

# 1 Chapter 3

1. Since  $P(A_1 \cap A_2) = 0.64 = (0.8)(0.8) = P(A_1)P(A_2)$ , we can conclude that  $A_1$  and  $A_2$  are independent; hence, the answer is (c).
2. (a)  $P(A \cap B) = P(B \cap A) = P(B|A) = (0.95)(0.05) = 0.0475$ .  
 (b)  $P(B) = P(B \cap A) + P(B \cap A^c) = 0.0475 + (0.03)(1 - 0.05) = 0.076$ .  
 (c)  $P(A|B) = P(A \cap B) / P(B) = 0.0475 / 0.076 = 0.625$ .
3. Let  $A$  be the event that an adult gets the flu and let  $B$  be the event that an adult gets the flu shot.  
 (a)  $P(A \cap B) = P(A|B)P(B) = (0.1)(0.42) = 0.042$ .  
 (b)  $P(A) = P(A \cap B) + P(A \cap B^c) = 0.042 + (0.7)(1 - 0.42) = 0.448$ .
4. Let  $X$  denote the number of people who have asthma. Then  $X \sim \text{Binomial}(n = 50, p = 0.2)$ . (Think about why!)

- (a)  $P(X = 19) = \binom{50}{19}(0.2)^{19}(0.8)^{50-19} \approx 0.001579$ .
- (b) The standard deviation is  $\sigma = \sqrt{np(1-p)} = \sqrt{(50)(0.2)(0.8)} = 2.828427$  and the mean/expected value is  $\mu = np = (50)(0.2) = 10$ . Hence, the  $z$ -score is  $(19 - 10) / 2.828427 \approx 3.181981$  which implies that 19 is a little over three standard deviations above the mean.
- (c) Using the empirical rule, we have that  $P(X \geq 19) \approx P(X \geq \mu + 3\sigma) \approx (1 - 0.997) / 2 = 0.003 / 2 = 0.0015$ . (Draw a picture!). The exact answer is

```
1 - pbinom(18, size = 50, prob = 0.2)
## [1] 0.002511203
```

5.  $E[X] = \sum_{r=1}^3 rP(X=r) = (1)(1/3) + (2)(1/3) + (3)(1/3) = 2$  envelopes.
6. (a)  $E[X] = \mu = \sum_{r=0}^4 rP(X=r) = (0)(0.2) + (1)(0.3) + (2)(0.3) + (3)(0.1) + (4)(0.1) = 1.6$  egg masses.  
 (b)  $Var[X] = \sum_{r=0}^4 (r - \mu)^2 P(X=r) = (0 - 1.6)^2 (0.2) + (1 - 1.6)^2 (0.3) + (2 - 1.6)^2 (0.3) + (3 - 1.6)^2 (0.1) + (4 - 1.6)^2 (0.1) = 1.44$ . Hence, the standard deviation is  $\sqrt{1.44} = 1.2$  egg masses.
7. Let  $A$  be the event that a subject is taking the drug (then  $A^c$  is the event that the subject is taking the placebo) and let  $B$  be the event that a subject improves.  
 (a)  $P(B \cap A) = P(B|A)P(A) = (0.6)(0.5) = 0.3$ .  
 (b)  $P(B) = P(B \cap A) + P(B \cap A^c) = 0.3 + (0.35)(0.5) = 0.475$ .

8. Let  $Y$  be a random variable denoting the number of chickens out of 20 with the bird flu. It then follows that  $Y \sim \text{Binomial}(n = 20, p = 0.1)$ .
- (a)  $P(Y = 5) = \binom{20}{5}(0.1)^5(0.9)^{15} = (15504)(10^{-5})(0.2058911) \approx 0.031921$ .
  - (b)  $E[Y] = np = (20)(0.1) = 2$  chickens.
  - (c)  $\sqrt{\text{Var}[Y]} = \sqrt{np(1-p)} = \sqrt{(20)(0.1)(0.9)} \approx 1.3416$  chickens.

9. Let  $X$  be a random variable denoting the number of frog eggs that hatch out of 100. Then,  $X \sim \text{Binomial}(n = 100, p = 0.87)$ . (Since the frog eggs hatch independently of each other!)

- (a)  $P(X = 80) = \binom{100}{80}(0.87)^{80}(0.13)^{20} \approx 0.01477606$ . You should be able to do this with a calculator, but in R we would just use

```
dbinom(80, size = 100, prob = 0.87)
## [1] 0.01477606
```

- (b) For a binomial random variable,  $E[X] = np = (100)(0.87) = 87$  eggs.
- (c) To use the empirical rule, we must first calculate the standard deviation of  $X$ . The standard deviation of a binomial random variable is given by  $\sqrt{\text{Var}[X]} = \sqrt{np(1-p)} = \sqrt{3.31} = 3.363034$ . Computing the  $z$ -score yields

$$Z = \frac{77 - 87}{3.363034} = -2.973505 \approx -3.$$

In other words, 77 eggs is about three standard deviations below the mean. So,  $P(X \leq 77) \approx$  the probability of being three or more standard deviations below the mean  $\approx 0.003/2 = 0.0015$ .

10. (a)  $\binom{20}{5} = \frac{20!}{5!(20-5)!} = \frac{(20)(19)(18)(17)(16)(15!)}{5!(15!)} = 15504$ .
- (b)  $P(X = 0) = \binom{5}{0} (7/20)^0 (1 - 7/20)^5 = 0.116$ .
11. Let  $X$  be a random variable that represents the number of white croaker fish with high mercury levels out of  $n = 100$ . It follows that  $X \sim \text{Binomial}(n = 100, p = 0.4)$ .
- (a)  $P(X = 100) = (0.4)^{100} \approx 1.6069 \times 10^{-40}$ .
  - (b)  $P(X = 45) = \binom{100}{45}(0.4)^{45}(1 - 0.4)^{55} = 0.0478$ .
  - (c)  $E[X] = np = (100)(0.4) = 40$  fish and  $\sqrt{\text{Var}[X]} = \sqrt{np(1-p)} = \sqrt{(100)(0.4)(0.6)} = 4.898979$  fish.
  - (d)  $P(X \geq 55) = 1 - P(X \leq 54) = 1 - [P(X = 0) + P(X = 1) + \dots + P(X = 54)]$ .  
In R, we get

```
1 - pbinom(54, size = 100, prob = 0.4)
## [1] 0.001710927
```

## 2 Chapter 4

1. No (it looks bimodal).
2. The best answer is (d); bimodal.
3. The population might consist of both males and females, and each of these subpopulations probably has its own mean.
4. The best answer is (a); 31.
5. The best answer is (c); 0.22.
6. The best answer is (e); 0.94.
7. It will remain the same. Go back to the properties about correlation and linear transformations!
8. The best answer is (d); 0.27.
9. The best answer is (d); 0.58.
10. Use the fact that  $X \sim N(\mu = 5.28, \sigma = 0.4)$ .

(a)

$$\begin{aligned}
 P(X > 5.4) &= 1 - P(X \leq 5.4) \\
 &= 1 - P\left(\frac{X - 5.28}{0.4} < \frac{5.4 - 5.28}{0.4}\right) \\
 &= 1 - P\left(Z < \frac{5.4 - 5.28}{0.4}\right) \\
 &= 1 - P(Z < 0.3) \\
 &= 1 - \Phi(0.3)
 \end{aligned}$$

In R, we get

```
1 - pnorm(0.3)
## [1] 0.3820886
```

(b)

$$\begin{aligned} P(5 < X < 6) &= P(X < 6) - P(X < 5) \\ &= P\left(\frac{X - 5.28}{0.4} < \frac{6 - 5.28}{0.4}\right) - P\left(\frac{X - 5.28}{0.4} < \frac{5 - 5.28}{0.4}\right) \\ &= P\left(Z < \frac{6 - 5.28}{0.4}\right) - P\left(Z < \frac{5 - 5.28}{0.4}\right) \\ &= P(Z < 1.8) - P(Z < -0.7) \\ &= \Phi(1.8) - \Phi(-0.7) \end{aligned}$$

In R, we get

```
pnorm(1.8) - pnorm(-0.7)
## [1] 0.722106
```

(c) The general formula for the  $p$ -th percentile, denoted  $x_p$ , of a normal distribution with mean  $\mu$  and standard deviation  $\sigma$  is

$$x_p = \mu + \sigma z_p,$$

where  $z_p$  is the  $p$ -th percentile of a standard normal distribution (which we can obtain using `qnorm(p)` in R). Hence, the 95-th percentile is  $x_{0.95} = 5.28 + 0.4z_{0.95}$ . Using `qnorm(0.95)` in R, we obtain  $x_{0.95} = 5.28 + 0.4(1.644854) = 5.937941$ .

(d) Since the data are a random sample from a normal distribution, we know that the sample mean also has a normal distribution; in particular,  $\bar{X} \sim N(\mu = 5.28, \sigma = 0.4/\sqrt{50})$ . Hence,

$$\begin{aligned} P(\bar{X} > 5.4) &= 1 - P(\bar{X} \leq 5.4) \\ &= 1 - P\left(\frac{\bar{X} - 5.28}{0.4/\sqrt{50}} < \frac{5.4 - 5.28}{0.4/\sqrt{50}}\right) \\ &= 1 - P\left(Z < \frac{5.4 - 5.28}{0.4/\sqrt{50}}\right) \\ &= 1 - P(Z < 2.12132) \\ &= 1 - \Phi(2.12132) \end{aligned}$$

In R, we get

```
1 - pnorm(2.12132)
## [1] 0.01694744
```

11. Use the fact that  $X \sim N(\mu = 170, \sigma = 20)$ .

(a)

$$\begin{aligned}P(X > 200) &= 1 - P(X \leq 200) \\&= 1 - P\left(\frac{X - 170}{20} < \frac{200 - 170}{20}\right) \\&= 1 - P\left(Z < \frac{200 - 170}{20}\right) \\&= 1 - P(Z < 1.5) \\&= 1 - \Phi(1.5)\end{aligned}$$

In R, we get

```
1 - pnorm(1.5)
## [1] 0.0668072
```

(b) Using the fact that  $\bar{X} \sim N(\mu = 170, \sigma = 20/\sqrt{20})$ , we get

$$\begin{aligned}P(\bar{X} > 200) &= 1 - P(\bar{X} \leq 200) \\&= 1 - P\left(\frac{\bar{X} - 170}{20/\sqrt{20}} < \frac{200 - 170}{20/\sqrt{20}}\right) \\&= 1 - P\left(Z < \frac{200 - 170}{20/\sqrt{20}}\right) \\&= 1 - P(Z < 6.708204) \\&= 1 - \Phi(6.708204)\end{aligned}$$

In R, we get

```
1 - pnorm(6.708204)
## [1] 9.851675e-12
```

(c)  $x_{0.95} = \mu + \sigma z_{0.95} = 170 + 20(1.644854) = 202.8971$  (mg/dL).

### 3 Chapter 5

1. Using the central limit theorem,  $\bar{X} \sim N(\mu = 19, \sigma = 7.8/\sqrt{30})$ . So,

$$\begin{aligned}P(\bar{X} > 21.3) &= 1 - P(\bar{X} \leq 21.3) \\&= 1 - P\left(\frac{\bar{X} - 19}{7.8/\sqrt{30}} < \frac{21.3 - 19}{7.8/\sqrt{30}}\right) \\&= 1 - P\left(Z < \frac{21.3 - 19}{7.8/\sqrt{30}}\right) \\&= 1 - P(Z < 1.615079) \\&= 1 - \Phi(1.615079)\end{aligned}$$

In R, we get

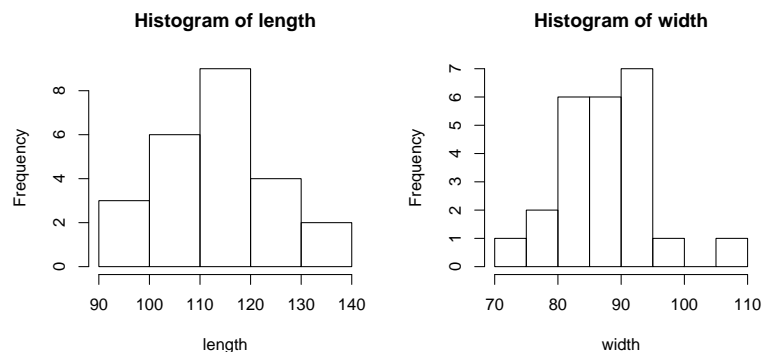
```
1 - pnorm(1.615079)
## [1] 0.05314679
```

2. (a) Running the R script `maleturtle.R`, a 99% confidence interval for the mean carapace length is (106.6246, 120.1254).
- (b) Running the R script `maleturtle.R`, a 99% confidence interval for the mean carapace width is (84.23794, 92.34539).
- (c) Based on the plot, it seems reasonable that this sample belongs to the Painted Turtle species (the true mean happens to be captured in the 99% confidence intervals).
- (d) Narrower, since we are decreasing our confidence.
- (e) The normal approximation does not seem unreasonable here, but more data is needed to get a clearer picture. In R, try

```
# Carapace length
length <- c(93, 94, 96, 101, 102, 103, 104, 106, 107,
            112, 113, 114, 116, 117, 117, 119, 120, 120, 121, 125,
            127, 128, 131, 135)

# Carapace width
width <- c(74, 78, 80, 84, 85, 81, 83, 83, 82, 89, 88,
           86, 90, 90, 91, 93, 89, 93, 95, 93, 96, 95, 95, 106)

# Histograms
par(mfrow = c(1, 2))
hist(length, br = 5)
hist(width, br = 5)
```



3. In R, you could use

```
# Path to data set on my laptop
path <- "C:\\Users\\greenweb\\Desktop\\Filing cabinet\\STT 6300\\Data sets\\bodytemp.csv"

# If you don't know how to find this, then just use: path <- file.choose()

# Load the data
bodytemp <- read.csv(path, header = TRUE)

# Temperature variable
temp <- bodytemp$temp

# Pulse rate
pulse <- bodytemp$pulse

# Part a)
t.test(temp, conf.level = 0.95)$conf.int

## [1] 98.12200 98.37646
## attr("conf.level")
## [1] 0.95

# Part b)
t.test(temp, conf.level = 0.99)$conf.int

## [1] 98.08111 98.41735
## attr("conf.level")
## [1] 0.99

# Part c)
#
# The 99% confidence interval for mean temperature is wider
# than the corresponding 95% confidence interval.

# Part d)
#
# No, since 98.6 is outside the range of both confidence
# intervals.

# Part e)
t.test(pulse, conf.level = 0.90)$conf.int

## [1] 72.73537 74.78771
## attr("conf.level")
## [1] 0.9
```

```
# Part f)
#
# False! Go back and read how we interpret confidence
# intervals!
```

4. The correct answer is c). Take a hard look at b) and try to determine why it is not the correct answer.
5. Skip. Extra credit.
6. The correct answer is c). For a given sample, the more confident you want to be, the wider your interval will be and vice versa.
7. The correct answer is f).

## 4 Chapter 6

**Problem 1** 1.  $H_0 : \mu = 1.2$  vs.  $H_1 : \mu \neq 1.2$

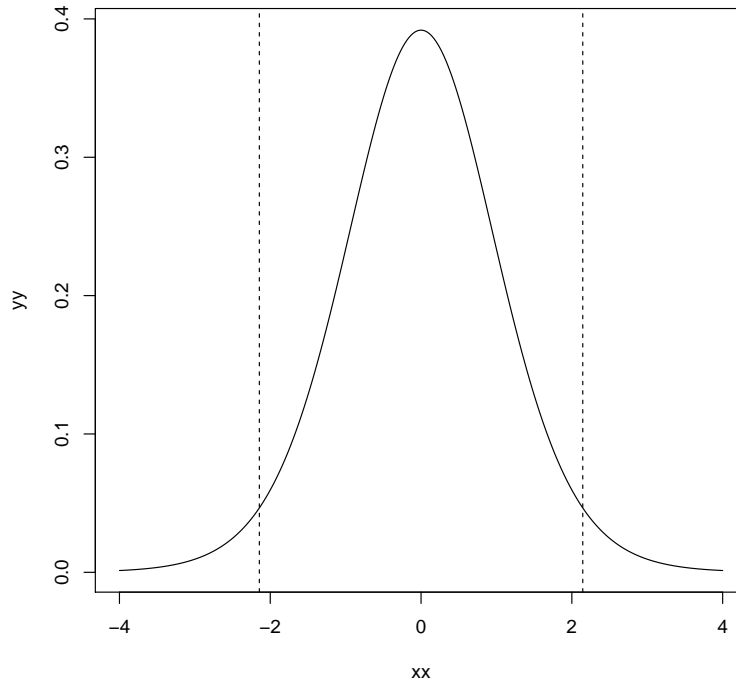
2. The type I error refers to the decision to reject the null hypothesis when it is true. In this case, concluding that the mean response time for rats injected with a unit dose of the experimental drug differs from 1.2 seconds, when in fact it does not.
3. The type II error refers to the decision to not reject the null hypothesis when it is false. In this case, concluding that the mean response time for rats injected with a unit dose of the experimental drug does not differ from 1.2 seconds, when in fact it does.
4. The test statistic is

$$\begin{aligned} t &= \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \\ &= \frac{1.05 - 1.2}{0.5/\sqrt{15}} \\ &= -1.161895 \end{aligned}$$

5. 

```
xx <- seq(from = -4, to = 4, length = 500)
yy <- dt(xx, df = 14)
plot(xx, yy, type = "l")
abline(v = c(-1, 1) * qt(0.975, df = 14), lty = 2)
```





6. Fail to reject  $H_0$  since  $|t| = 1.161895$  is not in the rejection region.

**Problem 2** The correct answer is b) (since  $0.01 < p \leq 0.05$ ).

**Problem 4** 1. The correct answer is that  $n$  will go UP!

2. The correct answer is that  $n$  will go DOWN!

3. The correct answer is that  $n$  will go UP!

**Problem 6** Here,  $\mu$  is the mean level of serum acid phosphatase in prostate cancer patients where the cancer has spread to surrounding lymph nodes. We wish to test the hypothesis

$$H_0 : \mu = 0.645 \quad \text{vs.} \quad H_1 : \mu > 0.645$$

The observed  $t$ -statistic is 2.5616 and the observed significance level (i.e., the  $p$ -value) is 0.009542. Hence, we would reject the null hypothesis at the 0.05 level of significance and conclude that the mean level of serum acid phosphatase in prostate cancer patients where the cancer has spread to surrounding lymph nodes is elevated.