
  param_str <- paste0(", n=", trials, ", p=", prob)
  main <- ifelse(is.null(main), paste0("Binomial distribution", param_str),
                 paste0(main, param_str))

  plot(0:trials, dbinom(0:trials, trials, prob = prob),
     xlab = xlab,
     ylab = ylab,
     type = "h",
     main = main)
  abline(v = successes, col = "red")

  mu <- trials * prob
  sigma <- sqrt(trials * prob * (1 - prob))
  is_normal_approx <- trials * prob >= 10 & trials * (1 - prob) >= 10
```

```r
print(paste0("mu: ", mu))
print(paste0("sigma: ", sigma))

print(paste0("Trials: ", trials))
print(paste0("Successes: ", successes))

prob.exact <- dbinom(successes, trials, prob)

prob.cumulative.lt <- pbinom(successes - 1, trials, prob)
# Normal approximation
prob.cumulative.lt.approx <- pnorm(successes, mu, sigma)
# Normal approximation with continuity correction.
prob.cumulative.lt.approx.corr <- pnorm(successes - 0.5, mu, sigma)

prob.cumulative.le <- pbinom(successes, trials, prob)
# Normal approximation
prob.cumulative.le.approx <- pnorm(successes, mu, sigma)
# Normal approximation with continuity correction.
prob.cumulative.le.approx.corr <- pnorm(successes + 0.5, mu, sigma)

prob.cumulative.gt <- pbinom(successes, trials, prob, lower.tail = F)
# Normal approximation
prob.cumulative.gt.approx <- pnorm(successes, mu, sigma, lower.tail = F)
# Normal approximation with continuity correction.
prob.cumulative.gt.approx.corr <- pnorm(successes + 0.5, mu, sigma, lower.tail = F)

prob.cumulative.ge <- pbinom(successes - 1, trials, prob, lower.tail = F)
# Normal approximation
prob.cumulative.ge.approx <- pnorm(successes, mu, sigma, lower.tail = F)
# Normal approximation with continuity correction.
prob.cumulative.ge.approx.corr <- pnorm(successes - 0.5, mu, sigma, lower.tail = F)

print(paste0("Binomial probability. P(X=x): ", round(prob.exact, 4)))

print(paste0("Cumulative probability. P(X<x): ", round(prob.cumulative.lt, 4)))
if (is_normal_approx) {
  print(paste0("    Normal approximation. P(X<x): ",
               round(prob.cumulative.lt.approx, 4)))
  print(paste0("    Normal approximation w/ continuity correction. P(<x): ",
               round(prob.cumulative.lt.approx.corr, 4)))
}

print(paste0("Cumulative probability. P(X<=x): ", round(prob.cumulative.le, 4)))
if (is_normal_approx) {
  print(paste0("    Normal approximation. P(X<=x): ",
               round(prob.cumulative.le.approx, 4)))
  print(paste0("    Normal approximation w/ continuity correction. P(<=x): ",
               round(prob.cumulative.le.approx.corr, 4)))
}

print(paste0("Cumulative probability. P(X>x): ", round(prob.cumulative.gt, 4)))
if (is_normal_approx) {
  print(paste0("    Normal approximation. P(X>x): ",
```

```
                  round(prob.cumulative.gt.approx, 4)))
    print(paste0("    Normal approximation w/ continuity correction. P(X>x): ",
                  round(prob.cumulative.gt.approx.corr, 4)))
  }

    print(paste0("Cumulative probability. P(X>=x): ", round(prob.cumulative.ge, 4)))
  if (is_normal_approx) {
    print(paste0("    Normal approximation. P(X>=x): ",
                  round(prob.cumulative.ge.approx, 4)))
    print(paste0("    Normal approximation w/ continuity correction. P(X>=x): ",
                  round(prob.cumulative.ge.approx.corr, 4)))
  }
}

binom_summary(52, 75, 0.60)
```




```
## [1] "mu: 45"
## [1] "sigma: 4.24264068711928"
## [1] "Trials: 75"
## [1] "Successes: 52"
## [1] "Binomial probability. P(X=x): 0.0244"
## [1] "Cumulative probability. P(X<x): 0.9389"
## [1] "    Normal approximation. P(X<x): 0.9505"
## [1] "    Normal approximation w/ continuity correction. P(<x): 0.9372"
## [1] "Cumulative probability. P(X<=x): 0.9633"
```

```
## [1] "    Normal approximation. P(X<=x): 0.9505"
## [1] "    Normal approximation w/ continuity correction. P(<=x): 0.9615"
## [1] "Cumulative probability. P(X>x): 0.0367"
## [1] "    Normal approximation. P(X>x): 0.0495"
## [1] "    Normal approximation w/ continuity correction. P(X>x): 0.0385"
## [1] "Cumulative probability. P(X>=x): 0.0611"
## [1] "    Normal approximation. P(X>=x): 0.0495"
## [1] "    Normal approximation w/ continuity correction. P(X>=x): 0.0628"
```

Cillian McNeill's notes

Example 1

Five terminals on an on-line computer system are at- tached to a communication line to the central computer system. The probability that any terminal is ready to transmit is 0.95.

Let X denote the number of ready terminals.

Example 2

A coin is tossed 10 times; success and failure are "heads" and "tails" respectively, each with probability, .5.

Let X be the number of heads (successes) obtained.

Example 3

It is known that 20% of integrated circuit chips on a production line are defective. To maintain and monitor the quality of the chips, a sample of twenty chips is selected at regular intervals for inspection.

Let X denote the number of defectives found in the sample.

Example 4

It is known that 1% of bits transmitted through a digital transmission are received in error. One hundred bits are transmitted each day.

Let X denote the number of bits found in error each day.

Calculating PDFS with `R`

Example 1

The probability of getting exactly 3 ready terminals in 5.

```
dbinom(x = 3, size =5, prob = .95)
```

```
## [1] 0.02143438
```

For all the probabilities.

```
x <- 0:5
dbinom(x, size= 5, prob =.95)
```

```
## [1] 0.0000003125 0.0000296875 0.0011281250 0.0214343750 0.2036265625
## [6] 0.7737809375
```

Example 2

Tossing a coin 10 times

```
x <- 0:10
round(dbinom(x, 10, .5), 4)
```

```
##  [1] 0.0010 0.0098 0.0439 0.1172 0.2051 0.2461 0.2051 0.1172 0.0439 0.0098
## [11] 0.0010
```

Interesting that the probability of getting 0 successes in 10 tosses is 1 in a thousand.

Example 3

```
options(scipen = 4)
x <- 0:20
dbinom(x, 20, 0.20)
```

```
##  [1] 1.152922e-02 5.764608e-02 1.369094e-01 2.053641e-01 2.181994e-01
##  [6] 1.745595e-01 1.090997e-01 5.454985e-02 2.216088e-02 7.386959e-03
## [11] 2.031414e-03 4.616849e-04 8.656592e-05 1.331783e-05 1.664729e-06
## [16] 1.664729e-07 1.300570e-08 7.650410e-10 3.187671e-11 8.388608e-13
## [21] 1.048576e-14
```

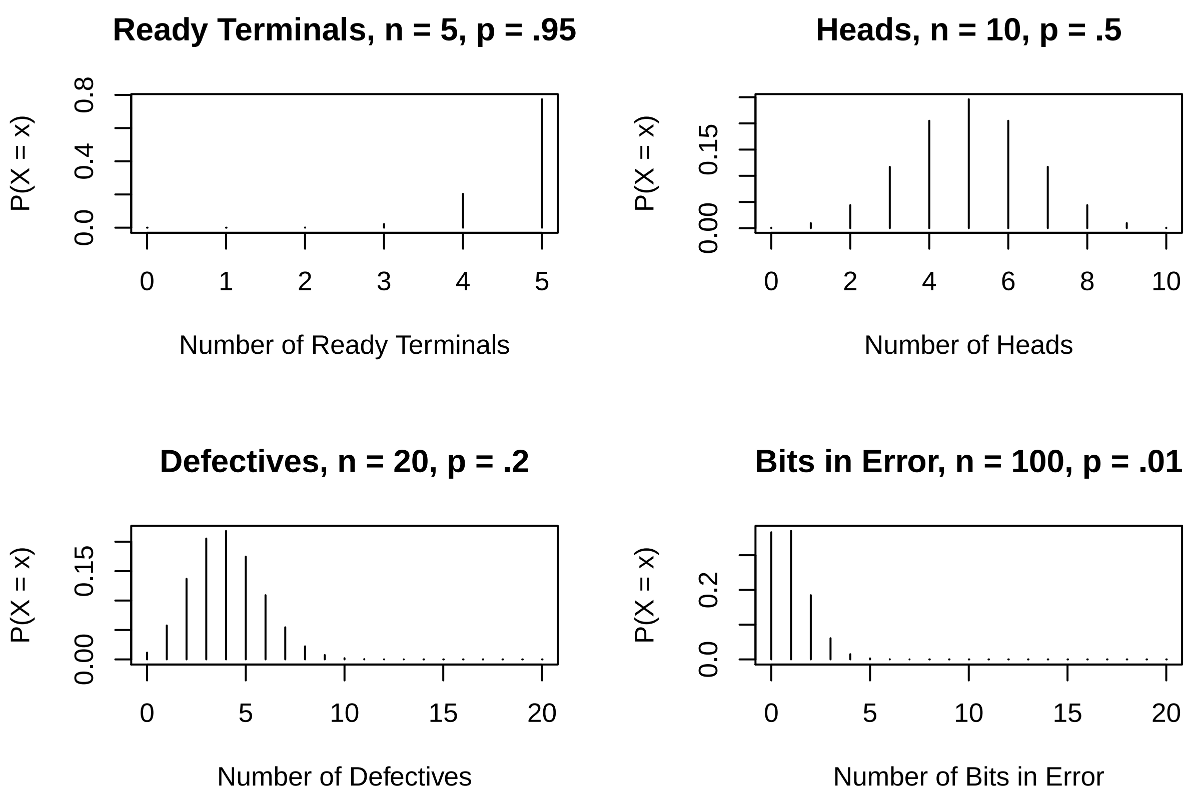
Example 4

```
x <- 0:100
dbinom(x, 100, 0.01)
```

```
##   [1]  3.660323e-01  3.697296e-01  1.848648e-01  6.099917e-02  1.494171e-02
##   [6]  2.897787e-03  4.634508e-04  6.286346e-05  7.381694e-06  7.621951e-07
##  [11]  7.006036e-08  5.790112e-09  4.337710e-10  2.965956e-11  1.861747e-12
##  [16]  1.078184e-13  5.785707e-15  2.887697e-16  1.344999e-17  5.863367e-19
##  [21]  2.398650e-20  9.230014e-22  3.347893e-23  1.146841e-24  3.716614e-26
##  [26]  1.141263e-27  3.325359e-29  9.206008e-31  2.424382e-32  6.079954e-34
##  [31]  1.453457e-35  3.315151e-37  7.220499e-39  1.502889e-40  2.991491e-42
##  [36]  5.698078e-44  1.039212e-45  1.815713e-47  3.040667e-49  4.882708e-51
##  [41]  7.521343e-53  1.111802e-54  1.577594e-56  2.149411e-58  2.812590e-60
##  [46]  3.535467e-62  4.269888e-64  4.955382e-66  5.526836e-68  5.924459e-70
##  [51]  6.103988e-72  6.044749e-74  5.753549e-76  5.263395e-78  4.627377e-80
##  [56]  3.909263e-82  3.173103e-84  2.474154e-86  1.852815e-88  1.332276e-90
##  [61]  9.195842e-93  6.090970e-95  3.870118e-97  2.357936e-99 1.376951e-101
##  [66] 7.703225e-104 4.126306e-106 2.115098e-108 1.036813e-110 4.856976e-113
##  [71] 2.172673e-115 9.273039e-118 3.772701e-120 1.461680e-122 5.387028e-125
##  [76] 1.886367e-127 6.267832e-130 1.973343e-132 5.877609e-135 1.653336e-137
##  [81] 4.383845e-140 1.093365e-142 2.558996e-145 5.605686e-148 1.145943e-150
##  [86] 2.178859e-153 3.838722e-156 6.239651e-159 9.310773e-162 1.268066e-164
##  [91] 1.565513e-167 1.737722e-170 1.717116e-173 1.492009e-176 1.122294e-179
##  [96] 7.159768e-183 3.766713e-186 1.568973e-189 4.851495e-193 9.900000e-197
## [101] 1.000000e-200
```

Plotting binomial PDFs

```
par(mfrow = c(2,2)) # multiframe

x<-0:5 #Example 11.1
plot(x, dbinom(x, size = 5, prob = .95),
xlab = "Number of Ready Terminals",
ylab = "P(X = x)", type = "h",
main = "Ready Terminals, n = 5, p = .95")

x<-0:10 #Example 11.2
plot(x, dbinom(x, size = 10, prob = .5),
xlab = "Number of Heads",
ylab = "P(X = x)", type = "h", main = "Heads, n = 10, p = .5")

x<-0:20 #Example 11.3
plot(x, dbinom(x, size = 20, prob = .2),
xlab = "Number of Defectives",
ylab = "P(X = x)", type = "h",
main = "Defectives, n = 20, p = .2")

x<-0:20 #Example 11.4. No need for full range here
plot(x, dbinom(x, size = 100, prob = .01),
                xlab = "Number of Bits in Error",
                ylab = "P(X = x)", type = "h",
                main = "Bits in Error, n = 100, p = .01")
```



Cumulative distribution function

Example 1

$P(less\ than\ or\ equal\ to\ 3\ terminals\ will\ be\ ready)$

$= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$

$= (.05)^5 + \binom{5}{1}(.95)^1(.05)^4 4 + \binom{5}{2}(.95)^2(.05)^3 + \binom{5}{3}(.95)^3(.05)^1$

$= .0000003 + .00003 + .00113 + .02143$

$\approx .0226$

Also

$= P(4\ or\ more\ terminals\ will\ be\ ready)$

$= P(X \geq 4)$

$= 1 - P(X \leq 3)$

$\approx 1 - .0226$

$= .9774$

Calculating CDFS with `R`

Example 1

$n = 5$, $p = .95$

$P(X \leq 3)$

```
pbinom(3, 5, .95)
```

```
## [1] 0.0225925
```

Example 2

$P(X > 3)$

```
1 - pbinom(3, 5, .95)
```

```
## [1] 0.9774075
```

```
pbinom(3, 5, .95, lower.tail = F)
```

```
## [1] 0.9774075
```

Example 3

$n = 20$, $p = .2$

$P(X \leq 4)$

```r
pbinom (4, size = 20, prob =.2)
```

```
## [1] 0.6296483
```

Example 4

$P(X > 4) = 1 - P(X \leq 4)$

```r
1 - pbinom(4, size = 20, prob = .2)
```

```
## [1] 0.3703517
```

```r
pbinom(4, size = 20, prob = .2, lower.tail = F)
```

```
## [1] 0.3703517
```

Example 5

For all cumulative probabilities

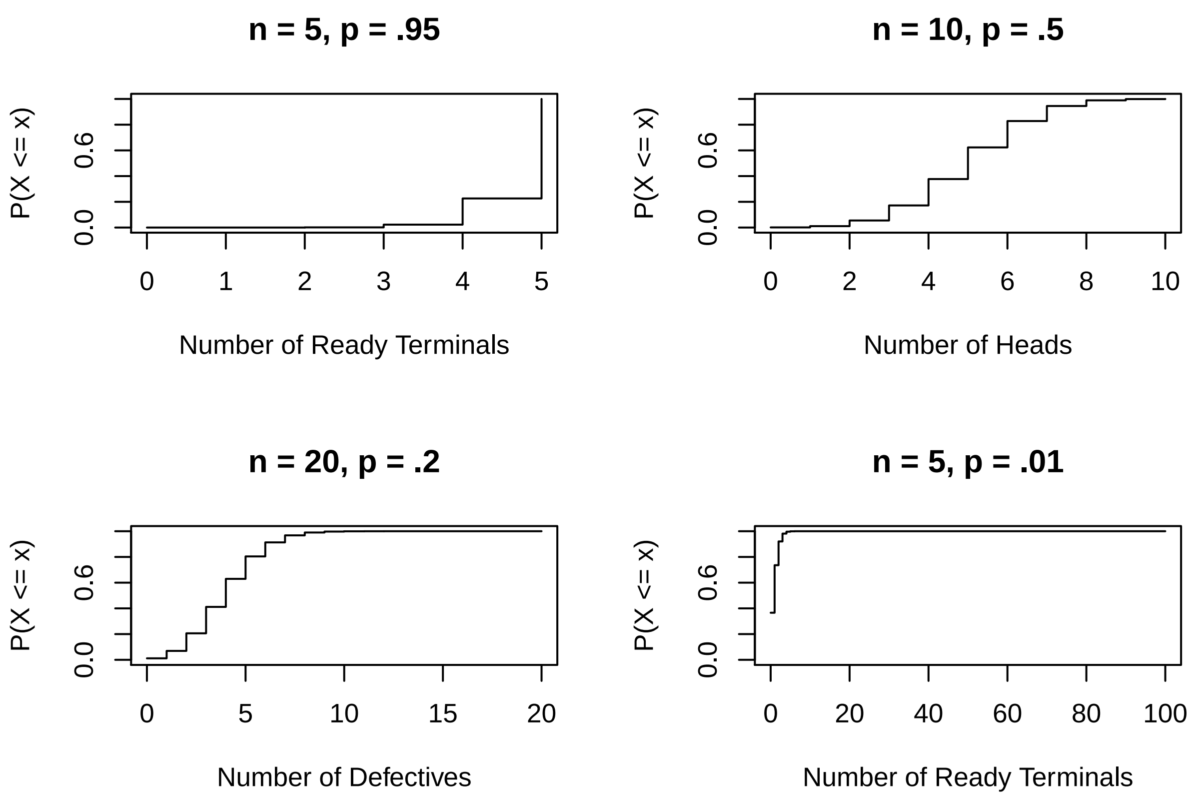
```r
x <- 0:20
prob <- pbinom(x, size = 20, prob= .2)

# Round to 4 decimal places.
round(prob, 4)
```

```
##  [1] 0.0115 0.0692 0.2061 0.4114 0.6296 0.8042 0.9133 0.9679 0.9900 0.9974
## [11] 0.9994 0.9999 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000
## [21] 1.0000
```

Plotting Binomial CDFs

```r
par(mfrow = c(2,2)) #multiframe
x = 0:5
plot(x, pbinom(x, size = 5, prob = .95),
     xlab = "Number of Ready Terminals",
     ylab = "P(X <= x)", ylim = c(0, 1),
     type = "s", main = "n = 5, p = .95")

x = 0:10
plot(x, pbinom(x, size = 10, prob = .5),
     xlab = "Number of Heads",
     ylab = "P(X <= x)", ylim = c(0, 1),
     type = "s", main = "n = 10, p = .5")

x = 0:20
plot(x, pbinom(x, size = 20, prob = .2),
     xlab = "Number of Defectives",
     ylab = "P(X <= x)", ylim = c(0, 1),
     type = "s", main = "n = 20, p = .2")
```

```r
x = 0:100
plot(x, pbinom(x, size = 100, prob = .01),
     xlab = "Number of Ready Terminals",
     ylab = "P(X <= x)", ylim = c(0, 1),
     type = "s", main = "n = 5, p = .01")
```



Quantile function

To be completed.

Binomial distribution

p. 142, ex. 4.1

What percent of a standard normal distribution $N(\mu = 0, \sigma = 1)$ is found in each region? Be sure to draw a graph.

I flailed unnecessarilyt at this. I computed a Z-score and wondered why it didn't match my `pnorm()` result. No, no. Z-score is not area. It's a standardized distance from the mean. Z-score in hand, input **that** into `pnorm()` to get the area bounded by the Z-score. However, here we're given a Standard Normal Distribution. We don't need to standardize the distribution, it's already standard.

a. 8.8507991%
b. 6.9436623%

c. 58.861454%
d. 4.5500264%

p. 142, ex. 4.2

What percent of a standard normal distribution $N(\mu = 0, \sigma = 1)$ is found in each region? Be sure to draw a graph.

a. 87.0761888%
b. 57.1423716%
c. $6.2209606 \times 10^{-14}$%
d. -60.7075077%

p. 142, ex. 4.3

Sophia who took the Graduate Record Examination (GRE) scored 160 on the Verbal Reasoning section and 157 on the Quantitative Reasoning section. The mean score for Verbal Reasoning section for all test takers was 151 with a standard deviation of 7, and the mean score for the Quantitative Reasoning was 153 with a standard deviation of 7.67. Suppose that both distributions are nearly normal.

- Write down the short-hand for these two normal distributions.
  - **Verbal reasoning:** $\mathcal{N}(\mu = 151, \sigma = 7)$
  - **Quantitative reasoning:** $\mathcal{N}(\mu = 153, \sigma = 7.67$
- What is Sophia's Z-score on the Verbal Reasoning section? On the Quantitative Reasoning section? Draw a standard normal distribution curve and mark these two Z-scores.
  - **Verbal reasoning:** Z_{VR} **= 1.2857143**
  - **Quantitative reasoning:** Z_{QR} **= 0.5215124**
- What do these Z-scores tell you? **The distance of her performance from the mean measured in standard deviation.**
- Relative to others, which section did she do better on? **Verbal reasoning.**
- Find her percentile scores for the two exams. **I want the lower tail area under the probability distribution bounded by her scores.**
  - **Verbal reasoning:** $Perc_{VR}$ **= 0.9007286**
  - **Quantitative reasoning:** $Perc_{QR}$ **= 0.6989951**
- What percent of the test takers did better than her on the Verbal Reasoning section? On the Quantitative Reasoning section?
  - **Verbal reasoning: 9.9271397%**
  - **Quantitative reasoning: 30.100494%**
- Explain why simply comparing raw scores from the two sections could lead to an incorrect conclusion as to which section a student did better on. **The two distributions differ in their mean and standard deviation.**
- If the distributions of the scores on these exams are not nearly normal, would your answers to parts (b) - (f) change? Explain your reasoning. **Yes, I would lack a way to standardize the distance from the mean.**

p. 142, ex. 4.4

In triathlons, it is common for racers to be placed into age and gender groups. Friends Leo and Mary both completed the Hermosa Beach Triathlon, where Leo competed in the *Men, Ages 30 - 34* group while Mary competed in the *Women, Ages 25 - 29* group. Leo completed the race in 1:22:28 (4948 seconds), while Mary completed the race in 1:31:53 (5513 seconds). Obviously Leo finished faster, but they are curious about how they did within their respective groups. Can you help them? Here is some information on the performance of their groups:

- The finishing times of the *Men, Ages 30 - 34* group has a mean of 4313 seconds with a standard deviation of 583 seconds.
- The finishing times of the *Women, Ages 25 - 29* group has a mean of 5261 seconds with a standard deviation of 807 seconds.
- The distributions of finishing times for both groups are approximately Normal.

Remember: a better performance corresponds to a faster finish.

- Write down the short-hand for these two normal distributions.
  - $\mathcal{N}(\mu = 4313, \sigma = 583)$
  - $\mathcal{N}(\mu = 5261, \sigma = 807)$
- What are the Z-scores for Leo's and Mary's finishing times? What do these Z-scores tell you?
  - Z_{Leo} **= 1.0891938**
  - Z_{Mary} **= 0.3122677**
  - **The distance of their performance from the mean measured in standard deviations.**
- Did Leo or Mary rank better in their respective groups? Explain your reasoning. **Mary did better than Leo. Her Z-score was lower. This Z-score measures duration of race. A lower Z-score corresponds to a better performance because it is a faster finish.**
- What percent of the triathletes did Leo finish faster than in his group?
  - **13.8034211%**
- What percent of the triathletes did Mary finish faster than in her group?
  - **37.7418559%**
- If the distributions of finishing times are not nearly normal, would your answers to parts (b)_(e) change? Explain your reasoning. **Yes, I would lack a way to standardize the distance from the mean.**

p. 142, ex. 4.5

In Exercise 4.3 we saw two distributions for GRE scores: $N(\mu = 151, \sigma = 7)$ for the verbal part of the exam and $N(\mu = 153, \sigma = 7.67)$ for the quantitative part. Use this information to compute each of the following:

Note: 1. get the Z-score with `qnorm()`. 2. Use population parameters and solve for x.

- The score of a student who scored in the 80^{th} percentile on the Quantitative Reasoning section.
  - **Z = 0.8416212**
  - $x = Z\sigma + \mu$
  - **Score = 159.4552349**
- The score of a student who scored worse than 70% of the test takers in the Verbal Reasoning section.
  - **Score = 148.9778481**

p. 143, ex. 4.6

In Exercise 4.4 we saw two distributions for triathlon times: N(μ=4313, σ=583) for *Men, Ages 30 - 34* and N($\mu = 5261$, σ=807) for the *Women, Ages 25 - 29* group. Times are listed in seconds. Use this information to compute each of the following:

- The cutoff time for the fastest 5% of athletes in the men's group, i.e. those who took the shortest 5% of time to finish.
  - **3354 seconds**
- The cutoff time for the slowest 10% of athletes in the women's group.
  - **6588 seconds**

p. 143, ex. 4.7

The average daily high temperature in June in LA is 77 degrees F with a standard deviation of 5 degrees F. Suppose that the temperatures in June closely follow a normal distribution.

- What is the probability of observing an 83 degrees F temperature or higher in LA during a randomly chosen day in June? **0.1150697**
- How cool are the coldest 10% of the days (days with lowest average high temperature) during June in LA? **71 degrees F.**

p. 143, ex. 4.8

The Capital Asset Pricing Model (CAPM) is a financial model that assumes returns on a portfolio are normally distributed. Suppose a portfolio has an average annual return of 14.7% (i.e. an average gain of 14.7%) with a standard deviation of 33%. A return of 0% means the value of the portfolio doesn't change, a negative return means that the portfolio loses money, and a positive return means that the portfolio gains money.

"Less than," "highest range," these require a cumulative distribution function, i.e. `pnorm()` and `qnorm()`.

- What percent of years does this portfolio lose money, i.e. have a return less than 0%? **32.8%**
- What is the cutoff for the highest 15% of annual returns with this portfolio? **Z-score times σ + μ. Gains of 49%.**

p. 143, ex. 4.9

Exercise 4.7 states that average daily high temperature in June in LA is 77 degrees F with a standard deviation of 5 degrees F, and it can be assumed that they to follow a normal distribution. We use the following equation to convert degrees F (Fahrenheit) to degrees C (Celsius):

$$C = (F - 32) \times \frac{5}{9}.$$

- Write the probability model for the distribution of temperature in degrees C in June in LA.

```r
celsius <- function(fahrenheit) {
  c <- (fahrenheit - 32) * (5 / 9)
  return(c)
}
mu.f <-77
sigma.f <- 5

mu.c <- celsius(mu.f)
## Nuh-uh.
## sigma.c <- celsius(sigma.f)
sigma.c <- 5 * (5 / 9)
```

- $\mathcal{N}(\mu = 25, \sigma = 2.7777778)$
- What is the probability of observing a 28 degrees C (which roughly corresponds to 83degrees F) temperature or higher in June in LA? Calculate using the degrees C model from part (a). **0.1400711**
- Did you get the same answer or different answers in part (b) of this question and part (a) of Exercise 4.7? Are you surprised? Explain. **At first, my answers differed greatly. My error was in the conversion of standard deviation from Fahrenheit to Celsius. I converted the number as if it were a temperature of 5 degrees, instead of a measure of interval. I fixed that and got a lot closer, but they still differ somewhat which surprises me because I thought I had all the units converted correctly and expected the cumulative distribution to be the same. Answer: I matched the book precisely. The answer key revealed that 28 degrees Celsius is not exactly 83 degrees Fahrenheit, it's 82.4. That's why I'm off. The problem set up lied!**
- Estimate the IQR of the temperatures (in degrees C) in June in LA.

```r
temp.25percent <- qnorm(0.25) * sigma.c + mu.c
temp.75percent <- qnorm(0.75) * sigma.c + mu.c
iqr.temp <- round(temp.75percent - temp.25percent, 2)
```

- **3.75 degrees Celsius.**

p. 143, ex. 4.10

Cholesterol levels for women aged 20 to 34 follow an approximately normal distribution with mean 185 milligrams per deciliter (mg/dl). Women with cholesterol levels above 220 mg/dl are considered to have high cholesterol and about 18.5% of women fall into this category. What is the standard deviation of the distribution of cholesterol levels for women aged 20 to~34? $\frac{x-\mu}{Z} = \sigma = \mathbf{39.0418739}$

p. 148, ex. 4.11

Determine if each trial can be considered an independent Bernoulli trial for the following situations.

- Cards dealt in a hand of poker. **No.**
- Outcome of each roll of a die. **No.**

Requirements:

- **Random**
- **Independent (I neglected to include this one)**

- **Two outcomes**

Answer: My answers are correct, but the book included a few more points about why, which I failed to notice.

- **Card dealing is not independent because it is sampling without replacement.**
- **Rolling dice could be Bernoulli if the events were simplified, for example {6, not 6}, but that would have to be specified.**

p. , ex. 4.12

In the following situations assume that half of the specified population is male and the other half is female.

- Suppose you're sampling from a room with 10 people. What is the probability of sampling two females in a row when sampling with replacement? What is the probability when sampling without replacement? $\frac{5}{10} \cdot \frac{4}{9} = \mathbf{0.2222222}$
- Now suppose you're sampling from a stadium with 10,000 people. What is the probability of sampling two females in a row when sampling with replacement? $\frac{5000}{10000} \cdot \frac{5000}{10000} = \mathbf{0.25}$ What is the probability when sampling without replacement? $\frac{5000}{10000} \cdot \frac{4999}{9999} = \mathbf{0.249975}$
- We often treat individuals who are sampled from a large population as independent. Using your findings from parts (a) and (b), explain whether or not this assumption is reasonable. **It is reasonable because the error is very low and presumably is within the precision required for the application. Here the error amouts to 0.000025.**

p. , ex. 4.13

A husband and wife both have brown eyes but carry genes that make it possible for their children to have brown eyes (probability 0.75), blue eyes (0.125), or green eyes (0.125).

- What is the probability the first blue-eyed child they have is their third child? Assume that the eye colors of the children are independent of each other. **Geometric distribution: 0.0957031**
- On average, how many children would such a pair of parents have before having a blue-eyed child? $\mu = \frac{1}{p} = \mathbf{8}$ What is the standard deviation of the number of children they would expect to have until the first blue-eyed child? $\sigma = \sqrt{\frac{1-p}{p^2}} = \mathbf{7.4833148}$

p. , ex. 4.14

A machine that produces a special type of transistor (a component of computers) has a 2% defective rate. The production is considered a random process where each transistor is independent of the others.

- What is the probability that the 10^{th} transistor produced is the first with a defect? **0.1992686**
- What is the probability that the machine produces no defective transistors in a batch of 100? **0.1326196**
- On average, how many transistors would you expect to be produced before the first with a defect? **50** What is the standard deviation? **49.4974747**
- Another machine that also produces transistors has a 5% defective rate where each transistor is produced independent of the others. On average how many transistors would you expect to be produced with this machine before the first with a defect? **20** What is the standard deviation? **19.4935887**
- Based on your answers to parts (c) and (d), how does increasing the probability of an event affect the mean and standard deviation of the wait time until success? **It reduces both the mean and the standard deviation.**

p. , ex. 4.15

Use the probability rules from Section 3.4 to derive the mean of a Bernoulli random variable, i.e. a random variable X that takes value 1 with probability p and value 0 with probability $1-p$. That is, compute the expected value of a generic Bernoulli random variable.

$\mu = E(X) = \sum_1^k x_i P(X = x_i)$

Here, $n = 2$, so

$x_1 P(X = x_1) + x_2 P(X = x_2) = 0 \cdot (1-p) + 1 \cdot p = p$

p. , ex. 4.16

Use the probability rules from Section 3.4 to derive the standard deviation of a Bernoulli random variable, i.e. a random variable X that takes value 1 with probability p and value 0 with probability $1-p$. That is, compute the square root of the variance of a generic Bernoulli random variable.

$\sigma^2 = \sum_{i=1}^k (x_1 - \mu)^2 P(X = x_i)$

$(x_1 - \mu)^2 P(X = x_1) + (x_2 - \mu)^2 P(X = x_2)$

$(0-p)^2 \cdot (1-p) + (1-p)^2 \cdot p$

$= p^2(1-p) + p(1 - 2p + p^2)$

$= -p^3 + p^2 + p - 2p^2 + p^3$

$= p(p + 1 - 2p)$

$= p(1-p)$

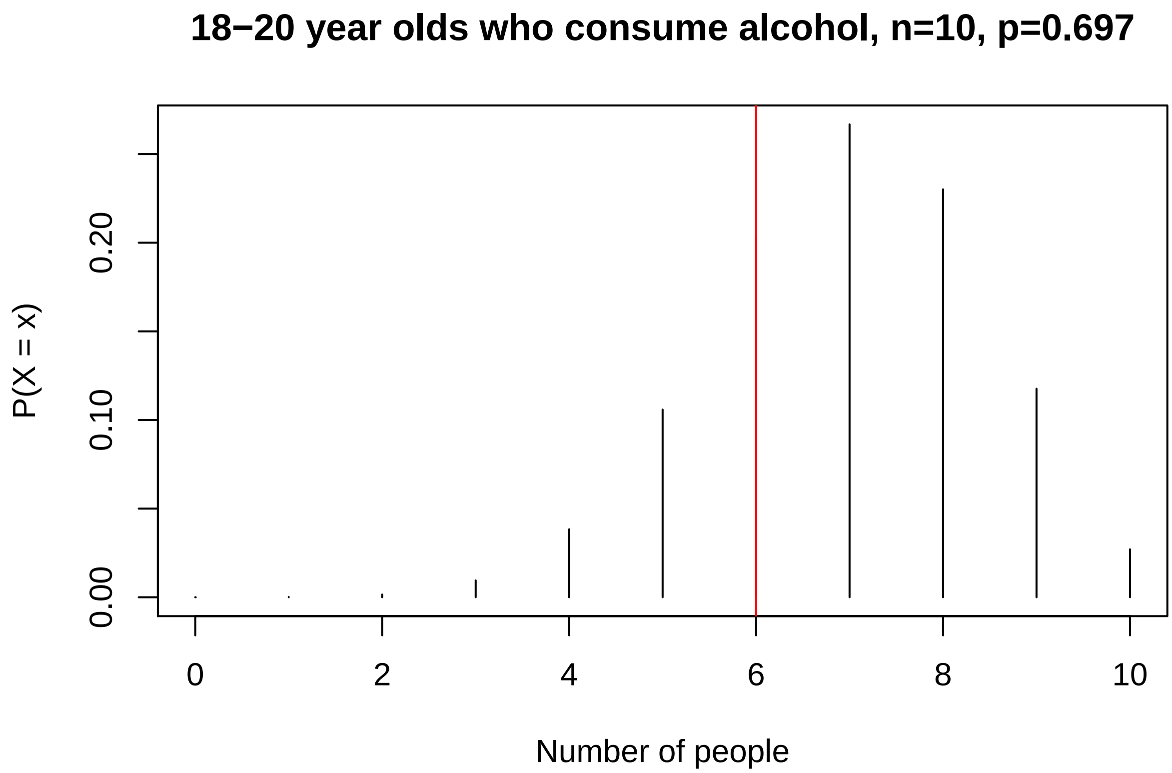
$s = \sqrt{p(1-p)}$

p. , ex. 4.17

Data collected by the Substance Abuse and Mental Health Services Administration (SAMSHA) suggests that 69.7% of 18-20 year olds consumed alcoholic beverages in any given year.

- Suppose a random sample of ten 18-20 year olds is taken. Is the use of the binomial distribution appropriate for calculating the probability that exactly six consumed alcoholic beverages? Explain. **Yes. 1. The trials are independent. 2. Number of trials is fixed at 6. 3. Each trial can be classified as a success or a failure, consumed or did not consume alcohol. 4. The probability of success is the same for each trial, 69.7%.**
- Calculate the probability that exactly 6 out of 10 randomly sampled 18- 20 year olds consumed an alcoholic drink.

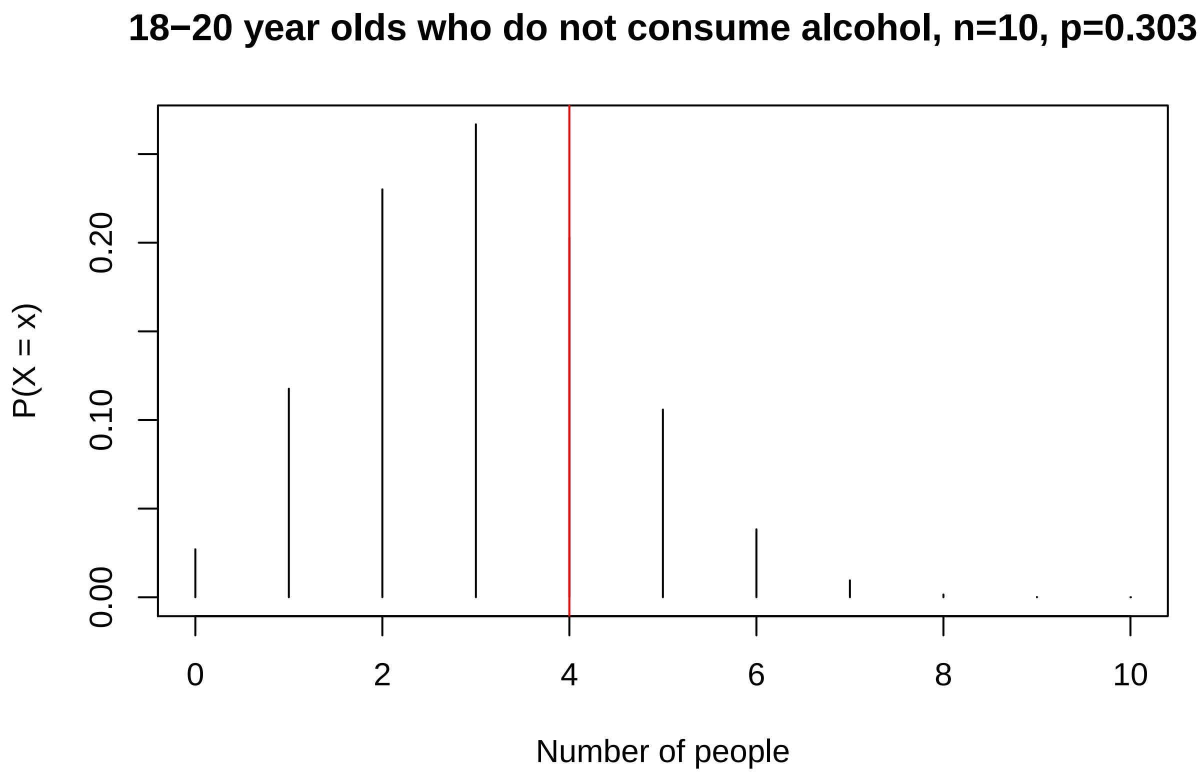
```
binom_summary(successes = 6, trials = 10, prob = 0.697,
              main = "18-20 year olds who consume alcohol",
              xlab = "Number of people")
```




```
## [1] "mu: 6.97"
## [1] "sigma: 1.45324120503102"
## [1] "Trials: 10"
## [1] "Successes: 6"
## [1] "Binomial probability. P(X=x): 0.2029"
## [1] "Cumulative probability. P(X<x): 0.1555"
## [1] "Cumulative probability. P(X<=x): 0.3584"
## [1] "Cumulative probability. P(X>x): 0.6416"
## [1] "Cumulative probability. P(X>=x): 0.8445"
```

- **0.2029488**
- **Don't do this, using `pbinom()`. 0.3584107. You don't want cumulative density, you want the probability density. cf. A Guide to dbinom, pbinom, qbinom, and rbinom in R**
- **Probability density with `dbinom`. 0.2029488**
- What is the probability that exactly four out of ten 18-20 year olds have *not* consumed an alcoholic beverage?

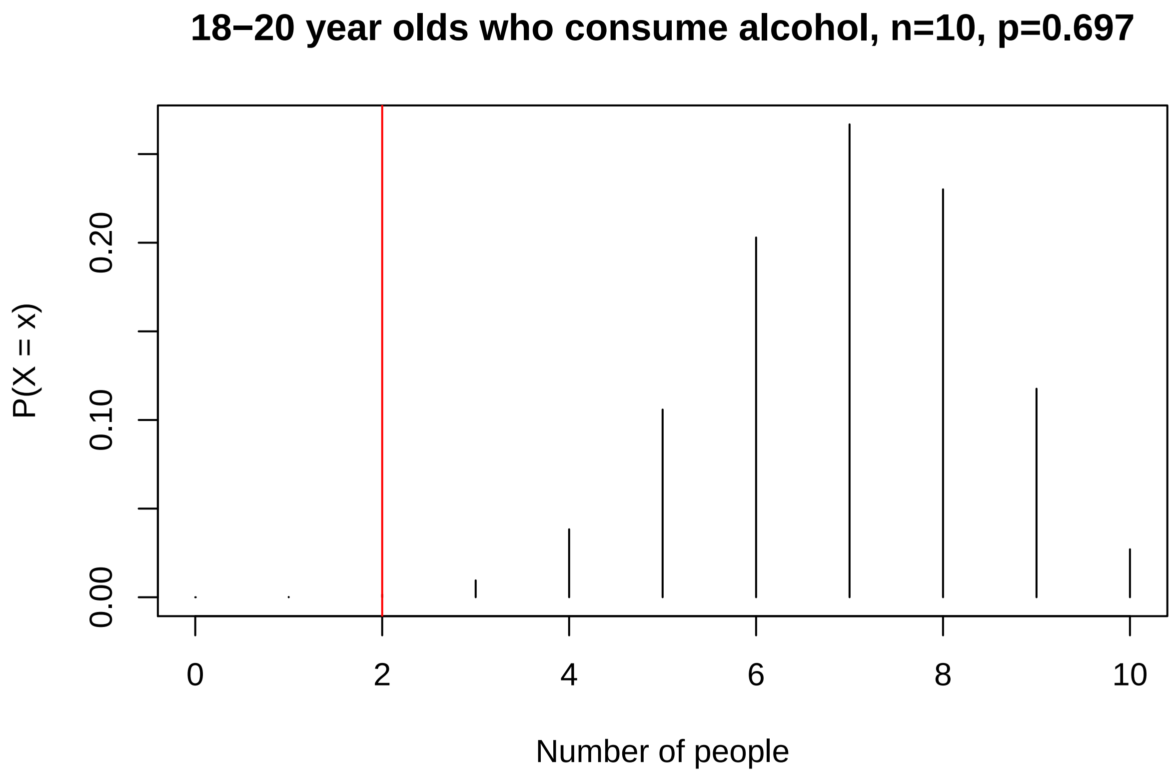
```
binom_summary(successes = 4, trials = 10, prob = 1 - 0.697,
              main = "18-20 year olds who do not consume alcohol",
              xlab = "Number of people")
```




```
## [1] "mu: 3.03"
## [1] "sigma: 1.45324120503102"
## [1] "Trials: 10"
## [1] "Successes: 4"
## [1] "Binomial probability. P(X=x): 0.2029"
## [1] "Cumulative probability. P(X<x): 0.6416"
## [1] "Cumulative probability. P(X<=x): 0.8445"
## [1] "Cumulative probability. P(X>x): 0.1555"
## [1] "Cumulative probability. P(X>=x): 0.3584"
```

- **I think I can reverse the probabilities of success and failure, in effect defining failure as Bernoulli success.** $\binom{10}{4}(1-0.697)^4 0.697^{10-4}$
- **0.2029488**
- What is the probability that at most 2 out of 5 randomly sampled 18-20 year olds have consumed alcoholic beverages?

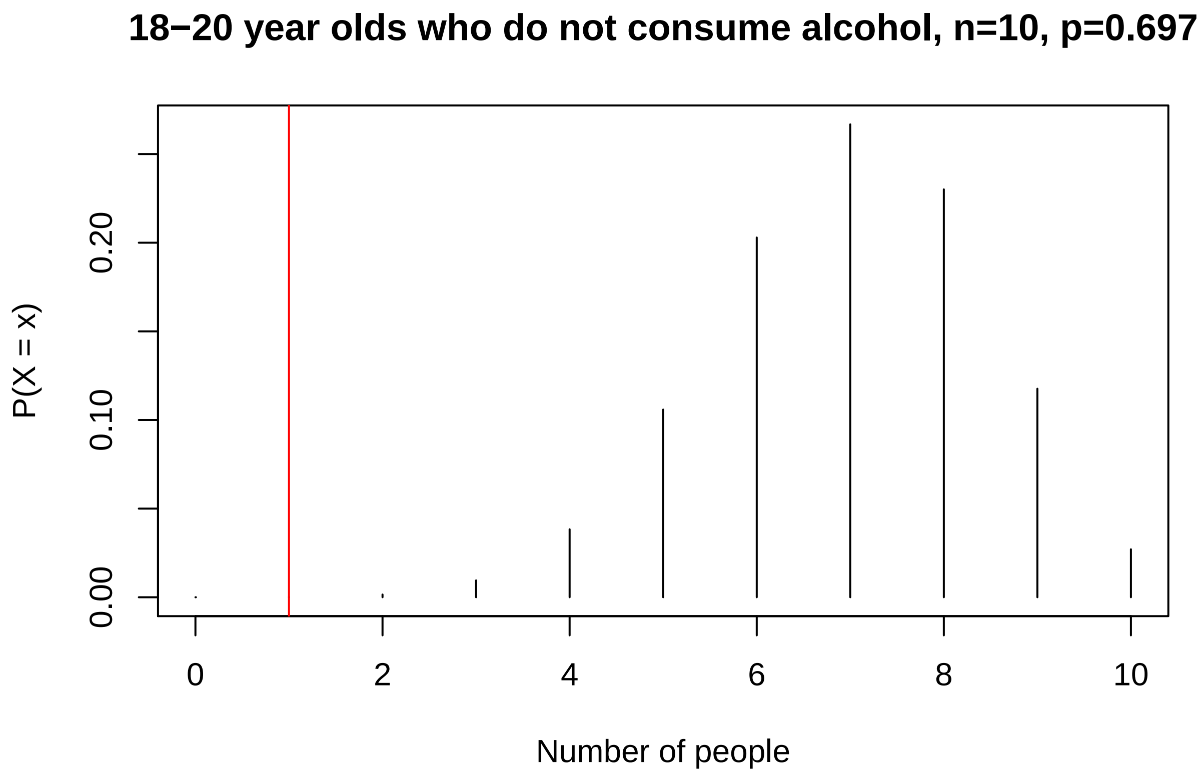
```
binom_summary(successes = 2, trials = 10, prob = 0.697,
              main = "18-20 year olds who consume alcohol",
              xlab = "Number of people")
```




```
## [1] "mu: 6.97"
## [1] "sigma: 1.45324120503102"
## [1] "Trials: 10"
## [1] "Successes: 2"
## [1] "Binomial probability. P(X=x): 0.0016"
## [1] "Cumulative probability. P(X<x): 0.0002"
## [1] "Cumulative probability. P(X<=x): 0.0017"
## [1] "Cumulative probability. P(X>x): 0.9983"
## [1] "Cumulative probability. P(X>=x): 0.9998"
```

Now it's time for CDF. 0.1670716 - What is the probability that at least 1 out of 5 randomly sampled 18-20 year olds have consumed alcoholic beverages?

```
binom_summary(successes = 1, trials = 10, prob = 0.697,
              main = "18-20 year olds who do not consume alcohol",
              xlab = "Number of people")
```



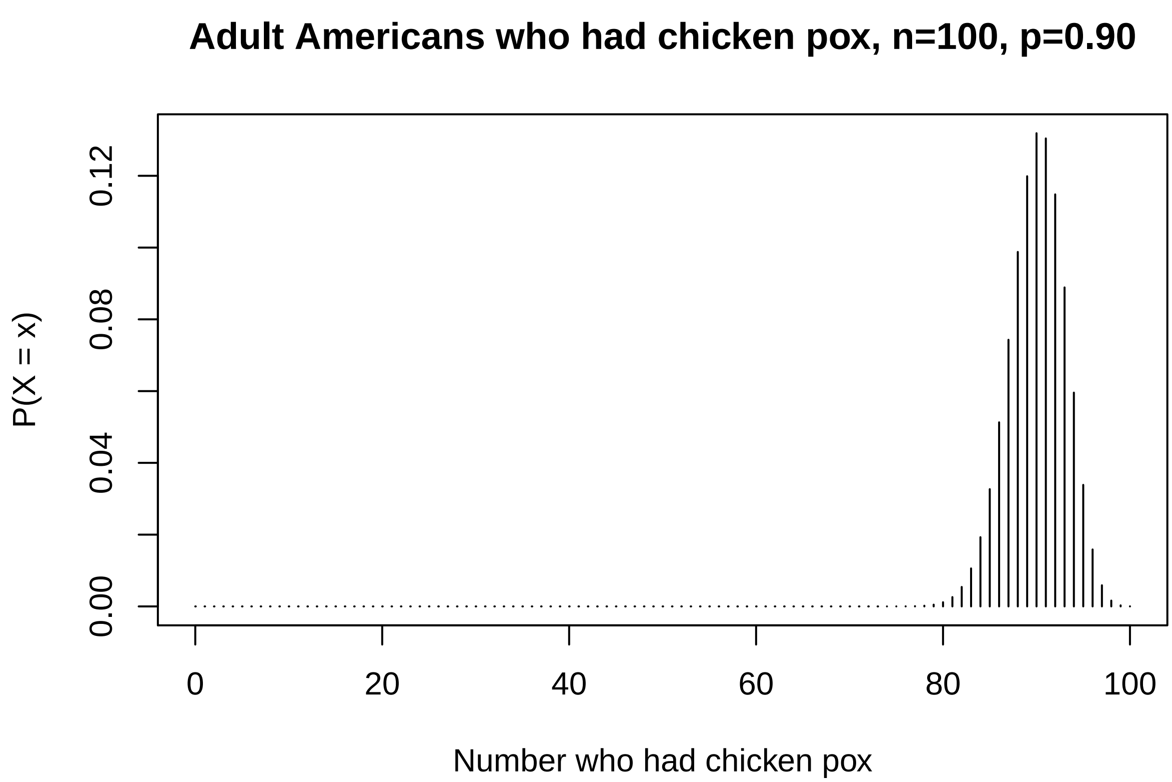

```
## [1] "mu: 6.97"
## [1] "sigma: 1.45324120503102"
## [1] "Trials: 10"
## [1] "Successes: 1"
## [1] "Binomial probability. P(X=x): 0.0002"
## [1] "Cumulative probability. P(X<x): 0"
## [1] "Cumulative probability. P(X<=x): 0.0002"
## [1] "Cumulative probability. P(X>x): 0.9998"
## [1] "Cumulative probability. P(X>=x): 1"
```

- **You can go for a complement here. 0.6424443. Well, I missed that. Try again. There are two mistakes here. First, you already forgot to distinguish between PDF and CDF. Second, you were looking at 4 choose 5. That's not it.**
- **We do want a complement. Compute the probability of getting, not exactly, 1 out of 5, but less than 1 out of 5. this includes observing 0 out of 5. Once you know that, then take its complement of that. You need to read the problems better. 0.997446. ANSWER: 0.997. I didn't get it right at first, but found other tutorials and similar practice problems**
- **Humility. No, it isn't the book. I found some supplemental binomial distribution exercises and understand my error about the complements. It all works better if I just write out the inequality. "At least 1" means $x \geq 1$. Thats the complement of $x < 1$ or $x \leq 1 - 1$. I know how to compute the probability of 0 successes. Therefore, 0.997446. That's it!**

p. , ex. 4.18

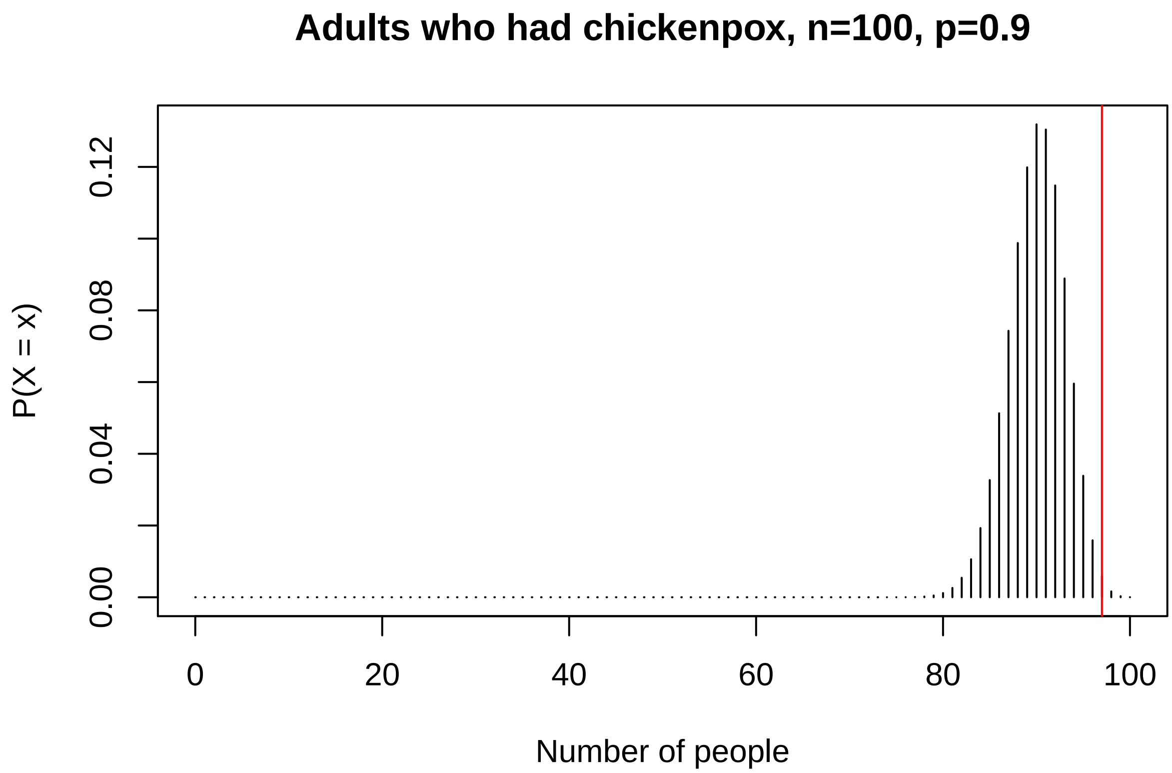
The National Vaccine Information Center estimates that 90% of Americans have had chickenpox by the time they reach adulthood.

```
plot(0:100, dbinom(x, 100, 0.90),
     xlab = "Number who had chicken pox",
     ylab = "P(X = x)",
     type = "h",
     main = "Adult Americans who had chicken pox, n=100, p=0.90")
```



- Suppose we take a random sample of 100 American adults. Is the use of the binomial distribution appropriate for calculating the probability that exactly 97 out of 100 randomly sampled American adults had chickenpox during childhood? Explain. **Yes. 1. Independent trials, 2. Fixed number of trials. 3. Success/Failure, 4. Same probability each trial.** - Calculate the probability that exactly 97 out of 100 randomly sampled American adults had chickenpox during childhood.

```
binom_summary(successes = 97, trials = 100, prob = 0.90,
              main = "Adults who had chickenpox",
              xlab = "Number of people")
```

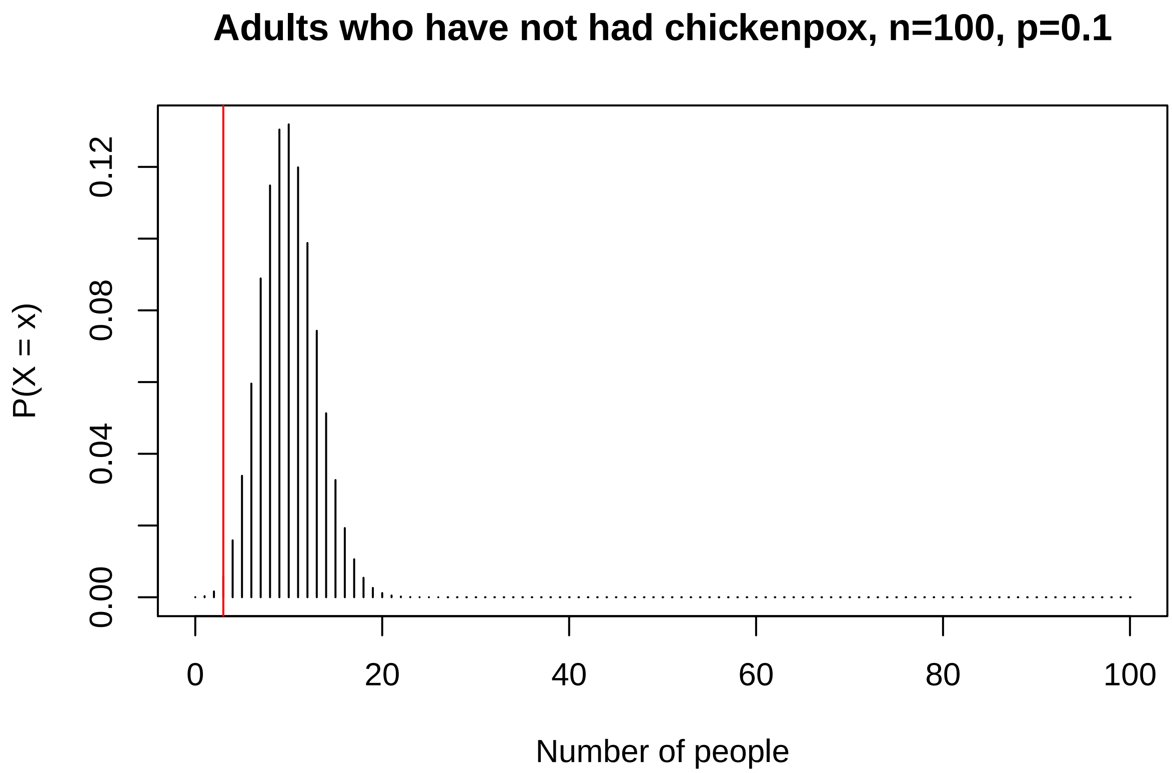



```
## [1] "mu: 90"
## [1] "sigma: 3"
## [1] "Trials: 100"
## [1] "Successes: 97"
## [1] "Binomial probability. P(X=x): 0.0059"
## [1] "Cumulative probability. P(X<x): 0.9922"
## [1] "Cumulative probability. P(X<=x): 0.9981"
## [1] "Cumulative probability. P(X>x): 0.0019"
## [1] "Cumulative probability. P(X>=x): 0.0078"
```

0.0058916

- What is the probability that exactly 3 out of a new sample of 100 American adults have *not* had chickenpox in their childhood?

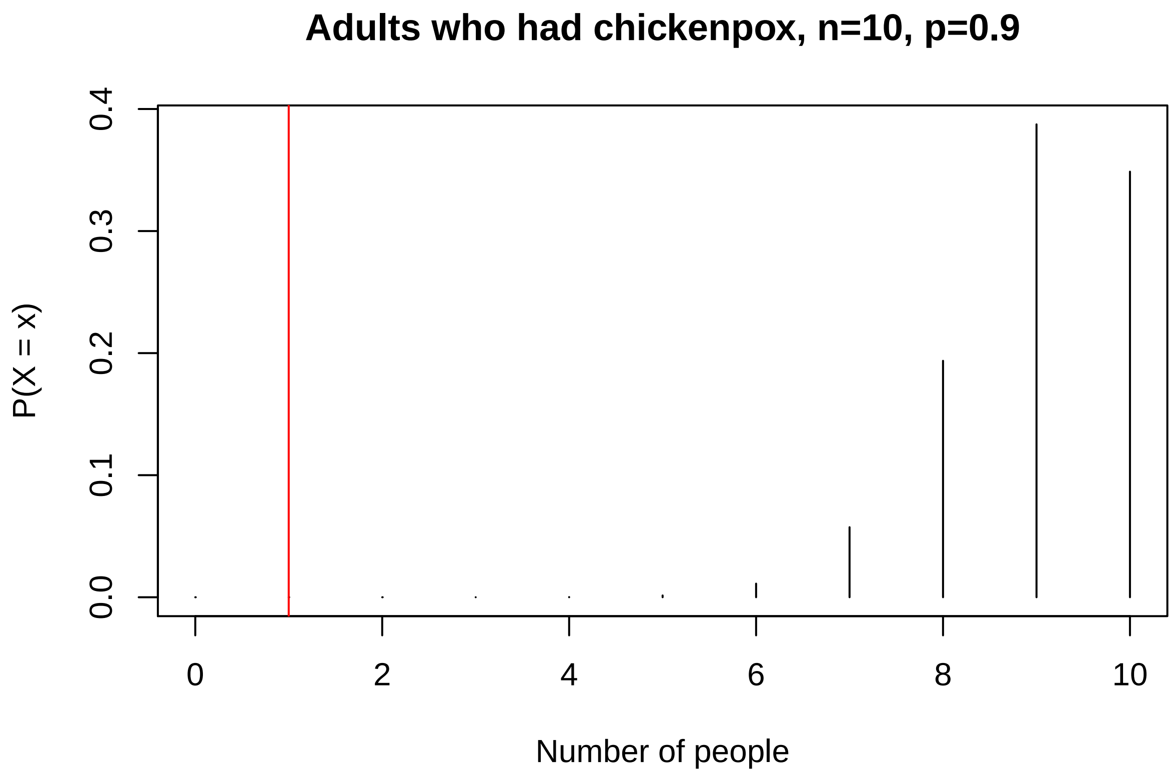
```r
binom_summary(successes = 3, trials = 100, prob = 1 - 0.90,
              main = "Adults who have not had chickenpox",
              xlab = "Number of people")
```




```
## [1] "mu: 10"
## [1] "sigma: 3"
## [1] "Trials: 100"
## [1] "Successes: 3"
## [1] "Binomial probability. P(X=x): 0.0059"
## [1] "Cumulative probability. P(X<x): 0.0019"
## [1] "Cumulative probability. P(X<=x): 0.0078"
## [1] "Cumulative probability. P(X>x): 0.9922"
## [1] "Cumulative probability. P(X>=x): 0.9981"
```

- **My original instinct is correct. cf. 4.17. 0.0058916**
- **And no, you can't do a complement. This isn't a cumulative distribution function. 0.9941084**
- What is the probability that at least 1 out of 10 randomly sampled American adults have had chickenpox?

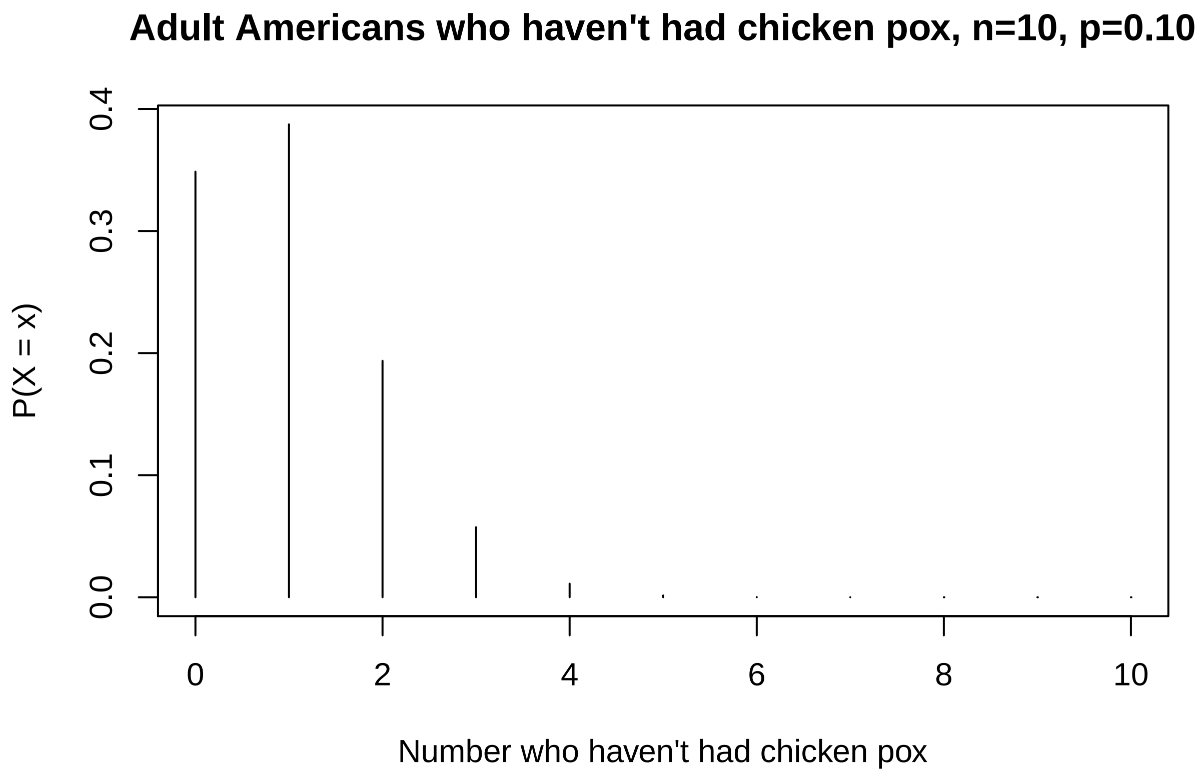
```
binom_summary(successes = 1, trials = 10, prob = 0.90,
              main = "Adults who had chickenpox",
              xlab = "Number of people")
```




```
## [1] "mu: 9"
## [1] "sigma: 0.948683298050514"
## [1] "Trials: 10"
## [1] "Successes: 1"
## [1] "Binomial probability. P(X=x): 0"
## [1] "Cumulative probability. P(X<x): 0"
## [1] "Cumulative probability. P(X<=x): 0"
## [1] "Cumulative probability. P(X>x): 1"
## [1] "Cumulative probability. P(X>=x): 1"
```

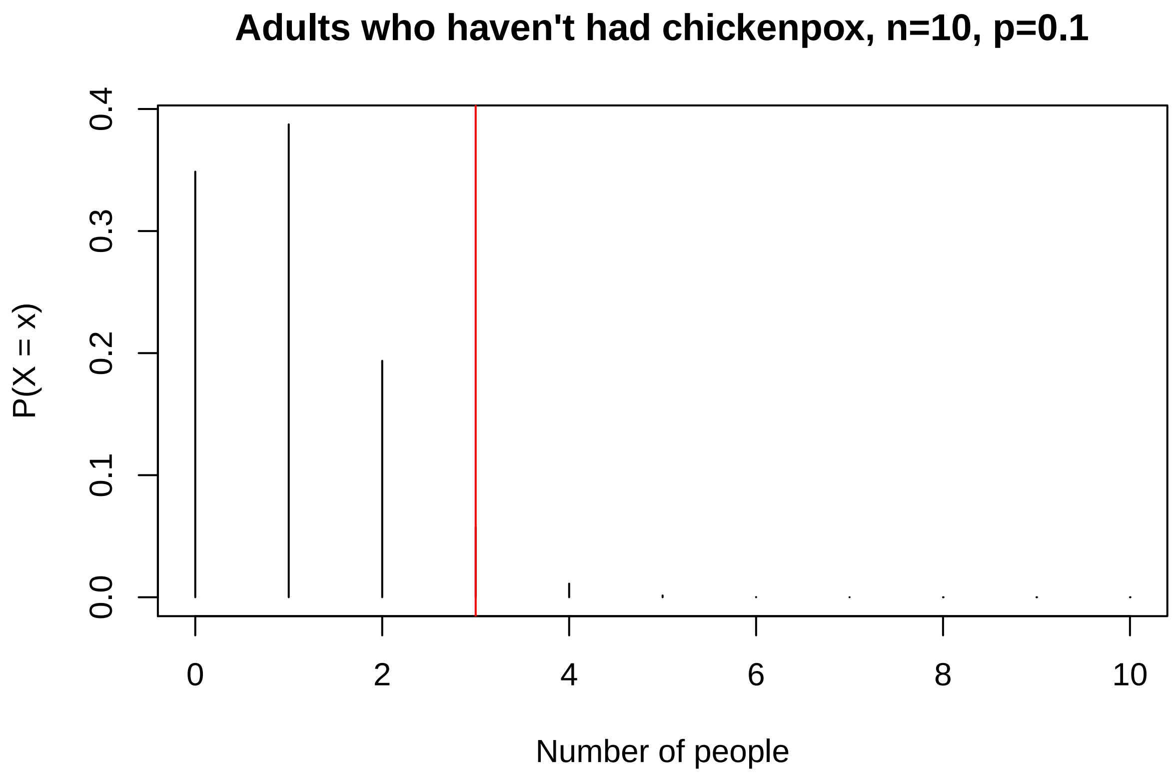
- What?? **I didn't understand this result at first, but it's obvious if you plot the distribution.** $\mathbf{P(X = 0)}$ **is so low, it rounds to 0. 1.** *Note: you got the right boundary for `pbinom()` in the complement here, but you regressed and wrote off-by-one errors later.*
- **1 this is right.**
- What is the probability that at most 3 out of 10 randomly sampled American adults have *not* had chickenpox?
  - First plot.

```
x <- 0:10
plot(x, dbinom(x, 10, 1 - 0.90),
     xlab = "Number who haven't had chicken pox",
     ylab = "P(X = x)",
     type = "h",
     main = "Adult Americans who haven't had chicken pox, n=10, p=0.10")
```



- Binary distribution summary.

```
binom_summary(successes = 3, trials = 10, prob = 1 - 0.90,
              main = "Adults who haven't had chickenpox",
              xlab = "Number of people")
```




```
## [1] "mu: 1"
## [1] "sigma: 0.948683298050514"
## [1] "Trials: 10"
## [1] "Successes: 3"
## [1] "Binomial probability. P(X=x): 0.0574"
## [1] "Cumulative probability. P(X<x): 0.9298"
## [1] "Cumulative probability. P(X<=x): 0.9872"
## [1] "Cumulative probability. P(X>x): 0.0128"
## [1] "Cumulative probability. P(X>=x): 0.0702"
```

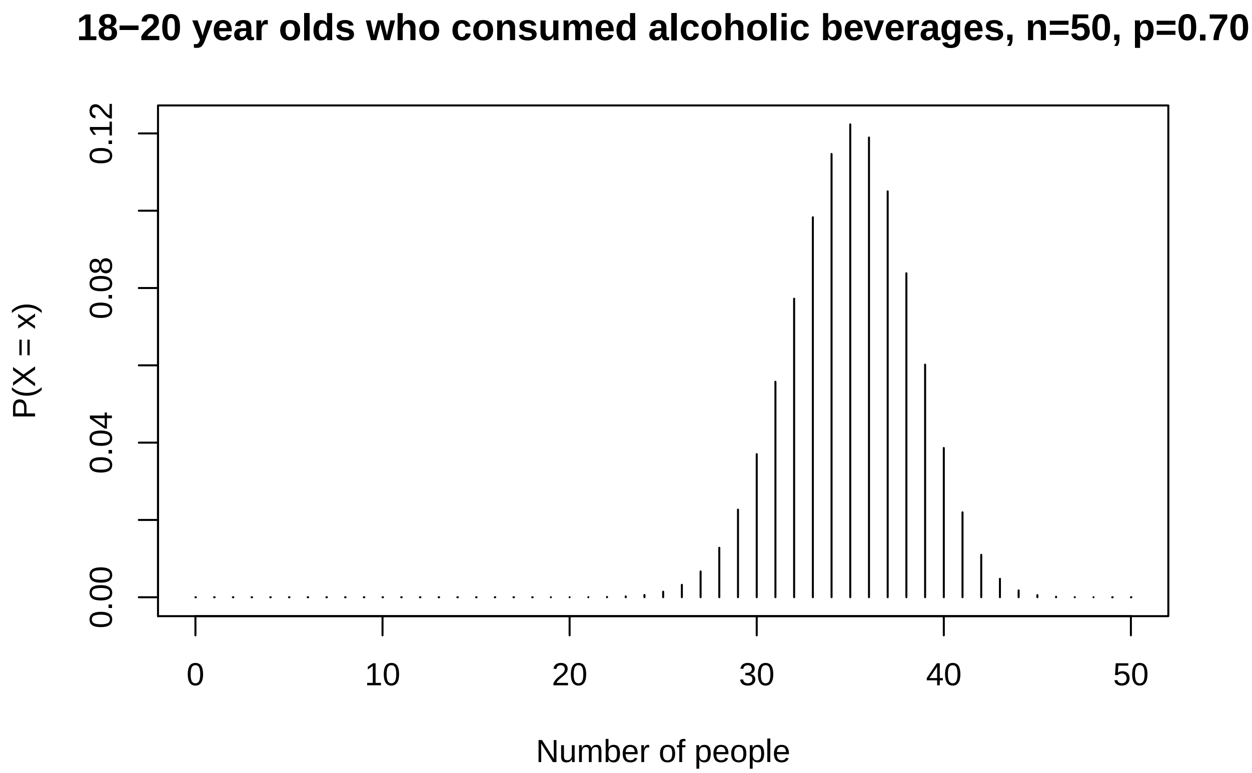
0.9872048

p. , ex. 4.19

We learned in Exercise 4.17 that about 70% of 18-20 year olds consumed alcoholic beverages in any given year. We now consider a random sample of fifty 18-20 year olds.

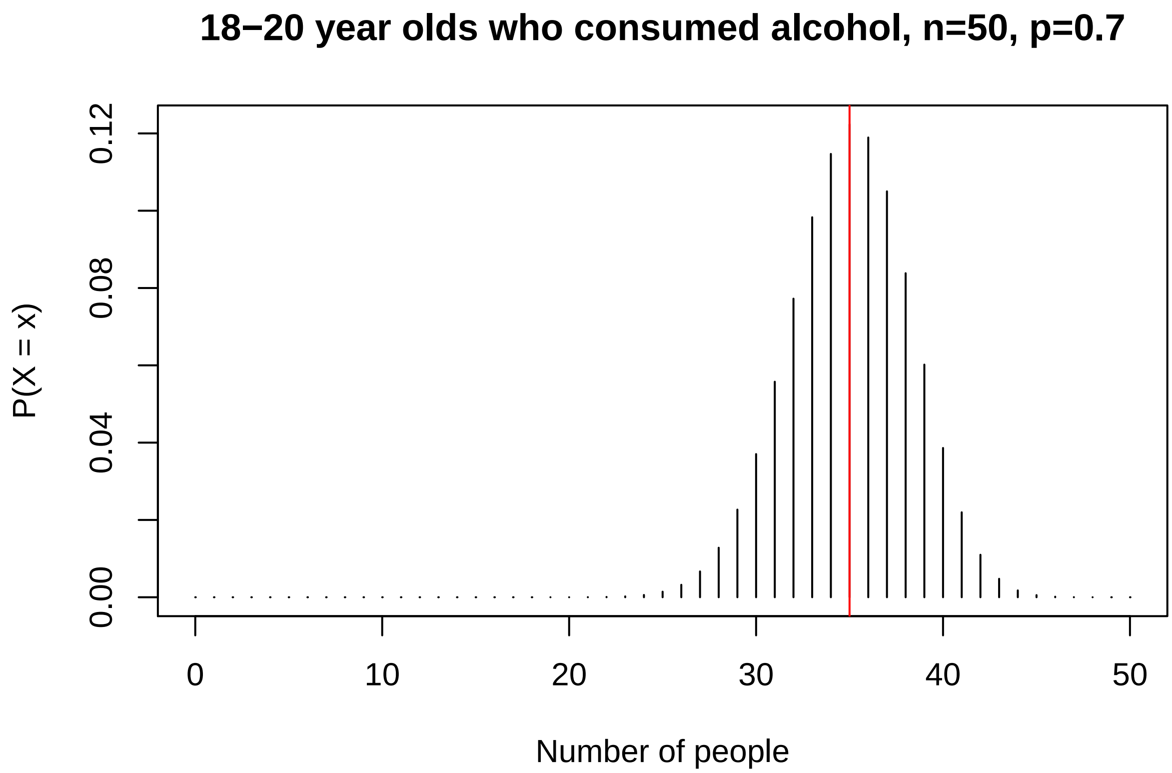
- First plot.

```
x <- 0:50
plot(x, dbinom(x, 50, 0.70),
     xlab = "Number of people",
     ylab = "P(X = x)",
     type = "h",
     main = "18-20 year olds who consumed alcoholic beverages, n=50, p=0.70")
```



- How many people would you expect to have consumed alcoholic beverages? And with what standard deviation?

```
binom_summary(successes = 35, trials = 50, prob = 0.70,
              main = "18-20 year olds who consumed alcohol",
              xlab = "Number of people")
```

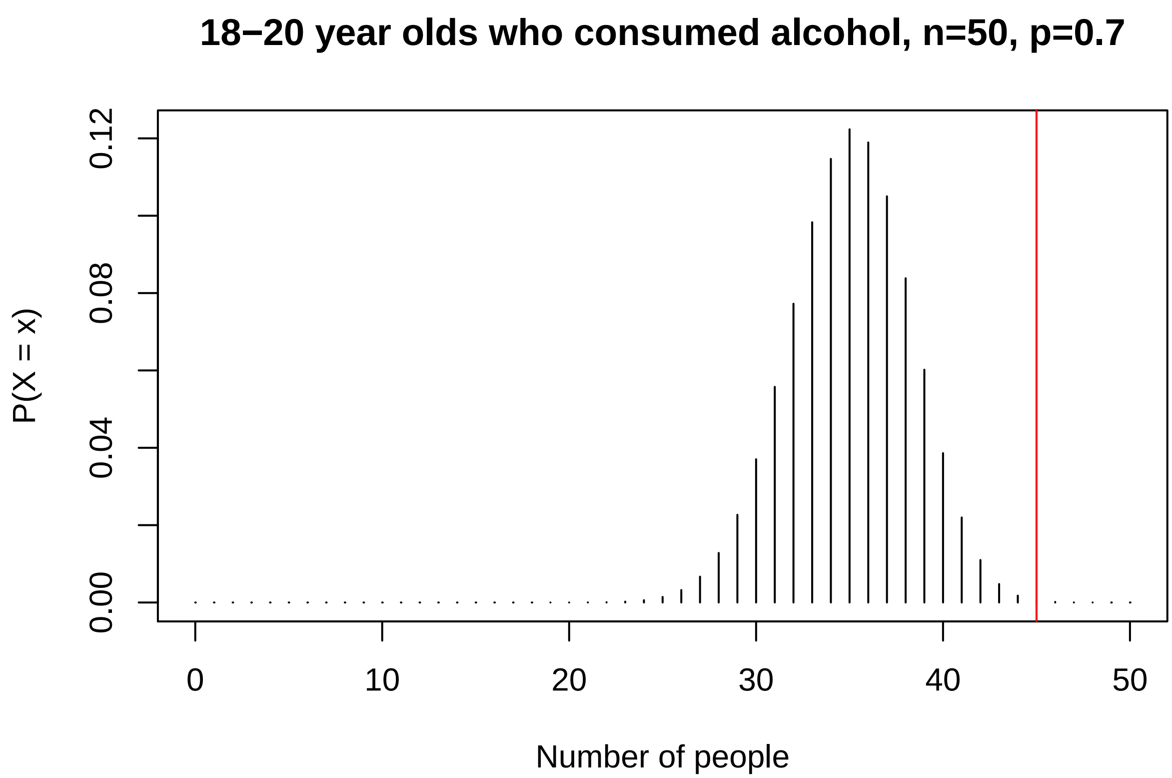



```
## [1] "mu: 35"
## [1] "sigma: 3.24037034920393"
## [1] "Trials: 50"
## [1] "Successes: 35"
## [1] "Binomial probability. P(X=x): 0.1223"
## [1] "Cumulative probability. P(X<x): 0.4308"
## [1] "    Normal approximation. P(X<x): 0.5"
## [1] "    Normal approximation w/ continuity correction. P(<x): 0.4387"
## [1] "Cumulative probability. P(X<=x): 0.5532"
## [1] "    Normal approximation. P(X<=x): 0.5"
## [1] "    Normal approximation w/ continuity correction. P(<=x): 0.5613"
## [1] "Cumulative probability. P(X>x): 0.4468"
## [1] "    Normal approximation. P(X>x): 0.5"
## [1] "    Normal approximation w/ continuity correction. P(X>x): 0.4387"
## [1] "Cumulative probability. P(X>=x): 0.5692"
## [1] "    Normal approximation. P(X>=x): 0.5"
## [1] "    Normal approximation w/ continuity correction. P(X>=x): 0.5613"
```

- $np = $ **35. ANSWER: the book comes up with 34.85. I don't know why.**
- $\sigma = \sqrt{np(1-p)} = \sqrt{npq} = $ **3.2403703.**
- Would you be surprised if there were 45 or more people who have consumed alcoholic beverages?
  - **Yes, it's very unlikely.**
  - **I based my answer on my plot. I should also know the book's approach using standard deviation. Thats what σ is for! $Z =$ 3.086067. The book came up with 3.12 because of its differing mean. 45 people lies more than 3 standard deviations from the mean, so it is unlikely.**

- What is the probability that 45 or more people in this sample have consumed alcoholic beverages? How does this probability relate to your answer to part (b)?

```
binom_summary(successes = 45, trials = 50, prob = 0.70,
              main = "18-20 year olds who consumed alcohol",
              xlab = "Number of people")
```




```
## [1] "mu: 35"
## [1] "sigma: 3.24037034920393"
## [1] "Trials: 50"
## [1] "Successes: 45"
## [1] "Binomial probability. P(X=x): 0.0006"
## [1] "Cumulative probability. P(X<x): 0.9993"
## [1] "    Normal approximation. P(X<x): 0.999"
## [1] "    Normal approximation w/ continuity correction. P(<x): 0.9983"
## [1] "Cumulative probability. P(X<=x): 0.9998"
## [1] "    Normal approximation. P(X<=x): 0.999"
## [1] "    Normal approximation w/ continuity correction. P(<=x): 0.9994"
## [1] "Cumulative probability. P(X>x): 0.0002"
## [1] "    Normal approximation. P(X>x): 0.001"
## [1] "    Normal approximation w/ continuity correction. P(X>x): 0.0006"
## [1] "Cumulative probability. P(X>=x): 0.0007"
## [1] "    Normal approximation. P(X>=x): 0.001"
## [1] "    Normal approximation w/ continuity correction. P(X>=x): 0.0017"
```

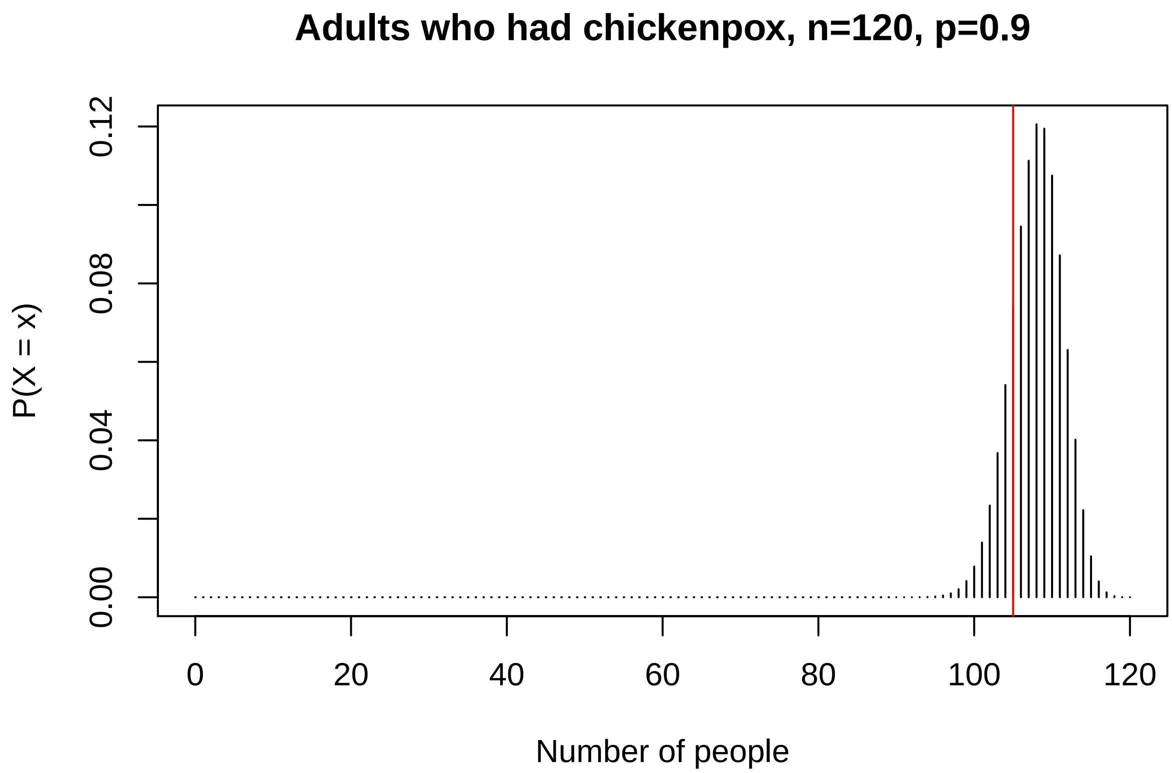
- **Correct. Calculate with sums of probabilities (`pbinom()`). 0.0007229. This low probability is the reason for my answer. I could see it in my plot. ANSWER: the book comes up with 0.0009. I'm trying to figure out if I have an off-by-1 error. I need to start from 45 - 1.**
- **Try the normal approximation. 0.0010141**
- **Try it with continuity correction. 0.0016852**
- My sums don't match my normal approximation very well, and everything is way off from the book's answers. Moving on, I'll get more opportunities to practice this in this chapter, Grinstead, and my workbooks coming in the mail. This website says my `pbinom()` calc is right: Binomial Probability Calculator.

p. , ex. 4.20

We learned in Exercise 4.18 that about 90% of American adults had chickenpox before adulthood. We now consider a random sample of 120 American adults.

- How many people in this sample would you expect to have had chickenpox in their childhood? $np =$ **108.** And with what standard deviation? $\sqrt{npq} =$ **3.2863353.**
- Would you be surprised if there were 105 people who have had chickenpox in their childhood? $Z =$ **32.8633535. No, the distance of 105 from the mean of 108 is well within the standard deviation of 33.**
- What is the probability that 105 or fewer people in this sample have had chickenpox in their childhood? How does this probability relate to your answer to part (b)?

```
binom_summary(successes = 105, trials = 120, prob = 0.90,
              main = "Adults who had chickenpox",
              xlab = "Number of people")
```




```
## [1] "mu: 108"
## [1] "sigma: 3.286335345031"
## [1] "Trials: 120"
## [1] "Successes: 105"
## [1] "Binomial probability. P(X=x): 0.0742"
## [1] "Cumulative probability. P(X<x): 0.144"
## [1] "    Normal approximation. P(X<x): 0.1807"
## [1] "    Normal approximation w/ continuity correction. P(<x): 0.1434"
## [1] "Cumulative probability. P(X<=x): 0.2182"
## [1] "    Normal approximation. P(X<=x): 0.1807"
## [1] "    Normal approximation w/ continuity correction. P(<=x): 0.2234"
## [1] "Cumulative probability. P(X>x): 0.7818"
## [1] "    Normal approximation. P(X>x): 0.8193"
## [1] "    Normal approximation w/ continuity correction. P(X>x): 0.7766"
## [1] "Cumulative probability. P(X>=x): 0.856"
## [1] "    Normal approximation. P(X>=x): 0.8193"
## [1] "    Normal approximation w/ continuity correction. P(X>=x): 0.8566"
```

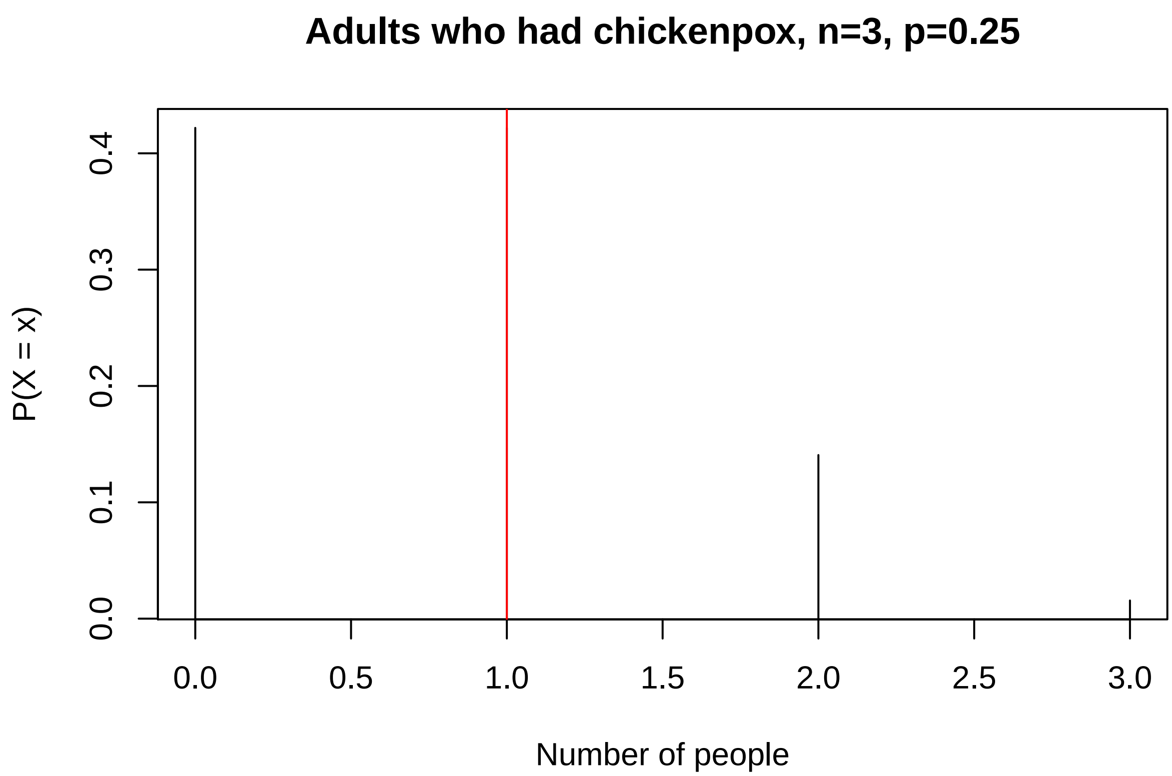
- **`pbinom()`. 0.2181634. Correct.**
- **Normal approximation. 0.1806552**
- **Normal approximation with continuity correction. 0.2234104**

p. , ex. 4.21

A dreidel is a four-sided spinning top with the Hebrew letters *nun*, *gimel*, *hei*, and *shin*, one on each side. Each side is equally likely to come up in a single spin of the dreidel. Suppose you spin a dreidel three times. Calculate the probability of getting

- at least one *nun*?

```
binom_summary(successes = 1, trials = 3, prob = 0.25,
              main = "Adults who had chickenpox",
              xlab = "Number of people")
```

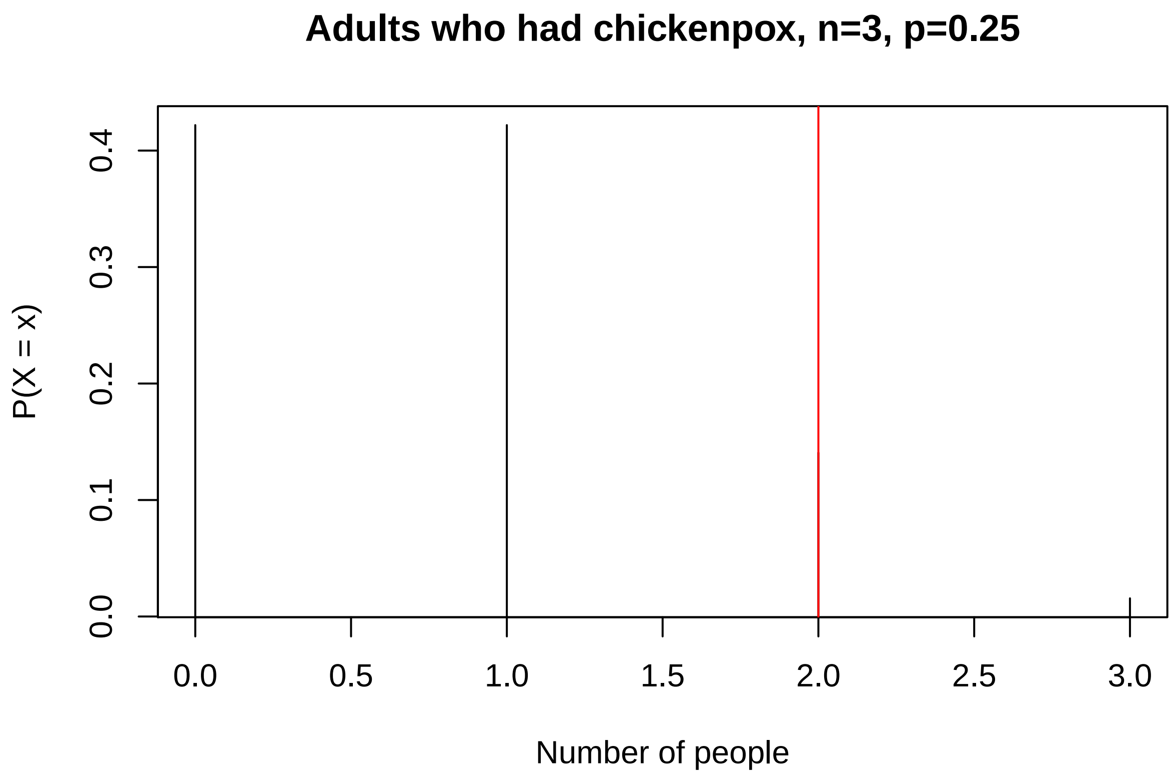



```
## [1] "mu: 0.75"
## [1] "sigma: 0.75"
## [1] "Trials: 3"
## [1] "Successes: 1"
## [1] "Binomial probability. P(X=x): 0.4219"
## [1] "Cumulative probability. P(X<x): 0.4219"
## [1] "Cumulative probability. P(X<=x): 0.8438"
## [1] "Cumulative probability. P(X>x): 0.1562"
## [1] "Cumulative probability. P(X>=x): 0.5781"
```

- **0.578125. Correct**
- **Normal approximation. Fails conditions. $np < 10 > n(1 - p).**

- exactly 2 *nuns*?

```r
binom_summary(successes = 2, trials = 3, prob = 0.25,
              main = "Adults who had chickenpox",
              xlab = "Number of people")
```




```
## [1] "mu: 0.75"
## [1] "sigma: 0.75"
## [1] "Trials: 3"
## [1] "Successes: 2"
## [1] "Binomial probability. P(X=x): 0.1406"
## [1] "Cumulative probability. P(X<x): 0.8438"
## [1] "Cumulative probability. P(X<=x): 0.9844"
## [1] "Cumulative probability. P(X>x): 0.0156"
## [1] "Cumulative probability. P(X>=x): 0.1562"
```

0.140625

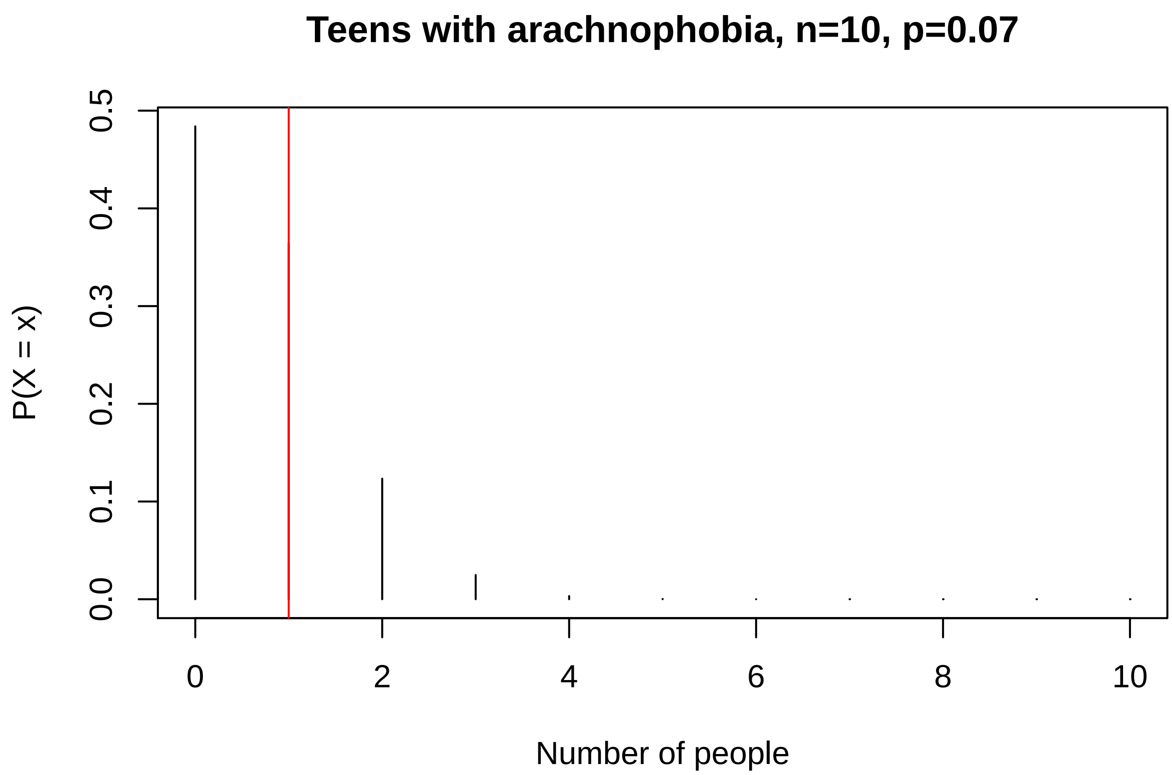
- exactly 1 *hei*? **0.421875**
- at most 2 *gimels*? **0.984375**

p. , ex. 4.22

A Gallup Poll found that 7% of teenagers (ages 13 to 17) suffer from arachnophobia and are extremely afraid of spiders. At a summer camp there are 10 teenagers sleeping in each tent. Assume that these 10 teenagers are independent of each other.

- Calculate the probability that at least one of them suffers from arachnophobia. **0.5160177**

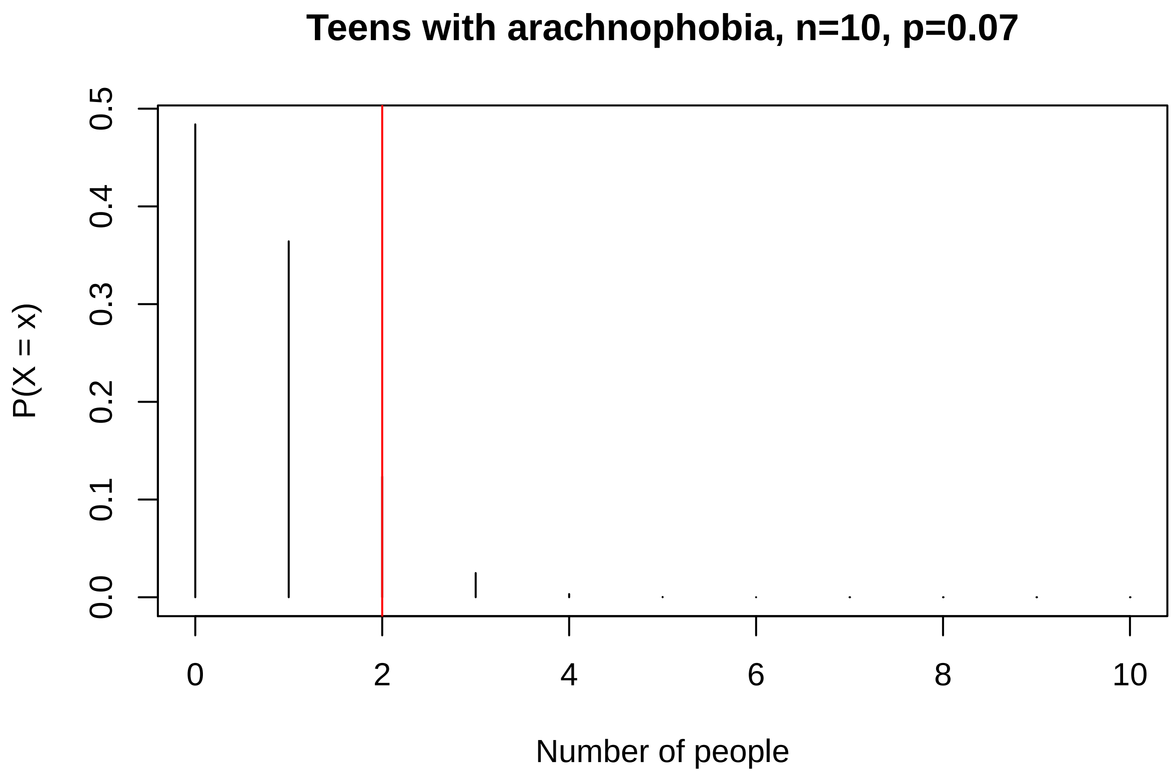
```r
binom_summary(successes = 1, trials = 10, prob = 0.07,
              main = "Teens with arachnophobia",
              xlab = "Number of people")
```




```
## [1] "mu: 0.7"
## [1] "sigma: 0.806845710157772"
## [1] "Trials: 10"
## [1] "Successes: 1"
## [1] "Binomial probability. P(X=x): 0.3643"
## [1] "Cumulative probability. P(X<x): 0.484"
## [1] "Cumulative probability. P(X<=x): 0.8483"
## [1] "Cumulative probability. P(X>x): 0.1517"
## [1] "Cumulative probability. P(X>=x): 0.516"
```

- Calculate the probability that exactly 2 of them suffer from arachnophobia. **0.1233878**

```r
binom_summary(successes = 2, trials = 10, prob = 0.07,
              main = "Teens with arachnophobia",
              xlab = "Number of people")
```




```
## [1] "mu: 0.7"
## [1] "sigma: 0.806845710157772"
## [1] "Trials: 10"
## [1] "Successes: 2"
## [1] "Binomial probability. P(X=x): 0.1234"
## [1] "Cumulative probability. P(X<x): 0.8483"
## [1] "Cumulative probability. P(X<=x): 0.9717"
## [1] "Cumulative probability. P(X>x): 0.0283"
## [1] "Cumulative probability. P(X>=x): 0.1517"
```

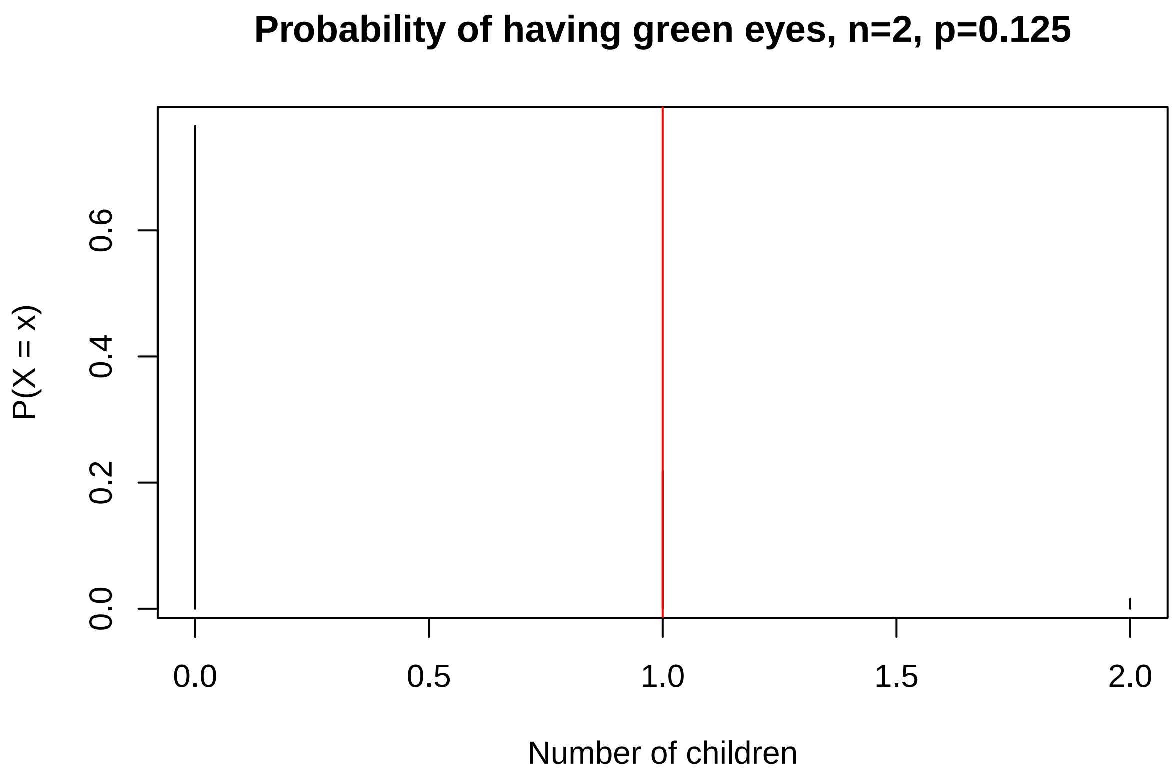
- Calculate the probability that at most 1 of them suffers from arachnophobia. **0.8482701**
- If the camp counselor wants to make sure no more than 1 teenager in each tent is afraid of spiders, does it seem reasonable for him to randomly assign teenagers to tents? **I'd be okay with probability of 0.84.**

p. , ex. 4.23

Exercise 4.13 introduces a husband and wife with brown eyes who have 0.75 probability of having children with brown eyes, 0.125 probability of having children with blue eyes, and 0.125 probability of having children with green eyes.

- What is the probability that their first child will have green eyes and the second will not? **0.109375**
- What is the probability that exactly one of their two children will have green eyes? **0.21875**

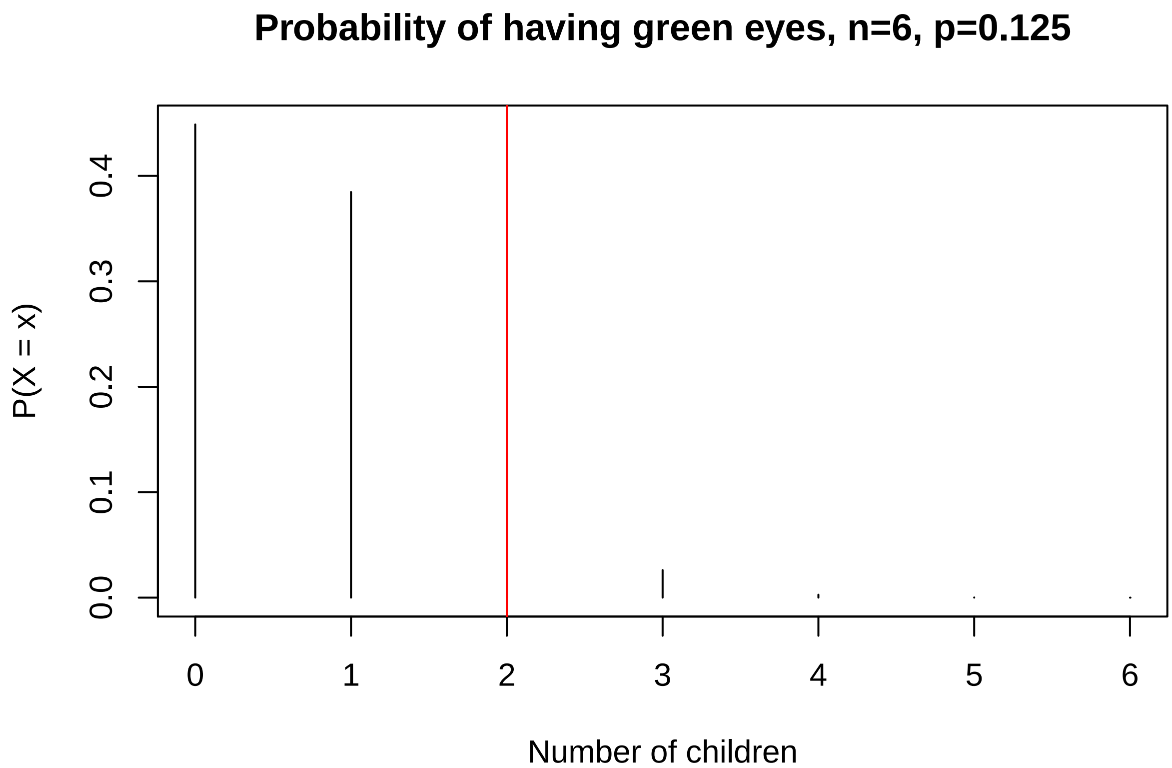
```r
binom_summary(successes = 1, trials = 2, prob = 0.125,
              main = "Probability of having green eyes",
              xlab = "Number of children")
```




```
## [1] "mu: 0.25"
## [1] "sigma: 0.467707173346743"
## [1] "Trials: 2"
## [1] "Successes: 1"
## [1] "Binomial probability. P(X=x): 0.2187"
## [1] "Cumulative probability. P(X<x): 0.7656"
## [1] "Cumulative probability. P(X<=x): 0.9844"
## [1] "Cumulative probability. P(X>x): 0.0156"
## [1] "Cumulative probability. P(X>=x): 0.2344"
```

- If they have six children, what is the probability that exactly two will have green eyes? **0.1373863**

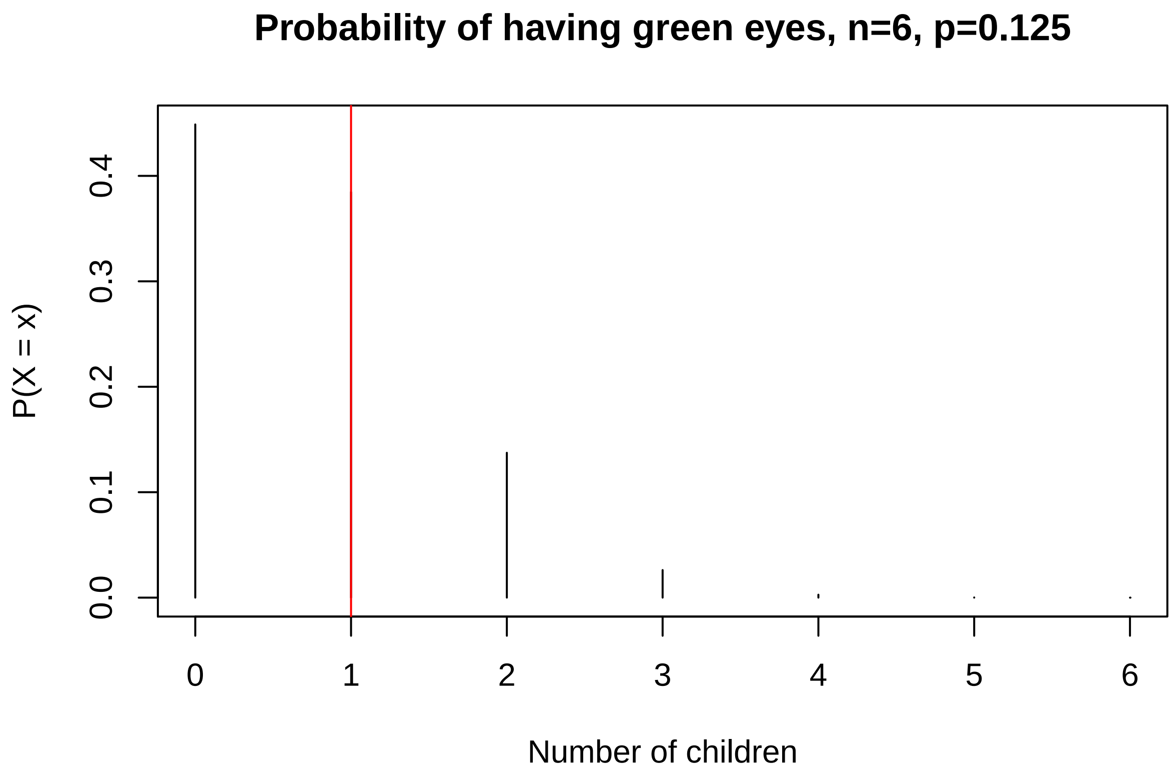
```r
binom_summary(successes = 2, trials = 6, prob = 0.125,
              main = "Probability of having green eyes",
              xlab = "Number of children")
```




```
## [1] "mu: 0.75"
## [1] "sigma: 0.810092587300983"
## [1] "Trials: 6"
## [1] "Successes: 2"
## [1] "Binomial probability. P(X=x): 0.1374"
## [1] "Cumulative probability. P(X<x): 0.8335"
## [1] "Cumulative probability. P(X<=x): 0.9709"
## [1] "Cumulative probability. P(X>x): 0.0291"
## [1] "Cumulative probability. P(X>=x): 0.1665"
```

- If they have six children, what is the probability that at least one will have green eyes?
  - **0.5512047**
  - **Easier calculation : 1 - probability that no child has green eyes 0.5512047**

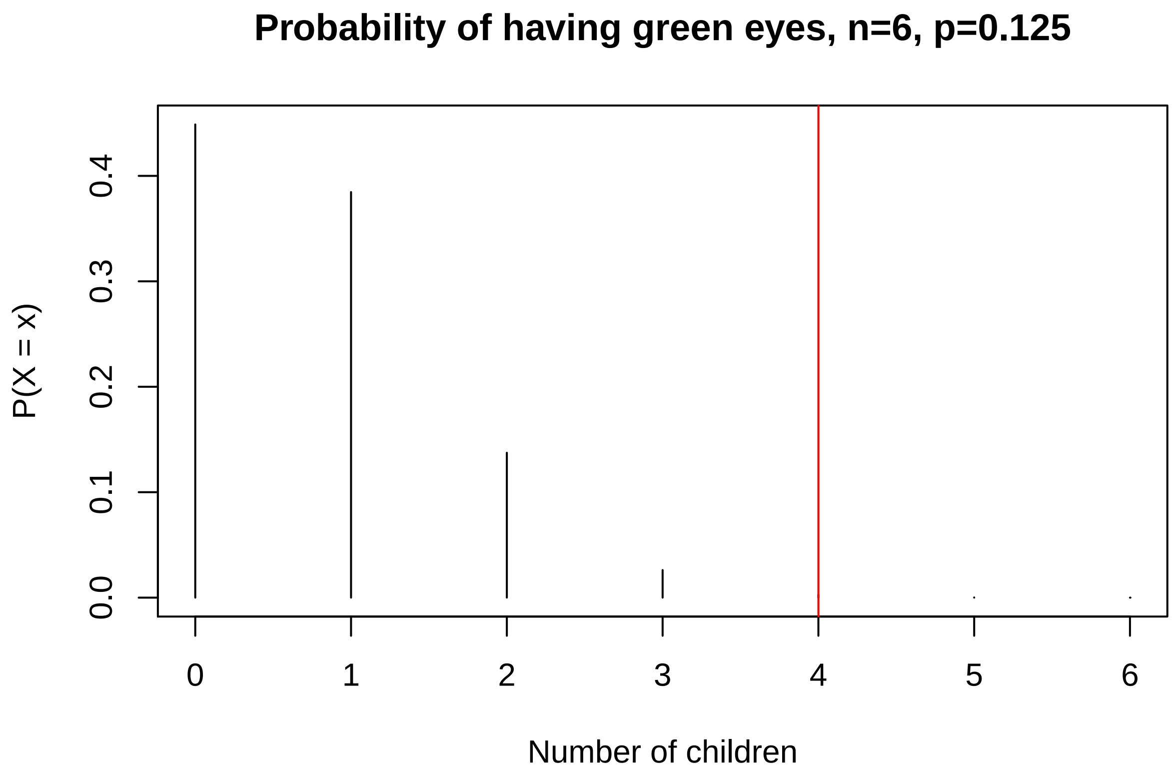
```
binom_summary(successes = 1, trials = 6, prob = 0.125,
              main = "Probability of having green eyes",
              xlab = "Number of children")
```




```
## [1] "mu: 0.75"
## [1] "sigma: 0.810092587300983"
## [1] "Trials: 6"
## [1] "Successes: 1"
## [1] "Binomial probability. P(X=x): 0.3847"
## [1] "Cumulative probability. P(X<x): 0.4488"
## [1] "Cumulative probability. P(X<=x): 0.8335"
## [1] "Cumulative probability. P(X>x): 0.1665"
## [1] "Cumulative probability. P(X>=x): 0.5512"
```

- What is the probability that the first green eyed child will be the 4^{th} child?
  - **Binomial: 0.0028038**
  - **Geometric: 0.0837402. This is correct. Why doesn't my binomial trick work?**

```r
binom_summary(successes = 4, trials = 6, prob = 0.125,
              main = "Probability of having green eyes",
              xlab = "Number of children")
```




```
## [1] "mu: 0.75"
## [1] "sigma: 0.810092587300983"
## [1] "Trials: 6"
## [1] "Successes: 4"
## [1] "Binomial probability. P(X=x): 0.0028"
## [1] "Cumulative probability. P(X<x): 0.997"
## [1] "Cumulative probability. P(X<=x): 0.9998"
## [1] "Cumulative probability. P(X>x): 0.0002"
## [1] "Cumulative probability. P(X>=x): 0.003"
```

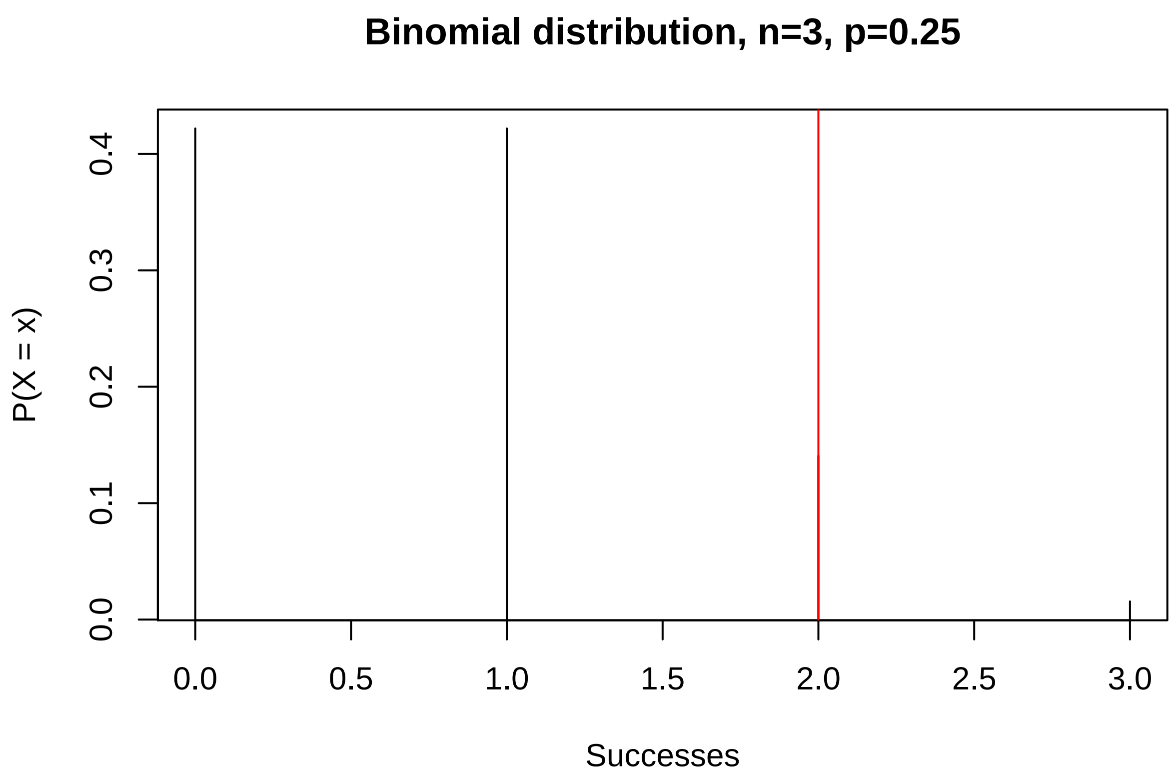
- Would it be considered unusual if only 2 out of their 6 children had brown eyes? $Z_2 =$ **-2.3570226. Yes, this lies more than 2 standard deviations from the mean.**

p. , ex. 4.24

Sickle cell anemia is a genetic blood disorder where red blood cells lose their flexibility and assume an abnormal, rigid, “sickle" shape, which results in a risk of various complications. If both parents are carriers of the disease, then a child has a 25% chance of having the disease, 50% chance of being a carrier, and 25% chance of neither having the disease nor being a carrier. If two parents who are carriers of the disease have 3 children, what is the probability that

- two will have the disease? **0.140625**

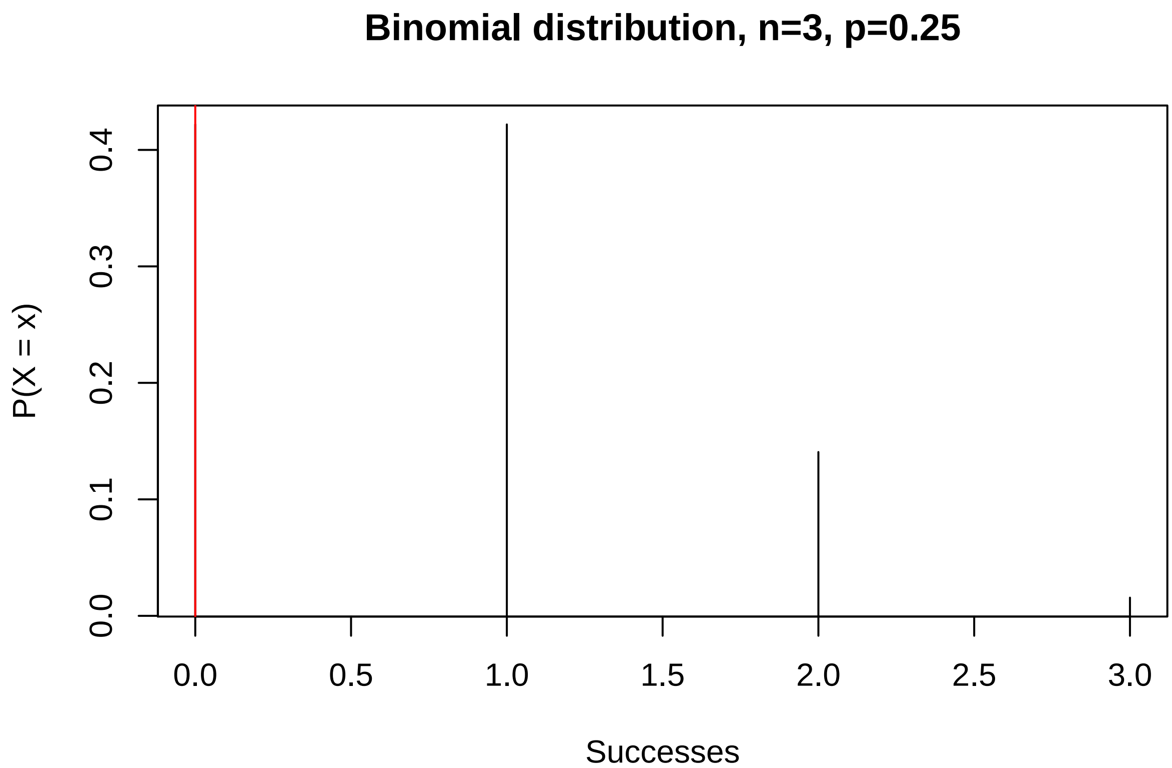
```r
binom_summary(2, 3, 0.25)
```




```
## [1] "mu: 0.75"
## [1] "sigma: 0.75"
## [1] "Trials: 3"
## [1] "Successes: 2"
## [1] "Binomial probability. P(X=x): 0.1406"
## [1] "Cumulative probability. P(X<x): 0.8438"
## [1] "Cumulative probability. P(X<=x): 0.9844"
## [1] "Cumulative probability. P(X>x): 0.0156"
## [1] "Cumulative probability. P(X>=x): 0.1562"
```

- none will have the disease? **0.421875**

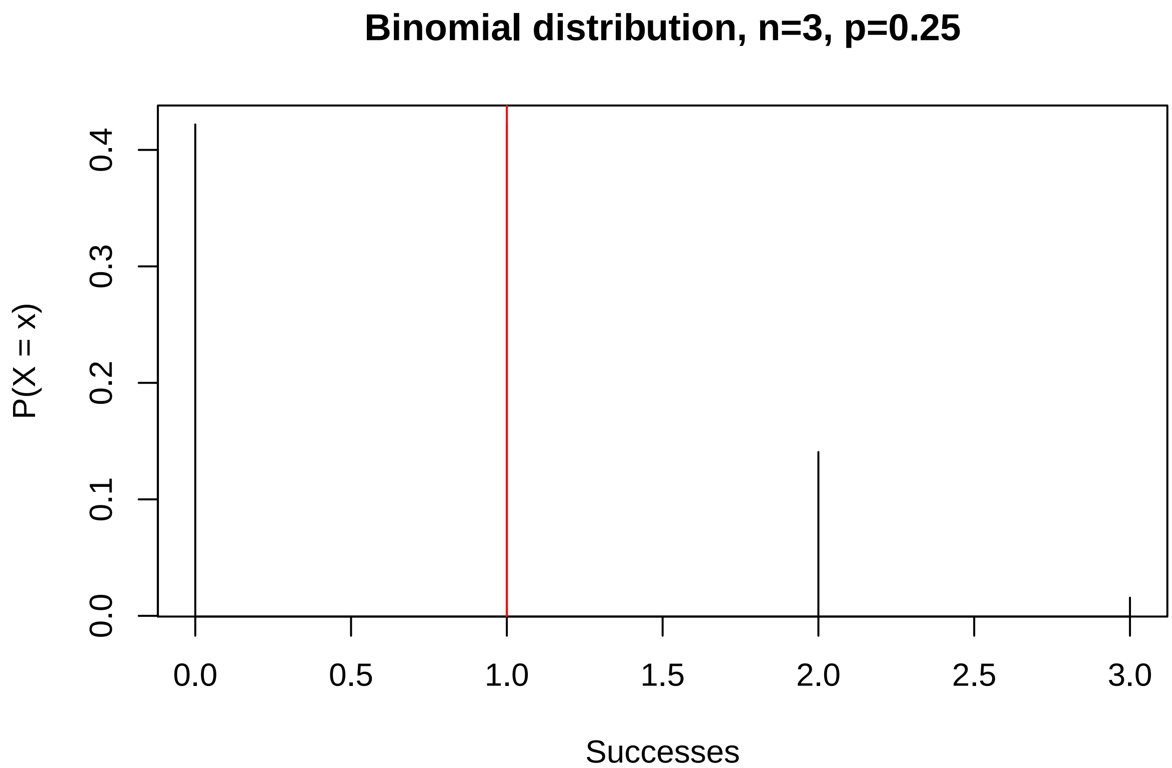
```r
binom_summary(0, 3, 0.25)
```




```
## [1] "mu: 0.75"
## [1] "sigma: 0.75"
## [1] "Trials: 3"
## [1] "Successes: 0"
## [1] "Binomial probability. P(X=x): 0.4219"
## [1] "Cumulative probability. P(X<x): 0"
## [1] "Cumulative probability. P(X<=x): 0.4219"
## [1] "Cumulative probability. P(X>x): 0.5781"
## [1] "Cumulative probability. P(X>=x): 1"
```

- at least one will neither have the disease nor be a carrier? **0.578125**

```
binom_summary(1, 3, 0.25)
```




```
## [1] "mu: 0.75"
## [1] "sigma: 0.75"
## [1] "Trials: 3"
## [1] "Successes: 1"
## [1] "Binomial probability. P(X=x): 0.4219"
## [1] "Cumulative probability. P(X<x): 0.4219"
## [1] "Cumulative probability. P(X<=x): 0.8438"
## [1] "Cumulative probability. P(X>x): 0.1562"
## [1] "Cumulative probability. P(X>=x): 0.5781"
```

- the first child with the disease will be the 3^{rd} child?
  - **Geometric by hand: 0.140625**
  - `dgeom()`**. Careful now, the first parameter of the function is the number of *failures*. : 0.140625**

p. , ex. 4.25

The formula for the number of ways to arrange n objects is $n! = n \times (n-1) \times \cdots \times 2 \times 1$. This exercise walks you through the derivation of this formula for a couple of special cases.

A small company has five employees: Anna, Ben, Carl, Damian, and Eddy. There are five parking spots in a row at the company, none of which are assigned, and each day the employees pull into a random parking spot. That is, all possible orderings of the cars in the row of spots are equally likely.

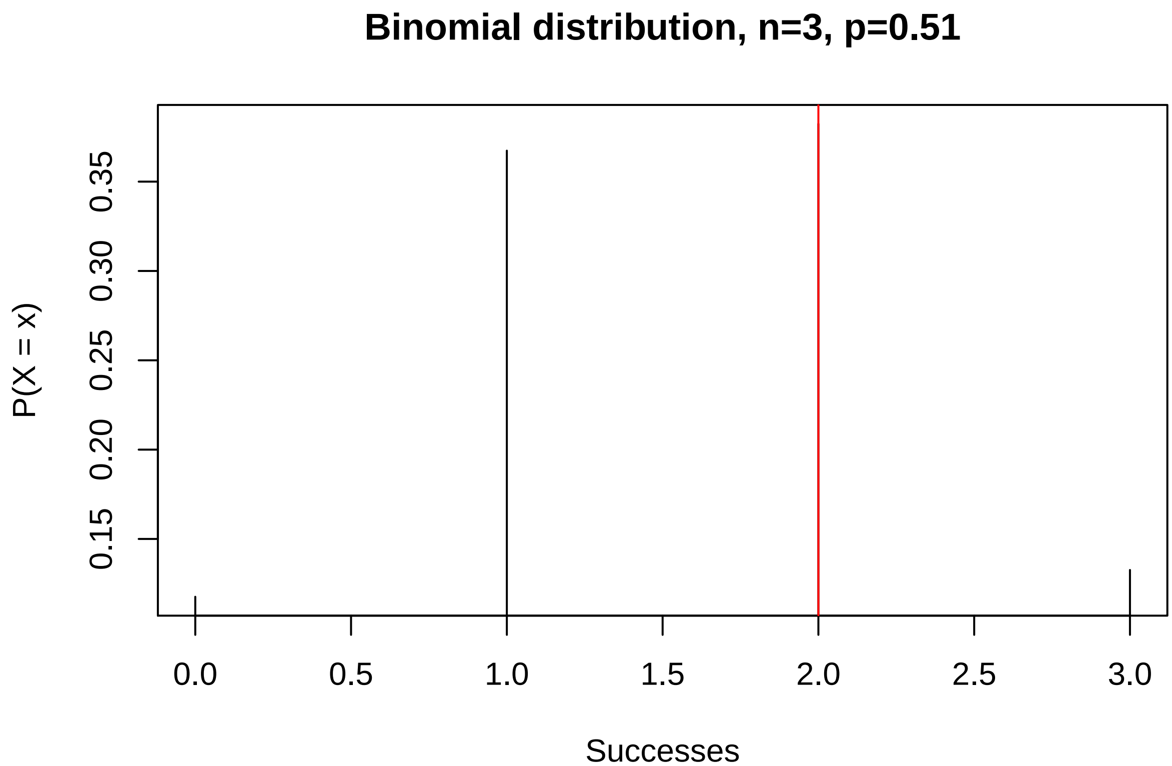
- On a given day, what is the probability that the employees park in alphabetical order? **This process is sampling without replacement. Therefore the probability is:** $\frac{1}{5} \cdot \frac{1}{4} \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{1}$ **= 0.0083333**
- If the alphabetical order has an equal chance of occurring relative to all other possible orderings, how many ways must there be to arrange the five cars? **A probability distribution must sum to 1. Since this ordering is equiprobable, the number of possible orderings must equal the inverse of the probability of a single ordering.** $\frac{1}{\frac{1}{5!}} = 5! =$ **120**
- Now consider a sample of 8 employees instead. How many possible ways are there to order these 8 employees' cars? $8! =$ **40320**

p. , ex. 4.26

While it is often assumed that the probabilities of having a boy or a girl are the same, the actual probability of having a boy is slightly higher at 0.51. Suppose a couple plans to have 3 kids.

- Use the binomial model to calculate the probability that two of them will be boys. **0.382347**

```
binom_summary(2, 3, 0.51)
```




```
## [1] "mu: 1.53"
## [1] "sigma: 0.865852181379709"
## [1] "Trials: 3"
## [1] "Successes: 2"
## [1] "Binomial probability. P(X=x): 0.3823"
## [1] "Cumulative probability. P(X<x): 0.485"
```

```
## [1] "Cumulative probability. P(X<=x): 0.8673"
## [1] "Cumulative probability. P(X>x): 0.1327"
## [1] "Cumulative probability. P(X>=x): 0.515"
```

- Write out all possible orderings of 3 children, 2 of whom are boys. Use these scenarios to calculate the same probability from part (a) but using the addition rule for disjoint outcomes. Confirm that your answers from parts (a) and (b) match.
  - $\Omega =$ **{BBG, BGB, GBB}**
  - $P(\omega_1)+P(\omega_2)+P(\omega_3) = 0.51\times0.49\times0.51+0.51\times0.49\times0.51+0.49\times0.51\times0.51 = 3\times0.51^2\times0.49$ **= 0.382347**
- If we wanted to calculate the probability that a couple who plans to have 8 kids will have 3 boys, briefly describe why the approach from part (b) would be more tedious than the approach from part (a). **Well, it is and it isn't. It would be more tedious if you calculate 8 terms. On the other hand, I simplified it algebraically and it isn't harder in that case.** $8\times0.51^3\times0.49^{8-3}$ **= 0.0299765 and `dbinom(3, 8, 0.51)` = 0.2098355. This isn't correct. Work Out the factorial division. This is an n choose k problem.**

p. , ex. 4.27

Calculate the following probabilities and indicate which probability distribution model is appropriate in each case. You roll a fair die 5 times. What is the probability of rolling

- the first 6 on the fifth roll? **Negative binomial.**
  - **Hand rolled: 0.0001286**
  - **Hand rolled 2: 0.0001286**
  - **`dnbinom()`: 0.0043413**
  - **Something is wrong.**
  - **I'll say. You want the geometric distribution.**
  - **Hand rolled: 0.0803755r**
  - **`dgeom()`: 0.0803755**
- exactly three 6s? **Binomial: 0.0321502**
- the third 6 on the fifth roll? **Negative binomial.**
  - **Michael Foley's beautiful demo. Do I know him from class? This: R function dnbinom(x, size, prob) is the probability of x failures prior to the rth success (note the difference) when the probability of success is prob.**
  - **0.0192901**
  - **0.0192901**

p. , ex. 4.28

Calculate the following probabilities and indicate which probability distribution model is appropriate in each case. A very good darts player can hit the bull's eye (red circle in the center of the dart board) 65% of the time. What is the probability that he

- hits the bullseye for the 10^{th} time on the 15^{th} try? **Negative binomial.**
  - **Hand-rolled. 0.1415591**
  - **`dnbinom()`. 0.1415591**
- hits the bullseye 10 times in 15 tries? **Binomial. 0.2123387**
- hits the first bullseye on the third try? **Geometric. 0.079625**

p. , ex. 4.29

For a sociology class project you are asked to conduct a survey on 20 students at your school. You decide to stand outside of your dorm's cafeteria and conduct the survey on a random sample of 20 students leaving the cafeteria after dinner one evening. Your dorm is comprised of 45% males and 55% females.

- Which probability model is most appropriate for calculating the probability that the 4^{th} person you survey is the 2^{nd} female? Explain. **Negative binomial, because it describes the probability of observing the k^{th} success on the n^{th} trial.**
- Compute the probability from part (a). **0.1853002**
- The three possible scenarios that lead to 4^{th} person you survey being the 2^{nd} female are

$$\{M, M, F, F\}, \{M, F, M, F\}, \{F, M, M, F\}$$

  One common feature among these scenarios is that the last trial is always female. In the first three trials there are 2 males and 1 female. Use the binomial coefficient to confirm that there are 3 ways of ordering 2 males and 1 female.

  - **3**
  - **3**

- Use the findings presented in part (c) to explain why the formula for the coefficient for the negative binomial is $\binom{n-1}{k-1}$ while the formula for the binomial coefficient is $\binom{n}{k}$. **The final trial isn't part of the permutation. Its success is defined as a requirement.**

p. , ex. 4.30

A not-so-skilled volleyball player has a 15% chance of making the serve, which involves hitting the ball so it passes over the net on a trajectory such that it will land in the opposing team's court. Suppose that her serves are independent of each other.

- What is the probability that on the 10^{th} try she will make her 3^{rd} successful serve?
  - **Hand rolled: 0.0389501**
  - **Hand rolled 2: 0.0389501**
  - **`dnbinom()`: 0.0389501**
- Suppose she has made two successful serves in nine attempts. What is the probability that her 10^{th} serve will be successful? **0.15**
- Even though parts (a) and (b) discuss the same scenario, the probabilities you calculated should be different. Can you explain the reason for this discrepancy? **Luck doesn't get stored. Believing otherwise is known as The Gambler's Fallacy.**

p. , ex. 4.31

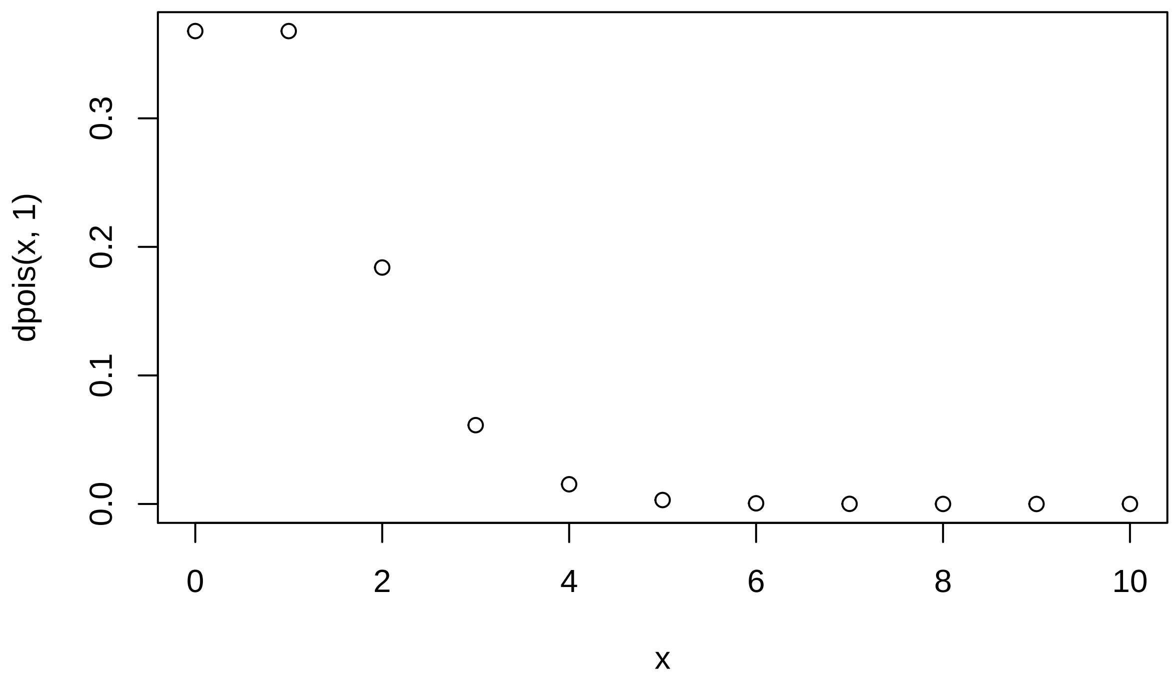
A coffee shop serves an average of 75 customers per hour during the morning rush.

- Which distribution have we studied that is most appropriate for calculating the probability of a given number of customers arriving within one hour during this time of day? **Poisson**
- What are the mean and the standard deviation of the number of customers this coffee shop serves in one hour during this time of day? $\mu = \lambda = 75$**.** $\sigma = \sqrt{\lambda} = \sqrt{75}$ **= 8.660254**
- Would it be considered unusually low if only 60 customers showed up to this coffee shop in one hour during this time of day? $Z = \frac{60-\mu}{\sigma}$ **=-1.7320508. No, the event falls within 2 standard deviations.**
- Calculate the probability that this coffee shop serves 70 customers in one hour during this time of day.
  - $P_{coffee}(70) = \frac{\lambda^k e^{-\lambda}}{k!} = \frac{\lambda^{70} e^{-70}}{75!}$ **= 0.0401603**
  - **0.0401603**

p. , ex. 4.32

A very skilled court stenographer makes one typographical error (typo) per hour on average.

```r
x <- 0:10
plot(x, dpois(x, 1))
```



- What probability distribution is most appropriate for calculating the probability of a given number of typos this stenographer makes in an hour? **Poisson.** - What are the mean and the standard deviation of the number of typos this stenographer makes? $\mu = 1$ $\sigma = 1$ - Would it be considered unusual if this stenographer made 4 typos in a given hour? $Z = \frac{4-1}{1} = 3$ Yes, 3 standard deviations from the mean is unusual. - Calculate the probability that this stenographer makes at most 2 typos in a given hour.

- **Hand rolled...**

```r
prob <- (1^0 * exp(-1)) / factorial(0) + (1^1 * exp(-1)) / factorial(1) + (1^2 * exp(-1)) / factorial(2
```

- **0.9196986**
- **`ppois()`: 0.9196986**

p. , ex. 4.33

For Monday through Thursday when there isn't a holiday, the average number of vehicles that visit a particular retailer between 2pm and 3pm each afternoon is 6.5, and the number of cars that show up on any given day follows a Poisson distribution.

- What is the probability that exactly 5 cars will show up next Monday?
  - **1 hour: 0.1453689**
  - **Are they asking for 8 hours?** 0.0015034**
- What is the probability that 0, 1, or 2 cars will show up next Monday between 2pm and 3pm? **0.0430359**
- There is an average of 11.7 people who visit during those same hours from vehicles. Is it likely that the number of people visiting by car during this hour is also Poisson? Explain. **Yes, the average equals λ. –WRONG**
  - **ANSWER: The book gives this answer. This lesson is important. I must think about the assumptions for each distribution. Every. Time.**

    The number of people per car is 11.7/6.5 = 1.8, meaning people are coming in small clusters. That is, if one person arrives, there's a chance that they brought one or more other people in their vehicle. This means individuals (the people) are not independent, even if the car arrivals are independent, and this breaks a core assumption for the Poisson distribution. That is, the number of people visiting between 2pm and 3pm would not follow a Poisson distribution.

p. , ex. 4.34

Occasionally an airline will lose a bag. Suppose a small airline has found it can reasonably model the number of bags lost each weekday using a Poisson model with a mean of 2.2 bags.

- What is the probability that the airline will lose no bags next Monday? **0.1108032**
- What is the probability that the airline will lose 0, 1, or 2 bags on next Monday? **0.6227137**
- Suppose the airline expands over the course of the next 3 years, doubling the number of flights it makes, and the CEO asks you if it's reasonable for them to continue using the Poisson model with a mean of 2.2. What is an appropriate recommendation? Explain. **The answer is no and there are two reasons.**
  1. The rate of bag losses (λ) is losses per weekday. If the flights are doubled, the best guess would be that the losses double. Therefore, the mean would be 4.4.
  2. Even an assumption of 4.4 is weak. What changes were made to infrastructure, personnel, and operations in order to scale to twice the number of flights? It would be better to measure bag losses for long enough to adopt a new mean.

p. , ex. 4.35

In the game of roulette, a wheel is spun and you place bets on where it will stop. One popular bet is that it will stop on a red slot; such a bet has an 18/38 chance of winning. If it stops on red, you double the money you bet. If not, you lose the money you bet. Suppose you play 3 times, each time with a $1 bet. Let Y represent the total amount won or lost. Write a probability model for Y.

- $Y = 2^k(\frac{18}{38})^k$. **0.8502697. Wrong: I didn't account for losses, which are calculated differently from wins.**
- **The book considers the sum of probabilities for: 0 wins + 1 win + 2 wins + 3 wins.**
  - 0 wins. -3. 0.1457938
  - 1 win. 0. 0.3936434
  - 2 wins. 3. 0.354279
  - 3 wins. 6. 0.1062837
  - The reason your probabilities were wrong the first time. You calculated the probability of each scenario okay, but you didn't multiply each by the number of ways that scenario could happen.

p. , ex. 4.36

The distribution of passenger vehicle speeds traveling on the Interstate 5 Freeway (I-5) in California is nearly normal with a mean of 72.6 miles/hour and a standard deviation of 4.78 miles/hour.

- What percent of passenger vehicles travel slower than 80 miles/hour? 93.92%
- What percent of passenger vehicles travel between 60 and 80 miles/hour? 0.52%
- How fast do the fastest 5% of passenger vehicles travel? 80.4624003 mph
- The speed limit on this stretch of the I-5 is 70 miles/hour. Approximate what percentage of the passenger vehicles travel above the speed limit on this stretch of the I-5. 29.32%

p. , ex. 4.37

Suppose a university announced that it admitted 2,500 students for the following year's freshman class. However, the university has dorm room spots for only 1,786 freshman students. If there is a 70% chance that an admitted student will decide to accept the offer and attend this university, what is the approximate probability that the university will not have enough dormitory room spots for the freshman class?

- **I failed to discern this is a normal approximation of a binomial.*
- **The book failed to discern that the interval is open on the bottom, at least in the description of the solution. The problem asked for the probability of insufficient insufficient rooms. With 1786 rooms, they run out on the 1787th student, not the 1786th. However, they got it when they computed a Z-score.**
- $\mu = np = 2500 \cdot 0.70 = 1750$
- $\sigma = \sqrt{npq} = \sqrt{2500 \cdot 0.70 \cdot 0.30} = 22.9128785$
- Probability $= 0.0580717$
- Probability (Z-score approach)

```
mu <- 2500 * 0.70
sigma <- sqrt(mu * 0.30)
Z <- (1787 - mu - 0.5) / sigma

mu
```

```
## [1] 1750
```

```
sigma
```

```
## [1] 22.91288
```

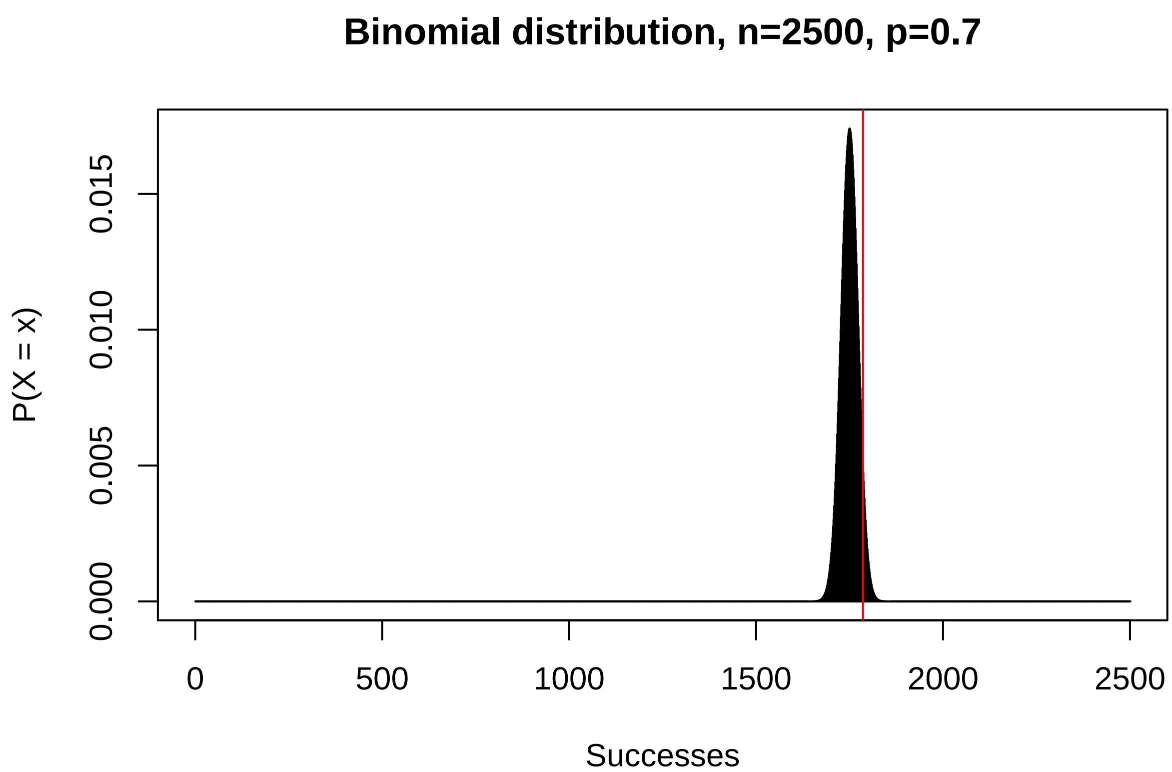
```
Z
```

```
## [1] 1.592991
```

```
p <- 1 - pnorm(Z)
p
```

```
## [1] 0.05558115
```

```r
binom_summary(1786, 2500, 0.70)
```




```
## [1] "mu: 1750"
## [1] "sigma: 22.9128784747792"
## [1] "Trials: 2500"
## [1] "Successes: 1786"
## [1] "Binomial probability. P(X=x): 0.0051"
## [1] "Cumulative probability. P(X<x): 0.9399"
## [1] "    Normal approximation. P(X<x): 0.9419"
## [1] "    Normal approximation w/ continuity correction. P(<x): 0.9394"
## [1] "Cumulative probability. P(X<=x): 0.9449"
## [1] "    Normal approximation. P(X<=x): 0.9419"
## [1] "    Normal approximation w/ continuity correction. P(<=x): 0.9444"
## [1] "Cumulative probability. P(X>x): 0.0551"
## [1] "    Normal approximation. P(X>x): 0.0581"
## [1] "    Normal approximation w/ continuity correction. P(X>x): 0.0556"
## [1] "Cumulative probability. P(X>=x): 0.0601"
## [1] "    Normal approximation. P(X>=x): 0.0581"
## [1] "    Normal approximation w/ continuity correction. P(X>=x): 0.0606"
```

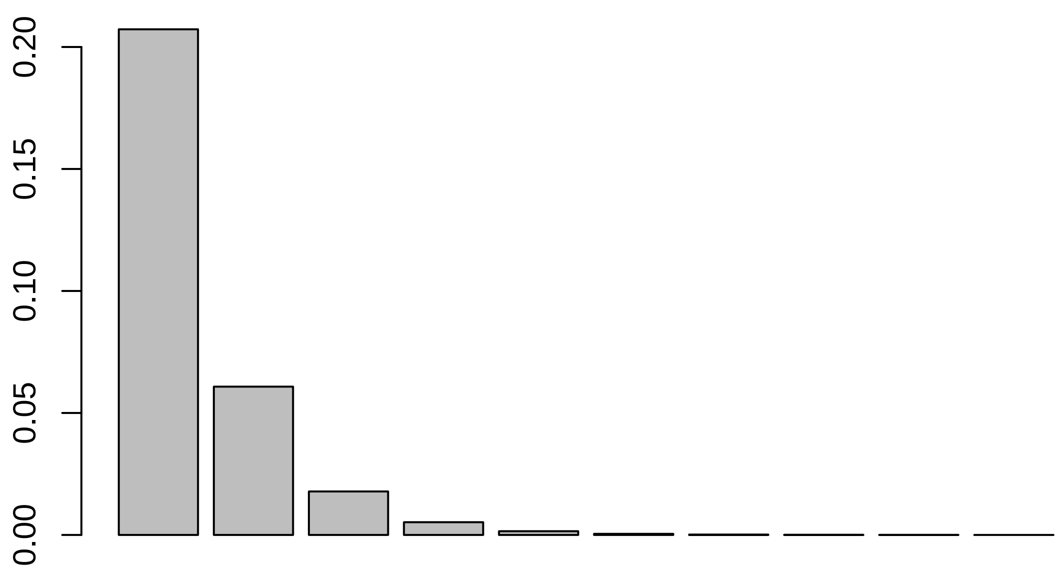
N.B.: I'm leaving the binomial for now and its normal approximation. There are too many questions about what to do at the boundaries of intervals. I think the Z-score approach is presented inconsistently in the text. I think my summary function was useful, but it has not paid off yet because of the inconsistency. I'll return to this with a fresh take on it when I go through my next workbook.

p. , ex. 4.38

Exercise 4.36 states that the distribution of speeds of cars traveling on the Interstate 5 Freeway (I-5) in California is nearly normal with a mean of 72.6 miles/hour and a standard deviation of 4.78 miles/hour. The speed limit on this stretch of the I-5 is 70 miles/hour.

- A highway patrol officer is hidden on the side of the freeway. What is the probability that 5~cars pass and none are speeding? Assume that the speeds of the cars are independent of each other.
  - 0.0021684
- On average, how many cars would the highway patrol officer expect to watch until the first car that is speeding? What is the standard deviation of the number of cars he would expect to watch?

```r
# Probability of speedig
p <- pnorm(70, 72.6, 4.78, lower.tail = F)
n <- 1:10
barplot(dgeom(n, p))
```




```r
# Average is expectation is mean.
mu <- 1 / p
mu
```

```
## [1] 1.414915
```

```
sd <- sqrt((1 - p)/ p^2)
sd
```

```
## [1] 0.7662046
```

- Expected cars to watch: 1.414915.
- σ: 0.7662046.

p. , ex. 4.39

Suppose a newspaper article states that the distribution of auto insurance premiums for residents of California is approximately normal with a mean of $1,650. The article also states that 25% of California residents pay more than $1,800.

- What is the Z-score that corresponds to the top 25% (or the 75^{th} percentile) of the standard normal distribution?

```
Z <- qnorm(0.75)
```

- 0.6744898.
- What is the mean insurance cost? What is the cutoff for the 75th percentile?
  - $1650.
  - $1800.
- Identify the standard deviation of insurance premiums in California.

```
sd <- (1800 - 1650) / Z
```

- 222.3903328. **ANSWER: The book differs from mine a little because they rounded Z. I kept the precision.

p. , ex. 4.40

SAT scores (out of 1600) are distributed normally with a mean of 1100 and a standard deviation of 200. Suppose a school council awards a certificate of excellence to all students who score at least 1350 on the SAT, and suppose we pick one of the recognized students at random. What is the probability this student's score will be at least 1500? (The material covered in Section 3.2 on conditional probability would be useful for this question.)

- $P(A|B) = \frac{P(A \text{ } and \text{ } B)}{P(B)} = \frac{P(X \geq 1500)}{P(X \geq 1300)}$ Note that in this case the probability of A and B is the same as the probability of B because of their given inequalities.
- 0.2153354

p. , ex. 4.41

Married women. The American Community Survey estimates that 47.1% of women ages 15 years and over are married.

(a) We randomly select three women between these ages. What is the probability that the third woman selected is the only one who is married?

- 0.1318051

(b) What is the probability that all three randomly selected women are married?

- 0.1044871

(c) On average, how many women would you expect to sample before selecting a married woman? What is the standard deviation?

- $\mu = \frac{1}{p} = 2.1231423$.
- $\sigma = \sqrt{\frac{1-p}{p^2}} = 1.544212$

(d) If the proportion of married women was actually 30%, how many women would you expect to sample before selecting a married woman? What is the standard deviation?

- $\mu = 3.3333333$.
- $\sigma = 2.7888668$

(e) Based on your answers to parts (c) and (d), how does decreasing the probability of an event affect the mean and standard deviation of the wait time until success?

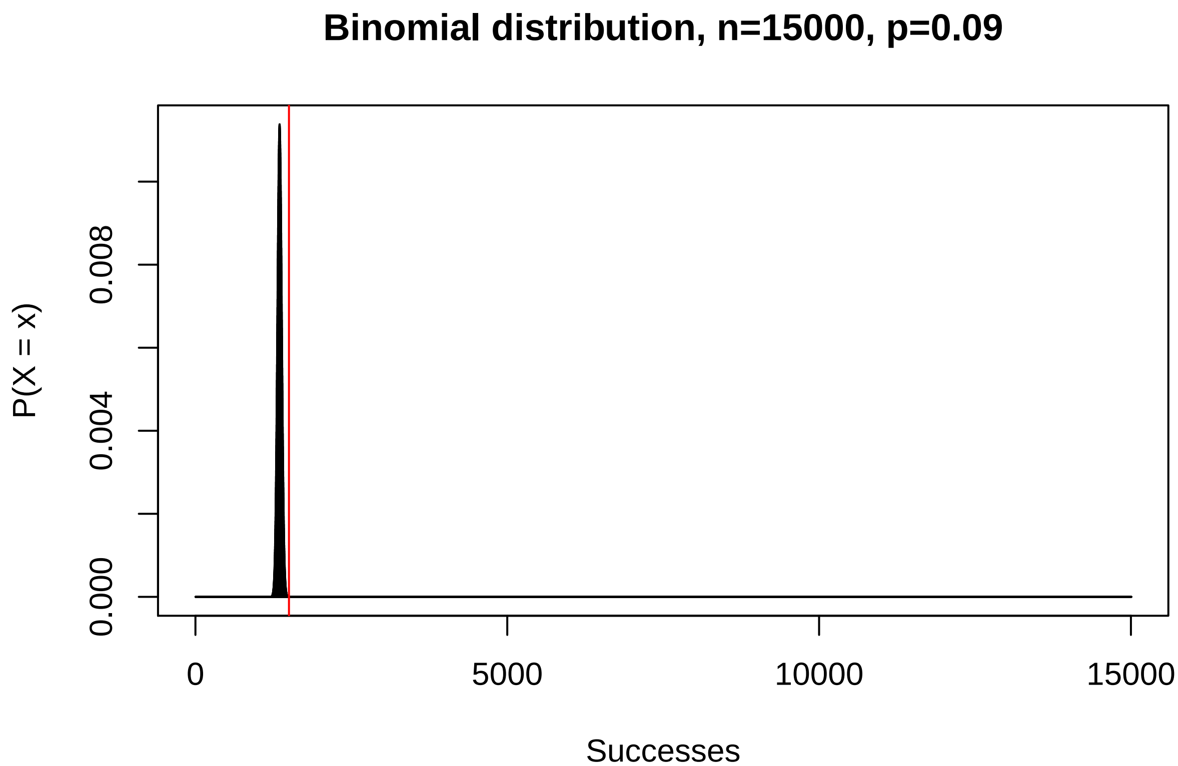
- Rarer events take a longer time before successes. Mean and standard deviation are higher.
- **I got all of these!**

p. , ex. 4.42

Pew Research reported that the typical response rate to their surveys is only 9%. If for a particular survey 15,000 households are contacted, what is the probability that at least 1,500 will agree to respond?

- Binomial: 0.0000133

```
binom_summary(successes = 1500, trials = 15000, prob = 0.09)
```




```
## [1] "mu: 1350"
## [1] "sigma: 35.0499643366438"
## [1] "Trials: 15000"
## [1] "Successes: 1500"
## [1] "Binomial probability. P(X=x): 0"
## [1] "Cumulative probability. P(X<x): 1"
## [1] "    Normal approximation. P(X<x): 1"
## [1] "    Normal approximation w/ continuity correction. P(<x): 1"
## [1] "Cumulative probability. P(X<=x): 1"
## [1] "    Normal approximation. P(X<=x): 1"
## [1] "    Normal approximation w/ continuity correction. P(<=x): 1"
## [1] "Cumulative probability. P(X>x): 0"
## [1] "    Normal approximation. P(X>x): 0"
## [1] "    Normal approximation w/ continuity correction. P(X>x): 0"
## [1] "Cumulative probability. P(X>=x): 0"
## [1] "    Normal approximation. P(X>=x): 0"
## [1] "    Normal approximation w/ continuity correction. P(X>=x): 0"
```

p. , ex. 4.43

Suppose weights of the checked baggage of airline passengers follow a nearly normal distribution with mean 45 pounds and standard deviation 3.2 pounds. Most airlines charge a fee for baggage that weigh in excess of 50 pounds. Determine what percent of airline passengers incur this fee.

- 0.0590851

p. , ex. 4.44

Heights of 10 year olds, regardless of gender, closely follow a normal distribution with mean 55 inches and standard deviation 6~inches.

- What is the probability that a randomly chosen 10 year old is shorter than 48 inches?
  - 0.1216725
- What is the probability that a randomly chosen 10 year old is between 60 and 65 inches?
  - 0.9522096
- If the tallest 10% of the class is considered "very tall", what is the height cutoff for "very tall"?
  - 62.6893094

p. , ex. 4.45

Suppose you're considering buying your expensive chemistry textbook on Ebay. Looking at past auctions suggests that the prices of this textbook follow an approximately normal distribution with mean \$89 and standard deviation \$15.

- What is the probability that a randomly selected auction for this book closes at more than \$100?
  - 0.2316776
- Ebay allows you to set your maximum bid price so that if someone outbids you on an auction you can automatically outbid them, up to the maximum bid price you set. If you are only bidding on one auction, what are the advantages and disadvantages of setting a bid price too high or too low? What if you are bidding on multiple auctions?
  - If you set it too high, you might go head to head with another bidder who did the same thing and you will both hit your maximum bid in an instant, rather than reevaluating your bid each time you see a response. If you set it too low, you lose the auction. If there are multiple auctions, those effects are multiplied.
- If you watched 10 auctions, roughly what percentile might you use for a maximum bid cutoff to be somewhat sure that you will win one of these ten auctions? Is it possible to find a cutoff point that will ensure that you win an auction?
  - The book had some opinions about these next answers. They're only opinions, and I'm not really committed to completing these parts.
- If you are willing to track up to ten auctions closely, about what price might you use as your maximum bid price if you want to be somewhat sure that you will buy one of these ten books?

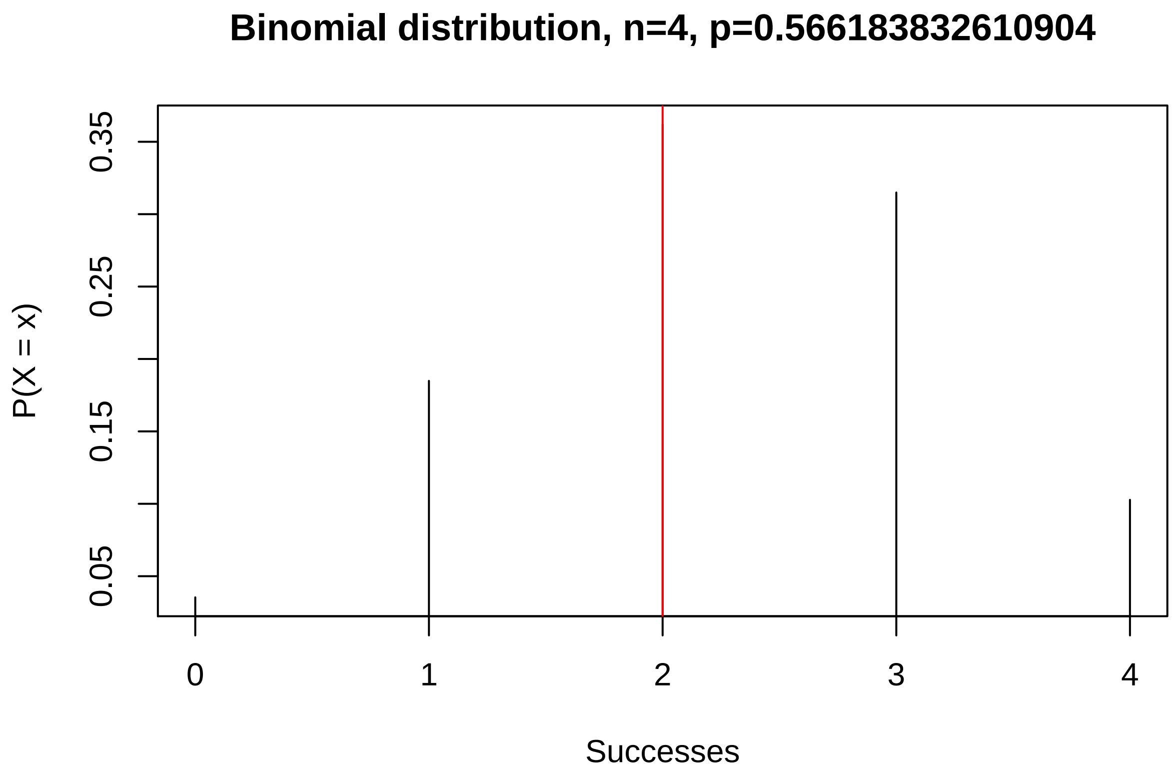
p. , ex. 4.46

Heights of 10 year olds, regardless of gender, closely follow a normal distribution with mean 55 inches and standard deviation 6~inches.

- The height requirement for *Batman the Ride* at Six Flags Magic Mountain is 54 inches. What percent of 10 year olds cannot go on this ride?
  - 43.3816167%
- Suppose there are four 10 year olds. What is the chance that at least two of them will be able to ride *Batman the Ride*?

- 1. **Doesn't look right.**

```r
binom_summary(2, 4, 1 - pnorm(54, 55, 6))
```




```
## [1] "mu: 2.26473533044361"
## [1] "sigma: 0.991200686644096"
## [1] "Trials: 4"
## [1] "Successes: 2"
## [1] "Binomial probability. P(X=x): 0.362"
## [1] "Cumulative probability. P(X<x): 0.2203"
## [1] "Cumulative probability. P(X<=x): 0.5823"
## [1] "Cumulative probability. P(X>x): 0.4177"
## [1] "Cumulative probability. P(X>=x): 0.7797"
```

- Suppose you work at the park to help them better understand their customers' demographics, and you are counting people as they enter the park. What is the chance that the first 10 year old you see who can ride *Batman the Ride* is the 3rd 10 year old who enters the park?
  - 0.204944
- What is the chance that the fifth 10 year old you see who can ride *Batman the Ride* is the 12th 10 year old who enters the park?
  - 0.0998584

p. , ex. 4.47

Heights of 10 year olds, regardless of gender, closely follow a normal distribution with mean 55 inches and standard deviation 6~inches.

- What fraction of 10 year olds are taller than 76 inches?
  - 0.0002326 **Correct**
- If there are 2,000 10 year olds entering Six Flags Magic Mountain in a single day, then compute the expected number of 10 year olds who are at least 76 inches tall. (You may assume the heights of the 10-year olds are independent.)
  - 0.4652582 **Looks correct. The book rounded the probability.**
- Using the binomial distribution, compute the probability that 0 of the 2,000 10 year olds will be at least 76 inches tall.
  - 0.372061 **Looks incorrect.**
- The number of 10 year olds who enter Six Flags Magic Mountain and are at least 76 inches tall in a given day follows a Poisson distribution with mean equal to the value found in part (b). Use the Poisson distribution to identify the probability no 10 year old will enter the park who is 76 inches or taller.

p. , ex. 4.48

In a multiple choice quiz there are 5 questions and 4 choices for each question (a, b, c, d). Robin has not studied for the quiz at all, and decides to randomly guess the answers. What is the probability that

- the first question she gets right is the 3^{rd} question?
- she gets exactly 3 or exactly 4 questions right?
- she gets the majority of the questions right?

I've just had enough and I want to go to another chapter.