R Introduction

We have course material to introduce you to R. You can find the pages we handed out in class plus the rewritten Appendices to the Course Notes from the Stat/For/Hort 571 Web page,

`http://www.stat.wisc.edu/~st571-1`

Click on the heading "R Software Introduction for Stat 571" for the overview or on "R Appendices for the Stat/For/Hort 571 Course Notes" to get the Appendices. Please bring up any questions you might have about installing or using R with your TA or your instructor.

Review

1. Random variable **X**: variable takes its values according to the outcomes of the experiment. For a discrete random variable X, the probability of an event { **X**= x} for a number x is denoted by $p(x)$.

2. The mean value or expected value of **X** is denoted by E(**X**), or μ or μ_X. For a discrete random variable, its mean value can be calculated by E(**X**)=$\sum xp(x)$
   (1) E(**X**+c)=E(**X**) +c (2) E(k**X**)=kE(**X**) (3) E(**X** + **Y**)=E(**X**) + E(**Y**)

3. The variance of **X** is denoted by Var(**X**) and can be calculated as Var(**X**)=E(**X**-μ)^2 or Var(**X**) = $\sum(x-\mu)^2 p(x)$ (discrete random variable).
   (1)Var(**X**+a)=Var(**X**) (2) Var(k**X**)=k^{2}Var(**X**)
   (3)Var(**X** +**Y**)=Var(**X**)+Var(**Y**) if **X** and **Y** are independent.
   (4)Var(**X** -**Y**)=Var(**X**)+Var(**Y**) if **X** and **Y** are independent.

4. Binomial Experiment
   (a) Each trial yields two outcomes, technically called success(S) and failure(F).
   (b) For each trial, the probability of success is the same and is denoted by p. The probability of failure is then 1-p for each trial and is denoted by q, so that p+q=1.
   (c) Trials are independent. The probability of success in a trial does not change given any information about the outcomes of other trials.

5. The Binomial Distribution
   (a) Notation :
   n = a fixed number of Bernoulli trials
   p = the probability of success in each trial
   x = the number of successes in n trials
   Then the random variable X is called a Binomial Random Variable. Its distribution is called a Binomial Distribution.
   (b) The binomial distribution with n trials and success probability p, is described by the equation

$$P[X=x] = \frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}$$

   (c) The mean and variance of X are $E(X) = np, \quad Var(X) = np(1-p)$.

6. The Normal Distribution
   (a) The normal distribution with a mean μ and a standard deviation σ is denoted by $N(\mu, \sigma^2)$.
   (b) The standard normal distribution is denoted by $N(0,1)$.
   (c) If X is normally distributed with μ and σ, then $Z = \frac{x-\mu}{\sigma}$ has the standard normal distribution.

Practice Problem

1. According to the Mendelian theory of inherited characteristics, a cross fertilization of related species of red and white flowered plants produces a generation whose offspring contain 25% red flowered plants. Suppose that a horticulturist wishes to cross 5 pairs of the cross-fertilized species. Of the resulting 5 offspring, what is the probability that
   (a) There will be no red flowered plants?
   (b) There will be 4 or more red flowered plants?

2. Consider flipping a coin. For the first flip there is a probability .5 of heads and .5 of tails. If the immediately preceding flip results in heads, the probability of heads on the coming flip is .8. If the immediately preceding flip results in tails, the probability of tails on the coming flip is .8. The coin is flipped a total of 3 times.
   (a) Let $X = \#$ of heads. Find the probability distribution of X.
   (b) Find the probability that the first flip results in heads.
   (c) Find the probability that the second flip results in heads.
   (d) Find the probability that the third flip results in heads.

3. Suppose we have observations from a $N(\mu, \sigma^2)$ distribution. Compute the following:

μ	σ	a	$\Pr(X \leq a)$	b	$Pr(X \geq b)$
10	6	??	.70	??	.32
-3	2	0	??		
33	2.5			31.81	??
-21	??	-15.12	0.90		

4. Using R,
   (a) Find $Pr(X \geq -6.2)$ if $X \sim N(-5, 8)$.
   (b) Find $Pr(-1.5 \leq X \leq 12.8)$ if $X \sim N(9, 25)$.
   (c) Find x_* so that $Pr(X \geq x_*) = .80$ if $X \sim N(12.5, 6.25)$.

5. The dry weight of organic matter in a particular tissue from soybean plants is known to be normally distributed with mean 28.2 mg and variance 8.7 mg^2.
   (a) What is the probability that the dry weight of organic matter in the particular tissue from a randomly selected soybean plant is less than 26.0 mg?
   (b) What is the value of dry weight such that 85% of the tissue have a dry weight lower than that value?
   (c) Find symmetric limits around the mean such that 90% of tissues will have dry weight between those limits.

Solution

1. Let X=the number of red flowered plants of the resulting 5 offspring, then $X \sim B(5, 0.25)$

   (a)
   $$Pr(X=0) = \frac{5!}{0!(5-0)!}(0.25)^0(1-0.25)^{5-0} = (0.75)^5 = 0.2377$$

   (b)
   $$\begin{aligned} Pr(X \geq 4) &= Pr(X=4) + Pr(X=5) \\ &= \frac{5!}{4!(5-4)!}(0.25)^4(1-0.25)^{5-4} + \frac{5!}{5!(5-5)!}(0.25)^5(1-0.25)^{5-5} \\ &= 5 \times (0.25)^4 \times 0.75 + (0.25)^5 = 0.0156 \end{aligned}$$

2. (a)
   $$\begin{aligned} P(X=0) &= P(TTT) = 0.5 \times 0.8 \times 0.8 = 0.32 \\ P(X=1) &= P(HTT \ \ or \ \ THT \ \ or \ \ TTH) = P(HTT) + P(THT) + P(TTH) \\ &= 0.5 \times 0.2 \times 0.8 + 0.5 \times 0.2 \times 0.2 + 0.5 \times 0.8 \times 0.2 = 0.18 \\ P(X=2) &= P(HHT \ \ or \ \ HTH \ \ or \ \ THH) = P(HHT) + P(HTH) + P(THH) \\ &= 0.5 \times 0.8 \times 0.2 + 0.5 \times 0.2 \times 0.2 + 0.5 \times 0.2 \times 0.8 = 0.18 \\ P(X=3) &= P(HHH) = 0.5 \times 0.8 \times 0.8 = 0.32 \end{aligned}$$

   (b)
   $$\begin{aligned} P(H_1) &= P(HHH) + P(HHT) + P(HTH) + P(HTT) \\ &= 0.5 \times 0.8 \times 0.8 + 0.5 \times 0.8 \times 0.2 + 0.5 \times 0.2 \times 0.2 + 0.5 \times 0.2 \times 0.8 = 0.5 \end{aligned}$$

   (c)
   $$\begin{aligned} P(H_2) &= P(HHH) + P(HHT) + (THH) + (THT) \\ &= 0.5 \times 0.8 \times 0.8 + 0.5 \times 0.8 \times 0.2 + 0.5 \times 0.2 \times 0.8 + 0.5 \times 0.2 \times 0.2 = 0.5 \end{aligned}$$

   (d)
   $$\begin{aligned} P(H_3) &= P(HHH) + P(HTH) + P(THH) + P(TTH) \\ &= 0.5 \times 0.8 \times 0.8 + 0.5 \times 0.2 \times 0.2 + 0.5 \times 0.2 \times 0.8 + 0.5 \times 0.8 \times 0.2 = 0.5 \end{aligned}$$

3. In the following calculations, $Z \sim N(0, 1)$

   - When $X \sim N(10, 6^2)$
   $$Pr(X \leq a) = 0.7 \implies Pr(X > a) = 0.3 \implies Pr(\frac{X-10}{6} > \frac{a-10}{6}) = 0.3$$
   $$\implies Pr(Z > \frac{a-10}{6}) = 0.3$$
   From the table ,$0.52 < \frac{a-10}{6} < 0.53$. Therefore, $13.12 < a < 13.18$.
   - When $X \sim N(10, 6^2)$
   $$Pr(X \geq b) = 0.32 \implies Pr(\frac{X-10}{6} \geq \frac{b-10}{6}) = 0.32 \implies Pr(Z \geq \frac{b-10}{6}) = 0.32$$
   From the table ,$0.46 < \frac{b-10}{6} < 0.47$. Therefore, $12.76 < a < 12.82$.

- When $X \sim N(-3, 2^2)$

$$Pr(X \leq 0) = Pr(\frac{X+3}{2} \leq \frac{3}{2}) = Pr(Z \leq 1.5) = 1 - Pr(Z > 1.5) = 1 - 0.668 = 09332$$

- When $X \sim N(33, (2.5)^2)$

$$Pr(X \geq 31.81) = Pr(\frac{X-33}{2.5} \geq \frac{31.81-33}{2.5}) = Pr(Z \geq -0.476) = 1 - Pr(Z \geq 0.476)$$

From the table, $0.6808 < Pr(X \geq 31.81) < 0.6844$.

- When $X \sim (-21, \sigma^2)$

$$P(X \leq -15.12) = 0.90 \implies P(\frac{X+21}{\sigma} \leq \frac{-15.12+21}{\sigma}) = 0.90 \implies P(Z \leq \frac{5.88}{\sigma}) = 0.90$$
$$\implies P(Z > \frac{5.88}{\sigma}) = 0.10$$

From the table, $1.28 < \frac{5.88}{\sigma} < 1.29$. Therefore,$4.56 < \sigma < 4.59$.

4. (a)

```
> pnorm(-6.2,mean=-5,sd=sqrt(8),lower.tail=F)
[1] 0.6643134
```

$Pr(X \geq -6.2) = 0.6643$

(b) $Pr(-1.5 \leq X \leq 12.8) = Pr(X \leq 12.8) - Pr(X \leq -1.5)$

```
> pnorm(12.8,mean=9,sd=5)-pnorm(-1.5,mean=9,sd=5)
[1] 0.7585083
```

$Pr(-1.5 \leq X \leq 12.8) = 0.7585$

(c)

```
> qnorm(0.80,mean=12.5,sd=2.5,lower.tail=F)
[1] 10.39595
```

$x_* = 10.396$

5. Let X=dry weight, then $X \sim N(28.2, 8.7)$

(a)

$$P(X < 26.0) = P(\frac{X-28.2}{\sqrt{8.7}} < \frac{26.0-28.2}{\sqrt{8.7}}) = P(Z < -0.7459) = P(Z > 0.7459)$$

From the table, $0.2266 < P(X < 26.0) < 0.2296$.

(b)

$$P(X < a) = 0.85 \implies P(\frac{X-28.2}{\sqrt{8.7}} < \frac{a-28.2}{\sqrt{8.7}}) = 0.85 \implies P(Z < \frac{a-28.2}{\sqrt{8.7}}) = 0.85$$
$$\implies P(Z > \frac{a-28.2}{\sqrt{8.7}}) = 0.15$$

From the table, $1.03 < \frac{a-28.2}{\sqrt{8.7}} < 1.04$. Therefore, $31.24 < a < 31.27$.

(c)

$$\begin{aligned} P(28.2 - a \leq X \leq 28.2 + a) = 0.90 &\implies P(-\frac{a}{\sqrt{8.7}} \leq \frac{X-28}{\sqrt{8.7}} \leq \frac{a}{\sqrt{8.7}} = 0.90 \\ &\implies P(-\frac{a}{\sqrt{8.7}} \leq Z \leq \frac{a}{\sqrt{8.7}}) = 0.90 \\ &\implies P(Z > \frac{a}{\sqrt{8.7}}) = (1-0.90)/2 = 0.05 \end{aligned}$$

From the table, $1.64 < \frac{a}{\sqrt{8.7}} < 1.65$. Therefore, $4.84 < a < 4.87$.